The EMBO Journal vol.13 no.7 pp.1541 – 1548, 1994

Conjugative transposition: Tn*916* integrase contains two independent DNA binding domains that recognize different DNA sequences

Fang Lu and Gordon Churchward[1]

Department of Microbiology and Immunology, Emory University, Atlanta, GA 30322, USA
[1]Corresponding author

Communicated by L.Caro

Transposition of the conjugative transposon Tn*916* requires the activity of a protein, called Int, which is related to members of the integrase family of site-specific recombinases. This family includes phage lambda integrase as well as the Cre, FLP and XerC/XerD recombinases. Different proteins, consisting of fragments of Tn*916* Int protein fused to the C-terminal end of maltose binding protein (MBP) were purified from *Escherichia coli*. DNase I protection experiments showed that MBP-INT proteins containing the C-terminal end of Int bound to the ends of the transposon and adjacent plasmid DNA. MBP-INT proteins containing the N-terminal end of Int bound to sequences within the transposon close to each end. Competition binding experiments showed that the sites recognized by the C- and N-terminal regions of Int did not compete with each other for binding to MBP-INT. We suggest that Tn*916* and related conjugative transposons are unique among members of the integrase family of site-specific recombination systems because the presence of two DNA binding domains in the Int protein might allow Int to bridge recombining sites, and this bridging seems to be the sole mechanism ensuring that only correctly aligned molecules undergo recombination.

Key words: conjugative transposition/DNA–protein interaction/integrase protein/site-specific recombination

Introduction

Conjugative transposons are genetic elements, first identified in Gram-positive bacteria, that during transposition transfer themselves from a donor cell to a recipient in a process which requires intercellular contact. The best studied examples of conjugative transposons are Tn*916*, a 16.4 kb element first found in the chromosome of *Enterococcus faecalis* (Franke and Clewell, 1981), and Tn*1545*, a 25.3 kb element originally identified in the chromosome of *Streptococcus pneumoniae* BM4200 (Courvalin and Carlier, 1986). They are remarkably promiscuous, and can conjugate between different species and genera of bacteria. Most of the conjugative transposons identified to date carry the *tetM* gene that specifies resistance to tetracycline. These elements are thus important in the spread of antibiotic resistance between Gram-positive pathogens. The properties of these transposons have recently been reviewed in detail (Scott, 1992, 1993). In Gram-positive bacteria, these elements can carry out either intracellular transposition at a frequency of 10^{-5} per donor (measured by mating out a conjugal plasmid into which the transposon has inserted) or intercellular conjugation at a lower frequency ($10^{-6}-10^{-8}$) (Franke and Clewell, 1981; Gawron-Burke and Clewell, 1982). In Gram-negative bacteria, while transposition can be observed, the predominant reaction is excision of the transposon, which leads to accumulation of a circular form (Scott *et al.*, 1988). The basis for this difference in behavior between Gram-positive and Gram-negative hosts is not understood.

These elements differ from other bacterial transposons in that insertion into a target site is not accompanied by a duplication of target sequences (Clewell *et al.*, 1988; Poyart-Salmeron *et al.*, 1989). Rather, a 6 bp segment of DNA, termed a coupling sequence, is brought in with the transposon. Subsequent excision of the transposon can either restore the original target sequence, or result in replacement of target DNA by the coupling sequence (Caparon and Scott, 1989). Nucleotide sequence analysis of target sequences and transposon-target junctions led to the proposal that these elements first excise and form a circular intermediate which can then reintegrate into a new site to complete the transposition process (Clewell *et al.*, 1988; Caparon and Scott, 1989). An excised circular form of the transposon can be detected either by gel electrophoresis or by PCR amplification across the joined transposon ends (Scott *et al.*, 1988; Caparon and Scott, 1989; Poyart-Salmeron *et al.*, 1989). A circular form of Tn*916* has been isolated from *E.coli*, then shown to integrate into different chromosomal sites upon transformation into *Bacillus subtilis* protoplasts, thus providing confirmation for the idea of a circular intermediate in transposition (Scott *et al.*, 1988). Currently it is thought that staggered endonucleolytic cleavage at the ends of the transposon followed by recircularization produces a circular intermediate with a short stretch of heteroduplex DNA between the juxtaposed transposon ends (Caparon and Scott, 1989). Such heteroduplexes have been demonstrated in the circular intermediate by nucleotide sequence analysis, resistance of the junction region to restriction endonuclease digestion and by following the inheritance of the two sequences during transposition (Caparon and Scott, 1989).

Genetic studies have shown that two genes located close to the ends of these transposons are implicated in transposition (Senghas *et al.*, 1988; Poyart-Salmeron *et al.*, 1989). On the basis of nucleotide sequence homologies with the genes of lambdoid phages, these genes have been termed *int* (for integrase) and *xis* (for excisionase) (Poyart-Salmeron *et al.*, 1989). The Tn*916* and Tn*1545* Int proteins are homologous to the integrase family of transposases, and contain the conserved arginine, histidine and tyrosine residues that are important for the DNA cleavage reaction catalyzed by these enzymes (Pargellis *et al.*, 1988; Parsons *et al.*, 1988, 1990; Poyart-Salmeron *et al.*, 1989). In Gram-

positive bacteria, the *int* gene is required for transposition (Storrs *et al.*, 1991; Bringel *et al.*, 1992). When an *int*⁻ transposon present in the chromosome is complemented by a cloned *int* gene integrated at a different chromosomal location, the *int* gene is required only in the donor cell for excision of the element, raising the possibility that the *int* gene product either is not required in the recipient for integration, or is transferred along with the transposon DNA during conjugation (Bringel *et al.*, 1992). In Gram-negative bacteria, excision requires the *int* gene product, and is stimulated by the *xis* gene product, but this latter gene is not essential for excision to occur (Poyart-Salmeron *et al.*, 1989). The nucleotide sequences of the *int* and *xis* genes of Tn*916* and Tn*1545* differ by one nucleotide, resulting in the presence of Lys at position 9 of Tn*916* Int protein and Arg at position 9 of Tn*1545* Int (Poyart-Salmeron *et al.*, 1989; Clewell *et al.*, 1991). By convention, the left end of Tn*916* corresponds to the right end of Tn*1545*. In this paper we use the Tn*916* convention. The *int* and *xis* genes are located at the left end of Tn*916*. The corresponding ends of the transposons are identical for at least 186 bp at the left end and 108 bp at the right (Clewell *et al.*, 1988; Poyart-Salmeron *et al.*, 1989). These end segments contain direct repeats of a sequence element, AGTAGTAAATT (termed DR-2), ~150 bp from the left end and 90 bp from the right end (Clewell *et al.*, 1988; Poyart-Salmeron *et al.*, 1989). The position of these DR-2 repeats relative to the ends of the transposon is shown in Figure 8. As expected from the similarity of the Int proteins and transposon ends, the Int protein of Tn*1545* can complement a Tn*916 int*⁻ mutant for transposition (Bringel *et al.*, 1992), and the Int protein of Tn*916* will catalyze the excision of a deletion derivative of Tn*1545* (see Results).

