The EMBO Journal vol.7 no.5 pp.1515 - 1526, 1988

Structural and functional analysis of Tn*4430*: identification of an integrase-like protein involved in the co-integrate-resolution process

Jacques Mahillon and Didier Lereclus[1]

Plant Genetic Systems, J.Plateaustraat 22, B-9000 Gent, Belgium, and [1]Laboratoire de Biochimie Microbienne, Institut Pasteur, Rue du Docteur Roux, 28, F-75724 Paris, Cedex 15, France

Communicated by R.Dedonder

The 4149-bp transposon Tn*4430* from *Bacillus thuringiensis* is delineated by 38-bp inverted repeats and codes for a 113-kd protein that shares homology with the transposases (TnpA) of Tn*3*, Tn*21* and Tn*501*. Through transpositional recombination, this protein generates the formation of co-integrates between both donor and target replicons, with duplication of Tn*4430* molecules. These features are characteristic of transposons of the Tn*3* family (class II elements). The second step of the transposition process, the co-integrate resolution, is mediated by a 32-kd protein. This protein (TnpI) displays regional similarities with site-specific recombinases of the integrase family, such as Int of bacteriophage λ, Cre of bacteriophage P1 or TnpA and TnpB of the Tn*554* transposon. Moreover, the 250-bp sequence upstream to the *tnpI* gene contains several structural features that are reminiscent of the *att*P attachment site of phage λ. This unique association between the integrase-like TnpI recombinase and the TnpA transposase qualifies Tn*4430* as a member of a new group within the class II mobile genetic elements.

Key words: Bacillus thuringiensis/class II transposon/integrase-like resolvase/Tn*4430*

Introduction

Bacillus thuringiensis is a Gram-positive entomopathogenic bacterium active against lepidopteran, dipteran or coleopteran larvae. Its toxicity is correlated with the presence of parasporal proteinaceous crystals that contain from one to more than three different δ-endotoxins (Whiteley and Schnepf, 1986). The genes coding for these toxins are located mainly on large plasmids and, for some strains, on the chromosome (Klier *et al.*, 1982; Kronstad *et al.*, 1983; Carlton and Gonzales, 1985; Lereclus *et al.*, 1985). Recently, several groups have reported the presence of two classes of repeated sequences (RS) (1.6 and 2.2 kb in size) adjacent to certain δ-endotoxin genes (Kronstad and Whiteley, 1984, 1986; Lereclus *et al.*, 1984; Mahillon *et al.*, 1985). Although the involvement of these RS in the mobility of δ-endotoxin genes remains speculative, structural analysis of the 1.6-kb RS has revealed a family of closely related insertion sequence elements (iso-IS*231*) sharing similarities with IS*4* from *Escherichia coli* (Mahillon *et al.*, 1985). Some of these IS*231* were also found to be structurally associated with a 4.2-kb mobile element, named Tn*4430* (Mahillon *et al.*, 1987).

This Tn*4430* transposon has already been studied in some detail. It is capable of *in vivo* insertions and adjacent deletions within the conjugative plasmid pAMβ1 when introduced in *B.thuringiensis* (Lereclus *et al.*, 1983, 1984). It harbours 38-bp terminal inverted repeats (IR) homologous to those of Tn*3*, Tn*21* or Tn*917*, and produces a 5-bp duplication upon transposition in *E.coli* (Lereclus *et al.*, 1986). Taken together, these results suggest that Tn*4430* is a class II transposable element. Yet, this transposon remains cryptic since no functions other than those required for transposition have been found.

Transposition of class II mobile elements (Kleckner, 1981) proceeds in two steps: a transpositional replicative recombination followed by a site-specific conservative recombination (Heffron, 1983; Grindley and Reed, 1985). The first event requires the product of the *tnpA* gene (transposase) and the presence of both terminal IR sequences. This replicative process leads to the fusion of donor and target replicons, with two copies of the Tn, in direct orientation, at their junctions (co-integrate). The second mechanism resolves the co-integrate by a site-specific, dsDNA exchange. This conservative recombination is mediated by the resolvase protein (TnpR) at the internal resolution sites (*res*) of both Tn copies. Thus far, all class II mobile elements transpose through this double recombination mechanism, in agreement with the symmetric transposition model proposed by Shapiro (1979) and Arthur and Sherratt (1979).

In this paper, we describe the complete structural organization of Tn*4430* and a functional analysis of its transposition in *E.coli*. The results show that, like other class II elements, Tn*4430* transposes through a two-step process with the formation of a co-integrate as an intermediate product. Although the elements required for the Tn*4430* transpositional recombination (*tnpA* and IR) are similar to those of other class II transposons, the protein involved in the Tn*4430*-co-integrate resolution does not belong to the resolvase-invertase family of site-specific recombinases. Amino acid comparison between this putative site-specific recombinase and several phage integrases indeed suggests that this protein (TnpI) belongs to the integrase family of site-specific recombinases (Argos *et al.*, 1986). Evolutionary implications and relationships between Tn*4430* and other mobile elements are discussed.

Results

Structural organization of Tn4430

The *B.thuringiensis* strain H1.1 (flagellar serotype H1) harbours three small plasmids (< 15 kb), one of which (pGI2) contains a copy of the Tn*4430* transposable element. This pGI2 plasmid has been cloned into pBR322, as a single *Sal*I restriction fragment, to give the pGI200 recombinant molecule. The Tn*4430* element defined as a 4.2-kb *Kpn*I fragment, represents 45% of the pGI2 cryptic plasmid

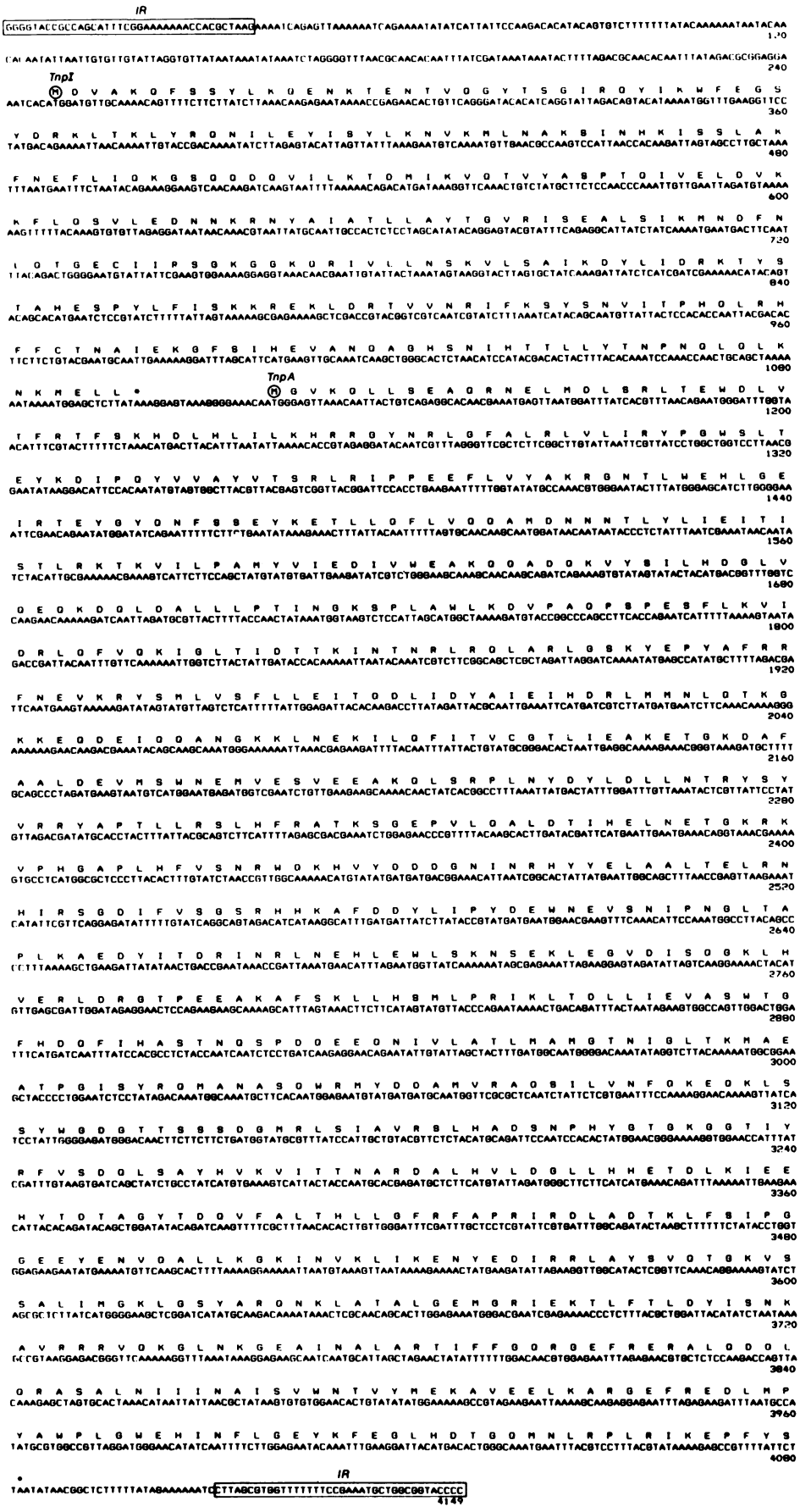


Fig. 1. Nucleotide sequence of Tn*4430*. The 4149-bp sequence of Tn*4430* contains 38-bp terminal IR, a putative resolvase gene (*tnpI*; from nucleotide 247 to 1101) and a transposase gene (*tnpA*; from nucleotide 1120 to 4083). The structural features upstream to the *tnpI* gene are described in Figure 9. In pGI2, the sequence 5′-TACAA-3′ flanks Tn*4430* in direct orientation (not shown).

