Acta Crystallographica Section F
**Structural Biology and Crystallization Communications**
ISSN 1744-3091

# Crystallization of a Mos1 transposase–inverted-repeat DNA complex: biochemical and preliminary crystallographic analyses

**Julia M. Richardson,* David J. Finnegan and Malcolm D. Walkinshaw**

School of Biological Sciences, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland

Correspondence e-mail: jrichard@staffmail.ed.ac.uk

Received 20 March 2007
Accepted 17 April 2007

A complex formed between Mos1 transposase and its inverted-repeat DNA has been crystallized. The crystals diffract to 3.25 Å resolution and exhibit monoclinic ($P2_1$) symmetry, with unit-cell parameters $a$ = 120.8, $b$ = 85.1, $c$ = 131.6 Å, $\beta$ = 99.3°. The X-ray diffraction data display noncrystallographic twofold symmetry and characteristic dsDNA diffraction at ∼3.3 Å. Biochemical analyses confirmed the presence of DNA and full-length protein in the crystals. The relationship between the axis of noncrystallographic symmetry, the unit-cell axes and the DNA diffraction pattern are discussed. The data are consistent with the previously proposed model of the paired-ends complex containing a dimer of the transposase.

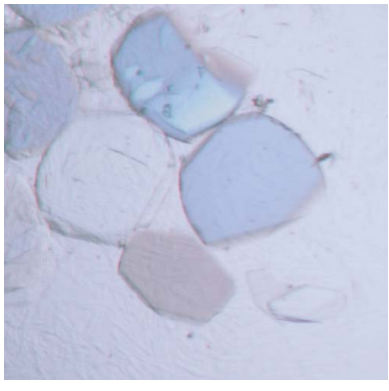
## 1. Introduction

Transposition of mobile DNA elements generates genetic diversity. Advantage has been taken of this process in the development of transposable elements as tools for insertional mutagenesis (Granger *et al.*, 2004), transgenesis (Robinson *et al.*, 2004) and as gene-therapy vectors (Ivics & Izsvak, 2006; Wilson *et al.*, 2007).

The eukaryotic *mariner* transposon *Mos1* is a member of the Tc1/*mariner* family of transposable elements and was first isolated from *Drosophila mauritiana*. Tc1/*mariner* elements are found in the genomes of most eukaryotes, including the human genome (Robertson & Zumpano, 1997). *Mos1* moves by a cut-and-paste mechanism involving a series of DNA-hydrolysis and transesterification reactions related to the integration of retroviruses (such as HIV-1) and the RAG1/2 recombinase-mediated rearrangement of immunoglobulin genes.

The sole requirement for *Mos1* transposition *in vitro* is the Mos1 transposase, a 40 665 Da enzyme encoded by a single gene within the 1.3 kbp element. In the first step of transposition, the transposase recognizes the 28-base-pair DNA inverted repeats (IRs) at either end of the element. This interaction is sequence-specific and is mediated by the N-terminal DNA-binding domain of the transposase (Zhang *et al.*, 2001; Augé-Gouillou *et al.*, 2001). Subsequently, the transposon ends are brought together to form a paired-ends complex (PEC) and the transposon is excised from flanking DNA. It has previously been shown (Dawson & Finnegan, 2003) that the 5′-end of the nontransferred strand (NTS) is nicked three bases within the element, whereas the 3′-end of the transferred strand (TS) is cleaved precisely at the junction of the transposon and flanking DNA. In the final steps, target DNA is recruited to the PEC and the transposon DNA is inserted at the 5′-end of a TA dinucleotide.

The structure of the bacterial Tn5 transposase with transposon DNA (Davies *et al.*, 2000) provided the first picture of a transposition intermediate. However, less is known of the structures of eukaryotic transposases or the complexes that they form with DNA. The structure of the bipartite DNA-binding domain of Tc3 transposase was determined in complex with IR DNA (Watkins *et al.*, 2004) and more recently the structure of the apo form of the N-terminally



truncated eukaryotic hAT transposase Hermes was solved (Hickman *et al.*, 2005). Previously, we determined the structure of the catalytic domain of Mos1 transposase (Richardson *et al.*, 2006) and showed that two divalent metal ions could bind in the active site. Invoking a two-metal mechanism for DNA hydrolysis, we argued that PEC intermediates containing protein dimers were sufficient to catalyse all the steps of *Mos1* transposition.

Here, we report the crystallization of a complex formed between full-length Mos1 transposase and its inverted-repeat DNA. The crystals diffract X-rays to 3.25 Å and the data display twofold noncrystallographic symmetry (NCS). These results are discussed in relation to proposed models of the PEC.

