5548–5554 Nucleic Acids Research, 1994, Vol. 22, No. 25

DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase

Sean D.Colloms+, Henri G.A.M.van Luenen and Ronald H.A.Plasterk*
The Netherlands Cancer Institute, Division of Molecular Biology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received October 18, 1994; Revised and Accepted November 18, 1994

ABSTRACT

Tc3 is a member of the Tc1/mariner family of transposable elements. All these elements have terminal inverted repeats, encode related transposases and insert exclusively into TA dinucleotides. We have studied the DNA binding properties of Tc3 transposase and found that an N-terminal domain of 65 amino acids binds specifically to two regions within the 462 bp Tc3 inverted repeat; one region is located at the end of the inverted repeat, the other is located ~180 bp from the end. Methylation interference experiments indicate that this N-terminal DNA binding domain of the Tc3 transposase interacts with nucleotides on one face of the DNA helix over adjacent major and minor grooves.

INTRODUCTION

Tc3 of *Caenorhabditis elegans* is one of the best characterized members of the Tc1/mariner family of transposable elements; a family of transposons found in a wide variety of organisms (1–15). Members of the Tc1/mariner family of transposons are distinguished by related transposase genes flanked by terminal inverted repeats of ~50 bp. Tc3 is an exception because it has terminal inverted repeats of 462 bp (16). The sequence of the terminal inverted repeats are not conserved between different elements except for the 4 most terminal nucleotides (5′-CA-GT-3′). Transposons belonging to the Tc1/mariner family always insert into the sequence TA, and this sequence is found as a direct repeat on both sides of the integrated element.

Forced expression of the Tc3 transposase protein (Tc3A) in *C.elegans* leads to excision and transposition of Tc3 (17). It has recently been shown that Tc3 is excised by a pair of staggered double-strand cleavages at both ends of the integrated transposon (18). The 3′ end of the excised transposon coincides with the last Tc3 nucleotide and carries a 3′ OH group. The 5′ end lacks the two most terminal transposon-encoded nucleotides. The 3′ transposon ends are thought to integrate into a target TA dinucleotide at positions staggered by 2 bp. After repair of the 4 nucleotide single-stranded gap at each transposon end, an intact copy of Tc3 flanked by a TA target duplication is produced.

In order to carry out the transposition reaction, the transposase (alone or as part of a multimeric complex) must first recognize and synapse the transposon termini before cleavage at the transposon ends. Furthermore, the transposase has to recognize the target DNA and insert the transposon ends with a 2 bp stagger. The present study is a first step in the analysis of the protein–DNA interactions required for these steps in the transposition reaction. The DNA binding properties of Tc1 transposase (Tc1A) have been studied in some detail (19). The N-terminal 63 amino acids of Tc1A bind specifically to the sequence between positions 5 and 26 bp from the transposon ends. A second, non-specific, DNA binding domain is located between amino acids 71 and 207 of Tc1A. It has previously been shown that Tc3 transposase binds specifically to the terminal sequences of Tc3 (17). Here we show that Tc3 transposase has two distinct binding activities. We use a South Western assay to demonstrate that the N-terminal 159 amino acids of Tc3A have a non-specific DNA binding activity, and we use electrophoretic mobility shift assays, DNase I footprinting and methylation interference experiments to characterize specific binding of an N-terminal domain of Tc3A to the Tc3 termini. We found that Tc3A, in contrast to Tc1A, binds to two regions in each repeat.

MATERIALS AND METHODS

Expression of Tc3A derivatives in *Escherichia coli*

To express full-length Tc3A, the Tc3 transposase gene, with its intron precisely removed by PCR (17), was inserted as a *NdeI–BglII* fragment into *NdeI* and *BamHI* cleaved pET3c (20), yielding pRP712. pRP712 was partially digested with *HindIII*, linear pRP712 was gel purified and 5′ overhanging ends were filled in with dNTPs and Klenow. The resulting blunt ends were ligated together to yield pSDC332, in which the *HindIII* site at codons 191 and 192 of *Tc3A* has been filled in. pSDC332 thus carries the normal *Tc3A* sequence up to codon 192, codon 193 has an alanine for glutamate substitution, which is followed immediately by a TGA stop codon. pSDC331 encoding $Tc3A_{1-159}$ was created by inserting the PCR product of primers AB1985 (5′-AAAAACATATGCCTCGAGGATCTGCCCTTT-

*To whom correspondence should be addressed

+Present address: University of Oxford, Unit of Microbiology, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK

CG) and AB4327 (5'-AAAAAAAGATCTTTACCAATAGTA-GCGGCAACCGTC) using pRP712 as template, as a *Nde*I−*Bgl*II fragment into the *Nde*I and *Bam*HI sites of pET3c. Plasmids expressing smaller derivatives of Tc3A were made by site-directed mutagenesis to produce two consecutive stop codons marked by the presence of either a *Nhe*I or a *Spe*I restriction site, at chosen positions within the *Tc3A* gene. The 0.6 kb *Xba*I−*Hin*dIII fragment from pRP712, containing the first 192 codons of *Tc3A*, was inserted into *Xba*I and *Hin*dIII-cleaved pMa5-8 to yield pSDC326. Gapped duplex molecules were made and site-directed mutagenesis was carried out as described by Stanssens *et al.* (21). Mutated *Tc3A* genes complete with stop codons at chosen positions were then inserted into *Xba*I and *Bam*HI-cleaved pET3c as a *Xba*I−*Hin*cII fragment, the *Bam*HI