(Mahillon *et al.*, 1988). The pGI200 plasmid and deletion derivatives (not shown) were used to determine the nucleotide sequence of Tn*4430* according to the method of Maxam and Gilbert (1980).

The Tn*4430* sequence (Figure 1), delineated by 38-bp IR sequences (Lereclus *et al.*, 1986) contains 4149 bp, with a G + C content of 35.2%. It is flanked by two direct repeats (DR) of 5 bp (5'-TACAA-3'). Since Tn*4430* has been shown to duplicate 5 bp upon transposition in *E.coli* (Lereclus *et al.*, 1986), the DR flanking Tn*4430* in the pGI2 plasmid most probably correspond to target DNA duplications.

A computer search for coding capacity revealed that within one frame, from nucleotide 199 to 4080, there is a single stop codon (-TAA- at position 1099) located in between two open reading frames (designated as ORF1 and ORF2, hereunder). No other ORF > 100 amino acids was detected in the five other frames. Proteins derived from these ORFs and the structural features observed in the region upstream of the first ORF (Figure 1) will be discussed later.

Tn4430 transpositional recombination

To characterize the function(s) involved in the Tn*4430* transpositional process, several insertion mutants were constructed by cloning the gene coding for a 3'5"-aminoglycoside phosphotransferase type III (APH-III gene, conferring resistance to kanamycin) into different restriction sites of the element (Table I and Figure 2). Transposition frequencies of these derivatives (pHTx), estimated through a conjugation–transposition assay using the conjugative plasmid pOX38, are summarized in Table II.

The transfer frequency of pBR322::Tn*4430* (pHT1) was estimated to be 4.6×10^{-6} per generation. This corresponds to the number of pOX38::pHT1 co-integrate molecules formed in the donor strain that were not resolved prior to conjugative transfer. It gives information not only on the transposition frequency but also on the co-integrate stability in *E.coli*. Derivative plasmids pHT3 and pHT6 have lost their capacity to transpose. These plasmids correspond to the APH-III insertions 3 and 6, both located within ORF2. However, as previously reported (Lereclus *et al.*, 1986), the derivative harbouring the APH-III insertion into the *Hpa*II site (pHT3) can be complemented with an intact copy of Tn*4430* provided *in trans* (Table II, line 8). The low transposition frequency observed in that case (as compared with pHT1) is presumably due to poor *trans*-complementation and/or to replication incompatibility between pHT3 and pHT44. All other insertions (5, 7, 8 and 9) allowed Tn*4430* derivatives to transpose, with frequencies one order of magnitude higher (pHT5, pHT7 and pHT8) or lower (pHT9) than pHT1. Although several explanations could account for

Table I. Plasmids used in this study

Plasmids	Size (kb)	Phenotype	Features[a]	References
pBT421	21	Ap,Cm	A *B.thuringiensis* berliner 1715 DNA fragment containing IS*231*, Tn*4430* and the 130 kd endotoxin gene, cloned into pHV33	Klier *et al.* (1982) Lereclus *et al.* (1984)
PGI2030 pGI2031	4	Ap	*tnpI* structural gene cloned into pLK57 under the control of λp_L (pGI2031), or not (pGI2030)	This paper
pHT1	8.5	Ap,Tc	pBR322::Tn*4430*	This paper
pHT3	10	Ap,Tc,Km	pBR322::Tn*4430*Ω3	Lereclus *et al.* (1986)
pHT5	10	Ap,Tc,Km	pBR322::Tn*4430*Ω5	This paper
pHT6	10	Ap,Tc,Km	pBR322::Tn*4430*Ω6	This paper
pHT7	10	Ap,Tc,Km	pBR322::Tn*4430*Ω7	This paper
pHT8	10	Ap,Tc,Km	pBR322::Tn*4430*Ω8	This paper
pHT9	10	Ap,Tc,Km	pBR322::Tn*4430*Ω9	This paper
pHT33	11.5	Tc,Cm	pHV33:Tn*4430*	This paper
pHT44	6.9	Ap	Tn*4430* *Kpn*I fragment cloned into pUC18, in both orientations	Lereclus *et al.* (1986)
pHT380	±57	Tra^+	pOX38::Tn*4430*	This paper
pHT385	±58.5	Km,Tra^+	pOX38::Tn*4430*Ω5	This paper
pHV33	7.3	Ap,Tc,Cm	pBR322+pC194	Primrose and Ehrlich (1981)
pLK57	3	Ap	Expression vector	Botterman and Zabeau (1987)
pOX38	±53	Tra^+	F derivative	Guyer *et al.* (1981)

[a]For constructions see Material and methods.

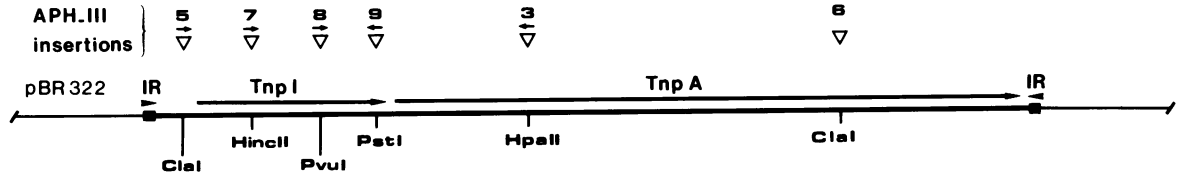


Fig. 2. Localization of the APH-III gene insertions within Tn*4430*. The APH-III gene has been inserted within different restriction sites (triangles) of the Tn*4430* element from pHT1. The resulting pHTx derivatives were designated by the number of the corresponding APH-III insertions. Small arrows refer to the APH-III gene orientation for each insertion, except for the number 6 where this orientation has not been determined. The features designated as IR, *tnpI* and *tnpA* correspond to those described in Figure 1.

Table II. Transposition frequency of the Tn*4430* derivatives[a]

Plasmids in donor strain (HB101)	Antibiotic selection for transconjugants	Transposition frequency per generation
1. pOX38 + pHT1	Ap, Tc, Rif	4.6×10^{-6}
2. pOX38 + pHT3	Km, Rif.	$<10^{-11}$
3. pOX38 + pHT5	Km, Rif.	1.4×10^{-5}
4. pOX38 + pHT6	Km, Rif.	$<10^{-11}$
5. pOX38 + pHT7	Km, Rif.	4×10^{-5}
6. pOX38 + pHT8	Km, Rif.	2.6×10^{-5}
7. pOX38 + pHT9	Km, Rif.	9.4×10^{-7}
8. pOX38 + pHT3 + pHT44	Km, Rif.	3.5×10^{-7}

[a]The transposition frequency of the Tn*4430* derivatives was determined by a mating-transposition assay between *E.coli* strains HB101 (donor) and LC916 (recipient, Rif^r), using the conjugative plasmid pOX38. This frequency is calculated by dividing the number of transconjugants by the conjugation frequency of pOX38 (4×10^{-1}) times the number of donor cell generations before selection (~ 20) and the number of donor cells.

these observations (see Discussion) these results indicate that these insertions do not inactivate a function required for transposition.

Intermolecular transpositional recombination mediated by class II elements results in the fusion of donor and target molecules with concomitant duplication of the transposable element (Heffron, 1983). To verify whether Tn*4430* generates such structures, plasmid preparations of transconjugants, presumed to contain these co-integrates (according to their resistance phenotype), were cut with *Bam*HI and hybridized to the ^{32}P-labelled *Kpn*I fragment of Tn*4430*. Since pOX38 contains two *Bam*HI sites and the unique *Bam*HI site of pHTx is located outside the Tn*4430*Ωx region (in the pBR322 moiety), duplication of Tn*4430*Ωx in the co-integrate would result in the appearance of two hybridizing bands. On the contrary, presence of a unique Tn*4430*-derivative copy within pOX38 would give a single hybridizing fragment. As shown in Figure 3, co-integrate molecules derived from the pHT7 and pHT9 transfer (lane 3 and 4 respectively) indeed contained two copies of the Tn*4430* derivatives. In both cases a third hybridizing band is also present in small amounts. This band corresponds to one of the two resolution products (pHT7 or pHT9), as revealed by their mobility similar to that of pHT3 (lane 1). The second resolution products (namely pOX38::Tn*4430*Ω7 or pOX38::Tn*4430*Ω9) cannot be seen in Figure 3 since their linear forms migrate with and are hindered by the pOX38 co-integrates. One should already note that resolution of the co-integrates ends up with a single Tn*4430* copy left within pOX38, as shown for pHT385 (lane 2).

These results strongly suggest that the transposition of Tn*4430* involves a replicative recombination step. As for Tn*3* (Heffron, 1983), Tn*4430* transposition probably takes place via a co-integrate intermediate. Moreover, one can speculate that at least one protein is required for the Tn*4430* transpositional recombination and that this protein is located within the second ORF. We therefore analysed Tn*4430* gene expression in both *E.coli* and *B.subtilis* minicells, using different plasmid constructions as shown in Figure 4.

