

Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C.elegans* in complex with transposon DNA

Gertie van Pouderooyen¹, Rene F.Ketting²,
Anastassis Perrakis¹, Ronald H.A.Plasterk²
and Titia K.Sixma^{1,3}

Divisions of ¹Molecular Carcinogenesis and ²Molecular Biology,
Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam,
The Netherlands

³Corresponding author
e-mail: sixma@nki.nl

The crystal structure of the complex between the N-terminal DNA-binding domain of Tc3 transposase and an oligomer of transposon DNA has been determined. The specific DNA-binding domain contains three α -helices, of which two form a helix–turn–helix (HTH) motif. The recognition of transposon DNA by the transposase is mediated through base-specific contacts and complementarity between protein and sequence-dependent deformations of the DNA. The HTH motif makes four base-specific contacts with the major groove, and the N-terminus makes three base-specific contacts with the minor groove. The DNA oligomer adopts a non-linear B-DNA conformation, made possible by a stretch of seven G:C base pairs at one end and a TATA sequence towards the other end. Extensive contacts (seven salt bridges and 16 hydrogen bonds) of the protein with the DNA backbone allow the protein to probe and recognize the sequence-dependent DNA deformation. The DNA-binding domain forms a dimer in the crystals. Each monomer binds a separate transposon end, implying that the dimer plays a role in synapsis, necessary for the simultaneous cleavage of both transposon termini.

Keywords: crystal structure/DNA binding/helix–turn–helix/Tc1/Tc3 transposase

Introduction

Tc3 of *Caenorhabditis elegans* is a member of the Tc1/*mariner* family of transposable elements. Members of that family are found in a wide variety of organisms, ranging from fungi to humans (Doak *et al.*, 1994). Transposable elements are small stretches of DNA that can move from one position in the genome to another. The proteins responsible for the excision and insertion of the transposon into the genome are generally encoded by the transposon sequences. In the Tc1/*mariner* family, the transposons encode only a single protein, the transposase, that is capable of performing the entire transposition reaction *in vitro* (Lampe *et al.*, 1996; Vos *et al.*, 1996). The Tc1/*mariner* transposase genes are flanked by terminal inverted repeats. The sequences of the terminal inverted repeats are not conserved between different elements, apart from the four most terminal nucleotides. Another shared

property of this family is that the transposon DNA is inserted into TA sequences of the host genome (for a review, see Plasterk, 1996).

The first step in transposition is the recognition of the transposon DNA by a specific DNA-binding domain of the transposase protein. For Tc3A, the Tc3 transposase, it has been shown that the N-terminal domain is responsible for this specific DNA binding. This domain binds a region ~12 bases away from the DNA cleavage site (van Luenen *et al.*, 1993; Colloms *et al.*, 1994) (Figure 1). Between the N-terminal specific DNA-binding domain and the catalytic domain, another DNA recognition domain is present. This second, more C-terminal domain recognizes DNA sequences located more towards the cleavage site (R.F.Ketting and R.H.A.Plasterk, unpublished results). Such a bipartite DNA binding has also been shown for the related Tc1 transposase (Tc1A) of *C.elegans* (Vos and Plasterk, 1994). In general, there is little sequence conservation in the N-terminal regions of Tc1/*mariner* transposases and no apparent conservation of the inverted repeat sequence. Thus, the specific DNA-binding domain of each transposase of this family recognizes the termini of its own transposon specifically, and the transposase protein of a given element will act only upon its own transposon ends.

Protein sequence alignments have revealed a weak similarity between the N-terminal region of the Tc1-like *Minos* element and a DNA-binding domain of the Pax/paired family, the paired domain, found in mammalian and *Drosophila* genes (Franz *et al.*, 1994). The paired domain is a conserved DNA-binding domain found in a set of transcription factors (Pax proteins) that play important roles in development. The significance of the similarity is enhanced by the identification of the bipartite nature of the paired DNA-binding domain (Czerny *et al.*, 1993; Xu *et al.*, 1995). The difference from the Tc3A bipartite DNA binding is that the C-terminal part of the paired domain could not be shown to bind to DNA (Xu *et al.*, 1995). However, in other members of the paired domain family (Pax proteins), a C-terminal part plays a role in site-specific recognition (Czerny *et al.*, 1993; Epstein *et al.*, 1994a,b; Xu *et al.*, 1995). Furthermore, secondary structure elements are similar for the Tc1A and Tc3A DNA-binding domains (as predicted with PHDsec, Rost and Sander, 1993).

Part of the catalytic domain is also involved in DNA binding. In a South-Western assay, it has been shown that a region between amino acids 98 and 159 shows non-specific DNA-binding activity (Colloms *et al.*, 1994). The catalytic domain of Tc3A shares sequence homology with the IS630-Tc1 family and contains a catalytic triad, DDE motif, of carboxyl groups (Figure 1) (Doak *et al.*, 1994), that is important for the Tc3 transposition activity (van Luenen *et al.*, 1994). Other polynucleotidyl transferases

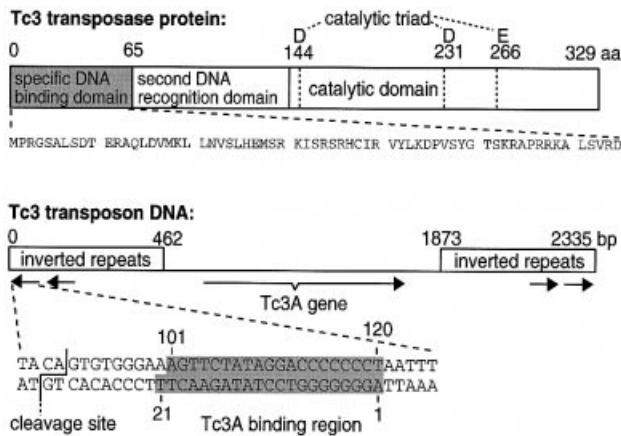


Fig. 1. Schematic representation of Tc3 transposase protein and Tc3 transposon DNA. Shaded boxes indicate which part of the protein and DNA were co-crystallized in this study. The numbering of the DNA oligomer used in this study is indicated. The arrows under the inverted repeats of the DNA indicate the two almost identical binding sites of Tc3A separated by ~180 bp at each transposon end (Colloms *et al.*, 1994). The binding site region used in this study (indicated by a gray box) is identical in these two sites. The function of the internal binding site is not very clear, because it does not seem essential for efficient transposition (H.G.A.M.van Luenen and R.H.A.Plasterk, personal communication). The sequence of the 65 amino acid N-terminal fragment of Tc3A differs at position 41 (Val instead of Glu) from the sequence in the GenBank database.

also use three (sometimes four) carboxyl groups to position one or two catalytic divalent metal ions (Yang and Steitz, 1995). This family includes ribonuclease H (Doolittle *et al.*, 1989), retroviral/retrotransposases (Fayet *et al.*, 1990; Kulkosky *et al.*, 1992), Mu transposase (Baker and Luo, 1994) and RuvC resolvase (Ariyoshi *et al.*, 1994). The crystal structures of the catalytic domains of several of these proteins (Katayanagi *et al.*, 1990, 1992; Yang *et al.*, 1990; Ariyoshi *et al.*, 1994; Dyda *et al.*, 1994; Rice and Mizuuchi, 1995) revealed that they share a structurally related core in which the catalytic carboxyl groups are oriented in a similar way. It is likely that the catalytic domain of the Tc3 transposase has a similar three-dimensional fold.

The detailed mechanism of transposition has many variations, for example in invoking single or double strand breaks of the host DNA, in the target site specificity or in the number of proteins involved (for a review, see Plasterk, 1995). Synapsis (assembly of multiple proteins, the transposon DNA ends and the target DNA) is thought to be required for proper transposition. This has been studied extensively for phage Mu, in which tetramerization of the transposase on the transposon ends is essential for transposition (for a review, see Chaconas *et al.*, 1996). Another example of the importance of synapsis has been shown for the Moloney murine leukemia virus. When one of the two ends of this viral DNA is mutated to block cleavage, cleavage is inhibited at the other (wild-type) end as well. This indicates that cleavage at either end (the first step in the integration process) only takes place when both ends contribute to the synapsis and are recognized by the integrase protein (Murphy and Goff, 1992).