Examination of a series of target sequences shows that the sites of integration of Tn*916* and Tn*1545* are not random, and are in fact similar to the nucleotide sequence of the joined transposon ends present in the excised, circular form of the transposon (Clewell *et al.*, 1988; Poyart-Salmeron *et al.*, 1989; Scott, 1993; Trieu-Cuot *et al.*, 1993). However, there seems to be an important difference in the mechanisms of recombination that occur in conjugative transposition and in bacteriophage lambda integration and excision. In the circular form of Tn*916*, the short coupling sequence lies between the ends of the transposon (Caparon and Scott, 1989). Since every excision can produce a circular form with different coupling sequences, this means that the central portion of the interacting transposon and target sequences, analogous to the overlap regions of *attB* and *attP* in the lambda system (Landy, 1989), can be different. Introduction of similar non-homology into the overlap regions of interacting lambda sites blocks recombination (Weisberg *et al.*, 1983; Nunes-Duby *et al.*, 1987; Kitts and Nash, 1988).

In order to begin the analysis of the recombination mechanisms that act during conjugative transposition, we have purified fused maltose binding proteins (MBPs) that carry different segments of Tn*916* integrase, and examined the interactions of these fused proteins with DNA segments containing either end of the transposon. The results show that Tn*916* integrase contains two independent DNA binding domains. One domain in the carboxyl end of the protein interacts with the end of the transposon and adjacent host DNA sequences. The other domain in the N-terminal end of the protein interacts with the DR-2 sequence elements located within the transposon (Figure 8).

Results

Purification of Tn916 integrase

Initially we attempted to purify native Tn*916* integrase. A 1628 bp *Mae*I–*Sau*3A fragment containing the Tn*916 int* gene was cloned into the vector plasmid pSP73 so that the *int* gene could be transcribed from a bacteriophage T7 promoter. T7 RNA polymerase was provided *in trans* from a phage lambda prophage carrying the T7 RNA polymerase under the control of the *lac* promoter (Studier and Moffatt, 1986). Even in the absence of inducer the cloned *int* gene could only be maintained if the bacterial host also contained a compatible plasmid carrying a gene encoding phage T7 lysozyme, a bifunctional protein that in addition to its enzymatic activity also binds and inhibits T7 RNA polymerase (Moffatt and Studier, 1987). Presumably, in the absence of T7 lysozyme inhibitor, even basal levels of expression of T7 RNA polymerase were sufficient to lead to the production of lethal amounts of Tn*916 int*. To test for activity of the cloned *int* gene *in vivo* we first cloned the defective derivative del4 of Tn*1545* (Poyart-Salmeron *et al.*, 1989) into a pSC101 plasmid. In this derivative, all Tn*1545* sequences except for 180 bp at one end and 950 bp at the other end of the transposon have been replaced with a DNA fragment encoding Kan^r. The resulting plasmid was introduced into the strain carrying the cloned *int* gene. Following induction of T7 RNA polymerase expression by addition of IPTG to the culture, total plasmid DNA was prepared from the culture and used as a template in a PCR assay to detect excision of the Tn*1545*del4 transposon. Oligonucleotide primers complementary to each end of the Tn*1545*del4 transposon were used to amplify a DNA fragment corresponding to the joined ends of the excised transposon. Such a fragment was amplified from templates prepared from cultures containing the cloned *int* gene, but not from cultures containing the pSP73 vector plasmid. Despite this evidence of *in vivo* activity of the cloned Tn*916 int* gene, attempts to detect DNA binding activity specific for DNA fragments containing the ends of Tn*916* and flanking bacterial DNA, using extracts prepared from cells carrying the cloned *int* gene, were unsuccessful.

In the course of these experiments we used a commercially available coupled *in vitro* transcription and translation system to analyze the synthesis of proteins from the pSP73 plasmid carrying the cloned *int* gene. Using either plasmid DNA, or mRNA produced by transcription with purified T7 RNA polymerase, we observed the synthesis of a protein of the size expected for Tn*916* Int protein, as well as a smaller protein. Neither was observed when the original pSP73 vector plasmid was used as template. We examined the DNA sequence of the Tn*916 int* gene and found an internal ATG codon at nucleotide 244 of the coding sequence. This codon was preceded at an appropriate distance by a sequence similar to a Shine–Dalgarno ribosome binding site. Translation initiating at this internal ATG codon could explain the production of the smaller protein observed in the *in vitro* translation experiments. In similar experiments using a DNA template encoding the Tn*1545* Int protein, several different products were observed (Poyart-Salmeron *et al.*, 1989).

On the basis of these results, we designed oligonucleotide

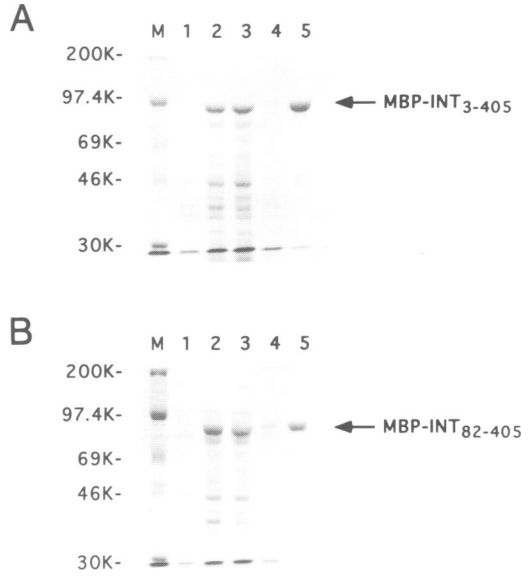

Fig. 1. SDS−PAGE of MBP-INT fusion proteins. (**A**) MBP-INT_{3-405}. (**B**) MBP-INT_{82-405}. Lane M, molecular weight markers. Lane 1, lysate from uninduced cells. Lane 2, lysate from induced cells. Lane 3, soluble crude extract. Lane 4, flowthrough from amylose column. Lane 5, peak fraction from elution of amylose column with maltose.

