Yan An, Yongfeng Lou† and
Yingwu Xu*

The Nurturing Station for the State Key
Laboratory of Subtropical Silviculture, Zhejiang
Agriculture and Forestry University, Lin'an,
Zhejiang 331300, People's Republic of China

† Present address: Key laboratory on the science
and technology of bamboo and rattan, Inter-
national centre for bamboo and rattan, Beijing
100102, People's republic of China.

Correspondence e-mail: yxu@zafu.edu.cn

Received 14 August 2013

Accepted 10 October 2013

Overexpression, crystallization and preliminary X-ray crystallographic analysis of release factor eRF1-1 from *Arabidopsis thaliana*

Peptide release factor 1 (RF1) regulates the termination of translation in protein synthesis by recognizing the stop codons. The eukaryotic RF1s (eRF1s) from *Arabidopsis thaliana* and human have different stop-codon preferences even though they share high sequence similarity. Based on known RF1 structures, it has been suggested that the specificity depends on both the local structure and the domain arrangement, but the lack of a structure of *Arabidopsis* eRF1 hinders a detailed comparison. To reveal the mechanism of stop-codon recognition and compare it with that of human eRF1, one of the three *Arabidopsis* eRF1s, AteRF1-1, was studied and a preliminary X-ray crystallographic analysis is reported here. The protein was overexpressed in *Escherichia coli* and crystallized at room temperature using the vapour-diffusion method. Crystals were grown from 1.6 M lithium sulfate, 0.1 M Tris-HCl pH 8.0, 2% (v/v) PEG 400 and diffracted to 3.77 Å resolution. The data were processed in point group 622, with unit-cell parameters $a = b = 136.6$, $c = 325.7$ Å.

1. Introduction

Termination of translation is a critical step in protein synthesis that takes place on the ribosome in response to a stop codon. This process requires a single class I release factor in eukaryotes (Konecki *et al.*, 1977; Frolova *et al.*, 1994) and two in bacteria: RF1 recognizes UAG and UAA, while RF2 recognizes UAA and UGA (Scolnick *et al.*, 1968). Class I release factors are termination-codon specific and promote the hydrolysis of peptidyl-tRNA, while class II release factors (RF3 and eRF3 in prokaryotes and eukaryotes, respectively) are not stop-codon specific (Zhouravleva *et al.*, 1996; Freistroffer *et al.*, 1997; Grentzmann *et al.*, 1998). Most organisms show a termination-codon preference. Yeast and mammals prefer UAA (Brown *et al.*, 1990; Bertram *et al.*, 2001) and UAG is favoured in both monocotyledon and dicotyledon plants (Angenon *et al.*, 1990), whereas eubacteria tend to use UAA but not UGG (Sharp *et al.*, 1988). This specificity has been mapped to the tripeptide motifs SPF and PAT in the release factors (Young *et al.*, 2010). A brief summary of the most favoured stop codons is shown in Fig. 1(a). It is believed that in eukaryotes, the N-terminal domain of eRF1, and in particular the tetrapeptide NIKS motif of human eRF1, is implicated in stop-codon recognition (Song *et al.*, 2000), and recent work suggests that atomic details of the interaction with the ribosome and mRNA and domain flexibility of eRF1 govern the recognition process (Petry *et al.*, 2005; Shin *et al.*, 2004).

Proteins of the eRF1 family share a high degree of similarity in amino-acid sequence. Whereas most eukaryotes possess a single class I eRF1, *Arabidopsis thaliana* has three (AteRF1-1, AteRF1-2 and AteRF1-3; Brown *et al.*, 1995) that share 85–94% identity (Chapman & Brown, 2004). All three have been shown to possess release-factor activity (Chapman & Brown, 2004), but AteRF1-1 is involved in cell elongation and radicle cell division (Petsch *et al.*, 2005), and AteRF1-2 affects glucose and phytohormone responses in modulating plant growth and development (Zhou *et al.*, 2010).

