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

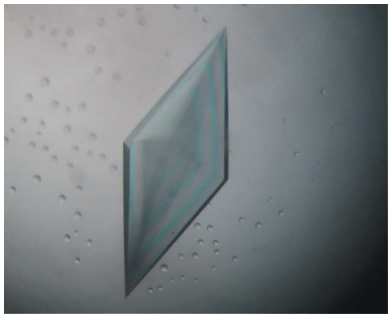
Cloning, purification, crystallization and preliminary X-ray crystallographic analysis of SET/TAF-Iβ ΔN from *Homo sapiens*

Zhen Xu,[a,b] Weili Yang,[a,b] Nuo Shi,[a,b] Yongxiang Gao,[a,b] Maikun Teng[a,b]* and Liwen Niu[a,b]*

[a]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230026, People's Republic of China, and [b]Key Laboratory of Structural Biology, Chinese Academy of Sciences, 96 Jinzhai Road, Hefei, Anhui 230026, People's Republic of China

Correspondence e-mail: mkteng@ustc.edu.cn, lwniu@ustc.edu.cn

Received 24 April 2010
Accepted 7 June 2010



The histone chaperone SET encoded by the SET gene, which is also known as template-activating factor Iβ (TAF-Iβ), is a multifunctional molecule that is involved in many biological phenomena such as histone binding, nucleosome assembly, chromatin remodelling, replication, transcription and apoptosis. A truncated SET/TAF-Iβ ΔN protein that lacked the first 22 residues of the N-terminus but contained the C-terminal acidic domain and an additional His_6 tag at the C-terminus was overexpressed in *Escherichia coli* and crystallized by the hanging-drop vapour-diffusion method using sodium acetate as precipitant at 283 K. The crystals diffracted to 2.7 Å resolution and belonged to space group $P4_32_12$.

1. Introduction

SET/TAF-Iβ belongs to the nucleosome-assembly protein (NAP) family, a group of histone chaperone-like proteins that play a variety of roles related to transcriptional control and DNA replication (Shikama *et al.*, 2000). The SET gene was initially isolated as a fusion gene with *can* in a case of acute undifferentiated leukaemia (AUL; Von Lindern *et al.*, 1992). SET/TAF-Iβ was subsequently characterized as a histone chaperone that has been suggested to be involved in alteration of chromatin structure (Kawase *et al.*, 1996). Further functional studies showed that SET/TAF-Iβ inhibits the acetylation of histone (Seo *et al.*, 2001) and DNA-binding transcription factor (Miyamoto *et al.*, 2003) as well as the methylation of DNA (Cervoni *et al.*, 2002). SET/TAF-Iβ functions as a regulator that is involved in replication (Matsumoto *et al.*, 1993), transcription (Matsumoto *et al.*, 1995), silencing (Cervoni *et al.*, 2002) and apoptosis (Fan *et al.*, 2003).

Full-length SET/TAF-Iβ (residues 1–277) consists of the NAP domain, which is conserved in NAP-family proteins, the N-terminus (residues 1–24) and a C-terminal acidic tail (residues 226–277). Previous research has indicated that one of the functional domains of SET/TAF-Iβ is its long acidic tail. The C-terminal acidic domain is required for DNA-replication (Nagata *et al.*, 1995) and transcription-stimulatory activity (Kawase *et al.*, 1996). Although it has been reported that the C-terminal acidic stretch of SET/TAF-Iβ is not necessary for its binding activity to histone H3 and H4, it is required for histone H2A–H2B binding activity (Muto *et al.*, 2007). Full-length SET/TAF-Iβ inhibits histone acetylation. In contrast, a SET/TAF-Iβ mutant lacking the entire C-terminal acidic tail was completely inactive (Seo *et al.*, 2001).

Three crystal structures of NAP-family proteins have been reported: NAP1 from *Saccharomyces cerevisiae* (PDB code 2ayu; Park & Luger, 2006), Vps75 from *S. cerevisiae* (PDB code 3dm7; Tang *et al.*, 2008) and NAP1 from *Plasmodium falciparum* (PDB code 3fs3; Gill *et al.*, 2009). All of those proteins contain the NAP domain. The crystal structure of human SET/TAF-Iβ ΔC (residues 1–225) has also been reported (PDB code 2e50; Muto *et al.*, 2007). Despite the fact that the C-terminal acidic domain may be disordered, it is presumed that this domain may influence the structure of SET/TAF-Iβ. In order to provide insight into the functional role of SET/TAF-Iβ containing the C-terminal acidic domain, we set out to determine the three-dimensional structure of SET/TAF-Iβ ΔN (residues 23–277) by X-ray crystallography. Here, we report the cloning, expression, purification,

crystallization and preliminary crystallographic analysis of SET/TAF-Iβ ΔN from *Homo sapiens*.

