The EMBO Journal vol.6 no.10 pp.3163–3169, 1987

Expression of proteins essential for IS*1* transposition: specific binding of InsA to the ends of IS*1*

Didier Zerbib, Michael Jakowec[1], Pierre Prentki[1], David J.Galas[1] and Michael Chandler

Centre de Recherche de Biochimie et Genetique Cellulaire du C.N.R.S., 118 Rte. de Narbonne, 31062 Toulouse Cedex, France, and [1]Department of Molecular Biology, University of Southern California, Los Angeles, CA 90089-1481, USA

Communicated by J.Davies

The insertion sequence IS*1* displays a complex array of open reading frames (ORF). In an attempt to identify those which encode polypeptide products, we have systematically placed each ORF under the control of the P1 promoter of phage lambda. In the expression system we used, only the product of the *ins*A gene was present in high enough amounts to be detected by polyacrylamide gel electrophoresis. The production of InsA was further increased in a first codon hook-up to phage T7 transcriptional and translational initiation signals. Cell extracts from InsA overproducers display a DNA binding activity specific for the ends of IS*1*. This activity was identified as the InsA protein itself.

Key words: DNA binding protein/transposable element/T7 promoter/*E. coli* extracts/gel retardation

Introduction

IS*1* is one of the smallest known functional insertion sequences. It is only 768 bp long (Johnsrud, 1979; Ohtsubo and Ohtsubo, 1978), exhibits eight open reading frames (ORF) of greater than 50 amino acids (six of which are indicated in Figure 1, see legend) and has been reported to carry a promoter at each extremity (Machida *et al.*, 1984a). It has been demonstrated that mutations in two of these reading frames result in the abolition of IS-mediated plasmid fusion (cointegrate formation; Machida *et al.*, 1984b) in *Escherichia coli*. We have recently confirmed this conclusion by performing site-directed mutagenesis in each of the ORFs of IS*1* (Jackowec *et al.* in preparation). These essential reading frames, *ins*A and *ins*B, are transcribed from the same strand of IS*1*, presumably from the promoter located at the left extremity (IRL) of the element (Figure 1). Little is known about the function of these genes and the fashion in which their expression is controlled. The presence of a rho-dependent termination signal located in the 3′ region of *ins*A or between the *ins*A and *ins*B coding sequences (Prentki *et al.*, 1986; Hübner *et al.*, 1987; Hollingshead, personal communication) suggests that expression of *ins*B may be controlled in part at the transcriptional level. A larger open reading frame (*ins*B′), which extends *ins*B by 42 amino acids at the NH_2 end resulting in an overlap of 78 bp with the 3′ end of *ins*A, is also apparent.

To directly analyse the expression and role of IS*1*-encoded genes, we have undertaken to clone and examine the expression of several of the IS*1* ORFs. In the study presented here, we report the cloning of several open reading frames and the expression and initial characterization of one IS*1*-encoded protein, InsA. We demonstrate, moreover, that InsA binds specifically to the end of IS*1*.

Results

Cloning and expression of the major IS1 ORF

Relevant sequence features of IS*1* are shown in Figure 1. To determine whether the major reading frames are capable of directing protein synthesis, we have inserted IS*1* restriction fragments into the expression vector pP1c236 (Remaut *et al.*, 1982) which places the P1 promoter of bacteriophage lambda just upstream of the inserts. In the first series of experiments, we constructed plasmids pOC7 (*ins*A), pOC8 (*ins*D), pOC9 (*ins*B) and pOC26 (*ins*B′, *ins*B), which are described in Table I.

These plasmids were examined for their ability to direct the synthesis of IS*1*-specified proteins. We have used three different systems: coupled *in vitro* transcription/translation (Zubay, 1973), minicells and maxicells. Examples of this analysis are shown in Figure 2. The only reading frame which was found to direct detectable levels of protein is InsA (pOC7; lanes 2 and 3). No detectable levels of InsB′ (pOC26; lane 6), InsB (pOC9; lane 5) or InsD (pOC8; lane 4) were observed in any of the three systems used. Some variation in minor bands can be seen. For a given plasmid, these tend to be associated with the particular expression system used to direct protein synthesis. In no case were these proteins specific to the cloned IS*1* sequences (data not shown).

One factor which complicates the interpretation of these results is that in all cases DNA external to IS*1* is also introduced into the expression vector. We have therefore constructed a second set of plasmids using pKH47::IS*1*.50 (Zerbib *et al.*, 1985), from which IS*1* can be obtained as an *Eco*RI fragment carrying only the *Eco*RI recognition sequences external to the element. Inserting this into the single *Eco*RI site of pP1c236 generated pZB5 (IS*1* transcribed in the direction *ins*A–*ins*B), and pZB4 (IS*1* transcribed in the direction *ins*C–*ins*D) (Table I). Plasmid pZB5 was subsequently deleted between the *Bam*HI site in the polylinker and either the *Bst*EII or *Mlu*I site within IS*1* to generate

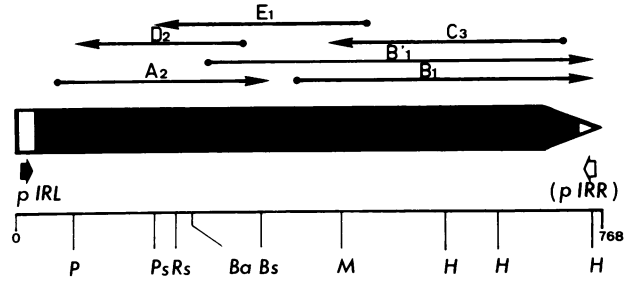


Fig. 1. Physical map of IS*1*. Sites for relevant restriction enzymes used in cloning are shown: P (*Pvu*II); Ps (*Pst*I); Rs (*Rsa*I); Ba (*Bal*I); Bs (*Bst*EII); M (*Mlu*I); H (*Hin*fI). The six major ORFs (A, B, B′, C, D and E) are also indicated together with their relative reading frames (1, 2 and 3). The IS*1* coordinates of these reading frames are: A, 56–329; B′, 250–750; B, 376–750; C, 719–432; D, 304–92; E, 468–202. Two additional reading frames, F (phase 3, coordinate 27–236) and G (phase 2, coordinate 353–508) have not been considered in detail in the text since they are carried by several of the clones and do not yield detectable products. pIRL and pIRR represent two promoters partially located within the terminal 23 bp inverted repeats of the element.

