Published online August 10, 2004

4306–4312 *Nucleic Acids Research, 2004, Vol. 32, No. 14*
doi:10.1093/nar/gkh770

Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA

Stephan Watkins, Gertie van Pouderoyen and Titia K. Sixma*

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

Received June 22, 2004; Revised and Accepted July 27, 2004

ABSTRACT

The bipartite DNA-binding domain of Tc3 transposase, Tc3A, was crystallized in complex with its transposon recognition sequence. In the structure the two DNA-binding domains form structurally related helix–turn–helix (HTH) motifs. They both bind to the major groove on a single DNA oligomer, separated by a linker that interacts closely with the minor groove. The structure resembles that of the transcription factor Pax6 DNA-binding domain, but the relative orientation of the HTH-domain is different. The DNA conformation is distorted, characterized by local narrowing of the minor groove and bends at both ends. The protein–DNA recognition takes place through base and backbone contacts, as well as shape-recognition of the distortions in the DNA. Charged interactions are primarily found in the N-terminal domain and the linker indicating that these may form the initial contact area. Two independent dimer interfaces could be relevant for bringing together transposon ends and for binding to a direct repeat site in the transposon end. In contrast to the Tn5 synaptic complex, the two Tc3A DNA-binding domains bind to a single Tc3 transposon end.

INTRODUCTION

Tc1/*mariner* and other transposon families are widespread in animals (1–4). Within the invertebrates, the Tc1/*mariners* are most prevalent, both in genomic copy number and number of different transposons within the same organism (5). The *Caenorhabditis elegans* genome contains six types of active transposons, of which Tc3 is one (6). The Tc1/*mariner* transposons encode a single protein, the transposase, and are flanked at either end by inverted terminal repeats (ITRs). These ITRs contain one or two binding sites for the transposase (2). The transposase is capable of performing the entire transposition reaction *in vitro* (7,8). It has a bipartite DNA-binding domain, which binds to either end of the transposon through base-specific recognition of the ITRs by its N-terminal region (9). This positions the catalytic domain of the transposase, such that cleavage of the transposon can occur at specific TA sites at the end of each ITR (see Figure 1). Biochemical evidence suggests that dimerization of transposase molecules is needed to completely cleave the ITR ends (10). Dimerization is also necessary to bring the two ends of the transposon DNA together. In a semi-random fashion, this DNA–protein dimer then inserts into a TA site (6,11,12). The insertion process shows weak specificity for certain TA sites dependent on nucleotides immediately 4 bp upstream of this target site (6).

The Tc1/*mariner* family encodes closely related transposases. These proteins are composed of three domains, the first two making up a bipartite DNA-binding domain, and the third a catalytic core domain containing a DDE motif found in many Mg^{2+}-dependent catalytic enzymes (see Figure 1) (13,14). The bipartite DNA-binding domain studied in this paper is thought to be responsible for both the specific recognition of the ITRs and the stabilization of the catalytic domain against the DNA substrate.

Bipartite DNA-binding domains occur across a large number of protein families, including the Tc1/*mariner* family and the structurally related Pax and Prd families of transcription domains (15). A well-studied example is the POU-family DNA-binding domains (16). Although overall amino acid sequences vary greatly, even within families, the secondary structures are conserved both in shape and function. All bipartite DNA-binding regions of proteins are characterized by having two domains separated by a linear stretch of amino acids. In most bipartite DNA-binding structures, the N-terminal domain is responsible for the recognition of specific DNA sequences, such as the ITRs for Tc1/*mariner* and other transposons. Within the Tc1/*mariner* family, bipartite DNA-binding regions also function to stabilize the whole protein against the DNA (14) allowing the DDE catalytic site to come into contact with the specific cleavage site.

We have solved the structure of the N-terminal DNA recognition domain of Tc3A in complex with the recognition region of the transposon ITR previously (17). This structure showed that DNA recognition occurs through a combination of base-specific contacts and DNA-shape recognition. A putative protein dimerization was seen between the N-terminal domains, which could be important for bringing the two ends of the transposon together. Here, we extended our study of the Tc3 transposase DNA recognition to the full bipartite DNA-binding domain

*To whom correspondence should be addressed. Tel: +31 20 5121959; Fax: +31 20 5121954; Email: t.sixma@nki.nl
Present address:
Gertie van Pouderoyen, Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

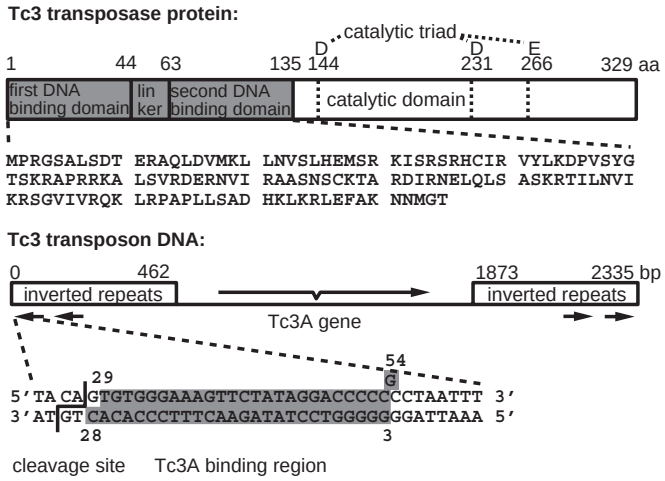

Figure 1. Schematic representation of Tc3 transposase protein (Tc3A) and Tc3 transposon DNA. Gray boxes indicate which part of the protein and DNA were co-crystallized in this study. The DNA numbering is indicated. G54 was used instead of the naturally occurring C to create an overhang that can form a base-pair with C28. The arrows under the inverted repeats of the DNA indicate the two almost identical binding sites of Tc3A separated by ~180 bp at each transposon end (9).

