Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary X-ray analysis of the novel DEAD protein BstDEAD from *Bacillus stearothermophilus*

Andrew B. Carmel and Brian W. Matthews*

Institute of Molecular Biology, Howard Hughes Medical Institute and Departments of Chemistry and Physics, 1229 University of Oregon, Eugene, OR 97403-1229, USA

Correspondence e-mail: brian@uoxray.uoregon.edu

Received 22 July 2003
Accepted 18 August 2003

DEAD proteins are members of a large and diverse family of RNA helicases that use energy from ATP hydrolysis to unwind short regions of duplex RNA. BstDEAD from *Bacillus stearothermophilus* is a 436-amino-acid protein and a representative member of the DEAD protein family. In addition to the mechanistic core common to DEAD proteins, BstDEAD has a unique ~60-amino-acid C-terminal extension that may denote a specific biological role. BstDEAD has been crystallized in space group $P4_{1/3}2_12$, with unit-cell parameters $a = b = 100.3$, $c = 110.6$ Å and one molecule per asymmetric unit. It is the first DEAD protein to be crystallized containing a unique extension outside of the core.

1. Introduction

The rearrangement, disruption or modification of structured RNA molecules and/or RNA–protein interactions has been attributed to the mechanistic activity of RNA helicases. Those classified within helicase superfamily II (SFII) include the DEAD, DEAH and DExH subfamilies, collectively known as DExD/H proteins and originally named for the Asp-Glu-Ala-Asp (DEAD) sequence motif (Linder *et al.*, 1989; Gorbalenya & Koonin, 1993; Tanner & Linder, 2001). It is generally believed that DExD/H proteins are nucleic acid-dependent ATPases and ATP-dependent RNA helicases, although only a small number of DExD/H proteins have demonstrated duplex RNA-unwinding activity *in vitro*. These proteins are characterized based on homology to a helicase 'core' containing eight highly conserved signature sequences that fold into two structurally similar domains (reviewed in Tanner & Linder, 2001). While the core includes the minimal determinants necessary for helicase activity, DExD/H proteins are often targeted for specific functions by unique extensions found on either termini. These extensions, while highly variable, may further define distinct subfamilies of DExD/H proteins all having related biological activity. Even as numerous DExD/H proteins have been identified by sequence homology, the biochemical activities and the precise biological roles of the majority of these proteins remain unknown.

The DEAD subfamily, characterized by the conserved Asp-Glu-Ala-Asp sequence, represents the largest and best-studied family of RNA helicases. Using a sequence-homology search, we identified what was at the time an uncharacterized member of the DEAD protein family from the thermophilic bacterium *Bacillus stearothermophilus*, called BstDEAD. BstDEAD is a 436-amino-acid protein (~50 kDa), with all eight of the conserved helicase motifs contained within an ~370-amino-acid core. The C-terminal domain of BstDEAD includes a highly basic ~60-amino-acid extension that may, through interactions with a yet-to-be-identified RNA substrate, target it for some unique biological activity. Furthermore, we observed that the C-terminal extension of BstDEAD is highly homologous to the C-terminal extensions of a small number of uncharacterized bacterial DEAD proteins. This C-terminal extension may define a novel subfamily of DEAD proteins, all of which may interact with the same RNA substrate.

In recent years, there has been considerable interest in DEAD proteins owing to their near-ubiquitous association with biological processes involving RNA. This interest notwithstanding, the crystal structures of only two DEAD proteins have been reported: yeast eIF4A (Caruthers *et al.*, 2000) and mjDEAD from *Methanococcus janaschii* (Story *et al.*, 2001). The overall folds of these proteins are highly conserved, consisting of two independent domains connected by a flexible linker. We report here the purification, crystallization and preliminary X-ray analysis of BstDEAD. This is the first report describing crystals of a DEAD protein with a significant extension that extends beyond the core.

2. Experimental procedures and results

2.1. Cloning, expression and purification

The 1308 bp gene coding for the full-length BstDEAD protein was originally identified by sequence homology in *B. stearothermophilus* using the previously characterized DEAD protein YxiN from *B. subtilis* as a search model (Kossen & Uhlenbeck, 1999). The BstDEAD

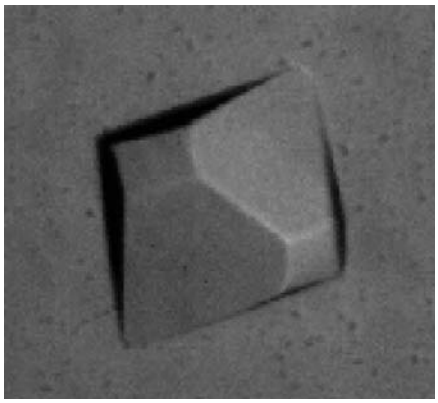

Figure 1
Crystal of BstDEAD. The longest dimension is approximately 0.3 mm.

