CMLS, Cell. Mol. Life Sci. 58 (2001) 403–419
1420-682X/01/030403-17 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS*21* and related elements

B. Berger and D. Haas*

Laboratoire de Biologie Microbienne, Université de Lausanne, CH-1015 Lausanne (Switzerland),
Fax +41216925635, e-mail: Dieter.Haas@lbm.unil.ch

Received 22 August 2000; returned for revision 17 October 2000; accepted 25 October 2000

Abstract. The bacterial insertion sequence IS*21* shares with many insertion sequences a two-step, reactive junction transposition pathway, for which a model is presented in this review: a reactive junction with abutted inverted repeats is first formed and subsequently integrated into the target DNA. The reactive junction occurs in IS*21*-IS*21* tandems and IS*21* minicircles. In addition, IS*21* shows a unique specialization of transposition functions. By alternative translation initiation, the transposase gene codes for two products: the transposase, capable of promoting both steps of the reactive junction pathway, and the cointegrase, which only promotes the integration of reactive junctions but with higher efficiency. This review also includes a survey of the IS*21* family and speculates on the possibility that other members present a similar transpositional specialization.

Key words. Transposase; cointegrase; insertion sequence; transposition; IS*21* family; IS*30*; IS*911*; retrovirus.

Discovery of the insertion sequence IS*21*

Transposition is defined as the translocation of a DNA fragment, the transposable element, from a donor site to a target site showing no sequence homology. Transpositional recombination events can promote different types of rearrangements such as the formation of deletions, inversions and replicon fusions [1]. When a conjugative plasmid undergoes replicon fusion with a bacterial chromosome, the chromosome can be mobilized from the origin of transfer provided by the plasmid, and conjugative chromosome transfer to recipient bacteria can occur. The ability of many conjugative plasmids to mobilize bacterial chromosomes is related to the activity of transposable elements [2]. The bacterial insertion sequence IS*21* was discovered during chromosome mobilization experiments in *Pseudomonas aeruginosa*. The broad-host-range plasmid R68.45, which promotes chromosome mobilization in *P. aeruginosa* and about 30 other bacterial species, owes this property to the presence of the tandemly repeated insertion sequence IS*21* [designated $(IS21)_2$]. The parental plasmid R68, which was first isolated from a clinical isolate of *P. aeruginosa*, carries a single copy of IS*21*. Spontaneous tandem duplication of IS*21* in R68 resulted in R68.45. Although R68.45 was obtained under laboratory conditions, there is evidence that $(IS21)_2$ can also form in nature [3–9]. R68.45 integrates into bacterial chromosomes at many different sites and thereby initiates chromosome transfer almost at random [2, 3, 9].

Transposition pathways of IS*21* and tandem formation

The transposition activities of IS*21* manifest themselves in different ways. A single copy of IS*21* can generate simple insertions, via a cut-and-paste (nonreplicative)

* Corresponding author.

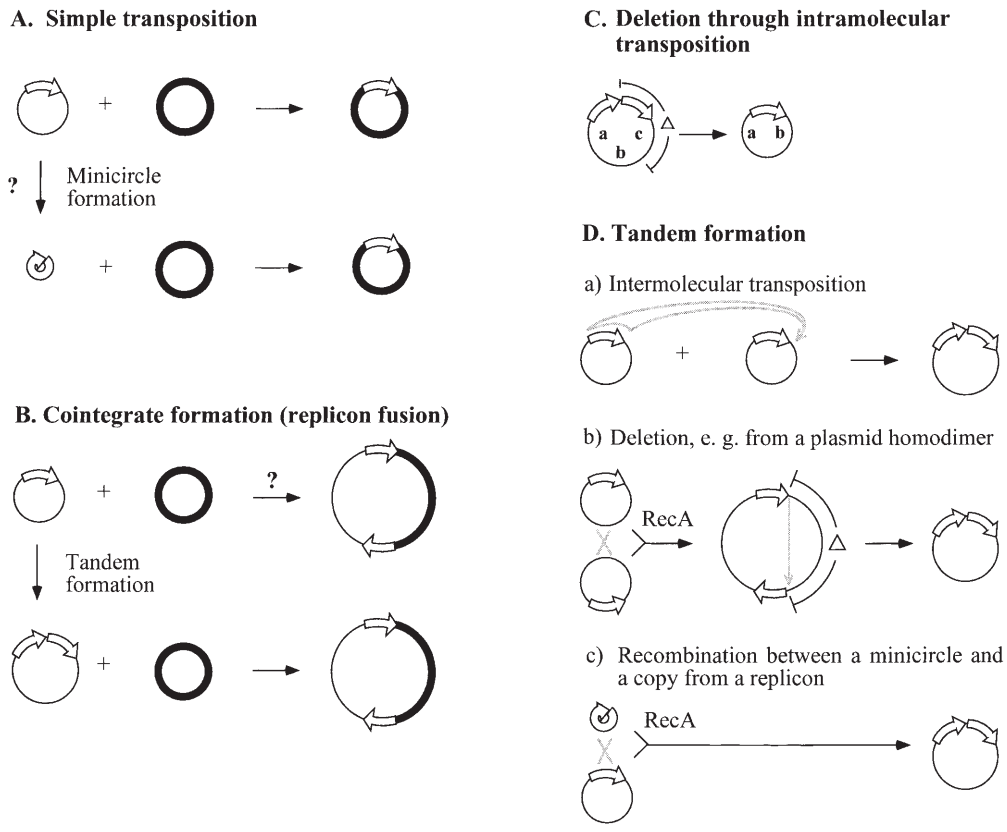


Figure 1. Overview of transposition pathways of IS*21*. Bold line, target replicon; thin line, donor replicon; Δ, deletion; arrows in grey represent site-specific recombination events; a cross stands for homologous recombination; figure modified from [39].

mechanism (fig. 1A) [10, 11]. This pathway is thought to involve circularization of IS*21* (fig. 1A, bottom part), an aspect that will be discussed in more detail at the end of this review. A single copy of IS*21* can also yield cointegrates at low frequency, probably through spontaneous tandem duplication of the element in the donor replicon before fusion with the target replicon; the resulting cointegrates show a copy of IS*21* in the same orientation at both junctions of the fused replicons (fig. 1B, upper part). By contrast, cointegrate formation (replicon fusion) between a replicon carrying $(IS21)_2$ and a target replicon is very efficient and also follows a cut-and-paste mechanism (fig. 1B, lower part) [10, 12]. The two or three nucleotides forming the IS*21*-IS*21* junction in the tandem are lost during the reaction [12, 13]. Both simple insertion and replicon fusion (fig. 1A, B) generate target duplications of 4 bp, rarely 5, 6 or 7 bp [13–15]. As in other transposition systems, the target duplications flanking the insert are formed by repair enzymes of the host [1, 16].

Although an $(IS21)_2$ structure is quite stable even in a $RecA^+$ host [8], deletions of one element can occur sometimes, and flanking DNA can be deleted concomitantly (fig. 1C), especially after conjugative transfer of an $(IS21)_2$ plasmid into a new host [8]. These deletion products are probably the consequences of intramolecular transposition of IS*21*.

Several models, which are not mutually exclusive, may explain IS*21* tandem formation. Simple intermolecular transposition of an IS*21* copy near one end of another IS*21* element is one explanation (fig. 1Da). Alternatively, an IS*21* element could transpose next to the copy situated on the other branch in a replication fork. Both mechanisms would involve site-specific recognition between two IS*21* termini [10, 17]. An IS*21* tandem can also result from a spontaneous deletion of a segment situated between two IS*21* copies present in direct orientation [10]. One way to generate a plasmid with two directly oriented IS elements is by RecA-dependent homodimerization (fig. 1Db). The deletion between two IS*21* elements would be a site-specific event as well. A further possibility leading to an $(IS21)_2$ structure would be a RecA-dependent recombination between an IS*21* copy carried by a replicon and an IS*21* minicircle (fig. 1Dc). At present, experimental evidence is available only for the deletion pathway (fig. 1Db) [10]. The other pathways (fig. 1Dac) have not been investigated because an appropriate selection system is lacking. In *P. aeruginosa*, the frequency of $(IS21)_2$ formation in R68 has been estimated to be $\leq 10^{-4}$ [2]. It emerges from the three proposed pathways of tan-

dem formation that in each case, at one time or another, two IS*21* ends come together in a site-specific recombination event.

Genetic organization of IS*21*

IS*21*, with 2131 bp, is among the largest insertion sequences. Its termini show inverted repeats (IRR and IRL) of 11 bp with one mismatch (fig. 2B, upper part). However, additional, less conserved sequence elements, termed multiple terminal repeats (MTRs), occur at both IS*21* termini (fig. 2AB) [15]. The consensus sequence for the MTRs of IS*21* is YRCCANYNNNRTNNNNCNNT, where Y and R represent pyrimidines and purines, respectively. Three MTRs are found at the left end, whereas two MTRs occur at the right end of IS*21* (fig. 2B, upper part). A potential integration host factor (IHF) recognition site is within MTR-L3 (fig. 2B, upper part).

When two IS*21* elements are organized in tandem, they are separated by a junction sequence of two or three variable nucleotides [11, 13]. The abutted IS*21* ends form a σ^{70} promoter whose −35 and −10 hexamers are located within IRR and MTR-L1, respectively (fig. 2B, upper part). This promoter drives the expression of the *istAB* genes in the downstream IS*21* element [13].

IS*21* contains two open reading frames (ORFs), the *istAB* genes, which are organized in an operon (fig. 2B). Both genes are necessary for efficient and accurate transposition activity [12, 13, 18]. They encode transposition and helper proteins, respectively, whose functions will be discussed below. As other transposable elements, IS*21* contains some additional short ORFs on both strands [19], but no product has been associated with them so far. The stop codon of *istA* and the start codon of *istB* overlap with one nucleotide in a stop-start configuration (fig. 2B), suggesting that expression of the *istA* and *istB* genes may involve translational coupling. By contrast, it seems highly unlikely that transposition of IS*21* depends on an IstAB fusion protein (produced by translational frameshifting), because separate expression of *istA* and *istB* does not affect the transpositional activity [11].

The function of the MTRs of IS*21* has not been assessed. However, in some other transposable elements such as bacteriophage Mu, Tn*7* and Tn*552*, MTRs are involved in the recognition of transposase and in the assembly of multimeric forms of this enzyme [20–25]. In the case of IS*21*, we note that the spacing of about 23 bp between the MTRs is such that the conserved sequence elements in the MTRs will be exposed to the same surface of the supercoiled DNA double helix [26, 27].

