RRM domain of human RBM7: purification, crystallization and structure determination

Nicholas Sofos, Mikael B. L. Winkler and Ditlev E. Brodersen*

Centre for mRNP Biogenesis and Metabolism, Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark. *Correspondence e-mail: deb@mbg.au.dk

Received 15 March 2016

Accepted 12 April 2016

Edited by I. Tanaka, Hokkaido University, Japan

Keywords: RNA degradation; human; NEXT complex; RRM domain; X-ray crystallography.

PDB reference: human RBM7 RRM domain, 5iqq

Supporting information: this article has supporting information at journals.iucr.org/f

RNA decay is an important process that is essential for controlling the abundance, quality and maturation of transcripts. In eukaryotes, RNA decay in the 3′–5′ direction is carried out by the exosome, an RNA-degradation machine that is conserved from yeast to humans. A range of cofactors stimulate the enzymatic activity of the exosome and serve as adapters for the many RNA substrates. In human cells, the exosome associates with the heterotrimeric nuclear exosome targeting (NEXT) complex consisting of the DExH-box helicase hMTR4, the zinc-finger protein hZCCHC8 and the RRM-type protein hRBM7. Here, the 2.5 Å resolution crystal structure of the RRM domain of human RBM7 is reported. Molecular replacement using a previously determined solution structure of RBM7 was unsuccessful. Instead, RBM8 and CBP20 RRM-domain crystal structures were used to successfully determine the RBM7 structure by molecular replacement. The structure reveals a ring-shaped pentameric assembly, which is most likely a consequence of crystal packing.

1. Introduction

RNA surveillance, maturation and decay in eukaryotic cells depend to a great extent on the 3′–5′ exonucleolytic and endonucleolytic activities of the RNA exosome complex (Chlebowski *et al.*, 2013; Schneider & Tollervey, 2013). Associated proteins regulate exosome activity and the recruitment of substrates, with the best characterized cofactors being the nuclear Trf4p/5p–Air1p/2p–Mtr4p polyadenylation (TRAMP) complex and the cytoplasmic Ski2p–Ski3p–Ski8p (SKI) complex in *Saccharomyces cerevisiae* (LaCava *et al.*, 2005; Wyers *et al.*, 2005; Vaňáčová *et al.*, 2005; Anderson & Parker, 1998; Brown *et al.*, 2000). In human cells, the nuclear RNA exosome is directed to many of its nucleoplasmic substrates by the nuclear exosome targeting (NEXT) complex, consisting of the DExH-box helicase and yeast Mtr4p homologue hMTR4/SKIV2L2, the RRM-protein RBM7 and the zinc-finger protein ZCCHC8 (Lubas *et al.*, 2011, 2015; Andersen *et al.*, 2013).

Accumulating evidence suggests a role for the NEXT complex in the surveillance of RNA polymerase II transcripts, including pre-mRNAs, PROMPTs, sn(o)RNAs, enhancer RNAs and replication-dependent histone RNAs (Andersen *et al.*, 2013; Hrossova *et al.*, 2015; Lubas *et al.*, 2011, 2015). RBM7 has been shown to be involved in targeting the NEXT complex to U-rich stretches in RNA, requiring a minimum of four pyrimidine nucleotides for efficient binding (Lubas *et al.*, 2015; Hrossova *et al.*, 2015). A recent NMR structure showed that the N-terminal domain of human RBM7 adopts a classical

Table 1

Data-collection and refinement statistics for human RBM7 1–91.

Values in parentheses are for the outer resolution shell.

Data collection	
PDB code	5iqq
Wavelength (Å)	0.9720
Resolution range (Å)	49.1–2.5 (2.6–2.5)
Space group	C121
Unit-cell parameters (Å, °)	$a = 115.71$, $b = 83.21$, $c = 60.83$, $\alpha = 90$, $\beta = 119.63$, $\gamma = 90$
Observed reflections	78050 (5716)
Unique reflections	17201 (1266)
R_{meas} (%)	17.7 (116.5)
$CC_{1/2}$ (%)	99.3 (47.7)
Multiplicity	4.5 (4.5)
Completeness (%)	96.8 (97.2)
Mean $I/\sigma(I)$	10.6 (1.5)
Refinement	
Reflections used in refinement	16761 (1632)
Reflections used for R_{free}	848 (98)
R (%)	21.6 (29.6)
R_{free} (%)	24.6 (35.7)
No. of non-H atoms	
Total	3521
Protein	3267
Solvent	254
No. of protein residues	405
R.m.s.d., bond lengths (Å)	0.004
R.m.s.d., bond angles (°)	0.70
Ramachandran favoured (%)	95.0
Ramachandran allowed (%)	3.0
Ramachandran outliers (%)	2.0
Average B factor (Å ²)	
Overall	45.9
Macromolecule	46.1
Solvent	44.2

