The EMBO Journal vol.13 no.24 pp.6125–6132, 1994

Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain

Jan C.Vos and Ronald H.A.Plasterk[1]

The Netherlands Cancer Institute, Division of Molecular Biology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[1]Corresponding author

Communicated by P.Borst

The Tc1 transposon of *Caenorhabditis elegans* is a member of the Tc1/mariner family of mobile elements. These elements have inverted terminal repeats that flank a single transposase gene. Here we show that Tc1 transposase, Tc1A, has a bipartite DNA binding domain related to the *paired* domain of mammalian and *Drosophila* genes. Both the DNA binding domain of Tc1A and the DNA binding site in the inverted repeat of Tc1 can be divided into two subdomains. Methylation interference studies demonstrate adjacent minor and major groove contacts at the inner part of the binding site by the N-terminal 68 amino acids of the DNA binding domain. In addition, Tc1A amino acids 69–142 are essential for major groove contacts at the outer part of the binding site. Recombinant Tc1A is found to be able to introduce a single strand nick at the 5′ end of the transposon *in vitro*. Furthermore, Tc1A can mediate a phosphoryl transfer reaction. A mutation in a DDE motif abolishes both endonucleolytic and phosphoryl transfer activities, suggesting that Tc1A carries a catalytic core common to retroviral integrases and IS transposases.

Key words: *Caenorhabditis elegans*/*paired* domain/Tc1/transposition

Introduction

Tc1 from the nematode *Caenorhabditis elegans* is a structurally simple transposon which belongs to the Tc1/mariner family of mobile elements found in a wide range of eukaryotes (Henikoff, 1992; Robertson, 1993; Radice *et al.*, 1994). The 1.6 kb long Tc1 element contains a transposase gene, *Tc1A*, flanked by 54 bp perfect inverted terminal repeats (Rosenzweig *et al.*, 1983a; Moerman and Waterston, 1989, for a review). Excision by double strand breaks at the transposon ends (Plasterk, 1991) results in a linear copy of the element (Ruan and Emmons, 1984), which can integrate elsewhere in the genome at a TA dinucleotide (Rosenzweig *et al.*, 1983b, Van Luenen and Plasterk, 1994). As can be inferred from work on the related *C.elegans* transposon Tc3, integration is accompanied by a TA target site duplication. Furthermore, excision leaves behind two nucleotides from the 5′ end of the transposon at the donor site, which after repair of the double strand gap results in characteristic footprints, previously thought to be the result of imprecise excision (Van Luenen *et al.*, 1994).

One can envisage several distinct structural and chemical steps during transposition of Tc1. First, Tc1A may bind to each end of the element and, perhaps helped by host factors, will bring the two ends together to form a synaptic complex. Second, two chemical steps have to be executed, which again may or may not require host factors. These steps are endonucleolytic cleavage at the junction between transposon and flanking DNA resulting in the excised intermediate, followed by a strand transfer reaction which couples the intermediate to a TA target sequence in the worm genome.

It has been shown that Tc1A can bind to the inverted repeat of Tc1 (Vos *et al.*, 1993). Similarly, the Tc3 transposase binds specifically to the inverted repeat of Tc3 (Van Luenen *et al.*, 1993). The minimal domain of Tc1A for binding to the inverted repeat is contained within the N-terminal 63 amino acids (Vos *et al.*, 1993). Within this domain a stretch of 39 amino acids was found to be 31% identical and 28% similar to an N-terminal domain of the IS30 transposase (Vos *et al.*, 1993). Furthermore, it has been reported that the Tc1-like Minos mobile element of *Drosophila hydei* contains an N-terminal domain with a weak similarity (17% identity and 49% similarity) to the *paired* DNA binding domain. The *paired* domain has been recognized as a conserved sequence motif in mammalian and *Drosophila* developmental genes (Bopp *et al.*, 1986; Walther *et al.*, 1991).

It has been suggested, based on sequence alignments, that the catalytic activity for cleavage and strand transfer resides within the Tc1 transposase. A possible common DDE motif has been recognized within the Tc1/mariner family (Doak *et al.*, 1994) which is shared with prokaryotic IS3 transposase and eukaryotic retroviral integrases (Fayet *et al.*, 1990; Kahn *et al.*, 1991). Characteristics of this motif are an aspartic acid and a glutamic acid separated by 34 or 35 residues, as well as a more N-terminal aspartic acid, which together are thought to co-ordinate a metal ion. Mutation of any of these residues in retroviral integrases abolishes catalytic activity (Drelich *et al.*, 1992; Engelman and Craigie, 1992; Kulkosky *et al.*, 1992; Van Gent *et al.*, 1992). The importance of these residues for the Tc1/mariner family has recently been demonstrated for Tc3 transposase in an *in vivo* assay (Van Luenen *et al.*, 1994).

In our attempts to establish an *in vitro* transposition system for Tc1, we have started the biochemical characterization of Tc1A purified from *Escherichia coli*. We have analyzed its binding to the inverted repeat and demonstrate that the protein has a bipartite DNA binding domain. Furthermore, we show that Tc1A has specific endonucleolytic cleavage as well as non-specific phosphoryl transfer activity. Mutation of the DDE motif

