Cell PRESS
Structure
Previews

# Repeat Recognition

Konstantinos Tripsianes[1,2] and Michael Sattler[1,2,*]
[1]Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
[2]Munich Center for Integrated Protein Science at Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany
*Correspondence: sattler@helmholtz-muenchen.de
DOI 10.1016/j.str.2010.09.006

**CUG-binding protein 1 (CUGBP1) is a ubiquitous RNA-binding protein implicated in altered RNA metabolism linked to myotonic dystrophy type 1. Crystal structures of the RRM domains in complex with cognate RNAs (Teplova et al., 2010) reveal molecular details for the selectivity of CUGBP1 toward GU-rich mRNA elements.**

Posttranscriptional regulation of gene expression involves various aspects, such as mRNA processing (splicing, capping and polyadenylation), mRNA export, and mRNA decay. These processes depend on the binding of multidomain RNA binding proteins or noncoding RNAs to *cis*-acting RNA sequences found adjacent to intron/exon splice sites and in 3′ and 5′ untranslated regions (UTRs) of eukaryotic pre-mRNA transcripts. Sequence variations in the regulatory elements can significantly alter posttranscriptional regulation of gene expression and are thus linked to human diseases.

The turnover of eukaryotic mRNAs is mostly modulated by conserved sequence elements located in the 3′-UTR. A well-characterized example is the AU-rich element (ARE), which is implicated in the regulation of mRNA half-life (Wilusz et al., 2001). The GU-rich regulatory element (GRE) represents a recently discovered sequence motif found in the 3′-UTR of certain mRNAs. The GRE consensus sequence UGUUUGUUUGU has been shown to mediate rapid decay of a number of GRE-containing transcripts in primary human T cells (Vlasova et al., 2008).

A key factor that recognizes GRE motifs and promotes mRNA degradation is the CUG-binding protein 1 (CUGBP1). This protein has been shown to associate with abnormally extended CUG repeats that occur in patients with myotonic dystrophy type 1 (Timchenko et al., 1996). CUGBP1 belongs to the CELF (CUGBP and ETR-3-like factors) family of proteins that have additional functions in modulating pre-mRNA splicing, enhancing translation, and triggering mRNA deadenylation (Barreau et al., 2006). CELF family members share a highly similar domain organization, comprising three RNA recognition motifs (RRM). RRM1 and RRM2 are adjacent to each other and separated by a long nonconserved linker from the C-terminal RRM3 (Figure 1A).

In this issue of *Structure*, Teplova et al. (2010) report on the crystal structures of RRM1 and the tandem RRM1/2 domains of CUGBP1 bound to GRE RNAs. In all structures, a single UGUU or UGUG repeat within the GRE is recognized by either RRM1 or RRM2. Both domains adopt the canonical RRM fold composed of a four-stranded β sheet, representing the typical RNA binding surface, flanked by two α helices (Figure 1B). The recognition of the UGUU/G motifs by the RRMs is almost identical, owing to conserved amino acids at key positions of the β sheet that are employed for RNA binding (Figure 1C). The bases of the third and fourth positions of the tetranucleotide motif (UU or UG) stack on two conserved phenylalanine rings of the RNP (ribonucleoprotein) sequence motifs of the RRM domains (Clery et al., 2008). A number of direct or water-mediated intermolecular hydrogen bonds contribute to the recognition of all four nucleotides. This hydrogen bond network is denser for the base edges of the central GU step, implying their importance for binding specificity (Teplova et al., 2010).

The authors note that the conformation of the first pair of nucleotides (UG) has structural similarity with left-handed Z-RNA. Such RNA conformations are found with alternating CG base pairs, and are characterized by a *syn* conformation of the guanine bases as a hallmark of left-handed helices (Popenda et al., 2004). A similar conformation is found for the CUGBP1-bound RNA, where the guanine in the second position of the UGU(U/G) RNA motifs displays a *syn* glycosidic torsion angle. This observation may explain the selectivity of the CUGBP1 RRMs toward GU-rich elements. Mutations targeting the guanine residue further support its role for the specificity of the interaction (Teplova et al., 2010).

Interestingly, the RNA recognition by RRM1 and RRM2 is distinct from the one reported for the RRM3 domain of CUGBP1 (Tsuda et al., 2009). RRM3 contains a conserved N-terminal extension that interacts with the β sheet and is important for RNA binding. The structure of RRM3 with $(UG)_3$ RNA revealed that the complex is formed via extensive stacking interactions, but none of the guanine residues was found in a *syn* conformation.