Transformation of *B.subtilis* minicells with the pHV33 shuttle plasmid led to the expression of the chloramphenicol acetyl-transferase (*cat*) gene (lane 1). Insertion of Tn*4430* into this vector (pHT33, lane 2) gave rise to the production

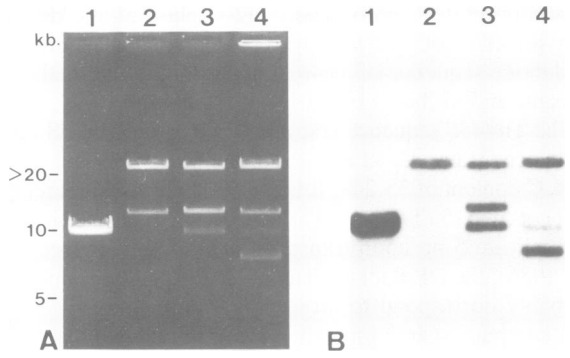


Fig. 3. Duplication of Tn*4430* derivatives, containing the APH-III gene insertion, in the co-integrate molecules formed between pOX38 and the pHT derivative plasmids. (**A**) Agarose gel electrophoresis of *Bam*H1 digested plasmid DNA. **Lane 1**, pHT3; **lane 2**, pHT385 (or pOX38::Tn*4430*Ω5); **lanes 3** and **4**, plasmid preparations of HB101 transconjugants (Ap^r, Tc^r, Km^r, Str^r) presumed to harbour a pOX38::Tn*4430*Ω7::pHT7 co-integrate and a pOX38::Tn*4430*Ω9::pHT9 co-integrate, respectively. The middle-band in lane 3 is a doublet. (**B**) Autoradiogram of the corresponding blots, hybridized with a ^{32}P-labelled *Kpn*I DNA fragment of Tn*4430*. In **lanes 3** and **4**, the hydridizing band migrating as pHT3 (**lane 1**) corresponds to the co-integrate-resolution products (pHT7 and pHT9 respectively) found in small amounts in these preparations.

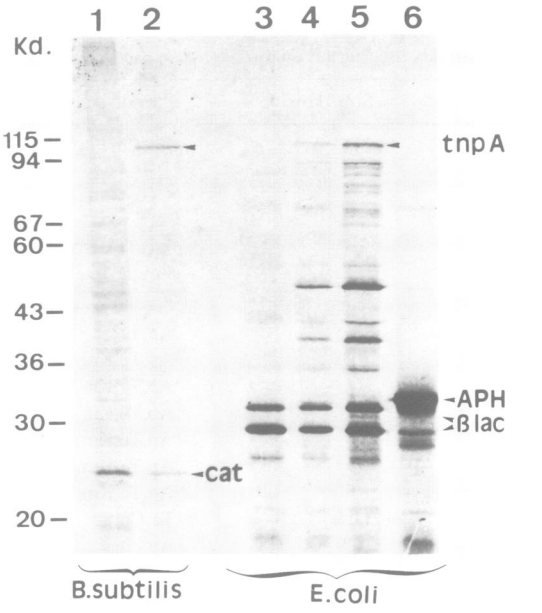


Fig. 4. Expression of the Tn*4430 tnpA* gene in *B.subtilis* and *E.coli* minicells. Autoradiogram of [^{35}S]L-methionine-labelled polypeptides from *B.subtilis* (**lanes 1** and **2**) and *E.coli* (**lanes 3–6**) minicells, resolved through a 7–15% gradient SDS–PAGE. The resident plasmids were: **lane 1**, pHV33; **lane 2**, pHT33; **lane 3**, pUC18; **lanes 4** and **5**, pHT44 containing Tn*4430* in both orientations; **lane 6**, pHT3. APH, cat and tnpA refer to the type III 3′5″-aminoglycoside phosphotransferase, the chloramphenicol acetyl transferase and the Tn*4430* transposase respectively. β-lac indicates the pre- and β-lactamase proteins.

of an additional protein with a mol. wt of ~110 000. Similarly in *E.coli*, pUC18 containing the 4.2-kb *Kpn*I fragment of Tn*4430* (plasmid pHT44), in both orientations, directed the synthesis of an additional 110-kd protein (lanes 4 and 5), as compared with *E.coli* containing the original vector (lane 3). The intermediate bands (from 35 to 95 kd)

A

	Tn*4430*	Tn*917*	Tn*501*	Tn*21*
Tn*917*	40.7	-	-	-
Tn*501*	37.6	31.8	-	-
Tn*21*	36.7	31.5	69.7	-
Tn*3*	29.2	24.9	29.5	28.6

B

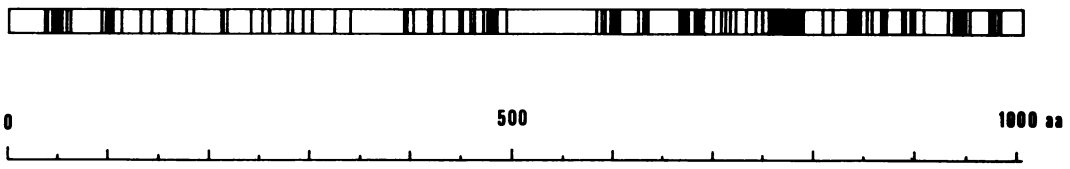


Fig. 5. Amino acid comparison of different TnpA proteins. (**A**) The TnpA protein sequences of Tn*3* (Heffron *et al.*, 1979), Tn*21* (Ward and Grinsted, 1987), Tn*501* (Brown *et al.*, 1985), Tn*917* (Shaw and Clewell, 1985) and Tn*4430* (Figure 1) were aligned so as to maximize their homology (not shown). From these comparisons, the percentage of homology was calculated for the 10 pairwise alignments. The values are given as the percentages of identical amino acids, for an average length of 1007 residues (several gaps have been introduced). Concerning Tn*917*, the protein taken as transposase in this comparison actually encompasses a 2.9-kb region of Tn*917*, including both ORF5 and ORF6 previously described by Shaw and Clewell (1985) (see further comments in the text). (**B**) Schematic representation of a consensus TnpA as defined by the amino acid comparison mentioned above. For the first 208 amino acids, the partially sequenced *tnpA* gene from Tn*2501* (Michiels *et al.*, 1987) was also taken into account. Each vertical bar corresponds to an amino acid conserved through the five (or six) TnpA proteins. Gaps introduced are not indicated.

observed with pHT44 correspond either to degradation products of the large protein or to secondary translational initiations and/or terminations. The fact that the insertion of the APH-III gene into the *Hpa*II restriction site of Tn*4430* (insertion 3, Figure 3) abolished the expression of the large protein (lane 6) confirmed that this protein originates from ORF2 and therefore is involved in the Tn*4430* co-integrate formation.

ORF2 contains two ATG codons within the first 300 nt, at positions 1120 and 1165 (Figure 1). Thirteen base pairs upstream to the former, the sequence 5′-ATAAAGGAG-3′ contains 8 bp (underlined) that are complementary to the 3′-OH terminus of the 16S rRNA of *B.subtilis* (McLaughlin *et al.*, 1981). Based on the Shine–Dalgarno model for initiation of translation (Shine and Dalgarno, 1974) this sequence can be considered as a potential ribosome binding site (RBS). No such potential RBS was found in front of the second ATG.

Taking the first ATG as the actual translation start codon, the predicted protein derived from ORF2 would contain 987 amino acids (listed in Figure 1), with a calculated mol. wt of 113 515. This prediction agrees well with the size determined for the large protein produced in *B.subtilis* and *E.coli* minicells (Figure 4).

Considering the above results, we propose the name transposase (TnpA) for this large protein. We compared the sequence of this putative transposase with those or other class II elements like Tn*3* (Heffron *et al.*, 1979), Tn*21* (Ward and Grinsted, 1987), Tn*501* (Brown *et al.*, 1985), Tn*917* (Shaw and Clewell, 1985) and the partially sequenced Tn*2501* (Michiels *et al.*, 1987). Concerning Tn*917*, the comparison was made with the amino acid sequence translated from a 2.9-kb region starting at nucleotide 2359 and ending at nucleotide 5284, according to the numbering of Shaw and Clewell (1985). This sequence encompasses both ORF5 and ORF6 previously described by Shaw and Clewell and requires several nucleotide shifts, with respect to the published sequence, to remain 'open'. Whether this large protein actually exists, rather than two smaller polypeptides, remains to be determined (D.Clewell, personal communication).

The different transposase sequences, mentioned above, were aligned so as to maximize their similarity (not shown). Based on these alignments, we deduced the pairwise percentages of common amino acids (Figure 5A) as well as the overall similarity (Figure 5B) observed between these proteins. The Tn*4430* TnpA shares similarities with all the other class II transposases. The similarities range from 40.7 down to 29.2% for Tn*917* and Tn*3* respectively. It should be noted that the level of similarity is not necessarily correlated with the evolutionary distance of their original bacterial hosts. Indeed, the transposases from the Gram-negative transposons Tn*501* and Tn*21* are more similar to those from the Gram-positive elements Tn*4430* and Tn*917* than that of Tn*3*. This could indicate that exchange of transposable elements between Gram-positive and Gram-negative bacteria can occur. Demonstration of *in vivo* transfer of genetic information between both groups (Trieu-Cuot *et al.*, 1985b) lends support to this hypothesis.

