The solution structure of Dead End bound to AU-rich RNA reveals an unprecedented mode of tandem RRM-RNA recognition required for mRNA regulation

Authors

Malgorzata M. Duszczyk[1*], Harry Wischnewski[2], Tamara Kazeeva[1], Fionna E. Loughlin[1,5], Christine von Schroetter[1,3], Ugo Pradère[4], Jonathan Hall[4], Constance Ciaudo[2*], Frédéric H.-T. Allain[1,3,6*]

Affiliations

[1]Institute of Molecular Biology and Biophysics, [2]Institute of Molecular Health Sciences, [3]Institute of Biochemistry, Department of Biology, [4]Institute of Pharmaceutical Sciences, Department of Chemistry and Applied Biosciences, ETH Zürich, 8093, Zürich, Switzerland [5]Present Address: Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, 3800 VIC, Australia [6]Lead Contact

Contact Information

*To whom correspondence should be addressed.
E-mail: m.m.duszczyk@gmail.com (M.M.D.), cciaudo@ethz.ch (C.C.), allain@bc.biol.ethz.ch (F.H.-T.A.)

Abstract

Dead End (DND1) is an RNA-binding protein essential for germline development through its role in post-transcriptional gene regulation. The molecular mechanisms behind selection and regulation of its targets are unknown. Here, we present the solution structure of DND1's tandem RNA Recognition Motifs (RRMs) bound to AU-rich RNA. The structure reveals how an NYAYUNN element is specifically recognized, reconciling seemingly contradictory sequence motifs discovered in recent genome-wide studies. RRM1 acts as a main binding platform, including atypical extensions to the canonical RRM fold. RRM2 acts cooperatively with RRM1, capping the RNA using an unusual binding pocket, leading to an unprecedented mode of tandem RRM-RNA recognition. We show that the consensus motif is sufficient to mediate upregulation of a reporter gene in human cells and that this process depends not only on RNA binding by the RRMs, but also on DND1's double-stranded RNA binding domain (dsRBD), which is dispensable for target binding *in cellulo*. Our results point to a model where DND1 target selection is mediated by a non-canonical mode of AU-rich RNA recognition by the tandem RRMs and a role for the dsRBD in the recruitment of effector complexes responsible for target regulation.

Introduction

Post-transcriptional gene regulation (PTGR) is orchestrated by an interplay between mRNA sequence and structure, and their dynamic interactions with RNA binding proteins (RBPs). RBPs instantly cover mRNA transcripts as they are transcribed and are essential for all aspects of RNA metabolism like maturation, transport, cellular localization and turnover. Differential gene expression patterns depend on tissue-specific RBP levels and their combined interactions with the transcriptome. Misregulation of this proces, due to mutations in the RBPs or RNA to which they bind, is at the origin of a plethora of genetic diseases (Corbett 2018). Understanding how RBPs specifically recognize their mRNA targets and how this is connected to their downstream fate is therefore crucial to understand the complex PTGR networks involved in health and disease.

The germline is composed of highly specialized cells which must preserve a totipotent genome through generations. It succeeds in this through a highly specialized RNA metabolism regulating a complex transcriptome (Ramsköld et al. 2009; Seydoux and Braun 2006). Dead End (DND1) is a germline-specific RBP. Conserved in vertebrates, it is essential for the development and migration of primordial germ cells (PGCs), pluripotent germline precursors, to the future reproductive organs. These processes occur early in embryogenesis by blocking the expression of somatic genes, controlled by extensive post-transcriptional regulation (Licatalosi 2016; Seydoux and Braun 2006). DND1 deletion causes loss of PGCs by reactivation of somatic gene expression patterns in zebrafish (Gross-Thebing et al. 2017; Weidinger et al. 2003). In mice, truncations of DND1 (the so-called 'Ter-mutation') lead to male sterility and the formation of testicular teratomas (Ruthig et al. 2019; Youngren et al. 2005).

DND1 performs a dual role in regulating gene expression: stabilizing some mRNA transcripts while destabilizing others. First, a direct role for DND1 in promoting RNA stability by inhibiting the action of miRNAs was shown for specific candidate mRNAs (p27/CDKN1B and LATS2 in a human tumor cell line and Nanos/TDRD7 in zebrafish embryos) (Kedde et al. 2007). DND1 binding to conserved U-rich regions close to miRNA seed sequences in the 3' untranslated regions (3'UTRs) of these targets potentially blocked their accessibility to the miRNA-induced silencing complex (miRISC) which rescued translation. Second, on a transcriptome-wide level, DND1 was shown to have an opposite effect and a wider role in germ cell PTGR by destabilizing a set of transcripts that must be cleared from developing

germ cells to ensure survival, through non-miRISC mediated recruitment of the CCR4-NOT deadenylase complex (Suzuki et al. 2016; Yamaji et al. 2017).

Transcriptome sequencing of the Ter mouse line, possessing a truncation in the DND1 gene, prior to the formation of teratomas, showed two groups of DND1 targets either up- or down-regulated compared to wild-type, involved in pluripotency and in differentiation respectively, in a sequential fashion, suggesting a dual and biphasic role for DND1. No enrichment for RNAs harboring a signature for miRNA-regulation was found in the wild-type versus Ter cell line, although previously described DND1 targets were re-identified (Ruthig et al. 2019).

Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) assays revealed that targets crosslinked to DND1 are enriched in a UUU/UUA triplet and are associated with apoptosis, inflammation and signaling pathways (Yamaji et al. 2017). An additional digestion optimized-RIP-seq (DO-RIP-seq) approach described [A/G/U]AU[C/G/U]A[A/U] as RNA sequence motif enriched in DND1 targets (Ruthig et al. 2019). Overall, these functional and genome-wide motif identification studies using high-throughput methods are partly contradictory. Thus, the molecular mechanisms of DND1 function and details on how DND1 achieves target specificity remain elusive.

Most RNA binding proteins exert their specific functions by combining several copies of RNA binding domains to increase specificity and affinity to their targets (Afroz et al. 2015; Singh and Valcárcel 2005). DND1 has a unique domain structure, carrying two RNA recognition motifs (RRMs) connected by a remarkably short, four-residue inter-domain linker preceding a double-stranded RNA binding domain (dsRBD) (Fig. 1A). This combination of domains is rare among RBPs and is only shared by DND1 and two other members of the hnRNPR-like subfamily (Marchler-Bauer et al. 2017) (Figs. S1A, S2). RRMs are the most abundant RNA binding domains in higher vertebrates, binding primarily single stranded RNA (Cléry et al. 2008; Gerstberger et al. 2014). Their conserved canonical fold is $\beta\alpha\beta\beta\alpha\beta$ – a four-stranded anti-parallel β-sheet packed on top of two alpha-helices. Their canonical mode of RNA binding involves stacking of two consecutive RNA bases by exposed aromatic residues from the two conserved consensus sequences (RNP1 & RNP2) in the first and third β-strands (Fig. S1B). The RRM uses several strategies to achieve specificity and affinity for its targets, namely by using extensions to the canonical fold with structural elements outside the β-sheet for RNA recognition and by employing several copies

of the domain. While RRM1 appears to be a canonical RRM, RRM2 on the contrary does not have any aromatic residues in RNP1 and RNP2 (Figs. 1A, S1B, S2). Although several structures of tandem RRM-RNA complexes have been determined (Afroz et al. 2014; Deo et al. 1999; Kanaar et al. 1995; Lukavsky et al. 2013; Teplova et al. 2010; Wang and Tanaka Hall 2001), the great majority of them contains combinations of two canonical RRMs. It is therefore intriguing to understand if and how the tandem RRMs of DND1 cooperate to specifically recognize their RNA targets, and if the dsRBD further influences RNA binding.

To address the question of how DND1 recognizes and represses its cellular targets at the molecular level, we first set out to understand the contribution of the three RNA binding domains of DND1 to target recognition. This demonstrated a crucial role for the RRMs, while removal of the dsRBD had no effect on target binding. We then determined the solution structure of the DND1 tandem RRMs in complex with an AU-rich RNA. Our structure reveals binding to seven nucleotides with limited sequence-specific interactions generating recognition of an NYAYUNN motif. Both RRMs participate in this RNA recognition via an unprecedented mode of cooperative binding. A canonical mode of binding by RRM1 is extended by additional secondary structure elements and contacts to RRM2 originating from α-helix 2, which is unique amongst known RRM-RNA interactions. Our structure explains the degeneracies in the motifs enriched in DND1 targets in recent genome-wide studies. Finally, we show, using luciferase-based assays, that while delivery of a miRNA-mimic represses a reporter gene, co-expression of DND1 increases translation of the reporter independently of miRNA-mimic delivery. We demonstrate that a single AU-rich element is necessary and sufficient for this effect, and that mutations in the RNA binding surface of DND1's tandem RRMs undo it. Surprisingly, also removal of the dsRBD returns translation to its original state, suggesting that the dsRBD is an essential effector domain, which likely recruits mRNA regulating factors. Immunoprecipitation followed by liquid chromatography-tandem mass spectrometry analysis of the interactome of DND1 targets in HEK293T reveals an enrichment for proteins involved in the stabilization of mRNAs like ELAVL1 and proteins localized in RNP granules. These results provide the first mechanistic and structural insights into the molecular mechanisms by which DND1 regulates a subset of mRNAs, thereby stabilizing the fate of germ cells.