## 2. Materials and methods

### 2.1. Preparation of the protein–DNA complex

The soluble single mutant (T216A) of the Mos1 transposase protein was expressed and purified as described previously (Richardson *et al.*, 2004) and concentrated to 448 µ*M* in 20 m*M* Tris pH 7.5, 0.25 *M* KCl and 1 m*M* DTT.

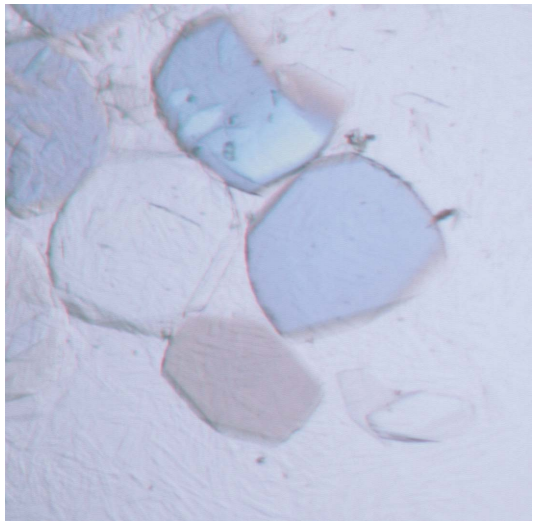


**Figure 1**
Crystals of the Mos1 transposase–inverted-repeat DNA complex.

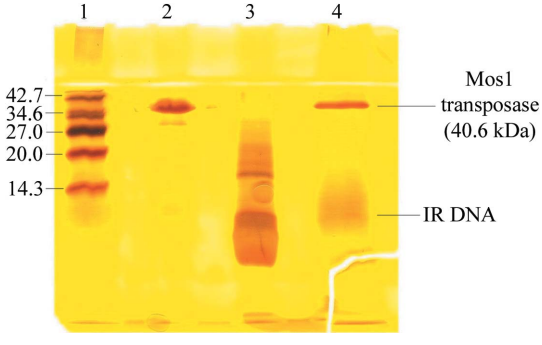



**Figure 2**
Silver-stained SDS polyacrylamide gel (15%) of dissolved crystals. Lane 1, molecular-weight markers; lane 2, purified Mos1 transposase; lane 3, IR DNA; lane 4, washed and dissolved crystals. The molecular weight of each protein in the molecular-weight markers is shown (in kDa) to the left of the gel. Diffusion of the DNA band is more pronounced in lane 4 than lane 3 because of the lower concentration of the sample and the presence of metal ions in the crystallization conditions.

Inverted-repeat oligonucleotides with sequences corresponding to the product of the double-strand break were chosen. The first strand had the full 28-base recognition sequence of the transferred strand of the right-hand IR, 5′-AAACGACATTTCATACTTGTACACC-TGA-3′, corresponding to the cleaved TS. The second strand had the 5′-modified 25-base sequence 5′-$PO_4$-GGTGTACAAGTATGAAA-TGTCGTTT-3′, corresponding to the cleaved NTS of the IR. The molecular weights of these strands are 8.5 and 7.8 kDa, respectively. These sequences, once annealed, generate double-stranded DNA (dsDNA) that is blunt at the inside end of the transposon sequence but staggered by three bases at the outside end, giving a structure equivalent to that of excised transposon DNA.

Reverse-phase HPLC-purified synthetic oligonucleotides were purchased from Eurogentec (Belgium) and confirmed by mass spectrometry to be 90–95% pure. Oligonucleotides were dissolved to 1 m*M* in TEN buffer (10 m*M* Tris pH 8.0, 1 m*M* EDTA, 0.1 *M* NaCl) and equal volumes of the TS and NTS DNA were mixed and annealed. The final concentration of dsDNA was 0.5 m*M*.

To prepare the protein–DNA complex, protein was added to excess dsDNA to give a final protein:DNA molar ratio of 1:1.5. Thus,