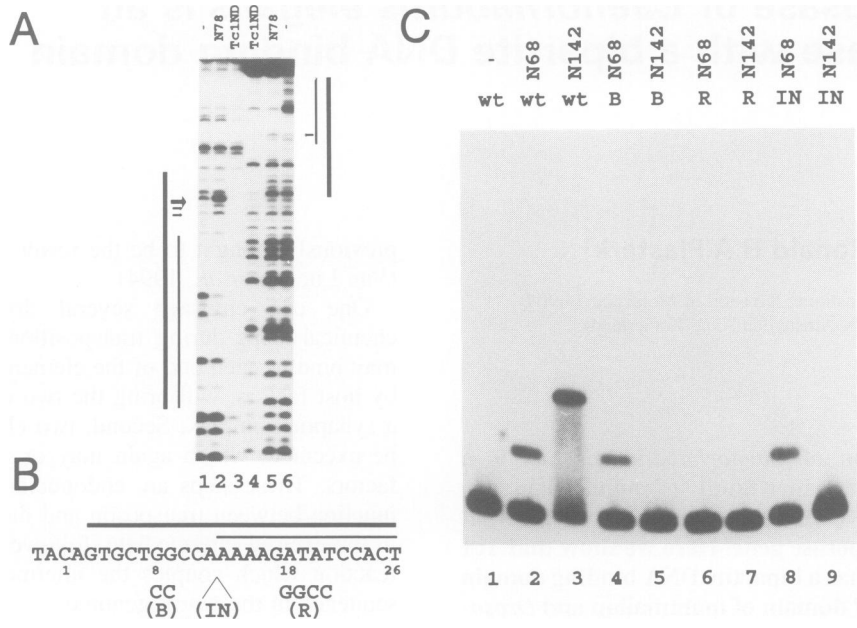

Fig. 1. Differential DNA binding by N-terminal derivatives of Tc1A. (**A**) DNase I protection assay on the 26 terminal base pairs of Tc1. A 3′ end-labeled DNA fragment containing the 26 terminal base pairs of Tc1 was footprinted in the presence of 1 μl of Tc1ND (lanes 3 and 4) or 1 μl of N78 (lanes 2 and 5) or in the absence of protein (lanes 1 and 6). Standard binding reactions contained 2 μg poly(dI–dC) competitor. DNase I products were separated on a sequencing gel. A probe labeled on the top strand (see B) was used in lanes 4–6, whereas a probe labeled on the bottom strand was used in lanes 1–3. The protected regions are indicated with thin (N78) and thick lines (Tc1ND) respectively. DNase I hypersensitive sites due to binding of N78 are marked by arrows. (**B**) Sequence of the 26 terminal base pairs of Tc1 (top strand) and mutations thereof. The protected areas are indicated as well as several sequence alterations. Mutation B is the substitution of GG by CC at positions 8 and 9. IN is an insertion of a T nucleotide between positions 12 and 13 and mutation R has the sequence GGCC at positions 18–21 instead of ATAT. (**C**) Gel retardation analysis of the ability of N-terminal derivatives to bind to wild-type or mutated Tc1 inverted repeat sequences. One microliter of polypeptides N68 or N142 was incubated with a 3′ end-labeled *Hin*dIII–*Sac*I fragment containing the terminal 26 bp of Tc1 or a mutation thereof in the presence of 2 μg poly(dI–dC) and 0.5 μg salmon sperm DNA using standard conditions. Complexes were resolved on a 4% polyacrylamide gel. The combination of polypeptide and probe is indicated on top of the lanes.

abolishes both activities *in vitro*, whereas DNA binding is not affected.

Results

Tc1 transposase has a bipartite DNA binding domain

Previously, we have shown that Tc1 transposase has an N-terminal DNA binding domain which specifically binds to the inverted repeat of Tc1 (Vos *et al.*, 1993). In defining the *cis* and *trans* requirements for binding, we noticed a difference in DNA binding properties of a polypeptide consisting of the first 78 amino acids of Tc1A (N78) and a longer N-terminal Tc1A derivative (Tc1ND), probably due to proteolysis, obtained from *E.coli* together with full-length Tc1A (Vos *et al.*, 1993; see also Figure 4). DNase I footprinting shows that the protected region obtained with N78 (Figure 1A, lanes 2 and 5) is considerably smaller than the one seen with Tc1ND (lanes 3 and 4). Also, N78 binding results in a strong DNase I hypersensitive site at position 8 in the bottom strand (lane 2). Minor hypersensitive sites are seen at positions 9 and 10 (bottom strand, lane 2) and at position 13 in the top strand (lane 5). We tested the *cis* requirements for binding of inverted repeat sequences of N-terminal derivatives of Tc1A in a gel retardation assay. Therefore, we expressed polypeptides corresponding to the first 68 (N68) or 142 (N142) amino acids of Tc1A, which by virtue of their C-terminal histidine tag (see Materials and methods) could be purified by chromatography on a Ni^{2+}–NTA resin (Hochuli *et al.*, 1988). Both N68 and N142 bind to the wild-type Tc1 sequence (Figure 1C, lanes 2 and 3 respectively). Mutation of base pairs 18–21 (see Figure 1B) abolishes binding of both N68 and N142 (lanes 6 and 7 respectively). In contrast, mutation of base pairs 8 and 9 does not affect N68 binding (lane 4), but is detrimental for binding of N142 (lane 5), although N142 includes N68. The same holds true for a base pair insertion between positions 12 and 13, which only impairs binding of N142, but not N68 (compare lanes 8 and 9). These results suggest that N142, in comparison with N68, contains an additional domain engaged in binding to the inverted repeat.

N-terminal deletions of Tc1A of 17, 31 or 70 amino acids did not show any affinity for inverted repeat sequences in a gel retardation assays (data not shown). Apparently, the more C-terminal domain is unable to bind DNA independently and therefore the N-terminus of this second domain cannot be mapped by deletion mutagenesis. Finally, the C-terminus of this domain is between amino acids 117 and 142 of Tc1A, because a polypeptide corresponding to the N-terminal 117 amino acids of Tc1A behaves like N68 (data not shown).

Methylation and missing base interference

To investigate which bases are involved in sequence-specific binding by Tc1A, we performed a methylation interference experiment. A probe representing the terminal inverted repeat sequence of Tc1 was treated with dimethyl