Table I. Construction of plasmids using the P1 promoter

IS1 Donor	Vector	Recombinant Plasmid	ORF	Proteins Observed
pTc 72.3	pPLC 236	pOC7	A	9.8 kD$_{(1,2,3)}$
pR 2.5 :: IS1		pOC 9	B	(-)$_{(3)}$
pR 2.4 :: IS1		pOC 8	D	(-)$_{(1,3)}$
pR2.5::IS1		pOC26	B',B	(-)$_{(3)}$
pKH47::IS1.50		pZB4 & pZB5	C,D,E A,B',B	(-)$_{(3)}$ 9.8 kD$_{(2,3)}$
pZB5		pZB6	A	9.8 kD$_{(2,3)}$
pZB5		pZB7	A	9.8 kD$_{(2,3)}$
pTc 72.24		pZB26	A,B',B	9.8 kD$_{(3)}$
pMJ5		pZB27	A$_{am}$,B',B	(-)$_{(3)}$

Relevant restriction sites are indicated: B, *Bam*HI; Bs, *Bst*EII; M, *Mlu*I; P, *Pvu*II; Ps, *Pst*I; R, *Eco*RI; Sa, *Sal*I; Ss, *Ssp*I. Sites used in these constructions are marked with a star. The heavy arrow indicates the vector-associated P1 promoter. The coordinates indicated are those of pBR322. Subscripts 1, 2, and 3 (last column) indicate minicell, maxicell and Zubay analysis respectively.

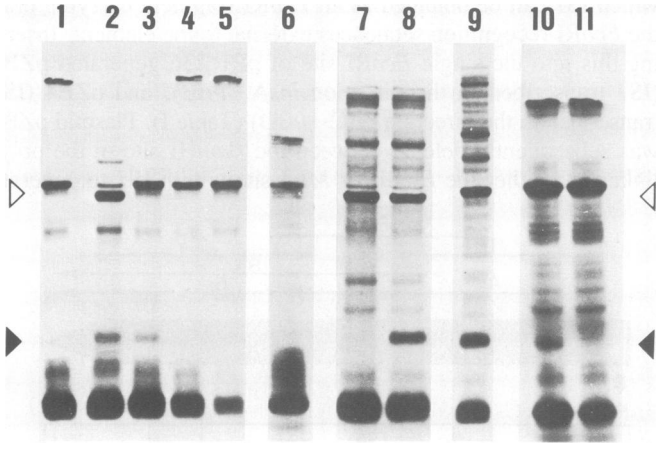


Fig. 2. Translation products of cloned IS1 reading frames. Results of electrophoresis of labelled proteins using a 15% acrylamide/urea/SDS gel. **Lane 1**: pP1c236 (vector); **lanes 2** and **3**: pOC7 (*ins*A); **lane 4**: pOC8 (*ins*D); **lane 5**: pOC9 (*ins*B); **lane 6**: pOC26 (*ins*B', *ins*B); **lane 7**: pZB4 (*ins*D, *ins*E); **lane 8**: pZB5 (*ins*A, *ins*B', *ins*B); **lane 9**: pZB6 (*ins*A); **lane 10**: pZB26 (*ins*A, *ins*B', *ins*B); **lane 11**: pZB27 (*ins*Aam, *ins*B', *ins*B). The structure of these plasmids is presented in Table I. This figure is a composite. **Lanes 1, 2, 3, 4** and **5**, **lane 6**, **lanes 7** and **8**, **lane 9**, and **lanes 10** and **11** are from separate gels. **Lane 2** shows expression using the minicell system; **lanes 1, 3, 4, 5, 6, 10** and **11** using the Zubay system; **lanes 7, 8** and **9** using the maxicell system. Vector specified beta-lactamase is indicated by arrows, and protein InsA by closed arrows.

pZB6 and pZB7 respectively (Table I). Both plasmids retain the *ins*A reading frame but lack downstream portions of IS*1*.

The patterns of protein synthesis directed by these plasmids in the maxicell system (Figure 2, lanes 7–9) are similar to those just described: a protein of 9.8 kd appears in all clones in which the *ins*A reading frame is in the correct orientation for P1-directed transcription (pZB5, lane 8; pZB6, lane 9). No additional protein is observed in clones encoding InsB'/InsB (lanes 6 and 8), InsB (lanes 5, 6 and 8), InsC, InsD or InsE (lanes 4 and 7). It is noteworthy that a 9.8 kd protein is apparent even in the case of pZB5 (lane 8) which carries the entire IS*1* in the correct orientation for transcription. The apparent mol. wt. of this protein is in good agreement with that predicted for InsA (9.884 kd) from its nucleotide sequence. That the production of the 9.8 kd protein is under control of the external P1 promoter is indicated by the absence of such a species in similar constructions without P1 (data not shown).

The 9.8 kd protein derives from the insA reading frame

To identify the 9.8 kd protein positively as InsA, we introduced an amber mutation by site-directed methods (a C to T transition at IS*1* coordinate 179) a codon 42 of the *ins*A gene in such a way that the corresponding mutation in the overlapping reading frame, *ins*D, is neutral. An additional mutation (A to G transition), neutral for both *ins*A and *ins*D, was introduced concomitantly at coordinate 175. Together, these mutations destroy the unique *Pst*I site in IS*1* and generate a second *Pvu*II site. The plasmid with the mutant IS*1* is pMJ5 (Materials and methods).

Both wild-type (from pTC72.24) and mutant (from pMJ5) *ins*A genes were inserted into pP1c236 to generate pZB26 and pZB27 respectively (Table I). Figure 2 (lanes 10 and 11) shows the pattern of plasmid directed protein synthesis in the Zubay system using a non-suppressing cell extract. Clearly, the wild-type *ins*A gene (pZB26; lane 10) directs synthesis of a 9.8 kd protein whereas the amber mutant (pZB27; lane 11) does not. We therefore conclude that the protein is InsA. The corresponding amber fragment truncated at codon 42 should carry no methionine residues and will therefore not be labelled.

Transcription termination within IS1

Insertions of IS*1* are polar in both orientations, and a rho-dependent transcription terminator (Das *et al.*, 1977; Reyes *et al.*, 1979) is located either in the 3′ extremity of *ins*A or between *ins*A and *ins*B (Prentki *et al.*, 1986; Hübner *et al.*, 1987; Hollingshead, personal communication). In addition, IS*1* transposition frequencies are influenced by rho (Datta and Rosner, 1987). Such an arrangement may interfere with expression of InsB′ and InsB proteins. To determine whether transcription termination alone could explain the absence of detectable proteins from reading frames other than *ins*A, we have introduced IS*1* and various IS*1* restriction fragments into an *in vivo* expression system which relies on a phage T7 promoter carried on a plasmid vector, and the supply of T7 RNA polymerase from a second, compatible plasmid (in this case, pGP1.2; Tabor and Richardson, 1985). We expect T7 RNA polymerase to traverse the host polymerase termination signal within IS*1*.

