Binding of the IS903 transposase to its inverted repeat *in vitro*

Keith M.Derbyshire and Nigel D.F.Grindley

Department of Molecular Biophysics and Biochemistry, Yale University,
New Haven, CT 06510, USA

Communicated by D.J.Finnegan

We have purified the transposase of IS903 in three different ways. We find that transposase expressed as a fusion protein with either glutathione-S-transferase or maltose-binding protein is soluble and can be purified rapidly using affinity chromatography. The third purification requires extracting the native transposase from an insoluble pellet using an alkaline pH buffer. All three proteins bind specifically to the ends of IS903 and give identical patterns of protection when challenged with DNase I. We have used the more stable fusion proteins to examine transposase–DNA interactions *in vitro*. Methylation interference experiments have identified critical bases for transposase binding; methylated purines that inhibit binding all lie within the inner part of the 18 bp inverted repeat (bp 7–16). Moreover, the positions and identities of these purines suggest that the transposase interacts with base pairs in adjacent major and minor grooves. Binding assays with mutant inverted repeats confirm that transposase binding is sensitive to sequence changes only within this inner region. We propose that the transposase binding site is limited to this domain of the inverted repeat. These data are consistent with our previous analysis of the behaviour of mutant ends *in vivo*, from which we postulated that the inverted repeat was composed of two functional domains; an inner binding domain (bp 6–18), which included a region of minor groove interactions, and an outer domain that was involved in a step subsequent to transposase binding.

Key words: insertion sequence/inverted repeat/minor groove interactions/transposase binding

Introduction

IS903 is a small bacterial insertion sequence (IS) of 1057 bp. It is flanked by perfect 18 bp inverted repeats and encodes a 307 amino acid transposase protein which carries out the transposition process (Grindley and Joyce, 1981). The inverted repeats are the only genetic element required *in cis* for transposition and it is assumed that they are recognized and cleaved at their ends by the transposase before being inserted into the target DNA. Thus the specific interaction of the transposase with the inverted repeat is critical during transposition and in maintaining the integrity of the transposon. However, the transposase must not only recognize and cleave site specifically at its inverted repeats, but also retain the ability to recognize a target DNA non-specifically and mediate a staggered cleavage for transposon insertion (reviewed in Galas and Chandler, 1989). In an

attempt to understand these different protein–DNA interactions we have begun an analysis of the transposase–inverted repeat interaction.

We have already described a comprehensive analysis of this interaction *in vivo* by examining the effect of mutant inverted repeats on transposition (Derbyshire *et al.*, 1987). This genetic analysis led us to propose that the IS903 inverted repeat was structurally complex, consisting of two functional domains. The larger of these domains, extending from position 6 to 18, was involved in transposase binding. The second domain consisted of the outer three base pairs of the inverted repeat and was involved in a step subsequent to binding. We also postulated, based on hierarchical effects of different changes at individual positions, that the transposase recognized part of the inverted repeat via minor groove contacts. Mutational analyses of the inverted repeats of IS1 and IS10 have subsequently suggested that they too have two functionally distinct domains within their inverted repeats (Zerbib *et al.*, 1990; Huisman *et al.*, 1989).

We have extended our analysis of transposase–DNA interactions by investigating the binding of the IS903 transposase to its inverted repeat *in vitro*. We have identified the regions of the inverted repeat required for transposase recognition and binding, by analyzing the effects of both purine methylation and base pair substitutions on complex formation. The data confirm our original predictions; the outer part of the inverted repeat is not involved in the initial binding of transposase to DNA; the DNA sequence essential for transposase binding is confined to the region between positions 6 and 18, and a portion of this sequence (13–16) appears to be recognized via interactions with the minor groove of the DNA.

Results

Expression and purification of active, soluble forms of IS903 transposase

The transposase proteins of bacterial IS elements have been notoriously difficult to purify in a soluble active form. For several years we have attempted to obtain active IS903 transposase without success. Our attempts have included overexpressing the native protein in a number of bacterial systems under the control of inducible promoters such as λP_L (Young *et al.*, 1983) and T7 (Studier *et al.*, 1990), using a variety of different conditions and times of induction. We have also tried the insect–baculovirus overexpression system (Smith *et al.*, 1983). In addition, we have made translational fusions to β -galactosidase (Germino and Bastia, 1984) and cII of phage λ (Nagai and Thørgensen, 1984). In each case we have run into similar problems; the protein is poorly expressed and is insoluble. Solubilization of the protein from the pellet using urea, guanidinium–HCl or sarkosyl (Frankel *et al.*, 1991), followed by renaturation, has also proved unsuccessful.

