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

Purification, crystallization and preliminary crystallographic analysis of the *Hermes* transposase

Zhanita N. Perez,[a] Primrose Musingarimi,[a]‡ Nancy L. Craig,[b] Fred Dyda[a] and Alison Burgess Hickman[a]*

[a]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, USA, and [b]Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, MD, USA

‡ Present address: London School of Economics and Political Science, Department of Social Policy, London, England.

Correspondence e-mail: ahickman@helix.nih.gov

Received 24 March 2005
Accepted 17 May 2005
Online 1 June 2005

DNA transposition is the movement of a defined segment of DNA from one location to another. Although the enzymes that catalyze transposition in bacterial systems have been well characterized, much less is known about the families of transposase enzymes that function in higher organisms. Active transposons have been identified in many insect species, providing tools for gene identification and offering the possibility of altering the genotypes of natural insect populations. One of these active transposons is *Hermes*, a 2749-base-pair element from *Musca domestica* that encodes its own transposase. An N-terminally deleted version of the *Hermes* transposase (residues 79–612) has been overexpressed and purified, and crystals that diffract to 2.1 Å resolution have been obtained at 277 K by the hanging-drop method.

1. Introduction

Hermes, a transposable element isolated from the genome of the housefly *Musca domestica* (Warren *et al.*, 1994), belongs to the *hAT* family (Kempken & Windhofer, 2001; Rubin *et al.*, 2001), one of ten superfamilies into which eukaryotic transposons can be classified (Kapitonov & Jurka, 2004). In general, *hAT* elements have short terminal inverted repeats (5–27 bp), generate 8 bp target-site duplications upon transposition and encode a single transposase protein that catalyzes the DNA breakage and rejoining reactions required for transposition (Kempken & Windhofer, 2001). All *hAT* transposases display significant amino-acid sequence similarity, with the highest primary structure conservation at their C-termini (Calvi *et al.*, 1991; Feldmar & Kunze, 1991).

Various functional and structural subdomains of the 612-residue *Hermes* transposase (Hermes; 70.1 kDa) have been determined or predicted (Fig. 1). For example, the N-terminus of Hermes contains residues important for nuclear localization (Michel & Atkinson, 2003) and has been proposed to contain a DNA-binding BED domain (residues 25–78; Aravind, 2000), while the C-terminus contains a sequence (residues 551–569) demonstrated to be important for multimerization (Michel *et al.*, 2003). It has been proposed that *hAT* transposases may carry a DSE catalytic triad (Bigot *et al.*, 1996), a variant of the DDE motif found in the active site of other transposases such as HIV-1 integrase and Mu transposase (Dyda *et al.*, 1994; Rice & Mizuuchi, 1995), and that two acidic residues in particular, Asp402 and Glu572, are necessary for transposition (Michel *et al.*, 2002). However, more recent biochemical characterization of Hermes *in vitro*, combined with secondary-structure

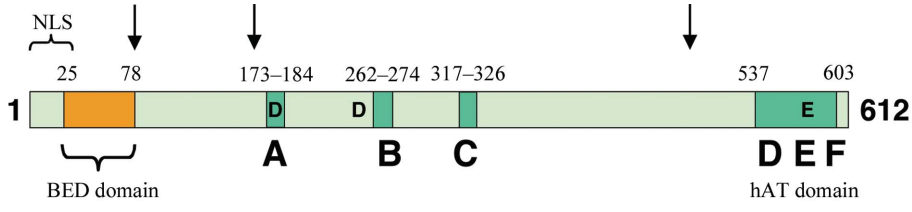

Figure 1
Schematic diagram of the *Hermes* transposase. Six conserved primary sequence blocks (designated A–F) are found in most *hAT* transposases (Rubin *et al.*, 2001). The three blocks D–F that cluster at the C-terminus are found within the *hAT* domain, which consists of residues 527–604. The three major proteolytic cleavage sites after residues 78, 162 and 482 are marked by arrows. Three important catalytic residues are Asp180, Asp248 and Glu572.

prediction, indicates that Hermes may be a member of the retroviral integrase superfamily and that residues Asp180, Asp248 and Glu572 are likely to form a metal-ion-binding site (Zhou *et al.*, 2004).

Our current understanding of the mechanisms of DNA transposition is derived from well studied prokaryotic and eukaryotic systems (Steiniger-White *et al.*, 2004; Chaconas & Harshey, 2002; Rio, 2002), although high-resolution structural data are currently limited to prokaryotic transposases. To date, only one full-length eukaryotic transposase, *Mos1* mariner from *Drosophila mauritiana*, has been crystallized (Richardson *et al.*, 2004). However, the structure of this mariner/Tc1 superfamily member has not yet been reported. Structural data are also available for the bipartite DNA-binding domain (residues 1–135) of the *Caenorhabditis elegans* Tc3 transposase in complex with transposon DNA (van Pouderoyen *et al.*, 1997; Watkins *et al.*, 2004). Although the structure provides insight into how the Tc3 transposase recognizes DNA, no structural information has been garnered about the remaining portion of this enzyme.