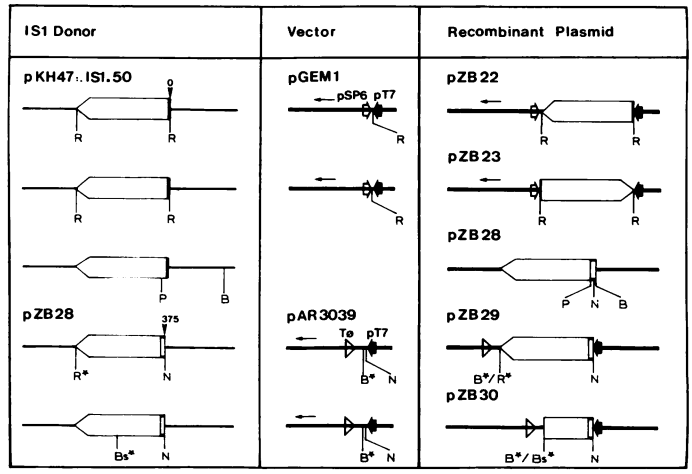
In a first series of experiments, the IS*1*-carrying *Eco*RI fragment of pKH47::IS*1*.50 was inserted in both orientations into the unique *Eco*RI site of the plasmid pGEM1 (Table II) to generate pZB22 (IS*1* transcribed in the *ins*A orientation) and pZB23 (opposite orientation), and the pattern of T7 directed expression was examined in whole cells (Figure 3). Only the results obtained with pZB22 are shown. Identical patterns were observed with pZB23 (data not shown). Surprisingly, in view of the results obtained using the P1 promoter of phage lambda, we do not observe any IS*1* specific translation products (data not shown), although T7 directed synthesis of beta-lactamase occurs (lane 3), even when a copy of IS*1* is inserted between the T7 promoter and the *bla* gene. Two types of promoter are available to direct beta-lactamase synthesis from these plasmids: the *bla* promoter normally resident on pBR322, and the T7 promoter. The high production of beta-lactamase at 30°C (lane 1) is presumably due to escape synthesis of T7 RNA polymerase at this temperature. At 42°C, a condition under which synthesis of T7 RNA polymerase is induced, it will be noted that a large fraction of the beta-lactamase remains in its higher mol. wt precursor form. This may be due to a saturation of the processing system. These results demonstrate that T7-directed transcription traverses IS*1* in both orientations, and suggests that the absence of IS*1* proteins is due to inefficient translation or differential stabilities of the mRNAs or protein products.

High level expression of InsA protein

To improve expression of InsA, we have constructed fusions which place the *ins*A reading frame under control of the strong T7 transcription/translation initiation signals in plasmid pAR3039 (Studier and Moffatt, 1986, and personal communication). In addition to the T7 promoter, this plasmid carries a T7 transcription terminator between the beta-lactamase reading frame and the cloning sites, thus strongly reducing transcription of the *bla* gene. The constructions were performed in two steps (see Materials and methods for details). They exchange the GTG initiation codon of *ins*A for an ATG, and position the beginning of the reading frame, either alone (pZB30) or with the remainder of IS*1* (pZB29), 9 bp downstream from the ribosome binding site carried by the vector.

When supplied with T7 RNA polymerase (from plasmid pGP1.2) *in trans*, pZB30 was found to direct rifampicin-resistant synthesis of a major, 9.8 kd polypeptide (Figure 4, lane 4). The presence of this protein can also be detected by silver staining (data not shown). Identical results were obtained with pZB29 (data not shown). In the case of pZB29, there is no evidence

Table II. Construction of plasmids using SP6 and T7 promoters



Restriction sites shown are as in the legend to Table I. In addition, the *Nde*I site which carries the ATG initiation codon is indicated (N), as are the T7 and SP6 promoters (closed and open arrows respectively) and the T7 transcription terminator (T_o). The coordinates indicated are those of pBR322.

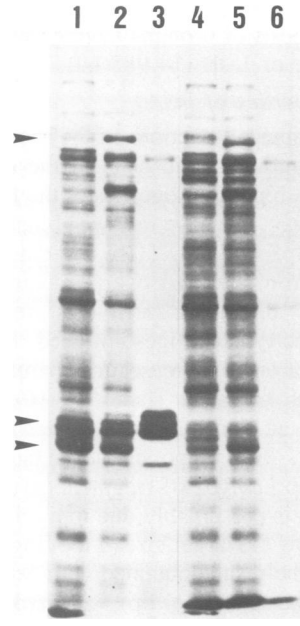


Fig. 3. T7 RNA polymerase is insensitive to the IS*1* transcription terminator. Separation of labelled proteins by 12.5% SDS–PAGE. **Lanes 1, 2** and **3**: pZB22 + pGP1.2; **lanes 4, 5** and **6**: pGp1.2 alone. **Lanes 1** and **4**: 30°C; **lanes 2** and **5**: 42°C; **lanes 3** and **6**: 42°C + rif. The arrows show the expected positions of T7 RNA polymerase, beta-lactamase and cI, in order of decreasing mol. wt.

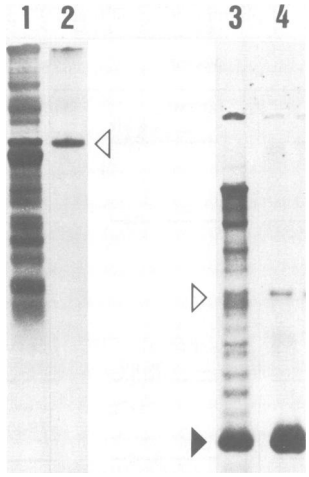


Fig. 4. High level expression of InsA. Separation of labelled proteins by electrophoresis on a 15% acrylamide/urea/SDS gel. **Lanes 1** and **2**: vector pAR3039 + pGP1.2; **lanes 3** and **4**: pZB30 + pGP1.2. **Lanes 1** and **3**: 42°C; **lanes 2** and **4**: 42°C + rif. Open arrows indicate beta-lactamase, closed arrow indicates the position of InsA.

of significant synthesis of proteins having the size expected for InsB′ (19.798 kd), or InsB (14.985 kd).

