Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 8312–8316, October 1991
Microbiology

Chlorobenzoate catabolic transposon Tn*5271* is a composite class I element with flanking class II insertion sequences

(transposase/Tn*3* family/evolution)

CINDY NAKATSU*, JAMES NG*, RAMA SINGH†‡, NEIL STRAUS†, AND CAMPBELL WYNDHAM*§

*Ottawa-Carleton Institute of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada; and †Department of Botany, University of Toronto, Toronto, ON M5S 2B3, Canada

Communicated by Allan M. Campbell, March 18, 1991 (received for review July 17, 1990)

ABSTRACT **The structure of a transposon specifying the biodegradation of chlorobenzoate contaminants is described. Tn*5271* is a 17-kilobase (kb) transposon that resides in the plasmid or chromosome of *Alcaligenes* sp. strain BR60 and allows this organism to grow on 3- and 4-chlorobenzoate. The transposon is flanked by a directly repeated sequence of 3201 base pairs (bp), which in turn is flanked by 110-bp inverted repeats. The 3.2-kb repeated sequence, designated IS1071, exists in multiple copies in the genome of *Alcaligenes* sp. strain BR60 and is involved in recombination of the catabolic genes into the chromosome of this strain. Sequence analysis revealed that the inverted repeat of IS1071 and the derived amino acid sequence of the single open reading frame within IS1071 are related to the inverted repeats and transposase (TnpA) proteins of the class II (Tn*3* family) transposable elements. The absence of a resolvase gene within IS1071 suggests that this element is capable of determining the first step in class II transposition only. This was confirmed by observations on the IS1071-dependent formation of stable cointegrates in a recombination-deficient *Escherichia coli*. These results support an evolutionary scheme in which the class II transposable elements descended from simple insertion sequences.**

Two of the most effective pressures driving natural selection in bacterial populations are the widespread use of antibiotics and the provision of unusual organic carbon sources in the form of environmental contaminants. In response to these pressures, bacteria have evolved varied mechanisms of resistance to antibiotics and different means of catabolizing unusual organic compounds. It is not surprising that the molecular mechanisms of adaptation to antibiotics and xenobiotics are similar. Plasmids have been implicated in the catabolism of complex organic compounds for two decades (1). The rearrangement of catabolic genes along with changes in their expression have been implicated as mechanisms in the evolution of bacteria exposed to recalcitrant carbon sources for almost as long (2). The discovery of transposable elements specifying resistance to antibiotics (3) was followed within a very few years by the first descriptions of transposition of the determinants for toluene degradation carried on plasmid pWWO (TOL) in *Pseudomonas putida* mt-2 (4–6).

Bacterial transposable elements fall into two well-defined structural classes (7). Class I elements include insertion sequences (IS), containing the genetic determinants for transposition only and composite transposons formed when flanking IS elements mobilize an intervening sequence (8). Class II elements, or the Tn*3* family transposons, are related by inverted repeat (IR) similarities, transposase amino acid sequence homologies, and transposition mechanism (9). The class II catabolic transposons that have been characterized encode lactose (Tn*951*; ref. 10) and toluene metabolism (Tn*4653*; ref. 11). Both are complex, nested class II elements. The naphthalene catabolic genes on plasmid NAH7 have recently been localized to a defective, class II transposon, Tn*4655*, which lacks a transposase function but has an intact resolution system (12).

Catabolic genes specifying the biodegradation of chlorinated aromatic compounds have been observed to undergo rearrangements; however, none of these has been shown to transpose as a defined element. Recently Tomasek *et al.* (13) mapped a 1477-base-pair (bp) IS element flanking 2,4,5-T catabolic genes of a *Pseudomonas cepacia* strain, suggesting a possible class I composite transposon structure.

