The Integrons In0, In2, and In5 Are Defective Transposon Derivatives

HEIDI J. BROWN,1,2 H. W. STOKES,1 AND RUTH M. HALL2*

School of Biological Sciences, Macquarie University, New South Wales 2109,1 and Sydney Laboratory, CSIRO Division of Biomolecular Engineering, North Ryde, New South Wales 2113,2 Australia

Received 5 February 1996/Accepted 24 May 1996

The class 1 integrons In0, In2, and In5, found in different locations in pVS1, Tn21, and pSCH884, have closely related structures. All three integrons contain an insertion sequence, IS1326, that is a new member of the IS2 family. IS1326 has caused deletions of adjacent 3′-CS and transposition module sequences, and all three integrons retain a complete copy of only one of four genes required for transposition of related transposons and are thus defective transposon derivatives. In2 contains an additional insertion sequence, IS1353, located within IS1326. IS1353 is a member of the IS3 family and appears to have been acquired after the integron was inserted into an ancestral mercury resistance transposon to create the ancestor of Tn21 and several other transposons that are close relatives of Tn21.

Many of the antibiotic resistance genes found in clinical isolates of gram-negative bacteria are known to be located in integrons (reviewed in references 12, 14, 15, and 31). The resistance genes are contained in discrete mobile elements known as gene cassettes and are integrated at a specific site, attI, in the integron. The integron encodes a site-specific recombination that belongs to the lambda integrase family, responsible for the insertion of cassettes at attI, and also provides the promoter responsible for expression of the cassette-encoded antibiotic resistance genes.

Three classes of integrons have been identified (see reference 31). However, the integrons most commonly isolated from antibiotic-resistant clinical isolates of members of the family Enterobacteriaceae and pseudomonads belong to class 1 and most of them contain the sulI sulfonamide resistance gene. These integrons appear to be mobile elements, as they are found in many different locations, in plasmids such as R46 and R388 and within transposons such as Tn21 and Tn1696 (16, 43). However, early studies aimed at identifying the sequences common to all integrons failed to identify features consistent with the classification of integrons as transposons (43). Two conserved regions located on either side of the integrated gene cassettes were identified. The 5′-CS and transposition module sequences identical to that found in the sulfI-containing integrons and two integrated cassettes, dfrB3 (dfrBIIe) and orfD. Tn402 also includes the complete qacE gene (26) and a 4-kb region that contains four ORFs, three of which were identified as encoding potential transposition functions, on the basis of similarities to known transposition proteins (30). Close relatives of these four genes are also found in the transposition module of mercury resistance transposon Tn5053, and it has been shown that three of the genes tinA, tinB, and tinQ, are essential for transposition and the fourth, tinR, is required for cointegrate resolution (18). Partial sequencing of In2 revealed that a 2.7-kb segment of the Tn402 transposition module is also present in In2 from Tn21, but only one complete gene (tinA) and a truncated form of a second gene (tinB) are present in this fragment (30). Thus, it appears that In2 may be a defective transposon that does not contain a complete set of transposition functions.

In In2, the 3′-CS sequences (13, 43) and the transposition module segment (30) are separated by a region of about 4 kb, and short inverted repeats (27 bp) have been identified at or close to the boundaries of this segment in In2 (45). The se-
quence of the left-hand end of this region has been determined (13), and identical sequences were found in In0 and In5, suggesting that In2, In0, and In5 share a common lineage. However, in In0 and In2, a deletion of 359 bases of the 3'-CS that are present in In5 has occurred (13). Together, these findings suggest that a further transposition or insertion sequence may be present in In2, In0, and In5.

To clarify the relationships between the structures found in different integrons, we examined the sequence of the region between the 3'-CS and transposition module regions in In2, In0, and In5. In all three integrons, an insertion sequence, IS1326, was found to be present, and in In2, a further insertion sequence, IS1353, inserted just inside one of the terminal inverted repeats of IS1326 was found. The regions adjacent to IS1326 were identical in In2 and In0, indicating that these two integrons have a recent common origin. In5 includes extra 3'-CS sequences to the left of IS1326 and has 439 bp less of the transposition module to the right, indicating that IS1326 has caused deletions of adjacent sequences. Possible routes for the evolution of the In2, In0, and In5 group from an integron with a backbone structure similar to that of Tn402 are discussed, and a coherent tree for the evolution of the Tn21 transposon family is presented.