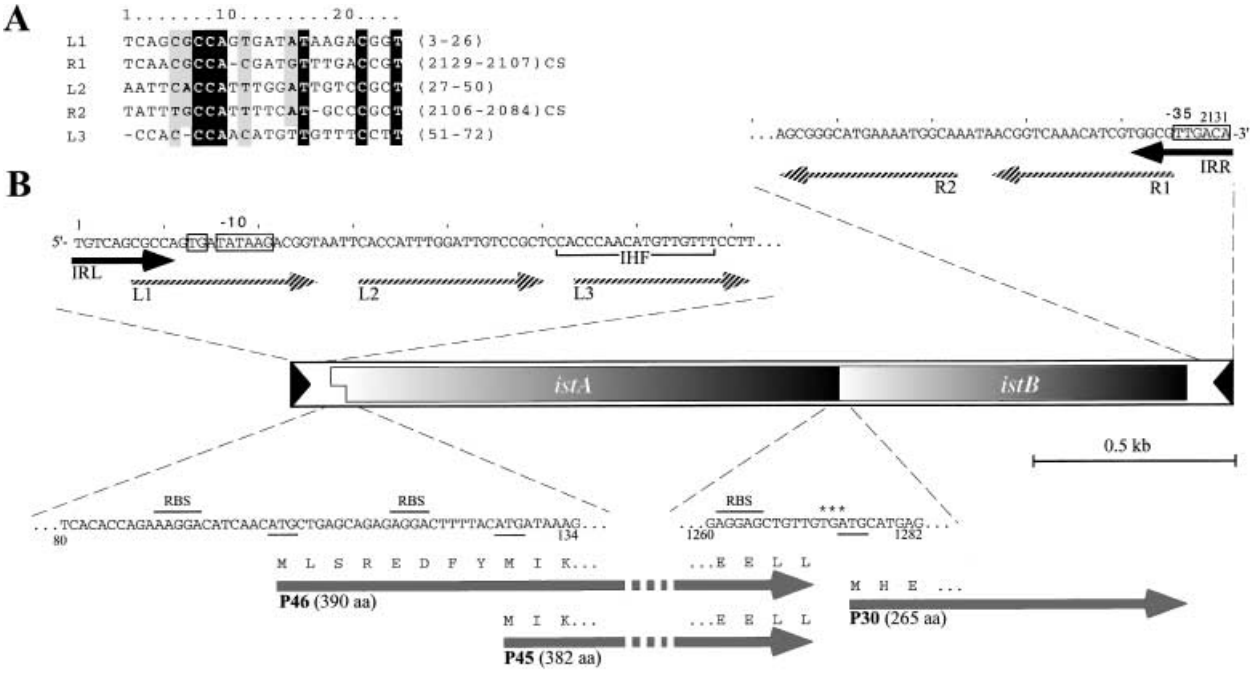


Figure 2. Insertion sequence IS*21*. (**A**) Alignment of the multiple terminal repeats (MTRs) together with their coordinates; CS, complementary strand; nucleotides outlined in black, identity; nucleotides outlined in grey, pyrimidine or purine; L3 is slightly less conserved and therefore not considered in the determination of the motif. (**B**) Upper part: termini of IS*21*; IRL and IRR, inverted repeats left and right; L1, L2, L3, R1 and R2, multiple terminal repeats; IHF, sequence resembling the consensus sequence for integration host factor binding; −10 and −35, sequences forming the promoter of the IR-IR junction. At the center: i*stA* and *istB*, the ORFs of IS*21*; black triangles represent the IRs. Lower part: products of the *istAB* genes; P46, the transposase coded by *istA*; P45, the cointegrase coded by *istA*; P30, the helper protein coded by *istB*; RBS, ribosome binding site; the start codons ATG are underlined; ***, stop codon.

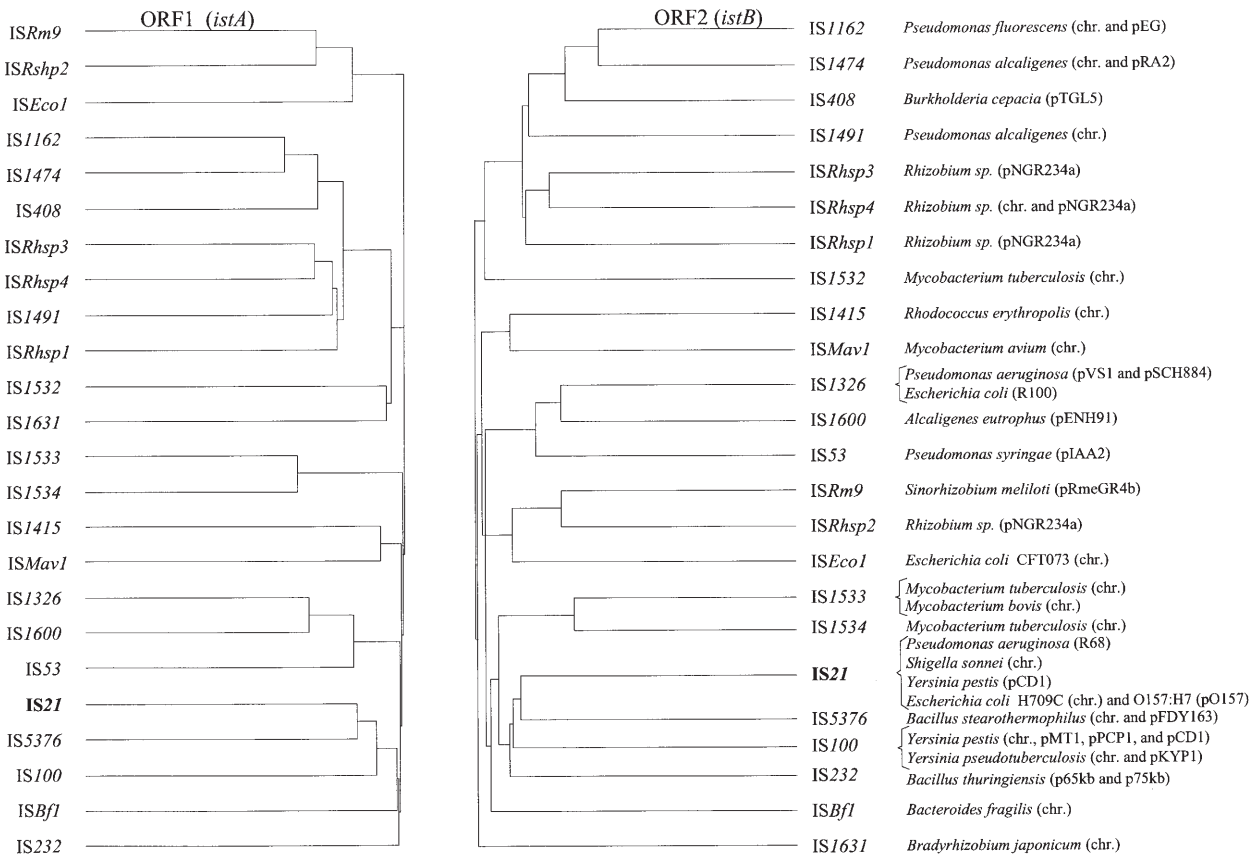


Figure 3. The IS*21* family. Dendrogram calculated by the Pileup program (GCG, University of Wisconsin, Madison) for both ORFs; the names of the bacteria refer to the organisms from which the elements have been isolated as a chromosomal copy (chr.) or a plasmid copy (name of the plasmid). The dendrograms obtained by Pileup represent the most similar pairs. The use of this approach is justified by the fact that the presence of divergent regions will not mask the close relatedness of the remainder of the sequences compared, especially of regions containing the active site.

When an IS*21*-IS*21* junction of a linearized plasmid was incubated with overexpressed IstA in a crude extract, about 30 nucleotides from both IS*21* ends were protected from nuclease attack [12], suggesting that MTR-L1 and MTR-R1 may interact with IstA during the replicon fusion reaction. If this interpretation is correct, then MTR-L1 might have an additional role in autoregulation of the *istAB* genes: IstA protein could block transcription by binding to MTR-L1 containing the –10 promoter sequence (fig. 2B).

A unique specialization of transposition functions in IS*21*: transposase and cointegrase

The *istA* gene is essential for both simple insertion and replicon fusion [11]. Two ATG start codons associated with ribosome binding sites (RBS) enable the *istA* gene to generate two products in the same reading frame, IstA(P46) and IstA(P45). These proteins of 46 kDa and 45 kDa, respectively, differ in their N-termini by eight amino acids (fig. 2B, lower part). Like transposases of several transposable elements, P46 and P45 show a helix-turn-helix motif at their N-termini (fig. 5A), probably promoting binding to DNA [13, 28]. The deduced amino acid sequences of IstA(P46) and IstA(P45) show a typical DDE motif (aspartate-122, aspartate-184, glutamate-230 in P46, cf. fig. 5A), which is part of the catalytic domain of retroviral integrases and numerous bacterial transposases [28, 29]. P46 and P45 have distinct functions. These were demonstrated with *istA* constructs that were engineered to express either IstA(P46) alone or IstA(P45) alone [11]. After overexpression in *Escherichia coli*, P45 promotes replicon fusion (cointegrate formation) at very high frequencies (10^{-1} to 10^{-2}) when the substrate is a reactive junction of two abutted IS*21* ends on a replicon and the target is a multicopy plasmid (e.g. fig. 1B, lower part) [11]. The optimal size of the IR-IR junction sequence is 2, 3 or 4 bp. With shorter or longer junction sequences (constructed artificially), the efficiency of replicon fusion decreases [11]. IS*21* ends joined via an artificial 4-bp (TATA) junction are slightly better substrates than those linked by a natural 2-bp (GT) or 3-bp (ACG) junction [11]. Because of this efficient enzymatic function in coin-

