

Ana P. G. Silva,^a Robert T. Byrne,^a Maria Chechik,^a Callum Smits,^a David G. Waterman^b and Alfred A. Antson^{a*}

^aYork Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, England, and ^bDiamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, England

Correspondence e-mail: fred@ysbl.york.ac.uk

Received 18 September 2008

Accepted 17 October 2008

Expression, purification, crystallization and preliminary X-ray studies of the TAN1 orthologue from *Methanothermobacter thermautotrophicus*

MTH909 is the *Methanothermobacter thermautotrophicus* orthologue of *Saccharomyces cerevisiae* TAN1, which is required for *N*⁴-acetylcytidine formation in tRNA. The protein consists of an N-terminal near-ferredoxin-like domain and a C-terminal THUMP domain. Unlike most other proteins containing the THUMP domain, TAN1 lacks any catalytic domains and has been proposed to form a complex with a catalytic protein that is capable of making base modifications. MTH909 has been cloned, overexpressed and purified. The molecule exists as a monomer in solution. X-ray data were collected to 2.85 Å resolution from a native crystal belonging to space group *P*6₁22 (or *P*6₅22), with unit-cell parameters *a* = 69.9, *c* = 408.5 Å.

1. Introduction

Following transcription by RNA polymerases, the bases of RNA may be subjected to post-transcriptional modification, which results in the presence of a diverse range of noncanonical bases in mature RNAs (Nakanishi & Nureki, 2005). Of the 107 currently known modifications, 91 have been shown to exist in transfer RNA, making it by far the most diversely modified RNA (Limbach *et al.*, 1994). The roles of these base modifications are varied, but they commonly improve the fidelity of translation (Agris, 2008) and the structural stability of tRNA (Helm, 2006).

Numerous sequence analyses and experimental studies have made it possible to identify different classes of RNA-modifying enzymes (Gustafsson *et al.*, 1996; Mueller & Palenchar, 1999). A number of these enzymes display a degree of modularity such that a catalytic core, which catalyses the modification, is fused to an RNA-binding domain, which binds the substrate. One such domain found in all three domains of life is the THUMP domain, which is named after the enzymes in which it has been identified: 4-thiouridine synthetases, methylases and pseudouridine synthases (Aravind & Koonin, 2001). Recently, the structures of three proteins containing the THUMP domain have been reported: the *Pyrococcus horikoshii* orthologue of 4-thiouridine synthetase (ThiI; Sugahara *et al.*, 2007), *Bacillus anthracis* ThiI (Waterman *et al.*, 2006) and human pseudouridine synthase Pus10 (McCleverty *et al.*, 2007). In each case the THUMP domain is positioned at the N-terminus of the protein. In ThiI, however, there is an additional N-terminal domain containing 60 residues, denominated the NFLD (near-ferredoxin-like domain), which is suggested to have a role in RNA binding (Waterman *et al.*, 2006). At present, no accurate structural information is available for the complex formed between the THUMP domain and RNA. Biochemical studies on the *P. abyssi* methyltransferase showed that the standalone THUMP domain has a far weaker affinity for tRNA than the full-length enzyme (Gabant *et al.*, 2006). Nevertheless, interactions between the THUMP domain and tRNA may contribute to targeting the tRNA substrate towards the catalytic domain as in other RNA-modifying enzymes (Ishitani *et al.*, 2003; Gabant *et al.*, 2006).



© 2008 International Union of Crystallography
All rights reserved

TAN1 (tRNA acetylation) from *Saccharomyces cerevisiae* has been shown *in vivo* to play a role in the formation of the modified base N^4 -acetylcytidine (ac^4C) at position 12 of tRNA^{Leu} and tRNA^{Ser}. Interestingly, *in vitro* studies showed that TAN1 alone was unable to catalyse the formation of the modification on its own yet was able to bind tRNA, an observation that is consistent with its lack of similarity to other acetyltransferases and the presence of the NFLD and THUMP domain (Johansson & Bystrom, 2004). Combined, these facts suggest that TAN1 is part of a larger RNA-modifying complex containing a separate catalytic module. Like *S. cerevisiae* TAN1, the *Methanothermobacter thermautotrophicus* orthologue MTH909 contains both an NFLD and a THUMP domain (with 13% sequence identity between the aligned regions), although the length of the polypeptide chain of the archaeal protein is somewhat shorter. MTH909 has a sequence identity of 14% or less with the closest homologues for which three-dimensional structures are available.

