The EMBO Journal vol.8 no.8 pp.2425 - 2433, 1989

Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn*1545*: homologies with other site-specific recombinases

Claire Poyart-Salmeron, Patrick Trieu-Cuot[1], Cécile Carlier and Patrice Courvalin

Unité des Agents Antibactériens, CNRS UA 271, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cédex 15, France

[1]To whom reprint requests should be addressed

Communicated by M.Hofnung

Excision is probably the initial and rate-limiting step of the movements of conjugative transposons of Gram-positive bacteria such as Tn*916* and Tn*1545*. We have shown, by molecular cloning and DNA sequencing, that a 2058 bp *Sau*3A right-junction fragment of transposon Tn*1545* specifies two gene products that are involved in the excision of the element. The DNA sequence of these genes, designated *orf1* and *orf2*, has been determined and the corresponding proteins, ORF1 and ORF2, have been identified in a bacterial cell-free coupled transcription-translation system. These proteins are freely diffusible since they are able to *trans*-complement *in vivo* a deletion derivative of Tn*1545* defective for excision. Using an *in vivo* complementation assay, we have demonstrated that ORF2 alone is able to catalyse excision and that ORF1 strongly stimulates the activity of ORF2. We also found that ORF1 and ORF2 display local homology with, respectively, proteins Xis and Int from lamboid phages, which suggests that these excision systems have a common origin. Based on the functional properties of the integrase of bacteriophage λ, on the analysis of the nucleotide sequence of the junction fragments and of the target before insertion and after excision, a model is proposed for ORF2-catalysed excision of Tn*1545* and related conjugative transposons.

Key words: conjugative transposon/excision/site-specific recombinases/*Streptococcus pneumoniae*/Tn*1545*

Introduction

The conjugative shuttle transposon Tn*1545*, originally detected in the chromosome of *Streptococcus pneumoniae* BM4200, confers resistance to kanamycin and structurally related aminoglycosides (*aphA-3*), to macrolide-lincosamide-streptogramin B type (MLS) antibiotics (*ermAM*) and to tetracycline (*tetM*) (Courvalin and Carlier, 1986). This 25.3 kb element is self-transferable to a large variety of Gram-positive bacteria where it is able to transpose to various sites and to induce mutations upon insertion (Courvalin and Carlier, 1986; Caillaud and Courvalin, 1987). Tn*1545* also transposes after cloning in *Escherichia coli* but conjugal transfer does not seem to occur (Courvalin and Carlier, 1987). Based on restriction analysis and functional properties, Tn*1545* appears to be related to the conjugative elements Tn*916*, Tn*918* and Tn*925* of *Enterococcus faecalis* (Franke and Clewell, 1981; Clewell *et al.*, 1985; Christie *et al.*, 1987) and to Tn*919* from *Streptococcus sanguis* (Fitzgerald and Clewell, 1985). These transposons confer resistance to tetracycline only (*tetM*) and their size varies from 16.4 kb (Tn*916*) to 23 kb (Tn*919*).

The nucleotide sequence of the extremities of Tn*1545* are almost identical to those of Tn*916* for at least 250 bp (Caillaud and Courvalin, 1987; Clewell *et al.*, 1988). Unlike most transposons, these elements are not flanked by terminal inverted repeated sequences, possess variable base pairs at their extremities and do not generate a duplication of the target DNA upon insertion. Another unusual property of Tn*1545* and Tn*916* is their ability to excise precisely from the target DNA (i.e. the targets are identical before insertion and after excision) in Gram-positive and Gram-negative hosts devoid of homologous recombination system. In the case of Tn*916*, however, the target after excision can differ from its original sequence by 3 bp (Clewell *et al.*, 1988). An excision-insertion mechanism involving a free non-replicative circular intermediate has been proposed for the movements (i.e. transposition and conjugation) of Tn*916* and related transposons (Gawron-Burke and Clewell, 1982, 1984; Clewell *et al.*, 1988). The model suggests that, following excision, the circular intermediate could undergo intracellular transposition to a new site, could be conjugatively transferred to a new host where it transposes, or could be lost from the progeny during cell division. In support of this model, a covalently closed circular form of Tn*916* produced *in vivo* and retaining the ability to undergo transposition has been recently identified (Scott *et al.*, 1988).

From these data it is clear that a better understanding of the migration of the conjugative transposons requires further knowledge on the excision process. Investigation of the effects of mutational insertions of Tn*5* into Tn*916* has revealed that a 1 kb-long region located near one end of the transposon was essential for excision (Senghas *et al.*, 1988). In this paper we report that the corresponding region in Tn*1545* encodes two gene products that are required for excision of this element. The DNA sequence of these genes, designated *orf1* and *orf2*, has been determined and the corresponding proteins, ORF1 and ORF2, were identified in a bacterial cell-free coupled transcription-translation system. These proteins are freely diffusible since they are able to *trans*-complement *in vivo* a deletion derivative of Tn*1545* defective for excision. We also demonstrate that ORF2 alone is able to catalyse excision and that ORF1 strongly stimulates the activity of ORF2. Finally, we show that ORF1 and ORF2 display local homology with Xis and Int proteins from lambdoid phages respectively, which suggests that these excision systems derive from a common ancestor.

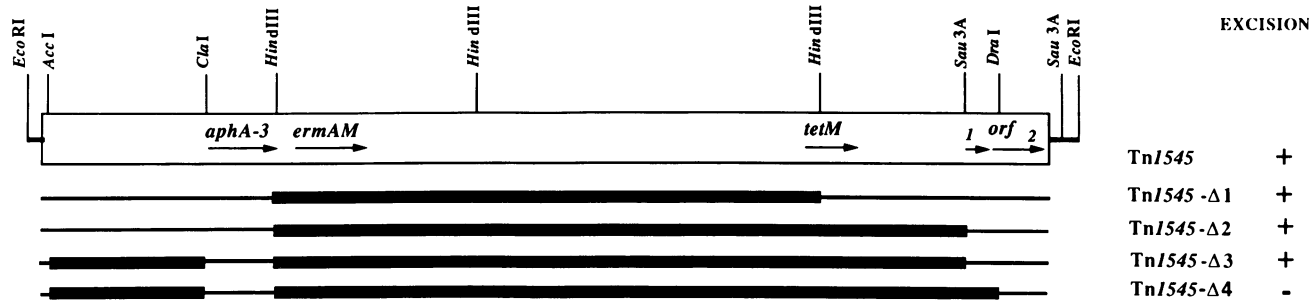

Fig. 1. Restriction and functional map of Tn*1545*.*aphA-3*, 3'-aminoglycoside phosphotransferase type III determinant; *ermAM*, erythromycin resistance methylase determinant; *tetM*, tetracycline resistance determinant; *orf1* and *2*, open reading frames 1 and 2 respectively. Arrows indicate direction and extent of transcription. The deleted regions leading to Tn*1545*-Δ1, -Δ2, -Δ3 and -Δ4 derivatives are represented by heavy black lines. Each derivative cloned into pUC18 was tested for excision in *E.coli* DH1: (+), excision; (−), no excision. The excision activity was deduced from the results presented in Figures 2 and 3.

