Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification and preliminary crystallographic studies of a single-point mutant of Mos1 mariner transposase

Julia M. Richardson,* Lei Zhang,‡ Severine Marcos,§ David J. Finnegan, Marjorie M. Harding, Paul Taylor and Malcolm D. Walkinshaw

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

‡ Present address: Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, CA 94609-1673, USA.
§ Present address: Equipe Cycle et Determination Cellulaire, UMR 762, Universite Paris VI, 9 Quai Saint Bernard, 75005 Paris, France.

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

Received 20 January 2004
Accepted 17 February 2004

A soluble single-point mutant of full-length Mos1 mariner transposase (MW = 40.7 kDa) has been overexpressed in *Escherichia coli*, purified to 95% homogeneity and crystallized. This provides the first example of the crystallization of a eukaryotic transposase. The native crystals diffract to 2.5 Å resolution and show tetragonal symmetry, with unit-cell parameters $a = b = 44.5$, $c = 205.6$ Å. Multiple-wavelength anomalous data from a selenomethionyl form of the protein and data from a heavy-atom derivative have been collected.

1. Introduction

The mariner element Mos1 (first isolated from *Drosophila mauritiana*) is a member of the mariner/Tc1 superfamily of transposable elements (Robertson, 1995). It encodes a 40.7 kDa transposase (containing 345 residues in the sequence shown in Fig. 1) that is the sole requirement for transposition of the element *in vitro*. This enzyme catalyses a set of DNA-cleavage and DNA-insertion reactions in a cut-and-paste mechanism. In the first step of this reaction, the transposase specifically recognizes and binds to the terminal inverted repeats of its own transposon *via* its N-terminal domain formed of the first 118 amino acids (Zhang *et al.*, 2001). The catalytic C-terminal domain mediates the subsequent excision of the transposon from the flanking DNA and its insertion into a target sequence. After disengagement of the protein from the DNA, the 5 bp single-strand gaps at each end are repaired. Mos1 transposase contains a variant, DD(34)D, of the catalytic DD(35)E motif (van Luenen *et al.*, 1994) that is required for divalent metal (Mg^{2+}) binding and is conserved among polynucleotide phosphotransferases. These include integrases, which are encoded by retroviruses and retrotransposons, and transposases of the bacterial elements Tn5 and Tn10 and other members of the mariner/Tc1 family.

The mechanism of transposition has been studied in detail for a number of prokaryotic elements, including Tn5 (Reznikoff, 2003) and Tn10 (Kennedy *et al.*, 1998). The crystal structure of Tn5 transposase in a synaptic complex provided the first structural insight into a transposase in its active DNA-bound state (Davies *et al.*, 2000). Other studies have revealed the structures of domains of bacterial transposases (Rice & Mizuuchi, 1995; Schumacher *et al.*, 1997) and related integrases (Chen *et al.*, 2000; Wang *et al.*, 2001). In contrast, less is known about the detailed mechanisms of eukaryotic transposition or the structures of the transposases involved. A recent biochemical analysis of the excision of DNA by Mos1 transposase (Dawson &