Results

DND1 binds CLIP/RIP targets *in cellulo* mainly through its RRM1

There is some ambiguity in the published RNA motifs targeted by DND1. It was first reported on the basis of data from reporter assays, that DND1 binds to U-rich regions of approximately 12 nucleotides in length (UUUUUCCUUAUUU and UUUUUACCUUUU) in the vicinity of miRNA seed sequences in the 3'UTR of the CDKN1B/p27 tumor suppressor mRNA (Kedde et al. 2007). Later, genome-wide PAR-CLIP studies defined a much shorter UUU/UUA triplet motif (Yamaji et al. 2017) and very recently a Digestion Optimized-RIP approach revealed [A/G/U]AU[C/G/U]A[A/U] as motif enriched in DND1 targets (Ruthig et al. 2019). A single RRM usually binds 4-6 nucleotides (Cléry et al. 2008). To understand how the two RRMs of DND1 cooperate with the dsRBD to engage an RNA target, we set out to define the contributions of each domain to RNA binding. We first validated published DND1 targets using RNA immunoprecipitation (RIP) from HEK293T cells transiently expressing either FLAG-tagged DND1 wild-type (WT) or mutants thereof. Mutant 1-235 lacks the dsRBD but includes the extended RRM2, making it longer than the Ter-mutation truncation (Youngren et al. 2005), which is located in the middle of RRM2 α-helix 2 (Fig. 1A); R98A contains a mutation of the conserved RNP1 sequence in RRM1 (Figs. 1A, S1B). The RIP was followed by quantitative PCR using primers for two DND1 targets revealed by PAR-CLIP and DO-RIP (Fig. 1B). Phorbol-12-Myristate-13-Acetate-Induced Protein 1 (PMAIP1), a cell cycle regulator promoting apoptosis, is the target with the highest number of normalized 3'UTR read counts in the PAR-CLIP dataset (Yamaji et al. 2017). Survival of Motor Neuron (SMN) protein is a small nuclear ribonucleoprotein (snRNP) assembly factor. Its pre-mRNA SMN1 is expressed at an order of magnitude lower level than PMAIP1 in HEK293 and is enriched in the DO-RIP dataset (Ruthig et al. 2019). As a negative control, we used solute carrier family 25 member 6 (SLC25A6), the mRNA of which is expressed at similar levels as PMAIP1 in HEK293T cells. In our RIP assays using both the full-length WT DND1 and a truncation mutant lacking the dsRBD, the enrichment over the input is higher for the two targets compared to the control (Figs. 1B, S3B). RIP using the full-length R98A mutant results in a level of target enrichment comparable to the negative control pulled down by the WT DND1. These data suggest that RRM1 is an essential element for RNA binding in cells, while the dsRBD appears to be dispensable for RNA binding, at least for these two abundantly expressed targets. The role of RRM2 could not be tested easily, as its RNA interaction surface could not be predicted considering the non-

canonical nature of this RRM. Therefore, we next investigated the contribution of the individual RRMs to RNA binding *in vitro*.

DND1's tandem RRMs cooperatively bind AU-rich RNA with high affinity

To identify high affinity RNA targets for the individual RRMs, we recombinantly expressed and purified the separate RRM domains, including their conserved N- and C-terminal extensions (Figs. 1A, S2). We then tested their binding to short oligonucleotides derived from the p27 target (Kedde et al. 2007) using isothermal titration calorimetry (ITC) to monitor thermodynamic changes upon ligand binding (Fig. 2). In addition, we used NMR titrations - series of two-dimensional ^{1}H-^{15}N HSQC spectra of protein with increasing RNA concentrations - to observe perturbations of backbone amide proton chemical shifts caused by RNA binding (Fig. 2B and S4). We found that RRM2 alone does not bind any of the oligonucleotides (Fig. S4A) and that only an oligonucleotide bearing a central adenosine (UUAUUU) has sufficient affinity to RRM1 to yield significant chemical shift perturbations in an NMR titration (Fig. S4B). The binding affinity of a slightly longer oligonucleotide CUUAUUUG to RRM1 is higher than 54 μM as measured by ITC (Fig. 2C). Surprisingly, the affinity for this sequence increased over 80-fold (0.68 μM K_D) when titrated with the tandem RRM domains showing the importance of RRM2 for RNA binding (Fig. 2A). NMR titration of UUAUUU to DND1 RRM12 confirmed the increased affinity with a change from fast to intermediate exchange on the chemical shift timescale, and the saturation of most residues at a 1:1 ratio (Fig. 2B). Large chemical shift changes throughout the ^{1}H-^{15}N HSQC spectrum indicated major structural rearrangements of the tandem RRMs upon RNA binding. Additional ITC and NMR titrations (Fig. S4) showed that only the tandem RRMs, not the single domains, have affinity to U-rich oligonucleotides devoid of adenosines, or with an adenosine at the periphery such as UUUUUAC (>35 μM K_D) and UUUUUCC (>100 μM K_D). The *in vitro* binding measurements are summarised in Table S2. Taken together the results above indicated that the DND1 tandem RRMs cooperatively bind AU-rich RNA targets of approximately 7-8 nucleotides in length, with the highest affinity to oligonucleotides comprising an adenosine in a central position.

The solution structure of DND1's tandem RRMs bound to CUUAUUUG RNA

In an effort to understand the mechanism of cooperative binding of the tandem RRMs of DND1, as well as their preference for AU-rich over U-rich RNAs, we solved the structure of the extended RRM12 in complex with the CUUAUUUG RNA target using NMR spectroscopy (Table 1, Fig. 3). The assignment procedure of protein and RNA and the structure calculation protocol is outlined in the Methods section. DND1 RRM12 showed poor solubility and in the final stages of purification was purified as a complex with the target RNA. Use of chemically-synthesized CUUAUUUG oligonucleotides, which were selectively labelled with ^{13}C–ribose units, was key to resolving the spectral overlap of critical residues, and aided the assignments of resonances and intermolecular NOEs (Fig. S5A). Hence, we could calculate a precise structural ensemble of this 27.5 kDa protein-RNA complex using unambiguous intra-protein (4947), intra-RNA (150) and intermolecular (103, Table S3) NOE-derived distance restraints. The isolated domains within this ensemble of 20 conformers converge to a root mean square deviation (RMSD) of 0.95 Å (RRM1), 0.57 Å (RRM2) and the $U_3A_4U_5U_6U_7$ RNA nucleotides to an RMSD of 0.6 Å for all heavy atoms (Figs. 3AB, Table 1). The global complex was initially less well-defined due to a lack of direct contacts between RRM1 and RRM2 and a limited number of restraints that could orient the domains through the RNA. We expected a well-defined orientation between RRM1 and RRM2 in the complex though, as ^{15}N T1 and T2 relaxation measurements indicated that the protein-RNA complex behaves as a rigid globular protein of approximately 30 kDa in solution, as opposed to two domains tumbling independently (Fig. S5B). We included 127 Residual Dipolar Couplings (RDCs) from amide bonds to increase precision and finally obtained a final ensemble of 20 conformers with an RMSD of 1.26 Å for all heavy atoms (Fig. 3C, Table 1).