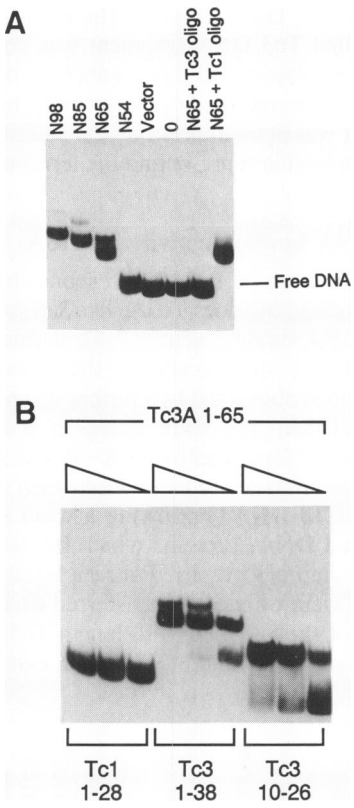

Figure 1. Sequence-specific DNA binding properties of Tc3A N-terminal derivatives. (**A**) *E.coli* extracts containing Tc3A derivatives were incubated in the presence of 1 μg of poly(dI−dC) with a ^{32}P-labelled 94 bp *Hin*dIII−*Eco*RI fragment from pSDC304, containing the 38 terminal bp of Tc3 within the pUC18 polylinker. Protein−DNA complexes were resolved by native electrophoresis on a 5% polyacrylamide gel. The bacterial extracts used (0.1 μl of each) contain $Tc3A_{1-98}$ (lane 1), $Tc3A_{1-85}$ (lane 2), $Tc3A_{1-65}$ (lane 3) or $Tc3A_{1-54}$ (lane 4). Lane 5 contains 0.1 μl of a control bacterial extract with no Tc3A derivatives and lane 6 contains no added protein. Lanes 7 and 8 are identical to lane 3 except for the addition of 10 pmol (an ~1000-fold excess over the labelled fragment) of unlabelled double-stranded oligonucleotide as competitor. The competitor in lane 6 represents the terminal 38 bp of Tc3, whereas the competitor in lane 7 represents the terminal 28 bp of Tc1. (**B**) Specificity of binding of $Tc3A_{1-65}$. 1.0 μl (lanes 1, 4, 7), 0.1 μl (lanes 2, 5, 8) or 0.01 μl (lanes 3, 6, 9) of a bacterial extract containing $Tc3A_{1-65}$ was incubated with a ^{32}P-end-labelled DNA fragment containing either the terminal 28 bp of Tc1 (lanes 1−3), the terminal 38 bp of Tc3 (lanes 4−6) or nucleotides 10−26 of Tc3 (lanes 7−9). DNA fragments were obtained as *Hin*dIII−*Eco*RI fragments from plasmids containing transposon terminal sequences inserted into the *Hin*cII site of pUC18. Protein−DNA complexes were resolved as in (A).

end having first been filled in with dNTPs and Klenow. AB3524 (5'-CATCTAAACGCTAGCCTTGACGCAAAGCTC) was used to make pSDC327 which encodes $Tc3A_{1-54}$, AB3525 (5'-CCGTGCGTGACTAGTAAAATGTGATT) was used in the construction of pSDC328 which encodes $Tc3A_{1-65}$, AB3526 (5'-GATATTCGCAACTAGTAACAATTGTCTG) was used in the construction of pSDC329 which encodes $Tc3A_{1-85}$ and AB3528 (5'-CCATCCTCAACTAGTAAAAACGATCTGG) was used in the construction of pSDC330 which encodes $Tc3A_{1-98}$.

For expression of Tc3A and the C-terminal deletion mutants of Tc3A, *E.coli* BL21 <DE3> pLysS was transformed with the above pET3c-derived expression plasmids. Cultures were grown in LB broth at 37°C until they reached an OD of 0.5 at 600 nm, IPTG was added to a final concentration of 0.5 mM and cultures were allowed to grow for a further 2 h at 37°C. Bacteria were harvested by centrifugation and resuspended in a buffer containing 50 mM HEPES, pH 7.6, 50 mM NaCl, 1 mM EDTA and 1 mM DTT. Cells were broken by sonication and insoluble material was removed by centrifugation at 20,000 *g* for 30 min. This routinely yielded crude extracts with a total protein concentration of 40 mg/ml, of which 1−5% was Tc3A derivative.

Gel retardation assays

For gel retardation assays the *Hin*dIII−*Eco*RI fragment of pSDC304, containing the oligonucleotide AB3104 (5'-T-CAAATTAGGGGGGGTCCTATAGAACTTTCCCACAC-TGTA) annealed to its complement AB3105 and inserted in the *Hin*cII site of pUC18, was labelled at both ends with [α-^{32}P]dATP using Klenow. Some experiments also used the *Hin*dIII−*Eco*RI fragment of pSDC335, containing the oligonucleotide AB4325 (5'-GGGTCCTATAGAACTTT3') and its complement AB4326 inserted in the *Hin*cII site of pUC18, or the *Hin*dIII−*Eco*RI fragment of a plasmid containing the terminal 28 bp of Tc1 in the *Hin*cII site of pUC18 (19) similarly end labelled. Binding reactions contained 10 fmol of labelled DNA in 10 μl of a buffer containing 25 mM HEPES, pH 7.6, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA, 2.5 mM spermidine, 10% glycerol and 0.1 mg/ml poly(dI−dC). The reaction was started by adding 0.01−1.0 μl of protein extract and incubated at 20°C for 20 min, then loaded directly onto a 5% polyacrylamide gel in 50 mM Tris−HCl, pH 8.0, 0.1 mM EDTA. Gels were run at 10 V/cm for 4 h at 4°C. Protein−DNA complexes were visualized by autoradiography.