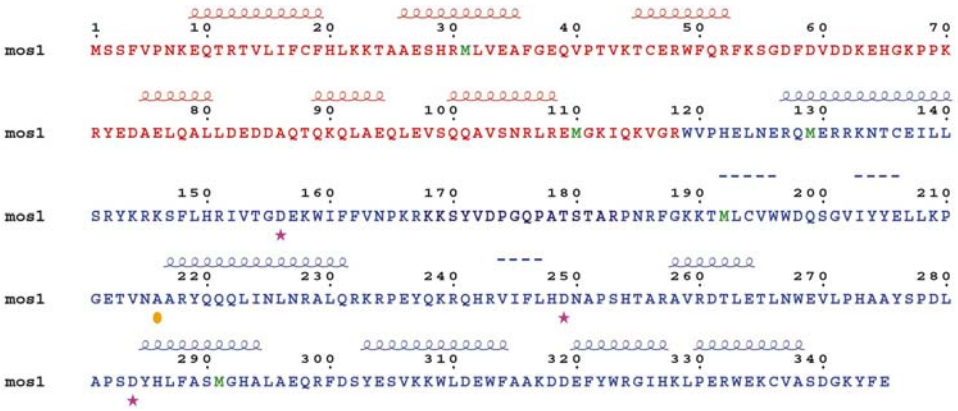

Figure 1
Sequence of Mos1 transposase. The N-terminal domain (residues 1–118) is coloured red and the C-terminal domain (residues 119–345) is coloured blue. An orange circle highlights the position of the single mutation T216A. Purple stars indicate the three active-site aspartic acid residues. Methionine residues are coloured green. The secondary structure predicted by the PredictProtein server is also shown. This diagram was created using the program *ESPript*.

Finnegan, 2003) highlighted differences in the mechanism of DNA cleavage by the enzyme compared with that by prokaryotic transposases. Moreover, similarities were found with the RAG1/2-mediated assembly of immunoglobulin and T-cell receptor genes in V(D)J recombination, consistent with the proposal that the modern immune system has evolved from a transposon more like Mos1 than bacterial transposons such as Tn5.

Structural analysis of Mos1 transposase has been precluded for many years owing to the insoluble nature of the protein when expressed in *Escherichia coli*. Here, we report the cloning, overexpression, purification and crystallization of a soluble single-point mutant of full-length Mos1 transposase.

2. Materials and methods

2.1. Protein cloning and overexpression

The single-point mutation (T216A) was isolated in a yeast two-hybrid screen designed to identify transposase derivatives that are able to interact more strongly with the wild-type protein. The sequence coding for the mutant protein was amplified by PCR and inserted into the *E. coli* expression vector pBCP378 (Velterop *et al.*, 1995) using *Nde*I and *Bam*HI restriction sites in the PCR primers. DNA sequencing confirmed the integrity of the cloned fragment. The recombinant plasmid was transformed into strain BL21(DE3) (Novagen) and plated out on LB agar containing carbenicillin (200 µg ml^{-1}). Single colonies were used to inoculate 50 ml of LB medium with carbenicillin. After overnight incubation at 303 K, these cultures were diluted into 500 ml LB medium (with 200 µg ml^{-1} carbenicillin). Expression of the transposase was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) when the cell culture grown at 310 K had attained late log phase (OD_{600} = 0.6). The mutant protein was shown by SDS–PAGE to be present in the soluble fraction of the total cell lysate. Expression conditions were optimized and the best level of expression was achieved using 1.0 m*M* IPTG for 4 h at 303 K.

For the production of selenomethionyl protein, the expression construct was transformed into the methionine-deficient *E. coli* strain B834(DE3). Single colonies grown on agar plates (with 200 µg ml^{-1} carbenicillin) were used to inoculate 50 ml overnight cultures of LB medium. These were pelleted and used to inoculate (to OD_{600} = 0.2) a minimal medium containing selenomethionine, along with the other 19 amino acids, and protein production was induced with IPTG as before. The subsequent extraction and purification of selenomethionyl protein proceeded as described below apart from the incorporation of 5 m*M* DTT in all buffers.