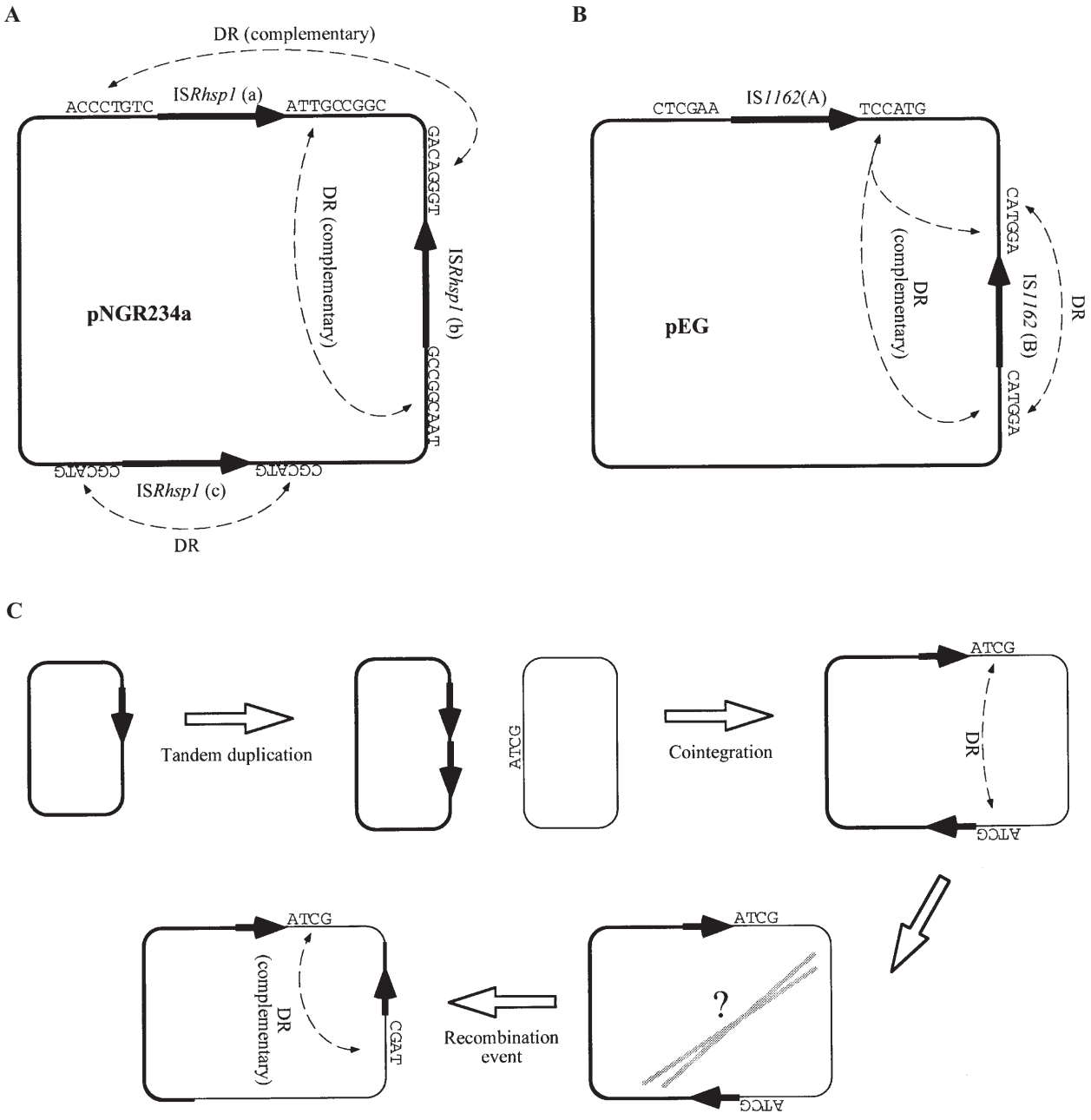
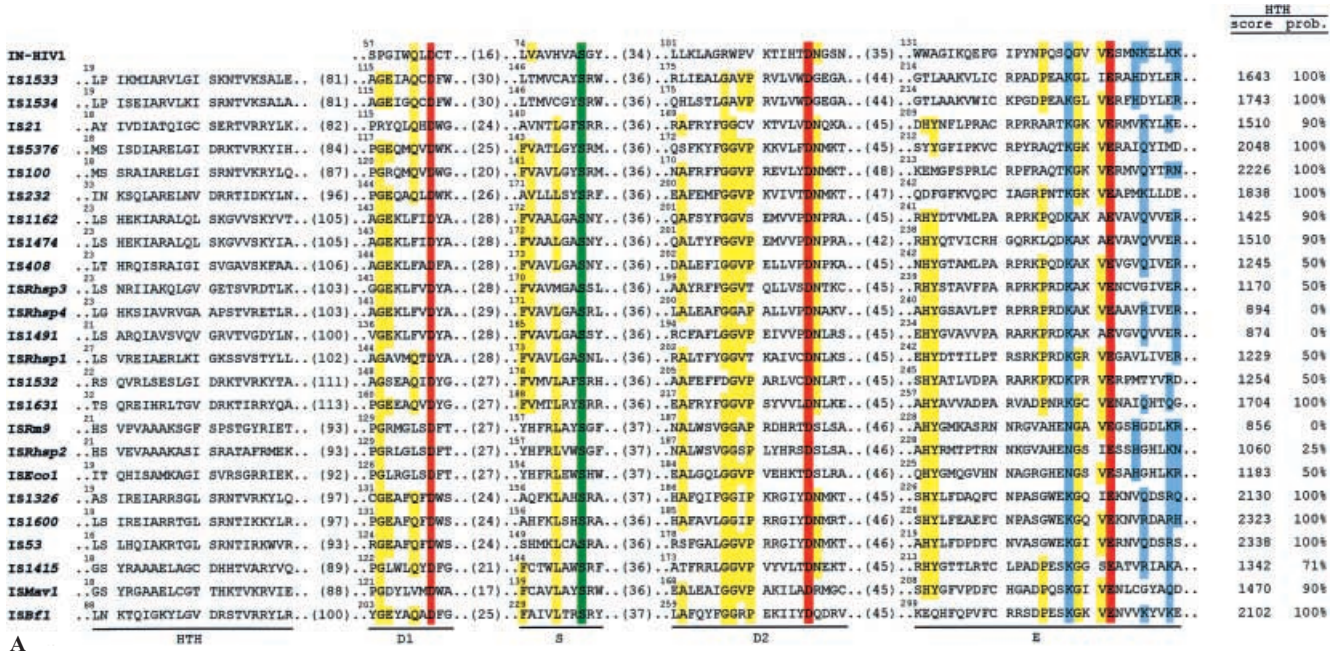


Figure 4. Cointegrate-like structures in the IS*21* family. (**A**) Localization and direct repeats of copies of IS*Rhsp1* on pNGR234a [66]. (**B**) Localization and direct repeats of copies of IS*1162* on pEG [58]. (**C**) Proposed molecular mechanisms to explain the complementary direct repeats. The insertion sequences are represented by arrows (leading from IRL to IRR); the sequences of the direct repeats (DR) are shown on the same strand; in the proposed model the tetranucleotide of the target site shown in the proposed model is arbitrary; ?, recombination event inverting a DNA segment.

tegrate formation, P45 has been named cointegrase. By contrast, when the substrate is composed of two distant IS*21* ends as in a single IS*21* element on a replicon, cointegrase has poor activity: simple insertions (fig. 1A) are formed at $\leq 10^{-6}$ by this enzyme [11]. P46 displays an equal and intermediate activity (ca. 10^{-3}) in both the simple transposition and the replicon fusion pathways (figs. 1A and 1B, lower part) [11]. For this reason, P46 has been designated transposase.

Mechanistically, the cleavage and strand transfer reactions catalyzed by cointegrase and transposase are probably very similar; we will discuss them at the end of this review. What then might be the role of the extra eight amino acids (MLSERDFY) at the N-terminus of trans-



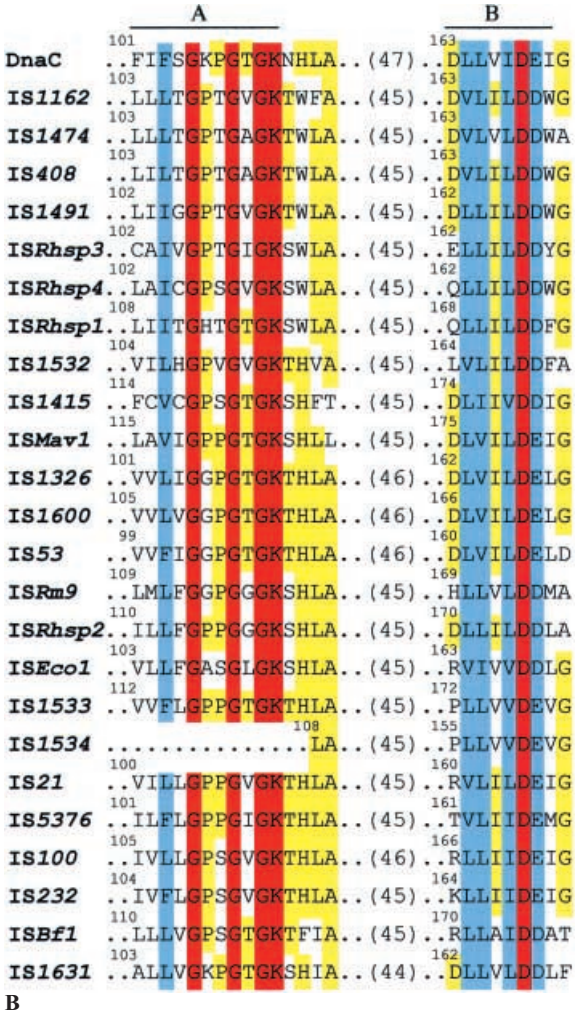


Figure 5. Alignments of the ORFs of the IS*21* family. (**A**) The amino acids of the DDE region of the ORF1 (*istA*) of the IS*21* family members and of IN-HIV1 [87]; HTH, helix-turn-helix motif; D1, D2 and D3, regions including the DDE triad; S, region including a well-conserved serine (see text); on the right, the score of the HTH motif calculated according to [86] and the probability of existence are shown; even if the scores obtained for IS*Rsp4* and IS*1491* predict no HTH motif, the analysis of these sequences by the program hthscan (GCG, University of Wisconsin, Madison) identifies them as probable HTH; the numbers in brackets represent the number of residues between the HTH motif and the first D, the first D and the S, the S and the second D, the second D and the E, respectively; the numbers above the sequences are the coordinates; yellow, residues conserved in the majority of cases; red, residues strictly conserved belonging to the DDE triad; green, serine strictly conserved; blue, basic residues neighboring the E. (**B**) Motifs of the ORF2 (*istB*) of the IS*21* family members and of DnaC from *E. coli* [32]; the numbers above the sequences are the coordinates; in brackets are the number of residues not represented; red, strictly conserved residues; blue, conserved similar residues; yellow, residues conserved in the majority of cases; the A domain of IS*1534* is deleted.

posase? These residues may help the transposase to bring two IS*21* ends close to each other, such that they can form an IR-IR junction which is subsequently inserted into the target DNA. The N-terminal residues of P46 and P45 might also influence the accessibility of the putative transposition donor complex to the target DNA.

Both the simple transposition (fig. 1A) and the replicon fusion (fig. 1B) pathways need a helper protein, IstB, encoded by IS*21*. This 30-kDa protein appears to have a function similar to that of MuB of bacteriophage Mu and TnsC of transposon Tn*7* [30, 31]. IstB has an ATP/GTP binding motif which is also found in DnaC [32] (cf. fig. 5B). The absence of IstB during replicon fusion drastically reduces the efficiency of the reaction and generates atypical target duplications and/or deletions in the target. Thus, IstB is needed for accurate strand transfer and capture of the target DNA [18]. Under optimal conditions the cointegration reaction proceeds at a frequency of 10^{-1} without being lethal to the *E. coli* host [11]. It is likely that this frequency represents an initial burst of activity and that a mechanism exists which prevents further insertions of IS*21* into the same target DNA, i.e. an immunity mechanism exists. Circumstantial evidence for immunity against multiple IS*21* insertion has been reported [17] and confirmed in our laboratory, but the mechanism involved has not been elucidated. By analogy with MuB of phage Mu and TnsC of Tn*7*, IstB might have a role in establishing immunity [23, 33–36].

Target selection of IS*21* in vivo and in vitro

The high replicon fusion activity of cointegrase, assisted by IstB, has been exploited to generate an in vitro system in which a reactive IS*21*-IS*21* junction is carried by a suicide plasmid [18]. Crude *E. coli* cell extracts containing overproduced cointegrase and IstB catalyze replicon fusion between the suicide construct and a multicopy target replicon at a frequency of ca. 10^{-3} (per target replicon) [15, 18].

The target specificity of IS*21* has been investigated by an analysis of replicon fusion products obtained in vivo and in vitro with IstA(P45) and IstB. More than 100 insertion sites have been sequenced [14, 15]. No consensus target sequence has become apparent, in contrast to an earlier proposal based on a small number of insertion sites [37]. However, the IS*21* transposition machinery appears to prefer targets displaying an 11-bp periodicity of the dinucleotide AA flanking the insertion sites. This sequence organization, which may confer anisotropic flexibility to the target DNA, suggests that insertions at the center of S-shaped DNA would be favored [B. Berger, T. Junier, D. Haas and A. Stasiak, unpublished results]. A more pronounced target preference showing an intrinsic S-shape has been revealed for IS*231A*, an element belonging to the IS*3* family [38].

A survey of the IS*21* family

IS*21* is the prototype of a growing family of bacterial insertion sequences. Since publication of previous reviews [39, 40], several new members have been added to the IS*21* family. Our first objective here is to describe the current status of this family.

Members of the IS*21* family are characterized by two *istA*- and *istB*-like ORFs, by terminal IRs of usually up to 50 bp which are not conserved within the family but generally end with CA, by the presence of MTRs of 15–20 bp at both ends, and by target duplications (direct repeats) of 4–9 nucleotides. These elements are listed in table 1. Elements whose amino acid sequences of the ORFs diverge by less than 5% are defined as isoforms. When dissimilarity is obviously the result of translation frameshifts (due to genuine mutations or possibly due to sequencing errors), nucleotide sequences showing less

Table 1. Origins of the insertion sequences of the IS*21* family.