South Western assays

Whole cell lysates of IPTG-induced *E.coli* BL21 <DE3> pLysS containing the following plasmids: pET3c, pSDC327, pSDC328, pSDC329, pSDC330, pSDC331, pSDC332 or pRP712, were separated by electrophoresis through a 12.5% SDS−polyacrylamide gel and transferred to nitrocellulose by electroblotting. The filter was processed essentially as described in Krogstad and Champoux (22), using pUC18 DNA labelled with [α-^{32}P]dATP by random priming (23) as probe.

DNase I footprinting assays

A 340 bp *Asn*I fragment from TR # 10 (16), containing 317 bp of the Tc3 right end and 22 bp of flanking *unc-22* sequences, was made blunt using Klenow, dTTP and dATP and inserted in both orientations into the *Hin*cII site of pUC18 to yield pSDC324 and pSDC325. To see protection of the most terminal

Tc3 sequences, pSDC324 was cleaved with *Bam*HI, labelled at the 3′ end with Klenow and [α-^{32}P]dATP or at the 5′ end with [α-^{32}P]ATP and T4 polynucleotide kinase, and subsequently cleaved with *Pst*I. To see protection of more internal Tc3 sequences, pSDC325 was similarly labelled at the *Bam*HI site and subsequently cleaved with *Pst*I. Labelled DNA fragments were purified away from the labelled vector fragment by electrophoresis through low melting point agarose. Binding reactions containing 200 c.p.s. labelled DNA and 0.4 μl of the appropriate crude extract were set up in a volume of 20 μl in the same buffer used for gel retardation reactions, except that spermidine was omitted. After 10 min of incubation at 20°C, 1 μl of DNase I (2.5 μg/ml in 100 mM $CaCl_2$ and 100 mM $MgCl_2$) was added. Reactions were stopped after 1 min by the addition of 5 μl stop solution containing 1% SDS and 300 mM EDTA. Reaction products were extracted once with phenol/chloroform, ethanol precipitated, resuspended in formamide loading mix and separated on an 8% polyacrylamide sequencing gel.

Methylation interference

The plasmids pSDC304 and pSDC305, containing both possible orientations of the oligonucleotides AB3104 and its complement AB3105 inserted in the pUC18 *Hin*cII site, were labelled with [α-^{32}P]dATP using Sequenase 2.0 (United States Biochemical) to fill in the *Hin*dIII site, and subsequently cleaved with *Eco*RI. DNA was ethanol precipitated and resuspended in 200 μl of 50 mM sodium cacodylate and 1 mM EDTA. Labelled DNA was methylated by the addition of 1 μl of DMS and the reactions were stopped after 5 min by adding 5 μl β-mercaptoethanol. DNA was precipitated twice by the addition of ammonium acetate to 0.8 M and 2.5 volumes of ethanol. Binding reactions were set up and run on a polyacrylamide gel as above, using ~3000 c.p.s. of methylated labelled DNA and 1 μl of crude extract containing either $Tc3A_{1-65}$ or $Tc3A_{1-85}$. Bound and unbound DNA was eluted from gel slices into 0.5 M ammonium acetate and 1 mM EDTA and precipitated with 2.5 volumes of ethanol. DNA was resuspended in 20 μl 20 mM potassium phosphate buffer, pH 7.0, and 1 mM EDTA and incubated at 90°C for 10 min. Reactions were placed briefly on ice, 2 μl of 1.0 M KOH was added and then incubated for a further 10 min at 90°C. Reactions were placed on ice, 180 μl of 10 mM Tris−HCl, pH 7.6, 1 mM EDTA and 20 μl of 3 M sodium acetate were added and DNA was precipitated with 2.5 volumes of ethanol. After a 70% ethanol wash, reaction products were resuspended in formamide loading mix and separated on a 12% polyacrylamide sequencing gel.