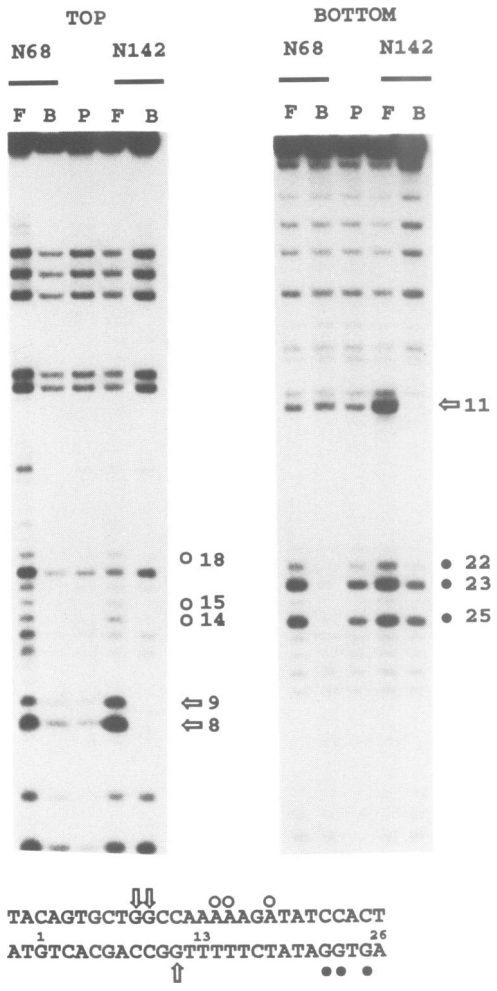

Fig. 2. DNA binding specificity of N68 and N142 analyzed by methylation interference. Each panel shows the comparison between the interference pattern obtained with either polypeptides N68 or N142 for one strand of the inverted repeat. P (probe) lanes contain the pattern of bands derived from the cleaved DMS-modified probe. B (bound) lanes show the products derived from isolated protein–DNA complexes, whereas F lanes represent the free probe. Methylated bases that interfere with binding for both polypeptides are indicated with circles. The strongest positions are indicated with closed circles. Methylated bases that interfere solely with binding by N142 are indicated with arrows. A summary of the interference data is given at the bottom of the figure.

sulfate (DMS), which modifies G and A residues in the major and minor groove respectively. The probe was incubated with either N68 or N142 and bound DNA was separated from free DNA on a native gel. DNA was purified from the gel, cleaved at the modified bases and run on a sequencing gel (Figure 2). On the bottom strand we detect interference at G22, G23 and G25 for both polypeptides, whereas a N142-specific interference is observed at G11. The top strand shows common interference at positions A14, A15 and A18. Furthermore, modification of G8 and G9 only interferes with binding of N142. In conclusion, it appears that the N-terminal 68 amino acids make major groove contacts with base pairs 22–25, as well as minor groove contacts with base pairs 14–18. N142 also interacts with the major groove at positions 8–11.

Next, we compared the binding specificity of the two polypeptides by missing base analysis. The inverted repeat probe was either modified with formic acid (purine-specific) or with hydrazine (pyrimidine-specific). Bound and free DNA were separated on a native gel, DNA was isolated, cleaved with piperidine and analyzed on a sequencing gel. Interferences were found for the N78 protein at positions 13–22 on the top strand and positions 14–25 on the bottom strand (Figure 3A). The strongest interferences are at positions 17–22 (top strand) and 17–24 (bottom strand). In contrast, N142 shows additional positions of interference (Figure 3B). The positions that are unique to N142 or are more enhanced relative to N78 are bases 7–13 (top strand) and 11–13 (bottom strand).

A summary of the binding results is given in Figure 3C. The N-terminal part of the bipartite domain of Tc1A binds to base pairs 12–26 of the inverted repeat, whereas additional specificity is obtained by the C-terminal part, which is reflected in recognition of base pairs 7–13.

Purification of the transposase

We have partially purified soluble Tc1A from *E.coli*, using a C-terminal histidine tag to separate full-length Tc1A from N-terminal derivatives which are produced as by-products in *E.coli* (Vos *et al.*, 1993). In parallel, a DDE mutant transposase was purified in which Asp247 was substituted by a Val through site-directed mutagenesis of the *Tc1A* gene. The corresponding mutation in Tc3A results in a transposase that is inactive *in vivo* (Van Luenen *et al.*, 1994), therefore we wished to test for possible effects of this mutation in Tc1A. The purified fractions were tested in a gel retardation assay using the Tc1 inverted repeat sequence as probe (Figure 4). As can be seen by comparison of the crude Tc1A extract (lane 3) with the control pET3c extract (lane 1), two Tc1A-specific complexes are detected. A minor one which has a low mobility and a major one which represents complexes containing N-terminal derivatives of Tc1A. These latter species migrate at a position comparable with the N142–DNA complex (lane 7). These complexes are specific, because they are not generated with a mutated Tc1A binding site (lane 2). The NTA eluates are enriched in full-length Tc1A, because the N-terminal derivatives, lacking the histidine tag, are removed (lane 4). As expected, the DDE mutant Tc1A[D247V] retains the ability to bind the Tc1A binding site (lane 5). The pET3c fraction does not contain any detectable specific DNA binding activity (lane 6). Note that the gel retardation assay gives an under-estimation of the amount of full-length Tc1A DNA binding activity as compared with DNase I footprinting (see below).

Endonucleolytic cleavage by Tc1A

To test for specific cleavage activity, we incubated a strand-specifically labeled DNA fragment containing inverted repeat sequences with Tc1A in the presence of Mn^{2+}. Several cleavage products were detected in the reaction with wild-type transposase (Figure 5, lane 3), of which the dominant product results from cleavage at the presumed 5′ excision site; the phosphate between the A and G 2 bp in from the end of Tc1. In contrast, using a probe labeled on the other strand, no cleavage was detected at the 3′ excision site (data not shown). This is unexpected, because of our double strand excision model (Plasterk,

