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

Yuan Lin,[a] Shilin Chen,[b] Shuyi Si[a]* and Yong Xie[b]*

[a]Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Tian Tan Xi Li, Beijing 100050, People's Republic of China, and [b]Institute of Medicinal Plant, Chinese Academy of Medical Sciences and Peking Union Medical College, 151 Malianwa North Road, Haidian District, Beijing 100193, People's Republic of China

Correspondence e-mail: sisyimb@hotmail.com, yxie@implad.ac.cn

Received 7 April 2011
Accepted 9 May 2011

Expression, purification, crystallization and preliminary X-ray crystallographic analysis of the hyperthermophilic nucleotidyltransferase TTHA1015 from *Thermus thermophilus* HB8

The *TTHA1015* gene from *Thermus thermophilus* HB8 encodes a hyperthermophilic nucleotidyltransferase. TTHA1015 has high homology to proteins belonging to two related families: the nucleotidyltransferase-domain superfamily and the DNA polymerase β-like family. However, no crystal structures of these proteins have been reported. Determination of the crystal structure of TTHA1015 will help in elucidation of its function and will be useful for understanding the relationship between the structure and the function of these homologous proteins. In this study, TTHA1015 was expressed, purified and crystallized. X-ray diffraction data were collected to 1.70 Å resolution. The crystal belonged to the monoclinic space group *C*2, with unit-cell parameters $a = 65.5$, $b = 34.7$, $c = 42.4$ Å, $\beta = 119.1^\circ$. There was one molecule per asymmetric unit, giving a Matthews coefficient of 1.86 Å^3 Da^{-1} and an approximate solvent content of 34%.

1. Introduction

The goal of genomic projects is to determine the complete genome of an organism and to provide information about the number of genes and the sequences of coding frames. *Thermus thermophilus* is a species of bacterium with important value in genetic engineering. Genomic projects on *T. thermophilus* are being performed in Japan using strain HB8 (Project ID 13202; Yokoyama *et al.*, 2000). This study revealed that the *TTHA1015* gene from *T. thermophilus* HB8 encodes a hyperthermophilic nucleotidyltransferase which contains one polypeptide chain of 98 residues (NCBI reference sequence YP_144281.1). Nucleotidyltransferases (NTs; EC 2.7.7) are transferase enzymes with phosphorus-containing groups that catalyze the transfer of a nucleotide to an acceptor hydroxyl group in a wide range of substrates (Betat *et al.*, 2010). More importantly, nucleotidyltransferases are utilized by all living systems and play roles in diverse biological processes such as polynucleotide synthesis and modification, small-molecule biosynthesis and metabolism, and antibiotic resistance (Bralley *et al.*, 2005; Neuenfeldt *et al.*, 2008).

The structural and functional details of TTHA1015 are still unknown. A protein sequence-homology search using the *Basic Local Alignment Search Tool* (*BLAST*) revealed that TTHA1015 has high homology to several proteins, in particular up to 71% identity to a DNA polymerase β-domain protein region (ZP_03496662.1) from *T. aquaticus* (Fig. 1). The *BLAST* search also revealed that all of the homologous proteins belonged to two related families: the nucleotidyltransferase-domain superfamily and the DNA polymerase β-like family (Aravind & Koonin, 1999; Holm & Sander, 1995). However, no crystal structures of these highly homologous proteins have been reported. To date, research on the three-dimensional structure of nucleotidyltransferases in *T. thermophilus* has been very limited. The function of TTHA1015 as a nucleotidyltransferase is still speculative and remains to be confirmed. Because protein structure is closely linked to protein function, structural genomics projects, which aim to determine the three-dimensional structure of proteins, may provide information on protein function (Bork & Koonin, 1998; Chothia & Lesk, 1986; Hegyi & Gerstein, 1999). Thus, research on the crystal structure of TTHA1015 will help to define its function and will provide other useful information for understanding the structure and the function of the homologous proteins.

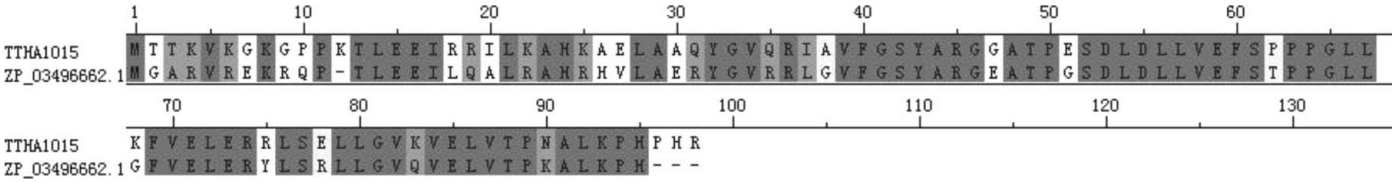

Figure 1
Amino-acid sequence alignment of TTHA1015 from *T. thermophilus* and a DNA polymerase β-domain protein region (ZP_03496662.1) from *T. aquaticus*. Conserved residues are shaded in dark grey and similar residues are shaded in light grey. Dashes indicate spacing in the amino-acid sequences for proper alignment. Numbers correspond to amino-acid positions.