RESULTS

Specific binding of Tc3A derivatives to transposon ends

Specific binding of a MBP−Tc3A fusion protein, purified from *E.coli*, to the terminal 38 bp of Tc3 has previously been described (17). To ascertain which region of the 329 amino acid Tc3A protein is responsible for binding to the transposon ends, Tc3A and a series of deletion derivatives of Tc3A were expressed in *E.coli*. A copy of the *Tc3A* gene, with the single intron precisely removed, was placed under the control of the bacteriophage T7 promoter in the vector pET3c (20). Stop codons were introduced at various positions within *Tc3A*, to make constructs from which only N-terminal portions of Tc3A were expressed. Using an electrophoretic mobility shift assay, we assessed the ability of bacterial extracts containing these Tc3A N-terminal derivatives to bind to Tc3 ends. The radiolabelled DNA fragment used contained the terminal 38 bp of Tc3 as well as the TA sequence which normally flanks Tc3, inserted within the pUC18 polylinker. A control bacterial extract in which no Tc3A derivative was present gave no retardation of the DNA fragment (Figure 1A, lane 5), and neither did an extract from bacteria expressing the N-terminal 54 amino acids of Tc3A ($Tc3A_{1-54}$) (Figure 1A, lane 4). Bacterial extracts containing $Tc3A_{1-98}$, $Tc3A_{1-85}$ and $Tc3A_{1-65}$ efficiently bound and retarded the radiolabelled DNA fragment (Figure 1A, lanes 1−3). As expected, the mobility of the protein−DNA complex decreased as the size of the Tc3A N-terminal derivative was increased. To determine whether this binding is specific for Tc3 terminal sequences, a fragment containing the terminal 28 bp of Tc1 was tested and found not to be bound by Tc3A N-terminal derivatives in this assay (Figure 1B, lanes 1−3). Further proof of the specificity of binding is shown in Figure 1A, lanes 7 and 8. The binding of $Tc3A_{1-65}$ to the radiolabelled Tc3 DNA fragment was competed by the addition of an excess of 40 bp unlabelled double-stranded oligonucleotide representing the 38 terminal bp of Tc3 plus flanking TA, but was not competed by the addition of a double-stranded oligonucleotide representing the terminal sequences of Tc1.

Non-specific DNA binding activity of Tc3A

The first 65 amino acids of Tc3A are responsible for sequence-specific DNA binding, but does Tc3A, like Tc1A (19), also have a non-specific DNA binding activity? To define the region of Tc3A responsible for such an activity the N-terminal deletion derivatives described above and two further deletion derivatives, $Tc3A_{1-159}$ and $Tc3A_{1-192}$, were tested in a South Western assay. Proteins were separated by SDS−PAGE, transferred to a nitrocellulose membrane, renatured and then incubated with radiolabelled pUC18 DNA. Following a wash step to remove any loosely bound DNA, proteins which bound pUC18 DNA were detected by autoradiography. Ponceau S staining confirmed that the proteins were efficiently transferred to and retained on the filter (data not shown). Only full-length Tc3A protein (329 amino acids), $Tc3A_{1-192}$ and, to a lesser extent, $Tc3A_{1-159}$

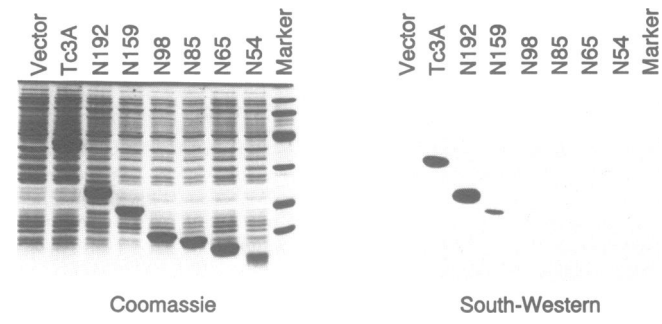

Figure 2. Non-specific DNA binding activity of Tc3A derivatives. Total cell lysates from *E.coli* expressing intact Tc3A (329 amino acids) (lane 2), $Tc3A_{1-192}$ (lane 3), $Tc3A_{1-159}$ (lane 4), $Tc3A_{1-98}$ (lane 5), $Tc3A_{1-85}$ (lane 6), $Tc3A_{1-65}$ (lane 7) and $Tc3A_{1-54}$ (lane 8) were separated by SDS−PAGE on duplicate 12.5% gels. One gel was stained with Coomassie brilliant blue (left panel). Proteins on the other gel were electroblotted to nitrocellulose and probed with random primed ^{32}P-labelled pUC18 DNA. Proteins which were able to bind DNA were detected by autoradiography (right panel). Lane 1 contains a control bacterial extract with no Tc3A derivative. Lane 8 contains markers of the following sizes: 14, 21, 31, 43, 66 and 97 kDa.

