Crystal structure of human Acinus RNA recognition motif domain

Humberto Fernandes 1, 2, Honorata Czapinska 1, Katarzyna Grudziaz 1, Janusz M. Bujnicki Corresp. 1, 3, Martyna Nowacka Corresp. 1

1 International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
2 Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland
3 Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland

Corresponding Authors: Janusz M. Bujnicki, Martyna Nowacka
Email address: lamb@genesilico.pl, mnowacka@genesilico.pl

Acinus is an abundant nuclear protein involved in apoptosis and splicing. It has been implicated in inducing apoptotic chromatin condensation and DNA fragmentation during programmed cell death. Acinus undergoes activation by proteolytic cleavage that produces a truncated p17 form that comprises only the RRM domain. We have determined the crystal structure of the human Acinus RRM domain at 1.65 Å resolution. It shows a classical four-stranded antiparallel β-sheet fold with two flanking α-helices and an additional, non-classical α-helix at the C-terminus, which harbors the caspase-3 target sequence that is cleaved during Acinus activation. In the structure, the C-terminal α-helix partially occludes the potential ligand binding surface of the β-sheet and hypothetically shields it from non-sequence specific interactions with RNA. Based on the comparison with other RRM-RNA complex structures, it is likely that the C-terminal α-helix changes its conformation with respect to the RRM core in order to enable RNA binding by Acinus.
Crystal structure of human Acinus RNA recognition motif domain

Humberto Fernandes\textsuperscript{1,2,4}, Honorata Czapinska\textsuperscript{1,4}, Katarzyna Grudziaz\textsuperscript{1}, Janusz M. Bujnicki\textsuperscript{1,3}, Martyna Nowacka\textsuperscript{1}

\textsuperscript{1} International Institute of Molecular and Cell Biology in Warsaw, Ks. Trojdena 4, 02-109 Warsaw, Poland
\textsuperscript{2} Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland
\textsuperscript{3} Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland
\textsuperscript{4} Co-first authors

Corresponding authors:
✉ Martyna Nowacka
International Institute of Molecular and Cell Biology in Warsaw, Ks. Trojdena 4, 02-109 Warsaw, Poland
e-mail: mnowacka@genesilico.pl
Current address:
University of Warsaw Biological and Chemical Research Centre, Zwirki i Wigury 101, 02-089, Warsaw, Poland
e-mail: mnowacka@chem.uw.edu.pl

✉ Janusz Bujnicki
International Institute of Molecular and Cell Biology in Warsaw, Ks. Trojdena 4, 02-109 Warsaw, Poland
Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland
e-mail: iamb@genesilico.pl

Abstract

Acinus is an abundant nuclear protein involved in apoptosis and splicing. It has been implicated in inducing apoptotic chromatin condensation and DNA fragmentation during programmed cell death. Acinus undergoes activation by proteolytic cleavage that produces a truncated p17 form that comprises only the RRM domain. We have determined the crystal structure of the human Acinus RRM domain at 1.65 Å resolution. It shows a classical four-stranded antiparallel β-sheet fold with two flanking α-helices and an additional, non-classical α-helix at the C-terminus, which harbors the caspase-3 target sequence that is cleaved during Acinus activation. In the structure,
the C-terminal α-helix partially occludes the potential ligand binding surface of the β-sheet and hypothetically shields it from non-sequence specific interactions with RNA. Based on the comparison with other RRM-RNA complex structures, it is likely that the C-terminal α-helix changes its conformation with respect to the RRM core in order to enable RNA binding by Acinus.

Introduction

The RNA recognition motif (RRM) is a small (approximately 90 residues) protein domain known for its characteristic βαββαβββαβ fold and frequent engagement in RNA binding. RRM-containing proteins are typically engaged in splicing, editing, export, degradation and regulation of translation (Maris, Dominguez & Allain, 2005). A single RRM typically recognizes a short continuous stretch of single-stranded RNA, usually less than eight ribonucleotides long (Afroz et al., 2015). Binding involves two conserved motifs termed RNP1 and RNP2, located in the two central β-strands, with [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] and [ILV]-[FY]-[ILV]-X-N-L consensus sequences, respectively (X stands for any amino acid) (Maris, Dominguez & Allain, 2005; Clery, Blatter & Allain, 2008). Prominent features of these motifs include conserved aromatic residues that are often involved in interactions with ribonucleotides of the RNA target sequence. In addition to the RRM core architecture, other structural elements may or may not be present: β-hairpins (Volpon et al., 2005), extra β-strands (Oberstrass et al., 2005; Blatter et al., 2015) or α-helices (Jacks et al., 2003). These additional elements, together with variable loops, can contribute to the recognition of an RNA target.

Acinus (apoptotic chromatin condensation inducer in the nucleus) is a protein involved in apoptosis and alternative splicing regulation. Human Acinus has three isoforms, Acinus-L, Acinus-S and Acinus-S’, which are most probably generated by alternative splicing and/or usage of alternative promoters (Sahara et al., 1999). The isoforms differ in size and N-termini (Fig. 1). The longest isoform L has a putative DNA binding domain, SAP, and two shorter isoforms, S and S’, have unique sequences at their N-termini (Sahara et al., 1999). All of them contain an RRM domain at the C-terminus and a conserved arginine-serine-repeat (RS repeat) region that are frequently present in splicing factors.