Our structure reveals a number of unusual structural features, including unprecedented modes of RRM:RNA recognition. The fold of RRM1 is non-canonical with the conserved N-terminal extension folding into a β-hairpin (β-1,0) packed on top of an N-terminal α-helix (α0). This structural element is tightly packed against the side of the RRM1 to create an extended surface (Fig. 3A), in an identical fashion as the extension of RRM1 in the RNA-binding protein hnRNPQ/SYNCRIP (Hobor et al. 2018). This extended RRM (eRRM) fold is conserved in all members of the hnRNPR-like family of RNA binding proteins (Figs. S1, S6). RRM2 preceeds the conserved C-terminal helical extension (α3) lacking

contacts to the core canonical RRM2 fold (Fig. 3B), as confirmed by the relaxation data which showed this helix tumbling independently in the tandem RRM-RNA complex (Fig. S5B). The RNA is bound in a canonical 5' to 3' fashion over β4 through to β2 using the RNP residues of RRM1, but is sandwiched between the two domains (Fig. 3D), burying the central UAUUU nucleotides in a positively charged channel (Fig. 3E). In addition to the primary RNA binding surface on RRM1, the conserved N-terminal hairpin extension is used to extend RNA binding for a longer sequence compared to canonical RRMs. Finally, and most surprisingly, this RNA binding is stabilized by RRM2 using an unprecedented binding pocket formed by RRM2 α-helix 2 and the edge of its β4, while the non-canonical β-sheet of RRM2, missing conserved aromatic RNP residues, is turned away from the RNA. Additional protein-RNA contacts are facilitated by the backbone and sidechains of the four-residue interdomain linker. This structure explains well the increase in RNA binding affinity of the tandem RRMs compared to RRM1 in isolation.

Structural details: specific readout by DND1's tandem RRMs

DND1's tandem RRMs directly bind primarily the central 5 nucleotides (bold) of CU**UAUUU**G RNA (Fig. 4). The intermolecular contacts are represented in a schematic fashion in Figure 4A. U_3 and A_4 are bound in a canonical fashion over the RRM1 β-sheet surface, their bases interacting with β4 and β1 respectively. The U_3 O2 is hydrogen-bonded to the C132 sidechain and its ribose contacts L130. This contact is probably not fully specific, since a cytidine would also be accepted in this position keeping this H-bond intact. The A_4 base is stacked on F61 in RNP2 and its ribose interacts with F100 from RNP1 in a canonical manner (Fig. 4B). This is suggested by the unusual chemical shifts of the carbons and protons of the A_4 ribose (Fig. S5A). The A_4 base is sequence-specifically recognized via hydrogen-bonds to its Hoogsteen edge (N6 amino and N7) from sidechains and main-chains of the highly conserved interdomain linker (R133, S134 and T135) (Figs. 4B, C). We observed variability in the H-bonding partners of A_4 in the structural ensemble, which reflects the exchange-broadening observed for the backbone resonances of these linker residues.

From U_5 onwards, the protein-RNA interactions deviate from canonical binding. While in a canonical RRM the U_5 base would be expected to stack onto RNP2 Y102, here U_5 rather stacks on A_4, as evidenced by the strong NOEs between A_4 H1' and U_5 H5 and H6 resonances, as well as weaker NOEs between A_4 H8 and the U_5 H6 resonance (Table S3). The ribose of U_5 is packed against Y102,

F100 and M90 (Fig. 4C) and its base functional groups are involved in hydrogen-bonding with the conserved interdomain linker backbone and sidechains (e.g. the S134 hydroxyl, T135 backbone amide and K137 NH3, Figs. 4B, D) and with the Y102 hydroxyl. Although some part of this H-bonding network would be lost, e.g. the contact to K137, a cytidine might also be accepted in this position, using an alternative hydrogen-bonding partner. Therefore, this contact is likely not fully sequence specific.

The most surprising element of the structure was the interaction between a highly conserved binding pocket of RRM2 and the specific recognition of U_6. U_6 lies under RRM2 α-helix 2 with its ribose ring in contact with A193, M194 and K197 and its base carbonyl O4 hydrogen-bonded by both the NH_3 of K196 and the sidechain HE1 of W215 (Fig. 4D). This double H-bond to O4 renders this contact fully sequence specific. Contacts to the U_6 and U_7 phosphate groups by H189 and K197 sidechains, respectively, further stabilize the U_6-RRM2 interaction, as evidenced by a large number of intermolecular NOEs (Fig. S5A). These interactions with RRM2 α-helix 2, holding the U_6 nucleotide like a claw, allow for an unusual reversal of the RNA backbone direction at U_6, which is rotated by 120 degrees around a vertical axis compared to U_5 (see Fig. 4D). The contacts to the U_6 phosphate group by R88 and Q39 of RRM1 (Fig. 4E) help position the two RRMs relative to each other, which explains their cooperative binding since both RRMs contribute to U_6 binding.

The U_7 phosphate is fixed by a salt bridge to K197 on RRM2, while the sidechain NH_2 of N37 on the tip of the N-terminal eRRM1 hairpin extension interacts with the G_8 phosphate (Fig. 4D). The U_7 base is not sequence-specifically recognized. Finally, the G_8 nucleotide is not well defined. Overall, in this conformation, the phosphate groups from U_3 to G_8 are hydrogen-bonded to one or two protein side-chains, with some originating from the RRM1 extension (N37 on β-strand -1 and Q39 on β-strand 0). Altogether this structure suggests that the recognition of a $N_2Y_3A_4Y_5U_6N_7N_8$ consensus sequence (Y is a pyrimidine) involves the tandem RRMs of DND1 as cytidines in the positions of U_3 and U_5 were accepted without significant disruption of the hydrogen-bonding network.

The binding topology on the eRRM1 and RRM2 is unprecedented with the involvement of a nucleotide binding pocket in α-helix 2 of an RRM. As the β-hairpin extension and RNA binding residues on eRRM1 are conserved in the hnRNPR-like family of RNA binding proteins (Figs. S1, S6),

it seems likely that the RNA binding mode of the eRRM1 is conserved. Although the interdomain linker, comprising amino acid residues TEK, is unique for DND1 within this family, it is possible that the linkers of the other family members might bind RNA in a similar fashion but using an alternative hydrogen-bonding network. Notably, DND1 is the only family member with a non-canonical β-sheet binding surface for RRM2, lacking aromatic residues.

Mutation of key RNA binding residues compromises the RNA interaction *in vitro*

To further understand the structure, we replaced several RNA interacting residues of the tandem RRMs of DND1 with alanine. We then tested the effect of these substitutions on RNA binding by NMR titrations, after confirming that the protein fold was not affected (Fig. S7A). The mutant of a canonical RRM1 RNP residue R98A failed to bind CUUAUUUG (Fig. S7B), confirming *in vitro* that the eRRM1 is the primary RNA binding surface of DND1, consistent with our RNA-IP and qPCR experiments (Figs. 1B, S3). Of note, other RRM1 RNP mutants could not be tested. Although we could express mutant F61A, it precipitated during the purification. Y102A could not be expressed in the soluble fraction at all, despite the fact that its equivalents have previously been used in several studies as RNA binding mutants (Goudarzi et al. 2012; Kedde et al. 2007; Mei et al. 2013; Slanchev et al. 2009). Although its solubility *in cellulo* and *in vivo* might be improved, its inactivity could result from an overall instability of the protein. Mutation of M90 on β2 of eRRM1, a residue interacting with U_5, also abolishes RNA binding (Fig. S7B). Although the linker T135 sidechain makes several specific contacts with A_4 or U_5, mutation of this residue to A does not reduce RNA binding. Most likely, other residues in the linker could compensate using an alternative hydrogen-bonding network for binding to U_5. K197A and W215F mutations of residues of the unusual RNA binding pocket of RRM2, resulted in weaker binding to RNA. Smaller chemical shift changes were seen in W215F compared to the WT upon RNA titration. In the K197A mutant, the NMR signals of the complete RRM2 α-helix 2 were exchange-broadened in the RNA-bound state, indicating a less stable RNA binding. We performed ITC experiments (Fig. S7C) to quantify the loss of RNA binding and obtained Kd values of 2.6 μM for K197A and >10 μM for W215F, compared to 0.7 μM for WT DND1 RRM12 binding to CUUAUUUG (Fig. 2A). These were significantly less than the Kd for RRM1 alone (>54 μM, Fig. 2C), suggesting that the single mutations do not fully destroy the RRM2 binding pocket. These assays confirmed that

the eRRM1 RNP is essential for RNA binding and that the novel RRM2 binding pocket stabilizes the primary eRRM1-RNA interaction.