2.2. Protein purification and analysis

Cells collected from 1 l cell culture by centrifugation (2500*g* for 20 min at 277 K) were frozen overnight and then suspended in 20 ml extraction buffer (buffer *A*; 25 m*M* NaH_2PO_4, 500 m*M* NaCl, 5 m*M* $MgCl_2$, 1 m*M* DTT pH 7.5) containing Complete protease-inhibitor mix (Roche). Lysozyme (100 µl at 100 mg ml^{-1} in buffer *A*) was added and the mixture was incubated on ice for 15 min and then sonicated. DNaseI (200 U, Sigma) was added and the suspension was incubated on ice for a further 30 min. The cell lysate was centrifuged at 2500*g* for 30 min at 277 K and the precipitate was discarded. Initially, the crude extract was applied onto a Poros HS cation-exchange column on a BioCAD Sprint FPLC system (Applied Biosystems) and eluted with a salt gradient (0.2–1 *M* NaCl). Fractions eluting at approximately 0.6 *M* NaCl were shown by SDS–PAGE to contain Mos1 transposase. These fractions were pooled, concentrated and exchanged into buffer *B* (25 m*M* Tris, 0.2 *M* NaCl, 1 m*M* DTT pH 7.5) in a Vivaspin 20 ml centrifugal concentrator with a 10 kDa cutoff. Subsequently, the protein was passed through a Superdex 200 HR10/30 gel-filtration column on an AKTA FPLC system (Amersham Pharmacia Biotech) in buffer *B* at 0.2 ml min^{-1}. In this way Mos1 transposase was purified to 95% homogeneity, as confirmed by SDS–PAGE, with a typical yield of 3 mg per litre of LB medium.

The mutant protein was shown in transposition assays and gel-retardation experiments to have identical activity to wild-type protein (Dawson & Zhang, personal communication). The molecular weight of purified transposase was found to be 40 662 Da by MALDI–TOF mass spectrometry, compared with the predicted weight of 40 665 Da taking into account the loss of the first methionine residue. The molecular weight of the selenomethionine-substituted protein was 40 925 Da. The difference in weight of 259 Da indicated 100% incorporation of five Se atoms (47 Da difference per selenium) plus one Mg atom (24 Da).

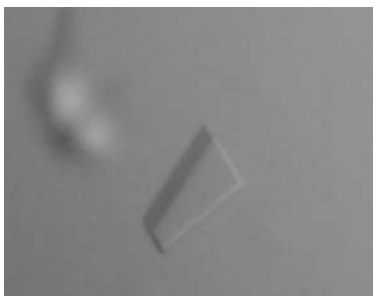

Figure 2
Crystals of Mos1 transposase.

2.3. Protein crystallization and preparation of heavy-atom derivatives

The hanging-drop vapour-diffusion method was used to grow crystals in 24-well Limbro plates. Plate-like crystals (shown in Fig. 2) with dimensions of up to 0.1 × 0.1 × 0.02 mm grew after three weeks at 290 K under the following conditions: the well solution consisted of 22–26%(*w*/*v*) PEG 4000, 0.1 *M* Tris buffer pH 7.5 and 5 m*M* $MgCl_2$ and the drop consisted of 2 µl protein solution at 13 mg ml^{-1} in buffer *B* plus 2 µl well solution. Selenomethionine-protein crystals were obtained by streak-seeding with a cat's whisker from a drop containing a crushed native crystal. A mercury derivative was obtained by soaking native crystals for 10 min in well solution containing 10 m*M* $KHgI_3$ followed by back-soaking into well solution. Prior to data collection, all crystals were briefly immersed in a cryoprotectant solution comprising 20%(*v*/*v*) glycerol in mother liquor and flash-cooled by plunging into liquid nitrogen. All diffraction data were collected at 100 K.

2.4. Biochemical analysis of the crystals

Native crystals were extracted from the hanging drops and washed in water before being dissolved in a 1:1 water:acetonitrile solution for N-terminal sequence analysis or 2%(*w*/*v*) SDS for analysis by PAGE and mass spectrometry. N-terminal sequence analysis of the native crystal (Procise sequencer, Applied Biosystems) predicted a sequence containing the residues (F/W)VPN. This could correspond to residues 4–7 with sequence FVPN and/or to residues 119–122 with sequence WVPH (see Fig. 1). MALDI–TOF mass spectrometry and SDS–PAGE (gel shown in Fig. 3) of the denatured and solubilized crystals indicated that the crystals contained or cleaved into four fragments: fragment *A* of 27.3 kDa, fragment *B* of 19.1 kDa, fragment *C* of 19.69 kDa and fragment *D* of ∼6 kDa. These fragments are consistent with the C-terminal domain (fragment *A*), residues Trp119–Glu345 (27.1 kDa), fragment *B* Ser2–Arg167 (19.72 kDa), fragment *C*

Table 1
Diffraction data statistics.

