Ji Yong Kang, Hyung Ho Lee,
Do Jin Kim, Sang Hee Han,
Olesya Kim, Hyoun Sook Kim,
Sang Jae Lee and Se Won Suh*

Department of Chemistry, College of Natural
Sciences, Seoul National University,
Seoul 151-742, South Korea

Correspondence e-mail: sewonsuh@snu.ac.kr

Received 4 September 2006

Accepted 13 October 2006

Overexpression, crystallization and preliminary X-ray crystallographic analysis of a putative transposase from *Thermoplasma acidophilum* encoded by the *Ta0474* gene

IS200 transposases, originally identified in *Salmonella typhimurium* LT2, are present in many bacteria and archaea and are distinct from other groups of transposases. To facilitate further structural comparisons among IS200-like transposases, structural analysis has been initiated of a putative transposase from *Thermoplasma acidophilum* encoded by the *Ta0474* gene. Its 137-residue polypeptide shows high levels of sequence similarity to other members of the IS200 transposase family. The protein was overexpressed in intact form in *Escherichia coli* and crystallized at 297 K using a reservoir solution consisting of 100 mM Na HEPES pH 7.5 and 20% (v/v) ethanol. X-ray diffraction data were collected to 1.78 Å. The crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 65.00$, $b = 34.07$, $c = 121.58$ Å, $\alpha = 90$, $\beta = 100.20$, $\gamma = 90^\circ$. Four monomers, representing two copies of a dimeric molecule, are present in the asymmetric unit, giving a crystal volume per protein weight (V_M) of $2.02 \text{ Å}^3 \text{ Da}^{-1}$ and a solvent content of 39.2%.

1. Introduction

Transposons are mobile DNA sequences that can insert themselves into nonhomologous target sites in the genome (Beuzon *et al.*, 2004). Insertion sequences, the smallest transposable elements, are capable of independent transposition because they contain a transposase-encoding gene. At least six protein families have been described that mediate transposition: DDE transposases, Y2 (or rolling-circle) transposases, tyrosine transposases, serine transposases, a fifth family that encodes a combination of reverse transcriptase and endonuclease (RT/En) activities and IS200 (or Y1) transposases (Curcio & Derbyshire, 2003; Ronning *et al.*, 2005). A distinct fold and a unique active site have recently been revealed for IS200 transposases (Ronning *et al.*, 2005; Lee *et al.*, 2006).

The insertion sequences IS605, IS606 and IS608 from *Helicobacter pylori* carry two open reading frames: *tnpA* and *tnpB* (Ton-Hoang *et al.*, 2005). These *tnpA* open reading frames encode unique transposases belonging to the IS200 family. The protein sequences of IS200 transposase-family members do not bear the hallmarks of other well characterized transposases (Ronning *et al.*, 2005). It has been proposed that IS200 transposases be designated as 'Y1 transposases' because a single tyrosine residue is absolutely conserved among the family members (corresponding to Tyr127 in the *H. pylori* IS608 transposase; Ronning *et al.*, 2005). The crystal structure of the *H. pylori* IS608 transposase revealed that it is structurally related to rolling-circle replication proteins (Ronning *et al.*, 2005). Subsequently, the structure of an IS200 transposase from *Sulfolobus solfataricus* (SSO1474) was determined in both Mn^{2+} -bound and Mn^{2+} -free forms (Lee *et al.*, 2006).

To facilitate further structure comparisons between IS200-like transposases, we initiated the structure determination of a putative transposase from *Thermoplasma acidophilum* (Ta0474) that shares high levels of amino-acid sequence identity with the two structurally characterized transposases of the IS200 family. The identity is 50% with *H. pylori* IS608 transposase and 40% with IS200 transposase from *S. solfataricus* (SSO1474). As the first step towards structure determination of the putative transposase from *T. acidophilum*

(Ta0474), we have overexpressed it in *Escherichia coli* and crystallized it. The crystals are suitable for structure determination at high resolution. The crystallization conditions and preliminary X-ray crystallographic data are reported here.

2. Experimental

2.1. Protein expression and purification

The Ta0474 gene (residues 1–137) was amplified by the polymerase chain reaction using the genomic DNA of *T. acidophilum* strain DSM 1728 as template. The forward and reverse oligonucleotide primers designed using the published genome sequence (Ruepp *et al.*, 2000) are 5'-G GAA TTC **CAT ATG** GAA ATC ATT AAT GTA GAA AAG GAA T-3' and 5'-CCG CCG **CTC GAG** TTA TCA TTT TCC CTT CTG ATC CTC TA-3', respectively. The bases in bold represent the *Nde*I and *Xho*I restriction-enzyme cleavage sites. The amplified DNA was inserted into the *Nde*I/*Xho*I-digested expression vector pET-21a(+) (Novagen). This vector construction yields the recombinant enzyme with no fusion tag. The protein was overexpressed in *E. coli* Rosetta2(DE3)pLysS cells. The cells were grown at 310 K to an OD₆₀₀ of 0.5 in Terrific Broth medium containing 50 µg ml⁻¹ ampicillin and protein expression was induced by 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG). Incubation was continued at 303 K for 24 h after IPTG induction and the cells were harvested by centrifugation at 5600g for 10 min at 277 K.

