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

Crystallization and preliminary X-ray diffraction of the DEAD-box protein Mss116p complexed with an RNA oligonucleotide and AMP-PNP

Mark Del Campo and Alan M. Lambowitz*

Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, TX 78712, USA

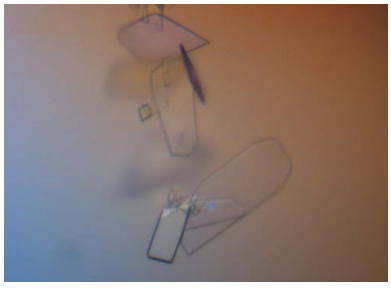
Correspondence e-mail: lambowitz@mail.utexas.edu

Received 12 May 2009
Accepted 10 July 2009

The *Saccharomyces cerevisiae* DEAD-box protein Mss116p is a general RNA chaperone which functions in mitochondrial group I and group II intron splicing, translation and RNA-end processing. For crystallization trials, full-length Mss116p and a C-terminally truncated protein (Mss116p/Δ598–664) were overproduced in *Escherichia coli* and purified to homogeneity. Mss116p exhibited low solubility in standard solutions ($\leq$1 mg ml^{-1}), but its solubility could be increased by adding 50 m*M* L-arginine plus 50 m*M* L-glutamate and 50% glycerol to achieve concentrations of $\sim$10 mg ml^{-1}. Initial crystals were obtained by the microbatch method in the presence of a U_{10} RNA oligonucleotide and the ATP analog AMP-PNP and were then improved by using seeding and sitting-drop vapor diffusion. A cryocooled crystal of Mss116p/Δ598–664 in complex with AMP-PNP and U_{10} belonged to space group $P2_12_12$, with unit-cell parameters a = 88.54, b = 126.52, c = 55.52 Å, and diffracted X-rays to beyond 1.9 Å resolution using synchrotron radiation from sector 21 at the Advanced Photon Source.

1. Introduction

DEAD-box proteins are a large and ubiquitous family of RNA helicases that function in RNA and ribonucleoprotein (RNP) structural rearrangements intrinsic to a variety of cellular processes, including translation, RNA degradation, RNA transport, RNA splicing and ribosome biogenesis (Cordin *et al.*, 2006; Jankowsky & Fairman, 2007; Linder, 2006). To effect structural rearrangements, DEAD-box proteins use the energy from cycles of ATP binding, hydrolysis and release of ADP and P_i products to unwind RNA duplexes by a mechanism involving local strand separation (Yang *et al.*, 2007; Yang & Jankowsky, 2006). The ATPase and RNA-strand separation activities are contained within a conserved helicase core consisting of two tandem RecA-like domains with 11 conserved motifs, including the eponymous D-E-A-D motif involved in ATP binding (Jankowsky & Fairman, 2007). In the absence of RNA and ATP, the two core domains, which are joined by a flexible linker, can exist in multiple open conformations, but cooperative binding of ATP and RNA leads to a closed conformation which carries out ATP hydrolysis and RNA unwinding (Jankowsky & Fairman, 2007). A structure believed to be that of the closed conformation with a bound single-stranded RNA (ssRNA) and the nonhydrolyzable ATP analog AMP-PNP was first visualized for the *Drosophila melanogaster* DEAD-box protein Vasa (Sengoku *et al.*, 2006) and subsequently for other DEAD-box proteins (Andersen *et al.*, 2006; Bono *et al.*, 2006; Collins *et al.*, 2009; von Moeller *et al.*, 2009). In this structure, the two core domains are brought together to form clefts that bind AMP-PNP and ssRNA on opposite sides of the protein (Sengoku *et al.*, 2006). The 11 conserved motifs reside at or near the interface between the domains and contribute to the binding of RNA and AMP-PNP or form interdomain contacts. In addition to the helicase core, many DEAD-box proteins contain N-terminal and/or C-terminal extensions, which differ between proteins and in some cases target the protein to specific RNA or RNP substrates *via* RNA–protein or protein–protein interactions (Cordin *et al.*, 2006). Structures of



© 2009 International Union of Crystallography
All rights reserved

DEAD-box proteins showing the helicase core acting together with such extensions are lacking.