Transposase sequences share 12.7% of perfectly conserved residues. As shown in Figure 5B, the common amino acids found are not randomly distributed but fall into two conserved regions. The N-terminal domain (from amino acid 1 to 500) contains 10.4% common amino acids whereas in the C-terminal region (from amino acid 580 to the end), 17.8% amino acids are conserved. As discussed by Ward and Grinsted (1987), this observation suggests that the C-terminal domain may be involved in reactions that are common to the different transposases (DNA cleavage and ligation), while the N-terminus could participate in more

Table III. Resolution frequency of the co-integrate formed between pOX38 and the pHT derivative plasmids

Co-integrates in donor strain (HB101)	Antibiotic selection for transconjugants	Resolution frequency[a]
1. pOX38::Tn*4430*Ω3::pHT3	Km, Sp	0.042
2. pOX38::Tn*4430*Ω5::pHT5	Km, Sp	0.040
3. pOX38::Tn*4430*Ω7::pHT7	Km, Sp	<0.005
4. pOX38::Tn*4430*Ω8::pHT8	Km, Sp	<0.005
5. pOX38::Tn*4430*Ω9::pHT9	Km, Sp	<0.005

[a]This frequency corresponds to the relative number of Ap^s, Tc^s clones with respect to Km^r Sp^r transconjugants. The *E.coli* HB101 and GY203 (Sp^r) were the donor and the recipient strains respectively. All the mating-experiments were repeated twice and the average number of transconjugants tested for antibiotic resistance was 100–150.

specific reactions (recognition of the IR). This assumption has recently been confirmed by Evans and Brown (1987) for Tn*21* and Tn*501*. Using hybrid transposases, these authors have indeed shown that the IR-recognition specificity of both transposases lay within their N-terminal region, between amino acids 28 and 216.

Terminal IRs of class II transposons have been considered as the specific sites used by the transposase for the transpositional recombination (Heffron, 1983; Grindley and Reed, 1985; Wishart *et al.*, 1985). As previously shown (Lereclus *et al.*, 1986), Tn*4430* is flanked by 38-bp IR sequences (Figure 1) that are related to those of other class II elements like Tn*3* or Tn*1721*. It was also observed that the Tn*4430* IRs shared greater similarity with those of the 'Tn*21* subgroup' elements (Tn*21*, Tn*1721* and Tn*501*) than with Tn*917*. This is surprising since the putative Tn*917* transposase is the most closely related one to that of Tn*4430*. The reasons for this discrepancy remain unknown.

Tn4430 co-integrate resolution

Considering that Tn*4430*, through its transpositional recombination features (*tnpA* and IR), belongs to class II of transposable elements, one would expect that this Tn also used a site-specific, co-integrate resolution system similar to the *tnpR–res* combination of Tn*3* and Tn*21*. To investigate this issue, we analysed the stability of the co-integrate molecules generated by Tn*4430* and its APH-III insertion derivatives (Figure 2), by measuring the percentage of pBR322 co-transfer in a mating assay. Resolutions of the pOX38::Tn*4430*Ωx::pHTx molecules would indeed result in the appearance of Ap^s Tc^s Km^r transconjugants. Co-integrate molecules derived from pHT3, pHT5 and pHT7 to pHT9 were first isolated by mating their host (LC916) with a HB101 recipient strain. Selection for Ap^r Tc^r Km^r transconjugants insured the elimination of molecules already resolved in LC916 (i.e. pOX38::Tn*4430*Ωx) which would bias the resolution frequency estimates. The resulting clones were thereafter mated with strain GY203 and the number of Ap^s Tc^s clones observed amongst the Km^r Sp^r transconjugants was used to measure the resolution frequencies (Table III).

Under these conditions, co-integrates derived from pHT3 and pHT5 were resolved with an estimated frequency of 4×10^{-2}. Concerning pHT7, pHT8 or pHT9, no resolution was detected above the detection threshold of 5×10^{-3}. These results lend support to the existence of a resolution mechanism in Tn*4430*.

Therefore, we analysed the plasmid contents of the Ap^r Tc^r Km^r Str^r HB101 clones (donor strain in the last conjugation) harbouring the different pHTx co-integrates. The results confirmed that both the pHT3- and pHT5-originated co-integrates have been partially resolved (Figure 6, lanes 8 and 5 respectively). It should be noted that the large amount of pHT3 and pHT5 plasmid DNA is due to their high copy number, relative to those of the pOX38 derivatives replicons. The results also showed that although the mating experiment did not reveal any detectable resolved molecules of the pHT7 and pHT9 co-integrates, the corresponding plasmid preparations contained small amounts of these molecules (Figure 6, lanes 6 and 7). Assuming that pHT7 and pHT9 are 20–50 times less abundant than pHT3 and pHT5, their resolution frequency would range between 8×10^{-4} and 2×10^{-3}.

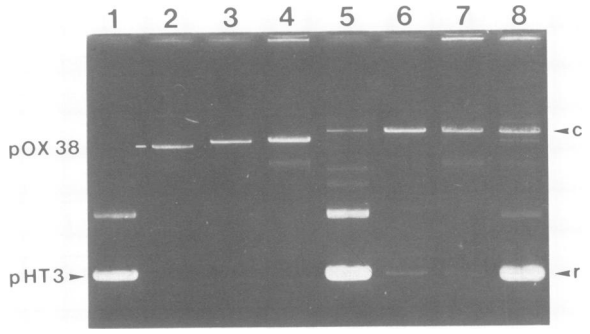


Fig. 6. Stability of the Tn*4430*Ωx co-integrates. Agarose gel electrophoresis of plasmid DNA: **lane 1**, pHT3; **lane 2**, pOX38; **lane 3**, pOX38::Tn*4430* (pHT380); **lane 4**, pOX38::Tn*4430*Ω5 isolated from an Ap^s, Tc^s, Km^r and Str^r HB101 transconjugant. Plasmids from **lanes 5–8** were isolated from Ap^r, Tc^r, Km^r and Str^r HB101 transconjugants: **lane 5**, pOX38::Tn*4430*Ω5::pHT5; **lane 6**, pOX38::Tn*4430*Ω7::pHT7; **lane 7**, pOX38::Tn*4430*Ω9::pHT9; **lane 8**, pOX38::Tn*4430*Ω3::pHT3. Symbols are: 'c' for co-integrate and 'r' for the resolution product (i.e. pHTx). The white line indicates the pOX38 plasmid. It should be noted that plasmid preparations from lanes 4, 7 and 8 were slightly contaminated with chromosomal DNA (e.g. the DNA band migrating below the pOX38 derivative in lane 4).

These figures are probably overestimates, since the antibiotics used for bacterial growth (ampicillin and tetracycline) might have preferentially selected for clones containing resolved plasmids (due to intraplasmidic homologous recombination, observed in *recA* genetic background, Fishel *et al.*, 1981; Laban and Cohen, 1981) whose copy number allowed increased expression of the antibiotic resistance gene(s).

These experimental data suggest that Tn*4430* can mediate resolution in *E.coli*, although at a low frequency. The process is hindered, at least partially, when the APH-III gene is inserted at the *Hin*cII, *Pvu*I or *Pst*I sites of Tn*4430*. Since these insertions are located within ORF1 of Tn*4430* (Figures 1 and 2), this might indicate the involvement of an ORF1-encoded protein in the resolution mechanism. Analysis of the Tn*4430* gene expression in both *E.coli* and *B.subtilis* minicells (Figure 4) did not reveal, however, such a protein (with an expected mol. wt of ~30 000, see below).

We therefore cloned a 1-kb *Fnu*DII–*Ssp*I fragment of Tn*4430* (from nucleotide 234 to 1238) into the pLK57 expression vector (Botterman and Zabeau, 1987), under the control of the λp_L promoter. The resulting plasmid, pGI2031, the vector pLK57 as well as a control recombinant pGI2030 (with the Tn*4430* fragment inserted in the opposite orientation) were used in an *E.coli in vitro* transcription–translation system. In parallel, a purified 1.1-kb *Ssp*I fragment of Tn*4430* (from nucleotide 127 to 1238) isolated from pHT1 was tested in the same experiment. As shown

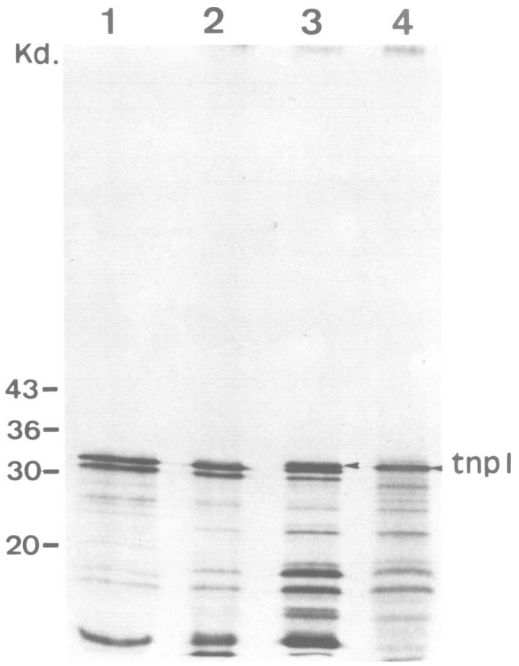


Fig. 7. Expression of the *tnpI* gene in an *E.coli in vitro* transcription–translation system. Autoradiogram of [^{35}S]L-methionine-labelled polypeptides, separated by a 7–15% gradient SDS–PAGE, specified by: **lane 1**, pLK57; **lane 2**, pGI2030; **lane 3**, pGI2031; **lane 4**, a 1.1-kb *Ssp*I fragment isolated from pHT1 and containing the *tnpI* gene. The two labelled polypeptides of ~30 kd in lanes 1, 2 and 3 correspond to the pre- and β-lactamase doublet. TnpI indicates the *tnpI* gene product.

in Figure 7, pGI2031 (lane 3) directs the expression of an additional 32-kd protein as compared to both control plasmids pLK57 and pGI2030 (lanes 1 and 2 respectively). Furthermore, transcription–translation of the purified *Ssp*I fragment led to the appearance of a protein displaying the same apparent size (Figure 7, lane 4). These results unequivocally demonstrate that the Tn*4430* ORF1 indeed codes for a protein of ~32 kd.