We have now found that IS903 transposase expressed as a C-terminal fusion to either glutathione-S-transferase (GST; Smith and Johnson, 1988) or to maltose-binding protein (MBP; Bedouelle and Duplay, 1988; Guan *et al.*, 1988) is both soluble and active. In each case ~50% of the fusion protein remains soluble after overexpression and it can then be purified rapidly using affinity chromatography. Although the overexpression and purification protocols with each system are very similar, the yields are much greater with the MBP fusion. The GST fusion protein appears to bind less well to its affinity resin than does the native GST protein suggesting that the transposase interferes with the fusion binding. We have not seen this problem with the MBP-transposase fusion protein. This may be because the MBP-transposase gene fusion is designed to encode a polyasparagine linker between the two protein domains, which may reduce interference with binding of the MBP moiety to its affinity resin (New England Biolabs).

Each of these fusion proteins can mediate transposition *in vivo* (although we cannot exclude the possibility that proteolytic processing of the transposase fusion protein is required for this activity). We have detected transposition events mediated by both fusion proteins, by using both a Lac⁺ papillation assay (unpublished data) and a mating out assay with the transposon supplied *in trans* on a second compatible plasmid (Derbyshire *et al.*, 1990). The papillation assay showed that formation of Lac⁺ papillae required the transposase fusion protein (data not shown). The mating out assay was used in a more quantitative fashion to measure levels of transposition under different induction conditions. In the absence of protein induction we could detect low levels of transposition, presumably due to basal level expression of the fusion. When protein expression was induced transposition frequencies increased 15- to 200-fold to between 5×10^{-5} and 1.4×10^{-4} per donor. Too high an induction of transposase expression in these experiments resulted in loss of cell viability. For this reason, transposase expression from the inducible Ptac promoter was kept to a low level by inducing with 10 μ M isopropyl- β -D-thiogalactoside (IPTG). Under these conditions fusion protein could not be detected in cell extracts by Coomassie staining of SDS-polyacrylamide gels.

We have also succeeded in solubilizing a small amount of the native transposase from an insoluble protein fraction. Native transposase was overexpressed using the T7 expression system (Studier *et al.*, 1990). After cell lysis the insoluble material, which included the transposase, was collected by centrifugation. This insoluble protein fraction was extracted with 50 mM CAPS pH 9.6, 1 mM dithiothreitol (DTT), 1 mM EDTA in the presence of S Sepharose FF beads. Approximately 5% of the transposase was solubilized into this solution and could be used for footprinting and band-shift analysis. Unfortunately, this activity was lost rapidly and over a few days at 4°C the protein precipitated. However, we have used this transposase-enriched fraction, to show that the native and fusion proteins have indistinguishable DNase I protection patterns.

DNase I protection analysis

DNase I footprinting demonstrated that each of the purified transposases binds specifically to the IS903 inverted repeat. Figure 1 shows the footprints of the native transposase and GST-transposase proteins. The transposase protects the

inverted repeat from positions 5 through 19 on the upper strand (with numbering starting at the end of the transposon). On the lower strand transposase protects the DNA from DNase I cleavage between bases 3 and 15 (data not shown but summarized in Figure 3). We observed two enhancements of DNase I cleavage between the bases immediately outside the inverted repeat on the lower strand.

Careful comparison of the DNase I protection patterns of each of the transposases showed that they were indistinguishable. This suggested that each of the fusion proteins bound to the inverted repeat in a similar way to the native protein. Consequently, we have carried out all of the following experiments with the more stable GST- or MBP-transposase fusion proteins.

Methylation interference studies with the GST-transposase

The interaction of the GST-transposase with its inverted repeat was examined by identifying those bases which interfered with binding when methylated. Dimethylsulfate alkylation occurs in the major groove at N-7 of guanine and in the minor groove at N-3 of adenine. Modification of either of these residues may interfere with binding by inhibiting a base contact or by changing the local DNA structure (Siebenlist and Gilbert, 1980). The GST-transposase fusion could bind inverted repeat-containing DNA while still attached to the glutathione affinity column. We took advantage of this to purify modified DNA bound by the fusion protein away from unbound DNA. Figure 2 shows the results of methylation interference experiments on both strands of the inverted repeat. Methylation of bases G7, G10, A14, A15 and A16 on the upper strand strongly interfered with binding. There was a weak interference with binding when A11 was methylated. On the lower strand, methylation of A13 interfered with binding whilst methylation of A8 and A9 had a weak effect on binding. These results along with the DNase I footprinting data are summarized in Figure 3. These data strongly suggested that the base-specific binding contacts made by the transposase were confined to the inner portion of the inverted repeat at bases 7 to 16. Quite

Fig. 1. DNase I footprinting of transposase-DNA complexes. The footprints were obtained using either the GST-transposase fusion or the native transposase. The amount of transposase (TNPase) added to each reaction is indicated. The DNA was 5'-labeled at the *Xho*I site of pKD1. The sequence lane (M) is an A+G Maxam-Gilbert reaction on the same DNA and the relevant sequence is marked.

strikingly, methylation of the G residues in the outer three G–C bp had no effect on DNA binding, consistent with our previous hypothesis that these three bases interact with the transposase in a step subsequent to DNA binding.