Results

Localization of the gene(s) involved in the excision of Tn1545

Transposon Tn*1545* excises at a very high frequency when cloned on a multi-copy plasmid in *recA E.coli* strains (Courvalin and Carlier, 1987). We took advantage of this property to localize the gene(s) involved in the excision of this element. Plasmid pIP804 is a non-haemolytic (Hly^-) derivative of the streptococcal plasmid pIP964 (Tra^+, Hly^+, 57.5 kb) obtained by insertion of Tn*1545* (Em^R, Km^R, Tc^R; 25.3 kb) into a 1.3 kb *Eco*RI fragment (Courvalin and Carlier, 1986). The resulting 26.6-kb *Eco*RI fragment of pIP804 (Figure 1) was purified, mixed with *Eco*RI-linearized and dephosphorylated pUC18 DNA, ligated, introduced into *E.coli* DH1 by transformation, and clones were selected on ampicillin. The plasmid content of 12 randomly selected transformants was analysed by agarose gel electrophoresis of crude bacterial lysates after digestion with *Eco*RI. As expected, all the clones were found to contain a single plasmid consisting of pUC18 plus the 1.3 kb *Eco*RI target fragment of pIP964 (Figure 2, lane 1). The nucleotide sequence of four target sites obtained after excision of Tn*1545* in independent cloning experiments was determined and compared with that of the target site prior to insertion. In three out of four clones studied, the nucleotide sequence of the target was identical before insertion and after excision, whereas that of the remaining clone differed by two substitutions (Figure 3). It therefore appears that excision of Tn*1545*, like that of Tn*916*, can generate nucleotide substitutions at the target site.

The availability of a restriction map of Tn*1545* enables the location of the region(s) necessary for excision by generating *in vitro* deletions (Caillaud *et al.*, 1987). The purified 7 kb left, and 6.8 kb right *Eco*RI−*Hin*dIII junction fragments of Tn*1545* in pIP804 (Figure 1) were mixed with *Eco*RI-linearized and dephosphorylated pUC18 DNA. After ligation, the mixture was transformed into *E.coli* DH1 and transformants were plated on ampicillin. The plasmid content of randomly selected clones was analysed as described above. The restriction profiles obtained with the majority of the clones (8 out of 12) consisted of pUC18 plus an insert that comigrated with the *Eco*RI target fragment of pIP964 (Figure 2, lane 2). The *Eco*RI inserts present in the remaining clones were found to consist of dimers of either the left or the right junction fragment self-ligated at their *Hin*dIII site. Analysis of target sequences obtained

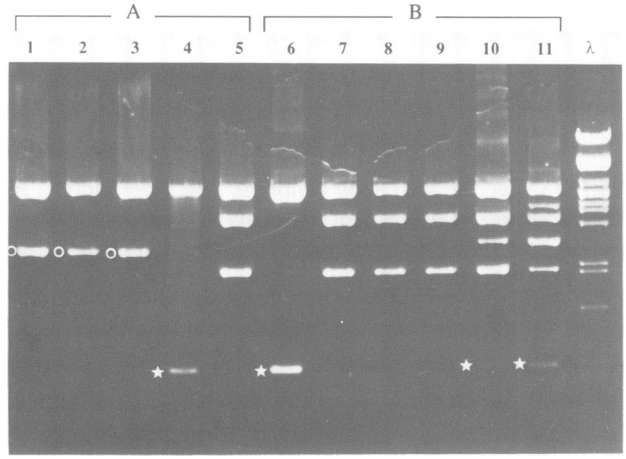

Fig. 2. Detection of excisase activity by analysis of restriction endonuclease patterns. Plasmids were introduced by transformation into *E.coli* DH1 (**A**) or *E.coli* DH1 harbouring the excision test-plasmid pUC18ΩTn*1545*-Δ4 (**B**). After overnight growth (**lanes 1−7** and **9**) or serial cultures for 100 generations (**lanes 8** and **10**), plasmid DNA was purified and digested with *Eco*RI (**lanes 1−3**), *Eco*RI plus *Hin*dIII (**lanes 4−10**) and electrophoresed. Incoming plasmid: **lane 1**, pUC18ΩTn*1545*; **lane 2**, pUC18ΩTn*1545*-Δ1; **lane 3**, pUC18ΩTn*1545*-Δ2; **lane 4**, pUC18ΩTn*1545*-Δ3; **lane 5**, pUC18ΩTn*1545*-Δ4; **lane 6**, pAT295 (*orf1*,*orf2*;+); **lane 7**, pAT296 (*orf1*,*orf2*;−); **lanes 8, 9**, pAT297 (*orf1*); **lanes 10,11**, pAT298 (*orf2*). Bacteriophage λ DNA digested by *Pst*I was used as molecular size standard. The 1.3 kb and 500 bp target fragments resulting from excision of Tn*1545* and deletion derivatives are indicated by open circles and asterisks respectively.

independently revealed that the transient Tn*1545*-deletion derivative constructed, Tn*1545*-Δ1, which consists of the juxtaposition of the two *Eco*RI−*Hin*dIII junction fragments of Tn*1545*, had excised like the parental element (Figure 3). This indicates that the two *Hin*dIII internal fragments are not necessary for excision. Various deletions were then generated *in vitro* into the *Eco*RI−*Hin*dIII junction fragments prior to their ligation into pUC18. The ability of each deletion transposon to excise *in vivo* was studied by monitoring the appearance of target DNA fragments in agarose gel electrophoresis and by determining their nucleotide sequence. The exact location and extent of every deletion is shown in Figure 1 and the results obtained are summarized in Figures 2 and 3. Tn*1545*-Δ3 (3.9 kb) is the smallest transient deletion derivative able to excise *in vivo* (Figure 2, lane 4). It consists of a 526 bp *Eco*RI−*Acc*I junction fragment that contains the left end (185 bp) of Tn*1545*, a 1.3 kb

*Cla*I−*Hin*dIII fragment that encodes kanamycin resistance (*aphA-3*) and a 2.06 kb *Sau*3A fragment that contains the right end (1800 bp) of Tn*1545*. To construct Tn*1545*-Δ3, we had to delete 800 bp of the original 1.3 kb target including the right *Eco*RI site, which leads to a new *Eco*RI−*Hin*dIII target of 500 bp. Element Tn*1545*-Δ4 (2.9 kb), which differs from Tn*1545*-Δ3 by a 1 kb deletion in the left portion of the *Sau*3A fragment (i.e. internal to the element), does not excise in the test system used (Figure 1, Figure 2, lane 5). These results indicate that the gene(s) involved in the excision process of Tn*1545* are located, at least partially, within the *Sau*3A junction fragment.