DNA binding properties of InsA

Since one of the principal functions of the transposition complex must be the recognition of the ends of the element, we have examined the DNA binding properties of the InsA protein using a gel retardation assay. We have followed binding of end-labelled DNA fragments of pKH47::IS*1*.50 by proteins in crude bacterial extracts obtained from the InsA overproducing pZB30/pGP1.2 T7 system. Results of these experiments are shown in Figure 5, together with the type of digestion used. In the presence of InsA we observe a clear increase in apparent mol. wt of the 185 bp *Eco*RI/*Pst*I fragment (lanes 8−10) and the 214 bp *Hin*dIII/*Pst*I fragment (lanes 2−4). These fragments carry the left end (IRL) of IS*1*. It is not possible from these results to detect a shift in the 593 bp *Eco*RI/*Pst*I fragment carrying the right end (IRR) of IS*1*. This is presumably due to a relatively small shift which is obscured by the larger size of the fragment (see below). Lanes 4−6 show the results obtained with *Ssp*I/*Eco*RI digested pKH47::IS*1*.50. The small (194 bp) *Ssp*I/*Eco*RI fragment of this plasmid has a size between that of the *Pst*I/*Eco*RI and *Pst*I/*Hin*dIII IRL-carrying fragments but carries no IS*1* sequences. It exhibits no specific retardation.

To define the position of the InsA binding site within the left end of IS*1* further, plasmid pKH47::IS*1*.50 was cleaved with *Hin*dIII, end-labelled and cleaved with *Hgi*AI, to generate an 85 bp fragment carrying 52 bp of IRL. The results (Figure 6, lanes 1−4) demonstrate a retardation equivalent to a size increase of 45 bp. Binding to the right end of IS*1* was examined by constructing a plasmid, pZB10, in which we inverted the IS*1* carried by pKH47::IS*1*.50 (by digestion with *Eco*RI, and religation). This plasmid was cleaved with *Hin*dIII, end-labelled, and cleaved with *Tth*111I, to generate an 88 bp fragment carrying 55 bp of IRR. Again, the results shown in Figure 6 (lanes 5−8) demonstrate a clear retardation of this fragment (increase in apparent size: 35 bp). Thus there is specific binding at sites located within the first 55 bp at both ends of IS*1*. It will be noted that the intensity of the bands decreases with increasing levels of extract. This

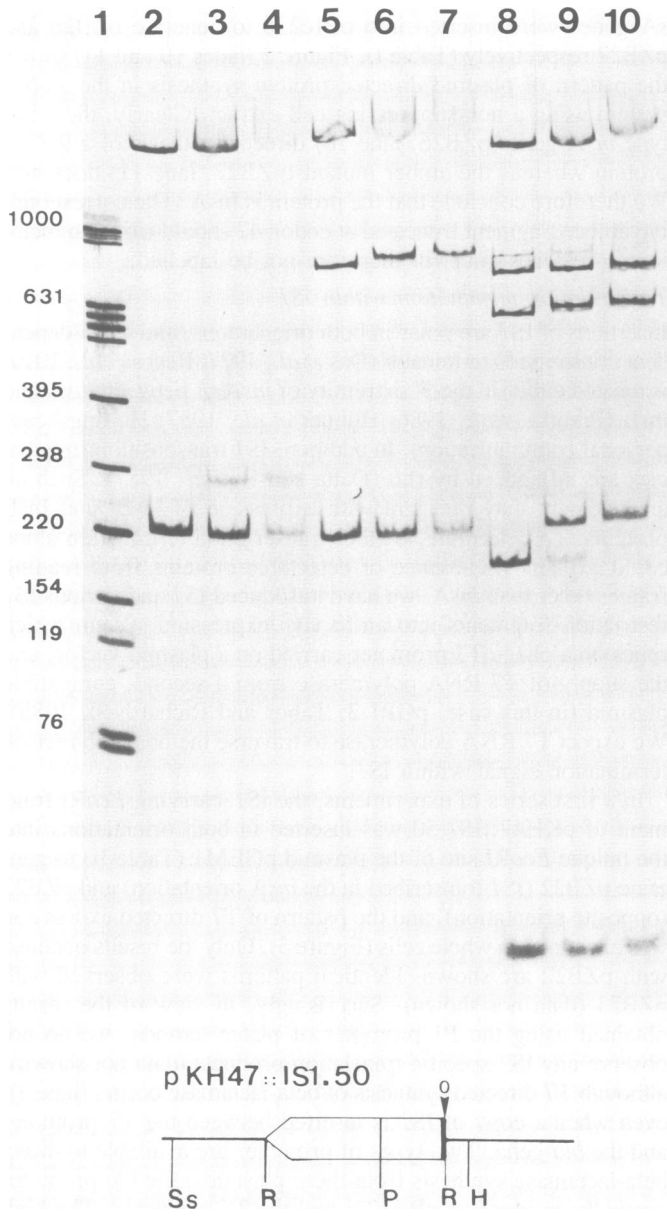


Fig. 5. Retardation of IS*1*-carrying DNA fragments of pKH47::IS*1*.50 by a cell extract enriched for InsA. **Lane 1**: pKH47::IS*1*.50 cleaved with *Eco*RI and *Hin*fI and end-labelled. **Lanes 2, 3** and **4**: *Hin*dIII/*Pst*I, end-labelled at the *Hin*dIII site; **lanes 5, 6** and **7**: *Eco*RI/*Ssp*I, end-labelled at the *Eco*RI site; **lanes 8, 9** and **10**: *Eco*RI/*Hin*dIII/*Pst*I, end-labelled at the *Eco*RI site. The restriction sites used are shown in the insert. The amount of extract added was: 0 μg, **lanes 2, 5** and **8**; 2.5 μg, **lanes 3, 6** and **9**; 5 μg, **lanes 4, 7** and **9**. Separation was on a 7.5% native acrylamide slab gel.

effect is as marked for non-specific as for specific (IS*1*-carrying) DNA fragments and is presumably due to the presence of non-specific DNA binding proteins in the extracts. The presence of such proteins would be expected not only to produce 'smearing' but also to prevent DNA from entering the gel.

