Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification and preliminary X-ray analysis of a fibrillarin homolog from *Methanococcus jannaschii*, a hyperthermophile

Hongming Wang, Hisao Yokota, Rosalind Kim and Sung-Hou Kim*

Department of Chemistry and Physical Biosciences Division, Lawrence Berkeley National Laboratory, University of California Berkeley, Berkeley, CA 94720-5230, USA

Correspondence e-mail: shk@lbl.gov

Received 18 March 1998
Accepted 27 May 1998

Fibrillarin plays a central role in ribosome biogenesis as a ribosomal RNA-processing protein. A *Methanococcus jannaschii* homolog of fibrillarin has been overexpressed, purified and crystallized. Crystals belong to the *C*2 space group with unit-cell parameters $a = 121.4$, $b = 43.2$, $c = 55.3$ Å, $\beta = 96.9^\circ$. Under flash-frozen conditions and using synchrotron radiation, the crystals diffract to 1.8 Å resolution. For structural determination, a selenomethionine derivative of the protein has also been crystallized.

1. Introduction

Ribosome biogenesis in eukaryotes takes place in a specialized nuclear structure, the nucleolus. Apart from the synthesis of ribosomal proteins and primary pre-ribosomal RNA (pre-rRNA) transcripts, ribosomal maturation requires three major activities: (i) processing of the pre-rRNA into smaller transcripts, (ii) modification of the processed pre-rRNA and (iii) assembly of the ribosomal proteins and rRNA into ribosomal subunits. These three activities are catalyzed by supramolecular complexes which are themselves ribonucleoproteins (Fournier & Maxwell, 1993). Recent biochemical analyses of these nucleolar complexes have revealed that they contain a family of small nucleolar RNAs (snoRNAs), ranging in size from 90 to 600 nucleotides, which associate with highly conserved proteins, including fibrillarin (Fournier & Maxwell, 1993).

Fibrillarin forms part of the dense-staining fibrillar regions of the nucleolus as detected in electron micrographs. Although the biochemical function of fibrillarin is not well understood, an *in vivo* mutational analysis in yeast has shown that fibrillarin mutants can affect each of the three post-transcriptional activities required for ribosomal maturation, including pre-rRNA cleavage, rRNA methylation and the actual assembly of the ribosomal subunits (Tollervey *et al.*, 1991, 1993). Evidence to support fibrillarin's role in RNA binding and modification comes from *in vitro* studies that show fibrillarin coprecipitates with small snoRNAs containing conserved sequences designated as C and D boxes (Baserga *et al.*, 1991). In many instances, these boxes are thought to guide rRNA methylation (Cavaille *et al.*, 1996; Tollervey, 1996).

Antibodies to fibrillarin have been found in the sera of patients with scleroderma, a human autoimmune disease that causes flaking and hardening of the skin (Arnett *et al.*, 1996). Similar antibodies are also elicited after exposure to toxic amounts of mercury and it has been shown that mercury binds fibrillarin at two cysteine residues (Pollard *et al.*, 1997). Structural information on fibrillarin may contribute to an understanding of both its apparent multifunctional role in ribosome biogenesis and its involvement in autoimmune disorders.

We have cloned a homolog of human fibrillarin from an archaeon, *M. jannaschii*, encoded by the gene designated *Mj*697 (Bult *et al.*, 1996; Amiri, 1994). Although the hyperthermophilic methanogen does not contain an organized nucleus (and hence, nucleolus), this 26 kDa protein shares 36% sequence identity and 68% sequence homology with human fibrillarin. It is, however, shorter than the human version (230 amino acids *versus* 321 amino acids in humans) and lacks the arginine- and glycine-rich sequence that may locate fibrillarin in the nucleolus (Schimmang *et al.*, 1989).

2. Experimental

2.1. Strains, plasmids and oligonucleotides

The *Mj*697 gene was amplified using the polymerase chain reaction (PCR) with *M. jannaschii* genomic DNA as the template and two oligonucleotide primers (A and B). Upstream primer A (5′-GCGGATCCATATGGAAGACATTAAAATCAAAGAG-3′) contains an *Nde*I restriction site (underlined) with an internal ATG translation initiation codon. Downstream primer B (5′-CCGGATCCCTCGAGTCATTTTCCCTCC-

CAAATACCTAC-3′) contains a *Bam*HI restriction site (underlined). The 717 base-pair blunt-ended PCR product was ligated into the *Sma*I site of Bluescript SK+ (Stratagene) and transformed into *E. coli* XL-1 Blue (Stratagene) with blue-white selection for inserts on Xgal plates containing ampicillin (50 mg ml^{-1}). White colonies were picked and the plasmid DNA was sequenced to confirm the identity of the *Mj*697 gene. An *Nde*I/*Bam*HI restriction fragment containing the gene was gel purified and ligated into the pET21a expression vector (Novagen) to construct the fibrillarin homolog expression vector, pET21a/Mj697.