The related DEAD-box proteins CYT-19 of *Neurospora crassa* and Mss116p of *Saccharomyces cerevisiae* have been used as model systems for investigating how DEAD-box proteins function on large natural RNA and RNP substrates. These proteins bind RNA and RNP substrates nonspecifically and function as general RNA chaperones (Mohr *et al.*, 2002; Huang *et al.*, 2005; Halls *et al.*, 2007; Del Campo *et al.*, 2009; Tijerina *et al.*, 2006). As such, they are needed for the efficient splicing of more than 13 different mitochondrial (mt) group I and group II introns, translational activation and RNA-end processing (Mohr *et al.*, 2002; Huang *et al.*, 2005). Besides having similar functions in their native cellular milieus, the two proteins are to a large degree functionally interchangeable, with CYT-19 suppressing all the defects in a *mss116*Δ strain and both recombinant proteins stimulating the splicing of diverse group I and group II introns *in vitro* (Huang *et al.*, 2005; Halls *et al.*, 2007; Mohr *et al.*, 2002, 2006; Solem *et al.*, 2006; Del Campo *et al.*, 2009).

The helicase cores of Mss116p and CYT-19 have ∼35% sequence identity to each other and ∼30% sequence identity to the helicase cores of other DEAD-box proteins of known structure (*e.g.* Vasa). In CYT-19 and Mss116p the helicase core is preceded by a short N-terminal extension (NTE) and is followed by a distinctive ∼150-residue C-terminal extension (CTE), which has predicted α-helical regions, and a basic tail, which is predicted to be unstructured (Mohr *et al.*, 2008). Truncations or point mutations in the CTE of CYT-19 or Mss116p inactivate RNA-dependent ATPase activity and decrease RNA-binding affinity, suggesting that this region may be required to stabilize the structure of the helicase core and could contribute to RNA binding (Mohr *et al.*, 2008). The basic tail is not essential for ATPase activity and is thought to contribute to nonspecific RNA binding, helping to tether the helicase core to large RNA substrates for multiple rounds of RNA unwinding (Grohman *et al.*, 2007; Mohr *et al.*, 2008).

Here, we report the use of high glycerol (50%) and L-arginine plus L-glutamate (50 m*M* each; Arg + Glu) to stabilize full-length Mss116p in solution at a concentration of 10 mg ml^{-1}. We crystallized full-length Mss116p in a ternary complex with a U_{10} RNA oligonucleotide (U_{10}) and AMP-PNP, but were unable to optimize these crystals. We obtained improved crystals of the ternary complex by using the Mss116p/Δ598–664 C-terminal truncation, which contains the NTE, the helicase core and the complete α-helical CTE, lacking only the putatively unstructured basic tail.

2. Experimental procedures

2.1. Cloning, expression and purification of Mss116p

Full-length Mss116p without its mt targeting sequence (amino-acid residues 1–36) was produced as a fusion to maltose-binding protein (MBP) from the plasmid pMAL-Mss116p. Plasmid construction and protein overexpression and purification have been described elsewhere (Del Campo *et al.*, 2007; Halls *et al.*, 2007; Mohr *et al.*, 2008). Briefly, MBP-Mss116p was overexpressed in *Escherichia coli* strain Rosetta 2 (EMD Biosciences) using ZYP-5052 autoinducing medium (Studier, 2005) at 295 K. Mss116p was purified using a protocol consisting of the following five steps performed at 277 K: (i) polyethyleneimine precipitation of nucleic acids; (ii) amylose column chromatography to purify the MBP-Mss116p fusion protein; (iii) tobacco etch virus (TEV) protease digestion to cleave MBP from Mss116p; (iv) heparin–Sepharose chromatography to separate Mss116p from MBP and (v) gel filtration through a Superdex-200 column to remove the remaining impurities. Only steps (iv) and (v) differ from previous descriptions (Halls *et al.*, 2007; Del Campo *et al.*, 2007), so they will be described here. After TEV protease cleavage, which leaves an N-terminal Gly-Ser preceding Mss116p residue 37, the sample was diluted to contain the same amount of KCl (300 m*M*) as the heparin column loading buffer (20 m*M* Tris–HCl pH 7.5, 300 m*M* KCl, 1 m*M* EDTA, 1 m*M* DTT and 10% glycerol) and then applied onto a Hi-Trap heparin HP column (GE Healthcare Biosciences). MBP and TEV protease did not bind to the heparin column under these conditions and Mss116p was eluted from the column with a linear gradient to 1 *M* KCl. The Mss116p peak fractions were then applied onto a Superdex 200 16/60 gel-filtration column (GE Healthcare Biosciences) in 20 m*M* Tris–HCl pH 7.5, 500 m*M* NaCl, 1 m*M* EDTA, 1 m*M* DTT, 50 m*M* Arg + Glu and 10% glycerol. The Mss116p peak fractions from this column were pooled and dialyzed into storage buffer (10 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl, 1 m*M* DTT, 50 m*M* Arg + Glu and 50% glycerol). Finally, Mss116p was concentrated to ∼10 mg ml^{-1} by centrifugation in an Amicon Ultra-4 centrifugal filter unit (Amicon) with a 30 kDa molecular-weight cutoff. The Mss116p concentration was determined by Bradford assay (Bio-Rad Laboratories) using bovine serum albumin (Pierce Biotechnology) as a standard.