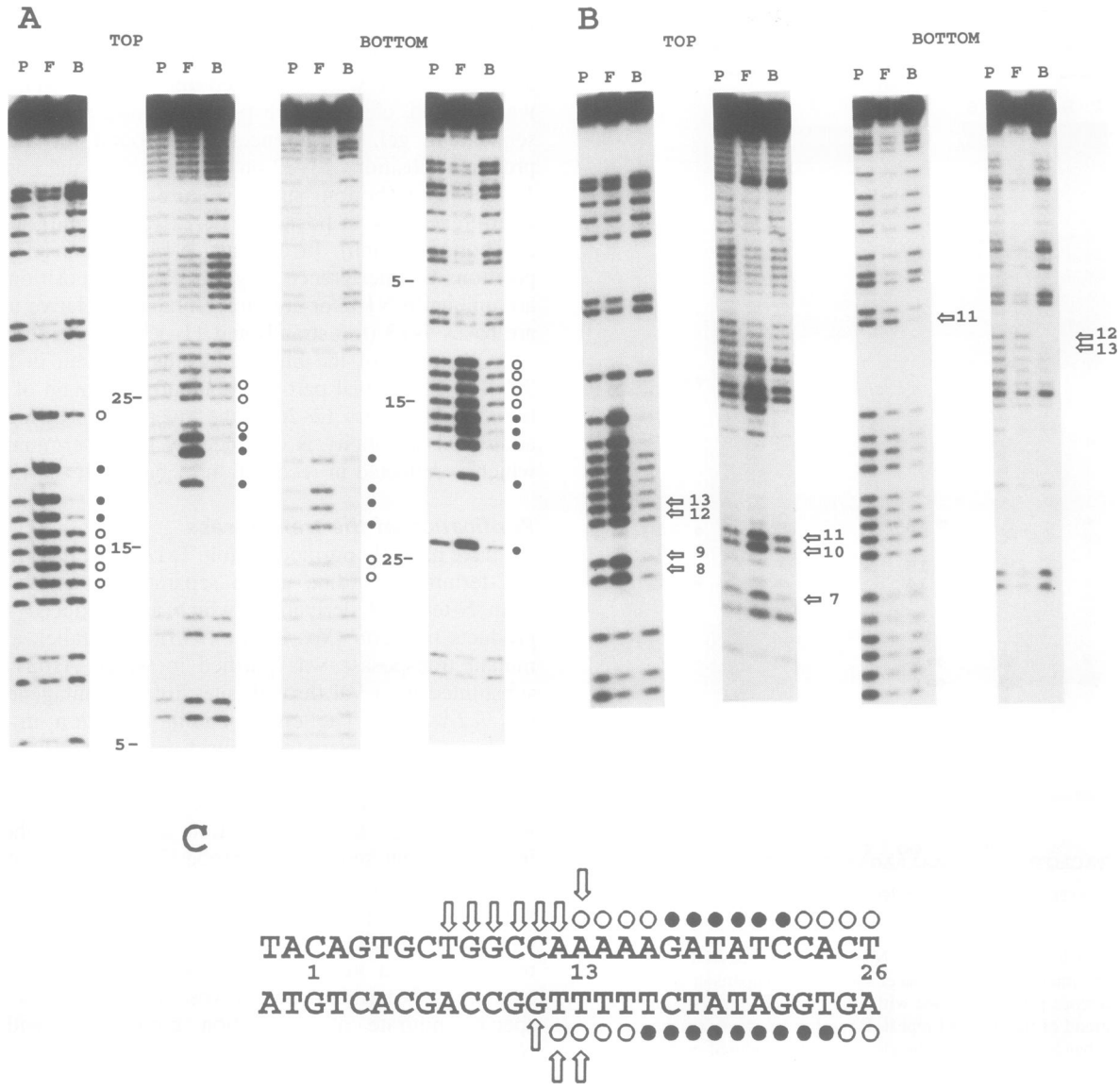

Fig. 3. DNA binding specificity of N78 and N142 analyzed by missing base interference. (**A**) Missing base interference analysis of the inverted repeat binding of N78. Each panel shows the data from one strand of the inverted repeat with either purine- (left) or pyrimidine-specific (right) modified probe (P). Free (F) and bound (B) DNA was isolated as described in Materials and methods. Circles indicate the positions where the absence of a base interferes with DNA binding. The strongest positions are marked by closed circles. The positions 5, 15 and 25 (see C) are indicated. (**B**) Missing base interference analysis of the inverted repeat binding of N142. The positions where the absence of a base specifically interferes with binding by N142 are indicated by arrows. See (A) for details. (**C**) Schematic representation of the missing base interference analyses.

1991; Van Luenen *et al.*, 1994). Thus, recombinant Tc1A gives rise to single strand nicks at the transposon ends. An identical result has been obtained with His-tagged Tc1A derived from transgenic nematodes (data not shown). A minor cleavage site, also visible in Figure 6, is in the DNA flanking the Tc1 at a sequence which is unrelated to the end of Tc1. Only a low level of cleavage was obtained in the presence of $MgCl_2$ instead of $MnCl_2$ (data not shown). One explanation for the absence of 3′ end cleavage is the use of a substrate which does not carry a complete transposon with two intact terminal repeats. It is conceivable that a proper synapsis between the two transposon ends is required for double strand cutting. However, we have not (yet) been able to detect 3′ cleavage using a template containing a complete Tc1 element. The cleavage activity we do detect is specific for Tc1A, because the control extract does not exhibit any cleavage activity (lane 1). Furthermore, the C-terminal deletion polypeptide N142 and, more interestingly, Tc1A[D247V] do not show any reactivity either (lanes 2 and 4 respectively). Therefore, nuclease activity requires an intact DDE motif.

DNase I footprinting analysis shows protection of the Tc1A binding site by both wild-type and mutant transposase (lanes 8 and 9), demonstrating again that Tc1A[D247V] is not impaired in DNA binding.

We tested the sequence specificity of cleavage at the site of excision in two ways. We mutated the binding site for transposase and we mutated the conserved TACAGT sequence at the junction between transposon and flanking DNA. This sequence is highly conserved within the Tc1/mariner family of transposons (Henikoff, 1992; Radice *et al.*, 1994). As can be seen in Figure 6A, mutation of

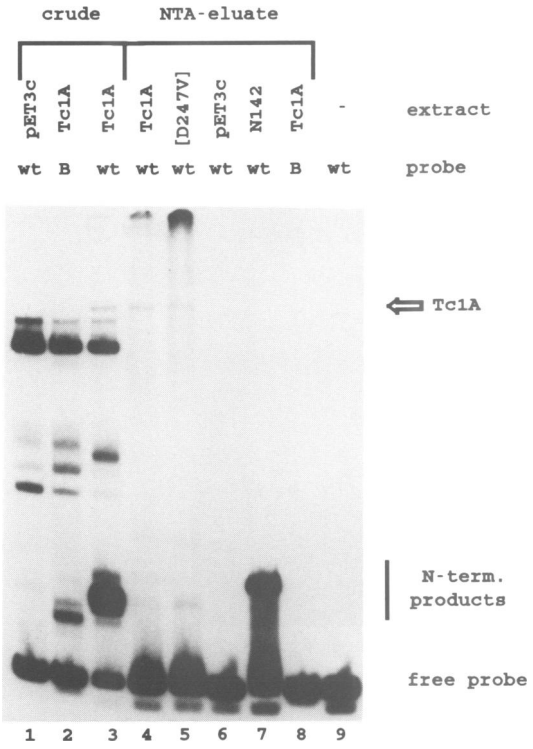

Fig. 4. Gel retardation analysis of recombinant Tc1A and Tc1A[D247V]. *E.coli* strain BL21(DE3) was transformed with either pET3c, pHT11 (expression of Tc1A) or pD247V (expression of Tc1A[D247V]). Extracts were prepared from induced cultures and passed on a Ni^{2+}–NTA resin. NTA eluates and crude extracts were tested for specific binding to a Tc1 inverted repeat probe. Standard binding reactions contained 2 µg poly(dI–dC) and 1 µl of crude lysates or 0.2 µl of NTA eluates. Complexes were analyzed on a 4% polyacrylamide gel. The usage of wild-type Tc1 inverted repeat sequence (wt) or mutant sequence B (see Figure 1B) as the probe is indicated at the top of the lanes, together with the source of protein. The position of the full-length Tc1A–DNA complex is indicated by an arrow.