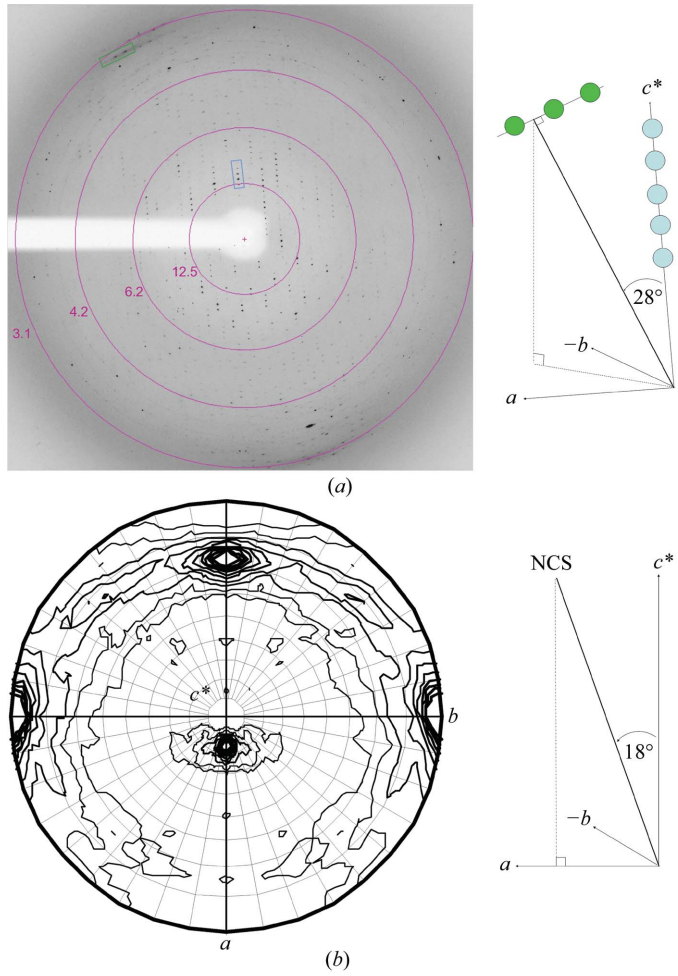



**Figure 3**
(*a*) Diffraction pattern taken over a 1° rotation about $\varphi$ with resolution rings marked in Å. A row of three strong Bragg reflections at 3.28–3.31 Å, arising from stacked bases in dsDNA, is highlighted by a green box. These reflections have indices (11 −11 32), (10 −10 33) and (9 −9 34). The row of reflections parallel to *c** with Miller indices (0 0 10), (0 0 11), (0 0 12), (0 0 13) and (0 0 14) is marked with a blue box. (*b*) Self-rotation function at $\chi$ = 180°. The *a*, *b* and *c** axes are marked. Peaks on the *b* axis indicate the crystallographic twofold screw symmetry, whereas peaks in the *ac** plane arise from noncrystallographic twofold symmetry.

**Table 1**
Diffraction data statistics.

Values in parentheses are for the highest resolution bin.

| | |
|---|---|
| Resolution (Å) | 29.00–3.25 (3.42–3.25) |
| Total observations | 108827 (14741) |
| Unique observations | 39353 (5508) |
| $R_{merge}$ (%) | 0.164 (0.441) |
| Completeness (%) | 94.2 (90.8) |
| $I/\sigma(I)$ | 6.9 (1.6) |
| Multiplicity | 2.8 (2.7) |

112 µl protein solution was added in 10 µl aliquots to 150 µl dsDNA. The concentration of the resulting complex was 95.4 µ*M*, if it is assumed that it comprised a dimer of protein and two IR DNA molecules.

### 2.2. Crystallization

Crystals were grown by vapour diffusion in hanging drops at 290 K in 24-well Linbro plates. The well solution consisted of 100 m*M* sodium citrate pH 6.0, 0.2 *M* KCl, 100 m*M* ammonium acetate and 5 m*M* $MgCl_2$. The hanging drop consisted of 2 µl fresh protein–DNA complex solution and 2 µl well solution. Crystals grew to maximum dimensions of 0.2 × 0.2 × 0.05 mm after 7–10 d and are shown in Fig. 1. Prior to data collection at 100 K, crystals were briefly immersed in cryoprotectant solution comprising 10% glycerol, 70% ammonium sulfate and 100 m*M* sodium citrate pH 6.0.

### 2.3. SDS–PAGE of dissolved crystals

Crystals were harvested from the hanging drop and washed three times in well solution before being dissolved in loading buffer for SDS–PAGE. Control samples of purified Mos1 transposase and double-stranded IR DNA were also prepared. All samples were heated to 373 K for 10 min and run on a 15% SDS polyacrylamide gel with molecular-weight markers (Fig. 2). The gel was silver-stained to visualize the DNA and protein using a protocol kindly provided by Dr A. Hickman (NIDDK, NIH). The gel clearly shows that the crystals contained both DNA and full-length protein, indicating that the tendency of the apo protein to undergo proteolysis (Richardson *et al.*, 2006) had been thwarted by the presence of DNA.