To confirm that the retardation observed in Figures 5 and 6 is due to binding of InsA, we have performed similar retardation assays using [^{35}S]methionione labelled cell extracts (with or without InsA) and unlabelled plasmid DNA fragments. The results are presented in Figure 7. Lane 1 shows the position of the four fragments of plasmid pKH47::IS*1*.50 cleaved with *Eco*RI, end-labelled, cleaved with *Pst*I and exposed to an un-

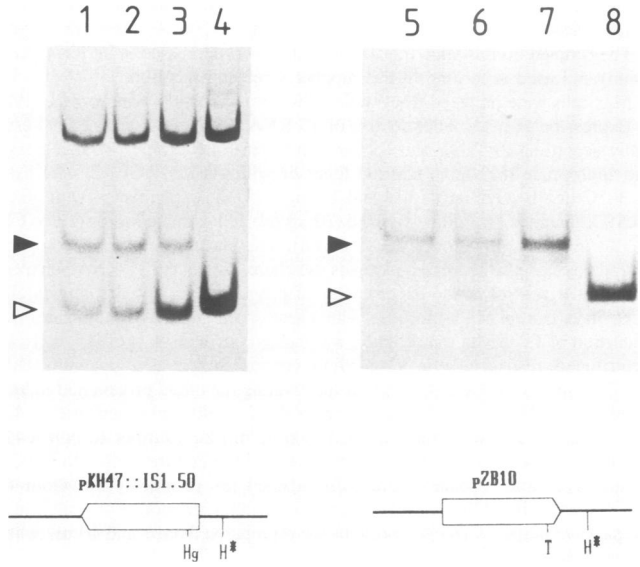


Fig. 6. Specificity of InsA binding to the ends of IS*1*. **Lanes 1–4** (IRL): pKH47::IS*1*.50, cleaved with *Hin*dIII, end-labelled and cleaved with *Hgi*AI; **lanes 5–7** (IRR): pZB10, cleaved with *Hin*dIII, end-labelled and cleaved with *Tth*111I. The restriction sites used are shown in the insert. The amount of extract added was: 0 μg, **lanes 4, 8**; 2.5 μg, **lanes 3, 7**; 5 μg, **lanes 2, 6**; 7.5 μg, **lanes 1, 5**. Electrophoresis was as in Figure 5.

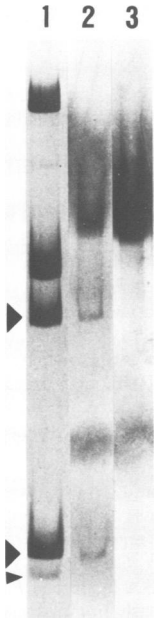


Fig. 7. Binding of labelled extracts to unlabelled IS*1*-carrying DNA fragments of pKH47::IS*1*.50. **Lane 1**: end-labelled DNA as in **lane 9** of Figure 5; **lanes 2** and **3**: unlabelled DNA digested with *Eco*RI, *Hin*dIII and *Pst*I together with labelled extract with or without InsA respectively. Arrows indicate the IS1-specific fragments. The small arrow indicates the position of the unbound IRL fragment. Separation was on a 3.75% native acrylamide slab gel.

labelled cell extract containing InsA (as in Figure 5, lane 9). Two of the four fragments (185 bp and 593 bp) carry IS*1* sequences. Note that the 185 bp fragment is detectably retarded in the presence of extract. Lane 2 shows the results of addition of a labelled InsA-containing cell extract to unlabelled *Eco*RI/*Pst*I-cleaved pKH47::IS*1*.50. It can be seen that of the four DNA fragments, only two, those carrying IS*1* sequences, are specifically bound by the extract containing labelled InsA. Moreover, the position of the lower band is identical to that of the retarded band observed with unlabelled extract and end-labelled DNA (lane 1). In the absence of exogenous DNA, no specific protein bands can be detected at these positions (data not shown). Specific labelling is not detected if the labelled extract does not carry InsA (lane 3). The labelled material in the upper part of the gel represents vector specified beta-lactamase present in this preparation.

Discussion

We have cloned and attempted to over-express all the major open reading frames carried by IS*1*. Of these, only one, *ins*A, was shown to direct detectable levels of a specific polypeptide (9.8 kd). Expression was observed when transcription was driven either by the P1 promoter of phage lambda (in minicells, maxicells or a coupled *in vitro* system; Figure 2) or by a T7 promoter (Figure 4) with and without the physical end of IS*1*. The identity of InsA with the 9.8 kd protein was confirmed by demonstrating that an IS*1* mutant carrying an amber codon in the *ins*A gene, but otherwise identical, does not direct its synthesis (Figure 2; lanes 10 and 11).

We have been unable to detect proteins specified by the other IS*1* reading frames. A protein product from at least one other reading frame (*ins*B) is implied by the isolation of nonsense mutations in the gene (Machida *et al.*, 1984b; Jakowec *et al.*, in preparation). Furthermore we have shown that InsB activity can depend, according to the position of the mutation, on the suppressing amino acid (to be published).

While the presence of a transcription terminator within IS*1* might explain the absence of other detectable proteins, this terminator is not recognized efficiently by T7 RNA polymerase: note the high level of T7-driven beta-lactamase synthesis from transcripts which must traverse IS*1* in either orientation (Figure 3). Expression of these genes may therefore be limited at the translational level and we note that, using SP6 directed IS*1* transcripts from the pGEM1 derivative plasmids pZB22 and pZB23, we have so far failed to obtain *in vitro* translation. Alternatively, the products may be unstable or they may migrate aberrantly in SDS gels. We are investigating these possibilities. It is possible that the low levels of expression implied by these results reflect 'down regulation' of these products by as yet unknown mechanisms, to keep transposition at low levels.