**MATERIALS AND METHODS**

**Bacterial strains and plasmids.** *Escherichia coli* JM109 (Δlac-proAB) ampE thi F’ (traD36 proAB lacI9780ΔZ ΔM13) was used to propagate plasmid DNA. The plasmids used are listed in Table 1. The prMH series of plasmids listed were constructed by randomly cloning fragments from EcorI, HindIII, and SphiI digests of the relevant parental plasmid (pVS1, Tn21, and pSCH884) into pUC19 (52) by using standard procedures (37). Plasmids containing the appropriate fragments were identified by restriction mapping and by sequencing of the 3'-CS that is present in Tn21 (bases 2626 to 2890 in Tn21) and is 274 bp longer than the transposition module of Tn21, which is also a class 1 integron, and the sequence identity of the terminus of In0 (bases 2626 to 2890 in Tn21) and Tn21. This 2.7-kb region lies within the transposition module of Tn402, which contains the ORFs encoding transposition functions (18, 30). Only 2.26 kb of the transposition module is present in In5. As these regions in-
include only the complete *tniA* gene and part of the *tniB* gene, it seems reasonable to conclude that the ancestor of In0 and In5 contained the complete transposition module, part of which was subsequently lost as the result of one or more IS326-mediated deletion events.

**Insertion sequence IS326 is a member of the IS21 family.** IS326 includes two long ORFs that overlap by 13 bp (Fig. 2). The polypeptides predicted by these ORFs were found to be related to the two polypeptides encoded by the *istA* and *istB* genes of IS21 (34) and to the corresponding polypeptides in three further known relatives of IS21, namely, IS232 (23), IS3376 (51), and IS100 (27). On the basis of the similarity of the IS326 ORFs to IstA and IstB from IS21, which have been shown to be required for transposition (33), the IS326 genes are also designated *istA* and *istB*. A detailed description of the features common to members of the IS21 family, including IS326, will be presented elsewhere (2a). Briefly, IstA includes an unusual D, D(45), E transposase motif and a helix-turn-helix motif and IstB has a distinctive set of NTP-binding domains. Inverted repeats of 27 bp are present at the boundaries of IS326 in In0, In2, and In5, as has been noted previously for In2 (45). However, as many of the mobile elements that encode a transposase with a D, D(35), E motif are bounded by 5'-TG and CA-3' (10, 28), we have assumed that IS326 is bounded by the same dinucleotides and that the inverted repeats are most likely to be 26 bp long.

**In2 is closely related to In0 but contains IS353 inserted within IS326.** The restriction map of In2 includes an extra segment not found in In0 and In5 (see Fig. 1). As this region has previously been shown to be absent from transposons such as Tn2411, which are otherwise very close relatives of Tn21 (20), the possibility that this extra segment corresponds to an insertion sequence was examined. The sequence of In2 from the EcoR1 site in IS326 to the outer, right-hand end of In2 was determined (GenBank accession number U42226), and this completes the sequences of In2 and Tn21. In2 contains a complete copy of IS326, but a further new insertion sequence, designated IS353, has been inserted just within the one end (right-hand end as shown in Fig. 3) of IS326. The part of the In2 sequence containing IS353 is shown in Fig. 4. IS353 is 1,613 bp long and is bounded by inverted repeats of 12 and 13 bp (boxed in Fig. 4). Comparison of the In2 sequence with that of In0 (Fig. 2) revealed that IS353 is flanked by a 2-bp direct duplication of the target sequence (indicated in Fig. 2). IS353 is a member of the IS3 family (see below). Beyond the end of IS326, the In2 sequence is identical to that of In0, and identity to the sequence of part of the transposition module of Tn402 extends up to the right-hand end of In2, as suggested by Rådström et al. (30), who sequenced several parts of this region. Thus, In2 is identical to In0, except that it includes an integrated *aadA1* gene cassette, that determines resistance to streptomycin and spectinomycin, and IS353.