DNA sequence of Tn1545-Δ3

The nucleotide sequence of the 1.34 kb *Cla*I−*Hin*dIII fragment that contains the kanamycin resistance gene *aphA-3* has been published (Caillaud *et al.*, 1987). We have determined the nucleotide sequence of the 526 bp *Eco*RI−*Acc*I and 2058 bp *Sau*3A fragments, which represent respectively the left and right junction fragments of the original transposon with the streptococcal plasmid pIP964.

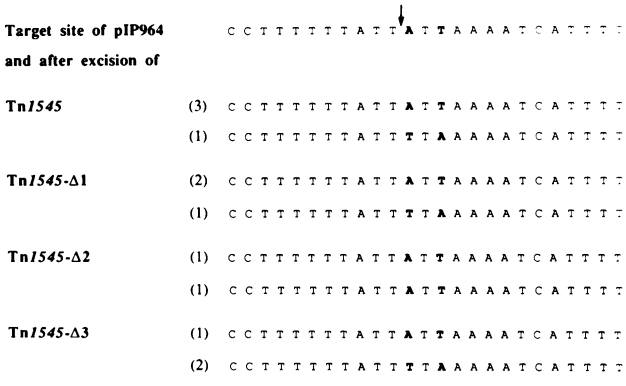

Fig. 3. Nucleotide sequence of the target site of pIP964 before insertion and after excision of Tn*1545* and deletion derivatives in *E.coli* DH1. The number of target sites sequenced is indicated in parentheses. Each target site was obtained after independent cloning experiments. The vertical arrow indicates the site of insertion of Tn*1545*. Variable bases are indicated in bold face lettering. Resequencing of the target site of pIP964 indicated that it contains a stretch of four adenine residues (ATTAAAATCA) instead of five (ATTAAAAATCA) as previously published (Caillaud and Courvalin, 1987).

The DNA sequence of the *Eco*RI−*Acc*I fragment is shown in Figure 4. Transposon Tn*1545* starts at coordinate 341 and ends at the *Acc*I site (coordinate 526). This 185 bp segment of Tn*1545* does not contain an open reading frame (ORF) longer than 100 bp on both strands of DNA or a putative translational initiation site. It is thus likely that this fragment does not direct protein synthesis.

The nucleotide sequence of the *Sau*3A fragment is shown in Figure 5. Analysis of the coding capacity reveals the presence of two ORFs that utilize the same reading frame: *orf1*, which extends from the ATG codon at coordinate 230 to the TAG codon at coordinate 431, and *orf2*, which extends from the ATG codon at coordinate 515 to the TAG codon at coordinate 1730. In both cases the initiator codon is preceded by a sequence characteristic of Gram-positive ribosome binding site (McLaughlin *et al.*, 1981). Open reading frame *orf1* could code for a protein of 67 amino acid residues having a predicted molecular mass of 8100 daltons. This protein, designated ORF1, is slightly basic (K + R = 13, D + E = 10). Within the 220 nucleotides upstream from *orf1* there is no significant homology with the sequences of promoters from Gram-positive bacteria. To determine whether ORF1 is an actual protein, the 357 bp *Taq*I fragment containing *orf1* in its entirety (Figure 5) was cloned in the *Acc*I site of pUC18 generating pAT293. In this construction, it is likely that *orf1* is transcribed from the *lac* promoter of pUC18. The proteins specified by pUC18 and pAT293 were characterized in an *in vitro* transcription−translation system from *E.coli*. One band of ~8000 daltons, which probably corresponds to ORF1, is encoded by pAT293 but not by pUC18 (Figure 6A). The second and larger open reading frame, *orf2*, potentially encodes a 405 amino acid protein with a calculated molecular mass of 46 925 daltons. The deduced peptide, ORF2, is highly basic (K + R = 72, D + E = 55) and was also characterized in an *in vitro* transcription−translation system. For this purpose, the 1750 bp long *Ssp*I−*Sau*3A fragment which contains *orf2* (Figure 5) was cloned in the *Sma*I−*Bam*HI sites of pUC18 to yield pAT294. The transcription of *orf2* in pAT294 is probably under the control of the *lac* promoter of pUC18. As shown in Figure 6B, pAT294 directs the expression of three additional proteins as compared to pUC18. We assume that the larger band corresponds to ORF2 since its apparent M_r of 48 000 daltons is in good agreement with that of 46 925 calculated from the predicted amino acid sequence. The reason for the presence of the two other additional bands

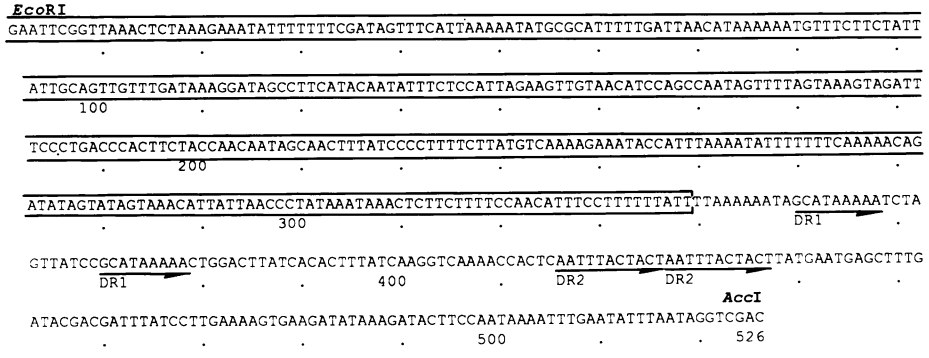

Fig. 4. DNA sequence of the 526 bp *Eco*RI−*Acc*I junction fragment containing the left end of Tn*1545* in pIP804 (pIP964::Tn*1545*) (Caillaud and Courvalin, 1987). Numbering begins at the first base pair in the sequence. The nucleotide sequence of pIP964 is framed. Direct repeats (DR) are depicted by horizontal arrows.

Fig. 5. DNA sequence of the 2058 bp *Sau*3A fragment containing the right end of Tn*1545* in pIP804 (pIP964::Tn*1545*) (Caillaud and Courvalin, 1987). Numbering begins at the first base pair in the sequence. The nucleotide sequence of pIP964 is framed. The ribosome binding sites (RBS) of *orf1* and *orf2* are underlined. The deduced amino acid sequences of ORF1 and ORF2 are presented above the DNA sequence. Domains I and II or ORF2 are delineated by vertical arrows. Direct repeats (DR) are depicted by horizontal arrows. Only relevant restriction sites are shown.