2. Materials and methods

2.1. Cloning

The SET gene (UniProtKB/Swiss-Prot accession No. Q01105-2) encodes the 32 kDa full-length SET/TAF-Iβ protein (residues 1–277). The gene fragment encoding residues 23–277 of SET/TAF-Iβ was amplified by polymerase chain reaction (PCR) from the human brain cDNA library using the oligonucleotide forward primer 5′-CAGTAGCA**CATATG**ACCTCAGAAAAAGAACAGCAAGAAG-3′ and the reverse primer 5′-GTC**CTCGAG**GTCATCTTCTCCTTCATCCTCCTC-3′ (Sangon). The PCR product was digested with *Nde*I and *Xho*I restriction endonucleases and cloned into pET-22b(+) vector (Novagen). The pET22b(+) vector adds an –LEHHHHHH tag at the C-terminus of the target protein for purification. Successful cloning was confirmed by DNA-sequencing analysis of the inserted ORF of SET/TAF-Iβ ΔN (Invitrogen Biotech).

2.2. Expression and purification

The recombinant plasmid was transformed into *Escherichia coli* BL21 (DE3) strain (Novagen) and the cells were grown in Luria–Bertani (LB) medium supplemented with ampicillin (50 µg ml^{-1}). The transformants were grown in LB medium at 310 K until they reached an optical density of 0.6 at 600 nm; expression was then induced with 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were grown at 310 K for 3 h after induction and harvested by centrifugation at 6000*g* and 277 K for 8 min. The bacterial pellets were resuspended in lysis buffer (500 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0 and 2% glycerol).

After the cells had been lysed by ultrasonication on ice, the lysate was clarified by centrifugation at 15 000*g* and 277 K for 30 min. The supernatant was loaded onto a nickel-chelating Sepharose Fast Flow column (GE Healthcare, USA) pre-equilibrated with loading buffer *A* (500 m*M* NaCl, 20 m*M* Tris–HCl, 10 m*M* imidazole pH 8.0). The column was washed with 20 column volumes of wash buffer (500 m*M* NaCl, 20 m*M* Tris–HCl, 50 m*M* imidazole pH 8.0) to remove contaminants. The target protein was eluted with elution buffer (500 m*M* NaCl, 20 m*M* Tris–HCl, 200 m*M* imidazole pH 8.0). The pooled fractions containing SET/TAF-Iβ ΔN protein were then exchanged into loading buffer *B* (1 *M* NaCl, 20 m*M* Tris–HCl pH 8.0) and loaded onto a HiPrep 16/10 Phenyl FF column (Amersham Biosciences, USA) pre-equilibrated with loading buffer *B*. The flowthrough fractions were applied onto a HiLoad 16/60 Superdex 200 prep-grade column (Amersham Biosciences, USA) pre-equilibrated with column buffer (200 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0). The apparent molecular weight (MW) of SET/TAF-Iβ ΔN was 143 kDa as estimated by gel filtration (data not shown). SET/TAF-Iβ ΔN was assumed to exist as a dimer of dimers. As the molecular shape of SET/TAF-Iβ may be not globular but rather extended, the apparent MW will be larger than the theoretical MW. The eluted fractions containing SET/TAF-Iβ ΔN were concentrated by centrifugal ultrafiltration (Millipore, 10 kDa cutoff). The final concentration of the target protein was 84 mg ml^{-1} as estimated using the BCA Protein Assay Kit (Pierce) with BSA as a standard. All purification steps were performed at 277 K and the result of each step was identified by SDS–PAGE analysis (Fig. 1). Finally, the purified protein was stored at 193 K for further use.

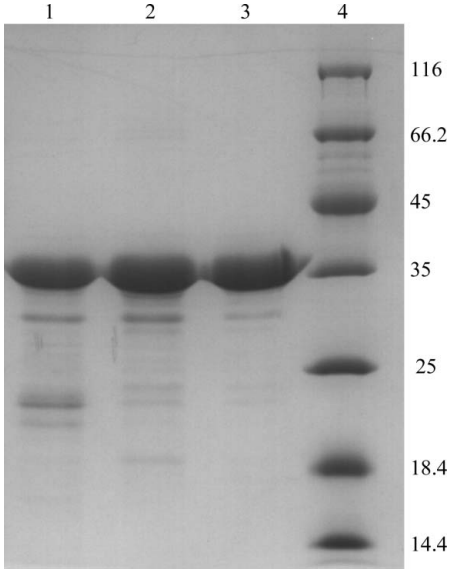


Figure 1
SDS–PAGE analysis of SET/TAF-Iβ ΔN at each stage of purification. About 10 µg protein was loaded per lane. Lane 1, eluant from nickel-chelating Sepharose Fast Flow column; lane 2, the flowthrough fraction after HiPrep 16/10 Phenyl FF column; lane 3, eluted fraction containing SET/TAF-Iβ ΔN from the HiLoad 16/60 Superdex 200 prep-grade column; lane 4, molecular-mass markers (kDa). The SDS–PAGE gel (12%) was stained with Coomassie Brilliant Blue R250.