complexed to a longer fragment of the ITR. By using this X-ray structure, we get insight into the arrangement of this domain on DNA and the relative orientation of the two domains. Comparison with other bipartite structures shows similarities in structure, but differences in recognition and complex formation with DNA.

MATERIALS AND METHODS

Plasmid pRP-1442 (made by George Verlaan), containing the Tc3(1–135) bipartite domain and a C-terminal 6 His-tag, was transfected into *Escherichia coli* BL21 DE3 cells and plated on carbenicillin plates. Single colonies were selected and grown in Luria–Bertani (LB) media overnight. An aliquot of 40 ml of overnight culture was used to inoculate one liter of LB media, and cultures were grown to an OD_{600} of 0.6 in ~4 h, at 37°C. Cultures were induced with 0.8 mM isopropyl-β-D-thiogalactopyranoside and grown for an additional 2 h. The cells were harvested and the protein was shown to be expressed solely in inclusion bodies.

Cells were lysed by flash freezing and solubilized in B-PER protein extraction buffer containing 0.25 mg/ml of lysozyme, 0.25 mg/ml DNase T and 0.25 mg/ml RNase. Inclusion bodies were purified by three subsequent washes in detergent buffer containing 20 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, 0.5% Triton X-100 and 0.5% NP-40, pH 7.6, followed by three washes in 3 M urea. Inclusion bodies were solubilized in 6 M urea, 100 mM NaCl, 10 mM 2-mercaptoethanol and 10 mM Tris, pH 8.0.

Purification of Tc3A(1–135) was done by Talon Bead© affinity chromatography, using a 40 mM imidazol wash with solubilization buffer, followed by elution with the same buffer containing 200 mM imidazol. Eluate containing the bipartite domain was then dialyzed against 20 mM sodium acetate, pH 5.5, 150 mM NaCl, 5% glycerol and 2 mM DTT with three subsequent buffer exchanges of 1:1000 for 4 h, 6 h and overnight. Dialyzed sample was centrifuged at 12 000 r.p.m. and the supernatant was loaded onto a Mono-S (Pharmacia) column and eluted with a 0–2 M NaCl gradient in dialysis buffer. Proteins were eluted at ~0.4 M NaCl, and ran as a single band on SDS–PAGE. Protein was concentrated to ~12 mg/ml using Centricon and Centriprep concentrators with a molecular weight cutoff of 10 kDa. High-performance liquid chromatography purified DNA oligos were purchased from Sigma-Genosys.

Protein–DNA complex was made by diluting equimolar samples with 20 mM Tris, pH 7.6, 5% glycerol, 1 mM EDTA, 2 mM DTT and 150 mM NaCl to a final concentration of 0.4 mM. The protein and DNA solutions were placed on ice for 15 min and then mixed rapidly and placed on ice for an additional 30 min. The complex solution was then concentrated to 5 mg/ml at 4°C using a Centriprep concentrator with a molecular weight cutoff of 10 kDa.

Crystals were grown by the hanging drop vapor diffusion method. The drops were made of equal amounts of complex solution and well solution (0.6 M sodium acetate, 10 mM DTT and 5% (v/v) glycerol, pH 5.5). Hexagonal crystals grew within 1 week with dimensions of 0.1–0.15 mm. The crystals were flash frozen in a cryo-buffer containing 30% glycerol and well solution. Crystals diffracted to 3.0 Å. Increased resolution was achieved by slowly raising the glycerol concentration in the well by 5% increments every 24 h to a final concentration of 35%. These crystals were frozen in liquid nitrogen directly from the drop, diffracting to 2.7 Å.

Data were collected at the Hamburg, EMBL outstation on beam line BW7A (λ = 1.0749 Å). Crystals belong to space group $P6_122$, with cell dimensions of $a = b = 93.7$ Å and $c = 255.6$ Å. Data were integrated using DENZO and SCALEPACK (18). The overall *B*-factor of the data is 101. Initial phases were found by molecular replacement with the N-terminal domain as model (17) using AMORE with a *R*-value of 0.47. Owing to the high solvent content (~77%) of the crystals, improvement on the electron density using the programs Arp/Warp (19) and Refmac5 (20,21) of CCP4 was very effective. The structure was built in O (22) and refined in Refmac5 using TLS parameters and bulk solvent flattening. The final *R*-factor was 23.3% and the *R*-free was 27.3% (Table 1). Structural comparisons and analysis were done in O, Whatif (23) and GRASP (24). Structural alignments were made in O. DNA analysis was done visually in O and using the program 3DNA (25) to determine the shifts and overall DNA geometry.