Introduction of an AU-rich motif into a reporter gene 3'UTR is necessary and sufficient for target repression by DND1

To investigate how the reduced RNA binding caused by these mutations affects the function of DND1 *in cellulo*, we tested these single amino-acid mutants in the context of the full-length protein in a luciferase-based gene reporter assay. Kedde et al. used such assays to show that DND1 protects certain mRNA targets from miRNA-mediated repression (Kedde et al. 2007). Thus, we transfected a psiCHECK2 dual luciferase plasmid bearing a partial p27/CDKN1B 3'UTR sequence into HEK293T cells. The partial UTR, cloned downstream from the Renilla luciferase ORF, contains two target sequences for miR-221 and two putative DND1 binding sites. This stretch of sequence was identical to that shown by Kedde et al. (Kedde et al. 2007) to be sufficient for protection of the target by DND1. Co-transfection of the reporter plasmid with a miR-221-3p mimic (miRIDIAN, Dharmacon) produced the expected decrease of luciferase activity compared to that observed with co-transfection of a negative control scrambled miRNA mimic (Fig. S8). Co-transfection of a plasmid expressing FLAG-tagged wild-type DND1 with the negative control miRNA increased luciferase expression from the reporter, demonstrating the functional effect of DND1 on this target. In contrast, FLAG-tagged wild-type DND1 showed only a mild, non-statistically significant effect on the repression of luciferase by miR-221. We concluded that DND1 was unable to counter the suppression of p27/CDKN1B 3'UTR by miR-221, at least under the conditions tested here. The effect of the R98A and 1-235 dsRBD truncation mutants did not differ significantly from that of WT DND1. We reasoned that possibly DND1 was not able to displace the miRNA that was present in abundance from the beginning of the experiment and possibly tightly bound to the target mRNA prior to a sufficiently high expression of the protein. The data suggested that this assay system was not suited to investigate the functional role of DND1 RNA binding on miRNA regulation. Therefore, we focused efforts on the inductive effect of DND1 alone. We inserted the full 3'UTR from the telomerase reversed transcriptase gene (TERT) into the psiCHECK2 dual luciferase plasmid. This mRNA was not expected to be targeted by DND1 as it lacks AU-rich regions and is not present in recent CLIP and RNA-IP DND1 target datasets (Ruthig et al. 2019; Yamaji et al. 2017). Upon transfection of this reporter into HEK293T cells together with expression plasmids for either wild-type or mutant full-length FLAG-tagged DND1, we do not observe any effect on luciferase activity (Fig. 5). Yet,

insertion of a single UAUUU (the central pentanucleotide bound to the tandem RRMs in our structure) into the TERT 3'UTR was sufficient to increase luciferase activity upon transfection of wild-type DND1. This increase became statistically significant if two consecutive UAUUU sequences (spaced by a UUUU tetranucleotide) were introduced. Transfection of the R98A eRRM1 RNP mutant or the truncation mutant lacking the dsRBD (DND1 1-235) raised the luciferase activity to that of the control level. These results indicate that the presence of an AU-rich motif in the 3'UTR is necessary and sufficient for DND1-mediated target regulation. While the dsRBD-truncation mutant was unable to rescue target expression here, our RNA-IP followed by q-RT-PCR (Figs. 1B, S3) showed that it still bound the RNA. Therefore, RNP-mediated motif recognition by the tandem RRMs of DND1 was necessary, but not sufficient for target regulation. This suggested that the dsRBD might have a downstream role in DND1-mediated gene regulation that is apparently independent of RNA binding.

The interactome of DND1 targets in HEK293T is enriched for mRNA stabilizing proteins and proteins present in RNP granules

To learn more about the context of the regulatory mechanism of DND1 binding, we transfected HEK293T cells with FLAG-tagged WT DND1, the R98A RRM12 RNA binding mutant and the 1-235 dsRBD truncation mutant. We then performed immunoprecipitation and Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometric (LC/ESI-MS/MS) analysis on three biological replicates, for identification and quantification of the pulled-down interactome (Fig. 6, Table S4). In total, 769 proteins were identified, of which 25 and 20 were found to be present at a statistically different levels for the R98A mutant and the 1-235 dsRBD truncation mutant, respectively, as compared to the interactome of WT DND1 targets (Table S4, t-test with FDR = 10%). Pleasingly, 29 from these 45 differentially present proteins were also found by Yamaji et al. (Yamaji et al. 2017) in the DND1 interactome from HEK293T cells. Notably however, our dataset (including all proteins equally pulled down by WT DND1 and mutants) did not include members of the CCR4-NOT complex, which deadenylates mRNAs as a first step in their degradation. This may have been due to the different method of DND1 over-expression in the two protocols: transient expression vs. a stable cell line, as well as different lysis and wash conditions. Seventeen proteins were present at statistically significant lower levels (least 1.5-fold change (FC)) in the pulldown with the R98A RNA binding mutant (see Fig. 6A, Table S4). This set included other RNA binding proteins that bind U-rich and AU-rich RNA (ELAVL1/HuR, HNRNPC). Gene Ontology analysis using the EnrichR tool (Chen et al. 2013) (Fig. S9) showed that this list was enriched for mRNA

stabilizing proteins (HNRNPC, ELAVL1/HuR, DHX9, SYNCRIP, LARP1 and PABPC1). Moreover, the list was also enriched for 'RNP granules as a cellular component' (due to PABPC1, LARP1, MOV10, DHX9, HNRNPL, DHX30). We validated ELAVL1, that protein present at the most statistically significant reduced level in the sample of the R98A RNA binding mutant, using western blot (Fig. S10). ELAVL1/HuR is well-known for its ability to stabilize mRNAs by binding to AU-rich elements (AREs) (Wang et al. 2013; Abdelmohsen and Gorospe 2010). One of its many functions is regulation of mESC differentiation (Wang et al. 2014). Just like DND1, it also has been reported to increase mRNA stability by blocking microRNA-targeting (Kundu et al. 2012). ELAVL1/HuR is present in stress granules (Jain et al. 2016) and is enriched in P-bodies (Hubstenberger et al. 2017), two important classes of cytoplasmic RNP granules. Cytoplasmic RNP granules are hotspots for post-transcriptional gene regulation, which depend upon complex networks of protein–RNA interactions, low-complexity protein sequences, and liquid–liquid phase separation (LLPS) for their formation (Tauber et al. 2020). Several other proteins enriched in the DND1 target HEK293T interactome are also enriched in stress granules and/or P-bodies (Table S4)(Jain et al. 2016; Hubstenberger et al. 2017). There is considerable overlap between stress, P-body and germ granules (RNP granules unique to the germline), although the latter is thought to be maintained as distinct RNP assemblies in germ cells (Voronina et al. 2011). Taken together, these findings provided evidence that DND1-dependent target regulation occurs in an RNP granule-like environment.

Nine proteins were present at statistically lower levels of at least 1.5FC in the pulldown of the 1-235 truncation mutant. This set was still GO-enriched in RNA binding terms (EIF4B, HNRNPC, MEX3A, DDX17, MATR3, TNRC6B). This suggested that the dsRBD might contribute to the binding of a subset of RNA targets regulated by these RBPs, or it recruits them to modulate the target specificity of DND1. Interestingly, TNRC6B is known to recruit CCR4-NOT. Finally, a surprising finding from the analysis by GO-term was enrichment of the 'ATPase activity' classification. DDX17 is an RNA helicase, while HSPA1A, HSPA6 and TRAP1 are chaperones belonging to the heat shock protein family that assists protein folding upon ATP hydrolysis; these proteins recognize hydrophobic surfaces of unfolded proteins and partially folded intermediates, thereby preventing protein aggregation (Morán Luengo et al. 2019). This suggested that even in the HEK293T cellular environment, there might have been folding issues with the DND1 dsRBD, a domain we were unable to obtain in soluble form for structural studies due to persistent aggregation during recombinant protein production. The ATPase activity of co-purified folding

chaperones could explain the remarkable assignment of enzymatic activity to the dsRBD (Liu and Collodi 2010; Aguero et al. 2017). Interestingly, TRAP1 is a mitochondrial isoform of cytosolic HSP90 (Altieri et al. 2012). Germ granular structures are often formed near mitochondria (Voronina et al. 2011) and it has been shown that signaling lipids on mitochondrial membranes are responsible for localization of components of intermitochondrial cement (IMC), a germ granular structure in mouse spermatocytes (Huang et al. 2011). This suggests that DND1 could be targeted to mitochondria in HEK293T cells and perhaps by a similar mechanism in developing germ cells. Altogether, the data in these differential interactome maps suggested that DND1 might be part of RNP granules localized near mitochondria regulating gene expression post-transcriptionally in germ cells.