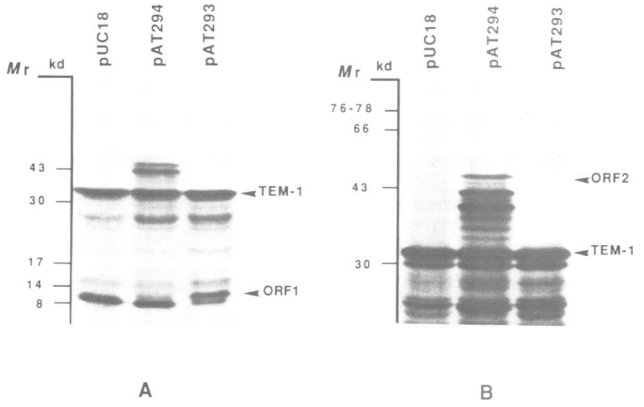

Fig. 6. Autoradiogram of [^{35}S]L-methionine labelled polypeptides specified *in vitro* by pAT293 (*orf1*), pAT294 (*orf2*) and pUC18. The proteins were electrophoresed in 15% (**A**) or 12% (**B**) polyacrylamide gels containing SDS. The positions of ORF1, ORF2 and TEM-1 (pre β-lactamase) are indicated.

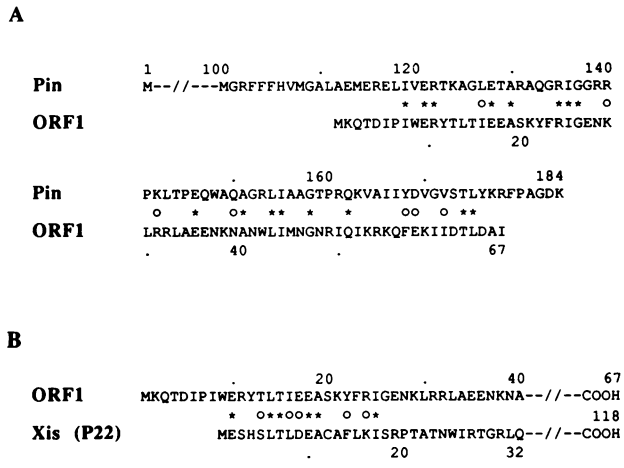

Fig. 7. Comparison of ORF1 amino acid sequence with those of (**A**) Pin from *E.coli* (Plasterk *et al.*, 1983) and (**B**) Xis from bacteriophage P22 (Leong *et al.*, 1986). Stars indicate identical residues and open circles indicate chemically similar residues: I-L-V-M, D-E, R-K, Q-N, S-T and F-Y (Dayhoff and Schwartz, 1978).

is not known. It is, however, likely that they both correspond to truncated versions of ORF2 since there is no large open reading frame on either strand of the *Ssp*I DNA fragment other than ORF2.

Homologies between site-specific recombinases and ORF1 and ORF2

To determine the possible functions of ORF1 and ORF2, we screened the pseqIP library for homologous sequences using the FAST program of Lipman and Pearson (1985). Sequence homology was found between ORF1 and the carboxyl terminal half of Pin from *E.coli*. The alignment in Figure 7A shows that of the 67 possible amino acid matches, 16 (24%) are identical and seven others (10.5%) are accepted replacements (Dayhoff and Schwartz, 1978), which gives a total of 35.5% homology. Pin belongs to the structurally homologous Hin-related site-specific recombinases (Garnier *et al.*, 1987) and catalyses the inversion of a 1800 bp region of the *E.coli* chromosome (Plasterk *et al.*, 1983). The homology between ORF1 and other Hin-related site-specific recombinases is lower than that with Pin. We also detected homology between the amino-terminal half of ORF1 and those of Xis excisionases of bacteriophage P22 (Figure 7B), and also to a lesser extent, of phages λ and φ80 (data not shown). These three Xis proteins are required for normal excision of the corresponding phage and homology among their amino acid sequences is limited to these positions in the amino-terminal region (Leong *et al.*, 1986).

No homology was found between ORF2, the Hin-related and Xis proteins, and the search for homologous sequences in the pseqIP library was unsuccessful. Because of the potential role of ORF2 as a site-specific recombinase, we looked for local regions of homology with the Int-related proteins. This group of proteins comprises phage-encoded integrases, transposon-encoded peptides (transposase and resolvase) and DNA invertases. Although these proteins display an overall sequence diversity, they can be aligned in their C-terminal halves where two domains (I and II), thought to be involved in their active site, are particularly well conserved (Argos *et al.*, 1986). In particular, a histidyl, an arginyl and a tyrosyl residue are perfectly conserved in domain II (Figure 8). This invariant tyrosyl residue is likely to be covalently linked to the DNA during recombination. Interestingly, these two conserved regions were found in ORF2 at the same relative position. Domain I extends from the tyrosine at coordinate 213 to the glutamic acid at coordinate 240 and domain II extends from the histidine at position 342 to the alanine at position 300 (Figures 5 and 8). This finding strongly suggests that ORF2 belongs to the family of Int-related site-specific recombinases.

In vivo properties of ORF1 and ORF2

To get an insight into the respective properties of ORF1 and ORF2, we investigated the possibility that excision of the cloned derivative element Tn*1545*-Δ4 (which lacks both *orf1* and *orf2*) could be achieved by providing ORF1 and/or ORF2 in *trans* on a separate plasmid. We therefore cloned the 2058 bp *Sau*3A fragment encoding ORF1 and ORF2 in both orientations in pHSG576, a low copy number plasmid compatible with pUC (Takeshita *et al.*, 1987). In plasmids pAT295 and pAT296, *orf1* and *orf2* are oriented clockwise (+) and counterwise (−) relative to the direction of transcription of the *lac* promoter of pHSG576 respectively. We also inserted in the same vector the 368 bp *Taq*I fragment encoding ORF1 (Figure 5) and the 1.8 kb *Ssp*I−*Sau*3A fragment encoding ORF2 (Figure 5) to yield pAT297 and pAT298, respectively. In both plasmids, *orf1* and *orf2* are oriented clockwise (+) relative to the external *lac* promoter. Plasmid pHSG576 and derivatives were independently introduced by transformation into *E.coli* DH1 harbouring the compatible excision tester plasmid pUC18Ω Tn*1545*-Δ4. The plasmid content of overnight transformant cultures grown in presence of ampicillin (selective marker of pUC18Ω Tn*1545*-Δ4) plus chloramphenicol (selective marker of pHSG576 derivatives) was analysed by agarose gel electrophoresis following digestion by *Eco*RI plus *Hin*dIII. Excision was deduced from the appearance of the specific 500 bp target DNA fragment and the results are summarized in Table I. Plasmid pAT295 (*orf1,orf2*;+) was able to *trans*-complement excision of Tn*1545*-Δ4 (Figure 2, lane 6), whereas pAT296 (*orf1,orf2*;−) was not (Figure 2, lane 7).