In 1988 we described an *Alcaligenes* sp. strain BR60, isolated from a chlorobenzoate contaminated landfill that undergoes deletion of chlorobenzoate catabolic genes from the indigenous plasmid pBRC60 (formerly pBR60) and exhibits recombination of plasmid sequences into the chromosome (14). We have used this host–plasmid association to study gene transfer in the environment (15). In the following report, we describe the structure of the chlorobenzoate catabolic transposon in *Alcaligenes* sp. strain BR60, the nucleotide sequence of the flanking direct repeats in this element, and their relationship to both class I and class II transposable elements.¶

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions. *Alcaligenes* sp. strains BR60 (pBRC60), BR40 (pBRC40), BR6053, and BR6020 have been described (14). For plasmid or total genomic DNA isolations, these strains were grown on 1% tryptone/0.5% yeast extract/0.5% NaCl, at 25°C. *Escherichia coli* strains JM109, HB101, and DH5α, with their plasmids pUC18, pUC19, pDPL13, pRK2013, or phage M13mp18 and M13mp19, have been described (16–19). A spontaneous rifampicin-resistant mutant of HB101 was also used. They were grown in Luria broth (19) at 37°C, with filter-sterilized ampicillin, kanamycin sulfate (Sigma), or rifampicin (Boehringer Mannheim) at 50, 50, or 200 μg/ml, respectively.

DNA Manipulations and Restriction Mapping. Plasmids pBRC60, pBRC40, and *E. coli* plasmids were isolated as described (14, 19). Single-stranded M13mp18 and M13mp19 were isolated by the method of Dale *et al.* (20). Total genomic DNA was prepared from *Alcaligenes* strains by the procedure of Wheatcroft and Watson (21). Genomic DNA, pBRC40, and pBRC60 were single or double digested with *Hin*dIII, *Eco*RI, *Pst* I, *Sal* I, *Nar* I, *Bam*HI, *Nru* I, *Bgl* II, *Sac*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IS, insertion sequence(s); IR, inverted repeat; IHF, integration host factor.

‡Present address: Institute for Marine Bioscience, National Research Council of Canada, Halifax, NS B3H 3Z1, Canada.
§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M65135).

I, and *Sma* I (Bethesda Research Laboratories) and resolved in 0.7% agarose in Tris borate/EDTA buffer (19). The catabolic region of pBRC60 was mapped using the 10 restriction enzymes listed above by hybridizing [α-^{32}P]dCTP-labeled (Amersham) fragments to pBRC60 digests immobilized on nylon (Hybond; Amersham) (14).

Insertion sequence IS1071 on a 4.1-kilobase (kb) *Nru* I fragment of pBRC40 was cloned into the *Nru* I site of pDPL13 (17) to give plasmid pBRN4029. The outside flanking ends of the two IS1071 sequences in Tn*5271* were cloned as 1.2- and 1.9-kb *Eco*RI/*Nru* I fragments into *Eco*RI/*Sma* I-cut pUC18 to give pBREN12 and pBREN19, respectively. The inside flanking ends were obtained from existing clones (pBRE11 and pBRH3; ref. 14). Transformations into *E. coli* JM109 or DH5α were carried out either by the method of Hanahan (22) or by electroporation with a Gene Pulser apparatus (Bio-Rad).

DNA Sequencing. *Hin*dIII fragment 4 cloned in pUC18 (pBRH4; ref. 14) was subcloned into the replicative forms of M13mp18 and M13mp19 (19) and then sequenced in both directions by using the commercial Sequenase kit (United States Biochemical) and deoxyadenosine 5'-[α-[^{35}S]thio]-triphosphate (Amersham). Purified plasmid pBRN4029 was then used for plasmid sequencing in both directions away from the H4 internal sequence. Plasmid sequencing was also used to characterize the termini of the two IS1071 elements cloned from pBRC60. Oligonucleotide primers were prepared by using a Cyclone DNA synthesizer (Biosearch). Sequence analyses were performed by using Microgenie version 6.0 (Beckman). Tn*5271*, IS1071, pBRC60, and pBRC40 are Plasmid Reference Center listings (E. Lederberg, Stanford).