Here, we present the results of a preliminary structural investigation of MTH909, which resulted in the collection of an essentially complete X-ray data set to 2.85 Å resolution from crystals of the native protein.

2. Materials and methods

2.1. Cloning and expression

The full-length *mth909* gene was amplified by PCR from the genomic DNA of *M. thermautotrophicus* (strain Delta H) using primers containing *Nde*I and *Hind*III sites. The insert was then cloned into pET28a (Merck Biosciences) using the same restriction-enzyme sites such that the recombinant protein would contain an N-terminal His tag with a thrombin cleavage site. *Escherichia coli* Rosetta (DE3) pLysS cells (Merck Biosciences) were transformed with the plasmid and single colonies from a Luria-Bertani (LB) agar plate containing 30 µg l⁻¹ kanamycin were used to inoculate a small LB-kanamycin culture, which was grown overnight at 310 K. 5 ml aliquots of the overnight culture were used the following day to inoculate 500 ml LB-kanamycin cultures. Cultures were grown at 310 K until an OD₆₀₀ of 0.6–0.8 was reached, at which point protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The temperature was reduced to 289 K and the cultures were shaken at 200 rev min⁻¹ overnight. The cells were harvested by centrifugation at 277 K.

2.2. Purification

Cells were resuspended in buffer A (0.5 M NaCl, 20 mM HEPES pH 8 and 20 mM imidazole) supplemented with a cocktail of protease inhibitors (Complete EDTA-free tablets, Roche). Cells were disrupted using a French press (10.3 MPa) and were centrifuged at 15 000 rev min⁻¹ for 1 h (Sorvall SS34) at 277 K; the supernatant was cleared with a 0.45 µm filter (Millipore). The filtrate was loaded onto a 5 ml HisTrap Q HP column (GE Healthcare) equilibrated with buffer A for Ni²⁺-affinity purification and the bound protein was eluted with an increasing proportion of buffer B (0.5 M NaCl, 20 mM HEPES pH 8 and 500 mM imidazole). Fractions containing MTH909 were pooled, buffer-exchanged into 50 mM NaCl and 20 mM HEPES pH 8 and then concentrated using Vivascience 5 kDa molecular-weight cutoff concentrators. The N-terminal His tag of the protein was removed using thrombin (BD Biosciences) in the reaction buffer (0.15 M NaCl, 25 mM CaCl₂ and 20 mM Tris pH 8.5) with a MTH909 concentration of 1.6 mg ml⁻¹ and 0.1 mU of thrombin per microgram of MTH909. The protein concentration was determined from the absorption at 280 nm using an extinction coefficient of 0.378 (ExpASY

Proteomics Server; <http://ca.expasy.org/>). The reaction was left for 24 h at room temperature in a rotator at 7 rev min⁻¹ and was confirmed to be complete by SDS-PAGE. The reaction mixture was loaded onto a 5 ml HisTrap Q HP column equilibrated with buffer C (250 mM NaCl and 20 mM Tris pH 8.5); buffer D (250 mM NaCl, 20 mM Tris pH 8.5 and 500 mM imidazole) was used for the gradient. SDS-PAGE showed that MTH909 eluted in the flowthrough fractions (consistent with the removal of the His tag) and these fractions were pooled and buffer-exchanged into buffer E (50 mM NaCl and 20 mM Tris pH 7.5). Pooled protein was loaded onto a Mono-Q column (GE Healthcare) pre-equilibrated with buffer E. The bound protein was eluted with an increasing proportion of buffer F (1 M NaCl and 20 mM Tris pH 7.5). Fractions containing MTH909 were buffer-exchanged into buffer E, concentrated to 10 mg ml⁻¹ and used for crystallization. The purity of MTH909 was analyzed by SDS-PAGE (Fig. 1a).