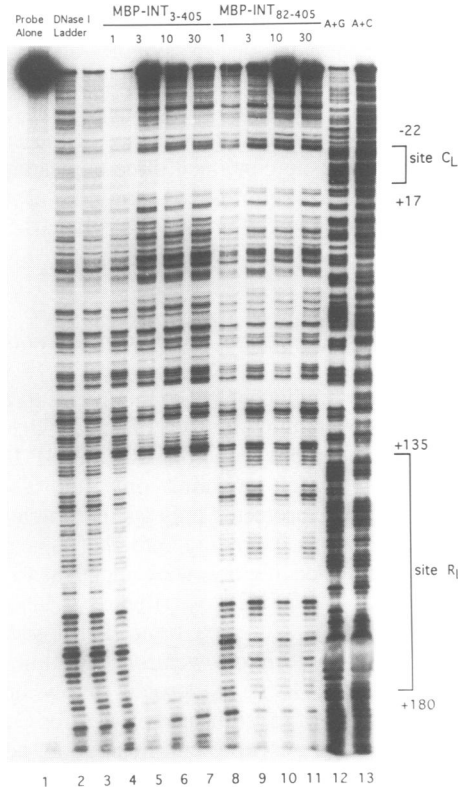

Fig. 2. DNase I protection assay using as radiolabelled probe DNA from the left end of Tn*916* and adjacent plasmid DNA. Lane 1, probe alone. Lanes 2 and 3, probe plus DNase I. Lanes 4−7, probe plus 1, 3, 10 or 30 ng/μl MBP-INT_{3-405} plus DNase I. Lanes 8−11, probe plus 1, 3, 10 or 30 ng/μl MBP-INT_{82-405} plus DNase I. Lanes 12 and 13, A+G and A>C Maxam−Gilbert sequence reactions.

primers that would allow us to amplify either the entire coding sequence of the *int* gene, or a portion beginning at nucleotide 244 and lacking the N-terminal end of the intact *int* gene. DNA fragments produced by PCR were cloned into the vector pMAL-c2 so that, following induction with IPTG, transcription initiating at a *lac* promoter would lead to the expression of a fused protein. The fused proteins would consist of MBP with either intact or a fragment of Tn*916* integrase at the C-terminal end. Plasmids were screened for the production of an appropriately sized protein. The junctions between the coding sequences of MBP and Tn*916* Int in several plasmids were determined and found to be identical. Plasmids designed to express a fused protein containing the intact integrase all contained an insert beginning at base 6 of the coding sequence of the *int* gene. Deletion of the first two codons was presumably due to treatment with S1 nuclease employed in the cloning procedure. Plasmids expressing the C-terminal end of Int all contained an insert beginning at base 244 of the coding sequence. One representative plasmid of each group was kept, designated pMBP-INT_{3-405} or pMBP-INT_{82-405}, and the sequence of the entire insert was determined and shown to be identical to the Tn*916 int* gene.

Cultures of strains carrying the pMBP-INT plasmids were induced for expression with IPTG, and soluble extracts were subjected to chromatography on amylose resin. Maltose binding protein fusions were recovered from the resin by elution with maltose. The results of these experiments for pMBP-INT_{3-405} and pMBP-INT_{82-405} are shown in Figure 1A and B respectively. In both cases the fused proteins were expressed at significant levels, constituting ~10% of total cellular proein (compare lanes 1 and 2). The fused proteins were soluble (lane 3) and readily purified (lanes 4 and 5). In each case 10−15 mg of protein was recovered from 800 ml of culture. The proteins were stable for at least 6 months at −80°C. Each fused protein contained a cleavage site for Factor X protease at the junction between the MBP and Int domains. Upon digestion with Factor X protease, MBP-INT_{82-405} accumulated, but MBP-INT_{3-405} underwent further proteolytic cleavage.

We tested both fused proteins for excision activity *in vivo*. Upon induction of MBP-INT_{3-405} expression in cells containing a plasmid carrying Tn*1545*del4 we were able to observe excision of the defective transposon (data not shown). Since this implied that the fused protein retained at least some biological activity we performed the rest of the experiments described in this paper using the fused proteins without cleavage and further purification of the Int fragments.

Interaction of Tn916 integrase and transposon DNA

In order to examine the interaction between Tn*916* Int protein and the transposon DNA, we amplified segments of DNA each containing one end of the transposon from Tn*1545*del4 (identical in sequence to the analogous regions of Tn*916*) joined to flanking target DNA sequences derived from the streptococcal plasmid pIP964. These fragments were radiolabelled and used as probes in DNase I protection experiments (Galas and Schmitz, 1978) using the purified MBP-INT proteins. Figure 2 shows the DNase I digestion pattern obtained in the presence and absence of integrase using as probe a DNA fragment containing the left end of the transposon. Comparison of lanes 2−7 shows that MBP-INT_{3-405} protected two distinct regions in the DNA fragment. One protected region spanned 40 bp (from +17 to −22; the left end of the transposon is designated +1)

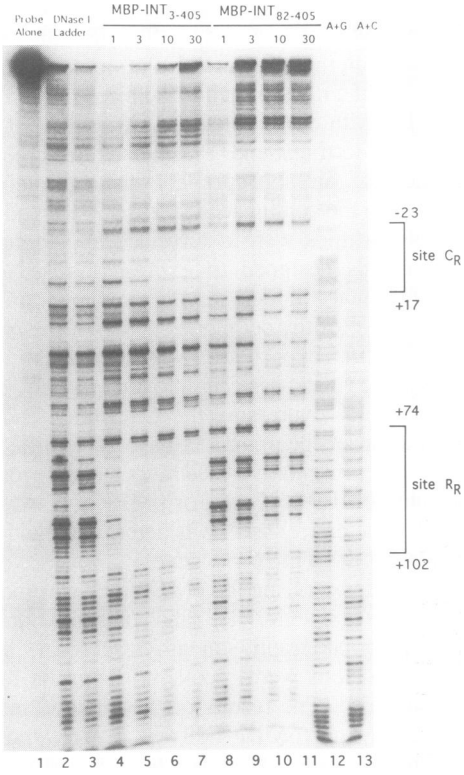

Fig. 3. DNase I protection assay using as radiolabelled probe DNA from the right end of Tn*916* and adjacent plasmid DNA. Lane 1, probe alone. Lanes 2 and 3, probe plus DNase I. Lanes 4–7, probe plus 1, 3, 10 or 30 ng/μl MBP-INT_{3-405} plus DNase I. Lanes 8–11, probe plus 1, 3, 10 or 30 ng/μl MBP-INT_{82-405} plus DNase I. Lanes 12 and 13, A+G and A>C Maxam–Gilbert sequence reactions.