IS element[a]	location[b]	Organism	Reference/ accession no.[c]	DR[d] (bp)	Comments on transpositione[e]
IS*21*	R68	*Pseudomonas aeruginosa*	[13] X14793	4 (5, 6, 7)	simple transposition,
	pO157	*E. coli* O157:H7	[41] AF74613 (T)		tandem formation,
	chr	*E. coli* H709C	[42]		cointegrate formation
	pB171	*E. coli* B171	[43] AB024946 (T)		(see text)
	chr.	*Salmonella enterica*	AJ242964 (T)		
	chr.	*Salmonella typhi*	AB029403 (T)		
	RK2	diverse	[44]		
	RP1				
(IS*8*)	RP4				
(IS*640*)	pCD1	*Yersinia pestis*	[45] AF053946 (T)		
(IS*21p*)	pCD1	*Yersinia pestis*	[46] AF074612 (T)		
(IS*640*)	chr.	*Shigella sonnei*	[47] X05956 (P)		
IS*53*	pIAA2	*Pseudomonas syringae*	[48] M83932	8	–

Table 1 (continued)

IS element[a]	Location[b]	Organism	Reference/ accession no.[c]	DR[d] (bp)	Comments on transposition[e]
IS*100*	chr.	*Yersinia pestis*	[49] AL031866	5	transposition [53]
	pMT1	*Yersinia pestis*	[45] AF053947 [50] AF074611		
	pPCP1	*Yersinia pestis*	[51] Z32853 [45] AF053945		
	pCD1	*Yersinia pestis*	[52] X78302 [45] AF053946 [46] AF074612		
	?	*Yersinia pestis*	L19030		
	chr.	*Yersinia pseudotuberculosis*	AJ236887		
	pKYP1	*Yersinia pseudotuberculosis*	[54] U59875		
	pB171	*E. coli* B171	[43] AB024946 (T)		
IS*232*	p65kb	*Bacillus thuringiensis*	[55] M38370	8, 6	simple transposition [55]
	p75kb	*Bacillus thuringiensis*	M77344		
IS*408*	pTGL5	*Burkholderia cepacia*	[56] L09108	8	transposition [57]
IS*1162*	pEG	*Pseudomonas fluorescens*	[58] X79443	6, 7	simple transposition [58]
	chr.	*Pseudomonas fluorescens*	[58]		replicon fusion without duplication of the element [58] cointegrate-like structure observed (see text)
IS*1326*	pVS1	*Pseudomonas aeruginosa*	[59] U38187	ND	–
	pSCH884	*Pseudomonas aeruginosa*	[59]		
	R100	*Escherichia coli*	AP000342		
IS*1415*	chr.	*Rhodococcus erythropolis*	[60] AF002247	5, 6	simple transposition [60]
IS1*474*	chr.	*Pseudomonas alcaligenes*	[61] U67315	4	replicon fusion, cointegrate formation with IS*1475*? [61]
IS*1475*	chr.	*Pseudomonas alcaligenes*	[61] U67315 (T)	ND	replicon fusion, cointegrate formation with IS*1474*? [61]
IS*1491*	chr.	*Pseudomonas alcaligenes*	[62] U84154	ND	simple transposition [62] replicon fusion, cointegrate formation? [62]
IS*1532*	chr.	*Mycobacterium tuberculosis*	[63] Z95389 to Z77165	ND	–
IS*1533*	chr.	*Mycobacterium tuberculosis*	[63] Z83858	5	–
	chr.	*Mycobacterium bovis*	[64] AJ238712		
IS*1534*	chr.	*Mycobacterium tuberculosis*	[63] Z95436	5	defective by deletion in ORF2?
IS*1546*	pRE4	*Pseudomonas putida*	[65]	ND	–
IS1600	pENH91	*Alcaligenes eutrophus*	[66, 67] D64144	ND	cointegrate-like structure observed, except for the DR [66]
IS*1631*	chr.	*Bradyrhizobium japonicum*	[68] AB011*021* (P)	ND	–
IS*5376*	pFDY163	*Bacillus stearothermophilus*	[69] X67861	5	transposition [69, 70]
IS*Atu1*	pTiA6NV	*Agrobacterium tumefaciens*	AF034854 (T)	ND	–
IS*Bf1*	chr.	*Bacteroides fragilis*	[71] U05888	ND	–
IS*Bf2*	chr.	*Bacteroides fragilis*	[71] U05886 (P)	ND	–
IS*Che1*	chr.	*Chelatobacter heintzii*	[72] L49438 (P)	ND	–
IS*Cpi1*	chr.	*Carnobacterium piscicola*	[73] L471*21* (T)	ND	–
IS*Eco1*	chr.	*Escherichia coli* CFT073	[74] AF081285	5	–
IS*Lla1*	chr.	*Lactococcus lactis*	AF064765 (P)	ND	–
IS*Mav1*	chr.	*Mycobacterium avium*	[75] AF125999 A63806	5	–
IS*Pae1*	chr.	*Pseudomonas aeruginosa*	AF087482 (T)	ND	–
IS*Psesp1*	chr.	*Pseudomonas* sp.	[67] AB019033 (T)	ND	–
IS*Rm9*	pRmeGR4b	*Sinorhizobium meliloti*	[76] Y13432	7	–

Table 1 (continued)

IS element[a]	Location[b]	Organism	Reference/ accession no.[c]	DR[d] (bp)	Comments on transposition[e]
IS*Rhsp1*	pNGR234a	*Rhizobium* sp.	[77] AE000081 AE000066 AE000097	6, 8, 9 (see text)	cointegrate-like structure observed (see text)
IS*Rhsp2*	pNGR234a	*Rhizobium* sp.	[77] AE000079 AE000086 AE000095	6, 7	–
IS*Rhsp3*	pNGR234a	*Rhizobium* sp.	[77] AE000099	8	–
IS*Rhsp4*	chr. pNGR234a	*Rhizobium* sp. *Rhizobium* sp.	[77] AE000090 [78]	7, 8, 9	defective by insertion of other IS elements?
IS*Sco1*	chr.	*Streptomyces coelicolor*	AL023861 (T?)	ND	–

[a] Elements shown in brackets are isoforms. IS*8* and IS*21* are probably identical.
[b] Location indicates plasmids or chromosomes (chr.).
[c] (P), partial DNA sequence; (T), truncated copy.
[d] DR, direct repeat = target duplication; ND, not determined.
[e] Experimental evidence established for the transpositional movements indicated.

than 10% divergence are considered as isoforms. Some of the elements of table 1 are truncated, incompletely sequenced or defective.

As the transposition activity of many members of the IS*21* family has not been tested (table 1), it is quite difficult to distinguish between sequencing errors and real mutations which could affect transposition activity. We are aware that the sequence corrections introduced by us (indicated by 'M' in table 2) favor the phylogeny to function.

In the elements listed in table 2, ORF1 (340–585 amino acid residues) is always longer than ORF2 (245–270 amino acid residues). The ORFs are frequently in an overlapping stop-start or start-stop configuration, arrangements that may allow translational coupling. In a few elements, even a superimposition of up to 115 bp has been reported. In no case does there seem to be a motif involved in a programmed frameshift between ORF1 and ORF2.

At both ends of the elements there are imperfect inverted repeats, which include two to seven variably conserved MTRs of about 20 bp each, with a spacing of generally 23–24 bp and occasionally 33–34 or 46–48 bp, i.e. about multiples of 11.3 bp, the supercoiled DNA double helix turn [26, 27]. The members of the IS*21* family show a CA dinucleotide or exceptionally a TA dinucleotide at the termini. The terminal CA exists in many bacterial elements which code for transposases carrying a DDE motif, and also at the termini of the integrated forms of retroviruses [28, 29]. In the case of IS*21*, a CA to GA transversion affects cointegrate formation by a factor of about 50, and a CA to CT transversion reduces activity to a nonmeasurable level [79]. Thus, the two ends observed in the IS*21* family differ from one another only in the penultimate, less important nucleotide. The dinucleotides CA and TA are both 'kinkable' [80], and this might facilitate cleavage at their 3′ end by integrase, transposase or cointegrase.

The IS*21*-IS*21* junction of R68.45 forms a strong promoter whose activity was demonstrated by an *istA-galK* fusion in the downstream element [13]. Such promoter sequences, discovered in other IS families as well [81], are expected for other members of the IS*21* family, e.g. for IS*1415* and IS*5376*. Potential promoters can also be identified upstream of ORF1, e.g. in IS*1326* and IS*53*, but they are less canonical. Whether such promoters exist in reality is uncertain because for no member of the IS*21* family, except for IS*21* itself, has a reactive IR-IR junction been described.

An alignment of ORF1 and ORF2 proteins of the IS*21* family was performed using the Pileup program (fig. 3). For both ORFs, similar subgroups of related elements appear, for instance IS*21* with IS*5376* and IS*100*, or IS*1326* with IS*1600* and IS*53*. The proximity of IS*21* (from various Gram-negative bacteria) and IS*5376* (from a Gram-positive bacterium) in the dendrogram suggests their mobility, probably via plasmid transfer. Another similar example is provided by the closely related elements IS*1532* (found in *Mycobacterium tuberculosis*) and IS*1631* (found in *Bradyrhizobium japonicum*) (fig. 3). On the other hand, IS elements from the same organism (e.g. IS*1532* and IS*1533* from *M. tuberculosis*) need not be closely related to each other.

Data on the transposition activities of IS*21* family members are scarce (see references in table 1). Apart from IS*21*, only IS*1162* has been shown to give simple insertion and replicon fusion, the latter without apparent duplication of the element. Simple insertion of some elements, e.g. IS*232* and IS*1491*, and sometimes replicon fusion, e.g. of IS*1474* and IS*1491*, have been observed, but repli-

Table 2. Properties of the elements of the IS*21* family.

Element	DNA [bp]	ORF1 [aa][a]	ORF2 [aa][a]	Alternative start in ORF1[b]	Intergenic ORF1-ORF2 structure[c]	Multiple terminal repeats[d]		Termini[d]	
						Left	Right	Left	Right
IS*21*	2131	390	265	ATG/ATG (8)	Stop-Start TGATG	1 2 3	1 2	CA	CA
IS*53*	2555	496 M	267 M	ATG/ATG (9)	superposition 8 bp	1 2	1 2	CA	CA
IS*100*	1954	340	259	ATG/ATG (7)	Stop-Start TGATG	1 2 3	1 2	CA	CA
IS*232*	2183	431	250	absence	superposition 8 bp	1 2 3	1 2	TA	CA
IS*408*	>2530	518	>231	ATG/ATG (7)	spacing 16 bp	1 2 3	1 2	CA	CA
IS*1162*	2634	558	249 M	ATG/ATG (7)	superposition 115 bp	1 2 3 4 5 6	12	CA	CA
IS*1326*	2470	507	260	GTG/GTG (3)	superposition 11 bp	1 2 3	1 2	CA	CA
IS*1415*	2580	513	263	ATG/ATG (7)	Stop-Start TGATG	1 2 3 4	1 2 3 4	CA	CA
IS*1474*	2633	558	249	ATG/ATG (7)	superposition 113 bp	1 2 3 4 5 6 7	1 2	CA	CA
IS*1491*	>2489	509	251	GTG/ATG (5)	spacing 17 bp	present, but the sequence of the element is apparently partial		?	?
IS*1532*	2645	511 M	251	ATG/ATG (7)	spacing 97 bp	1 2 3	1 2	CA	CA
IS*1533*	2213	413	266 M	absence	Stop-Start TGATG	1 2 3	1 2 3	CA	CA
IS*1534*	2130	411 M	248	absence	Stop-Start TGATG	1 2	1 2	CA	CA
IS*1600*	2520	518	264	GTG/GTG (3)	superposition 11 bp	1 2	1 2	CA	CA
IS*1631*	>2712	585	255	ATG/GTG (11)	spacing 4 bp	present, but the sequence of the element is apparently partial		?	?
IS*5376*	2107	400	251	ATG/ATG (8)	Start-Stop ATGA	1 2	1 2	CA	CA
IS*Bf1*	2787	582	263	long N-terminal extension	superposition 32 bp	?	?	CA	CA
IS*Eco1*	2407	499 M	250	absence	Start-Stop GTGA	1 2	1 2	TA	TA
IS*Mav1*	2536	500 M	264	ATG/ATG (7)	Start-Stop ATGA	1 2 3	1 2	CA	CA
IS*Rm9*	2798	502	270	absence	Start-Stop ATGA	1 2 3 4	1 2 3	CA	TA
IS*Rhsp1*	2623	516	263	ATG/ATG (11)	Start-Stop GTGA	1 2 3 4	1 2 3 4	CA	CA
IS*Rhsp2*	2641	504	298	absence	Start-Stop ATGA	1 2	1 2 3	CA	TA
IS*Rhsp3*	2786	514	248	TTG/ATG (7)	spacing 11 bp	1 2 3 4 5	1 2 3	CA	CA
IS*Rhsp4*	2599	516 M	245	ATG/ATG (7)	spacing 11 bp	1 2 3 4	1 2 3	CA	CA