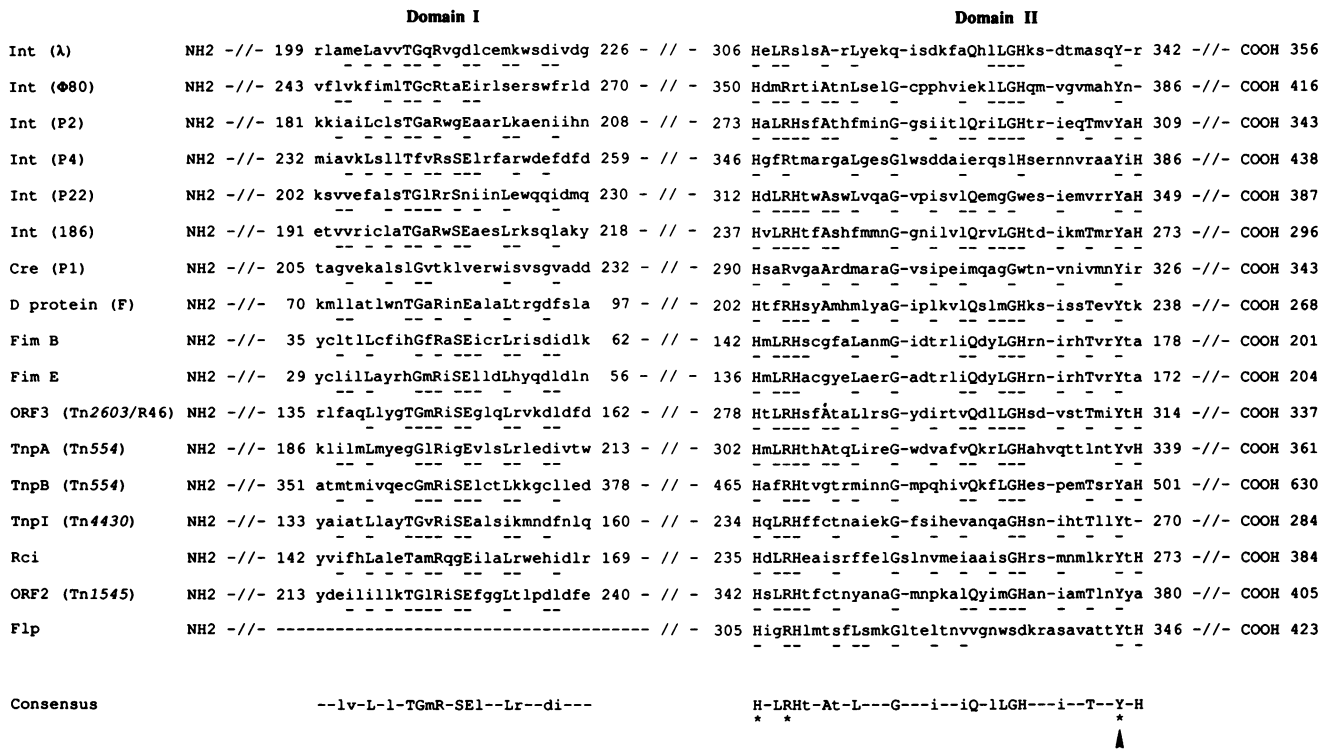

Fig. 8. Local homology among the integrase family of site-specific recombinases. Domains I and II are those described by Argos *et al.* (1986). The amino acid sequences of Int (P2), Int (P4) and Int (186) are from Argos *et al.* (1986); Int (λ) from Hoess *et al.* (1980); Int (Φ80) and Int (P22) from Leong *et al.* (1986); Cre (P1) from Sternberg *et al.* (1986); D protein (F) from Lane *et al.* (1986); Fim B and Fim E from Klemm (1986); ORF3 (R46/Tn*2603*) from Hall and Vockler (1987) and Ouellette and Roy (1987); Flp from Hartley and Donelson (1980); Tn*554* TnpA and TnpB from Murphy *et al.* (1985); Tn*4430* TnpI from Mahillon and Lereclus (1988); Rci from Kubo *et al.* (1988). Numbers indicate the position of the corresponding residues in the sequence of every protein. Gaps (−) were introduced to maximize homology. Positions at which a minimum of 9 out of 16 proteins contain the same amino acid are shown by underlined upper-case. Positions at which a minimum of 9 out of 17 proteins contain homologous amino acids are shown by underlined lower-case; chemically similar residues are: I-L-V-M, D-E, R-K, Q-N, S-T and F-Y (Dayhoff and Schwartz, 1978). Residues identical in all sequences are marked by an asterisk. Arrow indicates the tyrosine residue likely to be covalently linked to DNA during recombination.

These results indicate that the 2058 bp *Sau*3A fragment contains the gene(s) necessary for excision but probably not the upstream promoter(s). The co-existence of the excision-test plasmid and of pAT295 resulted in a high rate of excision as suggested by the intensity of the target DNA fragment (Figure 2, lane 6) and by the fact that, in three independent transformation experiments, 60% of the clones tested (88) were susceptible to kanamycin. It is worth noting that kanamycin susceptibility indicates that excision has affected all the copies of pUC18ΩTn*1545*-Δ4 within a single cell. On the contrary, no excision was observed after transformation of *E.coli* DH1 containing pUC18ΩTn*1545*-Δ4 with pAT297 (*orf1*) even after growth of these transformants for 100 generations in the presence of ampicillin plus chloramphenicol (Figure 2, lanes 8, 9). When similar transformation experiments were carried out with pAT298 (*orf2*), the target DNA fragment was observed just above the detection threshold (Figure 2, lane 10). However, serial cultures of these transformants for 100 generations in the presence of ampicillin plus chloramphenicol did result in the production of larger amounts of target DNA (Figure 2, lane 11) although all the clones tested (88/88) were still resistant to kanamycin. This indicates that, in all these clones, the tester plasmid was present in both excised and unexcised forms, as can also be deduced from restriction enzyme patterns (Figure 2, lanes 10, 11). Determination of target sequences obtained from complementation studies carried out with either pAT295 (*orf1,orf2*;+) or pAT298 (*orf2*) has revealed the presence of the two alternative forms of the excision products in both cases (data not shown). Taken together, these results demonstrate that ORF2 alone is able to catalyse excision of Tn*1545*-Δ4 and that ORF1 strongly stimulates the activity of ORF2.

Table I. Complementation experiments in *E.coli* DH1 harbouring the excision test plasmid pUC18ΩTn*1545*-Δ4

Incoming plasmid	Excision
pAT295 (*orf1,orf2*;+)	++++
pAT296 (*orf1,orf2*;−)	−
pAT297 (*orf1*)	−
pAT298 (*orf2*)	+
pHSG576	−

Plasmid pHSG576 and derivatives were introduced by transformation into *E.coli* DH1. The excision event was studied in transformants resistant to chloramphenicol (Cm) by monitoring the appearance of the specific 500 bp-long target fragment by agarose gel electrophoresis. (++++), (+) and (−) refer to the intensity of the target DNA fragment.