It was originally discovered that Acinus is activated by caspase-3, which recognizes its consensus DELD target sequence and cleaves it after D1093 (residue numbering of the Acinus-L isoform is assumed throughout the text). The truncated version of Acinus (called p17, residues 987-1093) covers the RRM domain and induces chromatin condensation at the terminal stages of apoptosis (Sahara et al., 1999; Hu et al., 2005). The cleavage of Acinus by caspase-3 can be inhibited by phosphorylation of S1180 and S1331 (Hu et al., 2005) or by the antiapoptotic AAC-11 (antiapoptosis clone 11) protein which directly binds to Acinus in the region comprising residues 840-918 (Rigou et al., 2009).

In other studies, knock downs of Acinus did not prevent apoptotic chromatin condensation but instead impaired DNA fragmentation suggesting involvement in DNA cleavage...
During programmed cell death (Joselin, Schulze-Osthoff & Schwerk, 2006; Rigou et al., 2009), later, Acinus was also found in ribonucleoprotein complexes: the spliceosome (Rappsilber et al., 2002; Zhou et al., 2002), the apoptosis- and splicing-associated protein (ASAP) complex (Schwerk et al., 2003; Murachelli et al., 2012) and the splicing-dependent multiprotein exon junction complex (EJC) (Tange et al., 2005). Consequently, its function in splicing became apparent. Acinus was shown to regulate splicing of a subset of apoptotic genes in human cells (Michelle et al., 2012) and to facilitate constitutive splicing of pre-mRNAs containing a weak, alternative 5′ splice site, favoring the mRNA isoform generated from this site (Rodor, 2016). Other results coupled Acinus to transcription and splicing of retinoic acid receptor (RAR)-dependent genes (Vucetic et al., 2008; Wang, Soprano & Soprano, 2015). Moreover, the RRM domain was found necessary for the enhancement of Acinus activity on the retinoic acid (RA) dependent splicing of a mini-gene from its weak 5′ splice site (Wang, Soprano & Soprano, 2015).

The RNA binding mode of Acinus remains unknown. Recently, however, Acinus-S was found to bind to a subset of suboptimal introns of pre-mRNAs, as well as to spliced mRNAs (Rodor et al., 2016). The analysis of iCLIP (individual-nucleotide resolution UV crosslinking and immunoprecipitation) data for Acinus-S demonstrated an enrichment of two sequence motifs: a U-rich sequence (in introns), and a “GAAGAA”-like motif (in exons). Additionally, it was shown that Acinus regulates the splicing of DFFA/ICAD that is an inhibitor of caspase-activated DNase and a major regulator of DNA fragmentation (Rodor et al., 2016).

Here, we report for the first time a crystal structure of the human Acinus RRM domain (AcRRM) at 1.65 Å resolution together with the analysis of its putative RNA binding mode. The structure of an AcRRM monomer shows a classic four-stranded β-sheet fold with two flanking α-helices and an additional non-classical α-helix at the C-terminus. In the absence of the target RNA, the C-terminal α-helix partially occludes the potential ligand binding site (RNP1 and RNP2).

**Materials and Methods**

**Cloning, expression and purification of Acinus RRM**

The protein fragment encoding the AcRRM domain was designed based on the secondary structure prediction of Acinus protein performed with the Genesilico meta-server (Kurowski & Bujnicki, 2003). The human Acinus cDNA (ID 9021673) was purchased from IMAGE Consortium Source BioScience. The DNA sequence encoding AcRRM (residues 1008-1100, Uniprot sequence ID Q9UKV3) was PCR amplified and subcloned into pGEX-4T1 (GE Healthcare Life Sciences). *E. coli* BL21(DE3) strain (New England BioLabs) was used to overexpress the glutathione S-transferase (GST) fusion-tagged AcRRM. Expression was carried out in LB medium, induced with 1 mM isopropyl-D-1-thiogalactopyranoside solution (IPTG) at OD₆₀₀ of 0.6 and conducted at 37°C with shaking at 200 rpm for 4 hours. The cells were harvested by centrifugation at 4000 x g, at 4°C for 20 minutes. The pellet was resuspended in 1x
PBS buffer, pH 7.9, supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma–Aldrich). The cells were subsequently lysed by French press at 18000 psi and the cell debris was removed by centrifugation at 20000 x g for 30 min. The fusion GST-AcRRM protein was purified by GST affinity with Glutathione-Agarose beads (Sigma-Aldrich) according to the manufacturer protocol (Suppl. Fig. 1A). The affinity-purified recombinant protein contained the thrombin recognition site and was subsequently cleaved with thrombin from bovine plasma (Sigma-Aldrich) to remove the GST tag (100 µg lyophilized thrombin was used to cleave 1 mg of fusion GST-AcRRM). The cleavage was performed for 2-4 hours at room temperature in buffer containing 50 mM Tris-HCl pH 9, and 50 mM NaCl. After cleavage, AcRRM maintained two additional amino acid residues (GS) at its N-terminus. AcRRM was separated from GST tag by heparin affinity using a HiTrap Heparin HP column (GE Healthcare Life Sciences) (Suppl. Fig. 1B), pre-equilibrated in buffer containing 50 mM Hepes-NaOH pH 8 and 1 mM DTT. The bound protein was eluted with gradient of 0 to 1 M NaCl in the 50 mM Hepes-NaOH pH 8 and 1 mM DTT buffer.