2.3. Crystallization

An initial search for crystallization conditions was performed using the hanging-drop vapour-diffusion method at 283 K in 24-well plates. The PEG/Ion, Crystal Screen and Crystal Screen 2 reagent kits (Hampton Research) were used for initial crystallization-condition screening. SET/TAF-Iβ ΔN containing an uncleaved His_6 tag at the C-terminus was used for crystallization. Hanging drops, each consisting of 1 µl protein solution (stored in buffer containing 200 m*M* NaCl, 20 m*M* Tris–HCl pH 8.0) and 1 µl reservoir solution, were equilibrated against 100 µl reservoir solution. Microcrystals of SET/TAF-Iβ ΔN appeared in condition Nos. 11 and 34 of Hampton Research Crystal Screen 2 after 5 d. After several rounds of optimization of the protein concentration, the pH of the buffer, the precipitant concentration and the ionic strength, large diffraction-quality crystals appeared at 283 K with an optimized protein concentration of 80 mg ml^{-1} (stored in buffer containing 200 m*M* NaCl,

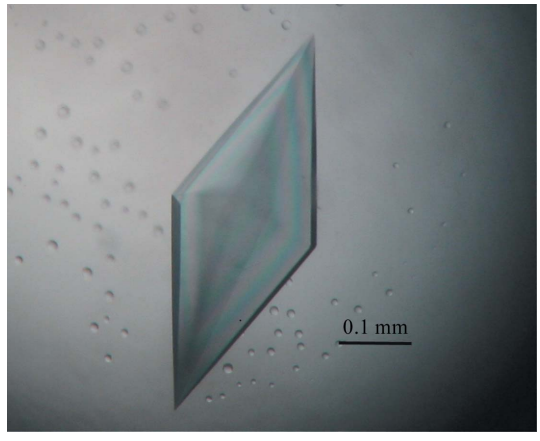


Figure 2
Photograph of an SET/TAF-Iβ ΔN crystal. The dimensions of this single crystal were about 0.6 × 0.2 × 0.2 mm.

Table 1
X-ray data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Source	SSRF beamline BL17U1
X-ray wavelength (Å)	0.97947
Space group	$P4_32_12$
Unit-cell parameters (Å)	$a = b = 114.75$, $c = 57.71$
Resolution range (Å)	50–2.70 (2.75–2.70)
Unique reflections	10722 (536)
Redundancy	7.3 (7.5)
Completeness (%)	96.5 (98.7)
Wilson B value (Å^2)	84.2
$R_{\rm merge}$† (%)	6.1 (37.7)
$R_{\rm meas}$ ‡ (%)	6.6 (40.5)
$\langle I/\sigma(I)\rangle$	33.6 (3.5)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm meas} = \sum_{hkl}[N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

20 m*M* Tris–HCl pH 8.0) using a reservoir solution consisting of 0.07 *M* cadmium sulfate hydrate, 0.1 *M* HEPES pH 7.8 and 1.0 *M* sodium acetate trihydrate. After several days, the well diffracting crystal had grown to typical dimensions of 0.6 × 0.2 × 0.2 mm (Fig. 2).

2.4. Diffraction data collection and processing

For X-ray data collection, a SET/TAF-Iβ ΔN crystal was harvested and soaked in a cryoprotectant solution consisting of reservoir solution with an additional 20%(*v*/*v*) glycerol for several seconds. The crystal was flash-cooled in liquid nitrogen and subjected to X-ray diffraction data collection at 100 K on synchrotron-radiation beamline BL17U1 using an MX225 CCD detector (MAR Research, Germany) at Shanghai Synchrotron Radiation Facility (SSRF). The crystal-to-detector distance was 200 mm. A complete diffraction data set containing 100 diffraction images was collected using a single SET/TAF-Iβ ΔN crystal with an oscillation angle of 1° per image. The wavelength was 0.97947 Å and the exposure time was 1 s per image. The diffraction data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Data-collection and processing statistics are listed in Table 1.