To understand how the Tc3 transposase recognizes its own transposon DNA ends, we studied the structure of the N-terminal specific DNA-binding domain (Tc3A-N) in complex with transposon DNA. This is the first crystal

structure of a DNA-binding region of a transposase in complex with DNA. With this, we can begin to understand the mode of recognition by transposases of their DNA substrates, and eventually the mode of binding and regulation of simultaneous cleavage at the two ends of the transposon.

Results and discussion

The overall structure of the protein-DNA complex

We have determined the X-ray structure of the 65 amino acid DNA-binding domain of Tc3 transposase (Tc3A-N) in complex with transposon DNA (Figure 2A) at 2.45 Å resolution. The parts of the transposase and transposon used in co-crystallization are indicated by shaded boxes in Figure 1. The structure shows that Tc3A-N contains three α -helices (residues 9–20, 25–32 and 36–44) typical of proteins sharing the helix–turn–helix (HTH) motif, such as homeodomains. The first helix is involved in dimerization of the protein domains in the crystal. Each monomer binds a DNA molecule. The second and third helix form the HTH motif and are involved in DNA recognition in the major groove. The N-terminus of the protein interacts with DNA in the locally narrowed minor groove. The C-terminus (residues 53–65 and the 6-histidine tag) is not visible in the final electron density, probably due to flexibility. The 20/21 DNA oligomer is in a non-linear B-DNA conformation and both the major and minor groove interact with the protein.

The HTH motif makes base-specific contacts in the major groove

Part of the recognition process of Tc3A for its own transposon DNA is mediated by direct readout of four bases by the HTH motif. The HTH motif, comprising helices 2 and 3, interacts in the major groove of the DNA. There are also extensive contacts with mainly one DNA backbone strand: 10 H-bonds and seven salt bridges (Figure 2B and C).

Within the HTH motif, the N-terminus of helix 2 is involved in extensive contacts with the phosphate groups in one DNA backbone strand. However, one side chain (His26) is making a purine-specific hydrogen bond to G7 (for base numbering see Figure 1). Backbone amides of Leu25 and His26 make hydrogen bonds to phosphate group 7. The dipole moment (Hol *et al.*, 1978) of helix 2 interacts favorably with phosphate group 7 as well. The contacts with one of the DNA backbones are consolidated by Ser24 (in the loop connecting helices 1 and 2) and Arg30, making hydrogen bonds to phosphate group 6. More salt bridges are made by His26 with phosphates 6 and 7 and by Arg30 with phosphates 5 and 6.

Two amino acids in the turn of the HTH motif make contacts with the other DNA backbone: Arg34 is hydrogen bonded to sugar 109 and salt bridged to phosphate 110. Ser35 donates a hydrogen to phosphate 110.

The N-terminal part of helix 3 (the recognition helix in the HTH motif) is involved in base-specific contacts with three bases (including one mediated via a water molecule) and several DNA backbone contacts. Arg36 makes one guanine- and one purine-specific hydrogen bond to the guanine base at position 8. Arg40 makes a thymine-specific hydrogen bond, via a water molecule, to T9.

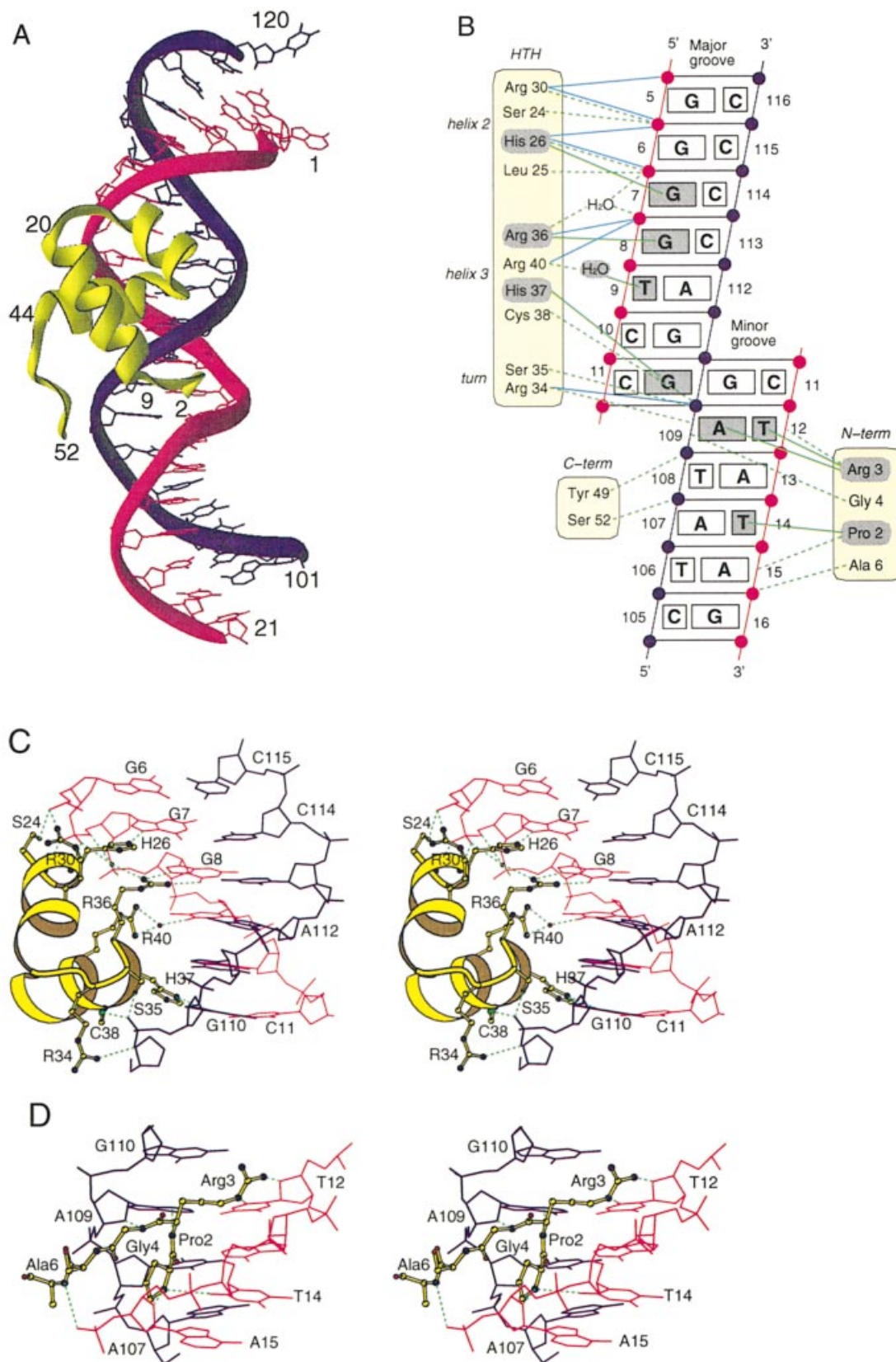


Fig. 2. Protein–DNA contacts. (A) A schematic view with ribbons drawn through the C α s of the Tc3A DNA-binding domain (yellow) and through the phosphate backbone of the DNA strands (blue and magenta). (B) Sketch summarizing the hydrogen bonding (indicated by green dotted lines, base-specific H-bonds by green solid lines) and salt bridging contacts (blue lines) between the Tc3A domain and the DNA. Gray boxes indicate residues involved in base-specific contacts. Hydrogen bonds are at a maximum distance of 3.5 Å and salt bridges at a maximum of 4.0 Å (Barlow and Thornton, 1983). (C) and (D) Stereo views (Kraulis, 1991) of the HTH DNA contacts in the major groove, and the N-terminus of Tc3A bound in the minor groove of DNA, respectively. Hydrogen bonds are indicated with green dotted lines.

His37 recognizes G110 via a hydrogen bond to the purine-specific N7 or to the guanine-specific O6 of G110. We are not able to distinguish this at the current resolution. Helix 3 makes additional hydrogen bonds through Arg36 via a water molecule to phosphates 7 and 8 and through Cys38 to phosphate 110. Salt bridges are made by both Arg36 and Arg40 to phosphate group 8.

Minor groove interactions

Additional base-specific recognition (of three bases closer to the DNA cleavage site) occurs by binding of the N-terminus in the minor groove. Six more hydrogen bonds are made with the DNA backbone by both the N- and C-termini of the domain (Figure 2).

