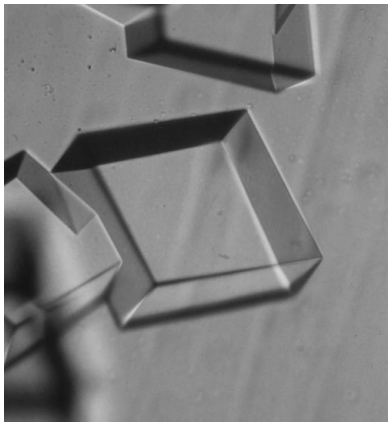
Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Guangteng Wu,[a] Qichen Huang,[a]* Youqi Tang,[a] Hideaki Unno[b] and Masami Kusunoki[b]

[a]State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry, Peking University, Beijing 100871, People's Republic of China, and [b]Research Center for Structural Biology, Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan

Correspondence e-mail: qchuang@chem.pku.edu.cn

Received 29 September 2004
Accepted 23 November 2004
Online 24 December 2004



Crystallization and preliminary crystallographic study of a recombinant predicted acetamidase/formamidase from the thermophile *Thermoanaerobacter tengcongensis*

No crystal structures are yet available for homologues of a predicted acetamidase/formamidase (Amds/Fmds) from the archaeon *Thermoanaerobacter tengcongensis*. The Amds/Fmds gene was cloned and expressed as a soluble protein in *Escherichia coli*. Native Amds/Fmds and its SeMet-substituted form were purified and crystallized by vapour diffusion in hanging drops at 296 K. The native crystals, which were grown in PEG 8000, belong to the monoclinic space group $P2_1$, with unit-cell parameters a = 41.23 (3), b = 152.88 (6), c = 100.26 (7) Å, β = 99.49 (3)°. The diffraction data were collected to 2.00 Å resolution using synchrotron radiation. Based on a predicted solvent content of 50%, a Matthews coefficient of 2.44 Å^3 Da^{-1} and two main peaks in the self-rotation function, the asymmetric unit is predicted to contain two dimers of the 32 kDa native protein. MAD data were collected for the SeMet protein, but the corresponding crystals display different unit-cell parameters and appear to contain four dimers in the asymmetric unit.

1. Introduction

Structural genomics aims towards determining a set of protein structures that will represent all domain folds present in the biosphere (Heinemann *et al.*, 2000). How to choose representative proteins for three-dimensional structure determination is an important issue in this research field. A convenient route towards fast structure-determination targets proteins from hyperthermophilic bacteria or archaea, because they can be easily purified from recombinant *Escherichia coli* cells and lend themselves especially to crystallization or NMR structure determination (Heinemann *et al.*, 2000).

In our group, genes from the thermophilic archaeon *Thermoanaerobacter tengcongensis* strain MB4T (Xue *et al.*, 2001) have been chosen for study. *T. tengcongensis*, which is native to China, was collected from the thermal spring at Tengchong (Yunnan, China). The complete sequence of its genome has been obtained (Genbank accession No. AE008691; Bao *et al.*, 2002). Several proteins from *T. tengcongensis* have been studied structurally and biochemically, such as thermostable hypoxanthine-guanine phosphoribosyltransferase (Chen *et al.*, 2003; You *et al.*, 2003), esterase (Zhang *et al.*, 2003), multisubunit membrane-bound [NiFe] hydrogenase and NADH-dependent Fe-only hydrogenase (Soboh *et al.*, 2004).

The predicted acetamidase/formamidase (Amds/Fmds) from *T. tengcongensis* contains 299 amino-acid residues, including eight methionines, and has a molecular weight of 31 880 Da. Its gene (Gene code No. TTE1919) is one of 20 genes from the genome of *T. tengcongensis* that were chosen for expression and structure determination by bioinformatics methods. Our criterion for targets of interest was that the proteins be likely to exhibit a novel protein fold and therefore add to the international structural genomics initiative. A BLASTP search (protein–protein BLAST; Altschul *et al.*, 1997) comparing the Amds/Fmds sequence with proteins of known structure revealed no strong matches; the longest aligned subsequence, between Amds/Fmds and 1sfo, contains 60 residues, of which 18 (30%) are identical. Therefore, we concluded that no crystal structures of Amds/Fmds homologues are currently available.