Discussion

Unique arrangement of DND1's tandem RRMs dictates conformation, orientation and accessibility of bound RNA

We determined the solution structure of the extended tandem RRMs of DND1 complexed with AU-rich RNA, which clarified the mechanism of target RNA recognition by this RNA binding protein that is essential for germ cell survival in higher vertebrates. Previously-solved tandem RRM-RNA complexes have shown that the two canonical β-sheet surfaces accommodating the RNA either form a cleft, like Sex-Lethal (Sxl)(Handa et al. 1999) and Human Antigen D (HuD – Fig. 7A)(Wang and Tanaka Hall 2001), or an extended surface (Poly-A Binding Protein (PABP – Fig. 7B)(Deo et al. 1999) and TAR DNA-binding protein 43 (TDP-43 – Fig. 7C)(Lukavsky et al. 2013)). In all of these structures, RNA stretches of 8–10 nucleotides are bound in a canonical fashion, with higher affinity and specificity than if a single domain is used. In all of the aforementioned tandem RRMs, the bound RNA adopts an extended conformation. The RRM-RRM orientation and RNA binding topology are rather different in our DND1-RNA structure, which is a result of four structural features found so far only in this complex: the lack of a canonical RNA binding surface for RRM2, an ultra-short inter-RRM linker, an extended RNA binding surface of eRRM1, and the presence of an unprecedented RNA binding pocket on α-helix 2 of RRM2. The complex embeds only a short canonical RNA binding stretch (U_3-A_4), which is followed by binding of U_5-U_7 in a highly unusual manner. Indeed, U_6 is bound by RRM2 α-helix2, resulting in a 120-degree rotation of U_6 compared to A_4 and the U_6 phosphate. U_7 ribose and G8 phosphate are bound by the tip of the eRRM1 extension. Binding by the linker residues

supports the RNA in this unique conformation, its short length being likely crucial to bridge nucleotides specifically recognized by RRM2 and eRRM1. The path of the RNA backbone is reversed, and the RNA is more compacted than in previously determined tandem RRM-RNA complexes; the U_3-G_8 phosphates span approximately 21-23 Å (Fig. 7D), while for example, in PABP an equivalent stretch of 6 nucleotides spans approximately 26-28 Å (Fig. 7B). Such backbone reversal capacity might help to fold the RNA, or alternatively, the tandem RRMs might be suited to recognize an RNA that is in an extended stem-loop conformation. Also, the central RNA residues are not solvent-accessible compared to other tandem RRM-RNA complexes. This structural feature would be consistent with DND1 possibly acting as a steric inhibitor of effector complexes targeting proximal binding sites, as suggested for the miRNA seed-complementary sequences targeted by miRISC in p27/CDKN1B and LATS2 (Kedde et al. 2007).

Structural extension of the eRRM1 increases RNA binding affinity and stabilizes a backbone turn in the recognized RNA.

While several extensions to the canonical RRM fold have been described, either extending the β-sheet surface by one or several strands or adding an α-helix at the N-or C-terminus (Afroz et al. 2015), the DND1 N-terminal extension of a β-hairpin packed on a third α-helix is so far restricted to the hnRNPR-like family of proteins (Figs. S1, S6). An X-ray structure of a similar eRRM from another member of this family, SYNCRIP/hnRNPQ, in its free form has been solved (Hobor et al. 2018). The latter is highly similar to the DND1 eRRM1 with the exception of the β3'/3'' hairpin and a small shift in the orientation of the N-terminal extension. These differences are likely due to RNA binding in our structure or, alternatively, the presence of an additional N-terminal acidic domain (AcD) found in the SYNCRIP structure. Our structure reveals that this β-hairpin packed on a third α-helix is essential for increasing the affinity to the RNA by fixing the backbone of U_6 and G_8 on the eRRM1 via Q39 and N37, respectively (Figs. 4D, E). Therefore, it is crucial for stabilizing the turn in the backbone observed in our complex. This is reminiscent of other extensions found in the RRM contributing to RNA binding like the β-hairpin found in FUS RRM (Loughlin et al. 2019) (Fig. 7E) and the fifth β-strand of PTB RRM2 (Oberstrass et al. 2005) (Fig. 7F).

The hrRRM2 presents a novel RNA binding pocket and its integrity is necessary for DND1 function.

We have shown that the primary RNA interaction interface of DND1 lies on eRRM1. It is the proximity of the second RRM, lacking a canonical RNA binding interface, but presenting a novel pocket for stabilization of this primary binding, that makes the RNA binding topology by the tandem RRMs unprecedented. Structures of several types of RRM domains without aromatic residues on the β-sheet surface have been described (Afroz et al. 2015). The qRRMs of hnRNPF use their β1/α1, β2/β3, and α2/β4 loops for recognition of G-rich sequences (Dominguez et al. 2010), while the ΨRRM of SRSF1 uses a conserved motif in the α-helix1 for purine-rich RNA binding (Clery et al. 2013). However, our structure is the first example of an RRM mediating RNA binding via α-helix 2. We propose to call an RRM using this interface for RNA binding the hrRRM, for hnRNPR-like family related RRM. We demonstrated the importance of the binding pocket on RRM2 by mutational analysis using *in vitro* binding assays and ITC (Fig. S7B and C). It is also supported by its almost full conservation, not only in DND1 (Fig. S2) but also other members of the hnRNPR-family (Fig. S6). Thus, its RNA binding mode is likely to be conserved. Our RRM2 structure is highly similar to that of the free RRM2 of RBM46 (RNA binding motif protein 46, PDB: 2DIS)(Dang et al. 2006), including the orientation of the residues of the novel binding pocket. The importance of this pocket for DND1 function was demonstrated in functional studies in zebrafish where the equivalent to the K197 mutant (K200T) was the only mutant outside of RRM1 causing loss of function (Slanchev et al. 2009). Most other loss-of-function mutants in this study can be explained using our structure. We have already described that the zebrafish Y104 RNP mutant (the equivalent of Y102 in humans) is unstable *in vitro*. Even if it would be stable *in vivo*, interactions with U_5 would be lost. The equivalents of Y72, F89 and H121 in zebrafish dnd1 are Y70, F87 and H119 in human DND1, which are important residues stabilizing the RRM fold. Y70 is particularly important for interaction between α-helices 0 and 1 in eRRM1, linking the core RRM fold and the N-terminal eRRM extension. Mutation of these residues most likely disrupts the fold of eRRM1. The only loss-of-function mutant that is not that easily explained is N94K, a mutant of the equivalent T92 in the human protein, situated in the β2-β3 loop. This residue is in close proximity to G_8 in our structure, but not well enough defined to interpret a specific interaction with the RNA. In the context of a longer RNA it could be involved in such specific binding. Finally, it should be mentioned that the Ter mutation, causing germ cell loss and testicular teratomas in mice (Youngren et al. 2005), is a truncation at the equivalent of R190 in α-

helix 2, further confirming that RRM2 and the dsRBD are essential domains for DND1 function. The novel binding pocket in RRM2 increases affinity and specificity to the readout of eRRM1 and creates a remarkable turn in the RNA backbone.

Limited sequence specificity leads to plasticity of RNA recognition

The RNA recognition observed in our structure unifies seemingly previously published contradictory data present concerning the RNA recognition elements found enriched in DND1 targets. In fact, a combination of a UUA/UUU triplet as enriched in CLIP (Yamaji et al. 2017) was used in our structure determination as the RNA target. The motif $Y_3A_4Y_5U_6N_7N_8$ derived from our structure also fitted with the [A/G/U]AU[C/G/U]A[A/U] motif enriched in RIP targets (Ruthig et al. 2019). Moreover, this motif may be interpreted as a repetition of the UAU motif, containing 2 adenosines that are specifically recognized in our structure. We have not tested binding to an oligonucleotide containing two spaced adenosines, but the increased affinity (avidity) of RBPs binding to repetitions of high affinity motifs has been demonstrated for several other RRM-containing proteins: hnRNPG (Moursy et al. 2014), PTB (Oberstrass et al. 2005), hnRNP C (Cieniková et al. 2014) and more recently ELAVL1/HuR (Ripin et al. 2019). Our structure also provides some insight into how the residues outside of the YAY motif could be recognized. For example, the binding pocket on RRM2 specifically recognizing U_6 in our structure (Fig. 4D) could not accommodate a cytidine without losing the double hydrogen-bond to O4. Overall, it appears that the tandem RRMs of DND1 demonstrate a certain plasticity for RNA recognition where a range of sequences can be bound, but a $Y_3A_4Y_5U_6$ is necessary for high affinity binding. Such high affinity binding could be a prerequisite for the activation of downstream processes such as the recruitment of regulatory factors (Webster et al. 2019). Here, we propose that the tandem RRMs bind RNA in a two-step mechanism. In a first step a range of sequences may be preselected by low affinity binding in order to attach DND1 to scan the 3'UTR (Fig. 7G panel a). Upon encountering a high-affinity YAYU element, DND1 pauses at the central adenosine (A_4), while RRM2 locks the uridine (U_6) in its hrRRM binding pocket which can then initiate downstream processes from a fixed position (Fig. 7G panel b).