Discussion

The current hypothesis for the nature of the movements of conjugative transposons such as Tn*916* and Tn*1545* implies that, regardless of whether donor and recipient replicons are

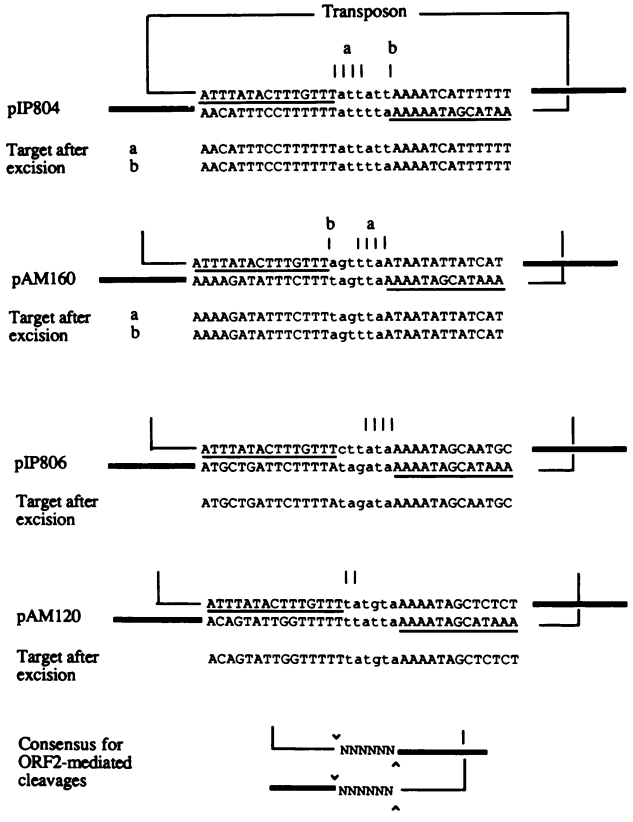

Fig. 9. Comparison of the nucleotide sequences of the junction fragments of Tn*1545* in plasmids pIP804, pIP806 (Caillaud and Courvalin, 1987), and of Tn*916* in pAM120 and pAM160 (Clewell *et al.*, 1988) and of the target site(s) after excision of the element. Invariable base pairs at the termini of the transposons are underlined. The 6-bp core sequence is written in lower-case. In every junction fragment, the possible recombination sites within the core sequence for excision are indicated by a vertical line. The DNA of Tn*1545* and Tn*916* is depicted by a heavy line and that of the target by a thin line. NNNNNN represents the 6 bp variable core; ∨ and ∧ indicates the positions of ORF2 cleavage sites on both strands of DNA.

within the same cell, excision from the donor replicon represents the initial and rate limiting step for transposition (Clewell and Gawron-Burke, 1986). In the present work we have shown, by molecular cloning and DNA sequencing, that the 2058 bp *Sau*3A right-junction fragment of Tn*1545* specifies two proteins, designated ORF1 and ORF2, that are involved in excision of the element. Using an *in vivo* complementation assay, we have demonstrated that ORF2 alone is able to catalyse excision and that ORF1 strongly stimulates the activity of ORF2. It was previously reported that the region of Tn*916* corresponding to that encoding ORF2 in Tn*1545* is necessary for excision (Senghas *et al.*, 1988). Taken together these results indicate that ORF2 is absolutely required for excision of this class of transposons. Proteins ORF1 and ORF2 are encoded by two open reading frames, *orf1* and *orf2*, separated by 84 bp. Both use the same reading phase and are transcribed towards the right end of the transposon. There is no stem−loop structure that could act as a transcriptional terminator downstream from *orf1* and *orf2*. No putative promoter sequence characteristic of genes from Gram-positive bacteria was detected upstream from these genes. It is therefore likely that *orf1* and *orf2* are cotranscribed from a promoter located upstream from the *Sau*3A site internal to the element. The hyperhaemolytic phenotype observed following certain insertions of Tn*1545* in pIP964 (Courvalin and Carlier, 1986) could result from transcriptional activation of *hly*, the gene directing synthesis of haemolysin, by this putative promoter. The initiation and termination sites of both *orf1* and *orf2* transcripts are currently being determined.

The genetic organization and the biological activity, at least in the excision process, of ORF1 and ORF2 is reminiscent of those of lambdoid phage-encoded proteins, Xis and Int respectively. This resemblance is strengthened by the finding that the corresponding proteins display local homologies (Figures 7B and 8). The λ Xis protein is a small basic polypeptide that binds co-operatively to two tandemly repeated sites in the P arm and thus stimulates the binding of Int (Yin *et al.*, 1985). Xis induces a DNA bend that presumably facilitates the formation of a productive synapsis (Thompson *et al.*, 1987). One can therefore reasonably speculate that the stimulatory effect of ORF1 on ORF2-catalysed excision is similar to that exerted by Xis on Int (λ). In this regard, the 11 bp tandem repeats in the left end of Tn*1545* (coordinate 416 in Figure 4) could constitute ORF1-binding site. ORF1 is not structurally related to other host-accessory proteins (IHF, HU and Fis) that stimulate numerous site-specific recombination systems in *E.coli*. However, computer-assisted comparison of ORF1 with the PseqIP amino acid sequence library showed that ORF1 displays significant homology with Pin from *E.coli* (Figure 7A). Pin belongs to the Hin-related family of site-specific recombinases thought to interact with DNA through a helix−turn−helix motif located at their carboxyl terminal ends (Garnier *et al.*, 1987). The fact that ORF1 displays homology with the carboxyl terminal half of Pin suggests that both proteins could interact with DNA in a similar fashion. Unfortunately, the two algorithms used (Levin *et al.*, 1986; Dodd and Egan, 1987) do not predict the presence of the helical motif at the carboxyl end of ORF1. It therefore remains to be determined whether ORF1 is an actual DNA binding protein.