RESULTS AND DISCUSSION

General description of the protein–DNA complex

The crystal structure of the Tc3 bipartite domain (1–135), complexed to a 26mer double-stranded DNA oligomer, was determined at 2.7 Å resolution. The parts of the transposase and transposon used in co-crystallization are indicated by gray boxes in Figure 1. The structure was solved by molecular replacement using the N-terminal domain structure [PDB code 1TCA (17)] as a model. The structure reveals two helix–turn–helix (HTH) protein domains separated by a linker of 18 amino acids, binding to the 26mer duplex of DNA, as shown in Figure 2. All DNA bases and amino acid residues 2–103 were defined in electron density, but electron density

Table 1. Crystallographic data and refinement statistics

Crystallographic data	
R merge (%)	9.2
$I/\sigma(I)$ (last shell)	20 (2.0)
Resolution limits (Å) (last shell)	20–2.7 (2.8–2.7)
Solvent content	77%
B-factor	101
Completeness (%) (last shell)	98.5 (98.9)
Refinement statistics	
Resolutions limits (Å)	12–2.7
Number of reflections	17820
R-factor (%)	23.3
Free *R*-factor (%)	27.3
r.m.s. deviation bonds (Å)	0.011
r.m.s. deviation angles (°)	1.90
Number of protein atoms	816
Number of DNA atoms	1061
Number of solvent atoms	7

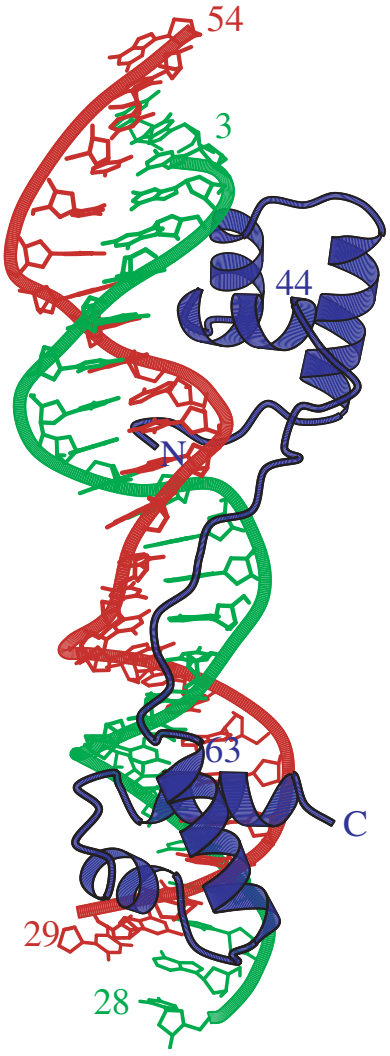

Figure 2. Tc3A(1–135) bound to transposon DNA. Tc3A(1–135) is shown schematically with ribbons drawn through the Cα atoms in blue. The N-terminal domain contains three α-helices (9–20, 25–32 and 36–44) and the C-terminal domain as well (63–75, 80–86 and 93–102). The DNA strands are shown schematically as ribbons through the phosphate backbone and the bases are shown as sticks (green and red).

for amino acid residues 1 and 104–135, and the C-terminal His-tag is missing. It is probable that the N-terminal methionine was removed from the protein, as shown for the earlier construct (17). The C-terminal residues are not visible in the electron density, probably due to the flexibility of these amino acids in the absence of the catalytic domain. The crystallographic data have a very high *B*-factor, probably due to the high solvent content (77%), which creates relatively weak packing interactions. The limited resolution, the high *B*-factor of the data and the missing C-terminal amino acids result in a relatively high *R*-factor (23.3%) and free *R*-factor (27.3%).

The Tc3A bipartite domain binding to DNA

The HTH fold of the N-terminal domain (residues 2–44) and of the C-terminal domain (residues 62–104) contain three α-helices each. The two domains can be superimposed with an r.m.s. difference of 2.0 Å using Cα atoms (6–11, 12–34, 35–46 on 60–65, 67–89, 92–103), showing the similarity. The two HTH domains are bound separately to the DNA oligomer, both binding in the major groove on opposite sides of the DNA. The linker crosses the DNA backbone, from the major to the minor groove, with 12 of its residues positioned between and parallel to the phosphate backbones of the minor groove.

The N-terminal domain shows more base-pair specificity in its DNA-recognition than the C-terminal domain (Figure 3A). As shown in the previous N-terminal structure, base-specific interactions are made by Arg-36 and His-37 within the major groove (11,17). A previously observed water-mediated contact of Arg-40 is not visible at this resolution. These interactions all require a narrow groove, provided by the observed change in the oligo to a more A-DNA form (26). There are also 10 hydrogen bonds and 7 salt bridges, forming mostly along the phosphate backbone, and the interface with DNA buries ∼960 $Å^2$ protein surface. These contacts are almost all preserved in this structure.