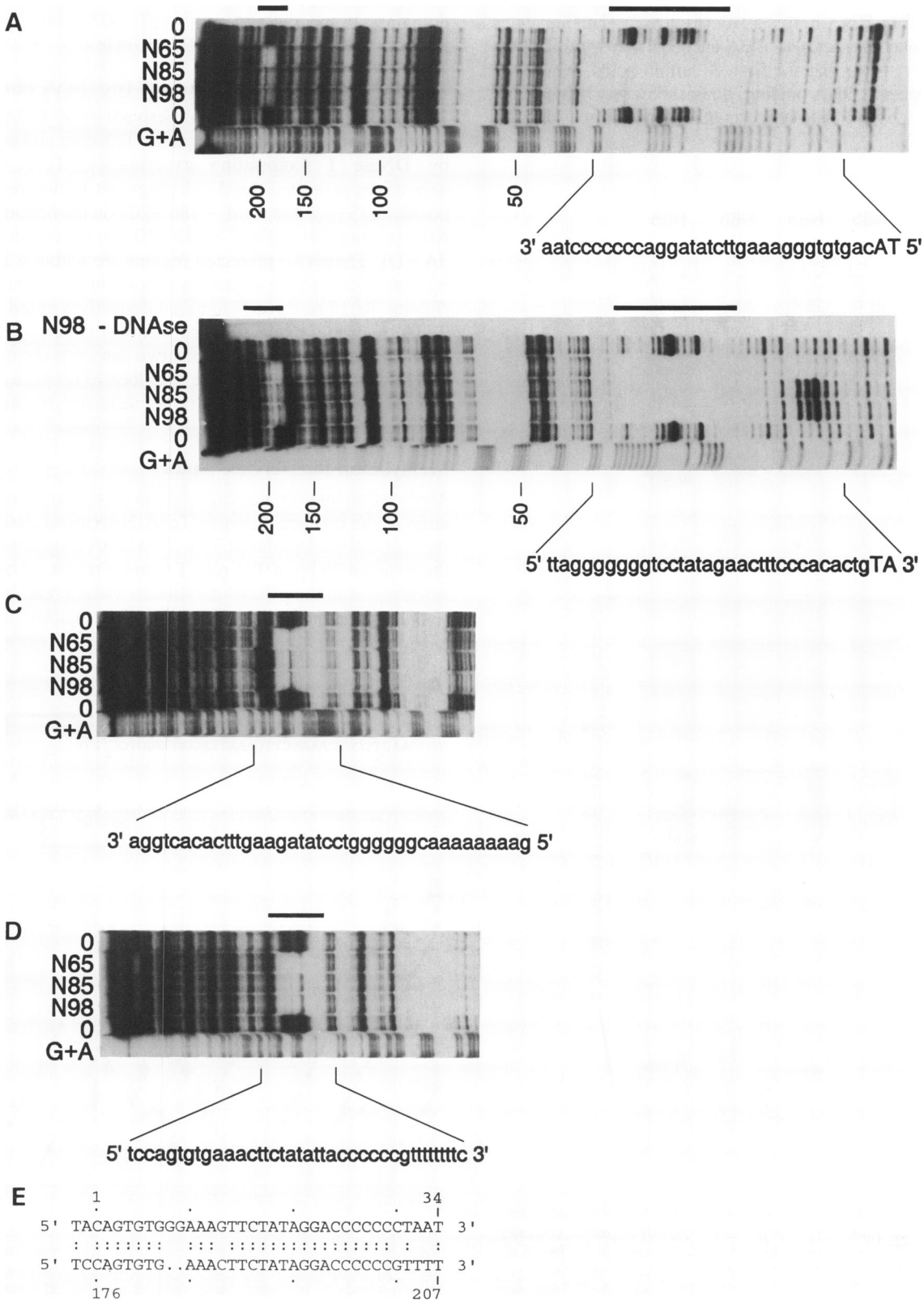

Figure 3. DNase I footprinting of Tc3A−DNA complexes. (**A** and **B**) DNase I footprints were obtained on the 353 bp *Bam*HI−*Pst*I fragment from pSDC324, containing the terminal 315 bp of the Tc3 right end. This DNA fragment was labelled either at the 3′ end (A) or the 5′ end (B) of the *Bam*HI site, 34 nucleotides from the transposon end. Reactions contained 0.4 μl of $Tc3A_{1-98}$, $Tc3A_{1-85}$ or $Tc3A_{1-65}$ or no added protein, as indicated (N98, N85, N65 and 0 respectively). Lane 7 in (A) (N98−DNase) is identical to lane 3 (N98) except that no DNase I has been added to the reaction. Maxam−Gilbert G + A sequencing reactions on the same DNA were run as markers. The sequence of the terminal 33 bp and the position relative to the terminus of Tc3 are indicated to the left of each panel. The two regions of protection are indicated by vertical bars. (**C** and **D**) To examine the protection in the region 186−203 bp from the transposon end, essentially the same DNA fragment as in (A) and (B) was labelled at the other end. The label was either on the 3′ end (C) or on the 5′ end (D). The lanes are marked as in (A) and (B). (**E**) Sequence comparison between the terminal and internal Tc3A binding sites.

could bind pUC18 DNA in this assay (Figure 2). This indicates that the first 159 amino acids of Tc3A encode a non-specific DNA binding activity. Note that the first 98 amino acids, which are sufficient for specific DNA binding, do not show any non-specific DNA binding activity. This suggests that there are two separate domains for specific and non-specific DNA binding located in the N-terminal half of the Tc3A protein.