The side chains of residues Pro2 and Arg3 are inserted in the minor groove. Since Met1 is missing (as shown by sequencing of the purified protein), Pro2 is at the N-terminus and thus positively charged. The two positively charged N-terminal residues bind like a thumb and finger between the negatively charged phosphate backbones of the locally relatively narrow minor groove (Figure 2D). Pro2 makes a pyrimidine-specific hydrogen bond to T14 and a hydrogen bond to sugar 15. The rest of Pro2 is in van der Waals contact with sugars 15 and 108. Arg3 makes one pyrimidine-specific hydrogen bond to T12 and one to sugar 12. The backbone amide of Arg3 most likely makes a purine-specific H-bond to A109, but a H-bond to sugar 109 cannot be excluded. The backbone amide of Gly4 is hydrogen bonded to sugar 109 and the backbone amide of Ala6 to phosphate 16.

The last ordered amino acids of the C-terminus (49–52) are close to the minor groove as well. Tyr49 OH and Ser52 O γ are donating hydrogens to phosphate groups 109 and 108 respectively. There is some electron density visible beyond Ser52 in the minor groove, although it is not continuous and cannot be interpreted, indicating that the flexible C-terminus is continuing along the minor groove towards the DNA cleavage site.

Protein dimer

The Tc3A domain behaves like a 25 kDa protein on a gel filtration column, while the mol. wt is only 8.3 kDa. The 19 flexible C-terminal amino acids are likely to cause a shift to higher molecular weight; however, the difference between the expected 8.3 kDa for a monomeric domain and the observed 25 kDa peak cannot be attributed to that only. It is more likely that Tc3A-N is forming a dimer in solution.

In the crystal, we observe protein dimers as well, in which each protein monomer binds one DNA molecule. Helix 1 is involved in the dimerization of two protein domains by a 2-fold crystallographic rotation. Each protein domain makes, at one side, contact with DNA and, at the opposite side, contact with a protein of a symmetry-related complex (Figure 3). The accessible surface of a monomer of the Tc3A domain is 3999 Å² (calculated with GRASP, Nicholls *et al.*, 1991). Upon dimerization with another Tc3A domain, 12% (475 Å²) of this surface is buried on each monomer. The contact, which involves mainly helix 1 of each domain, is predominantly hydrophobic (24 van der Waals contacts), and only two hydrogen bonds occur (between the carbonyl oxygen of Ala13 of both molecules and the N ϵ of Gln14 of the symmetry-related molecules). The dimer found in the

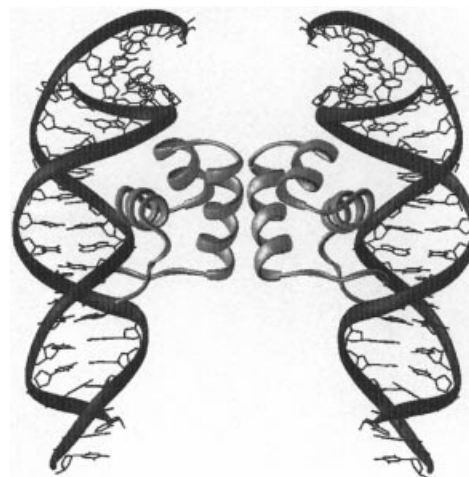


Fig. 3. Ribbon diagram displaying the dimer of the Tc3A DNA-binding domains with the two DNA oligomers of the transposon ends bound.

crystals, and presumably in solution, may also be present in active transposition complexes.

DNA conformation and recognition through sequence-dependent conformation

The DNA in the crystallized complex is not in a linear conformation (Figure 2A), but bent at both ends in different directions and planes. The bending of the DNA is stabilized (Strauss and Maher, 1994) by the large amount of positive charge on the protein, which is located mainly at the interface with DNA (Figure 4).

The conformation of the DNA was analyzed with the program CURVES (Lavery and Sklenar, 1989) using the global parameters. The average helical twist of 33.5° (10.7 residues per turn) and the average rise per base pair of 3.41 Å are typical for B-DNA. The bends of the DNA are reflected in increases in the roll and tilt angles of the base pairs and in deviations of the major and minor groove widths and depths, compared with average B-DNA (Stofer and Lavery, 1994). To define the sugar puckers, higher resolution data will be needed.

DNA sequences highly enriched in G:C base pairs have a tendency to form low twist angles (underwinding) and positive roll angles of the base pairs, resulting in a widening of the minor groove and narrowing of the major groove (Travers, 1993). The seven consecutive G:C base pairs (2:119–8:113) in this structure indeed have a relatively low twist angle (30.5° on average) and a positive roll angle (~12° per base pair). The roll angles of this stretch add up to an 82° bend of the DNA. In the crystal, one end of each DNA molecule bends into the minor groove of a symmetry-related DNA molecule, with the base pairs of the two DNA molecules almost perpendicular (Figure 5A). Part of this bending is visible in the top part of Figure 2A as a DNA bend to the right, but it is also partly away from the viewer.

At the position of the bent G:C stretch, Tc3A-N makes numerous hydrogen bonds and salt bridges to only one of the DNA backbones (phosphates 5–8, Figure 2B and C). These contacts would not have been possible if the DNA was more linear. It is not clear whether the bend is present in the DNA alone, to which the protein adapts itself, or

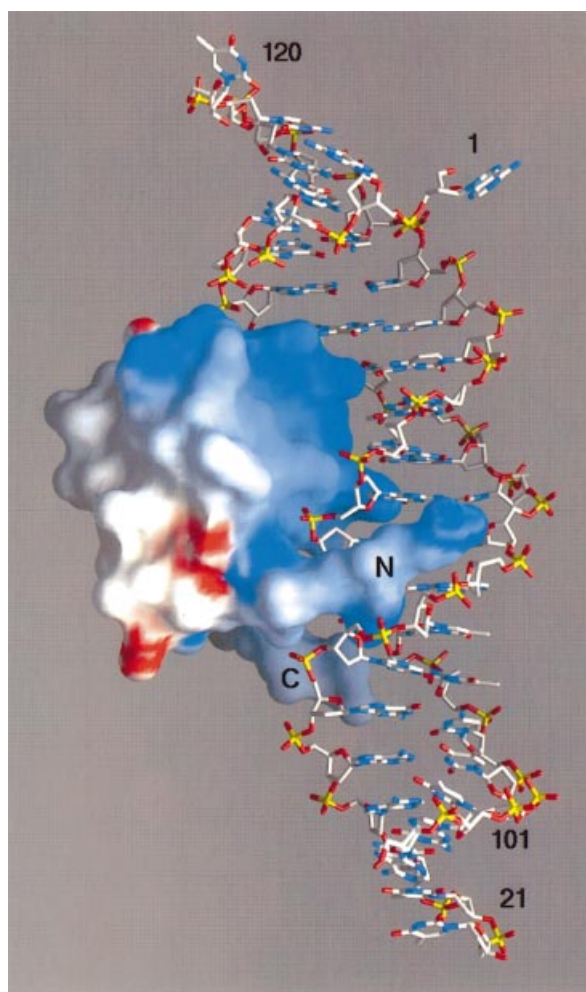


Fig. 4. The Tc3A domain bound to DNA. The protein is shown in an electrostatic surface representation with positively and negatively charged regions in blue and red respectively (GRASP-scale -10 to $+10$). DNA is shown in stick representation, with carbons in white, nitrogens in blue, oxygens in red and phosphors in yellow.

whether the bend is induced by the interaction with the protein. A third theoretical possibility, that the bend is caused or stabilized by the crystal contacts, seems unlikely given the exquisite complementarity of the protein and the deformed DNA.

The bending at the other end of the DNA oligomer (more towards the DNA cleavage site in the lower part of Figure 2A) is reflected in mainly negative roll and tilt angles (around base pairs 13:108 and 14:107) and a narrowing of the minor groove (4 \AA compared with 6 \AA for average B-DNA; Stofer and Lavery, 1994). The bending is $\sim 30^\circ$ and the direction is to the right in Figure 2A, and partly towards the viewer. This part of the DNA sequence contains AT bases, which have a tendency to form a narrow minor groove (Travers, 1993). In this narrow minor groove, the N-terminus is inserted, making hydrogen bonds and van der Waals contacts with both DNA backbones (Figure 2D). Such intimate protein–DNA contacts would not be possible with a wider minor groove. Again, the Tc3A domain is complementary to the sequence-dependent DNA conformation, contributing to the recognition by Tc3A of its own transposon DNA.