The 18 amino acid linker region (residues 45–62) connects the last helix of the N-terminal domain with the first helix of the C-terminal domain (Figure 3B). Residues 49–60 are close to the DNA. The DNA contact of the linker buries ∼600 $Å^2$ protein surface area. The linker backbone runs parallel to the minor groove. Most of the interactions of the linker with the DNA are non-specific van der Waals interactions, but several hydrogen bonds and salt bridges occur as indicated in Figure 3B. Some are with the phosphate backbone and some with the base pairs. Pyrimidine-specific contacts are made by Arg-54 NE to T42 O2 and Arg-57 NH1 to T39 O2. A purine-specific contact is present between Arg-54 NH2 and A43 N3. In the center of the linker there are five positively charged residues (Lys-53, Arg-54, Arg-57, Arg-58 and Lys-59) compensating the negative charge of the DNA phosphate backbones. The linker region spans approximately half a turn of the DNA.

In the C-terminal domain, two positively charged residues make purine-specific contacts. Lys-93 NZ hydrogen bonds with G32 N7 and Arg-94 NE and NH2 hydrogen bond with G34 O6 and N7, respectively. There are five contacts between the protein and the phosphate backbone: Ala-80 N with G32, Ser-92 N and OG with T20, Thr-95 OG1 with T20 and Arg-102 NH2 with T19 (see Figure 3C). The DNA contact of the