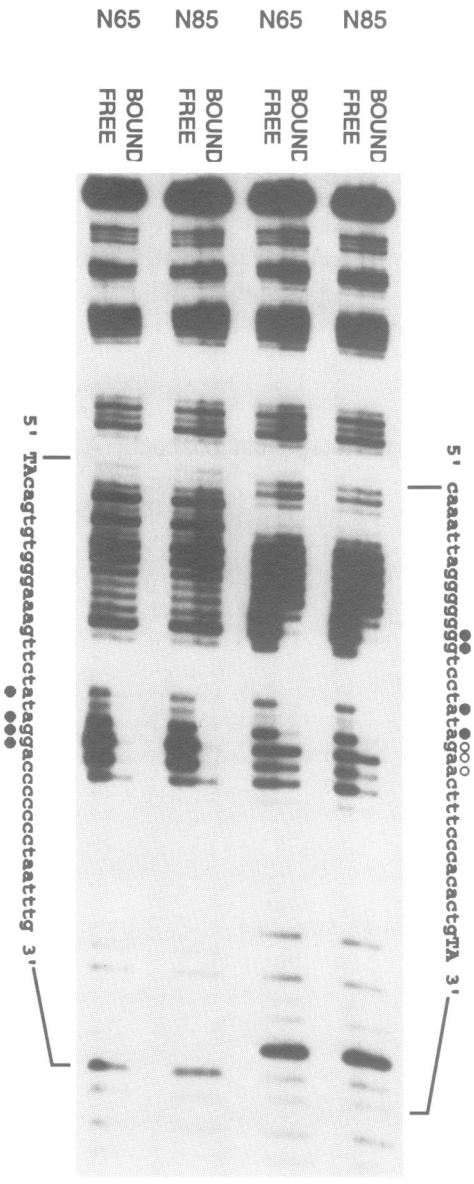

Figure 4. Methylation interference analysis of Tc3A−DNA interactions. Methylation interference was carried out on *Eco*RI−*Hin*dIII fragments, containing the terminal 38 nucleotides plus flanking TA of Tc3 inserted in the *Hin*cII site of the pUC18 polylinker, 3′-end-labelled with [α-^{32}P]dATP at the *Hin*dIII site. The DNA fragment used in lanes 1−4 (from pSDC305) contains the Tc3 sequences oriented with the terminal TA farthest from the labelled *Hin*dIII site; the DNA fragment used in lanes 5−8 (from pSDC304) contains the Tc3 sequences in the opposite orientation. Thus both strands of the Tc3 terminal DNA were analysed. Methylated DNA fragments were incubated with either $Tc3A_{1-65}$ or $Tc3_{1-85}$, as indicated. Bound and unbound DNA were separated on a native polyacrylamide gel, eluted from gel slices, cleaved at G and A positions by boiling in alkali and run on a sequencing gel. Tc3 terminal sequences are indicated at the side. Positions at which methylation interferes strongly with binding are shown by filled circles, positions with weaker interference are shown by open circles.

The Tc3 inverted repeat contains two Tc3A binding sites

The binding of the N-terminal derivatives of Tc3A to a DNA fragment containing 315 bp from the Tc3 right end was analyzed by DNase I footprinting experiments. Tc3A N-terminal derivatives protected this DNA fragment in two regions, from positions 13−32 and from ~186−203 on the bottom strand and from positions 9−28 and ~182−201 on the top strand (Figure 3A−D). These two protected regions are within a 26 out of 29 bp match between the sequences at positions 1−29 and 176−202 (Figure 3E). The three N-terminal derivatives tested, $Tc3A_{1-65}$, $Tc3A_{1-85}$ and $Tc3A_{1-98}$, gave identical patterns of protection from DNase I, except that the two larger proteins also gave several DNase I hypersensitive sites located between the terminal protected region and the TA target sequence on the top strand (Figure 3A and 5A).

Methylation interference studies

Methylation interference experiments were carried out to further analyze the contacts between Tc3A N-terminal derivatives and the terminal nucleotides of Tc3. Two DNA fragments containing the terminal 38 bp of Tc3 plus TA target repeat, in both possible orientations in the pUC18 polylinker, were 3′ end labelled with [α-^{32}P]dATP at the *Bam*HI site and methylated using DMS.

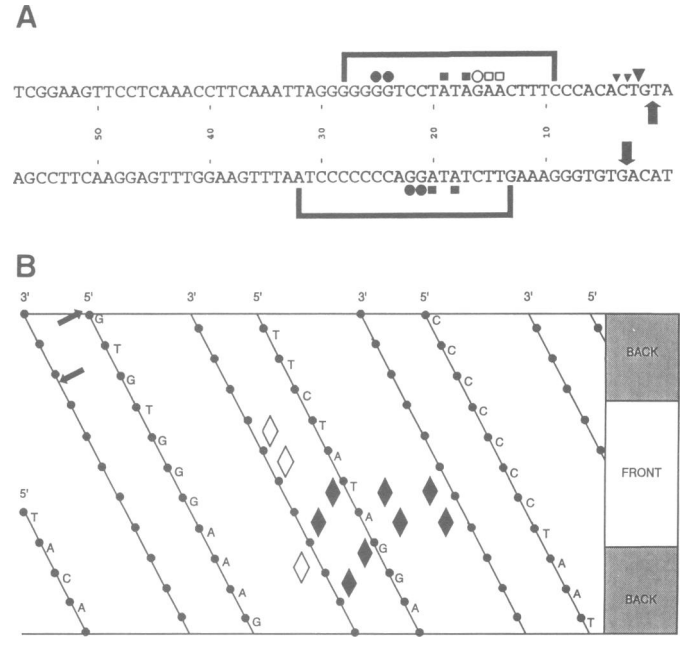

Figure 5. Schematic representation of the footprinting and methylation interference data for the terminal bp of Tc3. (**A**) The extent of the DNase I footprint is shown on both strands by horizontal bars; positions of enhanced cutting by DNase I in the presence of $Tc3A_{1-98}$ and $Tc3A_{1-85}$ are indicated by filled triangles. Methylation interference is indicated by circles (G positions) and squares (A positions). Filled symbols indicate strong interference, open symbols indicate weaker interference. The sites at which Tc3A is presumed to cut the Tc3 ends are shown by arrows. (**B**) Methylation interference results shown on a planar projection of the DNA helix. The sequence of one strand of the DNA is shown. Methylated bases which interfere with transposase binding are shown as diamonds. Filled diamonds indicate strongest interference. Tc3 excision sites are indicated by arrows