2. Materials and methods

2.1. Protein expression and purification

The gene encoding TTHA1015 was amplified by PCR from *T. thermophilus* HB8 genomic DNA and subcloned into the pET11a vector (Novagen). Two primers were constructed and used to amplify the full-length *TTHA1015* gene, using 5′-GGAATTCCATATGACAACGAAGGTCAAGGGGAAAGG-3′ as the sense primer, which introduces an *Nde*I restriction site (CATATG), and 5′-CGGATCCCTACCGATGTGGGTGGGG-3′ as the antisense primer, which introduces a *Bam*HI restriction site (GGATCC). TTHA1015 protein was expressed in *Escherichia coli* BL21 cells cultivated in LeMaster medium containing SeMet at 310 K. After disruption of the cells by sonication, the lysate was subjected to heat treatment at 343 K for 30 min and centrifuged to remove the denatured proteins. Purification of TTHA1015 protein from the supernatant was performed at 277 K using an ÄKTAexplorer system (GE Healthcare). The protein was loaded onto a HiTrap Phenyl column (5 ml) followed by a HiTrap SP column (5 ml) and a Mono S column (1 ml). Further purification was performed with a HiLoad 16/60 Superdex 75 column equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 150 m*M* NaCl and 2 m*M* DTT.

2.2. Crystallization

The purified TTHA1015 protein was concentrated to approximately 4.3 mg ml^{-1} using an Amicon Ultra-15 (MWCO 3000) filter unit (Millipore). Crystallization was carried out using the sitting-drop vapour-diffusion method at 293 K. Each sitting drop consisted of 1 µl protein solution and 1 µl reservoir solution. Crystal Screen, PEG/Ion and Index kits (Hampton Research) were used to establish initial crystallization conditions. Crystallites were obtained with condition No. 5 [200 m*M* $MgCl_2$, 20%(*v*/*v*) PEG 3350] of the PEG/Ion kit within 3 d at 293 K. The pH and the concentrations of PEG 3350 and $MgCl_2$ were then optimized. A single TTHA1015 crystal with approximate dimensions of 0.2 × 0.1 × 0.1 mm was obtained overnight from a sitting-drop mixture of protein solution and reservoir solution, which consisted of 28%(*v*/*v*) PEG 3350, 100 m*M* glycine pH 9.0 and 100 m*M* $MgCl_2$ (Fig. 2).

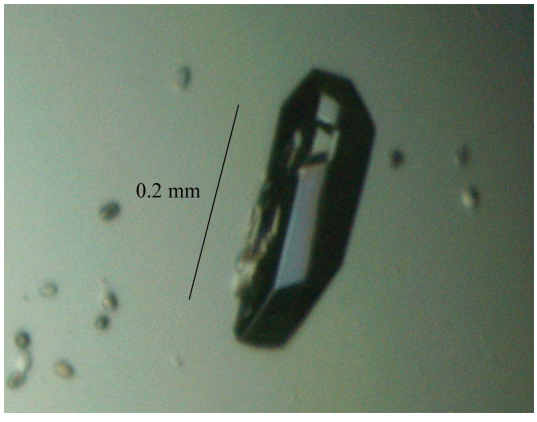

Figure 2
A typical crystal of TTHA1015, with approximate dimensions of 0.2 × 0.1 × 0.1 mm.

Table 1
Crystal parameters and X-ray diffraction data-collection statistics.

Values in parentheses are for the highest resolution shell.

Space group	*C*2
Unit-cell parameters (Å, °)	a = 62.5, b = 34.7, c = 42.4, β = 119.1
No. of molecules in asymmetric unit	1
V_M (Å^3 Da^{-1})	1.86
Solvent content (%)	34
Resolution range (Å)	50.00–1.70 (1.76–1.70)
No. of measured reflections	27176
No. of unique reflections	15711 (1262)
Completeness (%)	91.8 (74.2)
Average $I/\sigma(I)$	32.3 (4.4)
R_{merge}† (%)	2.6 (16.0)
Multiplicity	1.9 (1.7)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement of the intensity of reflection hkl, $\langle I(hkl)\rangle$ is the mean intensity of this reflection and the summation is the over all observed reflections.