Examination of the Tn*4430* ORF1 (Figure 1) showed two ATG within the first 100 codons, at positions 247 and 433 of the Tn*4430* DNA sequence. Proteins starting at these initiation codons would have a calculated mol. wt of 32 782 and 25 301 respectively. In addition, only the former is preceded by a sequence (5′-GACGCGGAGG-3′) from which 5 out of 10 bases (underlined) are complementary to the 3′-OH terminus of the 16S rRNA of *B.subtilis* (McLaughlin *et al.*, 1981).

These observations argue in favour of the first ATG as actual initiation codon. The sequence of the resulting 284-amino acid polypeptide is shown in Figure 1. Its calculated mol. wt of 32 782 is consistent with the 32-kd size estimated in the *in vitro* transcription–translation system (Figure 7).

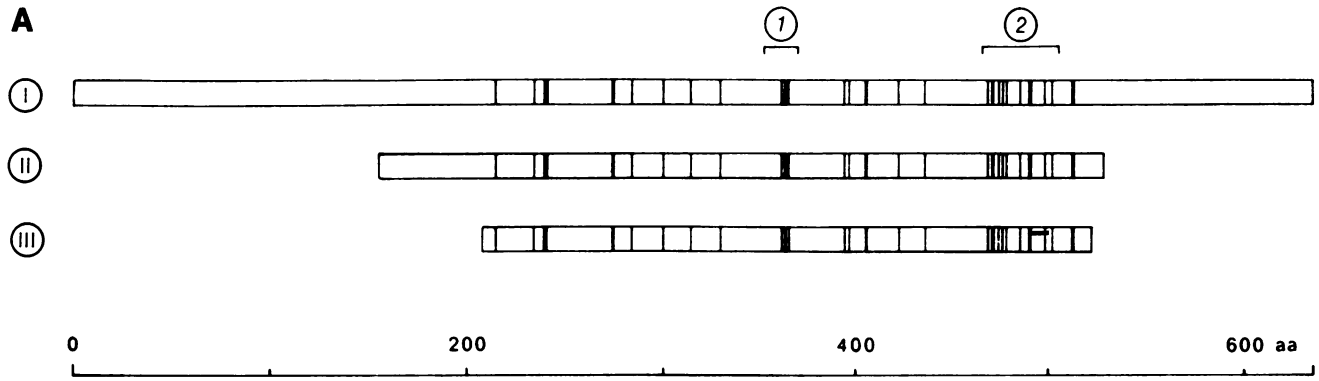
Because of the potential role of this protein as Tn*4430* site-specific recombinase, it was worth comparing its sequence with those of other class II element resolvases and with proteins involved in DNA inversion such as the Cin or Gin invertases of bacteriophages P1 and Mu. Both resolvases and invertases indeed belong to a common family of site-specific recombinases (Grindley *et al.*, 1985). Surprisingly, no similarity was found, even with those residues which are highly conserved among such recombinases. Using the FASTP program of Lipman and Pearson (1985), the NBRF library was then screened for possible protein similarity with the Tn*4430* ORF1 polypeptide. The search was unsuccessful. Therefore, we systematically compared the Tn*4430* ORF1 sequence with proteins encoded by other transposons such as the IS elements, the Tn*7* (Smith and Jones, 1986) and Tn*554* (Murphy *et al.*, 1985) transposons. This search revealed that the Tn*4430* ORF1 protein shared significant similarities only with two reading frames of the *Staphylococcus aureus* transposon Tn*554*, namely TnpA and TnpB. These frames encode proteins of 361 and 630 amino acids respectively. Although both TnpA and TnpB (as well as TnpC) proteins are required for the Tn*554* transposition (Murphy *et al.*, 1985), no precise functions have been allocated to these polypeptides.

Based on the sequence alignments (not shown), the pairwise similarities turned out to be 17.5, 24.8 and 20.7% for the Tn*554* TnpA–TnpB, Tn*4430* ORF1–Tn*554* TnpA and Tn*4430* ORF1–Tn*554* TnpB comparisons (with total lengths of 376, 314 and 314 residues respectively). The perfectly conserved amino acids (10.5%) are mainly distributed in the second half of the proteins (Figure 8A), particularly within a 40-residue C-terminal domain. A second, less-conserved region, was also found and is located in the centre of the Tn*4430* ORF1. Sequence alignments within these regions are shown in Figure 8B.

Recently, Argos *et al.* (1986) have shown that six bacteriophage integrases, the Cre protein in phage P1, and the FLP protein of the yeast 2μ plasmid are members of a new family of site-specific recombinases. Despite their large diversity, these eight proteins exhibit regional similarities, mostly in their C-terminus. Interestingly, the two most conserved regions of these integrases correspond to domains 1 and 2 found in the Tn*4430*–Tn*554* protein comparison, at the same relative position (Figure 8). Lastly, the D protein of the conjugative F plasmid, which acts as a transcriptional repressor and as a site-specific resolvase (Lane *et al.*, 1986) also turned out to belong to the integrase family, as suggested by its sharing of the two characteristic domains (Figure 8B). From the above results, one might suggest that the ORF1 protein of Tn*4430* is indeed a site-specific recombinase belonging to the family of integrases. Hence, we propose the name TnpI for this 284-amino acid protein encoded by the Tn*4430* ORF1. Based on these considerations, it should be possible to find a Tn*4430* DNA sequence that is specifically recognized as a resolution site by the TnpI. Although no experimental data are yet available, the structural features displayed by the sequence 5′ to the *tnpI* gene (Figures 1 and 9) suggest that this region could participate in DNA–protein interactions.

First, a 70-bp segment (from residue 77 to 146) contains six IR sequences that potentially can form a secondary structure, at the DNA and/or RNA level (Figure 9B). Second, two direct repeats of 16 bp are located 18 and 54 bases upstream of the *tnpI* start codon respectively. Third, the sequence 5′-CAACACAAT-3′ is reiterated four times in this 250-bp region, three times in the same orientation. Moreover, this 9-bp sequence is located both in the stem-loop structure and within the two 16-bp direct repeats. Finally, the sequence 5′-AAAATCAGA-3′ occurs twice in a 25-bp segment adjacent to the 38-bp left IR of Tn*4430*.

Several roles could be attributed to these striking features, in both recombinational and regulatory processes (see



B

Domain1

P2 (186/P22)	K	K	I	A	I	L	C	L	S	T	G	A	R	W	G	E	A	A	R	L	K	A	E	N	I
Lambda	-	L	A	M	E	L	A	V	V	T	G	Q	R	V	G	D	L	C	E	M	K	W	S	D	I
Ø80 (P4)	V	F	L	V	K	F	I	M	L	T	G	C	R	T	A	E	I	R	L	S	E	R	S	W	F
P1	T	A	G	V	E	K	A	L	S	L	G	V	T	K	L	V	E	R	W	I	S	V	S	G	V
F:D protein	K	M	L	L	A	T	L	W	N	T	G	A	R	I	N	E	A	L	A	L	T	R	G	D	F
Tn554 tnpA	K	L	I	L	M	L	M	Y	E	G	G	L	R	I	G	E	V	L	S	L	R	L	E	D	I
Tn554 tnpB	A	T	M	T	M	I	V	Q	E	C	G	M	R	I	S	E	L	C	T	L	K	K	G	C	L
Tn4430 tnpI	Y	A	I	A	T	L	L	A	Y	T	G	V	R	I	S	E	A	L	S	I	K	M	N	D	F

Domain2

P2 (186/P22)	H	A	L	R	H	S	F	A	T	H	F	M	I	N	G	G	S	I	I	T	L	Q	R	I	L	G	H	T	R	-	I	E	Q	T	M	V	Y	A	H
Lambda	H	E	L	R	S	L	S	A	-	R	L	Y	E	K	Q	I	S	D	K	F	A	Q	H	L	L	G	H	K	S	-	D	T	M	A	S	Q	Y	-	R
Ø80 (P4)	H	D	M	R	R	T	I	A	T	N	L	S	E	L	G	C	P	P	H	V	I	E	K	L	L	G	H	Q	M	-	V	G	V	M	A	H	Y	N	-
P1	H	S	A	R	V	G	A	A	R	D	M	A	R	A	G	V	S	I	P	E	I	M	Q	A	G	G	W	T	N	-	V	N	I	V	M	N	Y	I	R
F:D protein	H	T	F	R	H	S	Y	A	M	H	M	L	Y	A	G	I	P	L	K	V	L	Q	S	L	M	G	H	K	S	-	I	S	S	T	E	V	Y	T	K
Tn554 tnpA	H	M	L	R	H	T	H	A	T	Q	L	I	R	E	G	W	D	V	A	F	V	Q	K	R	L	G	H	A	H	V	Q	T	T	L	N	T	Y	V	H
Tn554 tnpB	H	A	F	R	H	T	V	G	T	R	M	I	N	N	G	M	P	Q	H	I	V	Q	K	F	L	G	H	E	S	-	P	E	M	T	S	R	Y	A	H
Tn4430 tnpI	H	Q	L	R	H	F	F	C	T	N	A	I	E	K	G	F	S	I	H	E	V	A	N	Q	A	G	H	S	N	-	I	H	T	T	L	L	Y	T	-

Fig. 8. Comparison of different proteins belonging to the integrase family of site-specific recombinases. (**A**) Conserved amino acids between the TnpA(II) and TnpB(I) proteins of Tn*554* (Murphy *et al.*, 1985) and the Tn*4430* TnpI(III). Each bar indicates the presence of an amino acid conserved through the three proteins. Gaps introduced to maximize the homologies are not indicated. Numbers 1 and 2, enclosed in open circles (top) indicate the most conserved domains within the proteins. These regions are depicted in (**B**), Amino acid sequence comparisons between several proteins of the integrase family. Domains 1 and 2 are those previously reported by Argos *et al.* (1986) for the Int proteins of phages P2, 186, P22, ϕ80, P4 and λ, and the Cre protein of phage P1. Protein sequences from the phages in parentheses are not shown since they are closely related to those described in this figure. Similar domains were found within the D protein of F plasmid (Lane *et al.*, 1986), both TnpA and TnpB of Tn*554* and the TnpI of Tn*4430*. Coordinates of the first domain are 70–94, 186–210, 351–375 and 133–157 for D protein of F, TnpA and TnpB of Tn*554* and TnpI of Tn*4430* respectively. Coordinates of the second domain are 202–239, 302–340, 465–502 and 234–270 for D protein of F, TnpA and TnpB of Tn*554* and TnpI of Tn*4430* respectively. Within each domain, the perfectly conserved amino acids are in boxes and the most conserved residues are underlined.