Binding of MBP–transposase to mutant inverted repeats

Complexes between the MBP–transposase protein and DNA fragments containing an IS end were readily detected using

Fig. 2. DNA binding of the GST–transposase analyzed by methylation interference. Each panel shows the data from one strand of the inverted repeat. The relevant sequence is shown at the side of each panel alongside the A+G sequence lane. Methylated bases which interfere with binding are indicated.

Fig. 3. A summary of the methylation interference experiments. **Top:** the IS903 inverted repeat is shown in bold. The adjacent sequences are shown in lower case. The methylated bases that interfere strongly (large star) or weakly (small star) with binding are shown on the linear DNA sequence. Also included, above and below the relevant strands, is the extent of DNase I protection. **Bottom:** in the planar projection of the DNA helix the top strand sequence of the inverted repeat is shown. Methylated bases which interfere with binding are shown as diamonds. Filled diamonds interfere most strongly. The two vertical arrows show the site of DNase I enhancement.

native PAGE (Figure 4A, lanes 2 and 3). Complex formation required the IS903 inverted repeat, since a control fragment (IR⁻) with the same DNA sequence as the test fragment, but lacking the inverted repeat and flanking restriction sites, was not retarded in the gel (Figure 4A, cf. lanes 1 and 3).

The hypothesis that the IS903 ends consist of two functionally distinct domains was derived from the *in vivo* study of IS903 derivatives with mutations in their terminal inverted repeats. The availability of active transposase protein and a simple assay for specific binding enabled us to test this hypothesis *in vitro*. Fragments containing mutant IS903 ends (test IR) were assayed with two concentrations of MBP–transposase. As an internal control a larger fragment containing a wild-type inverted repeat (IR⁺) was also included in each reaction. A specific shift with this control fragment served to rule out the presence of inhibitors in the binding reaction and provided a standard for measuring the relative affinity for the mutant site.

We examined transposase binding to ends carrying a selection of mutations throughout the inverted repeat; some of these results are shown in Figure 4. Mutations in the inner part of the inverted repeat either reduced (A6C; a change from the wild-type A to a C, at position 6 on the lower strand in Figure 3) or abolished (A8G, C10T, T14C and A17C) transposase binding. This is entirely consistent with our hypothesis that sequences important for specific binding of transposase are confined to the inner domain of the inverted repeat. Although the mutation T14A is within the inner domain, it had little influence on transposase binding. This was predicted from our genetic data, since this mutation had no effect on transposition. The binding of transposase to a T14A end but not to a T14C end, suggests that the protein can make equivalent interactions with either a T–A or an A–T base pair at this position, as would be expected for a minor groove but not a major groove interaction.

A second class of mutations (C1T, C2T, G3A and A5C), which are located in the outer part of the inverted repeat, had no effect on transposase binding (Figure 4A and data not shown). Since our previous analysis had shown that all of these mutations caused a substantial reduction in transposition (e.g. the mutation G3A reduces transposition to 0.05% of wild-type levels), we reasoned that these positions must be important for transposition. Our original genetic data led us to propose that at least positions 1–3 were involved in a step subsequent to transposase binding (Derbyshire *et al.*, 1987). The fact that transposase also binds normally to an A5C end suggests that the outer domain actually extends to base pair 5 of the inverted repeat. Mutations at position 4 of the inverted repeat had no effect on transposition *in vivo* and so the contribution of this base pair remains unclear.

MBP–transposase bends DNA

A characteristic of many DNA binding proteins is that they bend their recognition sites and in many cases this has been shown to be critical for the function of that protein. The MuA protein of the Mu transposon and the TnsB protein of Tn7 have both been shown to bend their cognate site (Kuo *et al.*, 1991; Arciszewska and Craig, 1991). We tested whether the MBP–transposase could bend DNA upon binding. A DNA fragment containing the IS903 inverted repeat was cloned into the pBend2 vector (Kim *et al.*, 1989), which is designed to allow a set of circularly permuted fragments of identical