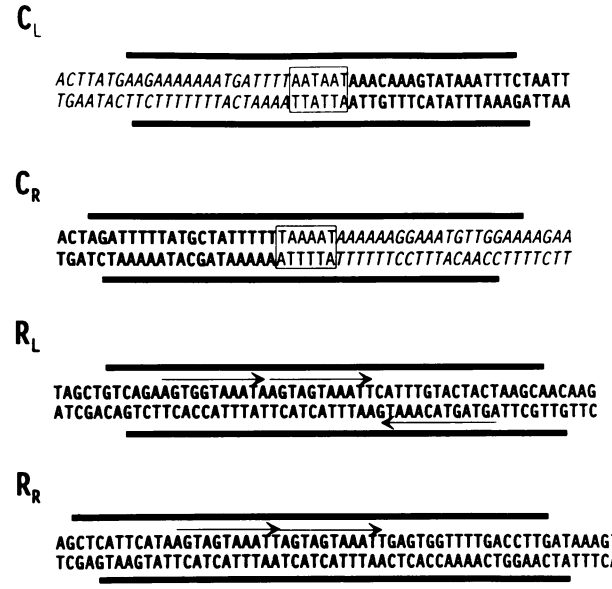

Fig. 4. DNA sequences protected by MBP-INT proteins. The limits of protection by MBP-INT proteins are designated by the bold lines above and below the sequences. For sites C_L and C_R, bold letters indicate transposon DNA, and italics indicate flanking plasmid DNA. The boxed sequences are the 6 bp coupling sequences. For sites R_L and R_R, the arrows show the DR-2 AGTAGTAAATT repeats.

at the end of the transposon, and extended past the 6 bp coupling sequence into the flanking bacterial target DNA. We arbitrarily named it site C_L (for coupling sequence region, left end). The second integrase-protected region spanned 45 bp (from +180 to +135). This second site lay entirely within the transposon and included three DR-2 repeats. We called it site R_L (for repeat region, left end).

A similar DNase I protection analysis was carried out using a DNA fragment containing the right end of the transposon and flanking target DNA (Figure 3, lanes 2–7). A region, C_R, of 40 bp at the end of the transposon extending past the coupling sequence into the flanking target DNA (from +17 to −23) was protected from DNase I cleavage. A second region, R_R, spanning 28 bp, with weak protection extending a further 14 bp towards site C_R, was also protected. Site R_R contained direct repeats similar to those found in site R_L. By comparing the degree of protection as a function of the concentration of MBP-INT_{3-405} in the reactions we estimated that the affinity of MBP-INT_{3-405} for R_L and R_R was similar, as were the affinities for sites C_L and C_R, while the affinity for R sites was 3- to 10-fold higher than that for C sites.

The DNase I protection experiments were repeated with the other strands labelled (data not shown). The results of all the protection experiments performed using MBP-INT_{3-405} are summarized in Figure 4 and the relative arrangement of the sites is shown in Figure 8. Examination of the sequence of the protected regions showed that sites C_L and C_R were similar, as were sites R_L and R_R, but the C and R sites appeared to be quite different. The C sites were characterized by tracts of T/A and A/T base-pairs surrounding the 6 bp coupling sequence at the end of the transposon. The R_L site contained three copies and the R_R site contained two copies of the DR-2 element AGTAGTAAATT (Clewell *et al.*, 1988; Poyart-Salmeron *et al.*, 1989).

Interaction of the C-terminal portion of Int with transposon DNA

We examined the binding of MBP-INT_{82-405} to the same DNA fragments containing the ends of Tn*1545*del4 used in the assays described above. In contrast to MBP-INT_{3-405}, MBP-INT_{82-405}, which contains only the C-terminal portion of Tn*916* Int, protected only a single region at each end of the transposon extending into the adjacent target DNA. These protected regions were identical to sites C_L and C_R bound by MBP-INT_{3-405} (Figure 2, lanes 8–11; Figure 3, lanes 8–11). Complete protection of sites C_L and C_R was observed at a 10-fold lower concentration of MBP-INT_{82-405} than MBP-INT_{3-405} protein implying that, if anything, the affinity of the truncated protein for C sites was higher than that of the intact protein. In addition, no binding of MBP-INT_{82-405} to sites R_L and R_R was observed although the affinity of MBP-INT_{3-405} for sites R_L and R_R was 10 times higher than for sites C_L and C_R. If the difference in DNA binding behavior of MBP-INT_{3-405} and MBP-INT_{82-405} were due to partial inactivation of a DNA binding domain that recognized both C and R sites, C-site binding activity should have been lost before R-site binding activity. We concluded that deletion of the N-terminal segment of Int specifically inactivated binding to site R_L and R_R.

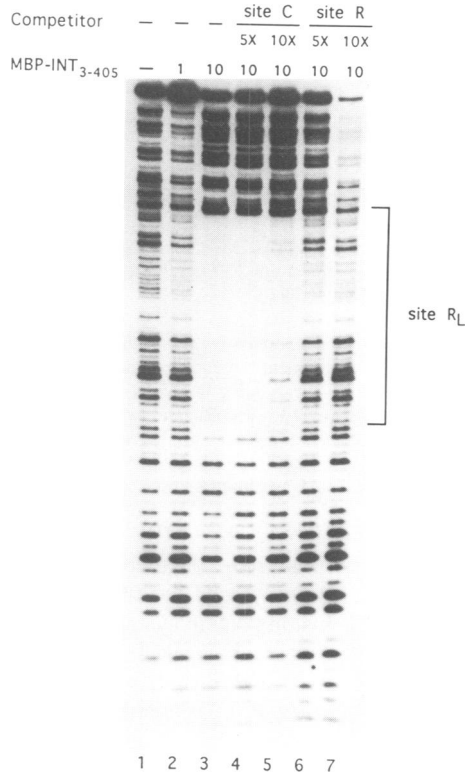

Fig. 5. Competition protection assay using radiolabelled site R_L as probe. Lane 1, probe plus DNase I. Lanes 2 and 3, probe plus 1 and 10 ng/μl MBP-INT_{3-405}. Lanes 4 and 5, probe plus 10 ng/μl MBP-INT_{3-405} plus 5- or 10-fold molar excess of competitor DNA fragment carrying site C_R. Lanes 6 and 7, probe plus 10 ng/μl MBP-INT_{3-405} plus 5- or 10-fold molar excess of competitor DNA fragment carrying site R_L.