RNA-recognition motif (RRM) α/β -sandwich fold with a β_1 – α_1 – β_2 – β_3 – α_2 – β_4 topology, which is a putative RNA-binding domain of about 90 amino-acid residues (Hrossova *et al.*, 2015; Maris *et al.*, 2005). The RBM7 RRM domain also contains the RNP1 [(RK)-G-(FY)-(GA)-(FY)-(ILV)-X-(FY)] and RNP2 [(ILV)-(FY)-(ILV)-X-N-L] signature sequences characteristic of the RRM family. The only deviation is the second residue of the RNP1 sequence of RBM7, which is Q and not G (Dreyfuss *et al.*, 1988; Hrossova *et al.*, 2015). RNA-binding studies have further shown that Phe13 and Phe52 of the first and third β -strands, respectively, are important for RNA binding, most likely through stacking interactions with nucleotide bases (Hrossova *et al.*, 2015; Maris *et al.*, 2005).

Here, we report the crystallization, structure determination and structural analysis of the crystal structure of the human RBM7 RRM domain determined to 2.5 Å resolution. The structure shows an unusual pentameric assembly, as well as an extended α_2 -helix and two short, tentative β -strands compared with the corresponding NMR model.

2. Materials and methods

2.1. Cloning, expression and purification

The part of the human RBM7 gene encoding residues 1–91 of the intact protein was inserted into the pET28M-6×His-SUMOTag expression vector (Tomecki *et al.*, 2015). The resulting plasmid was transformed into *Escherichia coli*

Rosetta 2 (DE3) competent cells, which were grown in LB medium supplemented with 34 $\mu\text{g ml}^{-1}$ chloramphenicol and 50 $\mu\text{g ml}^{-1}$ kanamycin at 37°C until the OD₆₀₀ reached 0.6–0.8. Cultures were chilled in ice water before expression at 18°C by the addition of 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) overnight. The cells were harvested and resuspended in 50 mM HEPES pH 7.0, 0.5 M KCl, 40 mM imidazole, 10 mM MgCl₂, 10% glycerol, 3 mM β -mercaptoethanol and protease-inhibitor cocktail tablets (Sigma–Aldrich) and were subsequently lysed by sonication and centrifuged at 30 000g for 30 min to remove cell debris. The supernatant was loaded onto a pre-equilibrated 5 ml Ni–NTA column (Qiagen). The bound protein was washed with 1 M KCl and eluted with 50 mM HEPES pH 7.0, 0.25 M KCl, 0.3 M imidazole, 10% (v/v) glycerol, 3 mM β -mercaptoethanol and subsequently exchanged into 50 mM HEPES pH 7.0, 0.3 M KCl, 10% glycerol, 5 mM β -mercaptoethanol using a HiPrep 26/10 desalting column (GE Healthcare). The His-SUMO tag was released by treatment with Ulp protease (a kind gift from A. Dziembowski, Warsaw University; Mossessova & Lima, 2000) and subsequently removed by reapplying the sample onto the Ni²⁺-affinity column. Finally, RBM7 1–91 was purified by gel-filtration chromatography on a Superdex 75 10/300 column (GE Healthcare) equilibrated in 20 mM HEPES pH 7.0, 150 mM KCl, 5 mM β -mercaptoethanol and was concentrated to $\sim 10 \text{ mg ml}^{-1}$ using a 5000 molecular-weight cutoff spin filter (Sartorius).