In addition to the predicted biological functions of targeted proteins, homology to human proteins was a criterion for selection. A

TBLASTN search (translated BLAST; Altschul *et al.*, 1997) against the est_human database predicted Amds/Fmds to be homologous to the human proteins derived from the following GenBank sequences: BQ939495, BM543991, D80153, D59772, AI469114 and BI762629.

In this paper, we describe the crystallization and preliminary X-ray diffraction analysis of Amds/Fmds from *T. tengcongensis*.

2. Expression and purification

The predicted Amds/Fmds was cloned and expressed as a soluble protein in *Escherichia coli* strain Rosetta (DE3) with a 20-residue His tag at its N-terminus. The SeMet-substituted protein was expressed in the methionine-auxotrophic strain B834 (DE3). After cell lysis, the native protein was purified by Ni–NTA His-binding affinity chromatography followed by gel filtration on a Sephacryl S-200 column in 50 m*M* Tris–HCl pH 7.6, 130 m*M* NaCl and 18 m*M* β-mercaptoethanol. The purification of the SeMet-substituted protein followed the same procedure and conditions, but with 20 m*M* DTT as antioxidant instead of β-mercaptoethanol, plus 0.2 m*M* EDTA as a chelator to complex metal ions that might assist oxidation (Doublié, 1997). The purity of the two recombinant proteins was characterized by SDS–PAGE.

3. Crystallization of native and SeMet-substituted Amds/Fmds

The crystallization of the predicted native Amds/Fmds and its SeMet-substituted form were carried out using the hanging-drop vapour-diffusion method. The initial crystallization screening for the native protein used the Crystal Screen 1 kit (Hampton Research). 1 µl protein solution (10 mg ml^{-1}) containing 50 m*M* Tris–HCl pH 7.6 with 130 m*M* NaCl and 12 m*M* β-mercaptoethanol was mixed with an equal volume of reservoir solution and equilibrated against 0.5 ml reservoir solution at 296 K. Rod-shaped microcrystals were obtained with a reservoir solution containing 0.2 *M* calcium acetate, 0.1 *M* sodium cacodylate and 18%(*w*/*v*) PEG 8000 pH 6.5. After further improvement, maximum dimensions of 0.6 × 0.5 × 0.08 mm were obtained in 3 d at pH 6.2. A crystal with dimensions of 0.3 × 0.3 × 0.1 mm is shown in Fig. 1.

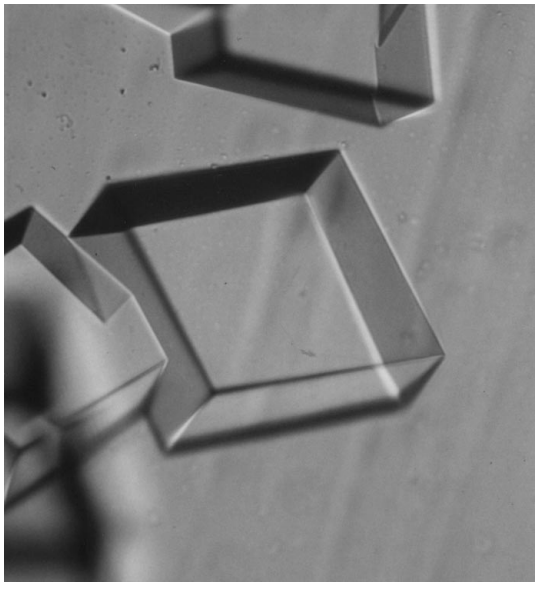


Figure 1
Crystal of the predicted Amds/Fmds.

Table 1
Diffraction data statistics of the predicted Amds/Fmds.

Values in square brackets refer to the highest resolution shell (2.07–2.00 Å).

X-ray source	Photon Factory, BL6A
Detector	ADSC Quantum 4R CCD
X-ray wavelength (Å)	1.000
Temperature (K)	100
Space group	$P2_1$
Unit-cell parameters†	
a (Å)	41.23 (3)
b (Å)	152.88 (6)
c (Å)	100.26 (7)
β (°)	99.49 (3)
Resolution limit (Å)	2.00
Total reflections	288176 [17545]
Unique reflections	81395 [6748]
Observed reflections [$I/\sigma(I) > 2$]	68743 [4382]
$R_{\rm merge}$‡ (%)	4.1 [24.6]
Completeness (%)	97.2 [80.5]
Completeness [$I/\sigma(I) > 2$] (%)	82.1 [52.3]
Multiplicity	3.5 [2.6]
$\langle I/\sigma(I)\rangle$	17.8 [2.5]

† Standard deviations are given in parentheses. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I(hkl)_i - \langle I(hkl)\rangle| / \sum_{hkl} I(hkl)$, where $I(hkl)_i$ is the *i*th measurement of the intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean intensity of reflection *hkl*.