The regional similarity observed between ORF2 and Int-related site-specific recombinases suggests that ORF2 belongs to this family (Argos *et al.*, 1986). Within this group of enzymes, the best characterized recombination events are those catalyzed by Int (λ), Cre (P1) and Flp from the 2 μ yeast plasmid for which *in vitro* systems have been developed. All three recombinases bind to specific sequences, generate staggered nicks with 5′ protruding ends of six (Cre), seven (Int) or eight (Flp) bases (Mizuucchi *et al.*, 1981; Craig and Nash, 1983; Andrews *et al.*, 1985; Hoess and Abremski, 1985), and become transiently attached to the newly formed 3′ phosphoryl terminus via an invariant tyrosine located in their C-terminal region (Figure 8) (Gronostajski and Sadowski, 1985; Pargellis *et al.*, 1988). Based on the functional properties of these enzymes and on the excision process of lambdoid phages, one can envisage the following mechanism for ORF2-catalysed excision of Tn*1545* and Tn*916*. The nucleotide sequences of four junction fragments were aligned so that reciprocal strand exchange generates the target(s) observed after excision (Figure 9). This allows determination of the left and right extremities of Tn*1545* and Tn*916* within which recombination does not occur. These 'invariable extremities' have been chosen to be as long as possible. In the proposed

synaptic complex, the invariable termini are separated by 6 bp, designated the core sequence. The possible recombination sites, within the core sequence for each junction fragment, were determined. We thus examined the possibility that a single 5′ staggered nick cleavage at either of these recombinant sites could generate the target(s) observed after excision. In pIP804 and pAM160, a staggered nick with 5′ protruding ends of 6 bp at the borders of the core sequence accounts for the presence of the two targets observed after excision (Figure 9). The unique target obtained with pIP806 and pAM120 can also result from such a cleavage. We therefore propose that, following specific binding to the ends of the transposon in the presence of ORF1, ORF2 generates 5′-hydroxyl protruding staggered nicks of 6 bp at one of the borders of the core sequence while being linked to the DNA through a 3′ phosphodiester bond. It is worth noting that, due to sequence identity between the targets and the left end of Tn*1545*, the core sequence could be 7 bp long. The presence of two *cis* DNA–protein complexes in a productive synapsis leads to an excised form of the target replicon and to a non-replicative, covalently closed circular form of the transposon. This model implies that DNA homology at the recombination site is not required for synapsis, as demonstrated for λ(Int) (Richet *et al.*, 1988), nor for complete strand-exchange reaction. Alternatively, only one strand of DNA at the recombination site could be resealed, leaving a nick in the other strand. Following excision, the heteroduplexes at the recombination sites could be removed by the mismatch repair system of the host or by replication for the target replicon. Elimination of either of the non-homologous DNA segment would result in the production of one of the two possible excision products independently in the target replicon and in the circular intermediate of the transposon. The degeneracy of the 6 bp core sequence is responsible for the presence of variable base pairs at the ends of Tn*1545* and Tn*916*. The presence of an additional adenine residue in the left extremity of Tn*1545* in pIP804 (AAAAATAGCATAA versus AAAATAGCATAA) might result from an error-prone polymerase during mismatch repair. Although this model accounts for all the previous structural observations, it needs to be supported by additional experimental data. Towards this goal, the development of an *in vitro* assay would greatly facilitate the understanding of ORF2-catalysed recombination. In particular, it will be of obvious interest to determine if ORF2 can also catalyse integration and, hence, if the direction of recombination (integration versus excision) is determined by the relative amounts of ORF1 and ORF2.

Certain strains of *Streptomyces ambofaciens* contain an 11 kb conjugative plasmid, pSAM2, that can be present in an integrated or a free state. It has been recently reported that the integration–excision system of pSAM2 is related to that of lambdoid phages and that it requires two enzymes that display local homology with Xis and Int (Boccard *et al.*, 1989a), and therefore to ORF1 and ORF2 (data not shown). All known integrations of pSAM2 involve a site-specific recombination event within a 58 bp *att* sequence present on both plasmid and chromosomal genomes (Boccard *et al.*, 1989b). The requirement of a large homologous sequence for integration of Tn*1545* or Tn*916* has not been reported. In fact, the exceptional transposition broad host range of Tn*1545*, numerous species of Gram-positive and Gram-negative bacteria, may reflect its capability to utilize efficiently degenerated *att* sequences for integration.

Materials and methods

Bacterial strains, media and antibiotics

Escherichia coli DH1 (Low, 1968) and JM83 (Messing, 1979) were used for plasmid construction and strain JM103 (Messing and Vieira, 1982) for phage preparation. The cells were grown in brain heart infusion broth or agar incubated at 37°C. Antibiotics for bacterial selection were used at the following concentrations (μg/ml): ampicillin, 100; kanamycin, 20; and chloramphenicol, 20.

Plasmids and DNA manipulations

Plasmids pUC18 and pUC19 (Messing and Vieira, 1982; Yanish-Perron *et al.*, 1985), and pHSG576 (Takeshita *et al.*, 1987) were used for cloning. Transformation of *E.coli*, large-scale and small-scale preparation of plasmid DNA, cleavage of DNA with restriction endonucleases, converting of recessed DNA ends to blunt ends, alkaline phosphatase treatment, ligation, analytical and preparative gel electrophoresis were done according to standard methods (Maniatis *et al.*, 1982).

DNA sequence determination

The restriction endonuclease generated fragments were subcloned into M13 bacteriophage vectors mp18 and mp19 (Yanish-Perron *et al.*, 1985). The DNA sequence was determined by the chain termination method (Sanger *et al.*, 1977).

Computer analysis of sequence data

The DNA sequences were arranged using DBCOMP and DBUTIL computer programs (Staden, 1980). DNA and protein sequences were compared using the algorithm of Wilbur and Lipman (1983). The FASTP program of Lipman and Pearson (1985) was used to search for protein similarities in the PseqIP library (Claverie and Bricault, 1986).

Analysis of plasmid-encoded proteins

The proteins specified by the recombinant plasmids were synthesized in an *E.coli in vitro* transcription–translation system (Zubay, 1980). Proteins were labelled with [^{35}S]L-methionine and processed for electrophoresis in an SDS-polyacrylamide gel as described (Laemmli, 1970).

Enzymes and biochemicals

Restriction enzymes, T4 DNA ligase, [^{35}S]L-methionine and [α-^{35}S]dATP were obtained from Amersham International or New England Biolabs (*Dra*I and *Ssp*I). Calf intestinal phosphatase and M13 phages were from Pharmacia. The Sequenase kit was from United Biochemical Corporation and the prokaryotic DNA-directed translation kit was from Amersham. Bacteriological supplies were purchased from Difco and antibiotics were provided by the following manufacturers: ampicillin and kanamycin, Bristol; chloramphenicol, Roussel-Uclaf.

Acknowledgements

We thank M.Chandler for continuous interest and helpful discussions during the course of this work and E.Richet for critical reading of the manuscript. This work was supported by a grant (ATP 'Bactéries Lactiques') from the Centre National de la Recherche Scientifique (CNRS). C.P.-S. was a recipient of a fellowship from the Fondation pour la Recherche Médicale (FRM).

References

Andrews,B.J., Proteau,G.A., Beatty,L.G. and Sadowski,P.D. (1985) *Cell*, **40**, 795–803.