gene was amplified using PCR from genomic *B. stearothermophilus* DNA. The isolated gene was inserted into the pET22b expression vector (Novagen) under control of the T7 promoter. Electrocompetent *Escherichia coli* BL21(DE3) gold cells (Stratagene) transformed with this plasmid were grown at 310 K following the method for selenomethionine incorporation described by Gassner *et al.* (1999). Overexpression of selenomethionine-substituted BstDEAD protein with a C-terminal hexahistidine tag was induced with IPTG (final concentration 1 m*M*) at an OD_{600} of 1.5–1.8 for 15 h at 303 K. Cells were harvested by centrifugation at 7000 rev min^{-1} for 15 min in a Beckman JA-20 rotor and resuspended in 20 m*M* Tris pH 7.9, 1 *M* KCl, 10 m*M* imidazole, 10%(*v*/*v*) glycerol, 0.1% Triton X-100. The cells were lysed by sonication and cell debris was removed by centrifugation at 17 000 rev min^{-1} for 40 min in a Beckman JA-20 rotor. The resulting supernatant containing soluble BstDEAD protein was loaded onto an Ni–NTA agarose (Qiagen) column. Purified BstDEAD protein was eluted using a 10–200 m*M* imidazole gradient followed by dialysis into 20 m*M* HEPES pH 7.5, 200 m*M* NaCl, 1 m*M* DTT, 1 m*M* EDTA. Further purification involved heating to 333 K for 30 min to denature extraneous proteins, followed by cooling on ice for 30 min and centrifugation for 1 h at 17 000 rev min^{-1} in a JA-20 rotor to remove contaminants. A typical 4 l preparation yielded ∼50 mg of pure protein, which was concentrated to ∼15 mg ml^{-1} prior to crystallization.

Table 1
Crystallographic data-collection and phasing statistics.

The four wavelengths, λ_1–λ_4, were based on a fluorescence scan of the Se *K* edge and correspond to the peak, inflection, high-energy remote and low-energy remote, respectively. Values in parentheses represent statistics for the highest resolution shell (3.28–3.20 Å).

Data set	λ_1	λ_2	λ_3	λ_4
Source	SSRL 9-2	SSRL 9-2	SSRL 9-2	SSRL 9-2
Temperature (K)	100	100	100	100
Wavelength (Å)	0.9792	0.9794	0.9185	0.992
Resolution (Å)	25.0–3.2	25.0–3.2	25.0–3.2	25.0–3.2
Measured reflections	72928	72954	71363	73031
Unique reflections	9752	9756	9733	9763
R_{sym} (%)	7.5 (36.3)	7.2 (33.0)	7.7 (36.1)	7.1 (33.0)
Completeness (%)	99.6 (99.6)	99.6 (100)	99.6 (100)	99.6 (99.6)
Anomalous completeness (%)	99.8 (100)	99.9 (100)	99.8 (100)	99.9 (100)
Average $I/\sigma(I)$	7.0 (2.1)	7.3 (2.3)	6.4 (1.9)	7.4 (2.2)

2.2. Crystallization

Prior to crystallization, the molecular weight of BstDEAD was determined using a Voyager-DE MALDI–TOF mass spectrometer (Perseptive Biosystems). The calculated molecular weight (∼51 300 Da) indicated essentially 100% SeMet incorporation. Crystals of full-length SeMet-substituted BstDEAD protein grew using the hanging-drop vapor-diffusion method. The initial crystallization conditions were determined using a grid screen with ammonium sulfate as a precipitant and buffers of varying pH. The largest BstDEAD crystals grew to ∼0.2 × 0.3 × 0.2 mm in 38–45%(*v*/*v* of saturated solution) ammonium sulfate, 40 m*M* MES pH 6.0 in 2–3 weeks at room temperature (Fig. 1).

2.3. Data collection and analysis

Prior to data collection, single crystals were rapidly swept through mother liquor containing ∼20%(*v*/*v*) glycerol as a cryoprotectant and were flash-frozen in liquid nitrogen at 100 K. Full-length BstDEAD crystallizes in space group $P4_{1/3}2_12$, with unit-cell parameters $a = b = 100.3$, $c = 110.6$ Å, and appears to contain one BstDEAD molecule per asymmetric unit ($V_M = 2.71$ Å^3 Da^{-1} with 54% solvent content; Matthews, 1968).

Diffraction data at four wavelengths were collected on beamline 9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) (Table 1). Diffraction data extended to ∼3.2 Å resolution and were processed using *MOSFLM*, *SCALA* and *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). *SOLVE* (Terwilliger & Berendzen, 1999) suggested the locations of ten of the 12 Se sites in space group $P4_12_12$, but these have not to date led to a model that can be satisfactorily refined. Efforts are ongoing to increase the resolution of the data and to improve the model.

We would like to thank the staff at the Stanford Synchrotron Radiation Laboratory for assistance with data collection. This work was supported in part by NIH grant GM20066 to BWM.

References

Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 13080–13085.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Gassner, N. C., Baase, W. A., Hausrath, A. C. & Matthews, B. W. (1999). *J. Mol. Biol.* **294**, 17–20.
Gorbalenya, A. E. & Koonin, E. V. (1993). *Curr. Opin. Struct. Biol.* **3**, 419–429.
Kossen, K. & Uhlenbeck, O. C. (1999). *Nucleic Acids Res.* **27**, 3811–3820.
Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., Slonimski, P. P. (1989). *Nature (London)*, **337**, 121–122.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Story, R. M., Li, H. & Abelson, J. N. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
Tanner, N. K. & Linder, P. (2001). *Mol. Cell*, **8**, 251–262.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.