Role of the dsRBD

We have shown that DND1's tandem RRMs, like the majority of RRMs, are relatively sequence-tolerant (Afroz et al. 2015). On the other hand, we know that linear sequence motifs are often insufficient to fully

capture RBP binding specificities (Dominguez et al. 2018). Specificity may be increased due to contextual features, either in the form of bipartite motifs, such as recently found for FUS (Loughlin et al. 2019), preference for a nucleotide composition flanking a high affinity linear motif, or due to the favoring of specific structural motifs adjacent to the linear motif. While the RNA binding surfaces of the tandem RRMs are highly conserved within the hnRNPR-like family of proteins, the sequences of the dsRBDs of DND1, RBM46 and ACF1 (APOBEC1 complementation factor) are rather less conserved (Fig. S6B). Thus, the highly specialized function of DND in germline development might originate from this domain. We have shown that the dsRBD is required for target regulation, either through direct or indirect recruitment of effector proteins, or by simply by displacing other factors, but also that the dsRBD is not essential for the binding of several highly abundant RNA targets (Figs. 1B, S3). Although the dsRBD of DND1 lacks some canonical RNA recognition residues (Masliah et al. 2013) (Fig. S6B), we cannot exclude that a non-canonical RNA binding surface contributes to the binding of a different set of targets. The dsRBD might increase the target specificity of DND1 by recognizing a stem-loop motif adjacent to the linear motif recognized by the tandem RRMs. While we know that 3'UTRs are highly structured not only *in vitro* but also *in vivo* (Beaudoin et al. 2018), it remains to be determined whether the 3'UTRs in the vicinity of the linear motifs targeted by DND1 are indeed enriched in secondary structure. Further structural studies should be undertaken to confirm that such structures can indeed be recognized by the full length DND1. Another possibility for increased target specificity of DND1 is cooperation with a binding partner. This was reported for NANOS2 (Suzuki et al. 2016), which was shown to interact with CCR4-NOT (Suzuki et al. 2010, 2012), or other germline-specific RNA binding proteins.

DND1 is a germline-specific AU-rich element binding protein

DND1 binds UAUU which is contained in AU-rich sequence elements (AREs) in 3'UTRs that have long been known to target mRNAs for rapid degradation in mammals. AREs are divided into three classes with respect to the copy number of the canonical pentamer AUUUA sequence: several copies are dispersed between U-rich regions in class I, are clustered in class II, while class III are predominantly U-rich and lack these canonical pentamers (Peng et al. 1996). More than 20 AU-rich RNA binding proteins (AUBPs) have been identified, that control the fate of ARE-containing mRNAs (Wu and Brewer 2012). Because DND1 CLIP and RIP-enriched targets do not necessarily contain the canonical ARE AUUUA pentamer target sequence, DND1 can be classified as a germline-specific AU-rich RBP (AUBP) targeting class III AREs. The recruitment of degradation machineries to mRNAs

for their destruction is a unifying mechanism between several AUBPs (Chen et al. 2001; Chou et al. 2006; Lykke-Andersen and Wagner 2005), which could be shared by DND1 in one of its dual roles. The opposing, stabilizing role might depend on a network of other stabilizing RNA binding proteins such as ELAVL1/HuR that we discovered in our pulldown. Interestingly, HuR shares with DND1 a possible function in stabilizing mRNA targets by inhibiting miRNA activity. Precisely how this functions is likely target- and context-dependent as it may depend on the presence and concentration of competing factors. As multiple AUBPs may modulate the stability and translation of a single ARE-mRNA, questions of functional redundancy and additivity or antagonism arise. It is likely that variations in the relative amounts of mRNAs, in the AUBPs present within a certain cell type or developmental stage and in the binding affinities, determine both the identity of the targeted mRNAs and their fate (Webster et al. 2019). Therefore, the fact that DND1 is specifically expressed in the germline will be a major contributing factor to its target specificity. The situation is made more complex by the granule environment, with a specific composition, where RNAs and RBPs are concentrated in a small volume, likely altering binding- and unbinding equilibria. Because of this highly complex context- and target- dependence, the composition of the DND1 interactome found in our pulldowns cannot be extrapolated to the process in developing germ cells. Thus, the interpretation of DND1 function based on reporter assays in non-native cell types should be addressed with caution. Similarly, the relevance of recognition motif derivation in large-scale DND1-RNA interaction studies performed in non-native cell types with transcriptomes differing from developing germ cells and using cross-linking that might overrepresent low affinity motifs should also be interpreted with caution. The structural and biophysical work in this study raises our understanding of the requirements for a high affinity RNA binding motif in DND1. In turn, this helps to reinterpret previous the results from previous studies in order to better understand the complex gene regulation networks during germline development and maintenance.

We have demonstrated here that DND1 prefers AU-rich over U-rich RNAs and that a central adenosine is required for high-affinity binding. The adenosine is specifically recognized by the eRRM1 binding pocket involving RNP2 and the interdomain linker (position 'N_1'). This contrasts with the RRM3 of another AUBP, HuR, that recognizes both AU-rich and U-rich sequences with similar affinities, the latter slightly higher due to an avidity effect (Ripin et al. 2019). Adenosines are bound by HuR RRM3 in two different positions: either on the periphery, or β-strand 3 using RNP1 (position 'N_2'). They are important to localize the protein at a precise position within the 3'UTR. Such a 'locking' mechanism has been also

proposed for the cytoplasmic polyadenylation element (CPE)-binding (CPEB) family of RNA-binding proteins. These RBPs bind the CPE sequence UUUUAU, which activates translation by cytoplasmic polyadenylation. The CPEB4 and CPEB1 tandem RRMs bind a UUUUA pentanucleotide sequence-specifically (Afroz et al. 2014). While the uridines are bound by RRM1, the adenosine is specifically recognized by RRM2 using RNP2 (position 'N_1'). RRM1 in isolation has a low affinity for U-rich RNAs and RRM2 does not bind U-rich sequences. Therefore, it is proposed that the protein is recruited to U-rich motifs through RRM1, after which it scans the 3'UTR in an open conformation until it encounters an adenosine in a consensus CPE and locks the protein onto this sequence. This is a similar mechanism as we now propose for DND1, although in our case the scanning for a high-affinity motif likely happens in a closed rather than open conformation since the isolated RRM1 does not bind U-rich sequences. This original mode of RNA target selection therefore could be a general mechanism for cytoplasmic RNA binding proteins regulating RNA via their 3' UTR.

In conclusion, we provide here the first structural and mechanistic insight into the molecular mechanisms by which the RNA binding protein DND1 regulates a subset of mRNAs and in turn might stabilize the fate of germ cells. Our results hint at a specialized function of the individual RNA binding domains of DND1, where the tandem RRMs are mainly responsible for target binding and the dsRBD for target regulation. This possibly occurs through the recruitment of regulatory factors, other RBPs that modulate the specificity of DND1, or displacement of competing RBPs. Our structure unifies DND1 RNA recognition elements recently found enriched in genome-wide interaction studies and facilitates understanding of loss-of-function mutants previously described in the literature. In addition, we have demonstrated an additional means by which an RNA recognition motif can recognize RNA, extending the repertoire of this versatile and abundant RNA binding domain.

Acknowledgments

We are grateful to Gunter Stier (BZH, Universität Heidelberg) and Dr. Arie Geerlof (PEPF Helmholtz Zentrum München) for providing us with expression plasmids and protocols that were developed at EMBL Heidelberg. We thank Dr. Julien Boudet for assistance with ITC measurements, Dr. Antoine Clery for cloning the RRM2 expression plasmid and assistance with ITC, Dr. Fred Damberger for assistance with NMR experiments and structure calculations, Dr. Markus Blatter and Dr. Yaroslav Nikolaev with

structure calculation and IT support respectively, Mauro Zimmermann for assistance with RNA preparation, Dr. Daniel Cirera-Salinas for cloning the psiCHECK-2-TERT plasmid and assistance with luciferase assays and Dr. Rajika Arora for assistance with the MS analysis. We thank Dr. Nina Ripin and Dr. Fred Damberger for critical reading of the manuscript. NMR was carried out at the Biomolecular NMR Spectroscopy Platform BNSP of the ETH Zurich. MS was carried out by the Functional Genomics Centre Zurich (FGCZ). This work was supported by the Promedica Foundation (to M.M.D.) and the Swiss National Science Fund SNF (120% Support Grant to M.M.D., NCCR RNA biology & disease and SNF grants 31003A_149921 and 31003A_170130 to F.H.-T.A., 31003A_173120 to C.C.) and the NCCR RNA biology & disease to JH.