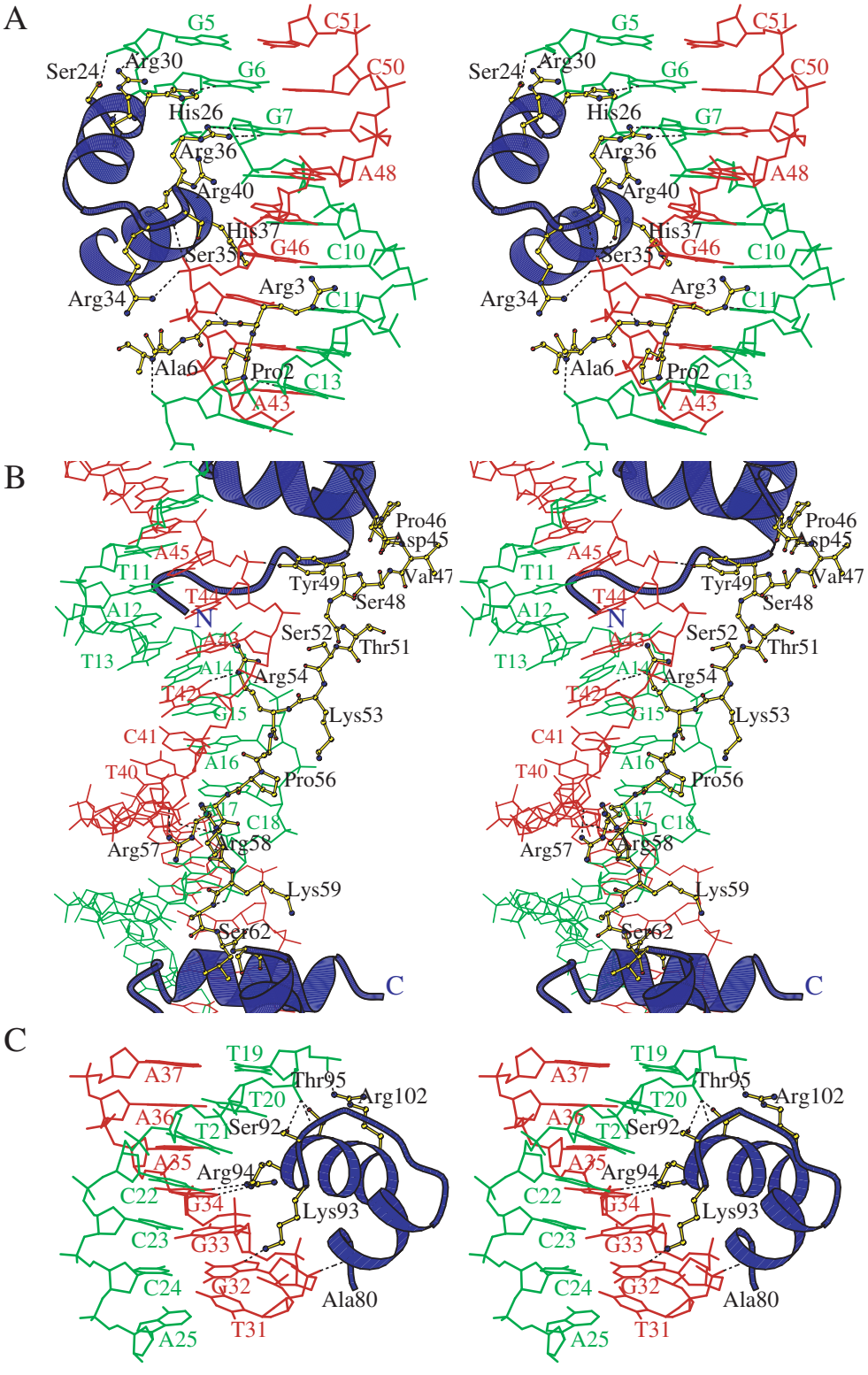

Figure 3. Stereo view of the protein–DNA contacts. (**A**) N-terminal domain–DNA contacts. Hydrogen and salt bridges (Pro-2 N to T13 O2, Arg-3 N and NE to A45 N3 and T11 O2, respectively, Gly-4 N to A45 O4′, Ala-6 N to G15 OP1, Ser-24 OG to G5 OP1, Leu-25 N to G6 OP1, His-26 N and NE2 to G6 OP2 and N7, respectively, Arg-30 NH2 to G5 OP2, Arg-34 NH1 to G46 OP1, Ser-35 N and OG to G46 OP2, Arg-36 NH1 and NH2 to G7 N7 and O6, respectively, and His-37 ND1 to G46 N7) are indicated with dotted lines. (**B**) The linker–DNA contacts. Hydrogen bonds and salt bridges (Tyr-49 OH to A45 OP1, Arg-54 NE to T42 O2, Arg-54 NH2 to A43 N3, Arg-57 NH1 to T39 O2, Arg-58 N to C41 OP1 and Ala-60 N to T19 OP1) are indicated with dotted lines. (**C**) C-terminal domain–DNA contacts. Hydrogen bonds and salt bridges (Ala-80 N to G32 OP1, Ser-92 N and OG to T20 OP2, Lys-93 NZ to G32 N7, Arg-94 NE and NH2 to G34 O6 and N7, respectively, Thr-95 OG1 to T20 OP2 and Arg-102 NH2 to T19 OP2) are indicated with dotted lines.

C-terminal domain buries ~500 Å^2 protein surface area. Upon conversion of the TLS parameters to individual *B*-factors, the N-terminal domain (B_{ave} = 70) is better defined than the C-terminal domain (B_{ave} = 117). These differences are due to domain movements, since they are absent after TLS refinement, and the remaining individual *B*-factors on the atoms are around 20 in all domains. The relatively high *B*-factors and the lower number of interactions for the C-terminal domain agree with previous experiments in which the N-terminal domain binds ITR DNA tightly, while the independent interaction of the C-terminal domain with ITR DNA is not measurable (14). It seems likely that the initial recognition between protein and DNA is dependent on the electrostatic complementarity of the N-terminal domain with the DNA, and later other regions start to interact.

The DNA structure

In the structure the 26mer DNA is substantially distorted from B-DNA. Compared to the previous structure of the N-terminal domain alone, the oligomer is 2 bp shorter at one end, and extends 8 bp further towards the transposon end, terminating 2 bp before the actual cleavage site (Figure 1). The common 17 bp have essentially the same conformation as was seen in the Tc3A N-terminal domain complex. There is a switch to A-DNA in a GC stretch (26) and a narrowing of the minor groove in a TATA region. In this region, a large twist and negative roll are observed. The similarity in DNA conformational parameters is striking, since the packing in the crystal is completely different in the two complexes. We conclude that the conformation observed is likely to be a feature of the Tc3A–DNA complex, and also that the binding of the second DNA-binding domain does not affect the conformation in this region.

The structure contains 8 bp that were not present in the previous structure and these show additional distortions. Their most striking feature is an extremely narrow minor groove in a region that contains three consecutive TA base-pairs. The DNA base-pairs show a negative roll in this region, with a large twist in the transition to three subsequent CG base-pairs. Each double-strand has a 3′C and a 5′G overhang, that stack end-to-end in the crystal, packing to form nearly perfect GC base-pairing. The 26mer DNA shows the overall bends at either end of the oligomer, resulting in a meandering multimer. Thus, the local distortions of the DNA do not add up to major global displacement from a straight helix.

Multimer formation

It is likely that similar to other transposases, the Tc1/*mariner* proteins function as dimers, since this helps to coordinate cleavage at either end of the transposon. Within the crystal structure two distinct dimeric interfaces are observed. The dimer interfaces occur between different symmetry related molecules, in both the N- and C-terminal domains (Figure 4). In one dimer interface the N-terminal helices of two symmetry-related molecules interact, shown in the previous structure of the N-terminal domain, and conserved in this crystal form. The C-terminal domains of an alternate set of symmetry-related molecules interact, starting at Arg-58. Comparisons of the respective buried surfaces of the two dimeric interfaces show that the N- and the C-terminal dimer bury a similar amount of the accessible surface (~500 Å^2 on each

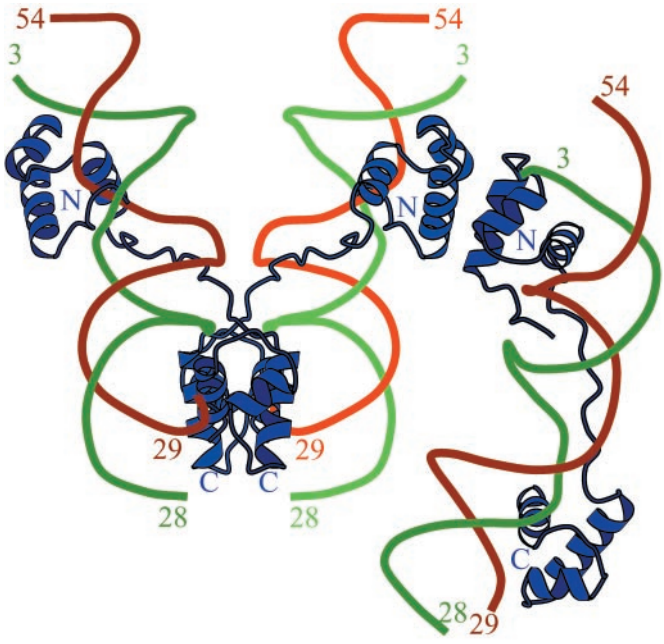

Figure 4. View of the dimerization of the N- and C-terminal domains. Tc3A(1–135) is shown schematically with ribbons drawn through the Cα atoms in blue. The DNA strands are shown schematically as ribbons through the phosphate backbone (green and red).