[a] Where modifications of the nucleotide sequence were introduced, "M" is added under the size of the ORF; aa, amino acid residues.
[b] Two start codons are given with the size difference (in amino acids residues) between the putative transposases and cointegrases.
[c] Description of the region including the stop of ORF1 and the start of ORF2.
[d] The numbers of the multiple terminal repeats are given only when the sequence of the IS termini is complete; ?, not determined.

con fusion products have not been examined. The proximity of some copies of IS*100* in *Yersinia pestis* or IS*1600* in *Alcaligenes eutrophus* suggests transposition of composite elements or cointegration, but the direct repeats normally generated by such mechanisms are not apparent. By contrast, the arrangement of IS*Rhsp1* and IS*1162* copies on plasmids pNGR234a and pEG, respectively, could be due to a cointegration mechanism. However, in both cases, the transposition event producing the cointegrates is probably followed by a recombination event inverting a segment carrying one element (fig. 4). These examples illustrate that simple transpositions and replicon fusions can occur among IS*21* family members, but these processes have mostly been inferred, not directly observed.

Does a cointegrase/transposase specialization exist in other members of the IS*21* family?

Here we analyze the genetic organization of the 5′ end of ORF1 in various elements, using IS*21* as a model. The *istA* gene of IS*21* uses two alternative start sites located 24 bp apart to produce transposase and its truncated form, cointegrase (fig. 2B) [11]. The difficulty in determining in silico whether alternative start sites exist for ORF1 arises from the fact that it is difficult to predict the effectiveness of potential ribosome binding sites and, moreover, in different organisms [82–84]. Furthermore, alternative start codons (GUG or UUG instead of AUG) may be used with various degrees of preference in different bacterial species: in *E. coli* AUG (83%) > GUG (14%) > UUG (3%); in *B. subtilis* AUG (78%) > UUG (13%) > GUG (9%) and in *M. tuberculosis* AUG (61%) > GUG (33%) >UUG (5%) [85].

In the absence of stringent criteria, inspection of ORF1 sequences suggests that most members of the IS*21* family probably use two alternative translation start sites, spaced 3 to 11 codons apart (table 2). The elements which are devoid of potential alternative sites of translation initiation for ORF1 represent a minority; two of them occur in *Mycobacterium spp.* (IS*1533* and IS*1534*) and four of them show at least one TA terminus (IS*232*, IS*Eco1*, IS*Rm9* and IS*Rhsp2*). The eight amino acid residues truncated in IS*21* cointegrase are mostly polar (fig. 2B). A similar preference is also found in ORF1 of IS*53*, IS*1415*, IS*1491*, IS*1631*, IS*Mav1* and IS*Rhsp1*, but other elements do not conform to this rule. Thus, the polarity observed at the N-terminus of IS*21* transposase could be fortuitous.

Domains of ORF1 and ORF2 proteins in the IS*21* family

For both *istA* products of IS*21* and for all ORF1 products of the members of the IS*21* family (except IS*Rm9*), a helix-turn-helix motif can be deduced in the N-terminal part (fig. 5A). Such a motif is predicted to provide DNA binding activity [86], which is consistent with the observation that numerous bacterial transposases have a DNA binding domain in their N-terminal part [28]. In the C-terminal part, few residues are strictly conserved among all elements of the family (a global alignment not shown, but see [39]).

The catalytic domain highlighted by the DDE triad, which is present in numerous bacterial transposases and retroviral integrases [28, 29, 87], is conserved in ORF1 of the IS*21* family (fig. 5A). A serine residue, which is situated halfway between the two aspartates in retroviral integrases (IN) such as IN HIV-1, is also conserved in the IS*21* family (fig. 5A). Thus, four regions can be defined (fig. 5A), one (S) including the conserved serine residue and the three others (D1, D2 and E) including the DDE triad and approximately corresponding to the regions referred to as N2, N3 and C1 by other authors [29, 88].

The importance of the DDE residues has been confirmed by directed mutagenesis of several retroviral integrases (for a review see [89]) and of transposases from Mu [90], IS*10* [91], IS*911* [29] and Tn*7* [92]. In the case of IN, mutations of D64, D116 or E152 abolish processing and strand transfer as well as disintegration, an enzymatic activity whose biological significance is not well defined and which dissociates integrated viral DNA from the target. In bacterial transposases, donor cleavage (equivalent to processing) and strand transfer are equally affected by mutations in the DDE triad, which is the catalytic core domain of transposases and integrases through coordination of essential divalent metal ions (especially Mg^{2+}) [29].

Although the conserved serine probably is not part of the catalytic site, S81A or S81G mutations in IN seriously impair the activity of the enzyme, probably by affecting folding [93, 94]. Structural analyses of IN of HIV-1 and MuA transposase show that this serine and its threonine homolog in MuA form part of a junction between two β sheets [95, 96]. However, it has been shown for Tn*7* that the analogous S304A mutation in TnsB (forming a heterodimeric transposase with TnsA) has no effect on transposition [92, 97].

Other residues of the regions D1, D2 and E are relatively well conserved in the IS*21* family (fig. 5A), and one can find hypothetical functions based on studies of IN. The Q62 residue of IN HIV-1, which interacts with the penultimate nucleotide C at the viral 5′ end [98], is present in about half of the elements of the family (fig. 5A). Residue Q148 of IN HIV-1, which also recognizes the 5′ end [99], has a lysine counterpart in most elements of the family (fig. 5A). Downstream from E152 of IN HIV-1 and located on the same side of the α helix [100], the residues K156 and K159 are involved in the positioning of the final dinucleotide CA [101]. Although these lysine residues are present in a few elements of the family (including IS*21* itself), other basic or polar residues, at identical or close positions, might play the same role. In addition, E152 appears to be involved in the specific recognition of the CA end by IN HIV-1 [99]. In the IS*21* family, the spacing between the first aspartate of the DDE triad and the conserved serine is variable, unlike the spacing separating the conserved serine from the second aspartate, which varies between 36 and 37 residues (fig. 5A). Whereas a number of other transposases and integrases have a DD(35)E arrangement, spacing between the second aspartate and glutamate is larger in IS*21* and related elements, extending from 42 to 48 residues (fig. 5A). It is possible that this spacing could form a loop playing a role in the control of the active site, as proposed for MuA where the motif is DD(55)E [96].

The ORF2 proteins of the IS*21* family, of which the IstB protein from IS*21* is the prototype, are better conserved than ORF1 proteins (fig. 5B; a total alignment of some elements is given in [39]). ORF2 proteins all have an ATP/GTP binding motif, similar to the motifs of the DnaA and DnaC proteins [32]. Other proteins of transposable elements showing the same type of motif are MuB of bacteriophage Mu [31, 102, 103] and TnsC of Tn*7* [104], for which ATP binding has been demonstrated, TnpB of Tn*552* [25, 105], and TniB of Tn*402* (=Tn*5090*) [106] and Tn*5053* [107, 108]. Like DnaC, these proteins are likely to form part of 'molecular matchmakers', which would use ATP for transient conformational changes promoting protein-DNA interactions [109]. These proteins can also be seen as ATP-dependent switches that control the assembly and activation of the transpososomes and may play a role in transposition immunity [35, 110].

Mechanistic aspects of the transposase/cointegrase specialization in the reactive junction pathway

Recent work on several bacterial insertion sequences has provided evidence for an intermediate, the IS minicircle, where the two ends of the transposable element are joined by a few base pairs [40]. IS minicircles and tandems share an important feature, i.e. an IR-IR junction. For IS*1*, a model including minicircle and tandem formation was proposed 20 years ago [111].

Several insertion sequences have been shown to form tandem and minicircle intermediates; tandem structures of IS*30*, IS*21*, IS*3*, IS*2* and IS*911* have been described [10, 13, 16, 112–115], whereas minicircles have been found for IS*911*, IS*21*, IS*3*, IS*1*, IS*2* and IS*30* [116–124]. These structures are both formed via an interaction between two separated IS ends, and the subsequent transpositional integration of these intermediates follows a similar pathway. These considerations lead us to propose a unifying model (fig. 6), which is based on findings reported for IS*911* and, to some extent, for IS*21* and IS*30*. We call this the reactive junction pathway. Similar models have previously been presented by others, especially by the groups of W. Arber, M. Chandler and F. Olasz [16, 29, 81, 113, 124].

IS*911* is a member of the IS*3* family [40]. The formation of IS*911* minicircles (fig. 6, 1st step) requires the OrfAB product, the transposase assembled by a programmed –1 frameshift occurring between two consecutive and partially superimposed ORFs [125], and probably some host functions to resolve the 'figure-eight' molecule, which results from the transfer of a single strand at one end of the IS element to the same strand close to the opposite IS terminus [117, 126]. Although linear forms of IS*911* originating from minicircles can give simple insertions in vitro [127], the main transposition pathway of IS*911* seems to proceed directly via minicircles [81, 128]. These are cleaved by the OrfAB product at the reactive IR-IR junction by two consecutive or simultaneous single strand cuts at the 3′ ends, which are then transferred to the target DNA (fig. 6, 2nd step). The integration of the minicircles is highly stimulated by the simultaneous presence of the OrfA product, which contains the helix-turn-helix motif but lacks the DD(35)E catalytic domain of OrfAB [81, 128]. Both OrfA and OrfAB contain a leucine zipper motif, which allows OrfAB oligomerization [129] and the regulation of IS*911* transposition through interaction of OrfA with OrfAB within a transposition complex [130]. Target capture appears to be an important function of OrfA [130]. Recently, tandem formation by site-specific recombination between the ends of two IS*911* copies on the same dimeric plasmid molecule has been shown to occur (as shown in fig. 1Db) and to produce the same type of IR-IR reactive junction as in minicircles [16]. In summary, the OrfAB product is not only necessary for the formation of minicircles but also for their integration into the target DNA.