Given the limited structural information available on eukaryotic transposases, we have cloned, overexpressed and crystallized a portion of Hermes with the objective of obtaining a high-resolution crystal structure. We anticipate the structural analysis of this transposase will provide insight into the mechanism by which *hAT* transposases catalyze DNA transposition.

2. Materials and methods

2.1. Protein cloning and overexpression

The full-length *Hermes* transposase (Hermes; residues 1–612) was subcloned into pET-15b (Novagen) for expression in *Escherichia coli* as an N-terminal His-tag fusion protein. Although we were able to obtain copious amounts of soluble protein, when it was subjected to analytical gel-filtration chromatography on a calibrated Superose 6 column (Pharmacia) full-length Hermes migrated at a position consistent with a large aggregate (>700 kDa). Consequently, in efforts to identify a version that was more suitable for structural studies, Hermes was subjected to limited proteolysis by papain and trypsin. The digested protein was analyzed by SDS–PAGE and a prominent stable cleavage product common to both proteases (MW $\simeq$ 45 kDa) was characterized by N-terminal sequencing and MALDI–TOF mass spectrometry. The results were consistent with the removal of residues 1–78 from the N-terminus and ~130 residues from the C-terminus (Fig. 1). Given the known importance of the C-terminal region, we elected to investigate the properties of a Hermes construct missing only the N-terminal 78 residues. Thus, primers were designed to remove codons 1–78 from full-length Hermes and DNA encoding residues 79–612 in pET-15b was generated using a QuikChange kit (Stratagene). Soluble protein was obtained using the same expression procedure as for full-length Hermes.

2.2. Protein purification and analysis

Soluble Hermes 79–612 (both wild-type and a single point mutant, S163G; see below) was obtained by expression in *Escherichia coli* BL21(DE3) cells which were grown at 310 K until OD_{600} = 0.6. Cells were then rapidly cooled on ice to 292 K and protein expression was induced by addition of IPTG to a final concentration of 0.5 m*M*.

Cells collected from 8 l culture were harvested 16–20 h post-induction. The pellet was resuspended in 300 m*M* NaCl, 12 m*M* phosphate pH 7.4, flash-frozen in liquid nitrogen and then stored at 193 K. Unless noted otherwise, all purification steps were performed at 277 K. After thawing, cells were lysed by sonication in the presence of 500 m*M* NaCl, 5 m*M* imidazole (Im), 25 m*M* Tris pH 7.5 and 2 m*M* β-mercaptoethanol (BME). Following centrifugation of the cell lysate at 100000*g* for 45 min, the supernatant was loaded onto a Hi-Trap metal-chelation column (Amersham Biosciences) previously equilibrated with $NiSO_4$. The column was washed extensively with 20 m*M* Tris pH 7.5, 2 m*M* Im and 500 m*M* NaCl followed by the same buffer containing 22 m*M* Im. Hermes 79–612 was eluted from the column using a gradient of 22–400 m*M* Im. After visualization on an SDS–PAGE gel, fractions containing Hermes 79–612 were combined and dialyzed against 20 m*M* Tris pH 7.5, 1 m*M* EDTA, 500 m*M* NaCl, 4 m*M* BME and 10%(*w*/*v*) glycerol. This was followed by dialysis against a single change of the same buffer containing 5 m*M* DTT in place of BME (TSK buffer). To remove the polyhistidine tag, 10 units of thrombin (Sigma) were added per milligram of protein and incubated overnight. Thrombin was removed by passage over a 1 ml benzamidine Sepharose 4B (Pharmacia) column. The recovered protein was concentrated to at least 10 mg ml^{-1} prior to size-exclusion chromatography on a TSK-Gel G3000SW column equilibrated in TSK buffer.

The elution profile from the TSK column (Fig. 2*a*) revealed three forms of Hermes 79–612, which has a molecular weight of 61 425 Da. Subsequent analysis of the three species on a calibrated analytical size-exclusion Superdex 200 column demonstrated that peaks 1, 2 and