surface). The C-terminal dimer produces only two symmetrical hydrogen bonds formed between the guanidinium group of Arg-58 and the carbonyl oxygen of Pro-56 on the symmetry-related molecule found in the linker region, while the N-terminal dimer has only one set of symmetric hydrogen bonds, as described previously (17). Thus both interfaces are relatively hydrophobic and at this stage the crystallographic data alone do not resolve whether they are relevant for *in vivo* function. However, the two interfaces interact with different molecules of DNA. It is conceivable that Tc3A works as a tetramer in the synaptic complex, in which two molecules are involved in binding each end of the ITR, and that two additional molecules bind to the almost identical internal binding site ~180 bp from the transposon end. This fits well with the data on the Sleeping Beauty transposon, which was shown to form tetramers, and to bind to the direct repeat regions (27). Such a multimerization step could be the mechanism through which these direct repeats affect the transposition rates (27).

Comparison to other bipartite DNA-binding domains

A comparison to known protein structures using the program DALI (28) shows that the Pax6 bipartite DNA-binding domain is very similar in structure. Structurally Tc3A belongs to the SCOP family of recombinase DNA-binding domains (29), and in particular to the paired domains, currently consists of Prd (30), Pax6 (31) and structurally similar homolog Pax5 (32), that each show a duplication of the HTH domain. Comparisons with Pax6 and Prd bipartite structures show the highest degree of similarity within the N-terminal regions, as described previously (30,31). Pax6 and Prd contain longer loops between the respective helixes in the HTH motifs, especially in the C-terminal domains. There is also an additional loop preceding

the N-terminal HTH motifs of Pax6 and Prd, not found in Tc3A. None of the other structures forms homo-dimers as Tc3A does, and instead bind to regulatory proteins from other families. This may account for some of the differences observed. However, the N- and C-terminal domains of Tc3A, Pax6 and Prd have different relative orientations (Figure 5).

Overall differences in the structures are also apparent in the mechanism of DNA recognition. Both the Pax6 and Prd bipartite domains bind to the same side of the DNA, with the linker region of Pax6 filling in less of the minor groove between the domains. This is surprising, as the Pax6 linker region contains two additional residues. The Tc3 bipartite structure has the N- and C-terminal HTH regions positioned on opposing sides of the DNA oligo. Similar rearrangements are seen in the POU family of bipartite DNA-binding domains (16), but these have