Values in parentheses are for the highest resolution shell.

Data set	Native	SeMet (inflection)	SeMet (peak)	SeMet (remote)	$KHgI_3$
Wavelength (Å)	1.10	0.98067	0.98017	1.000	1.005
Resolution (Å)	20–2.5 (2.57–2.5)	30–2.7 (2.83–2.7)	30–2.7 (2.83–2.7)	30–2.7 (2.83–2.7)	25–3.0 (3.09–3.0)
Unit-cell parameters (Å)					
$a = b$	44.5	44.6	44.6	44.6	43.9
c	205.6	205.5	205.5	205.5	204.2
Total observations	33917	43493	43267	42829	42991
Space group $P4_1/P4_3$					
Unique reflections	12439	10258	10237	10401	7409
R_{merge} (%)	5.1 (10.4)	8.1 (25.7)	8.8 (27.0)	13.7 (49.9)	7.1(10.2)
Completeness (%)	90.8 (95.8)	93.2 (98.3)	93.1 (97.7)	93.5 (98.0)	95.6 (88.3)
$I/\sigma(I)$	16.4 (6.4)	7.0 (2.9)	6.9 (2.8)	5.3 (1.5)	12.4 (10.4)
Space group $P4_12_12/P4_32_12$					
Unique reflections	7175	5984	5976	5966	4438
R_{merge} (%)	5.6 (10.9)	8.5 (27.3)	9.2 (28.6)	14.5 (53.1)	7.7 (11.2)
Completeness (%)	91.8 (97.7)	93.5 (98.3)	93.5 (97.7)	93.9 (98.1)	97.8 (95.3)
$I/\sigma(I)$	20.3 (8.3)	6.8 (2.7)	6.6 (2.6)	5.0 (1.4)	15.2 (13.1)

Pro184–Glu345 (19.273 kDa) and fragment *D* Trp119–Arg183 (7.87 kDa) or Trp119–Arg 167 (6.186 kDa). The *in vitro* cleavage of Mos1 adjacent to Arg119, Arg167 and Arg183 is consistent with the action of a protease and may reflect a role for protein degradation *in vivo*. For Tn5 it has been suggested that disengagement of the protein from DNA after transposition may be facilitated *in vivo* by proteolytic cleavage at Lys40 (Twining *et al.*, 2001).

2.5. X-ray data collection

Data were collected at the Daresbury SRS (station 14.2, various wavelengths) with an ADSC Q4 CCD detector and at the ESRF, Grenoble (station ID29, various wavelengths) with an ADSC Q210 CCD detector. In order to achieve the optimal alignment of the long *c* axis of the crystal with the spindle axis, crystals were either mounted in bent loops or positioned on a swing-arc goniometer head (Hampton Research). For MAD data, fluorescence spectra were measured to determine the absorption edge of selenium and data sets were measured at three wavelengths: the inflection point (0.98067 Å, λ_1) and the peak (0.98017 Å, λ_2) of the edge plus a low-energy remote wavelength (1.000 Å, λ_3). The mercury-derivative data were collected at a wavelength of 1.005 Å. All data were collected using a φ scan with a step size of 1.0° and indexing and scaling of data was performed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

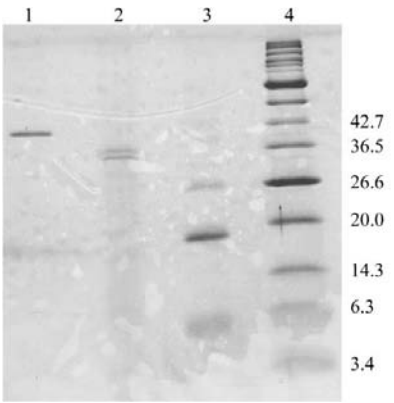

Figure 3
SDS–polyacrylamide gel (15%) of denatured and dissolved crystals of Mos1 transposase. Lane 1, purified Mos1 transposase; lane 2, crystallization drop after removal of crystals; lane 3, dissolved crystals; lane 4, molecular-weight markers (kDa).