A derivative of pMAL-Mss116p expressing the C-terminally truncated protein Mss116p/Δ598–664 was created similarly to other truncations described in Mohr *et al.* (2008). Mss116p/Δ598–664 was purified by the same method as full-length Mss116p (above). SeMet-labeled Mss116p/Δ598–664 was expressed using PASM-5052 auto-inducing medium (Studier, 2005) and purified in the same manner as the full-length protein.

2.2. Crystallization

To form an Mss116p–RNA–AMP-PNP complex, 90 μ*M* Mss116p (6.5 mg ml^{-1}) or Mss116p/Δ598–664 (5.8 mg ml^{-1}) was incubated with 180 μ*M* U_{10} (Integrated DNA Technologies), 1 m*M* AMP-PNP-Mg^{2+} and 2 m*M* $MgCl_2$ for at least 10 min on the desktop. Initial screening was performed by the microbatch method in 96-well round-bottom cell-culture plates (Corning). Plates were set up at ambient temperature by pipetting 0.5 μl complex into each well and then overlaying each drop with 30 μl paraffin oil. 0.5 μl screening reagent was then added to each well using a Phoenix crystallization robot (Art Robbins Instruments) and the plates were kept at 295 K. The reagents screened were from the Index, Crystal Screen, PEG/Ion, Silver Bullets and Natrix kits from Hampton Research and the Wizard and Cryo kits from Emerald BioSystems. Optimization of initial crystallization conditions was performed in 24-well sitting-drop plates using the microbatch method. During optimization, we found that crystal clusters could be grown at 295 K under paraffin oil using only 1 μl complex and 1 μl 6 m*M* $MgCl_2$ (as a precipitant). A crystal seed stock was made from these clusters by using the Seed Bead kit (Hampton Research). Improved single crystals were obtained using the sitting-drop vapor-diffusion method with drops consisting of 1 μl complex, 0.8 μl reagent and 0.2 μl of a 1:100 dilution of crystal seed. For each sitting drop, the reservoir solution was a 1:1 mixture of reagent and Mss116p storage buffer (see above) with a total volume of 250 or 500 μl. Crystals were removed from sitting drops with a nylon loop and flash-cooled immediately with liquid N_2. All crystallization plates were stored at either 295 or 288 K.

2.3. Data collection and processing

Synchrotron X-ray diffraction data from cryocooled crystals were collected on LS-CAT beamline 21-ID-D at the Advanced Photon

Source (APS), Argonne National Laboratory. For the SeMet data set, X-ray diffraction data for two separate parts of a single crystal were collected with the following parameters: detector distance = 340 mm, exposure time = 1 s, φ per image = 1°, total images = 616 (264 and 352 images from the two parts of the crystal). For the native data set, X-ray diffraction data for three separate parts of a single crystal were collected with the following parameters: detector distance = 280 mm, exposure time = 2 s, φ per image = 1°, total images = 438 (158, 100 and 180 images for the three parts of the crystal). Diffraction intensities were recorded on a MAR 300 CCD (MAR Research) and indexed and scaled with *HKL*-2000 (Otwinowski & Minor, 1997).