**IS353 is a member of the IS3 family.** Analysis of the IS353 ORFs longer than 50 amino acids revealed that on one strand there were two ORFs, orfA of 222 codons and orfB of 304 codons, spanning almost the entire insertion sequence (Fig. 4). Several ORFs were also identified on the other strand. Only ORFB was found to be related to known sequences. ORFB shares 20 to 37% identity with the ORFB products of IS3, IS911, and the other 32 known members of the IS3 family (10, 21, 28, 49), and several conserved regions found in the transposases of IS3 family members are shared by ORFB. In particular, the acidic residue motif D, D(35), E, which was originally identified in the IS3 family (10), is present in ORFB. This motif is also present in other bacterial transposases, including those of Tn552 (36) and the IS4 family (35), as well as in the integrases of many retrotransposons and retroviruses (10, 22), and is believed to be part of the catalytic domain of these transposases and integrases (see reference 38).

IS353 also exhibits an organization similar to that of members of the IS3 family. Most members of the IS3 family are bounded by 5'-TG and CA-3' (10, 28), and IS353 shares this feature. Members of this group generally code for two ORFs, one short ORF (orfA) that overlaps a larger downstream ORF (orfB), which is in phase 1 with respect to orfA. Expression studies with IS911 (29) and IS150 (48) have shown that these

---

**FIG. 1.** Restriction maps of *sul1*-type integrons In0, In2, and In5. Arrowheads indicate the 25-bp inverted repeats at the outer boundaries, and the vertical arrows indicate the point at which the In0 and In2 sequences diverge from that of In5. Solid lines below the maps indicate the extent of published sequences collated from the following sources: 1, Bissonnette and Roy (1); 2, Hall et al. (13); 3, Diver et al. (9); 4, Stokes and Hall (43); 5, Sundström et al. (44); 6, Brown et al. (2); 7, Shaw et al. (40). Dashed lines show the extent of sequences determined in this study. Restriction enzyme sites: B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; Ps, PstI; S, SalI; Sm, Smal; Sp, SphI.
elements produce three polypeptides, ORFA, ORFB, and fusion protein ORFAB. ORFAB, which appears to be the active transposase for IS911 (29), is a transframe product, resulting from a translational −1 frameshift that occurs in the region where orfA and orfB overlap. Specific signals induce the −1 frameshift; a heptanucleotide motif, AAAAAG, in the overlap region; and a downstream sequence capable of forming a strong secondary structure, such as a stem-loop structure (29, 48). These signals are also present in many other members of the IS3 family (3, 21, 49), and in the region of overlap between ORFA and ORFB in IS1353, the potential frameshift motif A6G is present and a pair of inverted repeats that have the potential to form a stem-loop structure are found downstream (Fig. 4). This suggests that IS1353, in common with other members of the IS3 family, also utilizes translational frameshifting. It is highly likely that in IS911, translation of ORFB is initiated at an ATT codon that lies immediately upstream of the A6G motif (29). In IS1353, there is also an ATT codon 3 bp upstream of the A6G motif; however, an ATG codon is also present six codons downstream, and which codon is in fact used for initiation of ORFB needs to be determined experimentally.

**DISCUSSION**

The sequence data presented here show that integrons In0, In2, and In5 all contain IS1326 and are also otherwise closely related (Fig. 3). The two differences between In0 and In2 are readily explained by known mechanisms for DNA insertion.

![Diagram](http://jb.asm.org/)
Whereas In0 contains no integrated cassettes (1), In2 contains the aadA1 cassette (44) that has been integrated at attI by the action of the integron integrase IntI1 (4). The further extra segment in In2 corresponds to an acquired insertion sequence, IS1353. There are only three differences between In0 and In5, one of which is the presence of an aacA cassette integrated at attI in In5 (40). The remaining differences are both deletions with one endpoint at an IS1326 boundary and can be explained by invoking deletion events originating at the boundaries of IS1326. These events are presumably mediated by the action of the insertion sequence-encoded transposition functions in a manner analogous to the formation of adjacent deletions by other insertion sequences and transposons (19).