Author Contributions

M.M.D., F.H.-T.A. and C.C. designed the experiments and wrote the paper. M.M.D. prepared samples, solved the structure and ran ITC experiments. H.W. performed luciferase, RIP and qPCR assays. T.K. prepared mutants and samples. F.E.L designed and performed the RRM2 titration experiments. C.v.S prepared mutants. U.P. synthesized selectively labeled RNA oligos. M.M.D., F.H.-T.A. H.W. and C.C. analyzed data. F.H.-T.A., J.H. and C.C. provided funding and resources.

References

Abdelmohsen K, Gorospe M. 2010. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdiscip Rev RNA* **1**: 214–229.

Afroz T, Cienikova Z, Cléry A. 2015. One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. *Methods Enzymol* **558**: 235–278.

Afroz T, Skrisovska L, Belloc E, Guillen-Boixet J, Mendez R, Allain FH-T. 2014. A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev* **28**: 1498–1514.

Aguero T, Jin Z, Chorghade S, Kalsotra A, King M Lou, Yang J. 2017. Maternal Dead-end 1 promotes translation of nanos1 by binding the eIF3 complex. *Development* **144**: 3755–3765.

Altieri DC, Stein GS, Lian JB, Languino LR. 2012. TRAP-1, the mitochondrial Hsp90. *Biochim Biophys Acta* **1823**: 767–773.

Beaudoin J-D, Novoa EM, Vejnar CE, Yartseva V, Takacs CM, Kellis M, Giraldez AJ. 2018. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat Struct Mol Biol* **25**: 677–686.

Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, et al. 2012. Amber 12. *Univ California, San Francisco*

Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ, Rance M. 2007. *Protein NMR Spectroscopy*. Elsevier Inc.

Chen C-Y, Gherzi R, Ong S-E, Chan EL, Raijmakers R, Pruijn GJM, Stoecklin G, Moroni C, Mann M,

Karin M. 2001. AU Binding Proteins Recruit the Exosome to Degrade ARE-Containing mRNAs. *Cell* **107**: 451–464.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**: 128.

Chou C-F, Mulky A, Maitra S, Lin W-J, Gherzi R, Kappes J, Chen C-Y. 2006. Tethering KSRP, a decay-promoting AU-rich element-binding protein, to mRNAs elicits mRNA decay. *Mol Cell Biol* **26**: 3695–706.

Cieniková Z, Damberger FF, Hall J, Allain FH-T, Maris C. 2014. Structural and Mechanistic Insights into Poly(uridine) Tract Recognition by the hnRNP C RNA Recognition Motif. *J Am Chem Soc* **136**: 14536–14544.

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

Clery A, Sinha R, Anczukow O, Corrionero A, Moursy A, Daubner GM, Valcarcel J, Krainer AR, Allain FH-T. 2013. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc Natl Acad Sci*.

Corbett AH. 2018. Post-transcriptional regulation of gene expression and human disease. *Curr Opin Cell Biol* **52**: 96–104.

Dang W, Muto Y, Inoue M, Kigawa T, Shirouzu M, Terada T, Yokoyama S. 2006. Solution structure of the RRM domain of unnamed protein product. *PDB ID 2DIS*.

Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**: 277–293.

Deo RC, Bonanno JB, Sonenberg N, Burley SK. 1999. Recognition of Polyadenylate RNA by the Poly(A)-Binding Protein. *Cell* **98**: 835–845.

Dominguez C, Fisette J-F, Chabot B, Allain FH-T. 2010. Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat Struct Mol Biol* **17**: 853–861.

Dominguez D, Freese P, Alexis MS, Su A, Hochman M, Palden T, Bazile C, Lambert NJ, Van Nostrand EL, Pratt GA, et al. 2018. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* **70**: 854-867.e9.

Gerstberger S, Hafner M, Tuschl T. 2014. A census of human RNA-binding proteins. *Nat Rev Genet* **15**: 829–845.

Goudarzi M, Banisch TU, Mobin MB, Maghelli N, Tarbashevich K, Strate I, van den Berg J, Blaser H, Bandemer S, Paluch E, et al. 2012. Identification and Regulation of a Molecular Module for Bleb-Based Cell Motility. *Dev Cell* **23**: 210–218.

Gross-Thebing T, Yigit S, Pfeiffer J, Reichman-Fried M, Bandemer J, Ruckert C, Rathmer C, Goudarzi M, Stehling M, Tarbashevich K, et al. 2017. The Vertebrate Protein Dead End Maintains Primordial Germ Cell Fate by Inhibiting Somatic Differentiation. *Dev Cell* **43**: 704-715.e5.

Güntert P, Buchner L. 2015. Combined automated NOE assignment and structure calculation with CYANA. *J Biomol NMR* **62**: 453–471.

Handa 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.

Herrmann T, Güntert P, Wüthrich K. 2002a. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* **24**: 171–89.

Herrmann T, Güntert P, Wüthrich K. 2002b. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J Mol Biol* **319**: 209–227.

Hobor F, Dallmann A, Ball NJ, Cicchini C, Battistelli C, Ogrodowicz RW, Christodoulou E, Martin SR, Castello A, Tripodi M, et al. 2018. A cryptic RNA-binding domain mediates Syncrip recognition and exosomal partitioning of miRNA targets. *Nat Commun* **9**: 831.

Huang H, Gao Q, Peng X, Choi S-Y, Sarma K, Ren H, Morris AJ, Frohman MA. 2011. piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell* **20**: 376–387.

Hubstenberger A, Courel M, Bénard M, Souquere S, Ernoult-Lange M, Chouaib R, Yi Z, Morlot J-B, Munier A, Fradet M, et al. 2017. P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol Cell* **68**: 144-157.e5.

Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. 2016. ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**: 487–498.

Kanaar R, Lee AL, Rudner DZ, Wemmer DE, Rio DC. 1995. Interaction of the sex-lethal RNA binding domains with RNA. *EMBO J* **14**: 4530–4539.

Kay LE, Torchia DA, Bax A. 1989. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**: 8972–8979.

Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JAF, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Ørom UA, et al. 2007. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**: 1273–86.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858.

Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**: W43-6.

Kundu P, Fabian MR, Sonenberg N, Bhattacharyya SN, Filipowicz W. 2012. HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Res* **40**: 5088–5100.

Laskowski R, Rullmann JA, MacArthur M, Kaptein R, Thornton J. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**: 477–486.

Lee W, Tonelli M, Markley JL. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**: 1325–1327.

Licatalosi DD. 2016. Roles of RNA-binding proteins and post-transcriptional regulation in driving male germ cell development in the mouse. In *Advances in Experimental Medicine and Biology*, Vol. 907 of, pp. 123–151, Springer, Cham

Liu W, Collodi P. 2010. Zebrafish dead end possesses ATPase activity that is required for primordial germ cell development. *FASEB J* **24**: 2641–2650.

Loughlin FE, Lukavsky PJ, Kazeeva T, Reber S, Hock EM, Colombo M, Von Schroetter C, Pauli P, Cléry A, Mühlemann O, et al. 2019. The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity. *Mol Cell* **73**: 490-504.e6.

Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, Buratti E, Baralle FE, Damberger FF, Allain FH-T. 2013. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol* **20**: 1443–1449.

Lykke-Andersen J, Wagner E. 2005. Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. *Genes Dev* **19**: 351–61.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200–D203.

Masliah G, Barraud P, Allain FH-T. 2013. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci* **70**: 1875–95.

Mei W, Jin Z, Lai F, Schwend T, Houston DW, King M Lou, Yang J. 2013. Maternal Dead-End1 is required for vegetal cortical microtubule assembly during Xenopus axis specification. *Development* **140**: 2334–44.

Morán Luengo T, Mayer MP, Rüdiger SGD. 2019. The Hsp70-Hsp90 Chaperone Cascade in Protein Folding. *Trends Cell Biol* **29**: 164—177.

Moursy A, Allain FH-T, Cléry A. 2014. Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic Acids Res* **42**: 6659–72.

Notredame C, Higgins DG, Heringa J. 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**: 205–217.

Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, et al. 2005. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**: 2054–7.

Peng SS, Chen CY, Shyu AB. 1996. Functional characterization of a non-AUUUA AU-rich element from the c-jun proto-oncogene mRNA: evidence for a novel class of AU-rich elements. *Mol Cell Biol* **16**: 1490–9.

Ramsköld D, Wang ET, Burge CB, Sandberg R. 2009. An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data ed. L.J. Jensen. *PLoS Comput Biol* **5**: e1000598.

Ripin N, Boudet J, Duszczyk MM, Hinniger A, Faller M, Krepl M, Gadi A, Schneider RJ, Šponer J, Meisner-Kober NC, et al. 2019. Molecular basis for AU-rich element recognition and dimerization by the HuR C-terminal RRM. *Proc Natl Acad Sci* **116**: 2935–2944.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**: W320–W324.