We have greatly increased the level of InsA synthesis by providing a strong T7 promoter, an efficient ribosome binding site and by substituting the GTG codon with an ATG initiation codon. Using crude cell extracts from cells carrying the InsA expressing plasmid, we have also demonstrated that InsA specifically recognizes the ends of IS*1*. Only IS*1*-carrying DNA fragments bind radioactively labelled InsA protein (Figure 7) and the sequences recognized have been identified with fragments containing only short end segments of IS*1* (Figure 6). We note that there is no significant difference in the intensity of the two bands representing InsA–DNA complexes, suggesting that InsA binds with a similar affinity to each fragment. The results also demonstrate that InsA does not bind detectably to the major IS*1* insertion hotspot in pBR322 (carried by the 193 bp pBR322 *Ssp*I/*Eco*RI fragment; see Figure 5). Since the only substantial homology between the ends of IS*1* resides in the 23 bp terminal inverted repeats, it is probable that InsA recognizes and binds sequences within these inverted repeats. The increase in apparent size of the complexes is relatively small. Much larger increases, resulting from bending of the DNA, have been observed upon specific binding

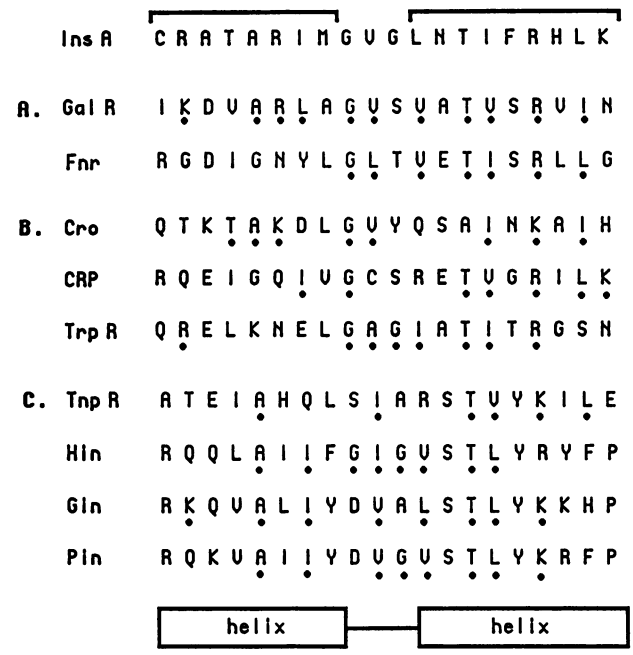


Fig. 8. Sequence comparison of the putative helix−turn−helix motif of InsA with the DNA binding domains of other bacterial proteins. Three sets of proteins are presented: (A) two repressors with striking similarities to InsA; (B) repressors whose crystal structure has been solved; (C) proteins involved in site-specific recombination. Sequences are taken from Pabo and Sauer (1984), except Hin (Zieg and Simon, 1980), Gin, and Pin (Plasterk *et al.*, 1983) and TnpR (Heffron *et al.*, 1979). Matches to InsA are indicated by dots. The hydrophobic residues Ala, Val, Ile and Leu are considered matches, as are Lys and Arg.

of the small basic IHF protein to fragments carrying the ends of IS*1* (Gamas *et al.*, 1987; Prentki *et al.*, 1987; H.Nash and A.Landy, personal communication). This suggests that InsA does not induce a major conformational change in the DNA upon binding.

Examination of the predicted amino acid sequence of InsA reveals the potential to form a helix−turn−helix structure characteristic of many DNA binding proteins (Pabo and Sauer, 1984; Figure 8). Model building has shown that the 28 amino acids at the carboxy terminus of InsA can assume a helix−turn−helix motif in which the carboxy-promixal helix is highly amphiphilic. The outward-facing residues are well placed to interact with the phosphate backbone and the major groove of B-DNA (Chandler and Galas, unpublished).

With the expression of InsA at high levels, we are now in a position to investigate the molecular details of the DNA binding properties of the protein and to probe the amino acid and DNA sequence specificities of binding. The present work also represents a significant step towards the development of an *in vitro* transposition system for IS*1*.

Materials and methods

Bacterial strains, growth conditions and plasmids

The bacterial strains used in this study were: DS410λ (Roulet *et al.*, 1984) for minicell production; CSR603 (Sancar *et al.*, 1981); and C600*rec*A$^-$ (our collection). The basic plasmids used in the various constructions shown in Tables I and II are as follows: pR2.4::IS*1* and pR2.5::IS*1* (Chandler and Galas, 1983); pTC72.3, pTC72.24 and pKH47::IS*1*.50 (Zerbib *et al.*, 1985); pP1c236 (Remaut *et al.*, 1981); pGEM1 (Promega Biotech); pGP1.2 (Tabor and Richardson, 1985); and pAR3039 (Studier and Moffat, 1986, and personal communication).

Strains were routinely grown in L broth or on L agar supplemented where appropriate with ampicillin (Ap; 50 μg/ml), tetracycline (Tc; 25 μg/ml), kanamycin (Km; 20 μg/ml), or streptomycin (Sm; 200 μg/ml).

Preparation, storage and use of minicells have been described previously (Roulet *et al.*, 1984).

The coupled transcription translation system was purchased in kit form (Amersham) and used according to the supplier's recommendations.

Maxicells were prepared from strain CSR603 according to Sancar *et al.* (1981).

Expression of InsA under control of T7 RNA polymerase was assayed essentially according to Tabor and Richardson (1985): C600*rec*A$^-$ strains carrying the appropriate derivative plasmid together with plasmid pGP1.2 were grown at 30°C in VB medium (Vogel and Bonner, 1956) supplemented with glucose (0.5%), thiamine (1 μg/ml), thymine (20 μg/ml) and casamino acids (Difco CSA, 0.5%) in the presence of IPTG (5 μg/ml). The plasmid pGP1.2 is compatible with the vector pBR322-based plasmids used here, carries the T7 RNA polymerase gene under control of the P1 promoter of phage lambda and the lambda cI857 gene under control of Plac together with a gene specifying resistance to kanamycin. Induction of T7 driven transcription was carried out at an A_{590} of 0.5: cells were centrifuged, resuspended in VB medium containing Met assay medium (Difco, 0.5%) instead of CSA at 42°C, thus inactivating the cI857 protein and inducing synthesis of T7 RNA polymerase from pGP1.2. Cells were incubated at 42°C for 30 min, treated with rifampicin (rif; 200 μg/ml) for a further 10 min at 42°C to inhibit transcription by host polymerase, and the culture shifted to 30°C for a further 20 min. Proteins were pulse-labelled for 10 min by the addition of L-[^{35}S]methionine at 20 μCi/ml (Amersham; >800 Ci/mmol). For analytical and preparative purposes labelled proteins were prepared from 1 and 10 ml cultures respectively.