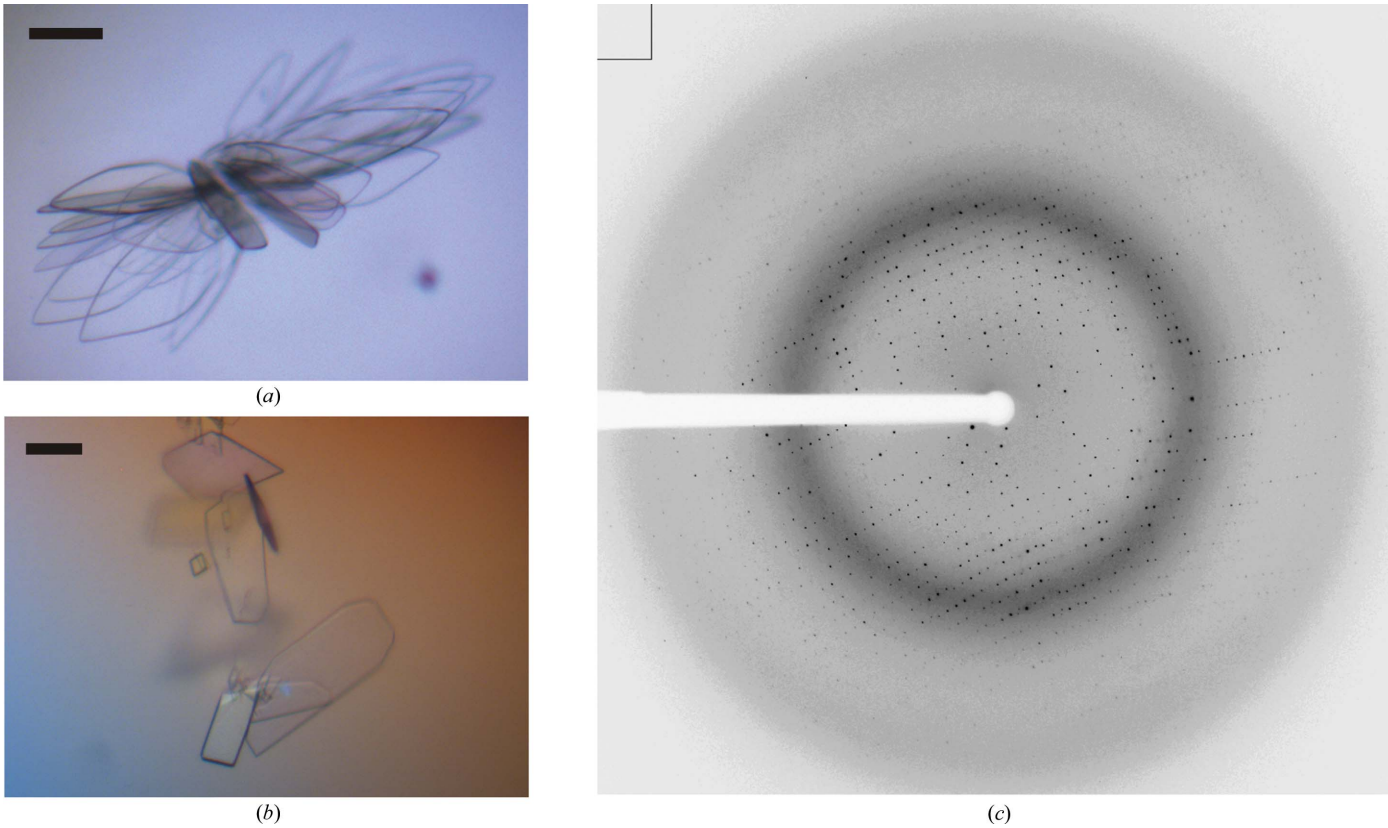
3. Results and discussion

Crystallization trials typically start with protein concentrated to 5–10 mg ml^{-1} in a buffer that contains the minimal solutes to keep the protein soluble. The MBP-Mss116p fusion protein was concentrated to 8 mg ml^{-1} in 20 m*M* Tris–HCl pH 7.5, 1 m*M* EDTA, 1 m*M* DTT (TED buffer) with 0.5 *M* KCl. However, the fusion protein did not yield any crystals even in the presence of U_{10} and AMP-PNP. When 8 mg ml^{-1} fusion protein was treated with TEV protease a hazy white precipitate formed, which was shown by SDS–PAGE to be Mss116p (data not shown). Because the MBP fusion protein was initially employed to improve the solubility of Mss116p during overexpression, it was unsurprising that its removal caused Mss116p to precipitate. We used a sitting-drop solubility screen to test whether glycerol, urea or sugars at different pHs might stabilize Mss116p after TEV protease cleavage of the MBP tag (Gruswitz *et al.*, 2005). Glycerol was the only additive that was found to keep Mss116p soluble after TEV protease cleavage (data not shown).