### 2.4. X-ray data collection

X-ray diffraction data were collected at the ESRF (beamline BM14) at a wavelength of 0.95 Å using a MAR CCD detector. All data were collected using a $\varphi$ scan with a step size of 1°. Indexing and integrating were performed with *MOSFLM* (Leslie, 1992) and the data were scaled with *SCALA* within the *CCP*4 software suite (Collaborative Computational Project, Number 4, 1994). The self-rotation function was generated using *MOLREP* (*CCP*4 suite) using data in the resolution range 25.0–4.0 Å and a search radius of 30 Å.

## 3. Results

### 3.1. Diffraction data

The crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters $a$ = 120.8, $b$ = 85.1, $c$ = 131.6 Å, $\beta$ = 99.3°. Diffraction data statistics are presented in Table 1 and one diffraction image is shown in Fig. 3(*a*). Strong diffraction peaks arising from the stacked bases of dsDNA are visible in the top left-hand corner at 3.32–3.2 Å resolution. The angle between the normal to the plane of this diffraction and the 00*l* row is 28°. This corresponds to the angle that the helical axis of DNA (perpendicular to the plane of the stacked bases) is oriented from the *c** axis.

Self-rotation analysis revealed the presence of one twofold NCS axis; no higher orders of symmetry were present. The self-rotation function at $\chi$ = 180° is shown in Fig. 3(*b*). The NCS axis is oriented 18° from *c** towards *a* in the *ac** plane (Fig. 3*b*). The NCS axis and the DNA helical axis are approximately parallel.

The diffraction pattern shows significant anisotropy: diffraction extends to ∼3.1 Å in the direction approximately parallel to the DNA helical axis, but is limited to 3.4 Å in the orthogonal directions.

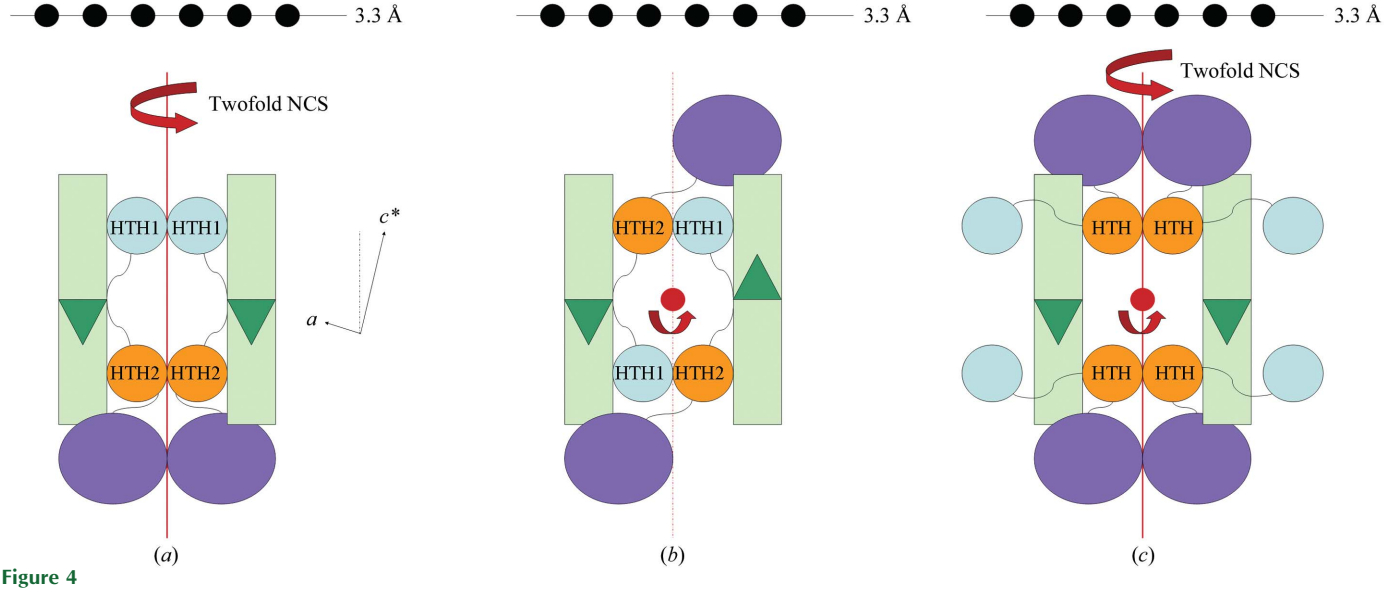



**Figure 4**
Schematic representations of three possible models of the Mos1 transposase paired-ends complex. Transposase dimers are bound *via* two helix–turn–helix motifs (HTH1 and HTH2) to (*a*) parallel or (*b*) antiparallel IR DNA molecules. Alternatively, parallel IR DNA molecules are bound in a tetrameric assembly (*c*) in which each of the four transposases contain only one helix–turn–helix motif (HTH) and two transposases bind to each palindromic IR. The transposase is represented by three circles: two DNA-binding motifs (pale blue and orange) and a catalytic domain (purple). The IR DNA is shown by a green rectangle. The twofold axes of symmetry that would arise in each assembly are marked by red lines or dots and filled arrows and the DNA diffraction at 3.3 Å is denoted by a line of black circles.