Unusual DNA–DNA crystal contacts

The DNA helix axis is roughly in the crystallographic b -direction. This is visible in the diffraction pattern as a fiber diffraction pattern at 3.4 \AA along the b -axis. Although the DNA is mainly parallel to the b -axis, its non-linearity causes unusual crystal contacts.

The A1 and T120 bases at one end of the oligomer do not form the expected base pair, but are flipped outwards. These bases, together with the last base pair (G2–C119), are interacting in the minor groove and are almost perpendicular to the base pairs of a symmetry-related DNA molecule (Figure 5A). Fraying of a DNA molecule at one end has been observed before (Joshua-Tor *et al.*, 1992), where the flipped-out bases are also interacting in the minor groove of a symmetry-related molecule. However, the base pairs of the symmetry-related molecule are not perpendicular, as in this case.

The other end of the DNA forms a shifted semi-continuous helix with the same end of a symmetry-related DNA molecule (Figure 5B). These two molecules are related by a 2-fold rotation axis in the a -direction (perpendicular to the paper in Figure 5B). The overhanging base T21 makes a triple helix by forming a Hoogsteen base pair with the symmetry-related base A101 in base pair A101:T20.

Comparison with other proteins

The three-dimensional fold of the Tc3A domain is a HTH fold. Proteins sharing this motif are the eukaryotic homeodomains and transcription factors, prokaryotic repressors and Hin recombinase, amongst others (for a review, see Luisi, 1995; Wintjens and Rooman, 1996). The Tc3A domain is most similar to the N-terminal part of the paired domain (Figure 6, Xu *et al.*, 1995). The r.m.s. difference of 51 C α s is 2.3 \AA , according to the 3D alignment of the DALI-server (<http://www.embl-heidelberg.de/dali>; Holm and Sander, 1993), and the r.m.s. difference of the C α s in the three helices and connecting loops is only 0.66 \AA . Other similar structures are, for example, the Oct-1 POU domain (Klemm *et al.*, 1994), C-myb DNA-binding domain (Ogata *et al.*, 1995), engrailed homeodomain (Kissinger *et al.*, 1990) and the paired homeodomain (Wilson *et al.*, 1995). The three helices and connecting loops have r.m.s. C α differences with Tc3A-N of 1.3 – 1.5 \AA .

Surprisingly, the recently determined Zn-containing N-terminal region of the functionally similar HIV integrase also contains a comparable HTH motif (Cai *et al.*, 1997; Eijkelenboom *et al.*, 1997). Superposition with the three helices and connecting loops of Tc3A-N results in a C α r.m.s. deviation of 2.0 \AA . The C-terminus, however, folds in a different direction, where it provides two of the ligands to the Zn ion, which is coordinated further by a histidine in helix 1 and a histidine in the loop between helices 1 and 2. Data on the function of this domain are not clear, and it is unknown whether it is involved in DNA binding, although the presence of the HTH domain is suggestive.

The main differences between the N-terminal part of the paired domain and Tc3A-N are located in their N- and C-termini. The few residues preceding the first helix in the Tc3A domain adopt a conformation different from that of the longer N-terminus of the paired domain, which

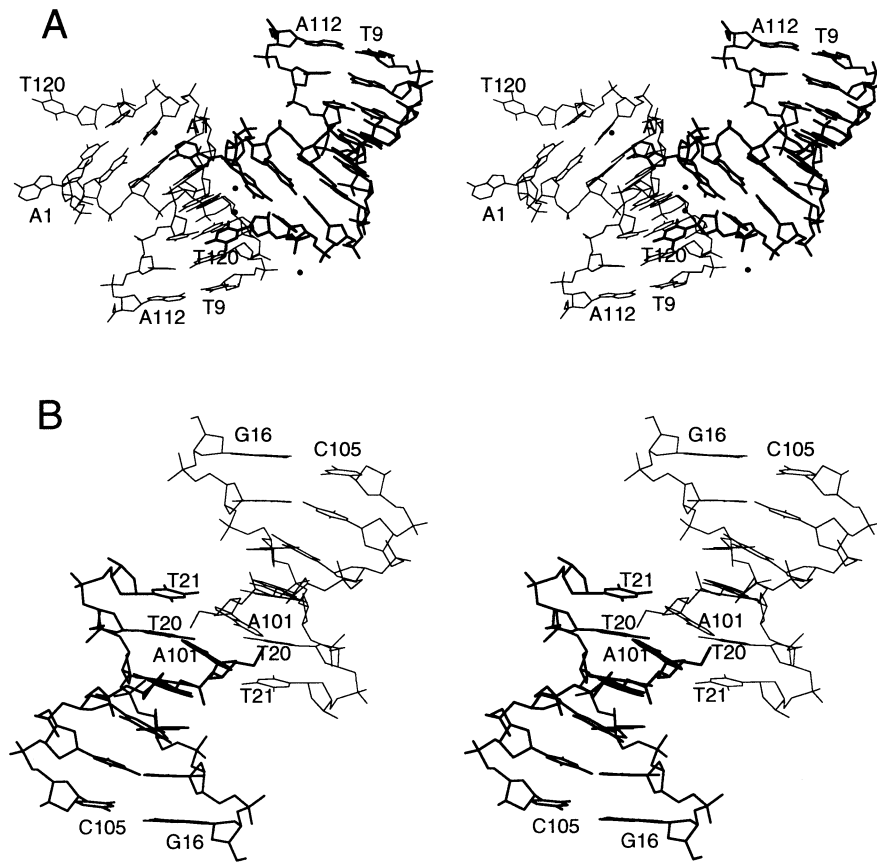


Fig. 5. Stereo view (Kraulis, 1991) of the DNA stacking in the crystal at both ends of the DNA oligomer. (A) View of one end along the 2-fold screw axis in the *c*-direction (water molecules are indicated as dots) and (B) view of the other end along the 2-fold axis in the *a*-direction.

forms a small β -sheet (Figure 6). Helix 3 is longer in the paired domain than in the Tc3A domain, but the direction of the C-termini is similar. In the paired domain, the C-terminus is connected to the second domain, but the linker is not clearly visible in the electron density (Xu *et al.*, 1995). In our crystals, only the loop towards the second DNA recognition domain (residues 53–65) is present, and it is invisible in the electron density as well.

Docking on DNA

It has been described earlier that there are several ways of docking an HTH motif on DNA (for example, see Suzuki and Gerstein, 1995; Wintjens and Rooman, 1996). The Tc3A-N protein is very similar to the N-terminal part of the paired domain and homeodomains. These have, however, a different way of docking on DNA. The homeodomains have only helix 3 (the recognition helix) inserted in the major groove, and the residues in the center of this helix are interacting with DNA. This recognition helix is relatively long and there are common features in the amino acid sequence playing a role in DNA interaction (Suzuki, 1993). The paired domain belongs to another family, which also includes the prokaryotic Hin recombinase and λ repressor. In this family, both helices 2 and 3 interact with DNA, and this family is distinguished by a relative short helix 2 of which the N-terminus interacts with the DNA backbone. Helix 3 is also interacting in the major groove, but at a different angle compared with the homeodomain family (Kissinger *et al.*, 1990; Xu *et al.*, 1995). The HTH motif of the Tc3A domain docks in a

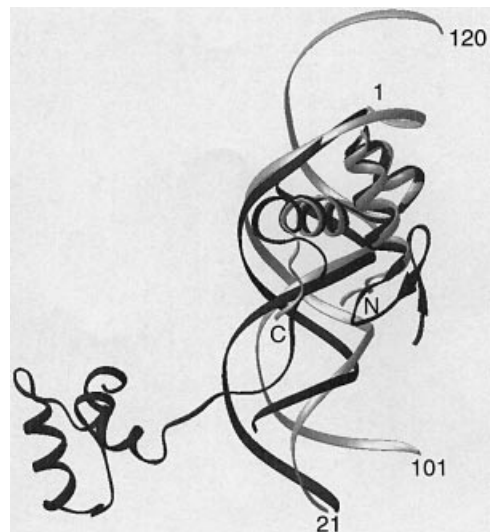


Fig. 6. A ribbon representation of the superposition of the Tc3A domain (light gray) and the equivalent part of the paired domain (dark gray), with their respective DNA molecules bound. Note the bending of the Tc3 oligomer in contrast to that bound to paired.

variant of the paired/Hin/ λ family, although the non-linear DNA in the Tc3A complex is exceptional.