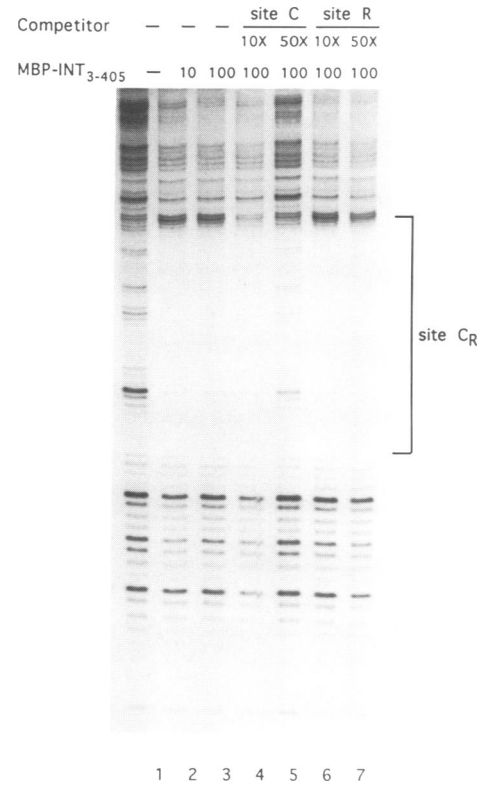

Fig. 6. Competition protection assay using radiolabelled site C_R as probe. Lane 1, probe plus DNase I. Lanes 2 and 3, probe plus 10 and 100 ng/μl MBP-INT_{3-405}. Lanes 4 and 5, probe plus 100 ng/50 μl MBP-INT_{3-405} plus 10- or 50-fold molar excess of competitor DNA fragment carrying site C_R. Lanes 6 and 7, probe plus 100 ng/μl MBP-INT_{3-405} plus 10- or 50-fold molar excess of competitor DNA fragment carrying site R_L.

Independent binding of Tn916 integrase to site C and site R

The different sequence patterns of sites C_L and C_R compared with sites R_L and R_R, as well as the lack of protection of sites R_L and R_R by MBP-INT_{82-405}, suggested that Tn*916* integrase might contain two DNA binding domains that interact independently with C and R sites. If this prediction were true, DNA fragments carrying a C and an R site separately should not compete with each other for binding MBP-INT_{3-405}. We therefore performed DNase I protection experiments using either a radiolabelled fragment containing a C site and an unlabelled fragment containing an R site as competitor, or a radiolabelled fragment containing an R site with an unlabelled fragment containing a C site as competitor. Two DNA fragments of similar size were amplified by PCR and used in these assays. One contained site R_L (from +71 to +221) and the other contained site C_R (from −103 to +66). MBP-INT_{3-405} was preincubated with an excess of unlabelled competitor fragment, then radiolabelled DNA was added to the reaction and the mixture was incubated for an additional 20 min, following which DNase I was added. In the case of binding to a radiolabelled fragment containing site R_L, the amount of unlabelled competitor DNA containing either site R_L or site C_R was 5- and 10-fold the amount of the labelled fragment. As shown in Figure 5, the protection by MBP-INT_{3-405} of site R_L was abolished by preincubation with excess unlabelled competitor containing site R_L (lanes 6 and 7), but was not affected by preincubation with unlabelled competitor containing site C_R (lanes 4 and 5). A similar experiment was carried out with labelled fragment containing site C_R as shown in Figure 6. Since MBP-INT_{3-405} had a 10-fold lower binding affinity to C sites compared with R sites, a higher concentration of the protein was required to achieve either partial or total protection of site C_R than that required for site R_L. Therefore more competitor DNA was required to compete away MBP-INT_{3-405} from site C_R than from site R_L. Protection against DNase I cleavage by MBP-INT bound to site C_R was abolished by preincubation of excess unlabelled homologous fragments containing site C_R (Figure 6, lanes 4 and 5), but not by heterologous fragments containing site R_L (Figure 6, lanes 6 and 7). On the basis of these experiments we concluded that DNA fragments containing either C or R sites did not compete with each other for binding to MBP-INT_{3-405}, and therefore Tn*916* integrase contained two independent DNA binding domains.

Interaction of the N-terminal end of Int with transposon DNA

Since MBP-INT_{82-405} was unable to bind to R sites, we suspected that the R site binding domain was located in the N-terminal portion of Int. To demonstrate this directly, we constructed a plasmid pMBP-INT_{3-78} which contained a segment of Tn*916* DNA encoding amino acids 3−78 of Int downstream of the gene encoding MBP. This plasmid produced a fused protein which was purified by

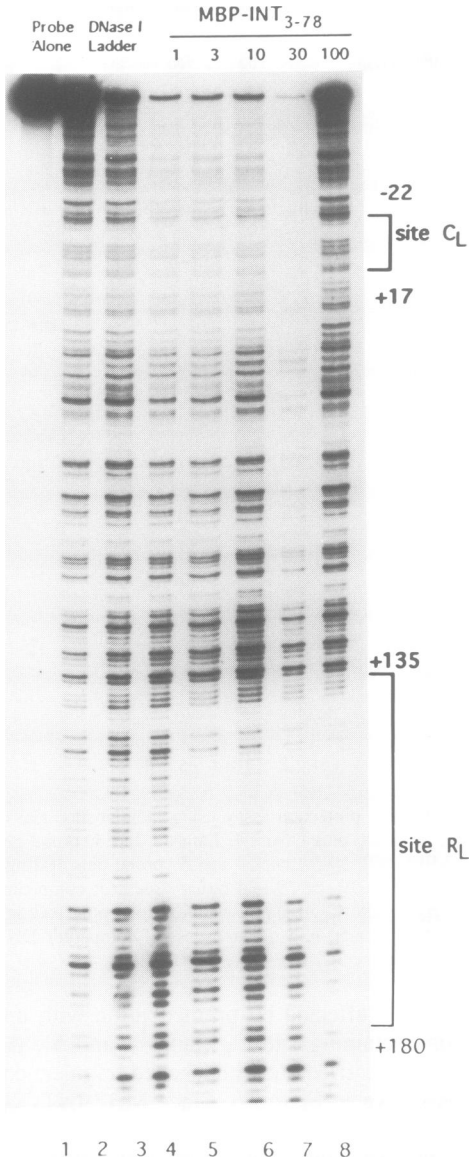

Fig. 7. DNase I protection assay using as radiolabelled probe DNA from the left end of Tn*916* and adjacent plasmid DNA. Lane 1, probe alone. Lanes 2 and 3, probe plus DNase I. Lanes 4–8, probe plus 1, 3, 10, 30 or 100 ng/μl MBP-INT$_{3-78}$ plus DNase I.

chromatography on amylose. The purified fused protein was incubated with a radiolabelled DNA fragment containing the left end of the transposon, and the resulting protein–DNA complexes were digested with DNase I. Figure 7 shows that the fused protein protected a region within site R_L, demonstrating that the N-terminal end of Int bound independently to an R site.