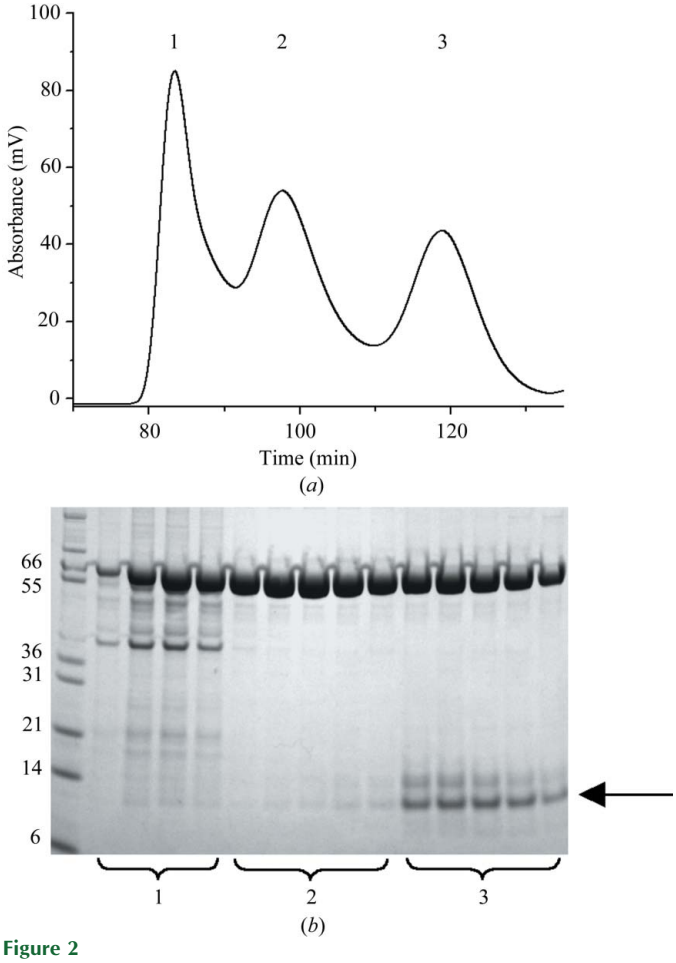

Figure 2
Elution profile of Hermes 79–612 S163G on a TSK-Gel G3000SW size-exclusion column. (*a*) Hermes 79–612 S163G elutes as three distinct species corresponding to a large-molecular-weight aggregate (peak 1), a hexamer (peak 2) and a smaller approximately dimeric species (peak 3). (*b*) SDS–PAGE analysis of the protein content of the three peaks. The arrow indicates the fragment in peak 3 subjected to MALDI–TOF (see text).

Table 1
Diffraction data statistics.

Values in parentheses are for the highest resolution shell (2.16–2.10 Å).

Wavelength (Å)	1.5418
Resolution (Å)	30–2.1
Total observations	155024
Unique reflections	41738
R_{merge} (%)	6.00 (27.1)
Completeness (%)	99.8 (99.9)
$I/\sigma(I)$	11.6 (4.3)

3 migrated with positions consistent with a high-molecular-weight aggregate, a species that was approximately hexameric and one that was approximately dimeric, respectively. The material in peak 2, but not that in peak 3, is active in *in vitro* assays of strand transfer and hairpin formation (Zhou & Craig, 2004).

When fractions corresponding to the three protein species were analyzed by SDS–PAGE (Fig. 2*b*), we noted that while the protein in peak 2 was quite homogenous, the dimeric protein consistently eluted with at least three other smaller proteins of ∼9–12 kDa. N-terminal sequencing indicated that the co-migrating contaminants all had the same N-terminal sequence, GSHMQSRE, corresponding to the four residues contributed by the remnants of the polyhistidine tag (GSHM) followed by the Hermes sequence beginning at residue 79. The most prominent contaminant, indicated by the arrow in Fig. 2(*b*), was analyzed by MALDI–TOF mass spectrometry using a PerSeptive Biosystems Voyager-DE instrument. The measured molecular weight was 9750 Da, consistent with a Hermes 79–162 fragment after removal of the polyhistidine tag. All attempts to remove the Hermes 79–162 fragment under conditions that maintained the native folded tertiary structure of the protein were unsuccessful. Efforts were subsequently made to eliminate the cleavage site by conservative mutation of Ile162 and Ser163. Although mutations failed to eliminate cleavage, a S163G mutant yielded considerably more soluble protein than the wild type and was therefore used for subsequent crystallization studies. This change in soluble protein recovery was the only reproducible difference between wild-type Hermes 79–612 and the S163G point mutant, which are otherwise indistinguishable in their purification properties, *in vitro* activities and crystallizability.

2.3. Protein crystallization

The proteins corresponding to material in peaks 2 and 3 were separately pooled, concentrated and used for crystallization trials. Although crystals were obtained of both species, only those corresponding to the approximately dimeric species were of diffraction quality. Initial crystals of the approximately dimeric species of Hermes 79–612 S163G were grown at 277 K by the hanging-drop method by mixing protein at 10 mg ml^{-1} in TSK buffer in a 1:1 ratio with 25%(*v*/*v*) polyethylene glycol 550 monomethyl ether (PEG 550 MME; Fluka), 0.1 *M* MES pH 6.5. Streak-seeding into a 1:1 mixture of protein and 23–25%(*v*/*v*) PEG 550 MME, 0.1 *M* MES pH 6.7 containing 0.25 *M* NaCl (total volume of 10 µl) yielded diffraction-quality crystals (Fig. 3*a*). Crystals were extracted from the hanging drop, washed and analyzed by SDS–PAGE (Fig. 3*b*); in addition to the expected Hermes 79–612 species, the persistent contaminants were also present. Crystals were cryoprotected by brief transfer into 10 m*M* Tris pH 7.5, 50 m*M* MES pH 6.5, 0.375 *M* NaCl, 2.5 m*M* DTT, 25%(*v*/*v*) PEG 550 MME, 10%(*w*/*v*) glycerol and frozen by rapid immersion in liquid propane.