There are nine residues (18%) in Tc3A-N that are identical to the N-terminal part of the paired domain. None of these residues are present at equivalent positions in the Hin recombinase (Feng *et al.*, 1994) or the λ repressor (Beamer and Pabo, 1992). Only three of these

residues (Arg30, Ser35 and Cys38 of Tc3A) are involved in similar DNA contacts. None of these are sequence specific; all three interact with the phosphate groups in the DNA backbone. The absence of conserved residues that interact with DNA is in contrast to the homeodomain family, which has anchor residues.

The locations within the HTH fold of the side chains that interact with the bases of the DNA are not very conserved either. Some side chains are located at the N-terminus of helix 3, but not at identical positions. Others are situated towards the center of helix 3, but again not at a fixed position. Tc3A-N has one exceptional side chain in helix 2 (His26) that is involved in a base contact.

Correlation with biochemical data

The double-stranded DNA oligomer used in this study was based on methylation interference (12 bp from G7 to A18) and footprinting (~20 bases on each strand) studies (Colloms *et al.*, 1994). The adenine and guanine bases which showed strong methylation interference are indeed in contact with the protein. A few bases with weak interference (G16, A17 and A18) are located where the flexible C-terminus of Tc3A-N is pointing. The middle of the 20/21 oligomer in the crystal is in contact with the protein, both in the major and minor groove, consistent with the prediction based on the methylation and footprinting studies (Colloms *et al.*, 1994).

The same 65 amino acid N-terminal domain with a histidine tag at the N-terminus (instead of the C-terminus as used in this study) did not bind transposon DNA, as shown by gel retardation assays (data not shown). The N-terminal residues of the protein (Pro2 and Arg3) are bound in the minor groove, and a histidine tag at the N-terminus would probably interfere with these contacts, making a protein–DNA complex impossible. Whether the N-terminal Met1 (absent in this study) is also absent *in vivo* in *C.elegans* is unclear. The base-specific contacts of Pro2 could be disturbed if Met1 were present, suggesting that Met1 is also absent *in vivo*.

We were not able to see the amino acids beyond Ser52, although some electron density in the minor groove suggests that the chain is continuing along the minor groove in analogy with the loop in the paired domain. A smaller N-terminal part of Tc3A (amino acids 1–54) has been shown, however, not to bind to Tc3 transposon DNA (Colloms *et al.*, 1994). This indicates that at least part of the floppy end (54–65) is essential for proper folding of this domain and/or DNA binding.

Comparison with Tc1 and Tc1A

Although the N-terminal DNA-binding domains of the related Tc3A and Tc1A show no obvious homology, secondary structure prediction (PHDsec, Rost and Sander, 1993) of the N-terminal domain of Tc1A predicts three α -helices (residues 12–23, 28–35 and 39–49), extremely similar to the Tc3A domain, strongly suggesting a HTH motif. Ten amino acids (out of the 65 of Tc3A) are identical in both proteins, when they are aligned on their (predicted) secondary structures. Most of these residues do not interact with DNA, except Arg34 and Ser35 in Tc3A (Arg37 and Ser38 in Tc1A), making DNA backbone contacts. The DNA backbone contact of Ser35 in Tc3A

was also observed at the equivalent position in the paired domain.

DNase footprints and methyl interference studies have shown that Tc1A and Tc3A bind to the transposon ends at (almost) identical distances from the DNA cleavage site (Vos *et al.*, 1993; Colloms *et al.*, 1994). However, the DNA sequence that is recognized by the Tc1A DNA-binding domain is not at all similar to the one recognized by Tc3A. Only three base pairs in the Tc1 and Tc3 transposon ends that are contacted by the N-terminal domains are identical, when aligned from the DNA cleavage site. These base pairs (T12:A109, A13:T108 and T14:A107 in our numbering system) are in the region of the narrow minor groove where the N-terminus of Tc3A is interacting. However, Tc1A has a three amino acid longer N-terminus when it is aligned on (predicted) secondary structure and, therefore, probably does not have an analogous mode of recognition.

Conclusions

Prediction of other transposase DNA recognition modes.

The crystal structure shows that recognition of Tc3 transposon DNA by the Tc3 transposase is based on a combination of sequence-specific hydrogen bonds as well as a remarkable complementarity between deformed DNA and the protein. In the complex, the transposon sequence has a non-linear character with irregular major and minor grooves. This DNA conformation is stabilized by extensive interactions with the protein, which could not be present if the DNA was linear. Thus, the protein either probes the pre-set deformation of the DNA, or possibly makes use of the ability of this particular DNA sequence to change its conformation as a response to this interaction. Further specificity is given by a number of base-specific hydrogen bonds, in both the major and the minor groove of the DNA.

As predicted, the Tc3A-N protein has the same fold as that found for the N-terminal part of the paired domain in the paired/Pax family of proteins. The two proteins also dock on DNA in the same location and orientation. However, very few specific contacts have been conserved, and the impressive complementarity to sequence-dependent deformation of Tc3 was not seen in the paired domain, where the DNA is almost linear. This limits the possibilities of predicting DNA binding for other transposases. For the Tc1/*mariner* family, the weak sequence homology and the similar secondary structure prediction results lead to the expectation that their N-terminal DNA-binding domains will have a similar HTH fold with a DNA docking mode analogous to the Tc3/paired/Hin recombinase/ λ repressor family. Detailed predictions on the mode of interaction with DNA are, however, not possible. In fact, analysis of the *C.elegans* Tc1 transposon DNA sequence would lead to a prediction that any potential complementarity of protein and DNA in the case of Tc1 should give a very different complex because the DNA sequence is very different from that in Tc3. However, it is likely that again a combination of sequence-dependent conformability with base-specific recognition plays an important role in DNA recognition by other members of the Tc1/*mariner* family.

Thus, the structure of the Tc3 transposon–transposase complex gives a good explanation of the specificity of recognition of Tc3, but does not help to predict other transposon–transposase complexes in detail.

Table I. Statistics on data collection^a and heavy atom refinement^b

Data set	Native1	Native2	CH ₃ HgCl	Hg(SCN) ₂	I-DNA
Temperature	4°C	-160°C	25°C	25°C	4°C
Resolution range (Å)	90–2.85	30–2.45	30–3.40	90–3.50	20–2.85
Wavelength (Å)	0.917	0.891	1.5418	1.5418	1.20
No. of observations	26 396	70 943	36 684	15 009	23 509
No. of unique reflections	5671	8089	3645	2991	5496
Completeness (last shell) (%)	92.1 (63.5)	89.7 (67.1)	99.6 (100.0)	87.5 (89.5)	91.6 (37.8)
R_{merge}^c (last shell) (%)	5.8 (42.5)	5.5 (43.8)	20.0 (69.2)	13.1 (53.7)	9.6 (53.5)
$I/\sigma(I)$ (last shell)	17.3 (2.0)	19.4 (1.9)	8.4 (1.9)	8.9 (2.2)	11.4 (1.5)
Resolution limit used in phasing (iso/ano) (Å)			3.2/4.0	3.2/4.2	2.85/5.0
No. of reflections (iso/ano) [$F/\sigma(F)>3$]			2616/1328	2285/1147	3937/758
Overall phasing power (iso/ano)			2.65/1.98	2.95/1.42	1.68/1.28

^aThe space group is C222₁, the cell dimensions for the frozen crystal ($a = 37.18$ Å, $b = 202.79$ Å, $c = 62.14$ Å) are slightly different from the non-frozen crystals ($a = 37.75$ Å, $b = 205.23$ Å, $c = 63.31$ Å).

^bOverall mean figure of merit (2.85 Å) 0.493 (4129 reflections)

^c $\sum |I - \langle I \rangle| / \sum I$.