Discussion section). In *E.coli*, however, insertion of the APH-III gene in the *Cla*I site (position 193) downstream to the first 16-bp DR, did not significantly modify the Tn*4430* resolution frequency, as compared to the APH-III insertion within the transposase gene (pHT5 and pHT3 derivatives respectively; Figure 2 and Table III). However, a residual promoter activity associated with the APH-III fragment (P.Trieu-Cuot, personal communication) might have increas-

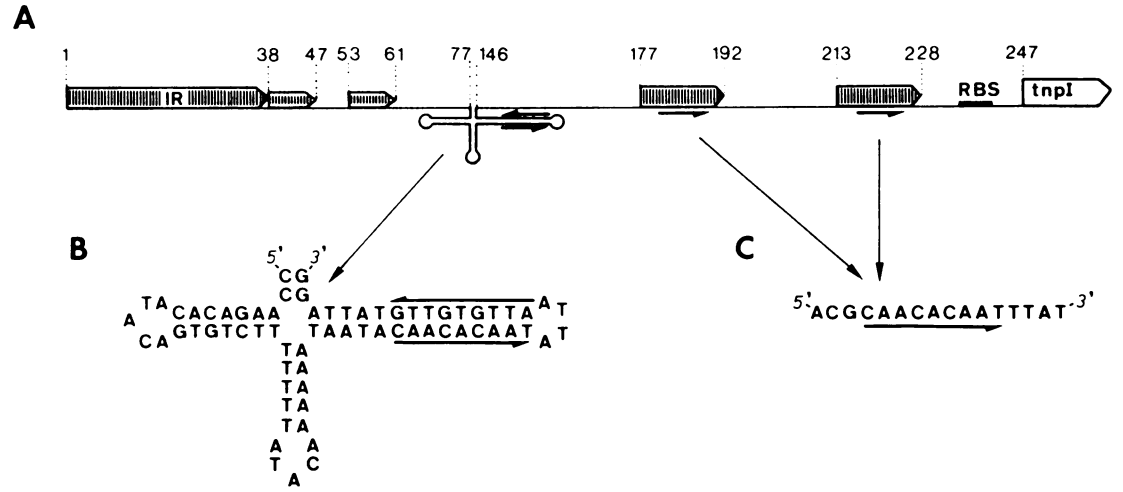


Fig. 9. Structural features of the sequence 5′ to the *tnpI* gene. (**A**) The sequence extends from nucleotide 1 to 250 (according to numbering in Figure 1). From left to right, the symbols are: large arrow with IR for the 38-bp left IR of Tn*4430*, the next two arrows for two 9-bp DR sequences (5′-AAAATCAGA-3′), the complex stem-and-loop structure for potential secondary configuration (details in B), the two arrows for the 16-bp DR sequences (detailed in C), RBS for potential ribosome binding site and the last open rectangle for the N-terminus of the *tnpI* gene. The half-arrows refer to the 9-bp sequence 5′-CAACACAAT-3′. (**B**) Possible secondary structure found between nucleotides 77 and 146 of Tn*4430*. (**C**) DNA sequence of the 16-bp DR present twice in Tn*4430*. The half arrows indicate the 5′-CAACACAAT-3′ sequences.

ed the *tnp*I gene expression which, in turn, could have counterbalanced the effects of the APH-III disruption.

Discussion

Tn4430 transposition in E.coli

Previously, we have shown that the Gram-positive transposon Tn*4430* was able to transpose in *E.coli* (Lereclus *et al.*, 1986). DNA sequence analysis at the insertion site revealed that this element was delineated by 38-bp IRs and generated 5-bp duplication at the target site. From these observations, it was postulated that Tn*4430* belongs to the Tn*3* family of transposable elements. In this paper, we describe the structural organization of Tn*4430* and, in *E.coli*, analyse the functional properties of its transposition. As is the case for other class II elements, Tn*4430* transposition proceeds in two steps. First, it generates the formation of co-integrate molecules between donor and target replicons. This mechanism is replicative (two copies of Tn*4430* are present in these co-integrates) and requires a 987-amino acid protein encoded by the element. Sequence similarity and functional analogy between this protein and other class II element transposases (TnpA) lend further support for considering this ORF2 protein as the actual Tn*4430* TnpA. Second, the Tn*4430*-originated co-integrates are unstable in *E.coli*. At low frequencies (~4%), these structures are resolved and give rise to the original replicon and the target molecule containing a new copy of the element. Similar low resolution frequencies have been observed for the mercury resistance transposon Tn*501* from *Pseudomonas* (Sherratt *et al.*, 1981).

Genetic experiments strongly suggest that the second 32-kd protein encoded by Tn*4430* is involved in this second recombination mechanism. However, examination of the plasmid preparations from transconjugants harbouring the pHTx co-integrates (Figure 6) indicates slight resolution in the case of pHT7 and pHT9 (lanes 6 and 7), although these events were not detected through the subsequent mating experiment (Table III). As we have already discussed, these resolutions might have occurred through homologous recombination between both Tn*4430*Ωx copies present in the co-integrate molecules.

Concerning the transcriptional regulation of the Tn*4430* genes, no obvious promoter-like sequences that match with known σ-specific consensus sequences of *E.coli* or *B.subtilis* promoters have been found. Nuclease S1 protection studies, using *E.coli* or *B.thuringiensis* RNA, would determine initiation and termination sites of the Tn*4430* transcript(s) and consequently, indicate whether both genes share the same mRNA molecules. In that regard, the low transposition frequency observed for pHT9 (Figure 2 and Table II) could result from insertion of the APH-III gene in a transcription regulatory sequence of the *tnpA* gene.

TnpI, a new type of transposon resolvase?

Class II elements are known to resolve the co-integrate intermediates through a conservative site-specific recombination mechanism mediated by a 20-kd resolvase acting on an internal resolution site (*res*) (Heffron, 1983). The most striking result observed with Tn*4430* is that although it displays homologies with transposons of the Tn*3* family, the 32-kd protein involved in its co-integrate resolution does not belong to the resolvase family of site-specific recombinases. In fact, the regional similarities observed between this protein and the newly described family of integrases (Argos *et al.*, 1986) suggest that this Tn*4430* protein belongs to this second group of site-specific recombinases. We propose the name TnpI in order to distinguish this putative Tn*4430* site-specific recombinase from those of the resolvase−invertase family.

Based on the properties displayed by the other proteins of the integrase family, particularly with the λ Int system, and because of the structural features observed in the region 5′ to the *tnpI* gene (Figure 9), we would like to propose a model of DNA−protein interactions for the regulation of Tn*4430* resolution.

Recently, it has been demonstrated that the operator of a *B.subtilis* phage φ105 repressor consists of three direct repeat sequences of 14 bp (Van Kaer *et al.*, 1987). It was also shown that only one of these sequences is required to give repression, albeit with lower efficiency. Since the two

16-bp DR of Tn*4430* (Figure 9A and C) are located in front of the *tnpI* gene, one can speculate that these DR act as operator sequences in the *tnpI* gene regulation.

Moreover, the 9-bp sequnence 5′-CAACACAAT-3′ located in those DR is reiterated twice in the complex secondary structure potentially formed between nucleotide 77 and 146 of Tn*4430* (Figure 9B). In the same region, between nucleotides 95 and 110, the sequence 5′-CTTTTTTTATAC-AAA-3′ contains 12 bp (underlined) that match to the 15-bp core of the λ*att*P site (Campbell, 1983). Assuming that the TnpI protein specifically recognizes the 9-bp sequence 5′-CAACACAAT-3′, protein−DNA interactions at this sequence might lead to both regulation of the *tnpI* gene expression (through the 16-bp DR) and to the formation of a complex required for site-specific recombination.

Finally, the consensus sequence of the known binding sites for the *E.coli* DNA-binding protein (called integration host factor or IHF) which is required for λ site-specific recombination, is found to overlap the 3′-end of both 16-bp DR. This could suggest the involvement of a host protein (in this case *B.thurigiensis*) in the resolution process. On the other hand, recognition of these bindings sites by the *E.coli* IHF might interfere with the Tn*4430* resolution process and account for the low resolution frequency observed in this host.