Fig. 4. Binding of MBP-transposase to different mutant inverted repeats analyzed by native gel electrophoresis. The control lanes (A, first panel; B, last panel) show the retardation with individual fragments. Note there is no shift with a fragment lacking an inverted repeat (IR^-). The positive control fragment (IR^+) is included in all experimental lanes. The tester fragments (Test IR) used throughout are identical except for the mutation shown above each panel. The amount of MBP-transposase (TNPASE) added to each reaction is shown above each lane.

size to be generated with the inverted repeat in different locations with respect to the fragment ends. Four DNA fragments, covering a range of inverted repeat locations, were used in a band-shift assay (Figure 5). As the distance from the inverted repeat to the closest DNA end was increased, the fragment was retarded more by the fusion protein; this is consistent with a DNA bend induced by the transposase (there appeared to be no natural bend in this DNA as shown by the similar mobilities of the free DNA fragments). This is not a dramatic bend as seen with some proteins (e.g. IHF, Thompson and Landy, 1988; CAP, Liu-Johnson *et al.*, 1986) and appeared qualitatively similar to that reported for TnsB of Tn7 (Arciszewska and Craig, 1991). The transposase-induced bend might explain the sites of enhanced DNase I cleavage seen just outside the inverted repeat (Figure 3). Such enhancements are often suggestive of a distortion in the DNA that opens up the minor groove and makes it more accessible to DNase I.

Discussion

We have exploited two expression systems that encode fusion proteins to purify the IS903 transposase. Each transposase fusion protein was expressed efficiently and was substantially soluble. Furthermore, by taking advantage of the fusion proteins' properties we can purify them rapidly by affinity chromatography. A priori there are no obvious reasons why these two particular systems should have fared better than the many other methods we have tried. One possible explanation is that in each case the transposase is the C-terminal component of the fusion; perhaps having an efficiently synthesized, soluble, N-terminal domain prevents aggregation of the transposase domains. We have used these fusion proteins to analyze the protein-inverted repeat interactions involved in DNA binding.

Transposase binds to the inner part of its inverted repeat

By combining the data from the methylation interference analysis, the footprinting experiments and the binding assays on mutant sites with our previous *in vivo* analysis of mutant

Fig. 5. Detection of DNA bending induced by the MBP-transposase fusion protein. Complexes between MBP-transposase (3 μ l) and 263 bp circularly permuted DNA fragments derived from pKD404 were analyzed by electrophoresis on a 5% native polyacrylamide gel. The distance between the inverted repeat and the nearest DNA end is as indicated above each lane. Note that there is no change in mobility of the unbound DNA as the fragment is permuted.

ends, we can build up a picture of the important contacts made by the transposase protein to its inverted repeat. Our *in vivo* analysis (Derbyshire *et al.*, 1987) had led us to propose that base pairs 6–18 were involved in DNA binding and this is consistent with the experiments described here. Firstly, methylation of purines in the inner part of the inverted repeat strongly inhibited binding (G7 to A16), implying that the transposase makes specific base contacts in this region. In contrast, binding of transposase was insensitive to methylation of the outer six base pairs suggesting that the transposase does not make significant contacts in the major groove for the outer three G–C base pairs or in the minor groove for residues 4–6. Secondly, the mutations that we examined which mapped within this inner domain either reduced (A6C), or abolished binding

(A8G to A17C) in our band-shift experiments. In contrast, transposase bound to ends mutated in the outer 5 bp with normal affinity, consistent with there being no base-specific protein–DNA contacts at these positions that are required for binding. Finally, although transposase protected the inner portion of the inverted repeat against DNase I cleavage, it failed to protect the outer end, suggesting that this outer portion is not in close contact with the transposase in the complex with a linear DNA fragment.

What are the boundaries of the binding domain?

Methylation of G7 interferes with transposase binding and so defines the outermost part of the inverted repeat to which we know transposase makes a direct binding contact. In the binding assay the transposase bound to the A5C end with wild-type affinity suggesting that A5 does not make a direct binding contact with the protein. The role of the base pair at position 6 is uncertain. The mutation A6C did reduce transposase binding but not as dramatically as other mutations in the binding domain. Furthermore, mutations at position 6 had a much more severe effect on transposition *in vivo* (Derbyshire *et al.*, 1987) than the *in vitro* binding data would indicate. Perhaps mutations at this position alter transposase binding in a way that prevents the transposase making further productive interactions with the transposon ends and this causes the more dramatic *in vivo* effect.

The innermost transposase-binding contact we can detect from the methylation interference data is to A16 in the minor groove. In addition, we saw that the mutation A17C dramatically reduced transposase binding in our band-shift assay. This is consistent with our *in vivo* assays which showed that mutations at position 17 reduced transposition up to 25-fold. Since mutations at position 18 had either a small (10-fold) or no effect on transposition *in vivo*, we propose that position 17 is the innermost part of the inverted repeat to which transposase makes significant contacts.