IS*30* is a bacterial insertion sequence for which tandem formation was reported earlier [112]. Members of the IS*30* family have one ORF [40]. This codes for a transposase whose N-terminal part interacts specifically with the IRs of the element [131] and whose C-terminal part comprises a DDE motif. The highly reactive and unstable structure $(IS30)_2$ promotes replicon fusion (cointegration) and many rearrangements, including loss of one of the elements from the dimer (dimer dissolution) and deletion of adjacent DNA [113, 124]. The formation of IS*30* minicircles as intermediates in the simple transposition pathway was recently demonstrated [124]. IS*30* shows a duality in its specificity of insertion, with a preference for targets adjacent to IRs [132] and a not very rigorous hotspot [37]. Mutant forms of IS*30* transposase with deletion of 31 or 81 amino acid residues at the C-terminus are defective for tandem formation, dimer dissolution and replicon fusion, but retain some transpositional activity with a preformed IS*30*-IS*30* junction as a substrate [133], illustrating the fact that the presence of an IR-IR junction can simplify the catalytic task of the transposase by bypassing the 1st step (fig. 6).

IS*21* tandems have been studied extensively, as they have high cointegration activity [13]. For unknown reasons, $(IS21)_2$ is more stable than tandems of other IS elements. This fact has facilitated studies on $(IS21)_2$-mediated cointegrate formation. Interestingly, the spacing between abutted IS elements in tandems and minicircles corresponds to the target duplication for IS*911* (3 bp) and IS*30* (2 bp), but this is not the case of the spacing observed in $(IS21)_2$ (2 or 3 bp instead of the standard 4-bp target duplication) [11]. It was realized early that a plasmid carrying $(IS21)_2$ could be considered as a circular transposable

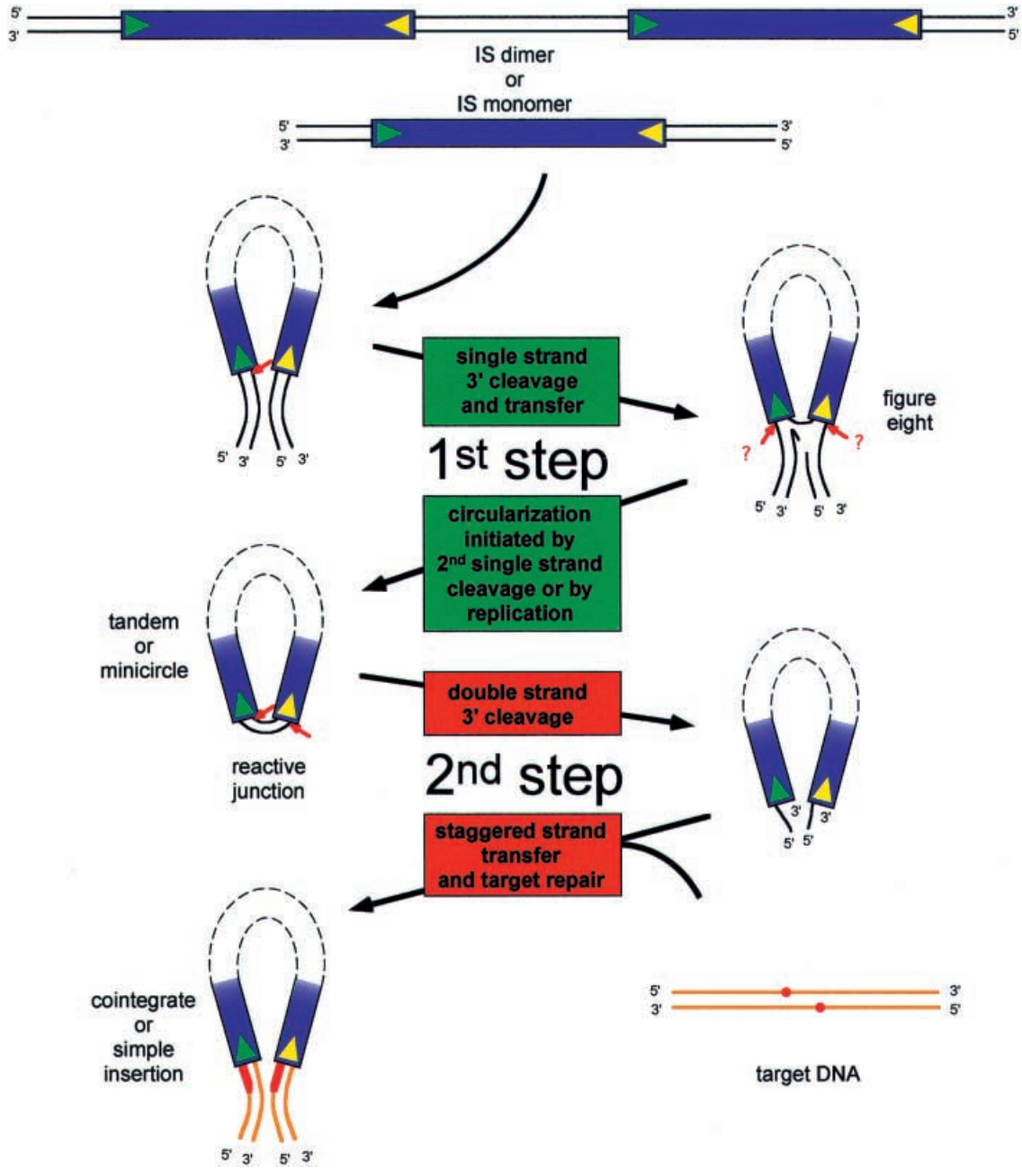


Figure 6. General model for the reactive junction transposition pathway. IS elements are shown in blue with yellow and green inverted repeats, backbone donor DNA in black, target DNA in orange, target duplications resulting from DNA repair of the target in red; red arrows show the single strand 3′ cleavages; red dots on the target DNA illustrate the staggered cuts.

element [10]. Such a structure has features similar to those of the circular intermediate structures (minicircles) postulated for other elements. Plasmids containing a single copy of IS*21* appear to release minicircular IS*21* in *E. coli* at about 10^{-3} with only the IstA function being required [118]. Although the structures of the putative IS*21* minicircles have not been entirely elucidated, we presume that they contain an IR-IR junction of the type found in $(IS21)_2$. This speculation is supported by analogy with tandems and minicircles of IS*911* and IS*30* [16, 124].

Transposase is 10,000 times more effective than cointegrase at promoting the simple transposition pathway of IS*21* [11]. This implies that transposase can bring together two ends of an IS*21* copy, e. g. by forming a figure eight (fig. 6, 1st step), whereas cointegrase cannot. The difference resides in the eight N-terminal amino acids of transposase, but remains unexplained mechanistically. Since the formation of an IR-IR junction in a tandem similarly requires the synapsis of two IS*21* ends – irrespective of the precise mechanism (fig. 1D) – we presume that transposase catalyzes this step.

Integration of an IR-IR junction into a target proceeds optimally in the presence of cointegrase and IstB [11, 18], and integration of natural minicircles and tandems can be expected to follow the same pattern (fig. 6, 2nd step). With respect to the insertion sequences discussed above, IS*30* transposase artificially truncated at the C-terminus retains cointegrase function and OrfA of IS*911*, like IstB, helps capture and integration into the target.

Integration of an IR-IR junction can also occur in the presence of transposase and IstB, albeit 100 times less efficiently than with cointegrase [11]. Therefore, the specialized cointegrase function can be seen as a means of accelerating the second step of the nonreplicative transposition pathway shown in figure 6, i.e. the reaction of two abutted IS ends with a target.

Concluding remarks

In nonreplicative transposition of insertion sequences and transposons, both strands of the transposable element are freed from the donor backbone at both ends of the element. This process can involve a hairpin intermediate with sequential cleavages of both strands as in Tn5 and Tn*10* transposition [134, 135], or the formation of structures with abutted IRs creating a reactive junction, as discussed here and in [136]. The reactive IS*21*-IS*21* tandem, once perceived as a curiosity in the world of insertion sequences [6, 10], now finds a place in the more general reactive junction pathway, which may be used not only by IS*21*, IS*911* and IS*30*, but also by other transposable elements.

Acknowledgements. We would like to thank M. Chandler, J. Mahillon and A. Stasiak for discussions, and V. Krishnapillai for reading and for comments on the manuscript. The work of B. Berger was supported by grants from the Swiss National Foundation (project 31-45896.95) and the International Brachet Foundation.

1 Craig N. L. (1996) Transposition. In: *Escherichia coli* and *Salmonella*, pp. 2339–2362, Neidhardt F. C. (ed.), ASM Press, Washington, DC
2 Reimmann C. and Haas D. (1993) Mobilization of chromosomes and nonconjugative plasmids by cointegrative mechanisms. In: Bacterial Conjugation, pp. 137–188, Clewell D. B. (ed.), Plenum Press, New York
3 Haas D. and Holloway B. W. (1976) R factor variants with enhanced sex factor activity in *Pseudomonas aeruginosa*. Mol. Gen. Genet. **144:** 243–251
4 Burkardt H. J., Riess G. and Pühler A. (1979) Relationship of group P1 plasmids revealed by heteroduplex experiments: RP1, RP4, R68 and RK2 are identical. J. Gen. Microbiol. **114:** 341–348
5 Riess G., Holloway B. W. and Pühler A. (1980) R68.45, a plasmid with chromosome mobilizing ability (Cma), carries a tandem duplication. Genet. Res. **36:** 99–109
6 Willetts N. S., Crowther C. and Holloway B. W. (1981) The insertion sequence IS*21* of R68.45 and the molecular basis for mobilization of the bacterial chromosome. Plasmid **6:** 30–52
7 Currier T. C. and Morgan M. K. (1982) Direct DNA repeat in plasmid R68.45 is associated with deletion formation and concomitant loss of chromosome mobilization ability. J. Bacteriol. **150:** 251–259
8 Haas D. and Riess G. (1983) Spontaneous deletions of the chromosome-mobilizing plasmid R68.45 in *Pseudomonas aeruginosa* PAO. Plasmid **9:** 42–52
9 Reimmann C., Rella M. and Haas D. (1988) Integration of replication-defective R68.45-like plasmids into the *Pseudomonas aeruginosa* chromosome. J. Gen. Microbiol. **134:** 1515–1523
10 Reimmann C. and Haas D. (1987) Mode of replicon fusion mediated by the duplicated insertion sequence IS*21* in *Escherichia coli.* Genetics **115:** 619–625
11 Schmid S., Seitz T. and Haas D. (1998) Cointegrase, a naturally occurring, truncated form of IS*21* transposase, catalyzes replicon fusion rather than simple insertion of IS*21*. J. Mol. Biol. **282:** 571–583
12 Reimmann C. and Haas D. (1990) The *istA* gene of insertion sequence IS*21* is essential for cleavage at the inner 3′ ends of tandemly repeated IS*21* elements in vitro. EMBO J. **9:** 4055–4063
13 Reimmann C., Moore R., Little S., Savioz A., Willetts N. S. and Haas D. (1989) Genetic structure, function and regulation of the transposable element IS*21*. Mol. Gen. Genet. **215:** 416–424
14 Seitz T., Berger B., Nguyen V. T., Tricot C., Villeret V., Schmid S. et al. (2000) Linker insertion mutagenesis based on IS*21* transposition: isolation of an AMP-insensitive variant of catabolic ornithine carbamoyltransferase from *Pseudomonas aeruginosa*. Protein Eng. **13:** 329–337
15 Berger B. (2000) La séquence d'insertion IS*21*: la famille de cet élément bactérien, sa spécificité d'insertion et son utilisation pour une méthode de linker insertion mutagenesis in vitro. PhD Thesis, University of Lausanne
16 Turlan C., Ton-Hoang B. and Chandler M. (2000) The role of tandem IS dimers in IS*911* transposition. Mol. Microbiol. **35:** 1312–1325
17 Danilevich V. N. and Kostiuchenko D. A. (1985) Immunity to repeated transposition of the insertion sequence IS*21*. [RUSSIAN]. Molekuliarnaia Biologiia **19:** 1242–1250
18 Schmid S., Berger B. and Haas D. (1999) Target joining of duplicated insertion sequence IS*21* is assisted by IstB protein in vitro. J. Bacteriol. **181:** 2286–2289
19 Galas D. J. and Chandler M. (1989) Bacterial insertion sequences. In: Mobile DNA, pp. 109–162, Berg D. E. and Howe M. M. (eds), ASM Press, Washington, DC
20 Craigie R., Mizuuchi M. and Mizuuchi K. (1984) Site-specific recognition of the bacteriophage Mu ends by the MuA protein. Cell **39:** 387–394
21 Allison R. G. and Chaconas G. (1992) Role of the A protein-binding sites in the in vitro transposition of Mu DNA. A complex circuit of interactions involving the Mu ends and the transpositional enhancer. J. Biol. Chem. **267:** 19963–19970
22 Baker T. A. and Mizuuchi K. (1992) DNA-promoted assembly of the active tetramer of the Mu transposase. Genes Dev. **6:** 2221–2232
23 Arciszewska L. K., Drake D. and Craig N. L. (1989) Transposon Tn7: *cis*-acting sequences in transposition and transposition immunity. J. Mol. Biol. **207:** 35–52
24 Arciszewska L. K., McKown R. L. and Craig N. L. (1991) Purification of TnsB, a transposition protein that binds to the ends of Tn7. J. Biol. Chem. **266:** 21736–21744
25 Rowland S. J., Sherratt D. J., Stark W. M. and Boocock M. R. (1995) Tn*552* transposase purification and in vitro activities. EMBO J. **14:** 196–205
26 Herzel H., Weiss O. and Trifonov E. N. (1998) Sequence periodicity in complete genomes of archaea suggests positive supercoiling. J. Biomol. Struct. Dyn. **16:** 341–345