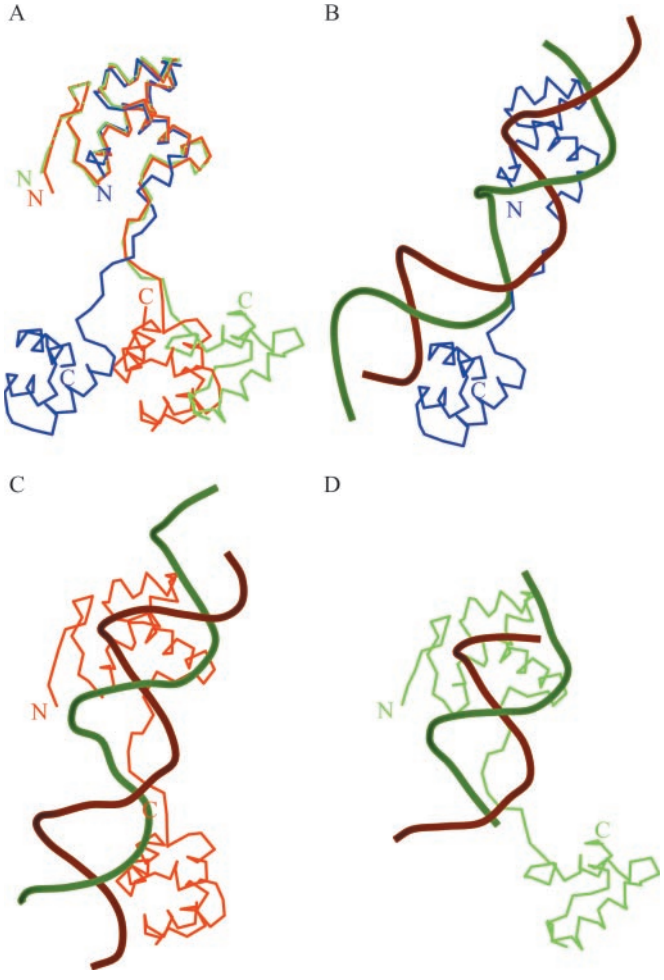

Figure 5. (**A**) Superposition of the N-terminal domains of Tc3A (blue), Pax6 (red) and Prd (green) shows that the different relative orientations of the C-terminal domains. The N- and C-termini are indicated with N and C, respectively. Superposition of the full N-terminal domains of Pax6 (13–57 and 62–67) and of Tc3A (2–46 and 47–52) shows an r.m.s. difference of 2.2 Å. Superposition of the N-terminal domains of Prd (13–55) and of Tc3A (2–44) shows an r.m.s. difference of 2.0 Å. In the C-terminal domains the helices (77–91, 93–104 and 114–128 of Pax6, 62–76, 77–88 and 90–104 of Tc3A) superimpose with r.m.s. difference of 0.8 Å. The C-terminal domain helices (77–87, 88–91, 94–104 and 114–124 of Prd, 62–72, 74–77, 78–88 and 90–100 of Tc3A) superimpose with r.m.s. difference of 1.2 Å. (**B**) Tc3A, (**C**) Pax6, and (**D**) Prd bound to DNA in the same view as (A) to show the different curvatures of the DNA.

flexible linkers, which make re-positioning easy. Here, we see that even with linkers that closely connect to the DNA variable positioning on the DNA of a bipartite DNA-binding domain is possible.

Comparisons of the DNA oligo's from the Pax6 and Prd structure show a number of differences. Only the Pax6 DNA was of similar length, 26 bp in all, while the Prd structure has a DNA oligo of 15 bp, filling only partially the DNA-binding pockets formed by the whole structure. The major difference between the Pax6 and Tc3 oligo's is in their respective conformations, with the Pax6 oligo being almost perfectly in the B-form. As shown in Figure 2, the Tc3 DNA has a bent conformation differing from B-DNA. The Prd oligo is also in the B-form, however the length of the oligo makes it hard to determine if a longer stretch of DNA would adopt the same conformation. The oligo from Pax6 is also lacking the GC-rich regions found at both ends of the Tc3 ITR, which may also account for the rigid nature of this oligo. Among these bipartite DNA-binding domains Tc3 is unusual in the extent of DNA distortion upon complex formation.

Comparison to other DDE transposases

Crystal structures of full-length Tn5 transposase with DNA have given insight to synaptic complex formation of a complete DDE-type transposase (24,33). This transposase has a similar catalytic domain to Tc3A, but the DNA-binding domains are differently arranged in the Tn5 transposase, with a specific DNA-binding domain at the N-terminus and a non-specific DNA-binding domain located at the C-terminus. When the DNA of the Tn5 complex is superimposed on that found in the Tc3 structure such that the DNA cleavage sites would overlap, the comparison could predict features of the Tc3 synaptic complex. In this superposition the space occupied by the catalytic domain of Tn5 is free, and a similar positioning of a catalytic domain at this transposon end would be feasible. Interestingly, the non-specific C-terminal DNA-binding domain of Tc3A is bound in the same area relative to the cleavage site as the C-terminal domain of Tn5. However, whereas the Tn5 N-terminal sequence recognition domain is binding a different DNA oligo, binding *in trans*, the N-terminal domain of Tc3A is bound *in cis*. Thus, the dimerization as observed in the Tn5 complex is not likely in Tc3A and the details of the synaps formation must be different in Tc3A. Similarly, the dimer formation that was observed in the crystal structure of HIV integrase catalytic-C-terminal domain (34) is different from those observed in the Tc3 crystal structures.

CONCLUSIONS

The bipartite DNA-binding domain of Tc3 transposase is shown to be composed of two HTH domains, separated by a linker. DNA recognition is present in all three fragments, but the N-terminal domain has the most extensive contacts. The two HTH domains are situated on opposite sides of the transposon DNA. They interact with a substantially distorted DNA oligomer, and the distortions are conserved between very different crystal forms, indicating that they are intrinsic to the complex. From this we conclude that the recognition of transposon DNA by Tc3A involves electrostatic complemen-

tarity, sequence-specific hydrogen bonds and the recognition of DNA distortion by shape complementarity.

In the crystal the bipartite DNA-binding domain is involved in two dimer interactions, with different neighbors. Each of these interfaces is relatively hydrophobic and could well belong to a functional multimer. In principle it is feasible that a tetramer of Tc3A molecules is necessary for the recognition of both the binding sites at both ITR ends in synaps formation. Although the observed arrangement of the Tc3A on the DNA is significantly different from the Tn5 DNA complex, in which the two DNA-binding domains bind *in trans* on the two ends of the transposon, a catalytic domain would comfortably fit at the cleavage site of the bound DNA. Further analysis however, will have to await a structure of the full-length transposase synaptic complex.

ACKNOWLEDGEMENTS

We thank George Verlaan and Ronald H.A. Plasterk for help in the project, Anastassis Perrakis for discussion, and Henri van Luenen, René Ketting and Ronald H.A. Plasterk for critical reading of the manuscript. We thank the staff at EMBL Hamburg for assisting with data collection. The structure has been submitted to the PDB with code 1U78. Funding was provided by NWO-CW, project 98016.