Although this model is attractive, the above assumptions remain to be confirmed. In that regard, development of an *in vitro* assay with the purified TnpI protein would be desirable. It would not only give information on the localization of the recombination site, but it would also provide a suitable system for testing the possible involvement of *B.thuringiensis* host factor(s) in the recombination process. Experiments designed to over-express and purify the Tn*4430* TnpI protein are under progress.

Evolutionary implications

To our knowledge, Tn*4430* is the first class II transposable element harbouring an integrase-like protein as putative resolvase. However, it would not be surprising that similar association will be observed in bacteria other than the Gram-positive *B.thuringiensis*. In that regard, the similarities observed between the Tn*4430* TnpI and its potential recombination site and the λ*att*−Int system is striking. A few years ago, Campbell (1981, 1983) speculated that the λ*att*P region (including both *int* and *xis* genes) are possibly originated from an ancestral transposon which he named Tnλ. According to his model, Campbell has postulated that *int* and *xis* genes were derived from a Tnλ parental gene (*tin*) that would have displayed co-integrate-resolution activities. The existence, in Tn*4430*, of such an 'integrase-like' resolvase belonging to the same family as λ*int* lends support to this evolution model of λ.

The *Staphylococcus aureus* transposon Tn*554* has several unique features: it does not bear terminal repeats, it does not generate target duplication upon transposition and transposes in a specific site of the *S.aureus* chromosome (Murphy and Löfdahl, 1984). The gene products of *tnpA, tnpB* and *tnpC* are all required for the Tn*554* transposition (Murphy *et al.*, 1985). Although this transposition behaviour of Tn*554* is strikingly different from that of Tn*4430*, both Tn*554* TnpA and TnpB proteins share similarities with the Tn*4430* TnpI, as well as with the conserved regions of the site-specific recombinases of the integrase family. Genetic and biochemical data on the precise role and function of these Tn*554* proteins will be of considerable value in attempting to understand the biological significance of these relationships.

Finally, the observation that the F plasmid D protein, which resolves miniF dimers (Lane *et al.*, 1986), is also presumably a member of the integrase family, indicates that the site-specific recombinases of this group are not restricted to phage or transposable element systems but may be part of the plasmid regulatory functions.

Materials and methods

Bacterial strains and growth conditions

The *E.coli* stains HB101 (*recA str*; Boyer and Roulland-Dussoix, 1969), LC916 (*recA rif*; Chandler and Galas, 1983) and GY203 (*recA Sp*; Morand and Devoret, 1977) were used in the mating experiments. *In vitro* plasmid constructions were achieved in strain HB101 with the exception of the pGI2030 and pGI2031 recombinant plasmids for which the *E.coli* strain K514 (Zabeau and Stanley, 1982) was used. Minicells were prepared from the strains *E.coli* AR1062 (Rambach and Hogness, 1977) and *B.subtilis* CU403 (Reeve *et al.*, 1973). The *E.coli* strains were grown at 37°C in Luria broth (Miller, 1972). Cultures for mating experiments were grown in brain heart infusion (BHI). Antibiotic concentrations for bacterial selection were: ampicillin, 100 μg/ml; chloramphenicol, 20 and 5 μg/ml for *B.subtilis* and *E.coli* strains respectively; rifampicin, 100 μg/ml; spectinomycin, 100 μg/ml; streptomycin, 100 μg/ml; and tetracycline, 5 μg/ml.

In vivo plasmid constructions

pHT380. The *E.coli* strain HB101 (pOX38) was first transformed with pHT421 (Klier *et al.*, 1982; Lereclus *et al.*, 1984) which contains an intact copy of Tn*4430*. A HB101 (pOX38 + pHT421) transformant was then mated with strain LC916, in order to obtain the conjugative co-integrate replicon pOX38::pHT421 through transposition of Tn*4430* into pOX38. The resolution product of this co-integrate (i.e. pOX38::Tn*4430*) was obtained by mating a Rif^r Tc^r Cm^r LC916 transconjugant with HB101. Among eight Str^r Tc^s Cm^s clones examined after conjugation, one (pHT380) contained a plasmid corresponding to pOX38::Tn*4430* as identified by its electrophoretic mobility in agarose gel (Figure 6, lane 3) and by hybridization with the ^{32}P-labelled *Kpn*I fragment of Tn*4430* (not shown).

pHT385. This plasmid was isolated from an Ap^s Tc^s Km^r Rif^r transconjugant resulting from the mating of LC916 (pOX38::Tn*4430*Ω5::pHT5) to the HB101 recipient strain. It corresponds to pOX38::Tn*4430*Ω5 as determined by its electrophoretic mobility (Figure 6, lane 4) and by hybridization with the ^{32}P-labelled *Kpn*I fragment of Tn*4430* (Figure 3, lane 2).

pHT1 and pHT33. pBR322 was introduced by transformation into strain HB101 containing pHT380. After mating the transformants with LC916, transconjugants were selected on rifampicin and tetracycline so as to isolate those containing the co-integrate replicon pHT380::pBR322. The resolution product of this co-integrate (i.e. pBR322::Tn*4430* or pHT1) was subsequently isolated through transformation of HB101 with the plasmid of LC916 and selection for tetracycline. The same protocol was used with the shuttle vector pHV33 and yielded the pHT33 plasmid. pHT1 confers resistance to both ampicillin and tetracycline whereas pHT33, in which Tn*4430* has transposed into the *bla* gene, is Tc^r and Cm^r.

In vitro plasmid constructions

pGI2030 and pGI2031. The pGI200 recombinant plasmid containing Tn*4430* (Mahillon *et al.*, 1988) was double-digested with *Fnu*DII and *Ssp*I and electrophoresed in a 0.7% agarose gel. A 1005-bp fragment (from nucleotide 234 to 1237, according to numbering in Figure 1) corresponding to the structural gene of *tnpI* plus its putative RBS, was eluted and ligated into the *Eco*RV restriction site of the pLK57 expression vector (Botterman and Zabeau, 1987), downstream to the λP_L promoter. Two types of recombinant plasmids were obtained: in pGI2031, the *tnpI* gene lies under the control of the lambda promoter whereas in pGI2030, it is in the inverted orientation.

Disruption plasmids pHT5 to pHT9. The 1.5-kb *Cla*I fragment from pAT21 (Trieu-Cuot and Courvalin, 1983) containing the APH-III gene was used to give Tn*4430* a genetic marker and to disrupt its DNA sequence at different places. The localization and/or the orientation of the APH-III insertions

(Figure 2) were verified through appropriate digestions with different restriction endoucleases.

pHT5 and pHT6. pHT1 was partially digested with *Cla*I and the DNA band corresponding to the linear pHT1 plasmid was recovered through agarose gel electrophoresis and ligated with the 1.5-kb *Cla*I fragment from pAT21. Among the Ap^r Km^r Tc^r recombinant plasmids obtained in strain HB101, pHT5 and pHT6 displayed the APH-III insertion at positions 193 and 3241 of Tn*4430* respectively.

pHT7. pHT1 was partially digested with *Hinc*II and the linear plasmid was purified on an agarose gel. The pAT21 *Cla*I fragment was filled in with the DNA polymerase I Klenow fragment and ligated to the linearized pHT1. The resulting plasmid pHT7 harboured the APH-III insertion at position 512 of Tn*4430*.

pHT8. pHT1 was partially digested with *Pvu*I and blunted with the Klenow fragment. The pAT21 *Cla*I fragment was filled in with the Klenow fragment. Both molecules were purified through agarose gel electrophoresis and ligated to yield pHT8. This recombinant plasmid contains the APH-III insertion at position 824 of Tn*4430*.

pHT9. The purified 1.5-kb *Cla*I fragment from pAT21 was filled in and ligated to a *Pst*I linker. After redigestion with the *Pst*I endonuclease, the fragment was purified through agarose gel electrophoresis and ligated with pHT1 partially digested with *Pst*I. Among the recombinant plasmids, pHT9 contained the APH-III insertion at position 1074 of Tn*4430*.

General techniques

The basic recombinant DNA procedures are described in Maniatis *et al.* (1982). The *E.coli* and *B.subtilis* strains were transformed according to the protocols described by Lederberg and Cohen (1974) and Anagnostopoulos and Spizizen (1961) respectively. Conjugations between *E.coli* strains were performed in BHI medium as previously reported (Lereclus *et al.*, 1986). In these conditions the number of generations of the donor cells before conjugation is ~20.

The *E.coli in vitro* transcription–translation system (Zubay, 1980) and *E.coli* minicells procedure (Rambach and Hogness, 1977) were used as described. Protocol for *B.subtilis* minicells preparation (Shivakumar *et al.*, 1979) was used with the following modifications. The medium used for bacterial growth (at 30°C) was: 1% bactotryptone, 1% nutrient broth, 0.5% yeast extract, 1% glucose, 0.5% NaCl, 10 mM $MgSO_4$, in a 100 mM phosphate buffer (pH7).

Transposition assays

The APH-III gene which confers resistance to kanamycin in *E.coli* and *B.subtilis* (Trieu-Cuot *et al.*, 1985a) was inserted at different sites of Tn*4430* on the pHT1 plasmid, leading to the derivatives pHT5 to pHT9 (for construction, see above). The pHT3 plasmid construction has previously been reported (Lereclus *et al.*, 1986). The generic name pHTx is used in the text to designate these derivatives.