2.2. Crystallization, data collection and structure determination

Human RBM7 1–91 was crystallized at 19°C using a reservoir solution consisting of 0.1 M HEPES pH 7.5, 1.4 M trisodium citrate with 0.2 + 0.2 μl sitting-drop vapour-diffusion drops. Rhombic crystals appeared within 4 d and grew to a maximum size of 100–150 μm in two dimensions. The crystals were transferred to a drop of mother liquor supplemented with 10% (v/v) glycerol and immediately flash-cooled in liquid nitrogen. Diffraction data extending to 2.5 Å resolution were collected on beamline I04 at Diamond Light Source, where 2400 fine-sliced diffraction images were collected with a 0.10° oscillation angle from a single crystal and were autoprocessed with *xia2* via *XDS/XSCALE* (Winter *et al.*, 2013; Kabsch, 2010). The crystals belonged to space group C121 with five molecules per asymmetric unit. Structure determination was carried out by molecular replacement in *Phaser* using an ensemble search model consisting of PDB entries 2j0q (Bono *et al.*, 2006) and 1h2v (Mazza *et al.*, 2002). The successful MR solution was subjected to one round of automatic model building and refinement in *ARP/wARP Classic* as implemented in the *CCP4* suite (Morris *et al.*, 2003; Perrakis *et al.*, 2001; Winn *et al.*, 2011). Subsequently, the model was manually rebuilt in *Coot* (Emsley *et al.*, 2010) and iteratively refined using *phenix.refine* (Afonine *et al.*, 2012) and *REFMAC* (Murshudov *et al.*, 2011). Data-collection, processing and refinement statistics are shown in Table 1.

3. Results and discussion

3.1. Purification of the human RBM7 RRM domain

Sequence analysis of human RBM7 using *GlobPlot* and *Phyre2* initially predicted the C-terminal region to be largely disordered, and thus residues 1–91 corresponding to the core RBM7 RRM domain were chosen for structural analysis (Kelley *et al.*, 2015; Linding *et al.*, 2003). The corresponding DNA sequence was inserted into pET28M-6×His-SUMOTag, giving rise to a protein with dual N-terminal 6×His affinity and SUMO solubility tags (Fig. 1*a*; see §2 for details). Following Ni²⁺-affinity purification and removal of the dual tag using Ulp protease, monodisperse protein was obtained using gel filtration on a high-resolution Superdex 75 10/300 column (Fig. 1*b*). During gel filtration, the RBM7 RRM domain elutes at 14.2 ml, corresponding to a theoretical molecular mass of approximately 12 kDa. Based on this, we conclude that the isolated RBM7 RRM domain is a monomer in solution.

3.2. Crystallization and structure determination

Rhombic crystals of the RBM7 RRM domain were obtained in 1.4 M citrate pH 7.5 at 19°C. The crystals diffracted to 2.5 Å resolution and belonged to space group C121 with five RBM7 RRM monomers per asymmetric unit. Surprisingly, structure determination by molecular replacement using a recently determined structure of a similar fragment of human RBM7 determined in solution by NMR (PDB entry 2m8h) was unsuccessful (Hrossova *et al.*, 2015). Instead, a 3D-BLAST structure search was performed using a minimal domain fragment of the NMR structure as a search model

(Yang & Tung, 2006). This search identified RBM8 from the crystal structure of the exon junction complex (PDB entry 2j0q; 28% sequence identity, 49% sequence similarity) and CBP20 derived from the human cap-binding complex (PDB entry 1h2v; 34% sequence identity, 51% sequence similarity) as the closest homologous structures in the PDB (Bono *et al.*, 2006; Mazza *et al.*, 2002). Using a search ensemble of these two structures, we were able to obtain a good molecular-replacement solution in *Phaser* with a Z-score of 11.4 (McCoy *et al.*, 2007). After one round of automatic rebuilding in *ARP/wARP* (Morris *et al.*, 2003; Perrakis *et al.*, 2001), the structure was completed by iterative refinement in *PHENIX* and *REFMAC* (Adams *et al.*, 2010; Murshudov *et al.*, 2011), resulting in final *R* and *R*_{free} values of 21.6 and 24.6%, respectively (Table 1). The final model contains five copies of the RBM7 RRM domain including residues 4–86 (chain A), 8–85 (chain B), 3–85 (chain C), 5–85 (chain D) and 6–85 (chain E).