The cell pellet was resuspended in ice-cold lysis buffer [20 mM Tris-HCl pH 7.9, 500 mM sodium chloride, 5 mM imidazole, 5% (v/v) glycerol and 1 mM phenylmethylsulfonylfluoride] and was lysed using an ultrasonic processor. Following heat treatment at 343 K for 10 min, the crude cell extract was centrifuged at 36 000g for 1 h at 277 K and the recombinant protein in the supernatant fraction was purified in three chromatographic steps. The first step utilized a heparin Sepharose XK-16 prep-grade column (12 ml; GE Healthcare) which had previously been equilibrated with 20 mM Tris-HCl pH 7.5 and 50 mM NaCl. The heparin column was used because Ta0474 is likely to interact with DNA. The supernatant fraction was diluted tenfold and applied onto the column. Upon eluting with a gradient of NaCl in the same buffer, Ta0474 transposase was eluted at ~1 M NaCl. After dialysis against 20 mM trisodium acetate pH 5.6 for 18 h, the transposase-containing fraction was applied onto a Source-S Sepharose prep-grade column (6 ml; GE Healthcare) which had previously been equilibrated with 20 mM trisodium acetate pH

5.6 and 50 mM NaCl. Upon eluting with a gradient of NaCl in the same buffer, the transposase was eluted at ~300 mM NaCl. Next, gel filtration was performed on a HiLoad XK-16 Superdex 200 prep-grade column (GE Healthcare) which was equilibrated and eluted with 20 mM Tris-HCl pH 7.5, 200 mM NaCl and 2 mM β-mercaptoethanol. The homogeneity of the purified protein was assessed by SDS-PAGE (Laemmli, 1970). The protein solution was concentrated using an YM10 ultrafiltration membrane (Millipore-Amicon). The protein concentration was estimated by measuring the absorbance at 280 nm.

2.2. Crystallization and dynamic light scattering

Crystallization was performed at 297 K by the hanging-drop vapour-diffusion method using 24-well VDX plates (Hampton Research). Initial crystallization conditions were established using screening kits from Hampton Research (Crystal Screens I and II, SaltRX, Index I and II, PEG/Ion and MembFac) and Emerald BioSystems (Wizard I and II). Each hanging drop was prepared on a siliconized cover slip by mixing 2 µl each of protein solution (4.4 mg ml⁻¹ in a buffer consisting of 20 mM Tris-HCl pH 7.5, 200 mM NaCl and 2 mM β-mercaptoethanol) and reservoir solution. The hanging drop was placed over 0.48 ml reservoir solution. Dynamic light-scattering experiments were performed on a model DynaPro-801 instrument from Wyatt (Santa Barbara, California, USA). The data were measured at 297 K with the protein at 1 mg ml⁻¹ concentration, dissolved in 20 mM Tris-HCl pH 7.5, 200 mM NaCl and 2 mM β-mercaptoethanol.

2.3. X-ray diffraction experiment

The crystals were flash-cooled using a cryoprotectant solution consisting of 100 mM Na HEPES pH 7.5, 20% (v/v) ethanol and 30% (v/v) glycerol. Crystals were soaked in 5 µl cryoprotectant solution for 10 s before being flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K on a Quantum 210 charge-coupled device area-detector system (Area Detector Systems Corporation, Poway, California, USA) at the BL-4A experimental station of the Pohang Light Source, Pohang, South Korea. The crystal was rotated through a total of 180° with a 1.0° oscillation range per frame. The raw data were processed and scaled using the HKL-2000 program package (Otwinowski & Minor, 1997).

For the molecular-replacement calculations, a dimer model of IS200 transposase from *S. solfataricus* (PDB code 2f4f; Lee *et al.*, 2006) was used as the search model. The program PHASER from the CCP4 program suite (Collaborative Computational Project, Number 4, 1994) was used to find one of the two dimers by a cross-rotational search followed by a translational search. The orientation and position of the first dimer was fixed and the second dimer was located using the MOLREP program from the CCP4 program suite. Manual model building was performed using the program COOT (Emsley & Cowtan, 2004). The model was refined with REFMAC from the CCP4 program package, including simulated annealing, positional refinement and individual B-factor refinement.

3. Results

The intact Ta0474 transposase from *T. acidophilum* was overexpressed in *E. coli* in soluble form with a yield of ~1.5 mg homogeneous protein per litre of culture. The molecular weight of the recombinant transposase was estimated to be ~39 kDa by dynamic light-scattering analysis, indicating that the enzyme exists as a homodimer in solution (calculated monomer weight 16 379 Da). This

Figure 1
Crystals of a putative transposase encoded by the Ta0474 gene of *T. acidophilum*.