27 Tomita M., Wada M. and Kawashima Y. (1999) ApA dinucleotide periodicity in prokaryote, eukaryote and organelle genomes. J. Mol. Evol. **49:** 182–192
28 Polard P. and Chandler M. (1995) Bacterial transposases and retroviral integrases. Mol. Microbiol. **15:** 13–23
29 Haren L., Ton-Hoang B. and Chandler M. (1999) Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. **53:** 245–281
30 Craig N. L. (1996) Transposon Tn7. Curr. Topics Microbiol Immunol 204: 27–48
31 Mizuuchi K. (1992) Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. **61:** 1011–1051
32 Koonin E. V. (1992) DnaC protein contains a modified ATP-binding motif and belongs to a novel family of ATPases including also DnaA. Nucleic Acids Res. **20:** 1997
33 Adzuma K. and Mizuuchi K. (1988) Target immunity of Mu transposition reflects a differential distribution of MuB protein. Cell **53:** 257–266
34 Adzuma K. and Mizuuchi K. (1989) Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. Cell **57:** 41–47
35 Yamauchi M. and Baker T. A. (1998) An ATP-ADP switch in MuB controls progression of the Mu transposition pathway. EMBO J. **17:** 5509–5518
36 Stellwagen A. E. and Craig N. L. (1997) Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. EMBO J. **16:** 6823–6834
37 Olasz F., Kiss J., Konig P., Buzas Z., Stalder R. and Arber W. (1998) Target specificity of insertion element IS*30*. Mol. Microbiol. **28:** 691–704
38 Hallet B., Rezsohazy R., Mahillon J. and Delcour J. (1994) IS*231A* insertion specificity: consensus sequence and DNA bending at the target site. Mol. Microbiol. **14:** 131–139
39 Haas D., Berger B., Schmid S., Seitz T. and Reimmann C. (1996) Insertion sequence IS*21*: related insertion sequence elements, transpositional mechanisms and application to linker insertion mutagenesis. In: Molecular Biology of Pseudomonads, pp. 238–249, Nakazawa T., Furukawa K., Haas D. and Silver S. (eds), ASM Press, Washington, DC
40 Mahillon J. and Chandler M. (1998) Insertion sequences. Microbiol. Mol. Biol. Rev. **62:** 725–774
41 Burland V., Shao Y., Perna N. T., Plunkett G., Sofia H. J. and Blattner F. R. (1998) The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. Nucleic Acids Res. **26:** 4196–4204
42 Kita K., Tsuda J., Kato T., Okamoto K., Yanase H. and Tanaka M. (1999) Evidence of horizontal transfer of the EcoO109I restriction- modification gene to *Escherichia coli* chromosomal DNA. J. Bacteriol. **181:** 6822–6827
43 Tobe T., Hayashi T., Han C. G., Schoolnik G. K., Ohtsubo E. and Sasakawa C. (1999) Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. Infect. Immun. **67:** 5455–5462
44 Pansegrau W., Lanka E., Barth P. T., Figurski D. H., Guiney D. G., Haas D. et al. (1994) Complete nucleotide sequence of Birmingham IncP α plasmids. Compilation and comparative analysis. J. Mol. Biol. **239:** 623–663
45 Hu P., Elliott J., McCready P., Skowronski E., Garnes J., Kobayashi A. et al. (1998) Structural organization of virulence-associated plasmids of *Yersinia pestis*. J. Bacteriol. **180:** 5192–5202
46 Perry R. D., Straley S. C., Fetherston J. D., Rose D. J., Gregor J. and Blattner F. R. (1998) DNA sequencing and analysis of the low-Ca^{2+}-response plasmid pCD1 of *Yersinia pestis* KIM5. Infect. Immun. **66:** 4611–4623
47 Matsutani S., Ohtsubo H., Maeda Y. and Ohtsubo E. (1987) Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196:** 445–455
48 Soby S., Kirkpatrick B. and Kosuge T. (1993) Characterization of an insertion sequence (IS*53*) located within IS*51* on the *iaa*-containing plasmid of *Pseudomonas syringae pv. savastanoi.* Plasmid **29:** 135–141
49 Buchrieser C., Prentice M. and Carniel E. (1998) The 102-kilobase unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. J. Bacteriol. **180:** 2321–2329
50 Lindler L. E., Plano G. V., Burland V., Mayhew G. F. and Blattner F. R. (1998) Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. Infect. Immun. **66:** 5731–5742
51 Podladchikova O. N., Dikhanov G. G., Rakin A. V. and Heesemann J. (1994) Nucleotide sequence and structural organization of *Yersinia pestis* insertion sequence IS*100*. FEMS Microbiol. Lett. **121:** 269–274
52 Filippov A. A., Oleinikov P. V., Motin V. L., Protsenko O. A. and Smirnov G. B. (1995) Sequencing of two *Yersinia pestis* IS elements, IS*285* and IS*100*. Contrib. Microbiol. Immunol. **13:** 306–309
53 Trukhachev A. L., Podladchikova O. N. and Dikhanov G. G. (1997) Molecular cloning and expression of IS*100* insertion element from *Yersinia pestis* in *Escherichia coli* cells [RUSSIAN]. Molekuliarnaia Biologiia **31:** 997–1001
54 McDonough K. A. and Hare J. M. (1997) Homology with a repeated *Yersinia pestis* DNA sequence IS*100* correlates with pesticin sensitivity in *Yersinia pseudotuberculosis*. J. Bacteriol. **179:** 2081–2085
55 Menou G., Mahillon J., Lecadet M. M. and Lereclus D. (1990) Structural and genetic organization of IS*232*, a new insertion sequence of *Bacillus thuringiensis.* J. Bacteriol. **172:** 6689–6696
56 Byrne A. M. and Lessie T. G. (1994) Characteristics of IS*401*, a new member of the IS*3* family implicated in plasmid rearrangements in *Pseudomonas cepacia*. Plasmid **31:** 138–147
57 Wood M. S., Lory C. and Lessie T. G. (1990) Activation of the *lac* genes of Tn*951* by insertion sequences from *Pseudomonas cepacia.* J. Bacteriol. **172:** 1719–1724
58 Solinas F., Marconi A. M., Ruzzi M. and Zennaro E. (1995) Characterization and sequence of a novel insertion sequence, IS*1162*, from *Pseudomonas fluorescens*. Gene **155:** 77–82
59 Brown H. J., Stokes H. W. and Hall R. M. (1996) The integrons In0, In2, and In5 are defective transposon derivatives. J. Bacteriol. **178:** 4429–4437
60 Nagy I., Schoofs G., Vanderleyden J. and De Mot R. R. (1997) Transposition of the IS*21*-related element IS*1415* in *Rhodococcus erythropolis*. J. Bacteriol. **179:** 4635–4638
61 Yeo C. C. and Poh C. L. (1997) Characterization of IS1474, an insertion sequence of the IS*21* family isolated from *Pseudomonas alcaligenes* NCIB 9867. FEMS Microbiol. Lett. **149:** 257–263
62 Yeo C. C., Wong D. T. and Poh C. L. (1998) IS*1491* from *Pseudomonas alcaligenes* NCIB 9867: characterization and distribution among *Pseudomonas species.* Plasmid **39:** 187–195
63 Gordon S. V., Heym B., Parkhill J., Barrell B. and Cole S. T. (1999) New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. Microbiology **145:** 881–892
64 Lefèvre P., Braibant M., Content J. and Gilot P. (1999) Characterization of a *Mycobacterium bovis* BCG insertion sequence related to the IS*21* family. FEMS Microbiol. Lett. **178:** 211–217
65 Eaton R. W., Selifonova O. V. and Gedney R. M. (1998) Isopropylbenzene catabolic pathway in *Pseudomonas putida* RE204: nucleotide sequence analysis of the *ipb* operon and neighboring DNA from pRE4. Biodegradation **9:** 119–132
66 Ogawa N. and Miyashita K. (1995) Recombination of a 3-chlorobenzoate catabolic plasmid from *Alcaligenes eutrophus*