AteRF1-1 has 72% identity to its human counterpart, implying high similarity in overall structure and fold. However, careful examination of the amino-acid sequences and of the human eRF1 crystal structure suggests that a change in domain orientation is possible. eRF1 is composed of three domains, N, M and C (Fig. 1b), in

© 2013 International Union of Crystallography
All rights reserved

Figure 1

(a) List of the most favoured stop codons. (b) Illustration of the AteRF1-1 domain organization based on sequence alignment with human eRF1 and its crystal structure. The expression construct covered amino acids 4–423.

a triangular arrangement. The highly conserved GGQ motif is located at the tip of the middle domain M and is linked to domains N and C by two linkers that define its orientation. Linker 1 (residues 141–144) adopts an extended conformation stabilized by side-chain hydrogen bonds between Arg10 and Asp142 in the human protein (PDB entry 1dt9; Song *et al.*, 2000), which are replaced by Lys9 and Ser141 in AteRF1-1. Arg10 is in the middle of a long N-terminal helix that fixes the orientation of domain C. Therefore, linker 1 may have to adjust its main-chain conformation to maintain the side-chain hydrogen bond in AteRF1-1 or, if the bond is not maintained, domain M (and therefore the GGQ motif) may have more freedom to adapt its orientation in the ribosome environment. In fact, a large-scale displacement of domain M has been observed for human eRF1 in complex with eRF3 (Cheng *et al.*, 2009), moving the GGQ motif by 44 Å as linker 1 rotates by 37°. A rearrangement of the middle domain is also seen between the free and the aEF1-bound form of an archaeal RF1 (Kobayashi *et al.*, 2012). Although the conformational changes may be triggered by binding to partners, they require flexibility in the linkers. Linker 2 (residues 273–278 in the human protein) adopts a deformed helical conformation which may be intrinsically flexible (Kobayashi *et al.*, 2012).

To explore the conformation change and its relationship to stop-codon recognition by AteRF1-1, we expressed this protein in *Escherichia coli*, purified it and initiated a structural investigation. Here, we report its crystallization conditions and the preliminary analysis of the X-ray diffraction data.

Figure 2

Crystals of eRF1-1 from *A. thaliana* grown using 1.6 M lithium sulfate, 100 mM Tris–HCl pH 8.0, 2% (v/v) PEG 400 to a size over 60 μm at room temperature.

2. Materials and methods

2.1. Cloning and expression

The AteRF1-1 gene was amplified by the polymerase chain reaction using *A. thaliana* cDNA as template. The forward and reverse primers were 5'-AAAAACGATGACGACAAGAACA and 5'-CTAATC-AAAGGCAGTCATGTCAAGC, respectively. They were designed using the GenBank sequence with accession No. U40217.1. The amplified DNA was cloned into the expression vector pEASY-E1 (TransGen Biotech, People's Republic of China) using the TA cloning strategy, resulting in a final construct consisting of residues 4–423 (Fig. 1b). The construct was sequenced and contained the expected gene with a hexahistidine tag at the N-terminus. The recombinant protein was overexpressed in *E. coli* BL21 (DE3) cells, which were grown at 310 K to an OD₆₀₀ of 0.6 in Luria broth culture medium containing 50 mg ml⁻¹ ampicillin. Protein expression was then induced using 0.4 mM isopropyl β-D-1-thiogalactopyranoside for 15 h at 293 K. The cells were harvested the next morning by centrifugation at 8000g.

2.2. Purification

The cell pellet from a 4 l culture was resuspended in 80 ml ice-cold lysis buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl, 10% glycerol) and sonicated for 10 min. The crude cell extract was centrifuged at 36 500g for 40 min at 277 K. We utilized the hexahistidine tag for affinity chromatography on Ni–NTA resin (GE Healthcare Life Sciences). 2 ml resin previously charged with Ni²⁺ was equilibrated with the lysis buffer and added to the supernatant. After incubation with slow shaking for 30 min on ice, the resin was washed with lysis buffer supplemented with a stepwise gradient of 10, 20 and 50 mM imidazole. The bound protein was eluted with elution buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 10% glycerol, 200 mM imidazole). The eluted sample was further purified by gel-filtration/size-exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 20 mM Tris–HCl pH 8.0, 200 mM NaCl. The protein was >95% pure when checked by SDS–PAGE. Its concentration was estimated by measuring the absorbance at 280 nm. The purified eRF1-1 was concentrated to a final concentration of 30 mg ml⁻¹ using an Amicon Ultra 30 K centrifugal filter device.