For unlabelled cell extracts, 300 ml of VB-cas cultures grown at 30°C were shifted to 42°C at an A_{590} of 0.5 for 40 min, followed by 2 h growth at 30°C in the absence of rif. The cells were collected by centrifugation, washed with 10 ml of fresh medium and resuspended in 1 ml of buffer I (20 mM Hepes, pH 8, 100 mM NaCl, 2 mM EDTA, pH 8, 7 mM beta-mercaptoethanol, 1 mM PMSF and 10% glycerol). The cells were broken by freeze−thawing and/or sonication and the lysate was cleared by centrifugation at 45 000 r.p.m. in a Beckman Ti50 rotor. The pellet was discarded and the supernatant was retained for further use. All manipulations were performed at 4°C. Protein concentrations were determined using the Biorad protein assay according to the recommendations of the suppliers

Gel electrophoresis

Restriction digests of plasmid DNA were routinely analysed on horizontal agarose slab gels (in Tris−borate buffer, pH 8, containing 0.2 mM EDTA; Maniatis *et al.*, 1982) or on vertical acrylamide gels (acrylamide/bis acrylamide = 30:0.8) in Tris−borate buffer, containing 10 mM EDTA, at room temperature.

Acrylamide gels used in the visualization of protein/DNA complexes (Fried and Crothers, 1981; Garner and Revzin, 1981) were identical except that 2.5 mM EDTA was included in the buffer and the samples were subjected to electrophoresis at room temperature for 3−4 h at 10 V/cm.

DNA chain termination sequencing reactions were analysed on standard 0.4 mm thick sequencing gels (Maxam and Gilbert, 1977).

The synthesis of labelled proteins was monitored by electrophoresis in low mol. wt SDS−urea polyacrylamide gels essentially as described (Focus 3:6, 1984) at pH 7.4 with a stacking gel containing 0.15% SDS. The loading buffer was 10 mM H_3PO_4, 8 M urea adjusted to pH 8.2 and supplemented with 5% SDS, 2.5% beta-mercaptoethanol and 0.01% bromophenol blue.

DNA manipulation

Plasmid DNA was isolated from L broth-grown cultures of C600*rec*A$^-$ for both preparative and analytical purposes using the alkaline lysis procedure (Ish-Horowicz and Burke, 1981). Large scale preparations were centrifuged through CsCl gradients in the presence of EtBr (Clewell and Helinski, 1969).

Restriction enzymes, T4 DNA ligase and *E. coli* DNA polymerase Klenow fragment were purchased variously from: Amersham, Boehringer, New England Biolabs or Bethesda Research Laboratories, and used as recommended.

DNA fragments were prepared by electrophoresis onto DEAE cellulose paper (NA-45, Schleicher and Schuell).

Chain termination DNA sequencing reactions (Sanger *et al.*, 1977) were performed directly on analytical preparations of supercoiled plasmid DNA (Zagursky *et al.*, 1985) using the appropriate oligonucleotide primer and AMV reverse transcriptase (*Life Science*).

Oligonucleotides were synthesized using an Applied Biosystems 381A DNA synthesizer, and purified on polyacrylamide−urea gels followed by electroelution into low mol. wt cut-off dialysis bags.

Plasmid construction

(a) Reading frames driven by the P1 promoter of lambda. The pBR322-based p1 expression vector pP1c236 (Remaut *et al.*, 1981) was used. The *ins*A-carrying clone, pOC7, was obtained by insertion of an *Eco*RI/*Bst*EII fragment of the plasmid pTC72.3 (Table I; Prentki *et al.*, 1986) between the *Eco*RI and *Bam*HI sites in the poly-linker of the vector. Both *Bst*EII and *Bam*HI ends were made flush by treatment with Klenow enzyme (Maniatis *et al.*, 1982) prior to insertion. This results in the reconstitution of the putative InsA termination codon contained with

the *Bst*EII site of IS*1*. Similarly, the *ins*B clone, pOC9, was generated by insertion of a *Bst*EII/*Sal*I fragment of the plasmid pR2.5::IS*1* (Table I) between the *Bam*HI and *Sal*I sites of the vector after rendering both *Bst*EII and *Bam*HI ends flush. This construction reconstitutes the pBR322 Tc^R gene, only part of which is present in the vector. Plasmid pOC8, carrying the *ins*D reading frame, was obtained by insertion of a *Bst*EII/*Sal*I fragment of the plasmid pR2.4::IS*1* (Table I) between the *Bam*HI and *Sal*I sites of the vector in a similar manner to pOC9. Reading frame *ins*B' was obtained by insertion of a *Pvu*II/*Sal*I fragment between the *Bam*HI (rendered flush) and *Sal*I sites of pP1c235 to give pOC26. The construction of additional P1 plasmid derivatives is described in the text.

(b) InsA amber mutation. The construction of the amber mutant will be described in detail elsewhere. Briefly, a *Sca*I/*Eco*RI fragment of plasmid pTC72.24 (Table I; Prentki *et al.*, 1986) was cloned between the *Sca*I and *Eco*RI sites of the vector pZ152 (a pBR322 plasmid carrying the origin of replication of phage M13; Zagursky and Berman, 1984). Single-stranded plasmid DNA was prepared from strain BW313, an *ung*$^-$ *dut*$^-$ strain (Kunkel, 1985), and subjected to oligonucleotide-directed mutagenesis with a 'complementary' mutant oligonucleotide as described (Kunkel, 1985). Following transformation of strain HB101, several clones exhibiting an additional *Pvu*II site within IS*1* were isolated and the presence of the correct mutation was confirmed by DNA sequencing (Sanger *et al.*, 1977). Plasmid pMJ5 carries a mutated IS*1* which is inactive in transposition in a *su*^{0} host but active in an *Su*II or *Su*III host (to be published). An *Ssp*I/*Bam*HI fragment of this plasmid carrying IS*1* was inserted between the *Eco*RI (rendered flush) and *Bam*HI sites of pP1c235 to generate pZB27. A similar construction using the wild-type IS*1* from pTC72.24 yielded pZB26 (Table I).