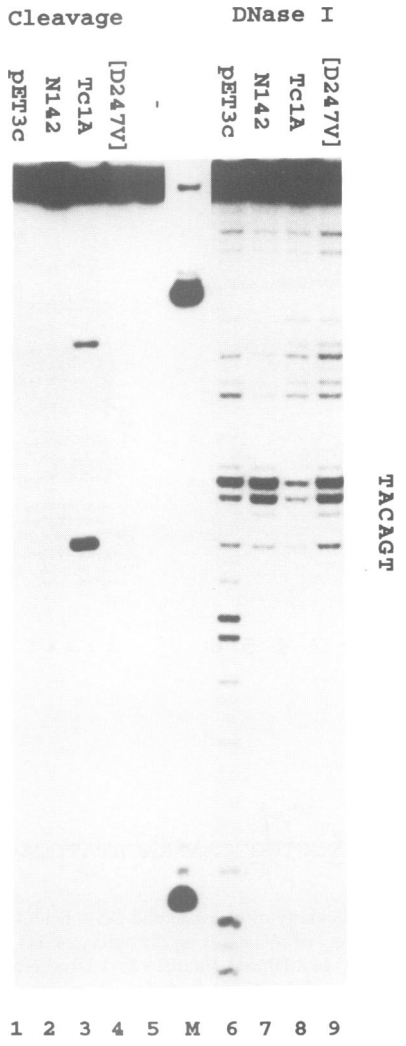

Fig. 5. Endonucleolytic and DNA binding activity of Tc1 transposase. A 3′ end-labeled *Hin*dIII–*Sac*I DNA probe containing the 26 terminal base pairs of Tc1 was incubated in the presence of 2 µg poly(dA–dT) and 4 mM $MnCl_2$ for 1 h at room temperature with 1 µl of NTA eluate (lanes 1–4). Alternatively, DNase I footprinting was performed using the same amounts of protein, competitor and probe (lanes 6–9). Products were separated on a sequencing gel. Lane 5 shows a cleavage reaction without the addition of an extract and lane M contains *Bam*HI and *Eco*RV cleavage products. The sequence TACAGT represents the TA target site and the first four nucleotides of Tc1. A minor cleavage site is at the *Bam*H1 site (GGATC↓C) in the sequence flanking the end of Tc1.

the transposase binding site does not affect cleavage at the junction site (lane 3), although binding to the inverted repeat is abolished (Figure 6B, lane 2). Therefore, it appears that stable transposase binding, as measured by DNase I footprinting, is not essential for cleavage. Apparently, we monitored the interaction between the catalytic site of Tc1A and the substrate DNA. This is further supported by the observation that the same cleavage is observed with Tc3 as substrate DNA, which has the same terminal four base pairs as Tc1. Furthermore, deletion of the N-terminal 70 amino acids of Tc1A does not abolish cleavage activity, although the resulting protein does not have any detectable specific binding affinity for Tc1 ends (data not shown). The sequence specificity of cleavage was further studied by mutations within the sequence TACAGT, which do not affect transposase binding (Figure 6B, lanes 3–5). Mutation of either TA (Figure 6A, lane 4) or GT (lane 6) considerably reduces cleavage at the 5′ excision site. Mutation of CA has no effect on cleavage at the 5′ junction (lane 5). The effect of these mutations on Tc1 transposition *in vivo* remains to be determined. Nevertheless, it is likely that the conserved TACAGT sequence is important during catalysis.

Tc1A has phosphoryl transfer activity

The retroviral integrase proteins (IN), which have a well-defined DDE motif, can execute a so-called disintegration reaction (Chow *et al.*, 1992). In this reaction, the 3′ hydroxyl of a 5′-labeled oligonucleotide attacks a phosphate within another oligonucleotide (schematically indicated at the bottom of Figure 7). For retroviral integrases it was found that this reaction could occur independently of the presence of specific sequences in the DNA substrate. Because Tc1A contains a DDE motif, we tested whether it has a similar phosphoryl transfer activity. Figure 7 shows that Tc1A can indeed perform this reaction (lane 3), although with a much lower efficiency than HIV-1 IN (lane 5). Activity correlates with the presence of an intact DDE motif, because Tc1A[D247V] is not active (lane 4).