RESULTS

The catabolic region of the 85-kb plasmid pBRC60 is shown in Fig. 1. Tn*5271* (17 kb) is flanked by 3.2-kb direct repeat sequences designated IS1071. Within the internal 10.6-kb unique region are sequences required for the growth of *Alcaligenes* sp. strain BR60 on chlorobenzoates. Deletion of the catabolic region from pBRC60 occurs at a frequency of 1.6×10^{-3} per cell per generation, in the absence of 3-chlorobenzoate, to give strain BR40 (14). The deletion plasmid pBRC40 (71 kb) retains one copy of IS1071 entirely within a 4.1-kb *Nru* I fragment generated by the deletion event. The DNA sequence of this copy of IS1071 is shown in Fig. 2. The sequence shown represents the product of an intramolecular recombination at an unknown point between the two direct repeats of IS1071 present on pBRC60. This was confirmed by comparison of the sequence in Fig. 2 with sequences at the four IS1071 termini of Fig. 1:

left outside,
5'-GCCGGTCTCGCGGTC GGGGTCTCCTCGTTT-3';
left inside,
5'-CGTCGAGCAGGGCAT GGGGTCTCCTCGTTT-3';
right inside,
5'-CGGGCAGCTTGTCGC GGGGTCTCCTCGTTT-3';
right outside,
5'-GGTTTTGCTCGTAAA GGGGTCTCCTCGTTT-3'.

The termini of IS1071 in Fig. 2 correspond to the left outside and right outside sequences of Tn*5271*. There are no duplicated sequences flanking the IS1071 termini.

The distribution of IS1071 in the genomes of *Alcaligenes* sp. strain BR60, strain BR6020 (mitomycin C cured), and strain BR6053 (chromosomal Tn*5271* recombinant) (14) was determined by digestion of genomic DNAs with *Bgl* II and probing with *Hin*dIII fragment 4 (IS1071 probe; data not shown). The BR60 genomic digest had seven fragments homologous to the IS1071 probe, ranging in size from 5 to 28 kb. Two of these (11 and 24 kb) corresponded to pBRC60 fragments containing the IS element (fragments *Bgl* II-2 and -3; Fig. 1). Strain BR6053 genomic digests also contained seven copies of IS1071, but the pBRC60 fragments were missing and two new chromosomal fragments were observed. Strain BR6020 genomic digests contained only three copies of IS1071 on chromosomal *Bgl* II fragments of 9, 19, and 22 kb.

DNA homology searches revealed that the 110-bp IR of IS1071 was related to the class II (Tn*3* family) IRs (Fig. 3). The regions of homology in the outer 38 bp have been recognized as DNA-binding and cutting sites for the class II transposases (25). There was no significant homology of the IS1071 IR sequence extending from 39 to 110 bp with sequences flanking the 38-bp IRs of the Tn*3* family transposons shown. However, an integration host factor (IHF) consensus binding sequence occurs at nucleotides 71–83 within the left and right IR of IS1071 (Fig. 3), at positions analogous to IHF binding sites near the Tn$\gamma\delta$ IR (23).

The large open reading frame within IS1071 (Fig. 2) spans 2910 bp, coding for a putative 970-amino acid polypeptide of molecular weight 108,413. This putative transposase exhibited significant homology to the class II transposases of Tn*3* and Tn*21* (Fig. 4). These alignments place greater weight on the known alignment of the Tn*3* and Tn*21* TnpA sequences. There was 21% identity of amino acids in comparing either IS1071 to Tn*3* or IS1071 to Tn*21*. Only 12% of amino acids were conserved for all three TnpA sequences. Including conservative substitutions raised similarities to 37% (with Tn*3*) or 39% (with Tn*21*).

The absence of the class II transposon signature of a 5-bp direct repeat of target DNA flanking IS1071 in pBRC60 and pBRC40 suggests that these copies of IS1071 are not the immediate products of transposition events. Indeed, this element may represent only the remnant of a once-functional class II transposon. The latter possibility was eliminated by characterizing transposition products of IS1071 in recombination-deficient mutants of *E. coli*. Plasmid pBRN4029 carrying a single copy of IS1071 formed cointegrates with the conjugative plasmid pRK2013 at a frequency of 5×10^{-4} per transconjugant in mating-out assays to a rifampicin-resistant *E. coli* HB101 recipient. Restriction mapping of these cointe-