Finally, AcRRM was purified by size exclusion chromatography on a Superdex 75 pg column (GE Healthcare Life Sciences) pre-equilibrated in 50 mM Hepes-NaOH pH 8, 100 mM NaCl, 1 mM DTT buffer (Suppl. Fig. 1C). The purified AcRRM protein was concentrated to 8 mg/ml using a 3000 molecular-weight cutoff Amicon Ultra-4 Centrifugal Filter Units (Merck Millipore).

**Crystallization**

Human AcRRM was crystallized at 18°C using 0.2 µl + 0.2 µl sitting drop vapor diffusion method. The Phoenix nano-dispensing robot (Art Robbins instruments) was used to set the crystallization drops with the Morpheus Screen (Molecular Dimensions), Index and Crystal Screen (Hampton Research) in 96 well crystallization plates (Hampton Research). The first needle-like crystals appeared within 2 days. The crystals collected from Index Screen 81 (G9) crystallization reagent (0.2 M Ammonium acetate, 100 mM Tris-HCl pH 8.5, 25% w/v Polyethylene glycol 3 350) were cryo-protected for 10 sec. in reservoir solution supplemented with 25% w/v Polyethylene glycol 400, flash-frozen, stored in liquid nitrogen and used for X-ray data collection.

**X-ray data collection and structure determination**

Native X-ray diffraction data sets extending to 1.65 Å resolution were collected using the CCD detector at beamline 14.2 at the Helmholtz-Zentrum Berlin (BESSY II, Germany) (Gerlach, Mueller & Weiss, 2016). A data set consisting of 120 frames was collected with an oscillation width of 1 degree per frame and a crystal-to-detector distance of 150 mm. The data were indexed and scaled using the XDS software package (Kabsch, 2010).

Primary molecular replacement searches were performed with the MrBUMP/Phaser pipeline (McCoy et al., 2007; Keegan & Winn, 2008) using RRM domain of Rna15 (PDB code 2X1F (Pancevac et al., 2010)) as a search model. It produced a marginal solution with two
monomers in the asymmetric unit, a Z score of 11.5 and a final log-likelihood gain (LLG) of 104. The model was then automatically rebuilt with ARP/wARP (Langer, 2008) excluding some resolution shells that were marked as ice rings. Re-inspection of diffraction images and the Wilson plot revealed two problematic resolution shells (2.29-2.22 and 1.94-1.90 Å). They were then excluded from a second iteration of indexing and scaling. Statistics for the diffraction data collection and (re)processing are presented in Table 1. Iterations of manual rebuilding, using the Coot program (Emsley & Cowtan, 2004), and restrained refinement using CCP4 REFMAC5 (Murshudov et al., 2011) that included TLS optimization with one group per monomer, produced the AcRRM model characterized by crystallographic R factor of 18.5% and R\text{free} of 22.4%. The data processing and refinement statistics are summarized in Table 1. Final model coordinates and the corresponding structure factors were deposited at Protein Data Bank with the 6G6S accession code.

Results

Overall structure

AcRRM exhibits the α/β sandwich fold (Fig. 2A and B). The core βαββαβ topology is followed by an additional C-terminal α-helix not present in the canonical RRM domains. The two molecules in the asymmetric unit of the crystal structure are very similar (rmsd of 0.25 Å for 91 Cα atoms), despite the fact that no non-crystallographic symmetry (NCS) restraints were used during the refinement. The N- and C-termini display good electron density for all residues of molecule A, but are not fully ordered in molecule B. AcRRM has three prolines within its sequence and two of them, P1022 and P1075, are in the cis-conformation in both molecules. The two AcRRM monomers are covalently bound in the crystal structure via a disulfide bond (Fig. 2B), that generates an interface area of approximately 500 Å for each of them (calculated by the PISA server (Krissinel, 2015)). For the two linked AcRRM monomers present in the crystal, the Complexation Significance Score (CSS) calculated by PISA is 1. However, computational modeling of the disulfide bond removal by replacement of C1083 with alanine, results in the CSS of 0. This indicates that in the reducing conditions the dimerization does not take place and the dimeric assembly is likely to be a crystal-packing artifact. The monomeric state of AcRRM in reducing condition was confirmed by gel filtration (Suppl. Fig. 2). The C1083 is conserved in AcRRMs of human, mouse and fly but in other RRM domains this position is generally not conserved (Suppl. Fig. 3).

RRM domains typically bind RNA via the aromatic residues of the RNP1 and RNP2 motifs located in the two central β-strands. These residues are also present in AcRRM. RNP1 comprises residues 1049 to 1056 [KSHCFVTY], with the deviations found at second, third and fourth amino acid of the consensus sequence (underlined). RNP2 comprises residues 1014 to 1019 [VHISNL] with deviation at the second amino acid of the consensus (Fig. 2A). The aromatic residues of RNPs are exposed at the β-sheet surface. In their close vicinity, there are additional aromatic amino acids (Y1094 and H1095) of the C-terminal α-helix packed against the
β-sheet on the opposite side to the α-helices 1 and 2. Furthermore, the C-terminal α-helix contains the DELD caspase-3 target site.