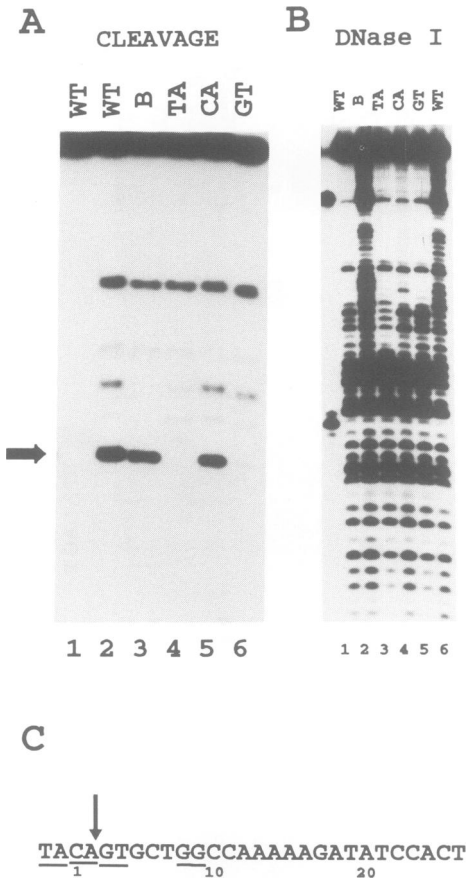

Fig. 6. Sequence specificity of cleavage and DNA binding by Tc1 transposase. (**A**) Effect of mutations on the ability of Tc1A to cleave at the 5′ junction. 3′ End-labeled *Hin*dIII–*Sac*I DNA probes were incubated with 1 µl of Tc1A NTA eluate for 1 h at room temperature. The probe contained the wild-type Tc1 inverted repeat sequence (lane WT) or a mutation as schematically shown at the bottom of the figure of the TA sequence into CG (lane TA), the CA sequence into TG (lane CA) or the GT sequence into AC (lane GT), or probe B (see Figure 1B). Lane 1 is a control reaction without extract. Cleavage products were analyzed on a sequencing gel. The arrow points to the cleavage product at the 5′ excision site. (**B**) DNase I footprinting analysis of the binding of Tc1A to the mutated Tc1 inverted repeats. The same probes as described in Figure 6A were used in a DNase I footprinting assay using 1 µl of Tc1A NTA eluate and 2 µl of poly(dA–dT) as competitor (lanes 1–5). Lane 6 is a reaction in the absence of extract. Cleavage products were analyzed on a sequencing gel. (**C**) 5′ Terminal sequence of the Tc1 inverted repeat and the flanking TA dinucleotide. The underlined sequences are mutated in the probes that are described in A. The arrow indicates the position of the 5′ excision site.

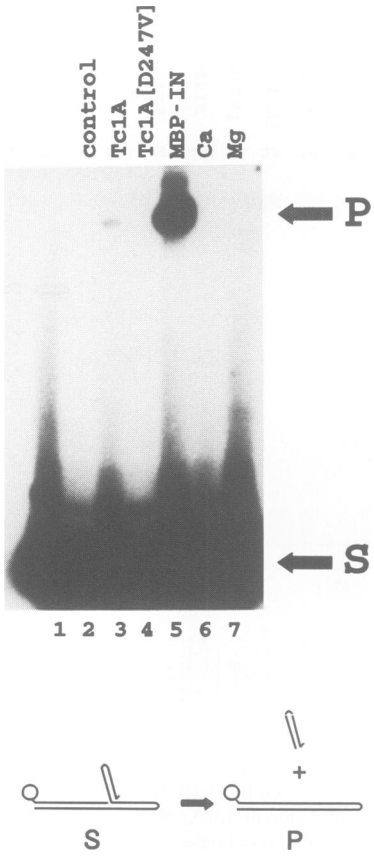

Fig. 7. Tc1 transposase has phosphoryl transfer activity. The phosphoryl transfer reaction is schematically given at the bottom of the figure. The open circle represents the 5′ ^{32}P label. The 3′ hydroxyl of the labeled strand in the substrate (marked S) is used in the phosphoryl transfer reaction to cleave the longer oligonucleotide and generate a labeled 40mer (marked P). The substrate was incubated for 1 h at room temperature in the presence of either 4 mM $MnCl_2$ (lanes 1–5), 10 mM $CaCl_2$ (lane 6) or 10 mM $MgCl_2$ (lane 7) and 1 µl of NTA elutes from pET3c (lane 2), pHT11 (lanes 3, 6 and 7) or pD247 or with 50 ng purified maltose binding protein–HIV-1 integrase fusion protein (lane 5). Products were analyzed on a 10% sequencing gel.

The reaction is dependent on Mn^{2+}, no product being generated in the presence of either Ca^{2+} or Mg^{2+}. As in the nuclease assay, an energy source such as ATP is not required (data not shown).

The phosphoryl transfer activity of Tc1A underscores the conservation of a catalytic domain present in proteins involved in transpositional recombination.

Discussion

Biochemical characterization of Tc1 transposase has resulted in the following main conclusions. The sequence-specific DNA binding domain of Tc1A has a bipartite structure. Furthermore, Tc1A is active in endonuclease cleavage and phosphoryl transfer. Both of these activities depend on an intact DDE motif.

The analysis of the transposase binding site of Tc1A revealed that in addition to the previously identified N-terminal domain of 63 amino acids, another part of the protein contributes to sequence-specific DNA binding. Mutational analysis of the binding site, as well as interference studies, define the bipartite recognition of the inverted repeat. The N-terminal part of the bipartite domain interacts with base pairs 12–25. The methylation interference data indicate that adjacent minor and major groove contacts are involved. The C-terminal part of the bipartite domain shows additional contacts with base pairs 7–12, which are apparently mainly via the major groove.

Protein sequence alignments have revealed a weak similarity between the transposase of the Tc1-like Minos element and the *paired* DNA binding domain found in mammalian and *Drosophila* genes (Franz *et al.*, 1994). The significance of the similarity is enhanced by the identification of the bipartite nature of the *paired* DNA binding domain (Czerny *et al.*, 1993) and the DNA binding study presented here for Tc1A. Especially striking is that

for both the *paired* domain and the Tc1A protein a comparable model can be drawn, in which the N-terminal half of the bipartite domain interacts with the 3′ part of the non-palindromic binding site and the C-terminal half binds to the 5′ part of the binding site (Czerny *et al.*, 1993). The *paired* domain was recognized as a conserved stretch of 128 amino acids (Bopp *et al.*, 1989), which could be classified in several subfamilies according to both sequence similarity and genomic organization (Walther *et al.*, 1991). The Tc1A bipartite domain cannot be grouped into those subfamilies and therefore may at best constitute a distantly related member of the *paired* family which has found its place in a transposition protein rather than a transcription factor.

The partial purification of soluble Tc1A from *E.coli* enabled us to investigate its catalytic activity. We could demonstrate single strand endonuclease activity, as well as phosphoryl transfer. Although these reactions do not represent a complete transposition event, it is the first example of any *in vitro* catalytic activity for a transposase of the Tc1/mariner family. Furthermore, the assays appear to be useful in the identification of pivotal amino acids in the protein. Because the endonuclease activity does not correlate with binding of Tc1A to the inverted repeat, it is likely that the catalytic domain is active independent of the DNA binding domain. This is also suggested by the observation that non-specific competitor DNA can inhibit cleavage of the inverted repeat probe, but does not interfere with sequence-specific binding to the probe (data not shown). The most prominent activity observed in the nuclease assay is cleavage at the 5′ excision site, whereas cleavage at the 3′ junction between transposon and flanking DNA is not detected. The pattern of cleavage products indicates sequence specificity in the endonuclease reaction. This is further supported by the effects of mutations in the conserved TACAGT sequence at the junction between transposon and flanking DNA. These mutations have no effect on the ability of Tc1A to bind to the inverted repeat. However, they do affect cleavage by Tc1A, which is in line with the picture which emerges from most transposable elements, where the outer few base pairs are important for catalytic steps in the transposition reaction, but not for binding of the transposase (Huisman *et al.*, 1989; Derbyshire and Grindley, 1992; Mizuuchi, 1992).