3.3. The human RBM7 RRM domain forms a pentameric assembly

The five RBM7 molecules per crystallographic asymmetric unit assemble into a pentameric ring structure with fivefold noncrystallographic symmetry (Fig. 2*a*). In the crystal packing, the pentamers are arranged in plates stacked on top of each other, with solvent channels running through the centre of the rings and little room between the individual plates (Fig. 2*b*). Neighbouring molecules only share 1–2 hydrogen bonds and a single hydrophobic interaction between Tyr76 and Pro64 (Fig. 3). An intersubunit hydrogen-bond interaction between

Figure 1

Purification of the human RBM7 RRM domain. (a) A 15% SDS-PAGE stained with Coomassie Brilliant Blue showing the progress of protein purification. His-SUMO-RBM7 1–91 was first isolated by Ni-NTA chromatography before removal of the His-SUMO tag by Ulp protease cleavage and subsequent separation by Ni-NTA chromatography. Lane M contains molecular-mass marker (labelled in kDa). (b) RBM7 1–91 elutes at 14.2 ml on a Superdex 75 10/300 column, corresponding to a molecular weight of ~12 kDa. The void volume (*V*₀) and elution volumes of standard proteins are shown above the chromatogram (*A*₂₈₀, blue; *A*₂₆₀, red).

the backbone carbonyl of Lys74 and the side-chain amido NH₂ group of Asn68 is consistent throughout the ring assembly,

whereas the Tyr76 and Glu60 side chains form hydrogen bonds between two pairs of adjacent molecules (Fig. 3). Analysis of

Figure 2

(a) The asymmetric unit of the RBM7 1–91 crystals contains a pentameric assembly of RRM domains. The model is shown in cartoon representation in various shades of green. (b) Crystal packing of the RBM7 1–91 pentamer structure. Solvent channels run through the centres of the pentamers (top), which form leaflets that are stacked on top of each other in the perpendicular direction (bottom).

Figure 3

Final σ_A -weighted $2DF_o - mF_c$ map contoured at 1σ , showing representative electron density at the interface between chains C (dark green) and D (light green). The enlarged view (right) shows details of the intersubunit interactions observed. Subscript letters indicate the chain to which the respective residues belong. Tyr76_D and Glu60_C form a hydrogen bond, Tyr76_D and Pro64_C engage in hydrophobic interactions, and the backbone carbonyl O atom of Lys74_D forms a hydrogen bond to the side-chain amide of Asn68_C.

Figure 4
(a) Cartoon overview of the human RBM7 RRM domain with secondary-structure elements labelled according to the notation shown in (c). β_A and β_B (orange) are short, tentative β -strands that are not part of the canonical RRM motif, with the latter being an extension of the β_4 strand. (b) Superposition of an ensemble of the 20 lowest energy NMR structures of the human RBM7 RRM domain (light grey) with chain D of the crystal structure (dark green), in the same orientation as in (a). (c) Amino-acid sequence of the N-terminal part of human RBM7 with the extent of the crystal structure shown above and helices (brown) and sheets (grey) annotated as in (a). β_A and β_B are shown in light grey with dashed outlines. The RNP2 and RNP1 motifs are indicated in the dashed boxes. Gln51, shown with dark blue shading, is a deviation from the RNP1 consensus sequence, which has a glycine at this position.

the contents of the asymmetric unit using PISA revealed that the protein interface areas of the pentamer average around $274 \text{ \AA}^2$, with a $\Delta G_{\text{interaction}}$ of around $-3.9 \text{ kcal mol}^{-1}$ (Krisinel & Henrick, 2007). Put together, the low number of interactions and the small size of the interface area strongly indicate that the pentamer is a crystal-packing artefact and thus is consistent with the behaviour of the human RBM7 RRM domain in solution.

3.4. Analysis of the RBM7 RRM fold

The human RBM7 RRM domain adopts a canonical RRM fold in all five molecules in the crystal structure, consisting of two α -helices packed against a four-stranded antiparallel β -sheet (Fig. 4a). In addition to the core RRM motif, the structure contains two short, tentative β -strands, which we denote β_A and β_B (Fig. 4c). The latter is an extension of the β_4 strand. These were not observed in the NMR structure and are not part of the classical RRM fold (Maris *et al.*, 2005). There are only minor structural differences between the five molecules in the asymmetric unit (r.m.s.d. on C^α atoms of $0.137 \text{ \AA}$), which are primarily located in the N-terminus (residues 4–10) and in the long loop between β_2 and β_3 . Compared with the NMR structure (Hrossova *et al.*, 2015), helix α_2 is longer by

four residues (residue 60–63) in the crystal structure, likely owing to the packing interaction between Pro64 and Tyr76 of the adjacent molecule that appears to stabilize the helix despite the presence of a proline. Finally, the loop regions at one end of the β -sheet are slightly further apart in the crystal structure compared with the NMR structure (Fig. 4b).