(c) Reading frames driven by T7 promoters. Insertion of IS*1* into the T7 expression vector pGEM1 (Promega Biotec) to generate pZB22 and pZB23 is described in the text. Plasmids pZB29 and pZB30 which carry *ins*A coupled to strong T7 transcription/translation initiation signals upstream from an efficient T7 transcription termination signal were constructed as follows: we synthesized two complementary oligonucleotides:

5'-G A T C C A T A T G G C T T C T G T T T C T A T C A G-3'
3'-. . . . G T A T A C C G A A G A C A A A G A T A G T C-5'

able to form a double-stranded DNA fragment terminated on one side by a *Bam*HI cohesive end, followed by an *Nde*I site (which carries the substitute ATG initiation codon), codons 2–6 of InsA, and is terminated at the other end by a *Pvu*II extremity normally present in IS*1*. The region of pKH47::IS*1*.50 (Table I; Zerbib *et al.*, 1985) between the *Bam*HI site and the resident IS*1* *Pvu*II site was replaced by the unphosphorylated complementary oligonucleotides using 'linker-tailing' (Lathe *et al.*, 1984) to generate pZB28. The modified IS*1* or the *ins*A reading frame (in which the InsA GTG codon has been replaced by ATG — carried within an *Nde*I site) were isolated respectively as *Nde*I/*Eco*RI and *Nde*I/*Bst*EII fragments and inserted between the *Nde*I and *Eco*RI cloning sites of the T7 vector plasmid pAR3039 (Studier and Moffat, 1986; and personal communication) to generate pZB29 and pZB30 respectively. 3' recessed extremities were rendered flush, where appropriate, using Klenow enzyme. Insertion at the *Nde*I site reconstitutes the substitute ATG initiation codon and places it 9 bp downstream from a strong ribosome binding site of the vector, which is itself located 28 bp downstream from the T7 promoter. Transcription of the downstream beta-lactamase gene of the vector is reduced (but not abolished) by the presence of a T7 transcription termination signal closely following the cloned gene(s). The structure of pZB29 and pZB30 was confirmed directly by DNA sequencing.

Protein–DNA interactions

Between 240 ng and 6 μg of crude protein extract were incubated with 3'-end labelled restriction fragments (50–200 ng) together with 2.5 μg of unlabelled salmon sperm DNA to reduce non-specific binding. The binding reaction was carried out in a final volume of 25 μl at 25°C for 25 min in a buffer containing 50 mM Tris–HCl (pH 7.4), 70 mM KCl, 1 mM EDTA, 1 mM beta-mercaptoethanol, 7 mM $MgCl_2$, 3 mM $CaCl_2$, 10% glycerol, 200 μg/ml BSA (Boehringer). The reaction was terminated by addition of 5 μl of loading buffer (1 mg/ml BSA, 50% glycerol, 0.01% xylene cyanol) and loaded immediately onto the appropriate gel, at room temperature, while the gel was under tension (4 V/cm). Electrophoresis was for 3–4 h at 10 V/cm.

Acknowledgements

We would like to thank M.Bertermier, O.Fayet and P.Gamas for fruitful discussions, and J.-L. Klein for the construction of pZB26. D.Z. was a recipient of a predoctoral grant from the M.R.T. (France). This work was supported by the C.N.R.S. (M.C.), a C.N.R.S./N.S.F. collaborative grant to M.C. and D.G. and N.I.H. grant AI1936 to D.G.

References

Chandler,M. and Galas,D. (1983) *J. Mol. Biol.*, **170**, 61–79.
Clewell,D.B. and Helinski,D. (1969) *Proc. Natl. Acad. Sci. USA.*, **62**, 1159–1166.
Das,A., Court,D., Gottesman,M. and Adhya,S.L. (1977) In Bukhari,A.I., Shapiro,J.A. and Adhya,S.L. (eds), *DNA Insertion Elements, Plasmids and Episomes.* Cold Spring Harbor Laboratory Press, New York, pp. 93–97.
Datta,A.R. and Rosner,J.L. (1987) *J. Bacteriol.*, **169**, 888–890.
Fried,M. and Crothers,D. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
Gamas,P., Chandler,M., Prentki,P. and Galas,D. (1987) *J. Mol. Biol.*, **195**, 261–272.
Garner,M.M. and Revzin,A. (1981) *Nucleic Acids Res.*, **9**, 3047–3060.
Heffron,F., McCarthy,J., Ohtsubo,M. and Ohtsubo,E. (1979) *Cell*, **18**, 1153–1163.
Hübner,P., Iida,S. and Arber,W. (1987) *Mol. Gen. Genet.*, **206**, 485–490.
Ish-Horowicz,D. and Burke,J. (1981) *Nucleic Acids Res.*, **9**, 2989–2999.
Johnsrud,L. (1979) *Mol. Gen. Genet.*, **169**, 213–218.
Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA.*, **82**, 488–492.
Lathe,R., Kieny,M.P., Skory,S. and Lecocq,J.P. (1984) *DNA*, **3**, 173–182.
Machida,Y., Machida,C. and Ohtsubo,E. (1984a) *J. Mol. Biol.*, **177**, 247–267.
Machida,Y., Machida,C. and Ohtsubo,E. (1984b) *J. Mol. Biol.*, **177**, 229–246.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.
Maxam,M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Ohtsubo,E. and Ohtsubo,E. (1978) *Proc. Natl. Acad. Sci. USA.*, **75**, 615–619.
Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293–321.
Plasterk,R.H.A., Brinkman,A. and van de Putte,P. (1983) *Proc. Natl. Acad. Sci. USA.*, **80**, 5355–5358.
Prentki,P., Teter,B., Chandler,M. and Galas,D. (1986) *J. Mol. Biol.*, **191**, 383–393.
Prentki,P., Chandler,M. and Galas,D.J. (1987) *EMBO J.*, **6**, 2479–2487.
Remaut,E., Stanssens,P. and Fiers,W. (1981) *Gene*, **15**, 81–93.
Reyes,O., Gottesman,M. and Adhya,S. (1979) *J. Bacteriol.*, **126**, 1108–1112.
Roulet,E., Allet,B. and Chandler,M. (1984) *Gene*, **28**, 65–72.
Sancar,A., Wharton,R.P., Seltzer,S., Kacinski,B.M., Clarke,N.D. and Rupp, W.D. (1981) *J. Mol. Biol.*, **148**, 45.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA.*, **74**, 5463–5467.
Studier,F.W. and Moffat,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
Tabor,S. and Richardson,C.C. (1985) *Proc. Natl. Acad. Sci. USA.*, **82**, 1047–1078.
Vogel,J.H. and Bonner,D.M. (1956) *J. Biol. Chem.*, **218**, 97.
Zagursky,R.J. and Berman,M.L. (1984) *Gene*, **27**, 183–191.
Zagursky,R.J., Baumeister,K., Lomax,N. and Berman,M. (1985) *Gene Anal. Tech.*, **2**, 89–94.
Zerbib,D., Gamas,P., Chandler,M., Prentki,P., Bass,S. and Galas,D. (1985) *J. Mol. Biol.*, **185**, 517–524.
Zieg,J. and Simon,M. (1980) *Proc. Natl. Acad. Sci. USA.*, **77**, 4196–4201.
Zubay,G. (1973) *Annu. Rev. Genet.*, **7**, 267–287.

Received on May 18, 1987; revised on June 30, 1987