IS903 transposase makes major and minor groove contacts

The methylation interference studies suggest that transposase makes both major and minor groove contacts with the inverted repeat (Figure 3). Furthermore, these contacts (G7, G10, A13, A14, A15, A16) all align along one face of the DNA helix implying that the transposase makes its most important contacts on one side of the DNA. We suggest that the weaker interferences resulting from methylation of A8, A9 and A11, which are on the opposite side of the DNA helix to the proposed binding site, represent indirect effects due to the residual positive charge on the methylated base or structural perturbations transmitted to the major groove or DNA backbone.

Strong inhibition of binding by methylation of G7 and G10 suggests that transposase makes contacts with both these bases via the major groove. If we examine the hierarchy of effects on transposition of mutations at T11, we note that it is also compatible with a major groove contact. Our *in vivo* analysis showed that T11G had only a minor effect on transposition while changes to A and C were much more severe (Derbyshire *et al.*, 1987). This implies that the transposase can make almost equivalent contacts to either a T–A or to a G–C base pair at this position. As Seeman *et al.* (1976) have pointed out, in the major groove the hydrogen bond-accepting carbonyl oxygens of T and G are

too similarly placed to allow discrimination between T and G, as are the hydrogen bond-donating amino groups of A and C. Thus if transposase were to form hydrogen bonds to one or both of these positions, but ignore the N-7 hydrogen bond acceptor of A (which provides unambiguous discrimination between A and C), it would distinguish poorly between T–A and G–C but would fail to recognize either A–T or C–G as we observed.

Methylation of adenines in the minor groove at positions 13–16 strongly interferes with transposase binding suggesting that the transposase recognizes this segment of the inverted repeat via the minor groove. This is supported by comparison of the effects of the two mutations at T14, which showed a striking difference in transposase binding; T14C abolished binding, whilst transposase bound a T14A end with a similar affinity to a wild-type end (Figure 4B). Even a single hydrogen bond interaction in the major groove should allow a protein to distinguish an A–T base pair from T–A (Seeman *et al.*, 1976). However, in the minor groove, the symmetrical location of the hydrogen bond acceptors of A and T will not allow discrimination between A–T and T–A. In contrast, a G–C base pair introduces an amino group into the center of the minor groove, which could interfere with DNA contacts. These results are also consistent with our *in vivo* experiments in which we had noted that the most acceptable substitutions (in terms of transposition) at positions 13–16 were A–T/T–A base pair reversals (Derbyshire *et al.*, 1987).

Interestingly, while many proteins recognize their binding sites via the major groove, there is a growing number of proteins that also make important minor groove contacts (for a recent review see Harrison, 1991). These include the homeodomains of eukaryotic transcription factors (Kissinger *et al.*, 1990), the resolvase and DNA invertase family of site-specific recombinases (Falvey and Grindley, 1987; Rimphanitchayakit *et al.*, 1989; Sluka *et al.*, 1987; Glasgow *et al.*, 1989), integration host factor and related proteins (Yang and Nash, 1989) and the eukaryotic TATA-binding factor, TFIID (Starr and Hawley, 1991; Lee *et al.*, 1991).

A common organization for transposon inverted repeats?

Our data support the hypothesis that the inner part of the IS903 repeat contains all the essential contacts for transposase binding and that the outer 5 bp, although important for transposition, are required not for transposase binding but for a subsequent step in the transposition process. A similar organization of two functional domains within an inverted repeat is now emerging as a common theme for other transposons. Genetic analysis of mutations in the end of IS10 supported the proposal that base pairs 6–13 of the inverted repeat were involved in sequence-specific contacts with the transposase and that the outer 3 bp constituted a functionally distinct domain (Huisman *et al.*, 1989). It has since been shown that changes in the outer domain cause accumulation of the excised transposon fragment, both *in vivo* and *in vitro*, implying that they are blocked in the integration step of transposition (D.Haniford and N.Kleckner, personal communication). The binding of the N-terminal domain (InsA) of the IS1 transposase to mutant inverted repeats has been examined *in vitro*. InsA binds normally to ends that are mutated in the outer 3 bp but its binding is sensitive to mutations in the inner portion of the inverted repeat,

suggesting that the InsA domain recognizes the inner part of the IS1 inverted repeat when binding (Zerbib *et al.*, 1990). The $\gamma\delta$ and Tn3 transposons also have two separate domains within their inverted repeats. Mutations and deletions of the outer part of the inverted repeat abolish transposition but still permit transposase binding and these mutants exhibit transposon immunity (Kans and Casadaban, 1989; Amemura *et al.*, 1990; Wiater and Grindley, 1990; E.May and N.D.F.Grindley, unpublished results).