4. Conclusion

We have crystallized and determined the crystal structure of the core RRM domain of human RBM7 covering residues 1–91 to $2.5 \text{ \AA}$ resolution. An ensemble of the homologous RRM-domain structures from CBP20 and RBM8 could successfully be used to determine the structure by molecular replacement. The crystallographic asymmetric unit contains five RRM domains arranged in a pentameric ring structure, but weak intermolecular interactions and a small intersubunit interface, as well as gel-filtration chromatography, strongly suggest that the pentamer is a crystal-packing artefact. The human RBM7 RRM structure is very similar to a recent structure determined in solution by NMR. However, small differences in the loop regions are observed, and α_2 is extended in the crystal structure compared with the solution structure. Finally, we observe

two short, tentative β -strands, β_A and β_B , which are not part of the classical RRM fold.

Acknowledgements

The authors would like to thank Andrzej Dziembowski and Anna Łabno for providing a full-length clone for human RBM7 in pET28M-6×His-SUMOtag, which was used to construct the RBM7 1–91 fragment, and Thomas Boesen for collecting the diffraction data, as well as the beamline staff at Diamond Light Source beamline I04 (proposal/visit No. mx9317-18). This work was funded by the Danish National Research Foundation's Centre for mRNP Biogenesis and Metabolism (grant No. DNRF58) and a Novo Nordisk Foundation grant (NNF15OC0017384) to DEB. The authors declare no competing financial interests.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Andersen, P. R. *et al.* (2013). *Nature Struct. Mol. Biol.* **20**, 1367–1376.
- Anderson, J. S. & Parker, R. P. (1998). *EMBO J.* **17**, 1497–1506.
- Bono, F., Ebert, J., Lorentzen, E. & Conti, E. (2006). *Cell*, **126**, 713–725.
- Brown, J. T., Bai, X. & Johnson, A. W. (2000). *RNA*, **6**, 449–457.
- Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. (2013). *Biochim. Biophys. Acta*, **1829**, 552–560.
- Dreyfuss, G., Swanson, M. S. & Piñol-Roma, S. (1988). *Trends Biochem. Sci.* **13**, 86–91.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Hrossova, D., Sikorsky, T., Potesil, D., Bartosovic, M., Pasulka, J., Zdrahal, Z., Stefl, R. & Vanacova, S. (2015). *Nucleic Acids Res.* **43**, 4236–4248.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. (2015). *Nature Protoc.* **10**, 845–858.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. & Tollervey, D. (2005). *Cell*, **121**, 713–724.
- Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. (2003). *Nucleic Acids Res.* **31**, 3701–3708.
- Lubas, M., Andersen, P. R., Schein, A., Dziembowski, A., Kudla, G. & Jensen, T. H. (2015). *Cell. Rep.* **10**, 178–192.
- Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., Andersen, J. S., Dziembowski, A. & Jensen, T. H. (2011). *Mol. Cell*, **43**, 624–637.
- Maris, C., Dominguez, C. & Allain, F. H. (2005). *FEBS J.* **272**, 2118–2131.
- Mazza, C., Segref, A., Mattaj, I. W. & Cusack, S. (2002). *EMBO J.* **21**, 5548–5557.
- 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.
- Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
- Mosessoova, E. & Lima, C. D. (2000). *Mol. Cell*, **5**, 865–876.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). *Acta Cryst.* **D57**, 1445–1450.
- Schneider, C. & Tollervey, D. (2013). *Trends Biochem. Sci.* **38**, 485–493.
- Tomecki, R., Drazkowska, K., Krawczyk, A., Kowalska, K. & Dziembowski, A. (2015). *Methods Mol. Biol.* **1259**, 417–452.
- Vaňáčová, Š., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G. & Keller, W. (2005). *PLoS Biol.* **3**, e189.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* **D69**, 1260–1273.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnauld, B., Devaux, F., Namane, A., Séraphin, B., Libri, D. & Jacquier, A. (2005). *Cell*, **121**, 725–737.
- Yang, J.-M. & Tung, C.-H. (2006). *Nucleic Acids Res.* **34**, 3646–3659.