2.3. Diffraction data collection

X-ray diffraction data from a single TTHA1015 crystal were collected on the BL5A structural biology beamline at the Photon Factory, High Energy Accelerator Research Organization (KEK), Japan. The TTHA1015 crystal was mounted on a rayon loop in a stream of liquid nitrogen at 100 K. Prior to data collection, the crystal was soaked in a cryoprotectant solution consisting of 50%(*v*/*v*) Paratone-N and 50%(*v*/*v*) paraffin to remove reservoir solution. The diffraction patterns were recorded on an ADSC Quantum 315 CCD detector. The wavelength, camera distance, oscillation range and exposure time were 1.000 Å, 180 mm, 1.0° and 8 s, respectively. A complete data set was collected as 180 images covering 180° in total. Diffraction intensity data were processed and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

3. Results and discussion

Recombinant TTHA1015 was expressed in *E. coli*. The molecular mass of the protein was approximately 10.8 kDa. X-ray diffraction data were collected at 1.70 Å resolution. The TTHA1015 crystal belonged to the monoclinic space group *C*2, with unit-cell parameters a = 65.5, b = 34.7, c = 42.4 Å, β = 119.1°. There was one molecule in the asymmetric unit, giving a Matthews coefficient of 1.86 Å^3 Da^{-1} (Matthews, 1968) and an approximate solvent content of 34%. The statistics of the diffraction data are shown in Table 1.

Molecular-replacement (MR) searches were performed using the *MrBUMP* program (Keegan & Winn, 2007) from the *CCP*4 program package (Winn *et al.*, 2011). The programs *MOLREP* (Vagin & Teplyakov, 2010) and *Phaser* (McCoy *et al.*, 2007) and the NMR solution structure of a putative minimal nucleotidyltransferase (chain *A*, PDB entry 1wot; S. Suzuki, H. Hatanaka, T. Hondoh, A. Okumura, Y. Kuroda, S. Kuramitsu, T. Shibata, Y. Inoue & S. Yokoyama, unpublished work) were used to search for a solution; however, a correct solution was not found. In addition, the following proteins (or) domains with more than 30% sequence identity to TTHA1015 were used as MR search models: a putative nucleotidyltransferase (NP_343093.1) from *Sulfolobus solfataricus* (PDB entry 2rff; 33.3% sequence identity; Joint Center for Structural Genomics, unpublished work), the C-terminal domain of MBF1 from *Trichoderma reesei* (PDB entry 2jvl; 34.0% sequence identity; Salinas *et al.*, 2009), the nucleotide-binding domain of the HI0073/HI0074 two-protein nucleotidyl transferase (PDB entry 1no5; 33.3% sequence identity; Lehmann *et al.*, 2005), kanamycin nucleotidyltransferase (PDB entry 1kan; 33.9% sequence identity; Pedersen *et al.*, 1995) and human phosphopantothenoylcysteine synthetase (PDB entry 1p9o; 33.3% sequence identity; Manoj *et al.*, 2003). As no correct MR solutions were found with these models, isomorphous replacement and multi-anomalous dispersion (MAD) methods are presently being utilized to solve the crystal structure of TTHA1015.

The authors are grateful to Drs Yamada and Matsugaki for the use of BL5A at KEK. This study was supported by the Research Foundation of IMPLAD, the Chinese Academy of Medical Science and Peking Union Medical College (to YX).

References

Aravind, L. & Koonin, E. V. (1999). *Nucleic Acids Res.* **27**, 1609–1618.
Betat, H., Rammelt, C. & Mörl, M. (2010). *Cell. Mol. Life Sci.* **67**, 1447–1463.
Bork, P. & Koonin, E. V. (1998). *Nature Genet.* **18**, 313–318.
Bralley, P., Chang, S. A. & Jones, G. H. (2005). *J. Bacteriol.* **187**, 5927–5936.
Chothia, C. & Lesk, A. M. (1986). *EMBO J.* **5**, 823–826.
Hegyi, H. & Gerstein, M. (1999). *J. Mol. Biol.* **288**, 147–164.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 345–347.
Keegan, R. M. & Winn, M. D. (2007). *Acta Cryst.* D**63**, 447–457.
Lehmann, C., Pullalarevu, S., Krajewski, W., Willis, M. A., Galkin, A., Howard, A. & Herzberg, O. (2005). *Proteins*, **60**, 807–811.
Manoj, N., Strauss, E., Begley, T. P. & Ealick, S. E. (2003). *Structure*, **11**, 927–936.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Neuenfeldt, A., Just, A., Betat, H. & Mörl, M. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 7953–7958.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pedersen, L. C., Benning, M. M. & Holden, H. M. (1995). *Biochemistry*, **34**, 13305–13311.
Salinas, R. K., Camilo, C. M., Tomaselli, S., Valencia, E. Y., Farah, C. S., El-Dorry, H. & Chambergo, F. S. (2009). *Proteins*, **75**, 518–523.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, 943–945.