For Tc3 it has been shown that mutation of the flanking TA dinucleotide into GC does not impair transposition *in vivo* (Van Luenen *et al.*, 1994), whereas our results show reduced 5′ cleavage *in vitro* when we mutate the TA into CG. Obvious explanations include a difference between Tc1 and Tc3 transposases and the different base substitutions introduced. Alternatively, 5′ cleavage may be a rate-limiting step in our *in vitro* assay, but not in the overall complete transposition reaction.

In analogy with other transposition systems, we believe that the 3′ end of the transposon, carrying a hydroxyl group, is used as a nucleophile in the strand transfer reaction (Van Luenen *et al.*, 1994). This ensures that no transposon sequences are lost in the transposition process, although 5′ end cleavage is 2 bp in from the end of Tc1.

To avoid non-productive transpositional recombination, the transposition steps are likely to be controlled, for instance in the recognition of the relative orientation of inverted repeats (see Mizuuchi, 1992). Therefore, the use of an isolated transposon end in our cleavage assay may not result in a productive DNA–protein complex, because a synapsis with two transposon ends is not formed. However, we have not (yet) been able to detect double strand excision of a complete Tc1 element *in vitro*. It is possible that another *C.elegans* protein may be required for double strand cleavage at the junction between transposon and host DNA. Alternatively, the C-terminal histidine tag on our recombinant transposase may interfere with assembly of a higher order protein–DNA complex or the recombinant protein may otherwise differ from the active transposase made in the worm.

The phosphoryl transfer reaction, or disintegration reaction, has been well studied for the retroviral integrase proteins (Chow *et al.*, 1992; Donzella *et al.*, 1993; Mazumber *et al.*, 1994). This reaction resembles integration, because it involves the attack on a phosphodiester bond by a 3′ hydroxyl group. The reaction appears to be rather sequence-independent and only requires the catalytic core of the integrase protein (Bushman *et al.*, 1993; Vink *et al.*, 1993). Mutation of any of the DDE amino acids abolishes the phosphoryl transfer reaction (Drelich *et al.*, 1992; Engelman *et al.*, 1992; Kulkosky *et al.*, 1992; Van Gent *et al.*, 1992). The ability of Tc1A to execute phosphoryl transfer like the retroviral integrase and the essential role of the DDE motif demonstrate the conservation of the mechanism of transpositional recombination.

Materials and methods

Expression and purification of six histidine-tagged transposase

Oligonucleotides containing six histidine triplets were used in a PCR to generate a transposase gene with an artificial C-terminus. Sequences and names of the oligonucleotides used are as follows: NTER1 (5′ GAGCCATATGGTAAAATCTGTTGGGTG); H68 (5′CCGTGGATCC-TTAGTGATGGTGGTGATGATGGGTTGTCACTCGAGGCC); H142 (5′CCGTGGATCCTTAGTGATGGTGGTGATGATGCGCTTTTGCC-CACGC); H343 (5′GCTGGATCCTTAGTGATGGTGGTGATGGTG-ATACTTTGTCGCGTATCCG). The relevant *Bam*HI and *Nde*I sites are underlined. PCR product were cloned as *Nde*I–*Bam*HI fragments into pET3c (Rosenberg *et al.*, 1987). In this way, a plasmid encoding full-length Tc1A was obtained (pHT11, oligonucleotides NTER1 and H343), as well as plasmids encoding either the first 68 (pN68His, oligonucleotides NTER1 and H68) or 142 (pN142His, oligonucleotides NTER1 and H142) amino acids of Tc1A. Site-directed mutagenesis was used to change codon 247 of Tc1A from aspartic acid to valine. Therefore, a *Bgl*II–*Eco*RV fragment of pHT11 was cloned into *Hin*dIII-blunted/*Bam*HI-restricted pMa/c5-8 (Stanssens *et al.*, 1989). The mismatch oligonucleotide was 5′CGTTTTCAGCAGGTTAACGATCCTAAGC. The plasmid (pD247V) carrying the mutant gene was identified by a newly generated *Hpa*I site.

Induction of the transposase or of N-terminal derivatives was in *E.coli* BL21(DE3) for 1–2 h at 30°C. Bacteria were sonicated in buffer H (25 mM HEPES, pH 7.5, 0.1 mM EDTA, 1 M NaCl, 15% glycerol, 0.25% Tween 20, 1 mM β-mercaptoethanol) and 10 mM imidazole (pH 8) was added to the soluble fraction before it was passed over a Ni^{2+}–NTA resin (Qiagen). Bound proteins were eluted with 100 mM imidazole in buffer H. Protein concentrations were ~2.5 mg/ml for the pET3c, pHT11 and pD247V extracts. Transposase was 5–10% of total protein. The protein concentrations of eluates for pN68His and pN142His were 0.1 mg/ml and 0.5 mg/ml and the N-terminal polypeptides were ~80% pure. The purity reflects the higher solubility of the N-terminal derivatives compared with the full-length transposase.

The purification of the N-terminal derivative which is expressed adventitiously together with full-length Tc1A has been described before (Vos *et al.*, 1993). The purification of the non-His-tagged N-terminal 78 amino acid polypeptide (Vos *et al.*, 1993) was achieved by chromatography of the soluble fraction of a pN78 extract on a S Sepharose FF (Pharmacia) column with 200 mM NaCl in H buffer. The polypeptide

was eluted with 1 M NaCl in H buffer, diluted twice with H buffer without NaCl and applied to a Blue Sepharose (Pharmacia) column. Elution was with H buffer containing 2 M NaCl. Final protein concentration was 0.2 mg/ml and the polypeptide was ~90% pure.

Methylation and missing base interference

Oligonucleotides (69mers) corresponding to the right end inverted repeat of Tc1, as well as the flanking sequence at the Tc1 insertion allele of *unc22* (st192) (Moerman and Waterston, 1989), were end-labeled with [γ-^{32}P]ATP. Ten picomoles of double-stranded oligonucleotides were modified in the presence of 5 µg sonicated salmon sperm DNA using formic acid (depurination), hydrazine (depyrimidation) or DMS (methylation of purines) according to Maxam and Gilbert (1980). Five picomoles of probe was incubated with 2 µl of purified N68, N78 or N142 in the presence of 2 µg poly(dI–dC), 25 mM Tris, pH 7.4, 1 mM DTT, 200 mM NaCl, 1 mM spermidine, 0.05 µg BSA in a total volume of 20 µl. Free DNA was separated from protein-bound DNA on a 4% native polyacrylamide gel in 1× TAE buffer. DNA was eluted from the gel and cleaved with piperidine (missing base interference) or with NaOH (methylation interference). Finally, DNA products were separated on a 7% sequencing gel.