2.3. Size-exclusion chromatography and multi-angle laser-light scattering (MALLS)

Chromatography was conducted on a Superdex 75 10/300 column (GE Healthcare; bed and void volumes of 24 and 8 ml, respectively) at a flow rate of 0.5 ml min⁻¹ in 250 mM NaCl and 20 mM Tris pH 7.5

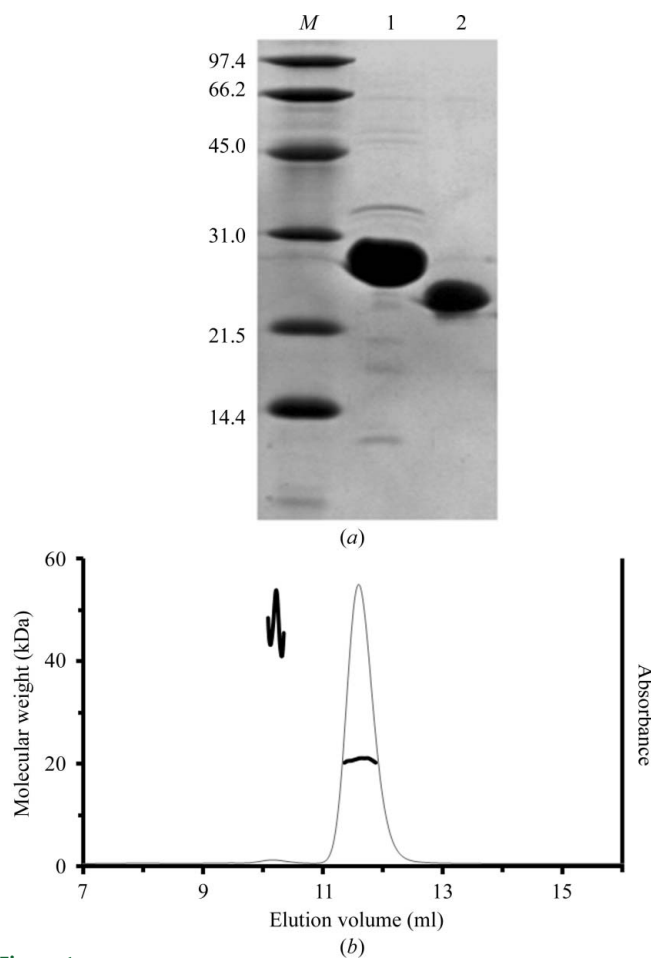


Figure 1 Purification and characterization of MTH909. (a) SDS-PAGE of purified MTH909 before (lane 1) and after (lane 2) the thrombin cleavage step. Lane M contains molecular-weight markers (kDa). (b) Analysis of the oligomerization state of the cleaved MTH909 by combination of size-exclusion chromatography with MALLS. The thin line corresponds to the absorbance monitored at 280 nm. The thick line shows the molecular weight calculated for the eluted species.

and the eluting species were monitored at 280 nm. Light-scattering data were recorded on an in-line Dawn Heleos II laser light-scattering instrument (Wyatt Technology) and the concentrations of the eluting species were measured using an in-line Optilab rEX refractometer (Wyatt Technology). A refractive-index increment (dn/dc) estimate of 0.184 ml g^{-1} was used for the protein (Folta-Stogniew & Williams, 1999).

2.4. Crystallization

Initial crystallization screening was carried out in 96-well sitting-drop plates using a Mosquito Nanolitre Pipetting robot (TTP Labtech) with the Index (Hampton Research), PACT (Newman *et al.*, 2005) and Clear Strategy Screens I and II (Brzozowski & Walton, 2001; Molecular Dimensions) crystallization screens. For crystallization, equal volumes of the protein and reservoir solutions (150 + 150 nl) were mixed and equilibrated against 80 μl reservoir solution at 293 K. Conditions in which small crystals grew were optimized using the same robotic procedure and the resulting crystals were tested using a Rigaku RU-H3R rotating-anode generator equipped with Osmic multilayer optics and a MAR Research MAR345 imaging-plate detector. The best crystal was frozen in liquid nitrogen after soaking in a cryoprotectant solution containing 0.15 M MgCl_2 , 0.1 M Tris pH 8, 15% (v/v) glycerol and 26% (v/v) PEG 6000.

2.5. Data collection and processing

X-ray data were collected from a single crystal at the ESRF station ID23-1. Data collection was performed at 100 K using an oscillation range of 0.5° per image with a total crystal rotation of 180° . The diffraction images were indexed and integrated with *DENZO* and scaled with *SCALEPACK* (Otwinowski & Minor, 1997). Initial attempts to determine the structure by molecular replacement were made using *BALBES* (Long *et al.*, 2008). These were followed by attempts using *MOLREP* (Vagin & Teplyakov, 1997) and *Phaser* (McCoy *et al.*, 2007) with the aligned regions of *B. anthracis* ThiI (PDB code 2c5s), *P. horikoshii* ThiI (PDB code 1vbk) and human Pus10 (PDB code 2v9k) as search models.