3. Results and discussion

As all crystallization trials using full-length SET/TAF-Iβ failed, the protein sequence of SET/TAF-Iβ was modified in order to improve its crystallizability. The N-terminal 24 amino acids of human SET/TAF-Iβ are not required for many activities such as DNA replication and transcription stimulation (Miyaji-Yamaguchi *et al.*, 1999), but the residues Thr23 and Ser24 are involved in a long α-helix (α2) which has a stable structure in the crystal structure of SET/TAF-Iβ ΔC (PDB code 2e50; Muto *et al.*, 2007). Therefore, a truncated form of SET/TAF-Iβ was designed for crystallization experiments that lacked 22 residues from the N-terminus.

The form which lacked 22 N-terminal residues was successfully cloned in *E. coli* with a fused hexahistidine tag at the C-terminus. The yield of pure protein is approximately 12 mg per litre of bacterial culture. The purity of SET/TAF-Iβ ΔN was more than 95% as estimated by SDS–PAGE analysis (Fig. 1). We obtained a crystal of SET/TAF-Iβ ΔN that diffracted to 2.70 Å resolution. The data-collection and processing statistics are summarized in Table 1. Preliminary crystallographic analysis gave a calculated Matthews coefficient ($V_{\rm M}$) of 3.18 Å^3 Da^{-1}, with a solvent content of 61.5%, assuming the presence of one molecule in the asymmetric unit (Matthews, 1968).

Molecular replacement was carried out with *MOLREP* (Vagin & Teplyakov, 1997) in the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) using the structure of residues 23–225 of *H. sapiens* SET/TAF-Iβ ΔC (PDB code 2e50; Muto *et al.*, 2007) as a search model. Furthermore, systematic absences indicated that the SET/TAF-IβΔN crystal belonged to space group $P4_32_12$ or its enantiomer $P4_12_12$. However, the *R* factor for space group $P4_12_12$ was always very high. For example, the initial *R* factors from molecular replacement were 0.488 for $P4_32_12$ and 0.562 for $P4_12_12$. Ultimately, the process of refinement confirmed space group $P4_32_12$ in this case. Further model building and refinement is in progress.

Financial support for this project was provided by research grants from the Chinese Ministry of Science and Technology (grant Nos. 2006CB806500, 2006CB910200 and 2006AA02A318) and the Chinese Academy of Sciences (grant No. KSCX2-YW-R-60).

References

Cervoni, N., Detich, N., Seo, S., Chakravarti, D. & Szyf, M. (2002). *J. Biol. Chem.* **277**, 25026–25031.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Fan, Z., Beresford, P. J., Oh, D. Y., Zhang, D. & Lieberman, J. (2003). *Cell*, **112**, 659–672.
Gill, J., Yogavel, M., Kumar, A., Belrhali, H., Jain, S. K., Rug, M., Brown, M., Maier, A. G. & Sharma, A. (2009). *J. Biol. Chem.* **284**, 10076–10087.
Kawase, H., Okuwaki, M., Miyaji, M., Ohba, R., Handa, H., Ishimi, Y., Fujii-Nakata, T., Kikuchi, A. & Nagata, K. (1996). *Genes Cells*, **1**, 1045–1056.
Matsumoto, K., Nagata, K., Ui, M. & Hanaoka, F. (1993). *J. Biol. Chem.* **268**, 10582–10587.
Matsumoto, K., Okuwaki, M., Kawase, H., Handa, H., Hanaoka, F. & Nagata, K. (1995). *J. Biol. Chem.* **270**, 9645–9650.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miyaji-Yamaguchi, M., Okuwaki, M. & Nagata, K. (1999). *J. Mol. Biol.* **290**, 547–557.
Miyamoto, S., Suzuki, T., Muto, S., Aizawa, K., Kimura, A., Mizuno, Y., Nagino, T., Imai, Y., Adachi, N., Horikoshi, M. & Nagai, R. (2003). *Mol. Cell. Biol.* **23**, 8528–8541.
Muto, S., Senda, M., Akai, Y., Sato, L., Suzuki, T., Nagai, R., Senda, T. & Horikoshi, M. (2007). *Proc. Natl Acad. Sci. USA*, **104,** 4285–4290.
Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A. & Matsumoto, K. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 4279–4283.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, Y.-J. & Luger, K. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 1248–1253.
Seo, S., McNamara, P., Heo, S., Turner, A., Lane, W. S. & Chakravarti, D. (2001). *Cell*, **104**, 119–130.
Shikama, N., Chan, H. M., Krstic Demonacos, M., Smith, L., Lee, C.-W., Cairns, W. & La Thangue, N. B. (2000). *Mol. Cell. Biol.* **20**, 8933–8943.
Tang, Y., Meeth, K., Jiang, E., Luo, C. & Marmorstein, R. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 12206–12211.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Von Lindern, M., Van Baal, S., Wiegant, J., Raap, A., Hagemeijer, A. & Grosveld, G. (1992). *Mol. Cell. Biol.* **12**, 3346–3355.