Discussion

Although attempts to detect specific DNA binding and to purify native Tn*916* Int protein were unsuccessful, we found that Tn*916* Int protein could be efficiently expressed and rapidly purified as a C-terminal fusion to MBP. MBP-INT proteins showed specific binding to sites overlapping the ends of the transposon as well as to sites within the transposon DNA. We also found, using a PCR assay to detect excised,

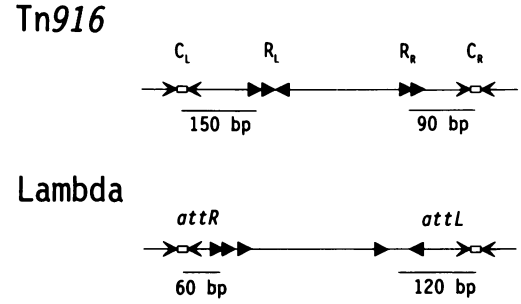

Fig. 8. Comparison of Int binding sites for Tn*916* and lambda. For Tn*916*, putative half sites for Int binding in sites C_L and C_R are indicated by open arrows surrounding the coupling sequence (open boxes). DR-2 repeats in sites R_L and R_R are represented by closed arrows. For lambda, core binding sites are represented by open arrows surrounding the overlap regions (open boxes). Arm-type integrase binding sites are represented by closed arrows.

circular transposon DNA, that MBP-INT$_{3-405}$ retained some excision activity *in vivo*. Whether or not this limited activity was due to proteolysis of the fusion protein is not clear. Similar MBP fusion protein systems have been used for the study of other recombinases. In the case of IS*903* transposase (Derbyshire and Grindley, 1992) and HIV integrase (Kulkosky *et al.*, 1992), the presence of MBP does not significantly change the DNA binding properties and catalytic activities of these recombinases.

DNase I protection assays revealed that the Tn*916* integrase bound to at least four sites on the integrated form of the transposon (Figure 8). Site C_L and C_R cover ~40 bp and extend ~20 bp from the left or right ends of the transposon into flanking plasmid DNA. These sites were characterized by stretches of T/A or A/T base-pairs flanking the 6 bp coupling sequence. We suggest that a monomer of Tn*916* Int protein recognizes each half site on either side of the coupling sequence, one composed of plasmid DNA and the other of transposon DNA. Since the plasmid half sites recognized by Tn*916* Int protein were originally joined to form the target sequence into which the transposon inserted, this suggests that target sequences are recognized by the Int protein in a similar way to the C sites. In other experiments not described here, we have found that both MBP-INT$_{3-405}$ and MBP-INT$_{82-405}$ bind specifically and protect from DNase I cleavage a 40 bp sequence surrounding a preferred target site. This target site contains sequences similar to the plasmid C_L and C_R half sites described here. The apparent affinity of binding to this target site was similar to that observed for sites C_L and C_R. Sites R_L and R_R were located within the transposon 150 and 90 bp respectively from the end of the transposon. These sites contained DR-2 repeats sharing the consensus AGTAGTAAATT (Clewell *et al.*, 1988; Poyart-Salmeron *et al.*, 1989). Based on the data, we suggest that a monomer of Int recognizes these repeats. All four integrase binding sites lie within a region which represents the minimal required sequences for transposition when integrase is supplied *in trans* (Trieu-Cuot *et al.*, 1991), implying that these sites are functionally important *in vivo*.

The difference in behavior of MBP-INT$_{3-405}$, MBP-INT$_{82-405}$ and MBP-INT$_{3-78}$, coupled with the results of the competition binding experiments, showed that MBP-INT contains two DNA binding domains that recognize C sites and R sites independently. In the N-terminal region of Int

protein, 33% of the first 30 amino acids are basic, consistent with the notion that the N-terminal region of Tn*916* Int can bind to DNA, and contains a domain that recognizes R sites. The C-terminal region of Int, which contains amino acids conserved among members of the integrase family of recombinases, including the Tyr residue involved in DNA strand cleavage, binds to C sites. During transposition, binding of a single Int molecule to the C site at one end of the transposon and the R site at the other end could serve to bring the transposon ends together to participate in the excision reaction. Similarly, binding of a single integrase protein to both transposon and target could serve to bring them together in the integration reaction.

During *in vitro* synthesis, shorter forms of both Tn*916* and Tn*1545* (Poyart-Salmeron *et al.*, 1989) Int proteins that contain the recognition site for C sites are produced. It is not known if such truncated forms of Int are normally produced *in vivo*, and what role they may play in transposition. However, if a multicopy plasmid encoding a C-terminal segment of Int is present in an *E.coli* strain carrying a Tn*916* transposon on a second plasmid, excision of the transposon is inhibited (D.Jaworski and D.Clewell, personal communication). Therefore it is possible that C-terminal segments of Int may play a regulatory role in transposition.

The most similar site-specific recombination system to Tn*916* is that of phage lambda (Landy, 1989). The arrangements of integrase binding sites for Tn*916* and the integrated lambda prophage are compared in Figure 8. Like Tn*916*, lambda integrase also recognizes two distinct classes of DNA sequences with independent DNA binding domains (Moitoso de Vargas *et al.*, 1988). The core-type sites, analogous to the proposed half sites in the Tn*916* C sites, share the consensus CAACTTNNT and are present as two inverted repeats at the end of the phage sequence (sites C and C′) and in the bacterial target site (sites B and B′), adjacent to the site where the DNA cleavage occurs. The 7 bp sequence between the sites of strand cleavage is called the overlap region (O). This corresponds to the 6 bp coupling sequences of Tn*916*. After integration, the core-type sites at the junctions between phage and bacterial DNA are present as BOC′ and COB′, which correspond to Tn*916* sites C_L and C_R described in this study. The lambda arm-type sites, analogous to R sites in Tn*916*, are characterized by the consensus sequence C/AAGTCACTAT and occur twice on the left end of lambda phage (attL) and three times on the right end (attR). The arm-type sites are 50−150 bp apart from the core-type sites, which is similar to the spacing between Tn*916* C and R sites.