The transposon Mu has three transposase (MuA) binding sites at each end and these assemble to form a nucleoprotein complex (the transpososome) whose structure is critical for transposition. Clearly, this is a much more complex situation than for those transposons described above, but if we examine the MuA–DNA interaction just at the outermost sites (which probably carry out similar functions to the inverted repeats) we note that there is again a striking similarity. DNA protection studies on linear DNA show that the MuA protein does not protect the outer 4 or 5 bp of the transposon (Craigie *et al.*, 1984), yet these outer base pairs become protected in the stable, nucleoprotein complex intermediates in Mu transposition suggesting that contact with this outer end of the transposon is important in later steps of transposition (Lavoie *et al.*, 1991; Mizuuchi *et al.*, 1991; Surette *et al.*, 1991; Kuo *et al.*, 1991).

Why are the inverted repeats of so many different transposons designed in this way? A plausible explanation is that the absence of close contact between the transposase catalytic site and its DNA cleavage site at the end of the transposon (the outer domain) may prevent premature cleavage at single ends of a transposon. The implication from our data and the experimental data with Mu, is that the transposase only comes in contact with the cleavage site once the two ends are in a complex. Formation of this nucleoprotein complex may induce a conformational change in the protein which realigns the transposase catalytic site with its DNA cleavage site—the ends of the transposon. Alternatively, each cleavage site may be recognized *in trans* (i.e. the transposase protomer bound to one end may recognize and cleave the transposon at the other end) so that it would be impossible to cleave at one end of a transposon without the other being present. Thus the separation of domains may ensure the efficient and correct formation of the transpososome before transposon cleavage occurs.

Materials and methods

Materials

Restriction enzymes and DNA modifying enzymes were purchased from New England Biolabs and Boehringer Mannheim and were used according to the manufacturer. Klenow fragment was a gift from C.Joyce and V.Derbyshire. The glutathione affinity resin was obtained from Sigma and the maltose-binding protein purification system was obtained from New England Biolabs.

Plasmids

The glutathione gene fusion was made in pGEX2T (Smith and Johnson, 1988). The transposase gene was cloned, in frame, between the *Bam*HI and *Eco*RI sites of the vector to generate pKD383. This gene fusion introduces an additional three amino acids between the thrombin cleavage site and the transposase N terminus. The maltose gene fusion was made in pMALc2 (New England Biolabs). A 2 kb *Bam*HI–*Pst*I fragment from pKD383 was cloned in frame into pMALc2 to generate pKD395. This gene fusion has an additional seven amino acids between the factor Xa cleavage site and the transposase N terminus. pKD371 is a derivative of pET3a (Studier *et al.*, 1990) containing the transposase gene under the control of a T7 promoter. The transposase gene was cloned from pKD162 (Derbyshire

et al., 1990) between the *Nde*I and *Bam*HI sites of pET3a. This plasmid was always induced in the presence of pLysS (Studier *et al.*, 1990). All transposase gene constructs have had a strong transcription terminator within the gene removed by oligo-directed mutagenesis, to improve expression. pKD404 was created by cloning a 143 bp *Xba*I–*Xho*I fragment from pKD100 (Derbyshire *et al.*, 1987) between the *Xba*I and *Sal*I sites of pBend2 (Kim *et al.*, 1989). This contains the inverted repeat and the first 120 bp of the kanamycin resistance gene.

DNA fragments

The fragment used for footprinting, methylation interference studies and the internal positive control in band-shift experiments was a 254 bp *Bam*HI–*Xho*I fragment from pKD1 (Weinert *et al.*, 1984). This fragment contains 112 bp from the inside end of IS903 and the first 134 bp of the kanamycin resistance gene from Tn903 (Grindley and Joyce, 1981). The negative control for band shifts was a 120 bp *Bam*HI–*Xho*I fragment from pKD1 containing the first part of the kanamycin resistance gene. The tester fragments containing mutant ends were generated from M13 derivatives that contained the 1.1 kb *Xba*I fragment from pKD100 clones with different mutant inverted repeats (Derbyshire *et al.*, 1987). These 167 bp *Hind*III–*Xho*I fragments contained the inverted repeat plus flanking restriction sites and the first 120 bp of the kanamycin gene. DNA fragments (263 bp) for determining the presence of a transposase-induced bend were generated by cleaving pKD404 with *Mlu*I, *Xho*I, *Asp*718 or *Bam*HI. DNA fragments were labeled either at their 5' end using T4 polynucleotide kinase and [γ -³²P]ATP or at their 3' end using DNA polymerase I (Klenow fragment) and [α -³²P]dATP. Fragments from the M13 derivatives were labeled by primer extension in the presence of [α -³²P]dATP before cutting with the appropriate enzymes.

