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

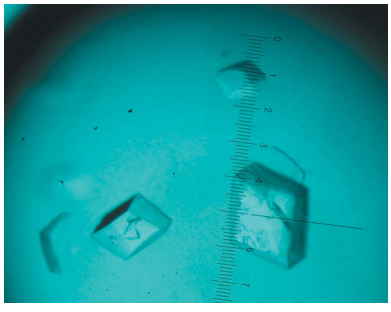
Crystallization and preliminary crystallographic analysis of the second RRM of Pub1 from *Saccharomyces cerevisiae*

Yingji Cui,[a,b] Heng Li,[a,b] Xu Li,[a,b] Yang Li,[a,b] Huihao Zhou,[a,b] Liwen Niu[a,b]* and Maikun Teng[a,b]*

[a]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230027, People's Republic of China, and [b]Key Laboratory of Structural Biology, Chinese Academy of Sciences, 96 Jinzhai Road, Hefei, Anhui 230027, People's Republic of China

Correspondence e-mail: lwniu@ustc.edu.cn, mkteng@ustc.edu.cn

Received 1 November 2008
Accepted 3 December 2008



mRNA stability is elaborately regulated by elements in the mRNA transcripts and their cognate RNA-binding proteins, which play important roles in regulating gene expression at the post-transcriptional level in eukaryotes. Poly(U)-binding protein 1 (Pub1), which is a major nuclear and cytoplasmic polyadenylated RNA-binding protein in *Saccharomyces cerevisiae*, is involved in the regulation of mRNA turnover as a *trans*-acting factor. It binds to transcripts containing the AU-rich element in order to protect them from degradation. Pub1 contains three RNA-recognition motifs (RRMs) which play significant roles in mRNA binding at AU-rich elements and stabilizer elements. In this study, the second RRM of Pub1 was crystallized by the hanging-drop vapour-diffusion method using polyethylene glycol 4000 as a precipitant at 283 K. An X-ray diffraction data set was collected using a single flash-cooled crystal that belonged to space group *H*3.

1. Introduction

In eukaryotes, mRNAs are first synthesized in the nucleus as pre-mRNAs using DNA as a template. After post-transcriptional processing, including 5′-end capping, splicing, 3′-end cleavage and polyadenylation, mature mRNAs are exported to the cytoplasm for protein synthesis and then degraded (Hollams *et al.*, 2002). Gene expression is thus a highly regulated process that involves regulation at both the transcriptional and post-transcriptional levels.

The turnover of mRNA is an elaborately controlled process that plays an important role in post-transcriptional regulation and is mediated by the interplay between the different *cis*-element sequences localized in the target mRNAs and the various *trans*-acting factors that interact with them (Duttagupta *et al.*, 2005). The AU-rich element (ARE) is a well studied *cis*-element that is found in the 3′-untranslated region (3′UTR) of many mammalian mRNAs and yeast transcripts (Vasudevan & Peltz, 2001). It has been demonstrated that ARE acts as a destabilizing element to mediate rapid mRNA turnover. There are a large number of RNA-binding proteins (RBPs) in metazoan eukaryotic cells, such as proteins that directly bind to nuclear polyadenylated RNAs, heterogeneous nuclear RNA-binding proteins (hnRNPs), those that associate with cytoplasmic mRNAs and mRNA-binding proteins (mRNPs). They can act as either enhancers or inhibitors by selectively binding to target motifs in the 3′UTR (Anderson *et al.*, 1993; Duttagupta *et al.*, 2005). The interplay between these RNA-binding proteins and distinct *cis*-elements in the target transcripts regulates mRNA stability (Apponi *et al.*, 2007).

Poly(U)-binding protein 1 (Pub1) is an important polyadenylated RNA-binding protein in *Saccharomyces cerevisiae* and is involved in the regulation of mRNA turnover as a *trans*-acting factor (Anderson *et al.*, 1993; Ruiz-Echevarria & Peltz, 2000; Vasudevan & Peltz, 2001). However, in contrast to other yeast RNA-binding proteins, Pub1 is detected both in the nucleus and the cytoplasm and there is little difference in its distribution patterns between the nucleus and the cytoplasm. This suggests that Pub1 acts as both an hnRNP and an mRNP (Anderson *et al.*, 1993). Similar to the ELAV-like protein