The synaptic complex. The complete Tc3 transposase molecule, as outlined in Figure 1, consists of a bipartite DNA-binding domain and a catalytic domain. Using the similarity to the paired domain structure, we can make some predictions about the second half of the bipartite DNA-binding domain of Tc3A. It is likely that the second half of the domain also contains an HTH motif, similar to that in paired, according to the weak sequence homology between the Tc1/*mariner* family and the paired domain. This prediction fits well with secondary structure prediction with PHDsec. In agreement with footprinting data, this second domain will probably bind to the DNA closer to the cleavage site, roughly in the same location as in the paired domain, since the C-terminal flexible region of Tc3A-N (53–65) runs in an equivalent direction to the flexible connecting loop in the paired domain. Nevertheless, the paired domain structure cannot make any prediction about the DNA-binding mode of the second HTH domain in Tc3A, since this region does not interact with DNA in paired form. Another family in which bipartite HTH domains occur is the family of eukaryotic homeodomain proteins. There the two HTH domains can interact with DNA in a variety of ways (for a review, see Tullius, 1995). It is conceivable that a similar variety will be seen in the Tc3/paired family.

To carry out the transposition reaction, the transposase must first recognize and synapse the two transposon termini before cleaving both the ends. The cleavage site in the transposon DNA has to be positioned properly within the active site of the catalytic domain of the transposase. Maybe the non-linear conformation of the transposon DNA, as seen in the crystal structure, plays a role in bringing the cleavage site to the active site on the catalytic domain of the protein. However, given the importance of the DNA deformation for specific recognition between Tc3A and its transposon, the DNA conformation could be quite different in other Tc1/*mariner* elements. Of the two DNA bends, the bend at the TATA sequence, which is both closer to the cleavage site and somewhat conserved in *C.elegans* Tc1 DNA, is more likely to play a role than the bending in the G:C stretch which is not conserved in Tc1 at all.

For actual synapse formation, multimerization of the protein is likely to be important. On the basis of sequence comparisons of several fish Tc1-like transposable elements,

a conserved, so-called leucine zipper motif has been indicated (Ivics *et al.*, 1996), with a suggested importance for dimerization/oligomerization of the proteins. However, the sequences referred to do not fold into a leucine zipper in the three-dimensional structure of Tc3A, but form part of the hydrophobic core of the protein, spread over the three helices.

In the crystal structure, we observe a crystallographic dimer of the Tc3A DNA-binding domains, in which each monomer binds one DNA molecule. It is plausible that this dimer provides the first glimpse of the way in which the Tc1 family brings the two transposon DNA ends together. Once the two ends are synapsed with a multimer of the protein, cleavage of the transposon can take place. In bacteriophage Mu, this DNA cleavage occurs *in trans*, i.e. the monomer that is bound to a DNA strand cleaves the opposite strand (Aldaz *et al.*, 1996; Savilahti and Muzuuchi, 1996; Yang *et al.*, 1996). In the crystallographic Tc3 dimer, the two DNA monomers are found in a parallel orientation, with the C-terminal regions of the Tc3A-N pointing in the same direction, indicating roughly where the second DNA recognition and catalytic domains will be located. It is possible that the catalytic domains of Tc3A also act *in trans*, with each monomer cutting the opposite strand of the DNA, thus ensuring that simultaneous cleavage of the two transposon ends takes place.

Materials and methods

Expression and purification

The N-terminal 65 amino acids of Tc3A with a C-terminal His tag were expressed in *Escherichia coli* from the pET3c-derived vector pRP1200. This vector was constructed from pSDC328 (Colloms *et al.*, 1994), by cloning the annealed oligonucleotides RK1 (5' CCATCACCATCACCA-TCACTAGA 3') and RK2 (5' AGCTTCTAGTGATGGTGATGGTGATGGTGA-TGG 3') into the *SpeI*-*HindIII* site to add the C-terminal His tag. Cultures of *E.coli* strain BL21(DE3)pLysS transformed with pRP1200 were grown at 37°C and protein expression was induced (at OD₆₀₀ = 0.6) by the addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) 4 h prior to harvesting. The cells were resuspended in lysis buffer (6 M guanidinium chloride, 100 mM NaCl, 50 mM Tris pH 7.5, 1 mM β-mercaptoethanol) and sonicated. Insoluble material was removed by centrifugation.

The supernatant of the lysate, to which 5 mM imidazole was added, was allowed to bind to Ni-NTA resin (Qiagen) overnight. The resin was loaded into a column and washed with five column volumes of lysis buffer containing 20 mM imidazole. The protein was eluted with 250 mM

imidazole in lysis buffer and 1% β -mercaptoethanol was added to the pooled fractions. The protein was then reduced fully, denatured at 37°C for 2 h and subsequently purified further on a gel filtration column (Superdex 75, High load 26/60, Pharmacia), in lysis buffer with 10 mM β -mercaptoethanol. The protein was renatured by dialyzing against 50 mM sodium acetate pH 4.8, 0.4 M NaCl, 10 mM β -mercaptoethanol, 10 mM EDTA at 4°C. The renatured protein was loaded on a cation exchange column (S-Sepharose Fast Flow, Pharmacia), washed with three column volumes of 0.6 M NaCl and eluted with 1.5 M NaCl. The pure protein (as judged from an SDS-PAGE analysis) was dialyzed against 20 mM sodium acetate pH 4.8, 0.1 M NaCl, 10 mM β -mercaptoethanol and concentrated to 5–7 mg/ml. The renatured protein gives a single peak on an analytical gel filtration column (Superdex 75, Pharmacia) in 20 mM sodium acetate pH 4.8, 0.2 M NaCl, 10 mM β -mercaptoethanol. This peak corresponds to a mol. wt of 25 kDa (the size of the monomer is 8.3 kDa). N-terminal sequencing of the purified protein showed that the N-terminal methionine was absent.

The 20- and 21-DNA oligomers (Figure 1) were synthesized on a

1 μ mol scale (Isogen) and purified on a 6 ml Resource Q column (Pharmacia) in 10 mM NaOH with an NaCl gradient. Initially, heavy precipitate formation upon mixing of protein and DNA occurred. Varying the salt concentration, pH or buffer did not affect precipitation, but variations in dilution and the ratio of DNA and protein prevented complete precipitation. The oligomers were annealed and mixed in a 1:1 molar ratio with the purified protein. Prior to mixing, the protein and DNA were diluted in 10 mM HEPES pH 7.5, 50 mM NaCl, 10 mM β -mercaptoethanol. The complex was 50-fold reconcentrated in Centricons to an $OD_{260\text{ nm}, 1\text{ cm}}$ of 70 (4.5 mg/ml complex).

Crystallization

Crystals were grown by the vapor diffusion method, from various precipitants and with different DNA oligomers, but most diffracted poorly. The best crystals (0.2×0.2×0.1 mm³ in size) used in this study grew within a week in hanging drops at 4°C. The reservoir contained 100 mM NaCl, 20 mM CaCl₂, 13–15% MPD and 10 mM dithiothreitol (DDT) buffered with 50 mM sodium acetate at pH 5.5. The drop contained 1 μ l of reservoir and 1 μ l of 4.5 mg/ml protein–DNA complex stock solution prior to equilibrium.

Data collection and processing

A diffraction data set of a native crystal was collected at 4°C using synchrotron radiation in beamline X11 at the EMBL Outstation at DESY in Hamburg, using a MARresearch image plate. A second crystal was soaked in cryo-protectant (mother liquor with the MPD concentration increased to 45%) for several minutes at 4°C and flash frozen in a nitrogen stream. A native data set was collected at –160°C in beamline BW7A. All data sets are anisotropic. While the resolution limit along the direction of the crystallographic *b*-axis is 2.85 Å and 2.45 Å for the 4 and –160°C native data sets respectively, the diffraction limits along the *a*- and *c*-direction are ~0.5 Å lower. The intensities were integrated using the program DENZO (Otwinowski, 1993). Regions with absent reflections in the *a*- and *c*-direction due to the anisotropy were excluded

Table II. Statistics on refinement

Data set	Native2
Resolution limits (Å)	20–2.45
No. of reflections in work set/test set	7708/372
No. of non-H protein atoms/DNA atoms	404/834
No. of water molecules	49
<i>R</i> -factor ^a (%) / <i>R</i> _{free} ^b (%)	23.4/31.8
R.m.s.d. in bond lengths (Å)	0.014
R.m.s.d. in bond angles (°)	2.368

^a $\sum||F_o| - |F_c|| / \sum|F_o|$.