The SeMet-substituted protein did not crystallize well using the same condition as for the native protein. The SeMet protein was concentrated to 7.5 mg ml^{-1} in 50 m*M* Tris–HCl, 130 m*M* NaCl, 20 m*M* DTT and 0.2 m*M* EDTA. Using sodium acetate instead of sodium cacodylate and 5 m*M* TCEP (Sigma, Product No. C4706) as an antioxidant in the reservoir solution, thin plate-shaped and needle-shaped crystals were obtained at pH 5.5–6.0. Single crystals with maximum dimensions of 0.6 × 0.3 × 0.02 mm were obtained in hanging drops in 3 d using 18%(*w*/*v*) PEG 8000, 0.2 *M* calcium acetate and 0.1 *M* sodium acetate pH 5.8 (Fig. 2).

4. Data collection and processing

Owing to the presence of 18%(*w*/*v*) PEG 8000, which functions as a cryoprotectant, using the precipitant solution worked well when cryocooling the crystals for data collection. The native and SeMet crystals were directly mounted in nylon CryoLoops (Hampton Research) and flash-frozen in liquid nitrogen and were then placed into a nitrogen stream at 100 K. Data were collected at beamline BL6A of the Photon Factory at the High Energy Acceleration Research Organization, Tsukuba, Japan with an ADSC Quantum 4R CCD camera (Watanabe *et al.*, 1995). Image data were processed with

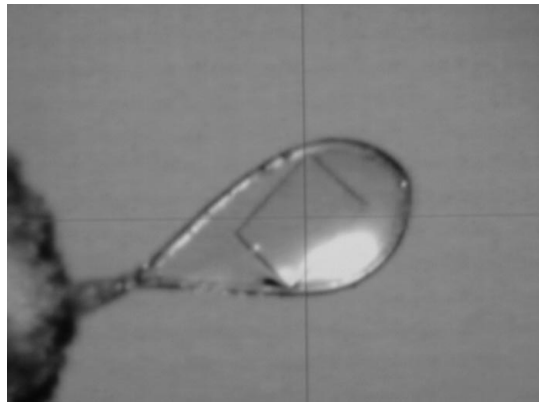


Figure 2
Crystal of SeMet-substituted Amds/Fmds frozen in a nylon cryoloop mounted on a goniometer in a nitrogen stream at 100 K.

Table 2
MAD data-collection statistics of SetMet-substituted predicted Amds/Fmds.

Values in square brackets refer to the highest resolution shell.

	Low-energy remote	Inflection	Peak	High-energy remote
X-ray wavelengh (Å)	1.0700	0.97938	0.97860	0.9000
Temperature (K)	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters†				
a (Å)	111.83 (2)	111.76 (2)	111.80 (3)	111.85 (4)
b (Å)	74.80 (2)	74.79 (2)	75.02 (2)	75.00 (3)
c (Å)	149.09 (6)	148.99 (7)	149.02 (8)	149.23 (11)
β (°)	100.56 (1)	100.52 (1)	100.53 (1)	100.57 (1)
Resolution limit (Å)	50.00–2.64 [2.73–2.64]	50.00–2.50 [2.59–2.50]	50.00–2.70 [2.80–2.70]	50.00–2.80 [2.90–2.80]
Total reflections	58236 [19603]	307468 [28353]	411538 [40020]	217301 [19529]
Unique reflections	71429 [6535]	84521 [8339]	67131 [6670]	60455 [5918]
Observed reflections [$I/\sigma(I) > 2$]	62185 [4319]	72419 [5269]	61300 [4956]	51968 [3793]
Completeness	99.0 [91.3]	99.9 [99.4]	99.9 [100]	99.9 [98.8]
Completeness [$I/\sigma(I) > 2$] (%)	96.2 [60.3]	85.6 [62.8]	91.2 [74.3]	85.9 [63.3]
Multiplicity	3.6 [3.0]	3.6 [3.4]	6.1 [6.0]	1.9 [1.8]
$R_{\rm merge}$‡ (%)	8.2 [22.2]	8.7 [31.8]	9.2 [25.6]	10.1 [31.1]
$\langle I/\sigma(I)\rangle$	9.1 [2.6]	8.6 [2.1]	11.7 [4.0]	7.7 [2.2]
$\Delta f'$		−8.60	−6.502	
$\Delta f''$		2.288	3.950	