Figure 3
(*a*) Crystals of Hermes 79–612 S163G. (*b*) SDS–PAGE analysis of dissolved crystals. Lane 1, molecular-weight markers (kDa); lane 2, dissolved crystals.

2.4. X-ray data collection

Native data were collected at 95 K on an R-AXIS IV image-plate detector mounted on a Rigaku RU-200 rotating-anode source operated at 50 kV/100 mA with multilayer focusing optics. Data were integrated and scaled internally using the *HKL* suite (Otwinowski & Minor, 1997).

3. Results and discussion

Crystals of the approximately dimeric form of Hermes 79–612 S163G were initially obtained under numerous conditions from 'The PEGS' screening suite from Nextal (Montreal, Canada) both at room temperature and 277 K. Optimization by the hanging-drop method at 277 K, combined with several rounds of streak-seeding, yielded crystals that nucleated overnight and continued to grow for up to a week to final maximum dimensions of 0.5 × 0.4 × 0.1 mm. The crystals belong to space group *C*2, with unit-cell parameters $a = 116.3$, $b = 84.9$, $c = 73.8$ Å, $\beta = 93.8^\circ$. Given the molecular weight of 79–612 S163G after thrombin cleavage, this corresponds to 49.2% solvent and a Matthews coefficient of 2.97 Å^3 Da^{-1}. Statistics for data collection are shown in Table 1.

Subsequent experiments have shown that the asymmetric unit contains one 79–612 monomer bound to a 79–162 fragment, consistent with the SDS–PAGE analysis of the dissolved crystals.

We thank C. Bradley for helpful comments on the manuscript.

References

Aravind, L. (2000). *Trends Biochem. Sci.* **25**, 421–423.
Bigot, Y., Augé-Gouillou, C. & Periquet, G. (1996). *Gene*, **174**, 265–271.
Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. (1991). *Cell*, **66**, 465–471.
Chaconas, G. & Harshey, R. M. (2002). *Mobile DNA II*, edited by N. L. Craig, R. Craigie, M. Gellert & A. M. Lambowitz, pp. 384–402. Washington DC: American Society of Microbiology.
Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). *Science*, **266**, 1981–1986.
Feldmar, S. & Kunze, R. (1991). *EMBO J.* **10**, 4003–4010.
Kapitonov, V. V. & Jurka, J. (2004). *DNA Cell Biol.* **23**, 311–324.
Kempken, F. & Windhofer, F. (2001). *Chromosoma*, **110**, 1–9.
Michel, K. & Atkinson, P. W. (2003). *J. Cell. Biochem.* **89**, 778–790.
Michel, K., O'Brochta, D. A. & Atkinson, P. W. (2002). *Gene*, **298**, 141–146.
Michel, K., O'Brochta, D. A. & Atkinson, P. W. (2003). *Insect Biochem. Mol. Biol.* **33**, 959–970.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pouderoyen, G. van, Ketting, R. F., Perrakis, A., Plasterk, R. H. & Sixma, T. K. (1997). *EMBO J.* **16**, 6044–6054.

Rice, P. & Mizuuchi, K. (1995). *Cell*, **82**, 209–220.
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.
Rio, D. C. (2002). *Mobile DNA II*, edited by N. L. Craig, R. Craigie, M. Gellert & A. M. Lambowitz, pp. 484–518. Washington DC: American Society of Microbiology.
Rubin, E., Lithwick, G. & Levy, A. A. (2001). *Genetics*, **158**, 949–957.
Steiniger-White, M., Rayment, I. & Reznikoff, W. S. (2004). *Curr. Opin. Struct. Biol.* **14**, 50–57.
Warren, W. D., Atkinson, P. W. & O'Brochta, D. A. (1994). *Genet. Res.* **64**, 87–97.
Watkins, S., van Pouderoyen, G. & Sixma, T. K. (2004). *Nucleic Acids Res.* **32**, 4306–4312.
Zhou, L. & Craig, N. L. (2004). Personal communication.
Zhou, L., Mitra, R., Atkinson, P. W., Hickman, A. B., Dyda, F. & Craig, N. L. (2004). *Nature (London)*, **432**, 995–1001.