^bBrunger, 1992b

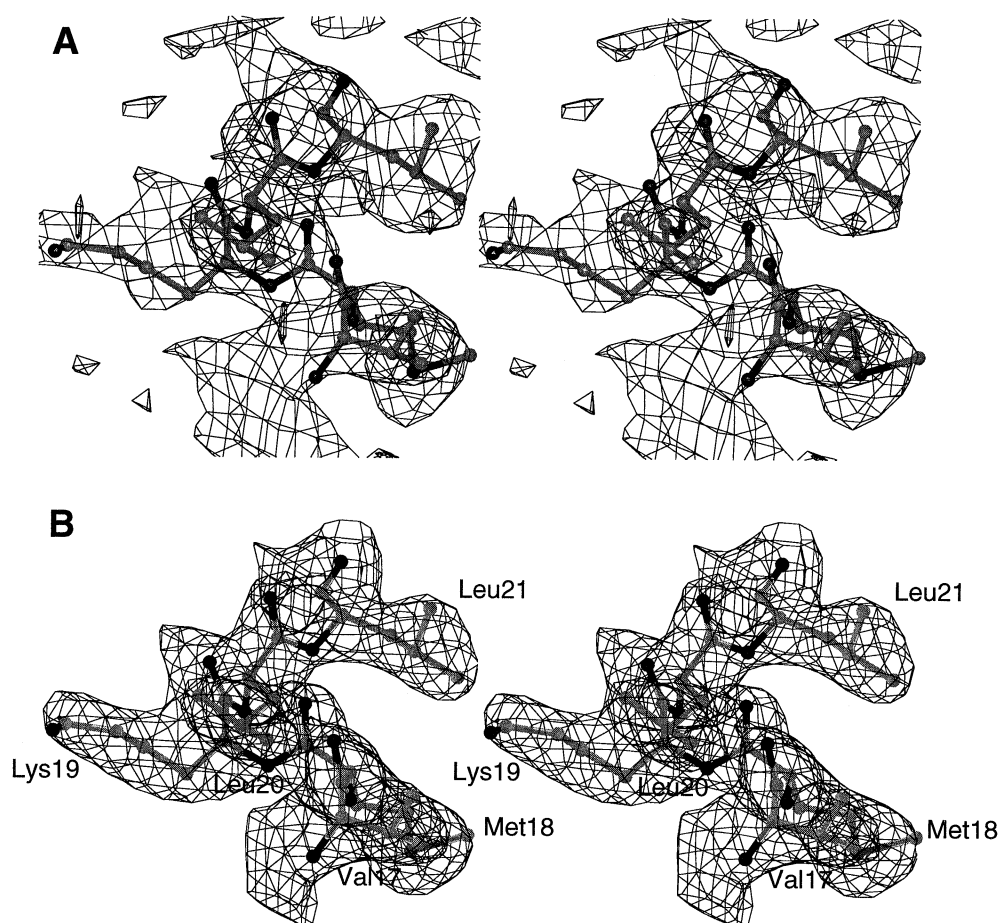


Fig. 7. Stereo diagram of the electron density for (A) the solvent-flattened MIRAS map and (B) the final $2F_o - F_o$ map, both contoured at 1σ . Carbon atoms are gray; oxygen, nitrogen and sulfur atoms are black.

from integration in the cryo data set. Reflections were scaled and merged using the program SCALEPACK (Otwinowski, 1993). Intensities were converted to amplitudes using the TRUNCATE program (French and Wilson, 1978). The statistics are given in Table I.

Heavy atom derivatives and phasing

Heavy atom derivatives were prepared by soaking crystals in solutions containing 1 mM methylmercury chloride or 3 mM mercury thiocyanate in mother liquor. A DNA oligomer in which one thymine (T9) was substituted for 5-iodouracil was also used to obtain crystals. Data sets for the first two derivatives were collected on the home-source (a Nonius FR591 rotating anode generator) with a Dip2000 image plate and for iodinated DNA in beamline X31 at the EMBL Outstation in Hamburg. Heavy atom-binding sites were determined from the difference Patterson map and confirmed in the anomalous difference Patterson map. The 4°C native data set was used as the reference set throughout phasing. All derivatives contained only one heavy atom per asymmetric unit. The multiple isomorphous replacement method with anomalous data (MIRAS) was used. MIRAS phases were calculated using the program PHASES (Furey and Swaminathan, 1990). Reflections with $F < 3\sigma(F)$ were rejected from phase refinement. The data were cut off at the resolution (see Table I) where the phasing power dropped below 1.0. Heavy atom refinement statistics are given in Table I. The solvent-flattening technique as implemented in PHASES was applied to improve the initial MIRAS map.

Model building and refinement

Most of the DNA and two α -helices of the protein were recognized in the resulting electron density map and built into the electron density using the program O (Jones *et al.*, 1991; Kleywegt and Jones, 1996). This model was translated to the slightly different unit cell of the frozen crystal. Five percent of the data were set aside as a test set (Brunger, 1992b). Rigid body and simulated annealing refinement against the cryo native data was performed with the XPLOR program (Brunger, 1992a). The R_{free} dropped from 55 to 44% with an R -factor of 33%. Still, the third, C-terminal, α -helix and most of the side chains were not recognizable in a $2F_o - F_c$ or SIGMAA-weighted (Read, 1986) electron density map. The partial model was refined with the program ARP (Lamzin and Wilson, 1993, 1997) using loose stereochemical restraints for the DNA and the existing protein region. In that procedure, reciprocal maximum-likelihood refinement using REFMAC (Murshudov *et al.*, 1997) was followed by automatic model update by ARP, for several cycles. ARP was instructed to introduce atoms as close as 1.2 Å from the existing protein or DNA atoms, to build missing model parts. After 50 cycles, the R_{free} dropped to 34%. The $2mF_o - DF_c$ SIGMAA-weighted electron density map showed clear density for most missing parts. This map was used to add the side chains and C-terminal residues to the XPLOR model. Amino acid residues 2–52 and the complete DNA were built and refined using simulated annealing ($R_{\text{free}} = 32.2\%$). Final refinement steps were performed with the refinement program TNT (Tronrud, 1996), which resulted in a final model with an R -factor of 23.4% and an R_{free} of 31.8% at 2.45 Å (refinement statistics are given in Table I).

The R -factor and R_{free} (Brunger, 1992b) are relatively high. This could be caused by the presence of the flexible C-terminus and/or the incompleteness of the data due to anisotropy. The model has good stereochemistry (Table II). The PROCHECK program (Morris *et al.*, 1992) showed that the main-chain dihedral angles for the majority of the residues lie in the most favorable regions of the Ramachandran plot and five in the additional allowed regions. The mean B -factor for all atoms has a high value of 48 Å² (protein atoms 41 Å², DNA atoms 51 Å², solvent atoms 53 Å²). The predicted B -factor from a Wilson plot (1949) is even higher, 58 Å². A stereo view of representative electron density is shown in Figure 7. The electron density of a $2F_o - F_c$ map contoured at 1σ covers the model almost completely. For a few side chains (with high B -factors) on the surface of the protein, no electron density is present, due to flexibility. The electron density at the end of the side chain of Arg3 is not very clear. This puts some doubt on the relative positions of Pro2 and Arg3, but refinement of the two alternative positions gave the best R -factors and stereochemistry with the current positions. The conformation of the last visible amino acid at the C-terminus, Ser52, is not completely clear in the electron density map. Two bases at one end of the double-stranded DNA oligomer are flipped outward. These do not fit in the electron density completely. The rest of the model fits very well in the electron density. The coordinates are submitted to the Brookhaven Protein Data Bank (1tc3).

Acknowledgements

We thank Dr Alexei Teplyakov for assisting with data collection, the members of the protein structure group at the NKI and Luca Jovine (MRC Laboratory of Molecular Biology, Cambridge) for helpful discussions. R.F.K. is supported by grant 700-35-210 from NWO/SON. A.P. is an EMBO fellow, ALTF-215/1995. We thank the European Union for support of the work at EMBL Hamburg through the HCMP Access to Large Scale Facilities grant, Contract Number CHGE-CT93-0040.