NH9 mediated by direct repeat elements. Appl. Environ. Microbiol. **61:** 3788–3795
67 Ogawa N. and Miyashita K. (1999) The chlorocatechol-catabolic transposon Tn*5707* of *Alcaligenes eutrophus* NH9, carrying a gene cluster highly homologous to that in the 1,2,4-trichlorobenzene-degrading bacterium *Pseudomonas* sp. strain P51, confers the ability to grow on 3-chlorobenzoate. Appl. Environ. Microbiol. **65:** 724–731
68 Isawa T., Sameshima R., Mitsui H. and Minamisawa K. (1999) IS*1631* occurrence in *Bradyrhizobium japonicum* highly reiterated sequence-possessing strains with high copy numbers of repeated sequences RSα and RSβ. Appl. Environ. Microbiol. **65:** 3493–3501
69 Xu K., He Z. Q., Mao Y. M., Sheng R. Q. and Sheng Z. J. (1993) On two transposable elements from *Bacillus stearothermophilus*. Plasmid **29:** 1–9
70 Xu K., Mao Y., Shen R. and Sheng Z. (1997) Study on transposition behavior of IS*5376* in *Bacillus stearothermophilus* [CHINESE]. I Chuan Hsueh Pao – Acta Genetica Sinica **24:** 178–182
71 Rogers, M. B., Bennett, T. K., Payne, C. M. and Smith, C. J. (1994) Insertional activation of *cepA* leads to high-level β-lactamase expression in *Bacteroides fragilis* clinical isolates. J. Bacteriol. **176:** 4376–4384
72 Xu Y., Mortimer M. W., Fisher T. S., Kahn M. L., Brockman F. J. and Xun L. (1997) Cloning, sequencing and analysis of a gene cluster from C*helatobacter heintzii* ATCC 29600 encoding nitrilotriacetate monooxygenase and NADH : flavin mononucleotide oxidoreductase. J. Bacteriol. **179:** 1112–1116
73 Quadri L. E., Kleerebezem M., Kuipers O. P., de Vos W. M., Roy K. L., Vederas J. C. et al. (1997) Characterization of a locus from *Carnobacterium piscicola* LV17B involved in bacteriocin production and immunity: evidence for global inducer-mediated transcriptional regulation. J. Bacteriol. **179:** 6163–6171
74 Guyer D. M., Kao J. S. and Mobley H. L. (1998) Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis and catheter-associated bacteriuria and from fecal samples. Infect. Immun. **66:** 4411–4417
75 Tizard M., Bull T., Millar D., Doran T., Martin H., Sumar N. et al. (1998) A low G+C content genetic island in *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* with homologous genes in *Mycobacterium tuberculosis*. Microbiology **144:** 3413–3423
76 Zekri S., Soto M. J. and Toro N. (1998) IS*Rm4-1* and IS*Rm9*, two novel insertion sequences from *Sinorhizobium meliloti*. Gene **207:** 93–96
77 Freiberg C., Fellay R., Bairoch A., Broughton W. J., Rosenthal A. and Perret X. (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. Nature **387:** 394–401
78 Perret X., Viprey V., Freiberg C. and Broughton W. J. (1997) Structure and evolution of NGRRS-1, a complex, repeated element in the genome of *Rhizobium* sp. strain NGR234. J. Bacteriol. **179:** 7488–7496
79 Seitz T. (1996) Linker insertion mutagenesis auf der Basis der Insertionssequenz IS*21*. PhD Thesis No. 11807, ETH-Zurich,
80 McNamara P. T., Bolshoy A., Trifonov E. N. and Harrington R. E. (1990) Sequence-dependent kinks induced in curved DNA. J. Biomol. Struct. Dyn. **8:** 529–538
81 Ton-Hoang B., Betermier M., Polard P. and Chandler M. (1997) Assembly of a strong promoter following IS*911* circularization and the role of circles in transposition. EMBO J. **16:** 3357–3371
82 Hannenhalli S. S., Hayes W. S., Hatzigeorgiou A. G. and Fickett J. W. (1999) Bacterial start site prediction. Nucleic Acids Res. **27:** 3577–3582
83 Vellanoweth R. L. and Rabinowitz J. C. (1992) The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. Mol. Microbiol. **6:** 1105–1114
84 de Smit M. H. and van Duin J. (1994) Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. J. Mol. Biol. **235:** 173–184
85 Rocha E. P., Danchin A. and Viari A. (1999) Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. Mol. Biol. Evol. **16:** 1219–1230
86 Dodd I. B. and Egan J. B. (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:** 5019–5026
87 Andrake, M. D. and Skalka, A. M. (1996) Retroviral integrase, putting the pieces together. J. Biol. Chem. **271:** 19633–19636
88 Rezsohazy R., Hallet B., Delcour J. and Mahillon J. (1993) The IS*4* family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. **9:** 1283–1295
89 Katz R. A. and Skalka A. M. (1994) The retroviral enzymes. Annu. Rev. Biochem. **63:** 133–173
90 Baker T. A. and Luo L. (1994) Identification of residues in the Mu transposase essential for catalysis. Proc. Natl. Acad. Sci. USA **91:** 6654–6658
91 Junop M. S. and Haniford D. B. (1997) Factors responsible for target site selection in Tn*10* transposition: a role for the DDE motif in target DNA capture. EMBO J. **16:** 2646–2655
92 Sarnovsky R. J., May E. W. and Craig N. L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. **15:** 6348–6361
93 van Gent D. C., Groeneger A. A. and Plasterk R. H. (1992) Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc. Natl. Acad. Sci. USA **89:** 9598–9602
94 Katz R. A., Mack J. P., Merkel G., Kulkosky J., Ge Z., Leis J. et al. (1992) Requirement for a conserved serine in both processing and joining activities of retroviral integrase. Proc. Natl. Acad. Sci. USA **89:** 6741–6745
95 Dyda F., Hickman A. B., Jenkins T. M., Engelman A., Craigie R. and Davies D. R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science **266:** 1981–1986
96 Rice P. and Mizuuchi K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. Cell **82:** 209–220
97 Hickman A. B., Li Y., Mathew S. V., May E. W., Craig N. L. and Dyda F. (2000) Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. Mol. Cell **5:** 1025–1034
98 Esposito D. and Craigie R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. EMBO J. **17:** 5832–5843
99 Gerton J. L., Ohgi S., Olsen M., DeRisi J. and Brown P. O. (1998) Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. J. Virol. **72:** 5046–5055
100 Goldgur Y., Dyda F., Hickman A. B., Jenkins T. M., Craigie R. and Davies, D. R. (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc. Natl. Acad. Sci. USA **95:** 9150–9154
101 Jenkins T. M., Esposito D., Engelman A. and Craigie R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-cross-linking. EMBO J. **16:** 6849–6859
102 Miller J. L., Anderson S. K., Fujita D. J., Chaconas G., Baldwin D. L. and Harshey R. M. (1984) The nucleotide sequence

of the B gene of bacteriophage Mu. Nucleic Acids Res. **12:** 8627–8638
103 Mizuuchi K. and Craigie R. (1986) Mechanism of bacteriophage Mu transposition. Annu. Rev. Genet. 20: 385–429
104 Gamas P. and Craig N. L. (1992) Purification and characterization of TnsC, a Tn7 transposition protein that binds ATP and DNA. Nucleic Acids Res. **20:** 2525–2532
105 Rowland S. J. and Dyke K. G. (1990) Tn*552*, a novel transposable element from *Staphylococcus aureus*. Mol. Microbiol. **4:** 961–975
106 Rådström P., Sköld O., Swedberg G., Flensburg J., Roy P. H. and Sundström L. (1994) Transposon Tn*5090* of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. J. Bacteriol. **176:** 3257–3268
107 Kholodii G. Y., Gorlenko Z. M., Lomovskaya O. L., Mindlin S. Z., Yurieva O. V. and Nikiforov V. G. (1993) Molecular characterization of an aberrant mercury resistance transposable element from an environmental *Acinetobacter* strain. Plasmid **30:** 303–308
108 Kholodii G. Y., Yurieva O. V., Lomovskaya O. L., Gorlenko Z. M., Mindlin S. Z. and Nikiforov V. G. (1993) Tn*5053*, a mercury resistance transposon with integron's ends. J. Mol. Biol. **230:** 1103–1107
109 Sancar A. and Hearst J. E. (1993) Molecular matchmakers. Science **259:** 1415–1420
110 Stellwagen A. E. and Craig N. L. (1998) Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. Trends Biochem. Sci. **23:** 486–490
111 Read H. A., Das Sarma S. and Jaskunas S. R. (1980) Fate of donor insertion sequence IS*1* during transposition. Proc. Natl. Acad. Sci. USA **77:** 2514–2518
112 Dalrymple B. (1987) Novel rearrangements of IS*30* carrying plasmids leading to the reactivation of gene expression. Mol. Gen. Genet. **207:** 413–420
113 Olasz F., Stalder R. and Arber W. (1993) Formation of the tandem repeat $(IS30)_2$ and its role in IS*30*-mediated transpositional DNA rearrangements. Mol. Gen. Genet. **239:** 177–187
114 Spielmann-Ryser J., Moser M., Kast P. and Weber H. (1991) Factors determining the frequency of plasmid cointegrate formation mediated by insertion sequence IS*3* from *Escherichia coli*. Mol. Gen. Genet. **226:** 441–448
115 Szeverenyi I., Bodoky T. and Olasz F. (1996) Isolation, characterization and transposition of an $(IS2)_2$ intermediate. Mol. Gen. Genet. **251:** 281–289
116 Polard P., Prère M. F., Fayet O. and Chandler M. (1992) Transposase-induced excision and circularization of the bacterial insertion sequence IS*911*. EMBO J. **11:** 5079–5090
117 Polard P. and Chandler M. (1995) An in vivo transposase-catalyzed single-stranded DNA circularization reaction. Genes Dev. **9:** 2846–2858
118 Schmid S. (1993) Die Insertionssequenz IS*21*: Entwicklung eines In-vitro-Integrationssystems. PhD Thesis No. 10197, ETH-Zürich
119 Sekine Y., Eisaki N. and Ohtsubo E. (1994) Translational control in production of transposase and in transposition of insertion sequence IS*3*. J. Mol. Biol. **235:** 1406–1420
120 Sekine Y., Aihara K. and Ohtsub, E. (1999) Linearization and transposition of circular molecules of insertion sequence IS*3*. J. Mol. Biol. **294:** 21–34
121 Turlan C. and Chandler M. (1995) IS*1*-mediated intramolecular rearrangements: formation of excised transposon circles and replicative deletions. EMBO J. **14:** 5410–5421
122 Sekine Y., Eisaki N., Kobayashi K. and Ohtsubo E. (1997) Isolation and characterization of IS*1* circles. Gene **191:** 183–190
123 Lewis L. A. and Grindley N. D. F. (1997) Two abundant intramolecular transposition products, resulting from reactions initiated at a single end, suggest that IS*2* transposes by an unconventional pathway. Mol. Microbiol. **25:** 517–529
124 Kiss J. and Olasz F. (1999) Formation and transposition of the covalently closed IS*30* circle: the relation between tandem dimers and monomeric circles. Mol. Microbiol. **34:** 37–52
125 Chandler M. and Fayet O. (1993) Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. **7:** 497–503
126 Polard P., Ton-Hoang B., Haren L., Betermier M., Walczak R. and Chandler M. (1996) IS*911*-mediated transpositional recombination in vitro. J. Mol. Biol. **264:** 68–81
127 Ton-Hoang B., Polard P., Haren L., Turlan C. and Chandler M. (1999) IS*911* transposon circles give rise to linear forms that can undergo integration in vitro. Mol. Microbiol. **32:** 617–627
128 Ton-Hoang B., Polard P. and Chandler M. (1998) Efficient transposition of IS*911* circles in vitro. EMBO J. **17:** 1169–1181
129 Haren L., Polard P., Ton-Hoang B. and Chandler M. (1998) Multiple oligomerisation domains in the IS*911* transposase: a leucine zipper motif is essential for activity. J. Mol. Biol. **283:** 29–41
130 Haren L., Normand C., Polard P., Alazard R. and Chandler M. (2000) IS*911* transposition is regulated by protein-protein interactions via a leucine zipper motif. J. Mol. Biol. **296:** 757–768
131 Stalder R., Caspers P., Olasz F. and Arber W. (1990) The N-terminal domain of the insertion sequence 30 transposase interacts specifically with the terminal inverted repeats of the element. J. Biol. Chem. **265:** 3757–3762
132 Olasz F., Farkas T., Kiss J., Arini A. and Arber W. (1997) Terminal inverted repeats of insertion sequence IS*30* serve as targets for transposition. J. Bacteriol. **179:** 7551–7558
133 Olasz F., Farkas T., Stalder R. and Arber W. (1997) Mutations in the carboxy-terminal part of IS*30* transposase affect the formation and dissolution of $(IS30)_2$ dimer. FEBS Lett. **413:** 453–461
134 Kennedy A. K., Guhathakurta A., Kleckner N. and Haniford D. B. (1998) Tn*10* transposition via a DNA hairpin intermediate. Cell **95:** 125–134
135 Bhasin A., Goryshin I. Y. and Reznikoff W. S. (1999) Hairpin formation in Tn*5* transposition. J. Biol. Chem. **274:** 37021–37029
136 Turlan C. and Chandler M. (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. Trends Microbiol. **8:** 268–274