2.2. Bacterial expression and protein purification

pET21a/Mj697 was transformed into *E. coli* BL21 (DE3) cells harboring the plasmid pSJS1240, which encodes tRNAs for the rare *E. coli* isoleucine (ATA) and arginine (AGA) codons (Kim *et al.*, 1999). For overexpression, cells were grown to an OD_{600} of 1.0 and induced with 0.5 m*M* isopropyl-β-D-thiogalactoside (IPTG) for 6 h. Cell pellets were harvested and sonicated in 25 m*M* Tris–HCl pH 7.5, 2 m*M* DTT, 1 m*M* EDTA, 10% glycerol and 1 m*M* PMSF. After centrifugation at 20000*g* for 20 min, the supernatant was heated at 358 K for 30 min. The heat-treated mixture was centrifuged at 40000*g* for 30 min in a Beckman Type Ti 60 rotor and the supernatant was applied onto a 1 ml Pharmacia High Trap Q column equilibrated with a buffer containing 50 m*M* Tris–HCl pH 7.9 and 50 m*M* NaCl. Most of the contaminating *E. coli* proteins bound to the column, whereas the fibrillarin homolog protein eluted in the flowthrough. The estimated yield was 35 mg of purified protein from 1 l of cells. The purity of the expressed fibrillarin homolog was confirmed by SDS–PAGE and electrospray mass spectrometry. The results of the electrospray mass spectrometry displayed a predominant peak at 25.971 kDa, corresponding to the predicted molecular weight (25.965 kDa).

2.3. Crystallization of native protein

For crystallization trials, the fibrillarin homolog protein was concentrated to 25 mg ml^{-1} using vacuum dialysis at 277 K against 50 m*M* Tris–HCl pH 7.9, 50 m*M* NaCl. Crystallization conditions were tested using the sparse-matrix sampling method (Jancarik & Kim, 1991) with commercially prepared reagents (Hampton Research). The hanging-drop vapor-diffusion method was used by mixing equal volumes (1 ml) of the protein and screening solutions and equilibrating against the corresponding reservoir solutions (500 ml) at room temperature. Monoclinic crystals were prepared using 20% PEG 4000 with 20% 2-propanol and 0.1 *M* sodium citrate pH 5.6. Crystals were grown for 3 d to dimensions of approximately 0.5 × 0.3 × 0.2 mm.

2.4. Expression, purification and crystallization of selenomethionine-substituted protein

The pET21a/Mj697 expression vector was transformed into *E. coli* B834 (DE3), a methionine auxotroph (Novagen). Transformed cells were grown at 310 K in M9 minimal media containing 45 mg ml^{-1} of selenomethionine, 50 mg ml^{-1} of ampicillin and 30 mg ml^{-1} of spectinomycin. Cells were induced at an OD_{600} of 0.7 with 0.5 m*M* IPTG for 6 h. Fibrillarin homolog protein containing selenomethionine (SeMet) was purified as described above. Electrospray mass spectrometry confirmed that all five methionine sites, including the first methionine, were substituted with selenomethionine. The SeMet crystals were grown under the same conditions as before, except that they were seeded with native crystals.

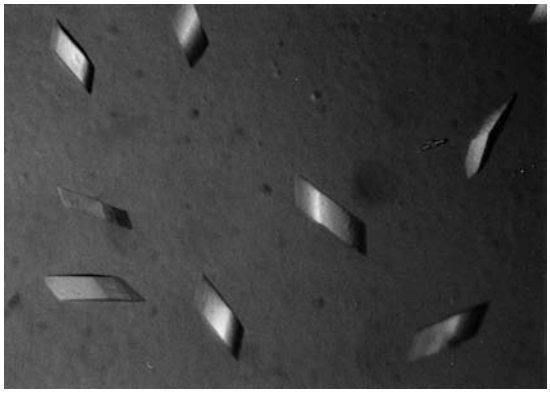

Figure 1
Native crystals of the *M. jannaschii* fibrillarin homolog grown from 20% PEG 4000, 20% 2-propanol, 0.1 *M* sodium citrate pH 5.6 have dimensions 0.5 × 0.3 × 0.2 mm. The space group is monoclinic *C*2.