Lambda and Tn*916* integrases differ from other members of the integrase family of recombinases in possessing two DNA binding domains. In lambda site-specific recombination, host factors such as IHF and FIS cooperate with lambda integrase to form complex wrapped protein−DNA structures which allow synapsis to occur (Bushman *et al.*, 1985; Gardner and Nash, 1986; Thompson *et al.*, 1986; Robertson and Nash, 1988). During synapsis between attL and attR, an integrase molecule bound by its N-terminal domain to an arm-type site can bind a core site at the other end of the prophage with its C-terminal domain, thus bridging the interacting prophage ends (Kim *et al.*, 1990; Kim and Landy, 1992). This has two consequences for recombination. First, interacting sites on chromosomal DNA that are separated by a considerable distance can be brought together in a stable complex. Second, the arrangement of the bridges can ensure that the recombining sites are aligned in the correct orientation. Excision of Tn*916* requires a similar interaction between the transposon ends and so it is not surprising that Tn*916* integrase also contains two independent DNA binding domains.

The ability of the integrase to bridge recombining sites is of additional importance in the Tn*916* system, because there seems to be no other mechanism to ensure that only correctly aligned molecules complete recombination. In other members of the integrase family, the region of the recombining site analogous to the overlap region of lambda is asymmetric. In lambda, a mismatch between interacting overlap sequences blocks recombination after the first round of strand cleavage because branch migration cannot occur (Nunes-Duby *et al.*, 1987; Nash and Robertson, 1989). In Cre- and FLP-mediated recombination, substitution of the asymmetric overlap region with a symmetrical sequence results in the formation of aberrant recombination products (inappropriate inversions and deletions) (Hoess *et al.*, 1986; Senecoff and Cox, 1986). Based on these results it has been proposed that in these systems, the role of asymmetry in core sequences is to ensure that only correctly aligned molecules complete recombination. In any Tn*916* transposition event, the coupling (overlap) sequences can be different and therefore it seems unlikely that recombination proceeds by a sequential mechanism involving branch migration. Therefore there would be no control of the recombination reaction at this level. In XerC/XerD-mediated recombination at a *dif* site, the essential asymmetry is provided by a requirement for two different recombinases to bind each half site (Blakely *et al.*, 1993). Our results show that a single recombinase can bind both half sites, leaving integrase bridging as the only way to ensure correct alignment of recombining sites.

As has been pointed out for Tn*1545* (Trieu-Cuot *et al.*, 1993), the lack of homology between the coupling sequences of interacting recombination sites has an important implication for the mechanism of recombination. If a sequential mechanism involving branch migration does not occur, it may be that Tn*916* excision and integration can only occur when all four DNA strands are cleaved at the same time. At present we have no data concerning the mechanism of strand cleavage by Tn*916* integrase, but for the integrase family as a whole, there is no evidence that strand cleavage must be concerted (Stark *et al.*, 1992). It may be that the lack of homology between the coupling sequences as well as some structural feature of the interacting DNA−protein complexes increases the probability of cleavage occurring simultaneously, permitting recombination, while at the same time allowing the transposon to insert into and excise from a larger number of sites than would be the case if the interacting sequences had to be identical. The purification of integrase described here should allow us to begin a biochemical analysis of Tn*916* transposition and to investigate the mechanism of strand cleavage.

Materials and methods

Reagents

Chemicals were purchased from Fisher Scientific and Sigma. Restriction enzymes were purchased from New England Biolabs and Boehringer

Mannheim. DNase I was from Worthington. The MBP fusion cloning and purification system was obtained from New England Biolabs.

Expression and purification of integrase

Three DNA fragments were amplified from an *MaeI*−*Sau*3A fragment of Tn*916* using as primer pairs the oligonucleotides 5′-ATGTCAGAAAAA-AGACGTGAC-3′ and 5′-AAATGAATTCGTACTACTAAGC-3′, 5′-A-TGACACTCTGCCAGCTTTA-3′ and 5′-AAATGAATTCGTACTAC-TAAGC-3′, and 5′-ATGTCAGAAAAAAGACGTGAC-3′ and 5′-CTG-GCAAAGCTTCATTTACTATCATACA-3′. PCR was carried out using standard reagent concentrations and 2 mM $MgCl_2$ with incubation temperatures and times of 1 min at 94°C, 1 min at 42°C and 1 min at 72°C for 40 cycles. DNA fragments were purified by electroelution from agarose gels. After treatment with S1 nuclease and *Eco*RI, the fragments were ligated to pMAL-c2 that had been treated with *Xmn*I and *Eco*RI to produce the plasmids pMBP-INT_{3-405}, pMBP-INT_{82-405} and pMBP-INT_{3-78}. Cultures of *E.coli* DH5α containing pMBP-INT plasmids were grown to an A_{600} of 0.5 and expression of fusion protein was induced by addition of 0.3 mM IPTG. After a further 2 h incubation, cells were harvested, suspended in buffer containing 10 mM Tris−Cl pH 7.4, 200 mM NaCl, 1 mM EDTA and 10 mM β-mercaptoethanol, lysed by passage through a French pressure cell, and centrifuged at 10 000 r.p.m. for 30 min. The supernatant was passed over an amylose column which was washed with 8 column volumes of the same buffer used to suspend the cells, and MBP fusion protein was eluted with 10 mM maltose as described by New England Biolabs. Protein concentrations were determined using the Bio-Rad DC protein assay kit.

Nuclease protection assays

Substrate fragments were prepared by PCR amplification of fragments from pUC18::Tn*1545*del4 and cloned into pUC18. The following primer pairs were used. Left end: 5′-CGCCAGGGTTTTCCCAGTCACGAC-3′ and 5′-GATCGGATCCGCTTGAATAAAGAGAAGC-3′. Right end: 5′-AG-CGGATAACAATTTCACACAGGA-3′ and 5′-AAATATTCAAGCTTT-ATTGGAAGT-3′. Site R_L: 5′-CTATGTGCTGTCGACTATTTT-3′ and 5′-GATCGGATCCGCTTGAATAAAGAGAAGC-3′. Site C_R: 5′-TGT-CAAAAGAATTCCCATTT-3′ and 5′-GAGTGGTGTCGACCTTGAT-A-3′. Plasmid DNA was prepared using the Promega Magic Maxipreps DNA purification system. To produce 3′ end-labelled fragments, plasmid DNA was digested with an appropriate restriction endonuclease, radiolabelled with ^{32}P using the Klenow fragment of *E.coli* DNA polymerase I and an appropriate [α-^{32}P]dNTP, digested with a second restriction endonuclease, and the radiolabelled fragment was purified by electrophoresis on a 1.2% agarose gel followed by electroelution and purification through a Nensorb 20 cartridge (NEN Research Products). For DNase I protection assays, various amounts of purified MBP-INT protein were incubated with 80 fmol radiolabelled DNA fragment at room temperature for 20 min in 50 μl BB buffer containing 10 mM Tris−Cl pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM β-mercaptoethanol and 0.1% NP40. One unit of DNase I was added and incubation was continued for 30 s or 1 min (in the absence of MBP-INT protein), or 2.5 min (in the presence of MBP-INT protein). The reaction was stopped by addition of 12.5 μl of stop solution (0.2 M EDTA, 1.5 M ammonium acetate, 120 μg/ml yeast RNA). The samples were then extracted with phenol:chloroform 1:1 and precipitated with ethanol. The pellets were suspended in 10 μl formamide dye mixture and heated at 90°C for 2 min prior to loading on a 6 or 8% sequencing gel. For competition assays, MBP-INT protein was incubated with unlabelled competitor DNA for 20 min prior to addition of radiolabelled probe DNA. Following addition of the probe DNA, incubation was continued for a further 20 min before addition of DNase I.