Argos,P., Landy,A., Abremski,K., Egan,S.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.M. (1986) *EMBO J.*, **5**, 433–440.

Boccard,F., Smokvina,T., Pernodet,J.-L., Friedmann,A. and Guérineau,M. (1989a) *EMBO J.*, **8**, 973–980.

Boccard,F., Smokvina,T., Pernodet,J.-L., Friedmann,A. and Guérineau,M. (1989b) *Plasmid*, **21**, 59–70.

Caillaud,F. and Courvalin,P. (1987) *Mol. Gen. Genet.*, **209**, 110–115.

Caillaud,F., Carlier,C. and Courvalin,P. (1987) *Plasmid*, **17**, 58–60.

Caillaud,F., Trieu-Cuot,P., Carlier,C. and Courvalin,P. (1987) *Mol. Gen. Genet.*, **207**, 509–513.
Christie,P.J., Korman,R.Z., Zahler,S.A., Adsit,J.C. and Dunny,G.M. (1987) *J. Bacteriol.*, **162**, 2529–2536.
Claverie,J.M. and Bricault,L. (1986) *Proteins Struct. Funct. Genet.*, **1**, 60–65.
Clewell,D.B. and Gawron-Burke,C. (1986) *Annu. Rev. Microbiol.*, **40**, 635–659.
Clewell,D.B., Ann,F.Y., White,B.A. and Gawron-Burke,C. (1985) *J. Bacteriol.*, **162**, 1212–1220.
Clewell,D.B., Flannagan,S.E., Ike,Y., Jones,J.M. and Gawron-Burke,C. (1988) *J. Bacteriol.*, **170**, 3046–3052.
Courvalin,P. and Carlier,C. (1986) *Mol. Gen. Genet.*, **205**, 291–297.
Courvalin,P. and Carlier,C. (1987) *Mol. Gen. Genet.*, **206**, 259–264.
Craig,N.L. and Nash,H.A. (1983) *Cell*, **35**, 795–803.
Dayhoff,M.O. and Schwartz,R.M. (1978) In Dayhoff,M.L. (ed.), *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington, DC, Vol. 5, pp. 353–358.
Dodd,I.B. and Egan,J.B. (1987) *J. Mol. Biol.*, **194**, 557–564.
Fitzgerald,G.F. and Clewell,D.B. (1985) *Infect. Immunol.*, **47**, 415–420.
Flamm,E.L. and Weisberg,R.A. (1985) *J. Mol. Biol.*, **183**, 117–128.
Franke,A.E. and Clewell,D.B. (1981) *J. Bacteriol.*, **145**, 494–502.
Garnier,T., Saurin,W. and Cole,S.T. (1987) *Mol. Microbiol.*, **3**, 371–376.
Gawron-Burke,C. and Clewell,D.B. (1982) *Nature*, **300**, 281–284.
Gawron-Burke,C. and Clewell,D.B. (1984) *J. Bacteriol.*, **159**, 214–221.
Gronostajski,R.M. and Sadowski,P.D. (1985) *Mol. Cell. Biol.*, **5**, 3274–3279.
Hall,R.M. and Vockler,C. (1987) *Nucleic Acids Res.*, **15**, 7491–7501.
Hartley,J.L. and Donelson,J.E. (1980) *Nature*, **286**, 860–864.
Hoess,R.H., Foeller,C., Bidwell,K. and Landy,A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 2482–2486.
Hoess,R.H. and Abremski,K. (1985) *J. Mol. Biol.*, **181**, 351–362.
Klemm,P. (1986) *EMBO J.*, **5**, 1389–1393.
Kubo,A., Kusukawa,A. and Komano,T. (1988) *Mol. Gen. Genet.*, **213**, 30–35.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Lane,D., de Feyter,R., Kennedy,M., Phua,S.H. and Semon,D. (1986) *Nucleic Acids Res.*, **14**, 9713–9728.
Leong,J.M., Nunes-Duby,S.E., Oser,A.B., Lesser,C.F., Youderian,P., Susskind,M.M. and Landy,A. (1986) *J. Mol. Biol.*, **189**, 603–616.
Levin,J.M., Robson,B. and Garnier,J. (1986) *FEBS Lett.*, **205**, 303–308.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Low,B. (1968) *Proc. Natl. Acad. Sci. USA*, **60**, 160–167.
McLaughlin,J.R., Murray,C.L. and Rabinowitz,J.C. (1981) *J. Biol. Chem.*, **256**, 11283–11291.
Mahillon,J. and Lereclus,D. (1988) *EMBO J.*, **7**, 1515–1526.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Martin,P., Trieu-Cuot,P. and Courvalin,P. (1986) *Nucleic Acids Res.*, **14**, 7047–7058.
Mertens,G., Klippel,A., Fuss,H., Blöcker,H., Franck,R. and Kahmann,R. (1988) *EMBO J.*, **7**, 1219–1227.
Messing,J. (1979) *Recombinant DNA. Tech. Bull.*, **1**, 43–44.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Mizuucchi,K., Weisberg,R., Enquist,L., Mizuucchi,M., Buraczyska,M., Foeller,C., Hus,P.-L., Ross,W. and Landy,A. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 429–437.
Murphy,E., Huwyler,L. and Freire Bastos,M.C. (1985) *EMBO J.*, **4**, 3357–3365.
Ouellette,M. and Roy,P. (1987) *Nucleic Acids Res.*, **15**, 1005.
Pargellis,C.A., Nunes-Düby,S.E., Moitoso de Vargas,L. and Landy,A. (1988) *J. Biol. Chem.*, **263**, 7678–7685.
Plasterk,R.H., Brinkman,A. and Van de Putte,P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5355–5358.
Richet,E., Abcarian,P. and Nash,H.A. (1988) *Cell*, **52**, 9–17.
Sanger,F., Nichklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scott,J.R., Kirchman,P.A. and Caparon,M.G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4809–4813.
Senghas,E., Jones,J.M., Yamamoto,M., Gawron-Burke,C. and Clewell,D.B. (1988) *J. Bacteriol.*, **170**, 245–249.
Staden,R. (1980) *Nucleic Acids Res.*, **8**, 3673–3694.
Sternberg,N., Sauer,B., Hoess,R.H. and Abremski,K. (1986) *J. Mol. Biol.*, **187**, 197–212.
Takeshita,S., Masashiro,S., Toba,M., Masahashi,W. and Hashimoto-Gotoh,T. (1987) *Gene*, **61**, 63–74.
Thompson,J.F,. Moitoso de Vargas,L., Koch,C., Kahmann,R. and Landy,A. (1987) *Cell*, **50**, 901–908.
Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.
Yanish-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Yin,S., Bushman,W. and Landy,A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1040–1044.
Zubay,G. (1980) *Methods Enzymol.*, **65**, 856–877.

Received on March 15, 1989; revised on May 5, 1989