Recent studies on Tn5053, which has a transposition module closely related to that of Tn402, have demonstrated that three genes, tniA, tniB, and tniQ, are essential for transposition of Tn5053 and that a further gene, tniR, that encodes a resolvase is required for efficient resolution of cointegrate intermediates (18). Because the complete tni module found in Tn402 is not present in any of the three integrons examined here, it seems reasonable to conclude that these integrons are all defective derivatives of an ancestral transposon and are unable to direct their own movement. However, the fact that In2 and In0, which contain identical 3'-CS and tni module segments, are found in independent locations, in Tn21 and in pVS1, respectively (see reference 13), suggests that movement can occur. As movement of Tn5053 derivatives with defects in any one of the tniA, tniB, and tniQ genes could be restored by supplying

FIG. 4. Nucleotide sequence of the In2 region containing IS1353. The IS1353 inverted repeats are boxed. The potential heptanucleotide frameshift signal A 6G is underlined, and the inverted repeats of the downstream potential stem-loop structure are indicated by arrows. Diamonds indicate the conserved acidic residues of the D, D(35), E motif. The terminal inverted repeat from IS1326 is underlined with an arrow.
the evolution of the sulI gene, which is found in the qacE of which contain the integrated gene cassettes (for the movement of the gene cassettes, the variable region -CS encoding the integrase responsible for site-specific events in the evolution of one of them. The acquisition of IS1353 by In2 is presumed to be a recent event (see below).

Evolution of the Tn21 transposon family. By using information from this study, together with data from previously published restriction maps of Tn21 and related transposons and heteroduplex analyses involving pairs of transposons from this family (11, 20, 24, 25, 38, 47), it is possible to construct a coherent model for the evolution of members of this family (Fig. 6). Only close relatives of Tn21 identified by the presence of the characteristic motif of EcoRI and HindIII sites found in the Tn21 transposition gene (tnpA, tnpR, and res) region are considered here as members of the Tn21 family. This group includes a little over half of the transposons listed as belonging to the Tn21 subgroup of the Tn3 transposon family in a recent review (see Table 1 in reference 11). Tn1401, Tn1406, and Tn1409 were not considered because the most recently published maps of these transposons (24) are not sufficiently accurate to permit analysis. One further transposon, Tn5086, that has subsequently been identified as a member of this group (45) was considered.

In the simplest evolutionary tree consistent with all of the published data, the first step in the evolution of this family is the insertion of an integron identical to In0 (or equivalent to In0) into the sulI gene cassette, for simplicity it has been assumed that this cassette was present in the incoming integron, Tn2411, rather than Tn21, is assumed to be the true ancestor of the integron-containing members of this family because the acquisition of IS1353 by In2 (in Tn21) appears to be a recent event that occurred after the movement of the integron to its present location. The region corresponding to IS1353 is not present in transposons that are otherwise very closely related to Tn21, such as Tn4 and Tn2411 (20, 25, 47), and it has been suggested that this region was acquired in Japan, as it is generally present in Tn21-like transposons isolated in Japan (11).

Most of the differences between Tn2411 and the known transposons in this family can be explained by site-specific integration of further cassettes into the integron or insertion of further transposons (e.g., Tn3 in Tn4 and IS1353 in Tn21) at various positions. Two further differences can be explained by IS1326-mediated events. The large deletion in Tn1831 (25) originates close to one boundary of IS1326, extends beyond the right-hand end of the integron into the mer region, and is likely to be due to a further IS1326-mediated deletion event. Trans-
poson Tn5086 contains all of the bases to the right of IS1326 in In2 and all but the last four bases to the left (TATC), which are replaced by two bases (CA) (45). The simplest explanation for this configuration is excision of IS1326 with imprecise rejoining and repair of the ends generated. A similar configuration is found in Tn2608 (47), which may also have arisen by this route.

Other members of the Tn21 subgroup are also known to include integrons, the best known being Tn1696, which contains the integron In4 (13, 43). However, Tn1696 is not a close relative of Tn2411 and the integron is located at a different position (38) and has a different structure, which includes IS6/100 (13). Thus, it appears that Tn1696 has a different evolutionary history and that integrons have been inserted into ancestral transposons containing imp and mer modules on more than one occasion. Studies to resolve the complete structure of In4 and Tn1696 are in progress.

ACKNOWLEDGMENTS

We thank Vilma Stanisich for communicating data prior to publication. Access to DNA databases and sequence analysis programs and assistance in their use were provided by the Australian National Genomic Information Service (A.N.G.I.S.). This work was supported by a grant from the National Health and Medical Research Council.

REFERENCES


40a. Stanisich, V. A. Personal communication.