Both glycerol (Sousa, 1995) and Arg + Glu (Golovanov *et al.*, 2004) have been shown to stabilize proteins and increase their solubility. To help stabilize Mss116p, 10% glycerol was added after purification of the fusion protein but before TEV protease cleavage and 50 m*M* Arg + Glu was added during the gel-filtration purification step. Mss116p could be spin-concentrated to a maximum of ∼2.5 mg ml^{-1} in TED buffer with 0.5 *M* KCl, 50 m*M* Arg + Glu and 10% glycerol, an approximately fivefold increase compared with the same buffer without Arg + Glu. To obtain a concentration of ∼10 mg ml^{-1}, Mss116p was dialyzed into TED buffer with 250 m*M* KCl, 50 m*M* Arg + Glu and 50% glycerol and then spin-concentrated. This buffer was altered slightly for crystallization trials (no EDTA and NaCl instead of KCl; see §2.2).

Initially, Mss116p was mixed with AMP-PNP and U_{10} and crystallization trials were set up with various commercial screens. We used the microbatch method to avoid drop swelling with sitting or hanging drops containing high concentrations of glycerol. Many conditions at pH 7–9.5 containing various polyethylene glycols yielded numerous tiny plate clusters throughout the drop after 1 d at 295 K. The appearance of these plate clusters required both AMP-PNP and U_{10}. However, optimization failed to produce larger single crystals.

To improve the ternary-complex crystals, we tested Mss116p/Δ598–664, a 67-residue C-terminal truncation of Mss116p which lacks the basic tail but is fully functional *in vivo* (Huang, 2004). In initial microbatch screening, Mss116p/Δ598–664 gave larger plate clusters under similar conditions as full-length Mss116p. During



(*a*) (*b*) (*c*)

Figure 1
Mss116p/Δ598–664–U_{10}–AMP-PNP crystals and diffraction pattern. (*a*) Plate clusters obtained by microbatch under oil. (*b*) Single plates obtained by seeding sitting drops. The scale bar is 0.1 mm in (*a*) and (*b*). (*c*) Diffraction pattern from a crystal in (*b*). The edge of the detector corresponds to 1.88 Å resolution.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Complex	AMP-PNP (SeMet)	AMP-PNP (native)
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	a = 88.17, b = 126.48, c = 55.54	a = 88.54, b = 126.52, c = 55.52
Wavelength (Å)	0.97931	0.91842
Total reflections	585879	761109
Unique reflections	24769	48438
Resolution (Å)	30.0–2.40 (2.44–2.40)	35.0–1.90 (1.94–1.90)
Completeness (%)	99.1 (91.9)	96.4 (82.7)
Mean $I/\sigma(I)$	57.6 (10.1)	30.6 (4.9)
$R_{\rm merge}$† (%)	8.5 (22.8)	7.7 (38.9)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

optimization, we found that microbatch drops consisting of 1 µl ternary complex plus 1 µl H_2O yielded similar large plate clusters over 2 d at 295 K. The plate clusters were more reproducible if 6 m*M* $MgCl_2$ was used instead of H_2O (Fig. 1*a*). To obtain single crystals, a portion of a single plate cluster was used to make a crystal seed stock, which was used to screen new sitting drops. Single rectangular plate crystals were obtained with either 0.2 *M* succinate pH 7.0 or 8% tacsimate pH 7.0 plus 15% PEG 3350 (Fig. 1*b*).

A diffraction experiment on APS beamline 21-ID-D showed that these rectangular plate crystals diffracted X-rays to better than 1.9 Å resolution (Fig. 1*c*). Data reduction and scaling to 1.9 Å showed that the crystals were orthorhombic, with unit-cell parameters a = 88.54, b = 126.52, c = 55.52 Å. The space group was narrowed to $P2_12_12$ by using *POINTLESS* (Evans, 2006). The Matthews coefficient $V_{\rm M}$ (Matthews, 1968) was 2.4 $Å^3\,Da^{-1}$, corresponding to a solvent content of 55%. Additional details of data collection are provided in Table 1.