HuR, the apparent mammalian orthologue of Pub1, Pub1 can bind to the ARE and ARE-like-sequence-containing transcripts to protect them from degradation *via* a deadenylation-dependent pathway in yeast (Vasudevan & Peltz, 2001; Apponi *et al.*, 2007). It has also been shown that Pub1 binds to stabilizer elements (STEs) located in the 5′UTR of YAP1 and GCN4 to protect them from turnover through the NMD pathway (Ruiz-Echevarria & Peltz, 2000). In conclusion, Pub1 can bind to at least two classes of stability element and may be a key *trans*-acting factor mediating the stability of multiple mRNAs in *S. cerevisiae* (Duttagupta *et al.*, 2005). Pub1 contains three RNA-recognition motifs (RRMs) which play crucial roles in mRNA binding at AREs and STEs. A human RNA-binding protein TIA-1, which is very similar to Pub1 in both sequence and domain structure, was found *via* a seqence-similarity search. The functions of the three RRMs of TIA-1 are different and recent studies have demonstrated that the second RRM of TIA-1, which shares 47% sequence identity to the second RRM of Pub1, is the major domain involved in TIA-1 binding to U-rich sequences (Kumar *et al.*, 2008; Forch *et al.*, 2002). The overall binding specificity is determined by the individual RRM. Therefore, we believe that the second RRM of Pub1 may also play a central role in poly(U) binding and determine the binding specificity of Pub1. In this study, we report the crystallization and preliminary crystallographic analysis of the second RRM of Pub1; the determination of its structure will provide us with an insight into its properties.

2. Materials and methods

2.1. Protein expression and purification

The DNA-coding sequence for the second RRM of yeast Pub1 (NCBI entry code NP_014382), corresponding to amino acids 161–240, was amplified from the *S. cerevisiae* genome by PCR using the following oligonucleotide primers containing artificial *Nde*I and *Xho*I sites (in bold): 5′-GTCTG**CATATG**ACATTTAACTTGTTTGTCGG-3′ (forward) and 5′-GAC**CTCGAG**CTTAGCAGCCCAGTTGATTC-3′ (reverse) (Sangon). The PCR product was digested with *Nde*I and *Xho*I and ligated into the *Nde*I and *Xho*I sites of the bacterial expression vector pET22b as a fusion protein with an –LEHHHHHH tag at the C-terminus. The cloning junctions were confirmed by DNA sequencing (Sangon). The plasmid containing the corresponding gene sequence was transformed into Rosetta (DE3) cells (Novagen). The cells containing the target plasmid were grown in Luria–Bertani (LB) medium supplemented with 50 mg ml^{-1} ampicillin at 310 K until the culture reached an OD_{600} of 0.6 and then induced with 1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) for a further 18 h at 289 K. The cells were harvested by centrifugation at 8000*g* for 5 min and then suspended in lysis buffer (50 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl). Following cell lysis by sonication on ice, the lysate was clarified by centrifugation at 12 000*g* for 30 min. The supernatant solution was applied onto a nickel-chelating column (GE Healthcare) pre-equilibrated with equilibration buffer (50 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl). The column was first washed with five column volumes of equilibration buffer (50 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl) and was then washed with washing buffer (50 m*M* Tris–HCl pH 8.0, 50 m*M* imidazole, 200 m*M* NaCl) to remove remaining impurities. The target protein was finally washed with elution buffer (50 m*M* Tris–HCl pH 8.0, 200 m*M* imidazole, 200 m*M* NaCl). The eluate was buffer-exchanged with equilibration buffer and loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with the same buffer. The flowthrough fraction, which contained the target protein, was buffer-exchanged with crystallization buffer (10 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl) and concentrated by centrifugal ultrafiltration (Millipore, 5 kDa cutoff) to 22 mg ml^{-1} as estimated using the BCA Protein Assay Kit (Pierce) for crystallization assays. The purity of the protein was greater than 95% by analysis on SDS–PAGE (Fig. 1) and the protein was judged to be suitable for crystallization. Finally, the purified protein was stored at 193 K for crystallization.

2.2. Crystallization

Crystallization experiments were performed at 283 K using the hanging-drop vapour-diffusion method; initial trials were carried out using Crystal Screens I and II, Index Screen (Hampton Research) and JBScreen Classic (Jena Bioscience). Hanging drops, each consisting of 1 µl reservoir solution and 1 µl protein solution, were equilibrated against 200 µl reservoir solution. After several rounds of optimization, a well diffracting single crystal grew after 48 h at 283 K using a protein concentration of 22 mg ml^{-1} and a reservoir solution comprising 25%(*w*/*v*) polyethylene glycol 4000 (PEG 4000), 100 m*M* sodium acetate pH 4.6. The well diffracting crystal grew to typical dimensions of 0.2 × 0.2 × 0.1 mm (Fig. 2).