These plasmids were introduced into the HB101 strain harbouring the conjugative F plasmid derivative pOX38 (Guyer *et al.*, 1981) to test, at first, the effect of the APH-III insertions on the transposition frequency of Tn*4430*. This frequency was estimated as follows: the HB101 strain containing pOX38 plus one of the pHT derivatives was mated with the LC916 recipient strain. The transcojugants were selected on LB medium containing kanamycin and rifampicin. Since the pHT derivatives cannot mediate autonomous transfer and since pOX38 is not able to mobilize those derivatives, any Km^r transconjugant should contain a pOX38 into which Tn*4430* has transposed. The transposition frequency per generation was determined as the number of Km^r transconjugants divided by the number of donor cells times the number of cell generations prior to selection (~20) and the conjugation frequency of pOX38 (4×10^{-1}). The Km^r transconjugants contained either a co-integrate molecule between pOX38 and the pHTx derivative or one of the resolution products, namely pOX38::Tn*4430*Ωx.

In the second step, the Tn*4430*Ωx generated co-integrates (pOX38::Tn*4430*Ωx::pHTx) resulting from the first mating experiment and conferring resistance to kanamycin, ampicillin and tetracycline to LC916 were isolated in the HB101 strain through a second mating. The effects of the APH-III insertions on the resolution frequency of the co-integrates were then determined by mating HB101 (pOX38::Tn*4430*Ωx::pHTx) with the GY203 strain. The transconjugants were selected on spectinomycin kanamycin LB medium and a few hundred of these clones were tested for their resistance to both ampicillin and tetracycline. The Tn*4430* resolution frequency was estimated by the number of Ap^s Tc^s colonies observed among the Km^r transconjugants tested.

Enzymes, media and chemicals

Restriction enzymes, T4 DNA ligase, bacterial alkaline phosphatase, T4 polynucleotide kinase, DNase I and DNA polymerase I (Klenow fragment) were obtained from either Boehringer Mannheim, New England Biolabs or Worthington. The *Pst*I-polylinker was purchased from New England Biolabs. Bacteriological supplies were purchased from Difco and antibiotics from Sigma.

Computer analysis

Computer analyses of DNA and protein sequences were made as previously described (Mahillon *et al.*, 1985). The Lipman and Pearson (1985) FASTP program was also used to search for protein similarities in the NBRF (National Biomedical Research Foundation) library.

Acknowledgements

We are deeply indebted to Josef Seurinck for his invaluable technical assistance in the DNA sequencing and to Ghislaine Menou for cooperation and excellent technical assistance. We are grateful to Professor J.Delcour and Dr M.-M.Lecadet for helpful discussions and encouragement and to Professor R.Dedonder in whose laboratory part of this work was conducted. We thank Guy Cornelis for communicating his results to us before publication. We also acknowledge Monique Van Nijlen and Hannelore Bombay for typing the manuscript. D.L. was supported by a grant from INRA.

References

Anagnostopoulos,C. and Spizizen,J. (1961) *J. Bacteriol.*, **81**, 741–746.
Argos,P., Landy,A., Abremski,K., Egan,J.B., Haggard-Ljungquist,E., Hoess,R.H., Kahn,M.L., Kalionis,B., Narayana,S.V.L., Pierson,L.S.,III, Sternberg,N. and Leong,J. (1986) *EMBO J.*, **5**, 433–440.
Arthur,A. and Sherratt,D. (1979) *Mol. Gen. Genet.*, **175**, 267–274.
Botterman,J. and Zabeau,M. (1987) *DNA*, **6**, 583–591.
Boyer,H. and Roulland-Dussoix,D. (1969) *J. Mol. Biol.*, **41**, 459–472.
Brown,N.L., Winnie,J.N., Fritzinger,D. and Pridmore,R.D. (1985) *Nucleic Acids Res.*, **13**, 5657–5669.
Campbell,A. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 1–9.
Campbell,A. (1983) In Shapiro,J.A. (ed.), *Mobile Genetic Elements.* Academic Press, New York, pp. 65–103.
Carlton,B.C. and Gonzales,J.M.,Jr (1985) In Hoch,J.A. and Setlow,P. (eds), *Molecular Biology of Microbial Differentiation.* American Society of Microbiology, Washington, DC, pp. 246–252.
Chandler,M. and Galas,D.J. (1983) *J. Mol. Biol.*, **170**, 61–91.
Evans,L.R. and Brown,N.L. (1987) *EMBO J.*, **6**, 2849–2853.
Fishel,R.A., James,A.A. and Kolodner,R. (1981) *Nature*, **194**, 184–186.
Grindley,N.D.F. and Reed,R.R. (1985) *Annu. Rev. Biochem.*, **54**, 863–896.
Grindley,N.D.F., Newman,B.J., Wiater,L.A. and Falvey,E.E. (1985) In Simon,M. and Herskowitz,I. (ed.), *Genome Rearrangement.* Alan R.Liss, NY, pp. 77–91.
Guyer,M.S., Reed,R.R., Steiz,J.A. and Low,K.B. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 135–140.
Heffron,F. (1983) In Shapiro,J.A. (ed.), *Mobile Genetic Elements.* Academic Press, New York, pp. 223–260.
Heffron,F., McCarthy,B.J., Ohtsubo,H. and Ohtsubo,E. (1979) *Cell*, **18**, 1153–1163.
Kleckner,N. (1981) *Annu. Rev. Genet.*, **15**, 341–404.
Klier,A., Fargette,F., Ribier,J. and Rapoport,G. (1982) *EMBO J.*, **1**, 791–799.
Kronstad,J.W. and Whiteley,H.R. (1984) *J. Bacteriol.*, **160**, 95–102.
Kronstad,J.W. and Whiteley,H.R. (1986) *Gene*, **43**, 29–40.
Kronstad,J.W., Schnepf,H.E. and Whiteley,H.R. (1983) *J. Bacteriol.*, **154**, 419–428.
Laban,A. and Cohen,A. (1981) *Mol. Gen. Genet.*, **184**, 200–207.
Lane,R., de Feyter,R., Kennedy,M., Phua,S.-H. and Semon,D. (1986) *Nucleic Acids Res.*, **14**, 9713–9728.
Lederberg,E.M. and Cohen,S.N. (1974) *J. Bacteriol.*, **119**, 1072–1074.
Lereclus,D., Menou,G. and Lecadet,M.-M. (1983) *Mol. Gen. Genet.*, **191**, 307–313.
Lereclus,D., Ribier,J., Klier,A., Menou,G. and Lecadet,M.-M. (1984) *EMBO J.*, **3**, 2561–2567.
Lereclus,D., Lecadet,M.-M., Klier,A., Ribier,J., Rapoport,G. and Dedonder,R. (1985) *Biochimie*, **67**, 91–99.
Lereclus,D., Mahillon,J., Menou,G. and Lecadet,M.-M. (1986) *Mol. Gen.*

Genet., **204**, 52–57.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Mahillon,J., Seurinck,J., Van Rompuy,L., Delcour,J. and Zabeau,M. (1985) *EMBO J.*, **4**, 3895–3899.
Mahillon,J., Seurinck,J., Delcour,J. and Zabeau,M. (1987) *Gene*, **51**, 185–194.
Mahillon,J., Hespel,F., Pierssens,A.-M. and Delcour,J. (1988) *Plasmid*, in press.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, a Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
McLaughlin,J.R., Murray,C.L. and Rabinowitz,J.C. (1981) *J. Biol. Chem.*, **256**, 11283–11291.
Michiels,T., Cornelis,G., Ellis,K. and Grinsted,J. (1987) *J. Bacteriol.*, **169**, 624–631.
Miller,J. (1972) *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Morand,P. and Devoret,R. (1977) *Mol. Gen. Genet.*, **157**, 61–67.
Murphy,E. and Löfdahl,S. (1984) *Nature*, **307**, 292–294.
Murphy,E., Huwyler,L. and do Carmo de Freire Bastos,M. (1985) *EMBO J.*, **4**, 3357–3365.
Primrose,S.B. and Ehrlich,S.D. (1981) *Plasmid*, **6**, 193–201.
Rambach,A. and Hogness,D.S. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5041–5045.
Reeve,J.N., Mendelson,N.H., Coyne,S.I., Hallock,L.L. and Cole,R.M. (1973) *J. Bacteriol.*, **114**, 860–873.
Sharpiro,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1933–1937.
Shaw,J.H. and Clewell,D.B. (1985) *J. Bacteriol.*, **164**, 782–796.
Sherratt,D., Arthur,A. and Burke,M. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 275–281.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
Shivakumar,A.G., Hahn,J. and Dubnau,D. (1979) *Plasmid*, **2**, 279–289.
Smith,G.M. and Jones,P. (1986) *Nucleic Acids. Res.*, **14**, 7915–7927.
Trieu-Cuot,P. and Courvalin,P. (1983) *Gene*, **23**, 331–341.
Trieu-Cuot,P., Klier,A. and Courvalin,P. (1985a) *Mol. Gen. Genet.*, **198**, 348–352.
Trieu-Cuot,P., Gerbaud,G., Lambert,T. and Courvalin,P. (1985b) *EMBO J.*, **4**, 3583–3587.
Van Kaer,L., Van Montagu,M. and Dhaese,P. (1987) *J. Mol. Biol.*, **197**, 55–67.
Ward,E. and Grinsted,J. (1987) *Nucleic Acids Res.*, **15**, 1799–1806.
Whiteley,H.R. and Schnepf,H.E. (1986) *Annu. Rev. Microbiol.*, **40**, 549–576.
Wishart,W.L., Broach,J.R. and Ohtsubo,E. (1985) *Nature*, **314**, 556–558.
Zabeau,M. and Stanley,K.K. (1982) *EMBO J.*, **1**, 1217–1224.
Zubay,G. (1980) *Methods Enzymol.*, **65**, 856–877.

Received on December 21, 1987; revised on February 19, 1988