DNase I footprinting, gel retardation, cleavage and phosphoryl transfer assays

The standard conditions for these various reactions were 25 mM Tris, pH 7.4, 1 mM spermidine, 0.05 µg/µl BSA, 1 mM DTT, 100 mM NaCl in 20 µl. Cleavage and phosphoryl transfer reactions included 4 mM $MnCl_2$ (or 10 mM $CaCl_2$ or $MgCl_2$ where indicated) and 2 µg poly(dA–dT). DNase I footprinting and gel retardations were performed in the absence of divalent cations and competitor DNA was present as indicated. After binding for 15 min on ice, 5 µl of 25% Ficoll were added for bandshift analysis and complexes were separated on a 4% native polyacrylamide gel. In the case of footprinting, 2 µl of a DNase I solution (0.01 mg/ml in 100 mM $CaCl_2$, 100 mM $MgCl_2$) was added to the binding reaction for 1–2 min at room temperature. Phosphoryl transfer reactions were carried out with 0.1 pmol of disintegration substrate (Vink *et al.*, 1994), whereas the other assays included 5 fmol of DNA fragments. These fragments were obtained from pUC19 derivatives into which 28mer oligonucleotides, representing the terminal 26 base pairs of Tc1 (or mutations thereof), were cloned in the *Hin*dII site in addition to the flanking TA target dinucleotides. These plasmids were digested with either *Pst*I and *Eco*RI or with *Sac*I and *Hin*dIII and the fragments were gel purified and end-labeled in a strand-specific manner with [α-^{32}P]dATP and Klenow.

Acknowledgements

We thank Fusinita van den Ent and Kees Vink for HIV IN protein and disintegration substrates. We thank our colleagues for discussions and Ivo De Baere, Piet Borst, Henri van Luenen, Kenny Simmen and Kees Vink for critically reading the manuscript. This work was supported by a Pionier grant of the Dutch Science Foundation NWO.

References

Bopp,D., Burri,M., Baumgartner,S., Frigerio,G. and Noll,M. (1986) *Cell*, **47**, 1033–1040.

Bushman,F.D., Engelman,A., Palmer,I., Wingfield,P. and Craigie,R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3428–3432.

Chow,S.A., Vincent,K.A., Ellison,V. and Brown,P.O. (1992) *Science*, **255**, 723–726.

Czerny,T., Schaffner,G. and Busslinger,M. (1993) *Genes Dev.*, **7**, 2048–2061.

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

Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 942–946.

Drelich,M., Wilhelm,R. and Mous,J. (1992) *Virology*, **118**, 459–468.

Donzella,G.A., Jonsson,C.B. and Roth,M.J. (1993) *J. Virol.*, **67**, 7077–7087.

Engelman,A. and Craigie,R. (1992) *J. Virol.*, **66**, 6361–6369.

Fayet,O., Ramond,P., Polard,P., Prére,M.F. and Chandler,M. (1990) *Mol. Microbiol.*, **4**, 1771–1777.

Franz,G., Loukeris,T.G., Dialektaki,G., Thompson,C.R.L. and Savakis,C. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 4746–4750.

Henikoff,S. (1992) *New Biol.*, **4**, 382–388.

Hochuli,E., Bannwarth,W., Döbeli,H., Gentz,R. and Stüber,D. (1988) *Bio/Technol.*, **6**, 1321–1325.

Huisman,O., Errada,P.E., Signon,L. and Kleckner,N. (1989) *EMBO J.*, **8**, 2101–2109.

Kahn,E., Mack,J.P.G., Katz,R.A., Kulkosky,J. and Skalka,A.M. (1991) *Nucleic Acids Res.*, **19**, 851–860.

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

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.

Mazumber,A., Engelman,A., Craigie,R., Fesen,M. and Pommier,Y. (1994) *Nucleic Acids Res.*, **22**, 1037–1043.

Mizuuchi,K. (1992) *Annu. Rev. Biochem.*, **61**, 1011–1051.

Moerman,D.G. and Waterston,R.H. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 537–556.

Plasterk,R.H.A. (1991) *EMBO J.*, **10**, 1919–1925.

Radice,A.R., Bugaj,B., Fitch,D.H.A. and Emmons,S.W. (1994) *Mol. Gen. Genet.*, in press.

Robertson,H.M. (1993) *Nature*, **311**, 485–486.

Rosenberg,A.H., Lade,B.N., Chui,D.-s., Lin,S.-W., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125–135.

Rosenzweig,B., Liao,L.W. and Hirsch,D. (1983a) *Nucleic Acids Res.*, **12**, 4201–4209.

Rosenzweig,B., Liao,L.W. and Hirsch,D. (1983b) *Nucleic Acids Res.*, **12**, 7137–7140.

Ruan,K. and Emmons,S.W. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 4018–4022.

Stanssens,P., Opsomer,C., McKeown,Y., Kramer,W., Zabeau,M. and Fritz,H.J. (1989) *Nucleic Acids Res.*, **17**, 4441–4454.

Van Gent,D.C, Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9598–9602.

Van Luenen,H.G.A.M. and Plasterk,R.H.A. (1994) *Nucleic Acids Res.*, **22**, 262–269.

Van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1993) *EMBO J.*, **12**, 2513–2520.

Van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1994) *Cell*, **79**, 293–301.

Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) *Nucleic Acids Res.*, **21**, 1419–1425.

Vink,C., van der Linden,K.H. and Plasterk,R.H.A. (1994) *J. Virol.*, **68**, 1468–1474.

Vos,J.C., van Luenen,H.G.A.M. and Plasterk,R.H.A. (1993) *Genes Dev.*, **7**, 1244–1253.

Walther,C., Guenet,J.-L., Simon,D., Deutsch,U., Jostes,B., Goulding,M.D., Plachov,D., Balling,R. and Gruss,P. (1991) *Genomics*, **11**, 424–434.

Received on July 27, 1994; revised on September 23, 1994