The AcRRM domain has a single conserved region on the surface, which correlates with a part of the protein that is predicted to be positively charged (Fig. 3, Suppl. Fig. 4). The region covers the RNP1 and RNP2 strands of the β-sheet and extends towards its edge. It correlates well with the RNA binding sites of the most similar RRM domains. Therefore, it seems plausible that Acinus binds its RNA target via this region in a way similar to other RRMs.

Comparison with other RRMs

The AcRRM crystal structure was compared with other RRM structures deposited in the Protein Data Bank using the DALI server (Holm & Rosenstrom, 2010). AcRRM was found to be structurally most similar to RRM of Rna15, a subunit of the 3′-end processing factor from *Saccharomyces cerevisiae* (PDB code: 2X1B (Pancevac et al., 2010). The βαββαβ core of both structures align with an rmsd of 1.6 Å (sequence identity of 17%) (Fig. 4A). The RRM of U1A (U1 small nuclear ribonucleoprotein A) was found to be the second closest structural relative of AcRRM (PDB: 5FJ4 (Huang, Wang & Lilley, 2016)). It harbors the C-terminal α-helix and aligns with an rmsd of 2.1 Å (sequence identity of 17%) (Fig. 4B).

There are two sequence insertions in AcRRM compared to the most similar known RRM structures (Fig 4 and 5A): the first insertion (residues 1040 and 1041) distorts the following β-strand making it shorter and increases the size of the loop-2 preceding β2. The second insertion (residues 1076 and 1078) is located in loop-5, making it larger than the corresponding loop in both other structures (RRMs of Rna15 and U1A) (Fig. 4 and 5A). In contrast, loop-3 is substantially shorter than in the similar structures and lacks the DXXT motif implied in the binding of uridine rich RNAs (Muto & Yokoyama, 2012). While loop-2 is located on the side of the RRM domain that usually does not participate in the RNA binding, the length of loops-3 and 5 might be functionally relevant.

Comparison with other RRM structures in complex with RNA

Focusing on the top hundred structures defined by Dali server (Holm & Rosenstrom, 2010) as most similar to AcRRM (Fig. 5B), 31 were RRMs in complex with RNA. They can be divided into two major groups: RRMs that bind single-stranded RNA and RRMs that bind an RNA hairpin.

The first group includes, among others, RRMs of splicing factors: (i) Tra2-β in complex with the (GAA)_2 sequence (PDB code 2RRA, (Tsuda et al., 2011) Fig. 5C), where the AGAA tetra-nucleotide is recognized through the N- and C-terminal extensions, as well as stacking interactions on the β-sheet surface; (ii) Sxl which binds to the 5′-UGUUUUUUU-3′ via a V-shaped, strongly electropositive cleft formed by β-sheets of two consecutive RRM domains (PDB code 1B7F, (Handa et al., 1999) Fig. 5D); (iii) HuD (ELAVL4) which binds (AU)-rich elements (AREs) in the 3′ untranslated regions of many short-lived mRNAs via RNP1 and C-terminus (PDB codes 1G2E and 1FX1 (Wang & Tanaka Hall, 2001)). The nucleotide sequences of all these ligands align poorly in the analyzed structures, but in all of them the RNA
is placed in proximity of RNP1 (contacts are made with U, G and A residues) and RNP2 (contacts are made with U, G and C residues).

The second group comprises multiple structures of U1A RRM in complex with the same RNA loop sequence (AUUGCACUCC) (Fig. 5E) and one structure of the Sxl RRM in complex with UUUGAGCACGUGA (PDB code 4QQB (Hennig et al., 2014)) (nucleotides that are in contact with RNPs are underlined).

**Putative RNA binding mode of AcRRM**

RRM domains are known for having diverse RNA binding modes (Maris, Dominguez & Allain, 2005; Clery, Blatter & Allain, 2008; Muto & Yokoyama, 2012; Daubner, Clery & Allain, 2013) that make the precise prediction of the Acinus interaction with the ligand very difficult. However, there are a few features that are likely to be conserved. In many RRMs, including the ones most similar to AcRRM, the RNA is bound at the face of the β-sheet with its 5'-end on the side of the β4 and 3'-end on the side of the β1-strand (Fig. 5). It seems likely that Acinus binds the RNA in analogous way. The RRMs usually use aromatic residues in position 2 of RNP2 (Fig. 5E) and 5 of RNP1 (Fig. 5C-E) for the stacking interactions with the RNA bases. H1015 and F1053 are present in these positions in AcRRM and are likely to fulfill the same roles (Fig. 5B). The aromatic residue in position 3 of RNP1 is often stacked against the ribose groups of the bound nucleic acid (Fig. 5D). This position is occupied by a H1051 in AcRRM that might have the same function (Fig. 5B). The first amino acid of RNP1 is often positively charged and sometimes interacts with the phosphate backbone. K1049 in AcRRM might play this role, but homologous residues of the similar RRMs are not involved in the interactions with the RNA backbone. While RNP1 and RNP2 mediated stacking interactions with the bases provide the RNA binding affinity, the sequence specificity determinants of RRMs are often located in surrounding secondary structure elements. The β4-strand and the flanking regions provide part of these interactions in RRMs similar to AcRRM (indicated with blue color in Fig. 5), but the protein-RNA hydrogen bonding patterns are too variable for precise predictions (Suppl. Fig. 5).