Acknowledgements

We thank F.Bringel for the gift of a clone of Tn*916* integrase, A.Glasgow and C.Moran for useful comments on the manuscript, D.Clewell for permission to cite unpublished results and J.Scott for her persistent encouragement and many fruitful discussions. This work was supported by grants from the Emory University Research Committee and Emory University Medical School.

References

Blakely,G., May,G., McCulloch,R., Arciszewska,L.K., Burke,M., Lovett,S. and Sherratt,D.J. (1993) *Cell*, **75**, 351−361.

Bringel,F.G., Van Alstine,G.L. and Scott,J.R. (1992) *J. Bacteriol.*, **174**, 4036−4041.

Bushman,W., Thompson,J.F., Vargas,L. and Landy,A. (1985) *Science*, **230**, 906−911.

Caparon,M.G. and Scott,J.R. (1989) *Cell*, **59**, 1027−1034.

Clewell,D.B., Flannagan,S.E., Ike,Y., Jones,J.M. and Gawron-Burke,C. (1988) *J. Bacteriol.*, **170**, 3046−3052.

Clewell,D.B., Flannagan,S.E., Zitzow,L.A., Su,Y.A., He,P., Senghas,E. and Weaver,K.E. (1991) In Dunny,G.M., Cleary,P.P. and McKay,L.L. (eds), *Genetics and Molecular Biology of Streptococci, Lactococci and Enterococci*. American Society for Microbiology, Washington, DC, pp. 39−44.

Courvalin,P. and Carlier,C. (1986) *Mol. Gen. Genet.*, **205**, 291−297.

Derbyshire,K.M. and Grindley,N.D.F. (1992) *EMBO J.*, **11**, 3449−3455.

Franke,A.E. and Clewell,D.B. (1981) *J. Bacteriol.*, **145**, 494−502.

Galas,D.J. and Schmitz,A. (1978) *Nucleic Acids Res.*, **5**, 3157−3170.

Gardner,J.F. and Nash,H.A. (1986) *J. Mol. Biol.*, **191**, 181−189.

Gawron-Burke,C. and Clewell,D.B. (1982) *Nature*, **300**, 281−284.

Hoess,R.H., Wierzbicki,A. and Abremski,K. (1986) *Nucleic Acids Res.*, **14**, 2287−2301.

Kim,S. and Landy,A. (1992) *Science*, **256**, 198−203.

Kim,S., Moitoso de Vargas,L., Nunes-Duby,S.E. and Landy,A. (1990) *Cell*, **63**, 773−781.

Kitts,P.A. and Nash,H.A. (1988) *J. Mol. Biol.*, **204**, 95−107.

Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G. and Skalka,A.M. (1992) *Mol. Cell. Biol.*, **12**, 2331−2338.

Landy,A. (1989) *Annu. Rev. Biochem.*, **58**, 913−949.

Moffatt,B.A. and Studier,F.W. (1987) *Cell*, **49**, 221−227.

Moitoso de Vargas,L., Pargellis,C.A., Hasan,N.M., Bushman,E.W. and Landy,A. (1988) *Cell*, **54**, 923−929.

Nash,H.A. and Robertson,C.A. (1989) *EMBO J.*, **11**, 3523−3533.

Nunes-Duby,S.E., Matsumoto,L. and Landy,A. (1987) *Cell*, **50**, 779−788.

Pargellis,C.A., Nunes-Duby,S.E., Moitoso de Vargas,L. and Landy,A. (1988) *J. Biol. Chem.*, **263**, 7678−7685.

Parsons,R.L., Prasad,P.V., Harshey,R.M. and Jayaram,M. (1988) *Mol. Cell. Biol.*, **8**, 3303−3310.

Parsons,R.L., Evans,B.R., Zheng,L. and Jayaram,M. (1990) *J. Biol. Chem.*, **265**, 4527−4533.

Poyart-Salmeron,C., Trieu-Cuot,P., Carlier,C. and Courvalin,P. (1989) *EMBO J.*, **8**, 2425−2433.

Robertson,C.A. and Nash,H.A. (1988) *J. Biol. Chem.*, **263**, 3554−3557.

Scott,J.R. (1992) *J. Bacteriol.*, **174**, 6005−6010.

Scott,J.R. (1993) In Sonenshein,A.L., Hoch,J.A. and Losick,R. (eds), *Bacillus subtilis and Other Gram-positive Bacteria*. American Society for Microbiology, Washington, DC.

Scott,J.R., Kirchman,P.A. and Caparon,M.G. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4809−4813.

Senecoff,J.F. and Cox,M.M. (1986) *J. Biol. Chem.*, **261**, 7380−7386.

Senghas,E., Jones,J.M., Yamamoto,M., Gawron-Burke,C. and Clewell,D.B. (1988) *J. Bacteriol.*, **170**, 245−249.

Stark,W.M., Boocock,M.R. and Sherratt,D.J. (1992) *Trends Genet.*, **8**, 432−439.

Storrs,M.J., Carlier,C., Poyart-Salmeron,C., Trieu-Cuot,P. and Courvalin,P. (1991) *J. Bacteriol.*, **173**, 4347−4352.

Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113−130.

Thompson,J.F., Waechter Brulla,D., Gumport,R.I., Gardner,J.F., Moitoso de Vargas,L. and Landy,A. (1986) *J. Bacteriol.*, **168**, 1343−1351.

Trieu-Cuot,P., Carlier,C., Poyart-Salmeron,C. and Courvalin,P. (1991) *Gene*, **106**, 21−27.

Trieu-Cuot,P., Poyart-Salmeron,C., Carlier,C. and Courvalin,P. (1993) *Mol. Microbiol.*, **8**, 179−185.

Weisberg,R.A., Enquist,L.W., Foeller,C. and Landy,A. (1983) *J. Mol. Biol.*, **181**, 189−197.

Received on November 30, 1993; revised on January 19, 1994