3. Results and discussion

3.1. Unit-cell determination

The diffraction data (see Table 1) were consistent with space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 44.5$, $c = 205.6$ Å. In this space group, with one molecule of intact protein in the asymmetric unit (eight molecules per unit cell), the value of V_M is 1.3 Å^3 Da^{-1} (1.2% solvent), which is physically unreasonable. An alternative explanation is that the space group is $P4_1$ or $P4_3$ with merohedral twinning (with twinning operator $h, k, l = k, h, -l$), giving rise to the apparent higher symmetry of the diffraction data. If one molecule of intact protein is present in this lower symmetry space group the value of V_M is 2.5 Å^3 Da^{-1} (50.6% solvent). Other forms of crystal twinning with lower symmetry monoclinic or triclinic space groups, possibly containing truncated protein, cannot be completely discounted until the structure has been fully refined.

3.2. Twinning tests

Analysis of the native data set for partial twinning with *CNS* suggested a twinning fraction of 0.46 and the Yeates twinning server suggested a value of 0.47. The cumulative intensity statistics for the acentric reflections followed typical patterns for non-twinned diffraction data. However, it is recognized that pseudo-centring or strong anisotropic diffraction, such as could occur from a plate-like crystal or from a crystal with one unit-cell edge much longer than the others, can affect these statistics.

3.3. Preliminary phasing

Combining the selenomethionine MAD data with the mercury-derivative data, the positions of four out of five possible selenium positions and two mercury positions were found with the program *SOLVE* (Terwilliger & Berendzen, 1999), assuming the space group to be $P4_1$ or $P4_3$. Whilst both these space groups are consistent with the crystal data and the positions found, only $P4_1$ provided an electron-density map with interpretable features.

We thank the staff of the ESRF and Daresbury SRS, along with Dr Jose Martinez-Oyanedel, for help with data collection. Dr Andy Cronshaw carried out the mass spectrometry and N-terminal sequence analysis. JMR is a Caledonian Research Foundation Research Fellow and LZ was supported by a Darwin Trust Studentship.

References

Chen, J. C. H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D. & Stroud, R. M. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8233–8238.
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.
Kennedy, A. K., Guhathakurta, A., Kleckner, N. & Haniford, D. B. (1998). *Cell*, **95**, 125–134.
Luenen, H. G. A. M. van, Colloms, S. D. & Plasterk, H. A. (1994). *Cell*, **79**, 293–301.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Reznikoff, W. S. (2003). *Mol. Microbiol.* **47**, 1199–1206.
Rice, P. & Mizuuchi, K. (1995). *Cell*, **82**, 209–220.
Robertson, H. M. (1995). *J. Insect Physiol.* **41**, 99–105.
Schumacher, S., Clubb, R. T., Cai, M., Mizuuchi, K., Clore, G. M. & Gronenborn, A. M. (1997). *EMBO J.* **16**, 7532–7541.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–681.
Twining, S. S., Goryshin, I. Y., Bhasin, A. & Reznikoff, W. S. (2001). *J. Biol. Chem.* **276**, 23135–23143.
Velterop, J. S., Dijkhuisen, M. A., van't Hof, R. & Poastma, P. W. (1995). *Gene*, **153**, 63–65.
Wang, J., Ling, H., Yang, W. & Craigie, R. (2001). *EMBO J.* **20**, 7333–7343.
Zhang, L., Dawson, A. & Finnegan, D. J. (2001). *Nucleic Acids Res.* **29**, 3566–3575.