Following incubation with $Tc3A_{1-65}$ and $Tc3A_{1-85}$, bound and free DNA were separated on a native gel, eluted from gel slices and then cleaved at the methylated positions by boiling in alkali. Cleavage products were separated on a sequencing gel and detected by autoradiography (Figure 4). Methylation of several different guanine and adenine residues, on both strands within the region covered by the DNase I footprint, interfered with binding by both $Tc3A_{1-65}$ and $Tc3A_{1-85}$, giving a very weak band in the bound fraction and a band of increased intensity in the unbound fraction. The interference patterns produced by $Tc3A_{1-65}$ and $Tc3A_{1-85}$, which were essentially identical, are summarized together with the DNase I footprinting data in Figure 5A.

Since guanine is methylated by DMS at the N-7 position in the major groove and adenine is methylated at N-3 in the minor groove, the methylation interference data suggest that the N-terminal portion of Tc3A makes contacts with residues in neighbouring major and minor grooves. These positions cluster together on one face of the DNA helix, as can be seen when the methylation interference data is drawn on a helical projection of the DNA (Figure 5B). In agreement with these footprinting results is the observation that positions 10–26 of the terminal inverted repeat of Tc3 are sufficient for recognition by the N-terminal DNA binding domain of Tc3A: an oligonucleotide representing positions 10–26 of the Tc3 inverted repeat, containing all of the bases at which methylation interferes with binding, when inserted into the pUC18 polylinker, was bound efficiently by $Tc3A_{1-65}$ (Figure 1B, last three lanes).

DISCUSSION

We have shown that the transposase of Tc3 has two distinct DNA binding activities: a non-specific DNA binding activity located within the first 159 amino acids of Tc3A and a sequence-specific DNA binding activity located within the first 65 amino acids of Tc3A. These results are similar to results previously found for the transposase of Tc1 (19). The first 63 amino acids of Tc1A are sufficient to bind specifically to Tc1 ends, while amino acids 71–207 of Tc1A contain the non-specific DNA binding activity of Tc1A. Furthermore, we have mapped the Tc3 transposase binding sites in the inverted repeat. It should be noted that the binding experiments were performed in an *in vitro* system using naked, linear DNA. The higher order structure of the DNA as it is *in vivo* could affect the binding of Tc3A.

The N-terminal DNA binding domains of Tc1 and Tc3 transposases show no obvious homology, either to each other or to any well-known DNA binding motif. We note generally that there is little conservation between N-terminal regions of different Tc1/mariner transposase proteins and also no apparent conservation of the inverted repeat sequence of these elements apart from the terminal four nucleotides. It seems likely that N-terminal regions of each transposase of this family specifically recognize the sequence of the terminal sequences of its own transposon, but that these sequences diverge between different transposons so that the transposase protein of a given element will only act at its own ends. The catalysis of the transposition reaction is probably more conserved among the different transposons using the conserved terminal 4 bp together with the DDE motif of the transposase protein. The DDE motif is conserved among transposases and integrases (24–26) and has been shown to be important for Tc3A activity (18).

The sequence of the terminal 4 bp is highly conserved among the Tc1/mariner class of elements, but these 4 bp are not required for recognition by the N-terminal DNA binding domains of Tc3 or Tc1 transposases nor are they protected from DNase I by these DNA binding domains (19; this paper). This suggests that the terminal 4 bp are important for steps in the transposition reaction that occur after the initial recognition of ends by transposase, and may involve recognition by the more highly conserved central regions of the transposase. This division of the inverted repeat into a terminal sequence important for catalytic steps and a more internal sequence needed for initial protein binding has also been observed for other transposons (IS903, IS1, IS10, Tn10, γδ, Tn3 and Mu) (27–34). The non-specific DNA binding activity that we observed might be involved in interactions between the transposase protein and target DNA, required for integration specifically into the sequence TA.

Tc3 is unique among the Tc1/mariner elements because it has extremely long, almost perfect inverted repeats. We have shown by DNase I footprinting that Tc3A binds to two almost identical sequences in the Tc3 inverted repeat: one located between positions 1 and 32 and the other located ~180 bp from the transposon end. It is not uncommon for transposon ends to contain multiple transposase binding sites, for example Mu has three MuA binding sites at each end (35) and many plant transposons have long arrays of transposase binding sites (36,37). Recent experiments indicate that the internal transposase binding sites are not required for transposition (H.G.A.M.van Luenen and R.H.A.Plasterk, unpublished observation). This raises two questions: why does Tc3 have such long, perfect, inverted repeats (462 bp) and why does each repeat contain two binding sites for transposase? One possibility is that the inverted repeats may have arisen by some aberrant transposition or recombination event and that Tc3 originated so recently that the two inverted repeats have not yet diverged. Alternatively the long inverted repeats of Tc3 may be maintained by continuing gene conversion events between different copies of Tc3 or by selection acting on the Tc3 sequence. We will further investigate the *cis* requirements for Tc3 transposition by changing the structure of the inverted repeats.

The major difference between the sequence at positions 1–32 and the sequence found 180 bp from the transposon ends is that the latter has a 2 bp deletion between the sequence 5′-CAGTG-3′ and the Tc3A binding site. This may explain why Tc3A cannot utilize the internal sites for transposition. This suggests that the helical phasing between the Tc3A binding site and the cleavage sites, on the opposite side of the helix, is critical for the formation of an active synaptic complex.