† Standard deviations are given in parentheses. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I(hkl)_i - \langle I(hkl)\rangle|/\sum_{hkl} I(hkl)$, where $I(hkl)_i$ is the ith measurement of the intensity of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl.

the *HKL*2000 suite (Otwinowski & Minor, 1997). Diffraction data statistics are summarized in Table 1 for the native Amds/Fmds protein and in Table 2 for its SeMet-substituted form.

The space group of the native crystal was determined to be monoclinic $P2_1$. Assuming the presence of two dimers, one tetramer or four monomers of Amds/Fmds in the asymmetric unit, the value of the Matthews constant $V_{\rm M}$ (Matthews, 1968) is 2.44 Å^3 Da^{-1}, corresponding to a solvent content of 50%, both of which are within the normal values for protein crystals.

The space group of the SeMet-substituted crystal also proved to be $P2_1$, but with different unit-cell parameters. Assuming the presence of four dimers, two tetramers, one octamer or eight monomers of the protein in the asymmetric unit, the value of the Matthews constant $V_{\rm M}$ is 2.41 Å^3 Da^{-1}, corresponding to a solvent content of 49%.

As the 299-residue sequence of the predicted Amds/Fmds contains eight methionines, there are 64 Se sites in the asymmetric unit of the SeMet crystal. The Se sites were determined with *SHELXD* (Schneider & Sheldrick, 2002). The 64 Se sites appear to be grouped into four clusters, consistent with the presence of four subunits in the asymmetric unit. Analysis of the self-rotation function shows two main peaks for the native crystal and four peaks for the SeMet crystal, suggesting that there are two dimers in the asymmetric unit of the native crystal and four dimers in the asymmetric unit of the SeMet crystal. Further model building and structure refinement are currently in progress.

We thank Professor Luhua Lai and Dr Jianfeng Pei, Institute of Physical Chemistry, Peking University, China for helping to choose the genes from the genome of *T. tengcongensis*, and Professor Ruseng Chen, Institute of Biophysics, Academia Sinica for providing the genomic DNA of *T. tengcongensis*. We also thank Drs Soichi Wakatsuki and Noriyuki Igarash for their kind help with data collection at the Photon Factory. This work was supported by grants from 'Structural Genomics' of the High Technology Development Program of China '863 Project of China' to QH and partially by Grants-in-Aid (Nos. 10558109 and 12480181) from the Japan Society for the Promotion of Science to MK.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bao, Q. *et al.* (2002). *Genome Res.* **12**, 689–700.
Chen, Q., You, D., Hu, M., Gu, X., Luo, M. & Lu, S. (2003). *Protein Expr. Purif.* **32**, 239–245.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–531.
Heinemann, U., Frevert, J., Hofmann, K. P., Illing, G., Maurer, C., Oschkinat, H. & Saenger, W. (2000). *Prog. Biophys. Mol. Biol.* **73**, 347–362.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–493.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Soboh, B., Linder, D. & Hedderich, R. (2004). *Microbiology*, **150**, 2451–2463.
Watanabe, N., Nakagawa, A., Adachi, S. & Sakabe, N. (1995). *Rev. Sci. Instrum.* **66**, 1824–1826.
Xue, Y., Xu, Y., Liu, Y., Ma, Y. & Zhou, P. (2001). *Int. J. Syst. Evol. Microbiol.* **51**, 1335–1341.
You, D., Chen, Q., Liang, Y., An, J., Li, R., Gu, X., Luo, M. & Su, X.-D. (2003). *Acta Cryst.* D**59**, 1863–1865.
Zhang, J., Liu, J., Zhou, J., Ren, Y., Dai, X. & Xiang, H. (2003). *Biotechnol. Lett.* **25**, 1463–1467.