2.3. X-ray diffraction data collection and preliminary analysis

Initial X-ray diffraction experiments were performed using in-house Cu $K\alpha$ X-rays generated by an RA-Micro007 rotating-anode

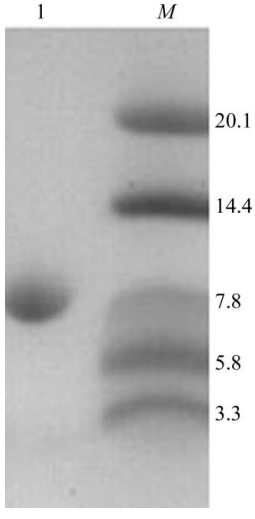


Figure 1
SDS–PAGE of Pub1 RRM2. Lane 1, purified fraction from the Superdex 75 column (greater than 95% purity); lane *M*, molecular-weight markers from BIOS (kDa).

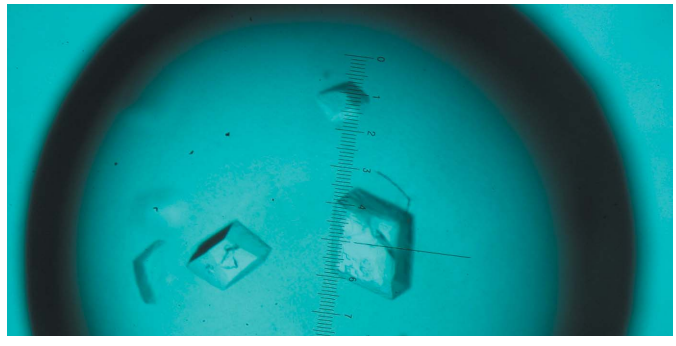


Figure 2
Photomicrograph of the crystal of the second RRM of Pub1. The dimensions of this single crystal were approximately 0.2 × 0.2 × 0.1 mm.

X-ray source (Rigaku) and diffraction images (Fig. 3) were collected using a MAR 345dtb imaging-plate detector (MAR Research). Prior to data collection, the crystal was transferred into cryoprotectant solution consisting of 35%(*w*/*v*) PEG 4000, 100 m*M* sodium acetate pH 4.6 and equilibrated for 3 min. The crystal was then flash-cooled to 100 K in a stream of cold nitrogen gas produced by an Oxford Cryosystems cooling device (Oxford Cryosystems Ltd). A complete diffraction data set consisting of 358 images with an oscillation angle of 1° per image was collected to 1.69 Å resolution at 100 K. Diffraction data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Data-collection and processing statistics are listed in Table 1.

Table 1
X-ray data-collection and processing statistics.

Values in parentheses are for the last shell.

Space group	*H*3
Unit-cell parameters (Å, °)	$a = b = 73.11$, $c = 41.12$, $\alpha = \beta = 90$, $\gamma = 120$
Resolution (Å)	36.56–1.69 (1.78–1.69)
Unique reflections	9164 (1358)
Redundancy	3.4 (3.2)
Completeness (%)	99.8 (99.9)
Average $I/\sigma(I)$	19 (4.2)
$R_{\rm merge}$† (%)	3.7 (29)
No. of molecules in unit cell (*Z*)	3
$V_{\rm M}$ (Å^3 Da^{-1})	2.35
Solvent content (%)	47.69

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations *i* of reflection *hkl*.

3. Results and discussion

The RNA-recognition motif, also known as the RNA-binding domain (RBD) or ribonucleoprotein domain (RNP), is an abundant protein motif that is found in many RNA-binding proteins in eukaryotes. It plays an important role in mRNA stability through binding to a specific sequence. By association with different types of protein domains, the RRM can also modulate the RNA-binding affinity and specificity of proteins. Moreover, biochemical and structural studies have shown that the RRM is also involved in protein–protein interaction (Maris *et al.*, 2005). To date, more than 100 RRM structures have been determined by either NMR or X-ray crystallography (Maris *et al.*, 2005). Of the structures reported, the closest homologue of the second RRM of Pub1 is the second RRM of the human protein TIA-1; the sequence identity between them is 47%. Although these RRM domains are usually composed of four β-strands forming an antiparallel β-sheet, with two α-helices packing against the β-sheet with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology, their binding affinity and specificity are very diverse. Further structures of RRMs are needed to fully understand the determinants of their specificity (Maris *et al.*, 2005). Moreover, the varying lengths of RRMs may result in differences in their structures and the mechanism of poly(U) binding has not yet been clearly described. The structure determination of the second RRM of Pub1 will provide us with further insight into its binding pattern. Furthermore, the functions of Pub1 and its orthologues have been little studied and no three-dimensional structures have been reported for them. Further biochemical and structural studies are needed to obtain a full picture of the roles of these proteins in the cell.