Teplova et al. (2010) report that the individual RRM1 and RRM2 domains bind with comparable affinities to UUGUU substrates. In contrast, the tandem RRM1/2 domains show increased affinity for an RNA comprising two sequential UUGUU motifs, thus indicating substantial binding cooperativity. Indeed, NMR studies confirm that in solution both RRMs interact with a tandem repeat RNA, containing two consecutive UGUU elements separated by one or two uracil residues. NMR relaxation data indicate that the two RRM domains tumble independently in the absence of RNA. However, the complex comprising the tandem RRM domains and the tandem RNA repeat is compact and rigid, consistent with the observed binding cooperativity (Figure 1D). Different structural and dynamical arrangements of tandem RRM domains have been observed (Lunde et al., 2007). In many cases, significant domain contacts stabilize a unique

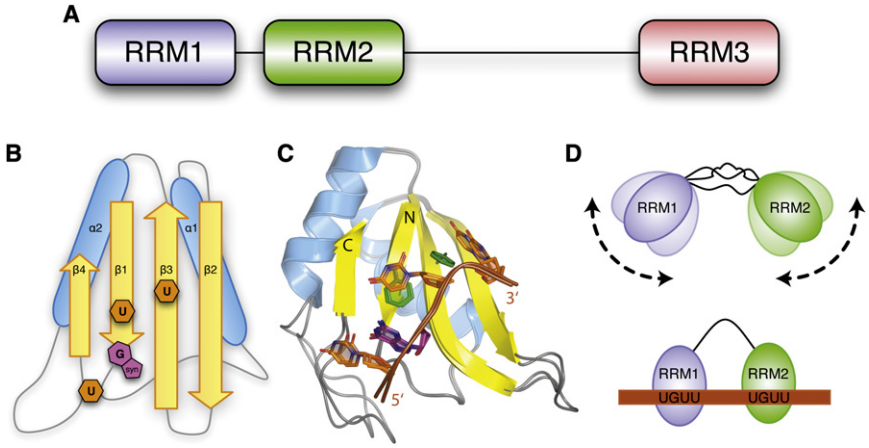


**Figure 1. Binding of CUGBP1 RRM1 and RRM2 Domains to UGUU RNA Motifs**
(A) Domain organization of the CUGBP1 protein.
(B) Structural topology of the CUGBP1 RRM1 and RRM2 domains. Strands are shown as yellow arrows and helices as cyan cylinders. The approximate positions of the RNA bases recognized are indicated schematically.
(C) Superposition of the CUGBP1 RRM1 and RRM2 crystal structures in complex with UGUU tetranucleotides. The conserved RNP1 and RNP2 phenylalanine residues that participate in stacking interactions with the RNA are shown as green sticks. Nucleotides are depicted as sticks as well, and colored orange for uracil residues and magenta for the *syn*-guanine residue. N and C termini of the RRMs and the 5′ and 3′ ends of the RNAs are indicated.
(D) Schematic representation of the domain arrangement of the tandem RRMs of CUGBP1. In the free form, the two domains tumble independently (top). Upon binding to a tandem UGUU repeat, a more compact domain arrangement is induced (bottom).

domain arrangement. In other cases weak or no interactions between the two domains are observed in the absence of the RNA ligand, and a stable conformation is only induced upon ligand binding. In such cases, crystallization may interfere with native domain arrangements that exist in solution (Simon et al., 2010). It is therefore important and advisable to combine the crystallographic data with solution techniques such as NMR spectroscopy, as has been done by Teplova et al. in the present work.

The structural analyses of the CUGBP1 RRM-RNA interactions (Teplova et al., 2010; Tsuda et al., 2009) set the stage for understanding the molecular functions of CELF family members. Yet several challenges lie ahead to reveal mechanistic details of the role of CELF proteins in GRE-mediated mRNA decay and alternative splicing. It will be important to determine if all RRMs are required for the diverse functions or whether perhaps the distinct RNA binding selectivities of RRM1/RRM2 and RRM3 contribute independently to different processes. It will also be interesting to identify other factors that may associate with GU-rich sequences, and if they also recognize an unusual conformation of the bound RNA, such as the *syn* guanine observed in the CUGBP1-RNA interaction. Finally, it will be critical to characterize the functional interplay between CUGBP1 and other splicing regulators that contribute to aberrant mRNA splicing observed in myotonic dystrophy 1.

## REFERENCES

Barreau, C., Paillard, L., Méreau, A., and Osborne, H.B. (2006). Biochimie. *88*, 515–525.

Clery, A., Blatter, M., and Allain, F.H. (2008). Curr. Opin. Struct. Biol. *18*, 290–298.

Lunde, B.M., Moore, C., and Varani, G. (2007). Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Popenda, M., Milecki, J., and Adamiak, R.W. (2004). Nucleic Acids Res. *32*, 4044–4054.

Simon, B., Madl, T., Mackereth, C.D., Nilges, M., and Sattler, M. (2010). Angew. Chem. Int. Ed Engl. *49*, 1967–1970.

Teplova, M., Song, J., Gaw, H.Y., Teplov, A., and Patel, D.J. (2010). Structure *18*, this issue, 1364–1377.

Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L., Roberts, R., Caskey, C.T., and Swanson, M.S. (1996). Nucleic Acids Res. *24*, 4407–4414.

Tsuda, K., Kuwasako, K., Takahashi, M., Someya, T., Inoue, M., Terada, T., Kobayashi, N., Shirouzu, M., Kigawa, T., Tanaka, A., et al. (2009). Nucleic Acids Res. *37*, 5151–5166.

Vlasova, I.A., Tahoe, N.M., Fan, D., Larsson, O., Rattenbacher, B., Sternjohn, J.R., Vasdewani, J., Karypis, G., Reilly, C.S., Bitterman, P.B., and Bohjanen, P.R. (2008). Mol. Cell *29*, 263–270.

Wilusz, C.J., Wormington, M., and Peltz, S.W. (2001). Nat. Rev. Mol. Cell Biol. *2*, 237–246.