2.5. X-ray data collection

Fully grown crystals (Fig. 1) were transferred to artificial mother liquor (0.1 *M* sodium citrate pH 5.6, 25% PEG 4000, 25% 2-propanol, 15% glycerol) for 30 min before flash-freezing and data collection. Preliminary X-ray data were collected at approximately 100 K on a Rigaku R-AXIS IIC imaging-plate system, using Cu $K\alpha$ radiation from a Rigaku RU200 HB rotating-anode generator operated at 40 kV and 100 mA. Native data using a synchrotron source was collected at the SSRL beamline 1–5. The crystal-to-image-plate distance was 153 mm and the oscillation range was 2°. Crystals diffracted to 2.3 Å on the R-AXIS system and to 1.8 Å at the Stanford Synchrotron Radiation Laboratory (SSRL). Autoindexing and data processing were performed with *DENZO* (Otwinowski, 1993).

Table 1
Data-collection statistics.

Data in parentheses refer to the last resolution bin.

Resolution (Å)	50.0–1.8 (1.91–1.83)
Total observations	105038
Unique reflections	31292 (3094)
R_{sym}† (%)	6.6 (45.9)
Completeness (%)	99.3 (99)

† $R_{sym} = \sum(|I - \bar{I}|)/\sum\bar{I}$.

3. Results and discussion

Phylogenetically, Archaea is closer to Eukarya than to Bacteria, and a hyperthermophilic Archaeon serves as a very useful source for proteins that are unstable and/or difficult to purify from eukaryotic organisms. We have shown that heterologous expression of a hyperthermophile homolog of human fibrillarin in *E. coli* cells containing plasmid-encoded tRNAs for two rare codons resulted in high levels of soluble recombinant protein. Purification of the protein was straightforward since the hyperthermophilic protein can withstand heat treatment that precipitates most *E. coli* proteins. Thereafter, a single elution on an anion-exchange column was required to yield crystallizable protein.

Preliminary diffraction data show that native crystals belong to the monoclinic *C*2 space group with cell parameters of $a = 121.4$, $b = 43.2$, $c = 55.3$ Å, $\beta = 96.9°$. V_m was calculated to be 2.78 Å^3 Da^{-1} with $Z = 4$, corresponding to one molecule per asymmetric unit. The solvent content was determined to be 55% using the Matthews equation (Matthews, 1968). The statistics from a complete data set collected at SSRL to 1.8 Å resolution are shown in Table 1.

We thank Dr David King for mass spectrometric analysis of the proteins, Stan Goldman for providing *M. jannaschii* genomic DNA, Dr Douglas Clark for *M. jannaschii* cells and Dr S. Sandler for the plasmid pSJS1240. We are also grateful to Jaru Jancarik and Kyeong Kyu Kim for help with the crystallization of the protein and

the preliminary X-ray data collection, respectively. The work described here was supported by the Director, Office of Energy Research, Office of Biological and Environmental Research at the US Department of Energy under contract No. DE-AC03-76SF00098 (RK and SHK).

References

Amiri, K. A. (1994). *J. Bacteriol.* **176**, 2124–2127.
Arnett, F. C., Reveille, J. D., Goldstein, R., Pollard, K. M., Leaird, K., Smith, E. A., Leroy, E. C. & Fritzler, M. J. (1996). *Arthritis Rheum.* **39**, 1151–1160.
Baserga, S. J., Yang, X. D. W. & Steitz, J. A. (1991). *EMBO J.* **10**, 2645–2651.
Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerad, L. M., Clayton, R. A., Goyne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R. & Venter, J. C. (1996). *Science*, **273**, 1058–1073.
Cavaille, J., Nicoloso, M. & Bachellerie, J. P. (1996). *Nature (London)*, **383**, 732–735.
Fournier, M. J. & Maxwell, E. S. (1993). *Trends Biochem. Sci.* **18**, 131.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. & Kim, S.-H. (1999). *Biotech. Lett.* In the press.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Pollard, K., Lee, D., Casiano, C., Bluthner, M., Johnston, M. & Tan, E. (1997). *J. Immunol.* **158**, 3521–3528.
Schimmang, T., Tollervey, D., Kern, H., Frank, R. & Hurt, E. C. (1989). *EMBO J.* **8**, 4015–4024.
Tollervey, D. (1996). *Science*, **273**, 1056–1057.
Tollervey, D., Lehtonen, H., Carmofonseca, M. & Hurt, E. C. (1991). *EMBO J.* **10**, 573–583.
Tollervey, D., Lehtonen, H., Jansen, R., Kern, H. & Hurt, E. C. (1993). *Cell*, **72**, 443–457.