2.3. Crystallization and X-ray data collection

AteRF1-1 was crystallized at 295 K using the sitting-drop vapour-diffusion method. The protein was first diluted to 20 mg ml⁻¹ in the gel-filtration equilibration buffer. Crystallization kits from Hampton Research and Emerald BioSystems were employed for the initial screening experiments. Typically, drops consisting of 0.3 μl protein solution and 0.3 μl reservoir solution were equilibrated against 350 μl reservoir solution.

Crystals appeared after 2 d using reservoir solution consisting of 2 M lithium sulfate, 0.1 M Tris–HCl pH 8.5, 2% (v/v) PEG 400 and were optimized to a size of over 60 μm with 1.6 M lithium sulfate, 0.1 M Tris–HCl pH 8.0, 2% (v/v) PEG 400 (Fig. 2). The crystals were transferred to the crystallization buffer containing 10% PEG 400 and 15% glycerol before being flash-cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K on beamline BL17U of the Shanghai Synchrotron Radiation Facility (SSRF), People's Republic of China. The crystals diffracted isotropically to about 3.7 Å resolution, yielding well separated sharp spots but weak intensity (Fig. 3). A total of 180 frames with 1° rotation per frame were collected from a single crystal. The raw data were processed using the HKL-2000

Figure 3

X-ray diffraction pattern. Data were collected using a CCD detector on beamline BL17U at the SSRF.

Table 1

Statistics of X-ray data collection.

Values in parentheses are for the outermost resolution shell.

Wavelength (Å)	0.979
Unit-cell parameters (Å)	$a = b = 136.6$, $c = 325.7$
Point group	622
Resolution (Å)	50–3.77 (3.94–3.77)
Unique reflections	19029 (2140)
Completeness (%)	99.8 (99.8)
Multiplicity	12.0 (12.0)
Average $I/\sigma(I)$	12.5 (3.9)
$R_{\text{merge}}^{\dagger}$	0.17 (0.48)
Mosaicity (°)	0.27

$^{\dagger} R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the measured intensity of the reflection with indices hkl .

program suite (Otwinowski & Minor, 1997). The statistics of data collection are summarized in Table 1.

3. Results

The open reading frame of eRF1-1 from *A. thaliana* contains 1311 base pairs coding for 436 amino-acid residues. The expression clone contained residues 4–423 (Fig. 1b) and a long N-terminal tag (MRGSHHHHHGMASELAL). Expression in *E. coli* yielded 8 mg purified protein per litre of culture, with a single band of ~47 kDa on SDS-PAGE matching the calculated molecular weight of 49 kDa. Size-exclusion chromatography suggests that the protein is a monomer in solution.

The recombinant protein readily formed crystals. A set of diffraction data was collected and processed to 3.77 Å resolution from a flash-cooled crystal. The data were integrated in point group 622 with a relatively large R_{merge} of 17%. Integration of the data in a lower symmetry group gave the same R_{merge} value, and analysis of the data using *phenx.xtriage* (Zwart, 2005) confirmed the choice of 622 on a Bayesian criterion (Schwarz, 1978). Based on the unit-cell parameters and the protein molecular weight, a solvent content of 58.6 or

44.8% is calculated if three or four copies of the monomer are present in the asymmetric unit, corresponding to a Matthews coefficient (Matthews, 1968) of 2.97 or 2.23 Å³ Da⁻¹, respectively. The largest Patterson peak was found in the native Patterson map at as low as 3.6% of the origin peak, suggesting that there was no translational pseudosymmetry. A self-rotation function was calculated by *MOLREP* (Vagin & Teplyakov, 2010) and *GLRF* (Tong & Rossman, 1990) to search for information in noncrystallographic symmetry. No significant peaks were revealed in the κ sections except for the peaks corresponding to the crystallographic symmetry. Furthermore, molecular replacement gave no clear solution when the human eRF1 structure (PDB entry 1dt9; Song *et al.*, 2000) was used as the input template to the *Phaser* module of the *PHENIX* program suite (Adams *et al.*, 2010; McCoy *et al.*, 2007). Splitting the model into two ensembles did not yield a solution. The domain arrangement in AteRF1 may be different from that in human eRF1, and the existence of multiple copies of the polypeptide chain in the asymmetric unit further increases the difficulty of the molecular replacement. In addition, the space group is uncertain; the sixfold may be a screw axis, although *phenx.xtriage* analysis did not show a space-group preference. (00 l) reflections were not observed in the X-ray data collection and no other apparent systematic absences were found. Table 1 summarizes the statistics of the data collection. The current work provides useful information for future optimization of the crystals and ultimately for the structural illustration of the specificity in the stop-codon recognition of AteRF1-1.