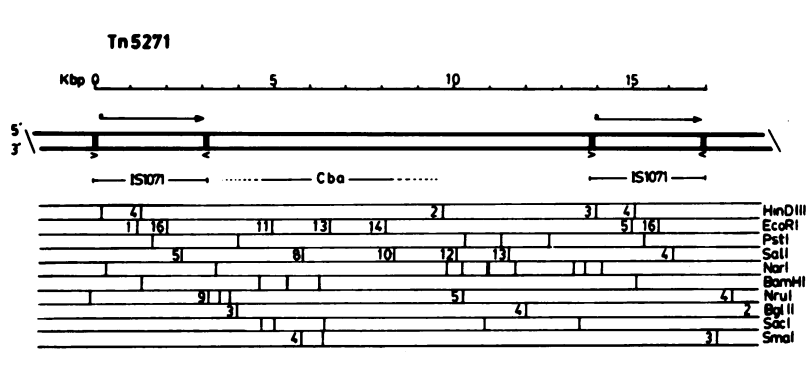

FIG. 1. Restriction enzyme digest map of the catabolic transposon Tn*5271* from *Alcaligenes* sp. strain BR60. Open reading frames within IS1071 are indicated by arrows. The positions of inverted repeat sequences are indicated by vertical bars and arrowheads. The location of the catabolic genes (—Cba—) has been determined by subcloning (unpublished data). Only cloned restriction fragments or fragments used as probes are numbered.