Some RRM domains can accommodate a spectrum of RNA sequences, and the actual contribution of RNPs to ligand discrimination may not always be exclusive (Clery, Blatter & Allain, 2008; Daubner, Clery & Allain, 2013). The determinants of sequence specific RNA recognition may be located in other structural elements of RRMs. For instance, RRM2 of Sxl (Handa et al., 1999) and RRM1 of PABP (Deo et al., 1999) interact with RNA via loops, in addition to RNPs. In extreme cases, the RRM domains do not bind RNA through the β-sheet at all but they do so through other structural elements, for instance qRRMs use loops-1, 3 and 5 (Dominguez, 2010). The comparison of AcRRM with similar RRM structures suggests that analogous elements might play a role in RNA binding (Fig. 5). These include the mentioned above β4 and its flanking regions (residues 1080-1090) that may provide sequence specific interactions. Alternatively, loop-1 (residues 1017-1025) and/or loop-3 (residues 1046-1051), that are immediately adjacent to the RNPs and characterized by high sequence conservation may be involved (Fig. 3B and 5B).
Acinus has an RNA recognition motif most similar to that of the RRM1 of Sex-lethal protein (Sxl) from Drosophila melanogaster (PDB code: 4QQB). They share 29% sequence identity. Sxl is involved in splicing and translation regulation during sex development in the fly (Samuels, Deshpande & Schedl, 1998) and binds to polypyrimidine tracts through both of its RRM domains connected by a flexible linker. The linker does not participate in RNA binding, however, upon binding it orders into a short, distorted α-helix (Crowder et al., 1999; Handa et al., 1999). Structurally, AcRRM was found to be most similar to RRMs of several other factors implicated in RNA processing: Rna15, U1A, Tra2-β and HuD.

There are several known cases of RRM oligomerization in vivo. For instance, the RRM3 of HuR and the RRM of RBPMS homodimerize through a helical region and bind to RNAs through their β-sheets, located opposite to the dimerization site (Scheiba et al., 2014; Teplova et al., 2016). Despite the fact that AcRRM crystalized as a homodimer, in which two AcRRM monomers are bound via C1083 of their β4, we believe that any potential oligomerization of Acinus protein is unlikely performed by RRM domain in the reducing environment of the nucleus.

In order to better understand the role of Acinus in splicing and apoptosis, more information is needed about its RNA binding mode. If AcRRM bound RNA in the way observed in related structures, the C-terminal α-helix would have to move away in order to make space for the RNA molecule. Such adjustment seems possible, and it is observed in some homologs with known structures. A good example is U1A RRM crystallized in complex with RNA (PDB code 4C4W (Huang & Lilley, 2014)). In this structure, the RRM domains that do not bind RNA have their C-terminal α-helix perfectly overlapping with the C-terminal α-helix of AcRRM (Fig. 6A). In the same crystal, in the RRM domains that bind RNA, the helix moves away to make space for the ligand (Fig. 6B). The C-terminal α-helix of U1A has been reported to cover the β-sheet surface and occlude RNA binding (Law et al., 2013). Similar observations have been made for the N-terminal RNA-binding domain of human polyadenylation factor CstF-64 (Perez Canadillas & Varani, 2003). However, the C-terminal α-helix of CstF-64 does not relocate but unfolds upon RNA binding and extends into the hinge domain. Consecutively, this conformational change may facilitate the assembly of the polyadenylation complex (Perez Canadillas & Varani, 2003). The C-terminal α-helices are present also in the C-terminal RRM of La protein (Jacks et al., 2003) and the RRM of GW182 protein (Eulalio et al., 2009).

Interestingly, the ten-amino acid-long C-terminal α-helix of AcRRM has a caspase-3 target sequence and can be truncated by this protease down to 6 amino acids. It has been shown that Acinus is cleaved from both the N- and C-termini in order to produce the p17 form of the protein, which plays a role in apoptosis and chromatin condensation (Sahara et al., 1999). Due to the fact that the C-terminal α-helix contains the DELD caspase-3 target site and two acidic side chains, it has a highly electronegative potential (Fig. 3C). Therefore, it could shield the
hydrophobic, electropositive surfaces in the absence of RNA (which could otherwise involve in non-sequence specific binding of the ligand). The function of the C-terminal \(\alpha\)-helix of AcRRM and the way in which the protease cleavage influences its conformation needs to be assessed by further biochemical and structural studies. For instance, it would be interesting to determine whether the C-terminal \(\alpha\)-helix is (i) relocated upon RNA binding, destabilized and cleaved by caspase-3 (ii) destabilized and cleaved by caspase-3 ahead of RNA binding via RNPs or, alternatively, (iii) C-terminal \(\alpha\)-helix occludes the RNA binding site on the \(\beta\)-sheet but RNA is bound through the loops?