References

- Aldaz, H., Schuster, E. and Baker, T.A. (1996) The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell*, **85**, 257–269.
- Ariyoshi, M., Vassilyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H. and Morikawa, K. (1994) Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. *Cell*, **78**, 1063–1072.
- Baker, T.A. and Luo, L. (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc. Natl Acad. Sci. USA*, **91**, 6654–6658.
- Barlow, D.J. and Thornton, J.M. (1983) Ion-pairs in proteins. *J. Mol. Biol.*, **168**, 867–885.
- Beamer, L.J. and Pabo, C.O. (1992) Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.*, **227**, 177–196.
- Brunger, A.T. (1992a) *X-PLOR Manual. Version 3.1*. Yale University Press, New Haven, CT.
- Brunger, A.T. (1992b) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.*, **4**, 567–577.
- Chaconas, G., Lavoie, B.D. and Watson, M.A. (1996) DNA transposition: jumping gene machine, some assembly required. *Curr. Biol.*, **6**, 817–820.
- Colloms, S.D., van Luenen, H.G.A.M. and Plasterk, R.H.A. (1994) DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase. *Nucleic Acids Res.*, **22**, 5548–5554.
- Czerny, T., Schaffner, G. and Busslinger, M. (1993) DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev.*, **7**, 2048–2061.
- Doak, T.G., Doerder, F.P., Jahn, C.L. and Herrick, G. (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common 'D35E' motif. *Proc. Natl Acad. Sci. USA*, **91**, 942–946.
- Doolittle, R.F., Feng, D.-F., Johnson, M.S., McClure, M.A. (1989) Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.*, **64**, 1–30.
- Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.
- Eijkelenboom, A.P.A.M. (1997) The solution structure of the N-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr. Biol.*, in press.
- Epstein, J., Cai, J., Glaser, T., Jepeal, L. and Maas, R. (1994a) Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J. Biol. Chem.*, **269**, 8355–8361.
- Epstein, J.A., Glaser, T., Cai, J., Jepeal, L., Walton, D.S. and Maas, R.L. (1994b) Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev.*, **8**, 2022–2034.
- Fayet, O., Ramond, P., Polard, P., Prere, M.F. and Chandler, M. (1990) Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol. Microbiol.*, **4**, 1771–1777.
- Feng, J.-A., Johnson, R.C. and Dickerson, R.E. (1994) Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science*, **263**, 348–355.
- Franz, G., Loukeris, T.G., Dialektaki, G., Thompson, C.R.L. and Charalambos, S. (1994) Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc. Natl Acad. Sci. USA*, **91**, 4746–4750.
- French, S. and Wilson, K. (1978) On the treatment of negative intensity observations. *Acta Crystallogr.*, **A34**, 517–525.

- Furey,W. and Swaminathan,S. (1990) PHASES—a program package for the processing and analysis of diffraction data from macromolecules. In *Abstracts of the American Crystallographica Association Meeting*, **18**, p. 73.
- Hol,W.G.J., van Duijnen,P.T. and Berendsen,H.J.C. (1978) The α -helix dipole and the properties of proteins. *Nature*, **273**, 443–446.
- Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
- Ivics,I., Izsavak,Z., Minter,A. and Hackett,P.B. (1996) Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl Acad. Sci. USA*, **93**, 5008–5013.
- Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Joshua-Tor,L., Frolow,F., Appella,E., Hope,H., Rabinovich,D. and Sussman,J.L. (1992) Three-dimensional structures of bulge-containing DNA fragments. *J. Mol. Biol.*, **225**, 397–431.
- Katayanagi,K., Miyagawa,M., Matsushima,M., Ishikawa,M., Kanaya,S., Ikehara,M., Matsuzaki,T. and Morikawa,K. (1990) Three-dimensional structure of ribonuclease H from *E.coli*. *Nature*, **347**, 306–309.
- Katayanagi,K., Miyagawa,M., Matsushima,M., Ishikawa,M., Kanaya,S., Nakamura,H., Ikehara,M., Matsuzaki,T. and Morikawa,K. (1992) Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.*, **223**, 1029–1052.
- Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.
- Klemm,J.D., Rould,M.A., Aurora,R., Herr,W. and Pabo,C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, **77**, 21–32.
- Kleywegt,G.J. and Jones,T.A. (1996) Efficient rebuilding of protein structures. *Acta Crystallogr.*, **D52**, 829–832.
- Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G. and Skalka,A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.*, **12**, 2331–2338.
- Lampe,D.J., Churchill,M.E.A. and Robertson,H.M. (1996) A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J.*, **15**, 5470–5479.
- Lamzin,V.S. and Wilson,K.S. (1993) Automated refinement of protein models. *Acta Crystallogr.*, **D49**, 129–147.
- Lamzin,V.S. and Wilson,K.S. (1997) Automated refinement for protein crystallography. *Methods Enzymol.*, **277**, in press.
- Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
- Luisi,B. (1995) DNA–protein interaction at high resolution. In Lilley,D.M.J. (ed.), *DNA–Protein: Structural Interactions*. Oxford University Press, Oxford, UK, pp. 1–48.
- Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins*, **12**, 345–364.
- Murphy,J.E. and Goff,S.P. (1992) A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase *in vivo*. *J. Virol.*, **66**, 5092–5095.
- Murshudov,G.N., Vagin,A.A. and Dodson,E.S. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240–255.
- Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Ogata,K. *et al.* (1995) Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Struct. Biol.*, **2**, 309–320.
- Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
- Plasterk,R.H.A. (1995) Mechanisms of DNA transposition. In Sherratt,D.J. (ed.), *Mobile Genetic Elements*. Oxford University Press, Oxford, UK, pp. 18–37.
- Plasterk,R.H.A. (1996) The Tc1/*mariner* transposon family. *Curr. Top. Microbiol. Immunol.*, **204**, 125–143.
- Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr.*, **A42**, 140–149.
- Rice,P. and Mizuuchi,K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209–220.
- Rost,B. and Sander,C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.*, **232**, 584–599.
- Savilahti,H. and Mizuuchi,K. (1996) Mu transpositional recombination: donor DNA cleavage and strand transfer in *trans* by the Mu transposase. *Cell*, **85**, 271–280.
- Stofer,E. and Lavery,R. (1994) Measuring the geometry of DNA grooves. *Biopolymers*, **34**, 337–346.
- Strauss,J.K. and Maher,L.J. (1994) DNA bending by asymmetric phosphate neutralization. *Science*, **266**, 1829–1834.
- Suzuki,M. (1993) Common features in DNA recognition helices of eukaryotic transcription factors. *EMBO J.*, **12**, 3221–3226.
- Suzuki,M. and Gerstein,M. (1995) Binding geometry of α -helices that recognize DNA. *Proteins*, **23**, 525–535.
- Travers,A. (1993) *DNA–Protein Interactions*. Chapman and Hall, London, UK.
- Tronrud,D.E. (1996) Knowledge-based B-factor restraints for the refinement of proteins. *J. Appl. Crystallogr.*, **29**, 100–104.
- Tullius,T. (1995) Homeodomains: together again for the first time. *Structure*, **15**, 1143–1145.
- van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1993) Mobilization of quiet, endogenous Tc3 transposons of *Caenorhabditis elegans* by forced expression of Tc3 transposase. *EMBO J.*, **12**, 2513–2520.
- van Luenen,H.G.A.M., Colloms,S.D. and Plasterk,R.H.A. (1994) The mechanism of transposition of Tc3 of *C.elegans*. *Cell*, **79**, 293–301.
- Vos,J.C. and Plasterk,R.H.A. (1994) Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA-binding domain. *EMBO J.*, **13**, 6125–6132.
- Vos,J.C., van Luenen,H.G.A.M. and Plasterk,R.H.A. (1993) Characterization of the *Caenorhabditis elegans* Tc1 transposase *in vivo* and *in vitro*. *Genes Dev.*, **7**, 1244–1253.
- Vos,J.C., De Baere,I. and Plasterk,R.H.A. (1996) Transposase is the only nematode protein required for *in vitro* transposition of Tc1. *Genes Dev.*, **10**, 755–761.
- Wilson,A.J.C. (1949) The probability distribution of X-ray intensities. *Acta Crystallogr.*, **2**, 318–321.
- Wilson,D.S., Guenther,B., Desplan,C. and Kuriyan,J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709–719.
- Wintjens,R. and Rooman,M. (1996) Structural classification of HTH DNA-binding domains and protein–DNA interaction modes. *J. Mol. Biol.*, **262**, 294–313.
- 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.
- Yang,J.-Y., Jayaram,M. and Harshey,R.M. (1996) Positional information within the Mu transposase tetramer: catalytic contributions of individual monomers. *Cell*, **85**, 447–455.
- Yang,W. and Steitz,T. (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*, **15**, 131–134.
- Yang,W., Hendrickson,W.A., Crouch,R.J. and Satow,Y. (1990) Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science*, **249**, 1398–1405.

Received on June 18, 1997; revised on July 10, 1997