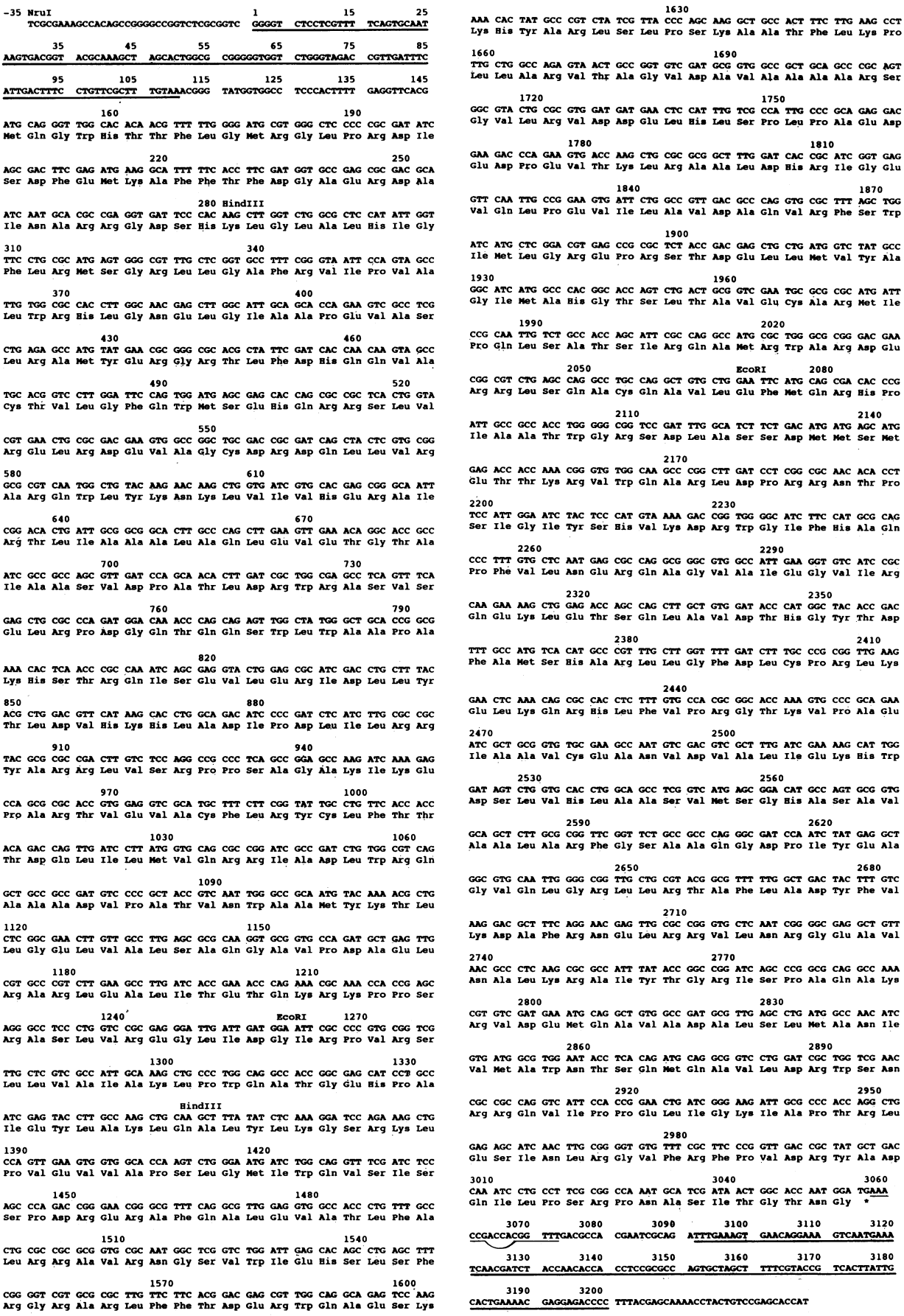

FIG. 2. Nucleotide sequence of IS1071. Numbering is relative to the first G in the IR sequence (underlined). Short sequences on either side of the insertion element are shown. The translation product of the open reading frame beginning at position 146 is shown. There is a potential ribosome binding site (GAGG) at position 135 and a stem–loop structure underlined beginning at the A in the stop codon at position 3058.

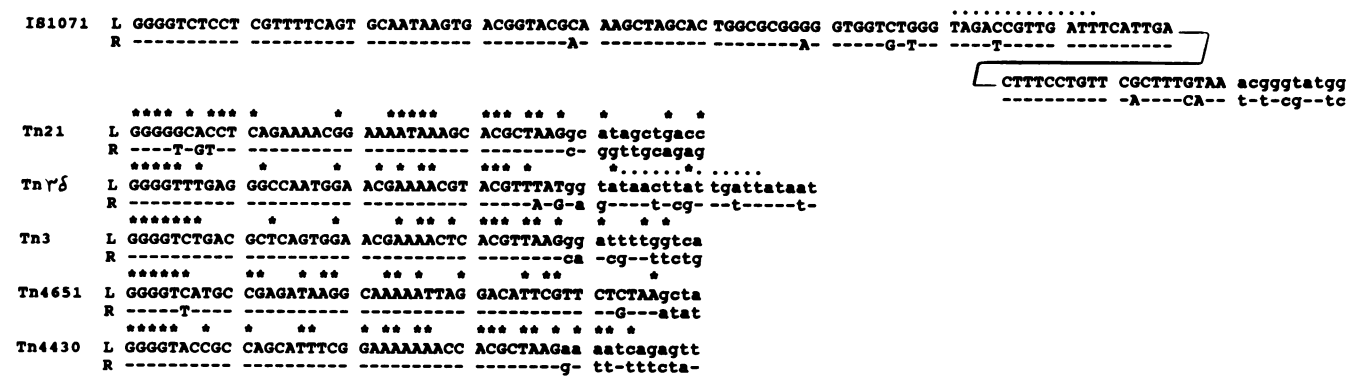

FIG. 3. Comparison of IR sequences defining the ends of class II transposable elements. L and R, left and right IRs, defined as distal and proximal to the *tnpA* gene (11). Nucleotides that differ in the right IR compared to the left are indicated. The ends of the IRs are indicated by a shift to lowercase letters. An asterisk above the sequence indicates identity at that position to the left IR of IS1071. Dotted lines mark sequences in IS1071 and Tn*γδ* homologous to the IHF binding site consensus sequence. Sources: Tn*21* and Tn*3*, GenBank; Tn*4651*, ref. 11; Tn*γδ*, ref. 23; Tn*4430*, ref. 24.

grates showed that they consist of the entire pBRN4029 replicon, flanked by direct repeats of IS1071, integrated into various sites on pRK2013 (data not shown). These cointegrates were quite stable in a *recA* host, although resolution was observed with subculturing in some cases.

DISCUSSION

The chlorobenzoate catabolic genes on plasmid pBRC60 are subject to deletion at a high frequency due to recombination between IS1071 copies on the plasmid. Recombination of the catabolic genes with the chromosome of BR60 also occurs at a high frequency (14) and is probably a result of recombination between plasmid and chromosomal IS1071 copies. This background of host-encoded recombination activity, combined with the fact that BR60 was recently isolated from an environment in which natural genetic exchange is likely to be important (15), makes it difficult to attribute the structure of Tn*5271* or the rearrangements of IS1071 in *Alcaligenes* sp.