**Conclusions**

We crystallized and determined the structure of the RRM domain of the human Acinus protein spanning residues 1008-1100, at a resolution of 1.65 Å. The classical \(\beta\alpha\beta\alpha\beta\) core of AcRRM is supplemented by an additional C-terminal \(\alpha\)-helix packed against the \(\beta\)-sheet at the side opposite to the other two helices. The aromatic residues of the RNPs in AcRRM are exposed to the cavity potentially creating an RNA-binding area which is occluded by the C-terminal \(\alpha\)-helix in the absence of the ligand. The additional \(\alpha\)-helix can be found in some other RRM domains (e.g. of U1A or La protein) but it is generally considered to be an auxiliary structural element that may block the RNA-binding site and has to change conformation to enable ligand binding. Interestingly, this \(\alpha\)-helix bears a caspase-3 target site and undergoes cleavage during apoptotic signaling. Further studies should address how caspase-3 cleavage changes the conformation of the AcRRM C-terminus and how it influences ligand binding.

**Acknowledgements**

We thank Dr Justyna Czarnecka and Dr Radosław Pluta for the X-ray diffraction data collection. We thank Dr Agnieszka Kiliszek and Dr Matthew Merski for critical reading of the manuscript. We thank HZB for the allocation of synchrotron radiation beamtime and the assistance during data collection. HC thanks the tutors of the DLS-CCP4 Data Collection and Structure Solution Workshop and in particular Dr Ronan Keegan for the help with the structure solution.

**References**

Afroz T, Cienikova Z, Clery A, Allain FH. 2015. One, Two, Three, Four! How Multiple RRM Read the Genome Sequence. *Methods Enzymol* 558:235-278.

Blatter M, Dunin-Horkawicz S, Grishina I, Maris C, Thore S, Maier T, Bindereif A, Bujnicki JM, Allain FH. 2015. The Signature of the Five-Stranded vRRM Fold Defined by Functional, Structural and Computational Analysis of the hnRNP L Protein. *J Mol Biol* 427:3001-3022.

Clery A, Blatter M, Allain FH. 2008. RNA recognition motives: boring? Not quite. *Curr Opin Struct Biol* 18:290-298.

Daubner GM, Clery A, Allain FH. 2013. RRM-RNA recognition: NMR or crystallography...and new findings. *Curr Opin Struct Biol* 23:100-108.
Deo RC, Bonanno JB, Sonenberg N, Burley SK. 1999. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* 98:835-845.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.

Eulalio A, Tritschler F, Buttn er R, Weichenrieder O, Izaurrealde E, Truffault V. 2009. The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. *Nucleic Acids Res* 37:2974-2983.

Gerlach M, Mueller U, Weiss MS. 2016. The MX beamlines BL14.1-3 at BESSY II. *Journal of large-scale research facilities* 2:1-6.

Hada n N, Nureki O, Kurimoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S. 1999. Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* 398:579-585.

Hennig J, Militti C, Popowicz GM, Wang I, Sonntag M, Geerlof A, Gabel F, Gebauer F, Sattler M. 2014. Structural basis for the assembly of the Sxl-Unr translation regulatory complex. *Nature* 515:287-290.

Holm L, Rosenstrom P. 2010. Dal i server: conservation mapping in 3D. *Nucleic Acids Res* 38:W545-549.

Huang L, Lilley DM. 2014. Structure of a rare non-standard sequence k-turn bound by L7Ae protein. *Nucleic Acids Res* 42:4734-4740.

Huang L, Wang J, Lilley DM. 2016. A critical base pair in k-turns determines the conformational class adopted, and correlates with biological function. *Nucleic Acids Res* 44:5390-5398.

Jacks A, Babon J, Kelly G, Manolaridis I, Cary PD, Curry S, Conte MR. 2003. Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure* 11:833-843.

Joselin AP, Schulze-Osthoff K, Schwerk C. 2006. Loss of Acinus inhibits oligonucleosomal DNA fragmentation but not chromatin condensation during apoptosis. *J Biol Chem* 281:12475-12484.

Kabsch W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr* 66:125-132.

Keegan RM, Winn MD. 2008. MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr D Biol Crystallogr* 64:119-124.

Krissinel E. 2015. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res* 43:W314-319.

Kurowski MA, Bujnicki JM. 2003. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res* 31:3305-3307.

Law MJ, Lee DS, Lee CS, Anglim PP, Haworth IS, Laird-Offringa IA. 2013. The role of the C-terminal helix of U1A protein in the interaction with U1hpII RNA. *Nucleic Acids Res* 41:7092-7100.

Maris C, Dominguez C, Allain FH. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272:2118-2131.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.

Michelle L, Cloutier A, Toutant J, Shkreta L, Thibault P, Durand M, Garneau D, Gendron D, Lapointe E, Couture S, Le Hir H, Klinck R, Elela SA, Prinos P, Chabot B. 2012. Proteins...
associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol Cell Biol* 32:954-967.