REFERENCES

1. Avancini,R.M., Walden,K.K. and Robertson,H.M. (1996) The genomes of most animals have multiple members of the Tc1 family of transposable elements. *Genetica*, **98**, 131–140.
2. Plasterk,R.H., Izsvak,Z. and Ivics,Z. (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet*., **15**, 326–332.
3. Curcio,M.J. and Derbyshire,K.M. (2003) The outs and ins of transposition: from mu to kangaroo. *Nature Rev. Mol. Cell Biol*., **4**, 865–877.
4. Ivics,Z., Kaufman,C.D., Zayed,H., Miskey,C., Walisko,O. and Izsvak,Z. (2004) The Sleeping Beauty transposable element: evolution, regulation and genetic applications. *Curr. Issues Mol. Biol*., **6**, 43–55.
5. Auge-Gouillou,C., Bigot,Y., Pollet,N., Hamelin,M.H., Meunier-Rotival,M. and Periquet,G. (1995) Human and other mammalian genomes contain transposons of the mariner family. *FEBS Lett*., **368**, 541–546.
6. van Luenen,H.G., Colloms,S.D. and Plasterk,R.H. (1994) The mechanism of transposition of Tc3 in *C. elegans*. *Cell*, **79**, 293–301.
7. Vos,J.C., De Baere,I. and Plasterk,R.H. (1996) Transposase is the only nematode protein required for *in vitro* transposition of Tc1. *Genes Dev*., **10**, 755–761.
8. Lampe,D.J., Churchill,M.E. and Robertson,H.M. (1996) A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J*., **15**, 5470–5479.
9. Colloms,S.D., van Luenen,H.G. and Plasterk,R.H. (1994) DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase. *Nucleic Acids Res*., **22**, 5548–5554.
10. Fischer,S.E., van Luenen,H.G. and Plasterk,R.H. (1999) *Cis* requirements for transposition of Tc1-like transposons in *C. elegans*. *Mol. Gen. Genet*., **262**, 268–274.
11. Ketting,R.F., Fischer,S.E. and Plasterk,R.H. (1997) Target choice determinants of the Tc1 transposon of *Caenorhabditis elegans*. *Nucleic Acids Res*., **25**, 4041–4047.
12. Vigdal,T.J., Kaufman,C.D., Izsvak,Z., Voytas,D.F. and Ivics,Z. (2002) Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/*mariner* transposable elements. *J. Mol. Biol*., **323**, 441–452.
13. Vos,J.C., van Luenen,H.G. and Plasterk,R.H. (1993) Characterization of the *Caenorhabditis elegans* Tc1 transposase *in vivo* and *in vitro*. *Genes Dev*., **7**, 1244–1253.
14. Vos,J.C. and Plasterk,R.H. (1994) Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. *EMBO J*., **13**, 6125–6132.
15. Glaser,T., Walton,D.S. and Maas,R.L. (1992) Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nat. Genet*., **2**, 232–239.
16. Phillips,K. and Luisi,B. (2000) The virtuoso of versatility: POU proteins that flex to fit. *J. Mol. Biol*., **302**, 1023–1039.
17. van Pouderoyen,G., Ketting,R.F., Perrakis,A., Plasterk,R.H. and Sixma,T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C.elegans* in complex with transposon DNA. *EMBO J*., **16**, 6044–6054.
18. Otwinowski,Z. and Minor,W. (eds) (1997) *Processing of X-Ray Data Collected in Oscillation Mode*. Academic Press, NY.
19. Perrakis,A., Harkiolaki,M., Wilson,K.S. and Lamzin,V.S. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr. D. Biol. Crystallogr*., **57**, 1445–1450.
20. Murshudov,G.N., Vagin,A.A., Lebedev,A., Wilson,K.S. and Dodson,E.J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr*., **55** (Pt 1), 247–255.
21. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr*., **57**, 122–133.
22. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
23. Hooft,R.W., Vriend,G., Sander,C. and Abola,E.E. (1996) Errors in protein structures. *Nature*, **381**, 272.
24. Davies,D.R., Goryshin,I.Y., Reznikoff,W.S. and Rayment,I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77–85.
25. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*., **31**, 5108–5121.
26. Lu,X.J., Shakked,Z. and Olson,W.K. (2000) A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol*., **300**, 819–840.
27. Izsvak,Z., Khare,D., Behlke,J., Heinemann,U., Plasterk,R.H. and Ivics,Z. (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. *J. Biol. Chem*., **277**, 34581–34588.
28. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol*., **233**, 123–138.
29. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol*., **247**, 536–540.
30. Xu,W., Rould,M.A., Jun,S., Desplan,C. and Pabo,C.O. (1995) Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell*, **80**, 639–650.
31. Xu,H.E., Rould,M.A., Xu,W., Epstein,J.A., Maas,R.L. and Pabo,C.O. (1999) Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and C-terminal subdomain in DNA binding. *Genes Dev*., **13**, 1263–1275.
32. Garvie,C.W., Hagman,J. and Wolberger,C. (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol. Cell*, **8**, 1267–1276.
33. Davies,D.R., Braam,L.M., Reznikoff,W.S. and Rayment,I. (1999) The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9 Å resolution. *J. Biol. Chem*., **274**, 11904–11913.
34. Wang,J.Y., Ling,H., Yang,W. and Craigie,R. (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J*., **20**, 7333–7343.