If the crystallized complex is assumed to contain two copies of the full-length transposase and two copies of IR DNA, the total molecular weight is 117 kDa. The Matthews coefficient suggests that either one or two copies of the PEC are present in the asymmetric unit, with $V_M$ values of 5.71 Å$^3$ Da$^{-1}$ (78.5% solvent) or 2.85 Å$^3$ Da$^{-1}$ (56.9% solvent), respectively.

## 4. Discussion

The diffraction data from the Mos1–IR DNA complex are consistent with our previously proposed model of the Mos1 transposase PEC, which contains two parallel IR DNA molecules and a dimer of transposase (Richardson *et al.*, 2006). Such an assembly would have twofold symmetry which is approximately parallel to the DNA helical axis (as indicated in Fig. 4*a*). If the DNA adopted an antiparallel arrangement (as in the Tn5 synaptic complex structure), the twofold symmetry axis would be perpendicular to the DNA helical axis (Fig. 4*b*). Others have proposed that IR ends are brought together by a transposase tetramer (Fig. 4*c*), with two binding sites for transposase on each palindromic IR DNA (Augé-Gouillou *et al.*, 2005). This assembly would be expected to display two perpendicular twofold-symmetry axes, one of which was parallel to the DNA helical axis.

Molecular replacement using the structure of the catalytic domain of Mos1 transposase (PDB code 2f7t) as the search model in *Phaser* (Storoni *et al.*, 2004) has yielded interpretable electron-density maps. Building and refinement of the structure is in progress.

We are grateful to the staff at the ESRF beamline BM14 for assistance with data collection and the MRC for funding data collection (grant No. BM14U-846). We also thank Dr Marjorie Harding for helpful comments on the manuscript and Dr Alison Burgess Hickman (NIDDK, NIH) for the silver-staining protocol. This work was funded by the Wellcome Trust (grant No. 074522).

## References

Augé-Gouillou, C., Brillet, B., Hamelin, M.-H. & Bigot, Y. (2005). *Mol. Cell. Biol.* **25**, 2861–2870.
Augé-Gouillou, C., Hamelin, M.-H., Demattei, M.-V., Periquet, G. & Bigot, Y. (2001). *Mol. Genet. Genomics*, **265**, 58–65.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, D. R., Goryshin, I. Y. & Reznikoff, W. S. (2000). *Science*, **289**, 77–85.
Dawson, A. & Finnegan, D. J. (2003). *Mol. Cell*, **11**, 225–235.
Granger, L., Martin, E. & Ségalat, L. (2004). *Nucleic Acids Res.* **32**, e117.
Hickman, A. B., Perez, Z. N., Zhou, L., Musingarimi, P., Ghirlando, R., Hinshaw, J. E., Craig, N. L. & Dyda, F. (2005). *Nature Struct. Mol. Biol.* **12**, 715–721.
Ivics, Z. & Izsvak, Z. (2006). *Curr. Gene Ther.* **6**, 593–607.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Richardson, J. M., Dawson, A., O'Hagan, N., Taylor, P., Finnegan, D. J. & Walkinshaw, M. D. (2006). *EMBO J.* **25**, 1324–1334.
Richardson, J. M., Zhang, L., Marcos, S., Finnegan, D. J., Harding, M. M., Taylor, P. & Walkinshaw, M. D. (2004). *Acta Cryst.* D**60**, 962–964.
Robertson, H. M. & Zumpano, K. L. (1997). *Gene*, **205**, 203–217.
Robinson, A. S., Franz, G. & Atkinson, P. W. (2004). *Insect Biochem. Mol. Biol.* **34**, 113–120.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Watkins, S., van Pouderoyen, G. & Sixma, T. K. (2004). *Nucleic Acids Res.* **32**, 4306–4312.
Wilson, M. H., Coates, C. J. & George, A. L. Jr (2007). *Mol. Ther.* **15**, 139–145.
Zhang, L., Dawson, A. & Finnegan, D. J. (2001). *Nucleic Acids Res.* **29**, 3566–3575.