Table 1

Data-collection statistics.

Values in parentheses are for the highest resolution shell.

X-ray source	Pohang Light Source beamline BL-4A
X-ray wavelength (Å)	0.97950
Temperature (K)	100
Space group	$P2_1$
Unit-cell parameters	
a (Å)	65.00
b (Å)	34.07
c (Å)	121.58
α (°)	90.00
β (°)	100.20
γ (°)	90.00
Resolution range (Å)	50–1.78 (1.84–1.78)
Total/unique reflections	588760/51361
$R_{\text{merge}}^{\dagger}$ (%)	7.8 (23.7)
Data completeness (%)	99.5 (96.9)
Average $I/\sigma(I)$	11.4 (6.8)
No. of dimers per asymmetric unit	2

$^{\dagger} R_{\text{merge}} = \sum_h \sum_i |I(h)_i - \langle I(h) \rangle| / \sum_h \sum_i I(h)_i$, where $I(h)_i$ is the intensity of the i th measurement of reflection h and $\langle I(h) \rangle$ is the mean value of $I(h)$ for all i measurements.

result is in agreement with the dimeric structures of IS200 transposase from *S. solfataricus* (Lee *et al.*, 2006) and IS608 transposase from *H. pylori* (Ronning *et al.*, 2005).

The best crystals were grown using reservoir solution consisting of 100 mM Na HEPES pH 7.5 and 20% (v/v) ethanol. Crystals grew to dimensions of $0.02 \times 0.04 \times 0.15$ mm within 2 d (Fig. 1). A set of X-ray diffraction data was collected to 1.78 Å resolution at 100 K. A total of 588 760 measured reflections were merged into 51 361 unique reflections, giving an R_{merge} of 7.8% and a completeness of 99.5%. The space group was determined to be $P2_1$ on the basis of systematic absences and symmetry of diffraction intensities. The unit-cell parameters are $a = 65.00$, $b = 34.07$, $c = 121.58$ Å, $\alpha = 90$, $\beta = 100.20$, $\gamma = 90^\circ$. Table 1 summarizes the statistics for data collection. If it is assumed that two copies of a dimeric molecule are present in the crystallographic asymmetric unit, the crystal volume per protein weight (V_M) is $2.02 \text{ Å}^3 \text{ Da}^{-1}$ and the solvent content is 39.2% (Matthews, 1968).

We have solved the structure by the molecular-replacement method using the dimer model of IS200 transposase from *S. solfataricus* (PDB code 2f4f; Lee *et al.*, 2006). The initial molecular-

replacement solution gave R_{work} and R_{free} values of 0.446 and 0.491, respectively, for 30.0–2.50 Å data. The model has been partially refined to crystallographic R_{work} and R_{free} values of 0.258 and 0.279, respectively, for 50.0–1.78 Å data and accounts for 544 residues (residues 1–136 in each chain) of the two copies of a dimeric molecule in the asymmetric unit. It reveals that Ta0474 transposase is structurally very similar to IS200 transposase from *S. solfataricus* (Lee *et al.*, 2006) and IS608 transposase from *H. pylori* (Ronning *et al.*, 2005). Ta0474 transposase (chain A) shows a root-mean-square deviation of 1.3 Å for 125 C α -atom pairs with IS200 transposase from *S. solfataricus* (PDB code 2f4f, chain A) and 1.2 Å for 128 C α -atom pairs with IS608 transposase from *H. pylori* (PDB code 2a6m, chain A). Further refinement is in progress and the structural details will be described in a separate publication.

We thank the staff at beamlines BL-4A and BL-6B of Pohang Light Source for assistance during X-ray experiments. This work was supported by a grant from the Korean Ministry of Science and Technology (NRL-2001). JYK, HHL, DJK, SHH, OK, HSK and SJL are recipients of BK21 fellowships from the Korean Ministry of Education and Human Resources Development.

References

- Beuzon, C. R., Chessa, D. & Casadesus, J. (2004). *Int. Microbiol.* **7**, 3–12.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Curcio, M. J. & Derbyshire, K. M. (2003). *Nature Rev. Mol. Cell Biol.* **4**, 865–877.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
- Lee, H. H., Yoon, J. Y., Kim, H. S., Kang, J. Y., Kim, K. H., Kim, D. J., Ha, J. Y., Mikami, B., Yoon, H. J. & Suh, S. W. (2006). *J. Biol. Chem.* **281**, 4261–4266.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Ronning, D. R., Guynet, C., Ton-Hoang, B., Perez, Z. N., Ghirlando, R., Chandler, M. & Dyda, F. (2005). *Mol. Cell*, **20**, 143–154.
- Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. & Baumeister, W. (2000). *Nature (London)*, **407**, 508–513.
- Ton-Hoang, B., Guynet, C., Ronning, D. R., Cointin-Marty, B., Dyda, F. & Chandler, M. (2005). *EMBO J.* **24**, 3325–3338.