Transposase overexpression and purification

The GST–transposase fusion protein was purified using a glutathione affinity column as described by Smith and Johnson (1988). The MBP–fusion protein was purified as described in the New England Biolabs protein fusion kit using an amylose affinity column. Induction of the proteins was carried out at room temperature (20°C) for 2 h. All buffers contained phenylmethylsulfonyl fluoride (PMSF) which was added fresh to a final concentration of 1 mM. Purified protein concentrations were ~20 μ g/ml for the GST–transposase fusion and 200 μ g/ml for the MBP–transposase fusion.

The native transposase was purified based on a procedure described by Hoess *et al.* (1988) and modified as follows; *Escherichia coli* BL21(DE3) (Studier *et al.*, 1990) containing pKD371 and pLysS was induced for expression by addition of IPTG to 0.4 mM and grown for a further 3 h. Cells (20 ml) were harvested and resuspended in 1/20 volume of 50 mM Tris–HCl pH 8, 1 mM EDTA, 1 mM DTT, 1 mM PMSF and 1 mg/ml of lysozyme and incubated on ice for 15 min. The cells were lysed by sonication and the insoluble material which included the transposase was collected by centrifugation. The insoluble pellet was then resuspended in 100 μ l of 50 mM CAPS–NaOH pH 9.6, 1 mM DTT, 1 mM EDTA. Two-hundred microliters of S sepharose FF beads (Pharmacia) in the same buffer were added and the slurry was gently mixed for 2 h at 4°C. Solubilized protein was then recovered in the supernatant after spinning down the beads and the remaining insoluble material in a microfuge for 5 min. This supernatant fraction was enriched for transposase and was used directly in footprinting and band-shift assays. We have since found that washing the pellet in 50 mM CAPS–NaOH pH 9.6 without the ion exchange resin, solubilizes some transposase.

DNase I footprinting

Transposase binding reactions were set up as follows: transposase was added to labeled DNA fragments in 10 μ l reactions containing 50 mM Tris–HCl pH 7.5, 10 mM MgCl₂, 1 mM DTT (NaCl was adjusted to 100 mM as appropriate). After binding at 37°C for 5 min and equilibration to room temperature for 3 min, DNase I was added to a final concentration of 3.3 μ g/ml and incubated for 1 min before adding 2 μ l of stop solution (100 mM EDTA, 1% SDS). The samples were then ethanol precipitated and resuspended in formamide dyes and the products separated on sequencing gels.

Methylation interference

The DNA was modified with dimethyl sulfate as described by Sienbenlist and Gilbert (1980). The GST–transposase fusion protein obtained from a 100 ml induction was bound to 100 μ l of glutathione beads as above. After extensive washing to remove unbound protein, the beads were washed three times in 1.5 ml of binding buffer (50 mM Tris–HCl pH 7.5, 10 mM MgCl₂, 100 mM NaCl, 1 mM DTT). Modified DNA, in 100 μ l of the same buffer, was then added to the beads, gently mixed and allowed to bind for 5 min at 37°C. The beads were spun down for 5 s in a bench microfuge

and the supernatant containing unbound DNA removed and kept. The beads were then washed with $4 \times 400 \mu\text{l}$ of binding buffer to remove free DNA. Specifically-bound DNA was removed from the protein by phenol extraction in the presence of 0.1% SDS. The bound and free fractions were ethanol precipitated, cleaved at modified positions by incubation at 90°C for 30 min in 10 mM sodium phosphate pH 8.0, 150 mM NaOH, and the products separated on a 6% sequencing gel.

Band-shift analysis

Transposase-DNA complexes were formed as described for DNase I footprinting except the reaction also contained 100 $\mu\text{g/ml}$ BSA. After the room temperature incubation 4 μl of loading buffer (5% Ficoll in 50 mM Tris-HCl pH 7.5, 10 mM MgCl_2 100 mM NaCl, 0.1% xylene cyanol FF) was added and the entire reaction was loaded on a 1.5 mm 5% polyacrylamide gel (polyacrylamide-bisacrylamide, 30:1, 100 mM Tris-borate pH 8.3, 1 mM EDTA). The gel was pre-run for 10 min at 120 V and left running whilst the samples were loaded. On completion of electrophoresis the gels were dried down and autoradiographed.

Acknowledgements

We would like to thank William Studier for the T7 expression system, David Toczyski for the pGEX plasmids, Marna Williams for the pBend plasmid, Max Fraser for help with the baculovirus system and Roger McMacken for suggestions regarding solubilization of insoluble transposase. We are grateful to Vicky Derbyshire, Cathy Joyce and Christian Parker for their comments on this manuscript. K.M.D. would like to acknowledge Vicky Derbyshire for encouragement and emotional support in his endless quest for active transposase. This work was supported by grant GM28470 from the NIH.