Although molecular replacement using another DEAD-box protein as a search model would have been a reasonable strategy to solve this data set, we found it easy to prepare crystals with SeMet-labeled protein in order to obtain phases experimentally. A SAD data set was collected from an SeMet-labeled derivative at the Se absorption edge. By using data to 2.4 Å resolution, *autoSHARP* (Vonrhein *et al.*, 2007) was able to locate all eight Se sites and produce an interpretable electron-density map that verified the space group with figures of merit (FOMs) of 0.43 and 0.11 for acentric and centric reflections, respectively. A SIRAS experiment with the native and Se data sets in *autoSHARP* produced a high-quality electron-density map for model building, with FOMs of 0.47 and 0.22 for acentric and centric reflections, respectively. Model building, refinement and structure description are reported elsewhere (Del Campo & Lambowitz, 2009).

We thank Paul Paukstelis for help with data collection and processing and for comments on the manuscript. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817). We are grateful to Joe Brunzelle at APS Sector 21 for help with data collection. This work was supported by NIH grant GM037951 to AML. MD was the recipient of a Ruth Kirschstein NRSA postdoctoral fellowship F32 GM076961 from the NIH.

References

Andersen, C. B., Ballut, L., Johansen, J. S., Chamieh, H., Nielsen, K. H., Oliveira, C. L., Pedersen, J. S., Séraphin, B., Le Hir, H. & Andersen, G. R. (2006). *Science*, **313**, 1968–1972.
Bono, F., Ebert, J., Lorentzen, E. & Conti, E. (2006). *Cell*, **126**, 713–725.
Collins, R., Karlberg, T., Lehtio, L., Schütz, P., van den Berg, S., Dahlgren, L. G., Hammarström, M., Weigelt, J. & Schüler, H. (2009). *J. Biol. Chem.* **284**, 10296–10300.
Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. (2006). *Gene*, **367**, 17–37.
Del Campo, M. & Lambowitz, A. M. (2009). In the press.
Del Campo, M., Mohr, S., Jiang, Y., Jia, H., Jankowsky, E. & Lambowitz, A. M. (2009). *J. Mol. Biol.* **389**, 674–693.
Del Campo, M., Tijerina, P., Bhaskaran, H., Mohr, S., Yang, Q., Jankowsky, E., Russell, R. & Lambowitz, A. M. (2007). *Mol. Cell*, **28**, 159–166.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L. Y. (2004). *J. Am. Chem. Soc.* **126**, 8933–8939.
Grohman, J. K., Del Campo, M., Bhaskaran, H., Tijerina, P., Lambowitz, A. M. & Russell, R. (2007). *Biochemistry*, **46**, 3013–3022.
Gruswitz, F., Frishman, M., Goldstein, B. M. & Wedekind, J. E. (2005). *Biotechniques*, **39**, 476–480.
Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E. & Lambowitz, A. M. (2007). *J. Mol. Biol.* **365**, 835–855.
Huang, H.-R. (2004). PhD thesis. University of Texas Southwestern Medical Center, Dallas, Texas, USA.
Huang, H.-R., Rowe, C. E., Mohr, S., Jiang, Y., Lambowitz, A. M. & Perlman, P. S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 163–168.
Jankowsky, E. & Fairman, M. E. (2007). *Curr. Opin. Struct. Biol.* **17**, 316–324.
Linder, P. (2006). *Nucleic Acids Res.* **34**, 4168–4180.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Moeller, H. von, Basquin, C. & Conti, E. (2009). *Nature Struct. Mol. Biol.* **16**, 247–254.
Mohr, G., Del Campo, M., Mohr, S., Yang, Q., Jia, H., Jankowsky, E. & Lambowitz, A. M. (2008). *J. Mol. Biol.* **375**, 1344–1364.
Mohr, S., Matsuura, M., Perlman, P. S. & Lambowitz, A. M. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 3569–3574.
Mohr, S., Stryker, J. M. & Lambowitz, A. M. (2002). *Cell*, **109**, 769–779.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. (2006). *Cell*, **125**, 287–300.
Solem, A., Zingler, N. & Pyle, A. M. (2006). *Mol. Cell*, **24**, 611–617.
Sousa, R. (1995). *Acta Cryst.* D**51**, 271–277.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Tijerina, P., Bhaskaran, H. & Russell, R. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 16698–16703.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. (2007). *Mol. Cell*, **28**, 253–263.
Yang, Q. & Jankowsky, E. (2006). *Nature Struct. Mol. Biol.* **13**, 981–986.