Murachelli AG, Ebert J, Basquin C, Le Hir H, Conti E. 2012. The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat Struct Mol Biol* 19:378-386.

Murshedov GN, Skubak P, Lebedev AA, Pannu NS, Steinhoff RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355-367.

Muto Y, Yokoyama S. 2012. Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems. *Wiley Interdiscip Rev RNA* 3:229-246.

Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmad Y, Pitsch S, Black DL, Allain FH. 2005. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 309:2054-2057.

Pancevac C, Goldstone DC, Ramos A, Taylor IA. 2010. Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors. *Nucleic Acids Res* 38:3119-3132.

Perez Canadillas JM, Varani G. 2003. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J* 22:2821-2830.

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12:1231-1245.

Rigou P, Piddubnyak V, Faye A, Rain JC, Michel L, Calvo F, Poyet JL. 2009. The antiapoptotic protein AAC-11 interacts with and regulates Acinus-mediated DNA fragmentation. *EMBO J* 28:1576-1588.

Rodor J, Pan Q, Blencowe BJ, Eyras E, Caceres JF. 2016. The RNA-binding profile of Acinus, a peripheral component of the exon junction complex, reveals its role in splicing regulation. *RNA* 22:1411-1426.

Sahara S, Aoto M, Eguchi Y, Imamoto N, Yoneda Y, Tsujimoto Y. 1999. Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation. *Nature* 401:168-173.

Schwerk C, Prasad J, Degenhardt K, Erdjument-Bromage H, White E, Tempst P, Kidd VJ, Manley JL, Lahti JM, Reinberg D. 2003. ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol Cell Biol* 23:2981-2990.

Tange TO, Shibuya T, Jurica MS, Moore MJ. 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* 11:1869-1883.

Teplova M, Farazi TA, Tuschi T, Patel DJ. 2016. Structural basis underlying CAC RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS. *Q Rev Biophys* 49:e1.

Tsuda K, Someya T, Kuwasako K, Takahashi M, He F, Unzai S, Inoue M, Harada T, Watanabe S, Terada T, Kobayashi N, Shirouzu M, Kigawa T, Tanaka A, Sugano S, Gunert P, Yokoyama S, Muto Y. 2011. Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. *Nucleic Acids Res* 39:1538-1553.
Volpon L, D'Orso I, Young CR, Frasch AC, Gehring K. 2005. NMR structural study of TcUBP1, a single RRM domain protein from Trypanosoma cruzi: contribution of a beta hairpin to RNA binding. *Biochemistry* 44:3708-3717.

Vucetic Z, Zhang Z, Zhao J, Wang F, Soprano KJ, Soprano DR. 2008. Acinus-S' represses retinoic acid receptor (RAR)-regulated gene expression through interaction with the B domains of RARs. *Mol Cell Biol* 28:2549-2558.

Wang F, Soprano KJ, Soprano DR. 2015. Role of Acinus in regulating retinoic acid-responsive gene pre-mRNA splicing. *J Cell Physiol* 230:791-801.

Wang X, Tanaka Hall TM. 2001. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol* 8:141-145.

Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* 419:182-185.
Figure 1

Schematic overview of three endogenous Acinus isoforms (L, S and S’) with SAP domain in green, RNA recognition motif (RRM) in blue, and RS-like domain in orange.

The different N-termini of Acinus S and S’ isoforms are indicated by amino acid sequence. The length of each isoform is given in brackets. The position of RRM domain in each isoform is given above the blue rectangles.
Figure 2

Overview of the human AcRRM domain (PDB code 6G6S).

(A) Amino-acid sequence of the human AcRRM with the α-helices (blue) and β-strands (grey) annotated above. The RNP2 and RNP1 motifs are indicated by the dashed boxes. (B) The cartoon overview of the secondary structure elements labelled as follows: yellow for RNP1 (on β-strand 3), orange for RNP2 (partially on β-strand 1), red for the C-terminal α-helix; with the electron density views enlarged for the region rich in aromatic residues and the disulfide bond between cysteines 1083 of different chains (most likely a crystal packing artifact). The secondary structure elements are labelled.
Figure 3

Conservation and predicted charge distribution mapped on the surface of the AcRRM domain.

(A) Overview of the human AcRRM domain colored according to the position in the sequence (RNP1 and RNP2 are located in the green and blue colored β-strands, the C-terminal α-helix is colored in red). (B) Amino acid sequence conservation of AcRRM calculated using the ConSurf server (Ashkenazy et al., 2010) and mapped on the solvent excluded surface of the protein. (C) The electrostatic potential of AcRRM calculated using the DelPhi server (Li et al., 2012) and mapped on the solvent accessible surface of the domain. Sequence alignment used for the calculation of ConSurf scores is presented in the Suppl. Fig. 4.
Figure 4

Superposition of the AcRRM structure with two most similar structures:

(A) RRM of Rna15 (PDB code 2x1b (Pancevac et al., 2010)) and (B) RRM of U1A (PDB code 5fj4 (Huang, Wang & Lilley, 2016)). AcRRM is colored as in Figure 2, and other RRMs are colored in light grey and light brown for panel A and B, respectively, with C-terminal features highlighted in darker shades. The caspase cleavage site in AcRRM is indicated by red arrow, and the two sequence insertions in AcRRM are indicated by black arrows.
Figure 5

Comparison of AcRRM with structurally related RRM domains crystallized in complex with RNA.