References

- Amemura, J., Ichickawa, H. and Ohtsubo, E. (1990) *Gene*, **88**, 21–24.
- Arciszewska, L. and Craig, N.L. (1991) *Nucleic Acids Res.*, **19**, 5021–5029.
- Bedouelle, H. and Duplay, P. (1988) *Eur. J. Biochem.*, **171**, 541–549.
- Craigie, R., Mizuuchi, M. and Mizuuchi, K. (1984) *Cell*, **39**, 387–394.
- Derbyshire, K.M., Hwang, L. and Grindley, N.D.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8049–8053.
- Derbyshire, K.M., Kramer, M. and Grindley, N.D.F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4048–4052.
- Falvey, E. and Grindley, N.D.F. (1987) *EMBO J.*, **6**, 815–821.
- Frankel, S., Sohn, R. and Leinwand, L. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1192–1196.
- Galas, D.J. and Chandler, M. (1989) *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 109–62.
- Germine, J. and Bastia, D. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4692–4696.
- Glasgow, A.C., Bruist, M.F. and Simon, M.I. (1989) *J. Biol. Chem.*, **264**, 10072–10082.
- Grindley, N.D.F. and Joyce, C.M. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 125–133.
- Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1988) *Gene*, **67**, 21–30.
- Harrison, S. (1991) *Nature*, **353**, 715–719.
- Hoess, A., Arthur, A.K., Wanner, G. and Fanning, E. (1988) *BioTechnology*, **6**, 1214–1217.
- Huisman, O., Errada, P.R., Signon, L. and Kleckner, N. (1989) *EMBO J.*, **8**, 2101–2109.
- Kans, J.A. and Casadaban, M.J. (1989) *J. Bacteriol.*, **171**, 1904–1914.
- Kim, J., Zwieb, C., Wu, C. and Adhya, S. (1989) *Gene*, **85**, 15–23.
- Kissinger, C.R., Liu, B., Martin-Bianco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell*, **63**, 579–590.
- Kuo, C.-F., Zou, A., Jayaram, M., Getzoff, E. and Hershey, R. (1991) *EMBO J.*, **10**, 1585–1591.
- Lavoie, B.D., Chan, B.S., Allison, R.G. and Chaconas, G. (1991) *EMBO J.*, **10**, 3051–3059.
- Lee, D.K., Horikoshi, M. and Roeder, R.G. (1991) *Cell*, **67**, 1241–1250.
- Liu-Johnson, H.-N., Gartenberg, M.R. and Crothers, D.M. (1988) *Cell*, **47**, 995–1008.
- Mizuuchi, M., Baker, T.A. and Mizuuchi, K. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9031–9035.
- Nagai, K. and Thørgensen, H.C. (1984) *Nature*, **309**, 810–812.
- Rimphanitchayakit, V., Hatfull, G.F. and Grindley, N.D.F. (1989) *Nucleic Acids Res.*, **17**, 1035–1050.
- Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804–808.
- Sienbenlist, U. and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 122–126.
- Sluka, J.P., Horvath, S.J., Bruist, M.F., Simon, M.I. and Dervan, P.B. (1987) *Science*, **238**, 1129–1132.
- Smith, D.B. and Johnson, K. (1988) *Gene*, **67**, 31–40.
- Smith, G.E., Summers, M.D. and Fraser, M. (1983) *Mol. Cell Biol.*, **3**, 2156–2165.
- Starr, D.B. and Hawley, D.K. (1991) *Cell*, **67**, 1231–1240.
- Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
- Surette, M.G., Harkness, T. and Chaconas, G. (1991) *J. Biol. Chem.*, **266**, 3118–3124.
- Thompson, J.F. and Landy, A. (1988) *Nucleic Acids Res.*, **16**, 9687–9705.
- Wiater, L.A. and Grindley, N.D.F. (1990) *J. Bacteriol.*, **172**, 4959–4963.
- Weinert, T.A., Derbyshire, K.H., Hughson, F.M. and Grindley, N.D.F. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 251–260.
- Yang, C.C. and Nash, H.A. (1989) *Cell*, **57**, 869–880.
- Young, D.F., Desselberger, U., Palese, P., Ferguson, B., Shatzman, A.R. and Rosenberg, M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6105–6109.
- Zerbib, D., Prentki, P., Gamas, P., Freund, E., Galas, D.J. and Chandler, M. (1990) *Mol. Microbiol.*, **4**, 1477–1486.

Received on May 7, 1992