FIG. 4. Alignment of TnpA sequences for Tn*3*, IS1071, and Tn*21*. Identical amino acids are indicated by asterisks; similar amino acids are indicated by colons. The numbers above each set of three sequence lines refer to the amino acid numbers for the IS1071 sequence. The Tn*3* sequence is 1004 and the Tn*21* sequence is 988 amino acids long. Data for the Tn*3* and Tn*21* transposase sequences were taken from GenBank.

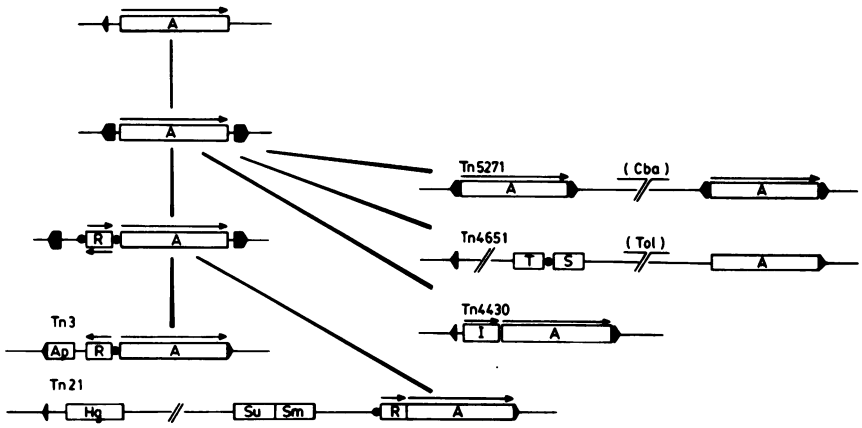

FIG. 5. Evolution of class II transposable elements. An ancestral element is indicated by a transposase gene (A) and its recognition sequence (—|—). IR sequences including the 38-bp recognition sequence are indicated by thick arrowheads (—▮—); the direction of transcription of *tnpA* (A), *tnpR* (R), and integrase (I) genes, where known, is indicated by arrows; ●, *res* site; Ap, Su, Sm, and Hg, ampicillin-, sulfonamide-, streptomycin-, and mercury-resistance loci, respectively; Cba and Tol, chlorobenzoate and toluene catabolic loci, respectively; T and S, trans-acting resolution genes for the toluene transposon.

strain BR60 to transposition events. However, we have observed transposition of the cloned IS1071 element to generate cointegrates in a recombination-deficient *E. coli*. Therefore, it would seem likely that transposition has played a part in the mobilization of catabolic genes in Tn*5271*. We are currently characterizing transposition products of both Tn*5271* and IS1071 in recombination-deficient hosts in order to define target sites.

Class II transposable elements share structural features that clearly delineate them from other mobile genetic elements (9). IS1071 has inverted repeat sequences and a transposase sequence that place it among the class II elements; however, it has no sequence homologous to class II resolvases. A search of the internal 10.6-kb sequence of the transposon (unpublished data) has failed to reveal a candidate for a resolvase-like gene. Given the complementary nature of the activities of transposase and resolvase, and the evidence that resolvase catalyzes site-specific recombination between cis-oriented *res* sites (26), it is unlikely that a resolvase gene linked to a specific transposon would exist outside of the IR defining the element. We therefore conclude that IS1071 is a true IS carrying the coding information for replicative transposition only. Our observation of the IS1071-dependent formation of stable cointegrates in a *recA* host supports this conclusion.

The transposase alignments of IS1071 with Tn*3* and Tn*21* revealed only 12% conservation of amino acids, comparable to the identity observed when Tn*3*, Tn*21*, Tn*501*, Tn*917*, Tn*2501*, and Tn*4430* sequences were all aligned (24). These transposons are therefore representative of the known evolutionary limits of the class II elements. Schmitt *et al.* (27) have proposed a scheme for the evolution of the class II transposons based on the observation that a single copy of an IR sequence and a transposase are the minimal requirements for transposition. Inverted repeats of the recognition sequence flanking the transposase may have been acquired by a process such as strand exchange during hairpin loop replication. Such an element would be structurally identical to IS1071 (Fig. 5). One prediction of this scheme is that the ancestral IRs, because of the nonspecific events occurring at a hairpin loop replication fork, may have been longer than the minimum transposase recognition sequence of 38 bp. This is the case observed with IS1071 (110-bp IR), Tn*2501* (48-bp IR), and the toluene transposon Tn*4651* (46-bp IR). Acquisition of site-specific recombination systems (resolvase or integrase), antibiotic-resistance determinants, or catabolic genes by an element like IS1071 would give rise to the class II transposons depicted in Fig. 5 (24, 27, 28).