3. Results and discussion

3.1. Expression and purification

MTH909 was successfully cloned and overexpressed in a soluble form. The protein was purified to homogeneity by a three-step procedure involving two rounds of Ni^{2+} -affinity chromatography sepa-



Figure 2
The best crystal of MTH909, from which the X-ray data were collected.

rated by cleavage of the His tag and followed by a final step of anion-exchange chromatography (Fig. 1a). The molecular weight of the protein was determined by mass spectrometry to be 20 596 Da, which is in agreement with the calculated value of 20 577 Da for the full-length protein containing the GSH residues of the thrombin cleavage site.

3.2. Oligomerization state of MTH909

MTH909 eluted largely as a single peak during size-exclusion chromatography (Fig. 1b). According to MALLS measurements, the major peak contains species with an average molecular weight of $20.9 \pm 0.4 \text{ kDa}$, which is close to the theoretical molecular weight of 20.6 kDa for the monomer of MTH909. The eluted species in the very minor peak have an average molecular weight of $46.6 \pm 4.7 \text{ kDa}$, which corresponds to either dimers of MTH909 or impurities (Fig. 1a). These data show that under the tested conditions MTH909 exists almost exclusively as a monomer in solution.

3.3. Crystallization

The initial crystallization trials resulted in crystals being obtained under four different conditions. Two of these conditions in which the crystals were very small were optimized further. The crystals grown with 0.1 M Tris pH 8.5, 2 M ammonium sulfate (Index condition No. 6) could not be reproduced, while optimization of the crystals grown with 0.1 M HEPES pH 7.5, 0.2 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 25% (w/v) PEG 3350 (Index condition No. 84) led to crystals with poor diffraction quality (4.5 Å resolution). The two other conditions both produced single crystals that were large enough for data collection. The crystal grown with 0.1 M bis-tris pH 6.5 and 2 M ammonium sulfate (Index condition No. 4) diffracted to 3.5 Å resolution but the diffraction pattern was 'split', indicating that it was not a single crystal. The crystal grown with 0.2 M MgCl_2 , 0.1 M Tris pH 8 and 20% (w/v) PEG 6000 (PACT condition D10) had dimensions of $0.1 \times 0.1 \times 0.1 \text{ mm}$ (Fig. 2) and

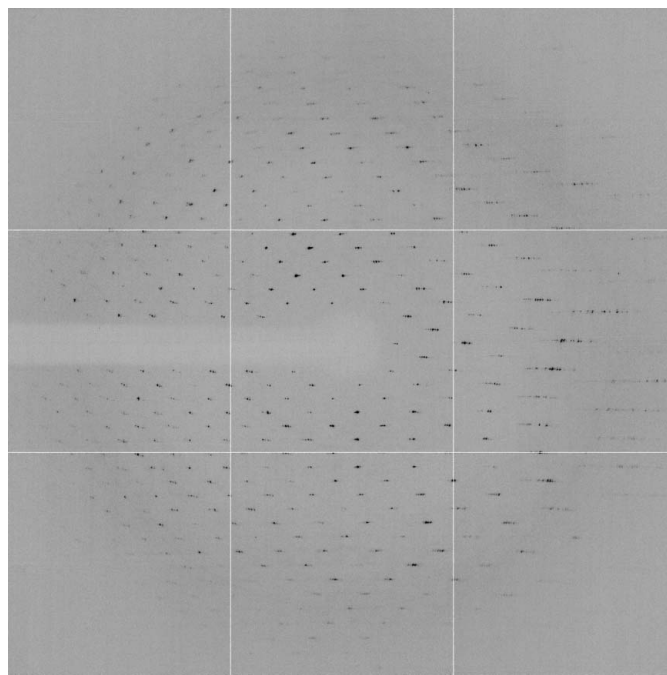


Figure 3
The diffraction image obtained at the ESRF station ID23-1. Diffraction data were collected at a wavelength of 0.9184 \AA with a crystal-to-detector distance of 488 mm. The resolution at the edges of the image is 3.0 Å.

Table 1

X-ray data statistics for the crystal of native MTH909.

Values in parentheses are for the highest resolution shell.