ACKNOWLEDGEMENTS

We thank P.Anderson for the plasmid TR # 10 from which all Tc3 sequences were derived. We thank P.Borstand and J.C.Vos for critical reading of the manuscript. We thank J.van Eerden for technical assistance. This work was supported by a Pionier grant of the Netherlands Science Foundation NWO to R.H.A.P. and a long-term fellowship from the European Molecular Biology Foundation to S.D.C.

REFERENCES

1. Harris,L.J., Baillie,D.L. and Rose,A.M. (1988) Nucleic Acids Res., **16**, 5991–5998.
2. Tausta,S.L. and Klobutcher,L.A. (1989) Cell, **59**, 1019–1026.

3. Harris,L.J., Prasad,S.S. and Rose,A.M. (1990) J. Mol. Evol., *30*, 359–369.
4. Brezinsky,L., Wang,G.V.L., Humphreys,T. and Hunt,J. (1990) Nucleic Acids Res., **18**, 2053–2059.
5. Franz,G. and Savakis,C. (1991) Nucleic Acids Res., **19**, 6646.
6. Abad,P., Quiles,C., Tares,S., Piotte,C., Castagnone-Sereno,P., Abadon,M. and Dalmasso,A. (1991) J. Mol. Evol., **33**, 251–258.
7. Henikoff,S. (1992) New Biologist, **4**, 382–388.
8. Heierhorst,J., Lederis,K. and Richter,D. (1992) Proc. Natl. Acad. Sci. USA, **89**, 6798–6802.
9. Robertson,H.M., Lampe,D.J. and Macleod,E.G. (1992) Nucleic Acids Res., **20**, 6409.
10. Tudor,M., Lobocka,M., Goodell,M., Pettitt,J. and O'Hare,K. (1992) Mol. Gen. Genet., **232**, 126–134.
11. Daboussi,M., Langin,T. and Brygoo,Y. (1992) Mol. Gen. Genet., **232**, 12–16.
12. Robertson,H.M. (1993) Nature, **362**, 241–245.
13. Garcia-Fernandez,J., Marfany,G., Baguna,J. and Salo,E. (1993) Nature, **364**, 109–110.
14. Caizzi,R., Caggese,C. and Pimpinelli,S. (1993) Genetics, **133**, 335–345.
15. Radice,A.R., Bugaj,B., Fitch,D.H.A. and Emmons,S.W. (1994) Mol. Gen. Genet., in press.
16. Collins,J., Forbes,E. and Anderson,P. (1989) Genetics, **121**, 47–55.
17. Van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1993) EMBO J., **12**, 2513–2520.
18. Van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1994) Cell, **79**, 293–301.
19. Vos,J.C., Van Luenen,H.G.A.M. and Plasterk,R.H.A. (1993) Genes Dev., **7**, 1244–1253.
20. Rosenberg,A.H., Lade,B.N., Chui,D., Lin,S., Dunn,J.J. and Studier,F.W. (1987) Gene, **56**, 125–135.
21. Stanssens,P., Opsomer,C., McKeown,Y.M., Kramer,W., Zabeau,M. and Fritz,H. (1989) Nucleic Acids Res., **17**, 4441–4454.
22. Krogstad,P.A. and Champoux,J.J. (1990) J. Virol., **64**, 2796–2801.
23. Feinberger,A.P. and Vogelstein,B. (1983) Anal. Biochem., **132**, 6–13.
24. Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G. (1994) Proc. Natl. Acad. Sci. USA, **91**, 942–946.
25. Fayet,O., Ramond,P., Polard,P., Prére,M.F. and Chandler,M. (1990) Mol. Microbiol., **4**, 1771–1777.
26. Khan,E., Mack,J.P.G., Katz,R.A., Kulkosky,J. and Skalka,A.M. (1990) Nucleic Acids Res., **19**, 851–860.
27. Huisman,O., Errada,P.R., Signon,L. and Kleckner,N. (1989) EMBO J., **8**, 2101–2109.
28. Zerbib,D., Prentki,P., Gamas,P., Freund,E., Galas,D.J. and Chandler,M. (1990) Mol. Microbiol., **4**, 1477–1486.
29. Derbyshire,K.M. and Grindley,N.D.F. (1992) EMBO J., **11**, 3449–3455.
30. Kans,J.A. and Casadaban,M.J. (1989) J. Bacteriol., **171**, 1904–1914.
31. Amemura,J., Ichikawa,H. and Ohtsubo,E. (1990) Gene, **88**, 21–24.
32. Waiter,L.A. and Grindley,N.D.F. (1988) EMBO J., **7**, 1907–1911.
33. Haniford,D. and Kleckner,N. (1994) EMBO J., **13**, 3401–3411.
34. Mizuuchi,K. (1992) Annu. Rev. Biochem., **61**, 1011–1051.
35. Craigie,R., Mizuuchi,M. and Mizuuchi,K. (1984) Cell, **39**, 387–394.
36. Kunze,R. and Starlinger,P. (1989) EMBO J., **8**, 3177–3185.
37. Gierl,A., Lutticke,S. and Saedler,H. (1988) EMBO J., **7**, 4045–4053.