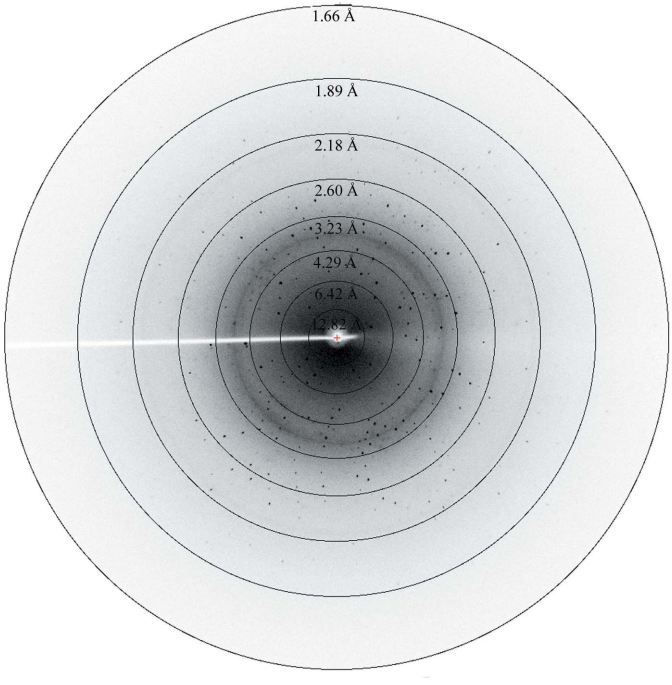


Figure 3
Image of the diffraction pattern of the second RRM of Pub1.

Three RRMs of Pub1 were cloned and expressed, but only the second RRM was overexpressed in a soluble form in *E. coli*. The molecular weight of the second RRM was 10 kDa as estimated by size-exclusion chromatograpy, which is in good agreement with the calculated value of 9.90 kDa. This suggests that the protein exists as a monomer in solution. We have obtained a crystal of the second RRM of Pub1; the crystal belonged to space group *H*3. The calculated Matthews coefficient ($V_{\rm M}$) of 2.35 Å^3 Da^{-1}, with a solvent content of 47.69%, indicates the presence of one molecule in each asymmetric unit (Matthews, 1968). The molecular-replacement method was performed using the structure of the second RRM of human protein TIA-1 (Kumar *et al.*, 2008) as a search model. Further refinement of the structure is under way.

This work was supported by grants from the Chinese National Natural Science Foundation (Nos. 30121001, 30025012 and 30571066), the Chinese Ministry of Science and Technology (Nos. 2006CB806500, 2006CB910200 and 2006AA02A318), the Chinese Academy of Sciences (No. KSCX2-YW-R-60) and the Chinese Ministry of Education (No. 20070358025) to LN and MT.

References

Anderson, J. T., Paddy, M. R. & Swanson, M. S. (1993). *Mol. Cell. Biol.* **13**, 6102–6113.
Apponi, L. H., Kelly, S. M., Harreman, M. T., Lehner, A. N., Corbett, A. H. & Valentini, S. R. (2007). *Mol. Cell. Biol.* **27**, 6569–6579.
Duttagupta, R., Tian, B., Wilusz, C. J., Khounh, D. T., Soteropoulos, P., Ouyang, M., Dougherty, J. P. & Peltz, S. W. (2005). *Mol. Cell. Biol.* **25**, 5499–5513.
Forch, P., Puig, O., Martinez, C., Seraphin, B. & Valcarcel, J. (2002). *EMBO J.* **21**, 6882–6892.
Hollams, E. M., Giles, K. M., Thomson, A. M. & Leedman, P. J. (2002). *Neurochem. Res.* **27**, 957–980.
Kumar, A. O., Swenson, M. C., Benning, M. M. & Kielkopf, C. L. (2008). *Biochem. Biophys. Res. Commun.* **367**, 813–819.
Maris, C., Dominguez, C. & Allain, F. H. (2005). *FEBS J.* **272**, 2118–2131.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ruiz-Echevarria, M. J. & Peltz, S. W. (2000). *Cell*, **101**, 741–751.
Vasudevan, S. & Peltz, S. W. (2001). *Mol. Cell*, **7**, 1191–1200.