(A) Sequence alignment of AcRRM and five structurally most similar RRM domains. RNP1 and RNP2 are marked with yellow and orange shading, respectively. The additional fragment, comprising β-strand 4 and flanking regions, that might provide sequence specific contacts with RNA is shown in purple. AcRRM secondary structure elements (L1-5 – loops-1 to 5) and sequence conservation based on the scores calculated with the ConSurf program (Suppl. Fig. 4) (Ashkenazy et al., 2010) are given above the alignment. (B-E) Comparison of the putative RNA binding site of (B) AcRRM with the known RNA complexes of similar RRMs: (C) Tra2-β, (D) Sxl and (E) U1A. AcRRM is colored according to the ConSurf conservation scores, the other RRMs are colored in light grey apart from RNP2 (orange), RNP1 (yellow) and the additional fragment covering β4 (purple). The caspase cleavage site in AcRRM is indicated by red arrow, the cysteine residue forming a disulphide bond in the crystal by green arrow. In panels C-E, only the RNA fragments interacting with RNP1 and RNP2 are shown and the numbers 1-5 correspond to the position in RNP1 and RNP2. The thin black lines indicate the RNA backbone. The complete picture of discussed RRM-RNA interactions is provided in Suppl. Fig. 5.
Figure 6

Superposition of AcRRM with the RRM of U1A protein (PDB code 4c4w (Huang & Lilley, 2014)):

(A) without RNA (B) in complex with RNA (dark brown color), with demonstration of the C-terminal α-helix relocation upon RNA binding. AcRRM is colored as in Figure 2, and U1A in light grey with C-terminal α-helix highlighted in dark grey. The caspase cleavage site in AcRRM is indicated by red arrow (for better clarity only in panel A).
Table 1 (on next page)

Data collection and structure refinement statistics

Values in parenthesis are for the outer resolution shell.

\(^a\) \(\langle l/\sigma (l) \rangle\) is the mean signal-to-noise ratio, where \(l\) is the integrated intensity of a measured reflection and \(\sigma (l)\) is the estimated error in the measurement.

\(^b\) \(R_{\text{meas}} = 100 \times \sum_{hkl} \left\{ \frac{N(hkl)}{[N(hkl) - 1]} \right\}^{1/2} \sum_i l_i(hkl) - \langle l(hkl) \rangle / \sum_{hkl} \sum_i l_i(hkl)\), where \(l_i(hkl)\) is the \(i\)th observed intensity of reflection \(hkl\), \(\langle l(hkl) \rangle\) is the average of symmetry-related observations of reflection \(hkl\) and \(N(hkl)\) is the multiplicity.

\(^c\) \(R_{\text{work}} = 100 \times \sum_{hkl} \| F_{\text{obs}} \| - \| F_{\text{calc}} \| / \sum_{hkl} \| F_{\text{obs}} \|\), where \(F_{\text{obs}}\) and \(F_{\text{calc}}\) are observed and calculated structure-factor amplitudes, respectively.

\(^d\) The \(R_{\text{free}}\) value was calculated as for \(R_{\text{work}}\) using only an unrefined randomly chosen subset of reflection data (5%).
## Data collection

| Parameter                                         | Value         |
|---------------------------------------------------|---------------|
| Diffraction source                                | BESSY 14.2    |
| Wavelength (Å)                                    | 0.918410      |
| Rotation range per image (°)                      | 1             |
| Total rotation range (°)                          | 120           |
| Crystal-to-detector distance (mm)                 | 150           |
| Space group                                       | $P 2_1 2_1 2_1$ |
| Unit cell parameters (Å)                          | 30.4 67.9 80.1 |
| Mosaicity (°)                                     | 0.32          |
| Resolution range (Å)                              | 80.7-1.65 (1.74-1.65) |
| Excluded resolution ranges (Å)                    | 2.29-2.22, 1.94-1.90 |
| Total No. of reflections                          | 90411         |
| No. of unique reflections                         | 19003         |
| Completeness (%)                                  | 91.6 (98.5)   |
| Multiplicity                                      | 4.76 (4.80)   |
| <$I/\sigma(I)>$ a                                 | 16.3 (1.97)   |
| $R_{\text{meas}}$ (%) b                          | 7.1 (86.4)    |

## Refinement

| Program                                           | REFMAC 5.8.0189 |
|---------------------------------------------------|-----------------|
| Overall mean B-factor (Å$^2$) for protein chains  | 16.0, 15.2      |
| Overall mean B-factor (Å$^2$) for solvent          | 30.2            |
| $R_{\text{work}}$ (%) c                           | 18.5            |
| $R_{\text{free}}$ (%) d                           | 22.4            |
| Rms deviations from ideal                         |                 |
| Bond lengths (Å)                                  | 0.017           |
| Bond angles (°)                                   | 1.8             |
| Residues in Ramachandran plot (%)                 |                 |
| Favored                                           | 99              |
| Outliers                                          | 0               |