X-ray source	ID23-1, ESRF
Wavelength (Å)	0.9184
Temperature (K)	100
Space group	$P6_122$ or $P6_522$
Unit-cell parameters (Å)	$a = b = 69.9$, $c = 408.5$
Resolution range (Å)	25–2.85 (2.95–2.85)
No. of unique reflections	14804 (1296)
R_{merge}^\dagger (%)	10.0 (55.6)
Completeness (%)	98.6 (89.3)
Redundancy	9.8 (5.9)
Average $I/\sigma(I)$	19.3 (2.4)
Wilson B factor (Å ²)	37

$^\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the measured intensity of each reflection and $\langle I(hkl) \rangle$ is the intensity averaged over multiple observations of symmetry-related reflections.

diffracted to 3.2 Å resolution in-house. From this crystal, essentially complete data to 2.85 Å resolution were collected (Fig. 3).

3.4. X-ray data collection and preliminary analysis

Analysis of merging statistics and systematic absences indicated that the crystals belonged to space group $P6_122$ (or $P6_522$), with unit-cell parameters $a = b = 69.9$, $c = 408.5$ Å. Cumulative intensity distributions indicated no twinning. Data-collection and processing statistics are presented in Table 1. Three molecules in the asymmetric unit would result in a solvent content of 46.8% (Matthews, 1968). The self-rotation function did not reveal any significant peaks apart from those resulting from the crystallographic symmetry, suggesting that the potential noncrystallographic rotation axes could be parallel to the crystallographic axes. Attempts to solve the structure by molecular replacement using the known structures of THUMP domain-containing proteins proved unsuccessful, most probably because of the very low sequence identity between MTH909 and the search models (sequence identities of 14% or less).

In conclusion, MTH909 was successfully purified, crystallized and X-ray data from a native crystal were collected to 2.85 Å resolution. Future attempts to solve the structure will be focused on crystallizing SeMet MTH909 or preparing heavy-atom derivatives. Comparison of the structure of MTH909 with structures of other proteins containing

the THUMP domain will provide further understanding of its RNA-binding ability.

We would like to thank Johan Turkenburg and Sam Hart for help during data collection and the ESRF (Grenoble) for excellent data-collection facilities. This work was supported by Fundação para a Ciência e a Tecnologia, Portugal (studentship SFRH/BD/17372/2004 to APGS), by the BBSRC (PhD studentship to RTB) and by the Wellcome Trust (fellowship 081916 to AAA).

References

- Agris, P. F. (2008). *EMBO Rep.* **9**, 629–635.
- Aravind, L. & Koonin, E. V. (2001). *Trends Biochem. Sci.* **26**, 215–217.
- Brzozowski, A. M. & Walton, J. (2001). *J. Appl. Cryst.* **34**, 97–101.
- Folta-Stogniew, E. & Williams, K. (1999). *J. Biomol. Tech.* **10**, 51–63.
- Gabant, G., Auxilien, S., Tuszynska, I., Locard, M., Gajda, M. J., Chaussinand, G., Fernandez, B., Dedieu, A., Grosjean, H., Golinelli-Pimpaneau, B., Bujnicki, J. M. & Armengaud, J. (2006). *Nucleic Acids Res.* **34**, 2483–2494.
- Gustafsson, C., Reid, R., Greene, P. J. & Santi, D. V. (1996). *Nucleic Acids Res.* **24**, 3756–3762.
- Helm, M. (2006). *Nucleic Acids Res.* **34**, 721–733.
- Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. & Yokoyama, S. (2003). *Cell*, **113**, 383–394.
- Johansson, M. J. & Bystrom, A. S. (2004). *RNA*, **10**, 712–719.
- Limbach, P. A., Crain, P. F. & McCloskey, J. A. (1994). *Nucleic Acids Res.* **22**, 2183–2196.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 125–132.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCleverty, C. J., Hornsby, M., Spraggon, G. & Kreuzsch, A. (2007). *J. Mol. Biol.* **373**, 1243–1254.
- 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.
- Mueller, E. G. & Palenchar, P. M. (1999). *Protein Sci.* **8**, 2424–2427.
- Nakanishi, K. & Nureki, O. (2005). *Mol. Cell*, **19**, 157–166.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* **D61**, 1426–1431.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Sugahara, M., Murai, S., Sugahara, M. & Kunishima, N. (2007). *Acta Cryst.* **F63**, 56–58.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Waterman, D. G., Ortiz-Lombardia, M., Fogg, M. J., Koonin, E. V. & Antson, A. A. (2006). *J. Mol. Biol.* **356**, 97–110.