We thank the staff of beamline BL17U at the Shanghai Synchrotron Radiation Facility for assistance during X-ray data collection. This work was supported by the National Natural Science Foundation of China (grant No. 31270715) and the Foundation of the Zhejiang Agriculture and Forestry University (grant No. 2010FR072).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Angenon, G., Van Montagu, M. & Depicker, A. (1990). *FEBS Lett.* **271**, 144–146.

- Bertram, G., Innes, S., Minella, O., Richardson, J. & Stansfield, I. (2001). *Microbiology*, **147**, 255–269.
- Brown, C. M., Quigley, F. R. & Miller, W. A. (1995). *Plant Physiol.* **110**, 336.
- Brown, C. M., Stockwell, P. A., Trotman, C. N. A. & Tate, W. P. (1990). *Nucleic Acids Res.* **18**, 6339–6345.
- Chapman, B. & Brown, C. (2004). *Gene*, **341**, 219–225.
- Cheng, Z., Saito, K., Pisarev, A. V., Wada, M., Pisareva, V. P., Pestova, T. V., Gajda, M., Round, A., Kong, C., Lim, M., Nakamura, Y., Svergun, D. I., Ito, K. & Song, H. (2009). *Genes Dev.* **23**, 1106–1118.
- Freistrotter, D. V., Pavlov, M. Y., MacDougall, J., Buckingham, R. H. & Ehrenberg, M. (1997). *EMBO J.* **16**, 4126–4133.
- Frolova, L., Le Goff, X., Rasmussen, H. H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A.-L., Celis, J. E., Philippe, M., Justesen, J. & Kisselev, L. (1994). *Nature (London)*, **372**, 701–703.
- Grentzmann, G., Kelly, P. J., Laalami, S., Shuda, M., Firpo, M. A., Cenatiempo, Y. & Kaji, A. (1998). *RNA*, **4**, 973–983.
- Kobayashi, K., Saito, K., Ishitani, R., Ito, K. & Nureki, O. (2012). *Nucleic Acids Res.* **40**, 9319–9328.
- Konecki, D. S., Aune, K. C., Tate, W. & Caskey, C. T. (1977). *J. Biol. Chem.* **252**, 4514–4520.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- 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.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Petry, S., Brodersen, D. E., Murphy, F. V. IV, Dunham, C. M., Selmer, M., Tarry, M. J., Kelley, A. C. & Ramakrishnan, V. (2005). *Cell*, **123**, 1255–1266.
- Petsch, K. A., Mylne, J. & Botella, J. R. (2005). *Plant Physiol.* **139**, 115–126.
- Schwarz, G. (1978). *Ann. Stat.* **6**, 461–464.
- Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. (1968). *Proc. Natl Acad. Sci. USA*, **61**, 768–774.
- Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H. & Wright, F. (1988). *Nucleic Acids Res.* **16**, 8207–8211.
- Shin, D. H., Brandsen, J., Jancarik, J., Yokota, H., Kim, R. & Kim, S.-H. (2004). *J. Mol. Biol.* **341**, 227–239.
- Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A. & Barford, D. (2000). *Cell*, **100**, 311–321.
- Tong, L. & Rossmann, M. G. (1990). *Acta Cryst. A* **46**, 783–792.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.
- Young, D. J., Edgar, C. D., Poole, E. S. & Tate, W. P. (2010). *RNA*, **16**, 1623–1633.
- Zhou, X., Cooke, P. & Li, L. (2010). *J. Exp. Bot.* **61**, 357–367.
- Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L. & Philippe, M. (1996). *EMBO J.* **14**, 4065–4072.
- Zwart, P. H. (2005). *Acta Cryst. D* **61**, 1437–1448.