The class II transposons have been very successful in promoting the fitness of a range of prokaryotes under very different environmental conditions. With the discovery of IS1071, there is now good evidence to suggest that the transposase gene alone confers a selective advantage on its host under some environmental conditions.

The assistance of Mary Barbara, Ellen Schwartzel, and Maysoon Salih in this work is greatly appreciated. Research was supported by grants to C.W. from the Natural Sciences and Engineering Research Council of Canada.

1. Chakrabarty, A. M. & Gunsalus, I. C. (1971) *Genetics* **68**, S10.
2. Bayley, S. A., Duggleby, C. J., Worsey, M. J., Williams, P. A., Hardy, K. G. & Broda, P. (1977) *Mol. Gen. Genet.* **154**, 203–204.
3. Hedges, R. W. & Jacob, A. F. (1974) *Mol. Gen. Genet.* **132**, 31–40.
4. Jacoby, G. A., Rogers, J. E., Jacob, A. F. & Hedges, R. W. (1978) *Nature (London)* **274**, 179–180.
5. Nakazawa, T., Hayashi, E., Yokota, T., Ebina, Y. & Nakazawa, A. (1978) *J. Bacteriol.* **134**, 270–277.
6. Chakrabarty, A. M., Friello, D. A. & Bopp, L. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3109–3112.
7. Kleckner, N. (1981) *Annu. Rev. Genet.* **15**, 341–404.
8. Sylvanen, M. (1988) in *Genetic Recombination*, eds. Kucherlapati, R. & Smith, G. R. (Am. Soc. Microbiol., Washington), pp. 331–356.
9. Sherratt, D. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 163–184.
10. Michiels, T., Cornelis, G., Ellis, K. & Grinsted, J. (1987) *J. Bacteriol.* **169**, 624–631.
11. Tsuda, M., Minegishi, K.-I. & Iino, T. (1989) *J. Bacteriol.* **171**, 1386–1393.
12. Tsuda, M. & Iino, T. (1990) *Mol. Gen. Genet.* **223**, 33–39.
13. Tomasek, P. H., Frantz, B., Sangodkar, U. M. X., Haugland, R. A. & Chakrabarty, A. M. (1989) *Gene* **76**, 227–238.
14. Wyndham, R. C., Singh, R. K. & Straus, N. A. (1988) *Arch. Microbiol.* **150**, 237–243.
15. Fulthorpe, R. R. & Wyndham, R. C. (1991) *Appl. Environ. Microbiol.* **57**, 1546–1553.
16. Yannisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
17. Gendel, S., Straus, N., Pulleyblank, D. & Williams, J. (1983) *J. Bacteriol.* **156**, 148–154.
18. Figurski, D. H. & Helinski, D. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1648–1652.
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
20. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–46.
21. Wheatcroft, R. & Watson, R. J. (1988) *Appl. Environ. Microbiol.* **54**, 574–576.
22. Hanahan, D. (1984) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Washington), p. 115.
23. Wiater, L. A. & Grindley, N. D. F. (1988) *EMBO J.* **7**, 1907–1911.
24. Mahillon, J. & Lereclus, D. (1988) *EMBO J.* **7**, 1515–1526.
25. Ichikawa, H., Ikeda, K., Amemura, J. & Ohtsubo, E. (1990) *Gene* **86**, 11–17.
26. Krasnow, M. A. & Cozzarelli, N. R. (1983) *Cell* **32**, 1313–1324.
27. Schmitt, R., Rogowsky, P., Halford, S. E. & Grinsted, J. (1985) in *Evolution of Prokaryotes*, eds. Schleifer, K. H. & Stackebrandt, E. (Academic, London), pp. 91–104.
28. Mercier, J., Lachapelle, J., Couture, F., Lafond, M., Vezina, G., Boissinot, M. & Levesque, R. C. (1990) *J. Bacteriol.* **172**, 3745–3757.