Rückert M, Otting G. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* **122**: 7793–7797.

Ruthig VA, Friedersdorf MB, Garness JA, Munger SC, Bunce C, Keene JD, Capel B. 2019. The RNA-binding protein DND1 acts sequentially as a negative regulator of pluripotency and a positive regulator of epigenetic modifiers required for germ cell reprogramming. *Development* **146**: dev175950.

Ryter JM, Schultz SC. 1998. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* **17**: 7505–13.

Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. 2002a. TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci* **95**: 13585–13590.

Salzmann M, Pervushin K, Wüthrich K, Wider G, Senn H. 2002b. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J Am Chem Soc* **121**: 844–848.

Seydoux G, Braun RE. 2006. Pathway to Totipotency: Lessons from Germ Cells. *Cell* **127**: 891–904.

Shen Y, Delaglio F, Cornilescu G, Bax A. 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**: 213–223.

Singh R, Valcárcel J. 2005. Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol* **12**: 645–653.

Slanchev K, Stebler J, Goudarzi M, Cojocaru V, Weidinger G, Raz E. 2009. Control of Dead end localization and activity--implications for the function of the protein in antagonizing miRNA function. *Mech Dev* **126**: 270–7.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr*

Purif **41**: 207–34.

Suzuki A, Igarashi K, Aisaki K -i., Kanno J, Saga Y. 2010. NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. *Proc Natl Acad Sci* **107**: 3594–3599.

Suzuki A, Niimi Y, Shinmyozu K, Zhou Z, Kiso M, Saga Y. 2016. Dead end1 is an essential partner of NANOS2 for selective binding of target RNAs in male germ cell development. *EMBO Rep* **17**: 37–46.

Suzuki A, Saba R, Miyoshi K, Morita Y, Saga Y. 2012. Interaction between NANOS2 and the CCR4-NOT Deadenylation Complex Is Essential for Male Germ Cell Development in Mouse ed. A.J. Cooney. *PLoS One* **7**: e33558.

Szklarczyk A, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47 (D1)**: D607-D613.

Tauber D, Tauber G, Parker R. 2020. Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends Biochem Sci* **45**: 764–778.

Teplova M, Song J, Gaw HY, Teplov A, Patel DJ. 2010. Structural Insights into RNA Recognition by the Alternate-Splicing Regulator CUG-Binding Protein 1. *Structure* **18**: 1364–1377.

Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, Markley JL, Fox BG. 2005. Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr Purif* **40**: 268–278.

Vogt M, Taylor V. 2013. Cross-linked RNA Immunoprecipitation. *BIO-PROTOCOL* **3**.

Voronina E, Seydoux G, Sassone-Corsi P, Nagamori I. 2011. RNA granules in germ cells. *Cold Spring Harb Perspect Biol* **3**.

Wang J, Guo Y, Chu H, Guan Y, Bi J, Wang B. 2013. Multiple functions of the RNA-binding protein HuR in cancer progression, treatment responses and prognosis. *Int J Mol Sci* **14**: 10015–10041.

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.

Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC. 2014. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**: 191–198.

Webster MW, Stowell JA, Passmore LA. 2019. RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife* **8**.

Weidinger G, Stebler J, Slanchev K, Dumstrei K, Wise C, Lovell-Badge R, Thisse C, Thisse B, Raz E. 2003. dead end, a Novel Vertebrate Germ Plasm Component, Is Required for Zebrafish Primordial Germ Cell Migration and Survival. *Curr Biol* **13**: 1429–1434.

Wenter P, Reymond L, Auweter SD, Allain FH-T, Pitsch S. 2006. Short, synthetic and selectively 13C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes. *Nucleic Acids Res* **34**: e79.

Wu X, Brewer G. 2012. The regulation of mRNA stability in mammalian cells: 2.0. *Gene* **500**: 10–21.

Yamaji M, Jishage M, Meyer C, Suryawanshi H, Der E, Yamaji M, Garzia A, Morozov P, Manickavel S, McFarland HL, et al. 2017. DND1 maintains germline stem cells via recruitment of the CCR4–NOT complex to target mRNAs. *Nature* **543**: 568–572.

Youngren KK, Coveney D, Peng X, Bhattacharya C, Schmidt LS, Nickerson ML, Lamb BT, Deng JM, Behringer RR, Capel B, et al. 2005. The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature* **435**: 360–364.

Zhu G, Xia Y, Nicholson LK, Sze KH. 2000. Protein Dynamics Measurements by TROSY-Based NMR Experiments. *J Magn Reson* **143**: 423–426.

Figures and legends

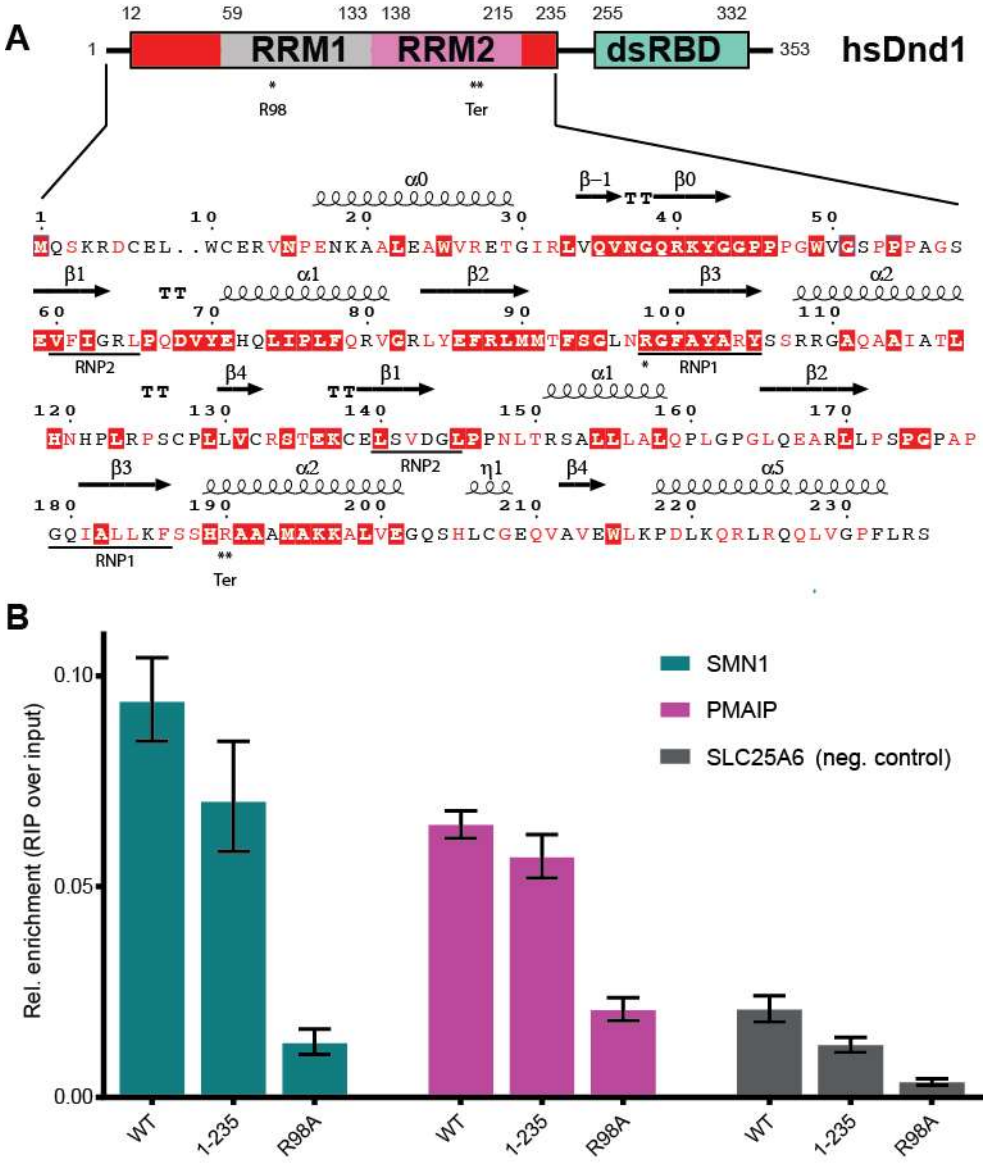

Figure 1. DND1 binds RNA targets mainly through its RRMs

A) Domain structure of DND1 and sequence of the N-terminal part of the human protein (Uniprot Q8IYX4) ending at the C-terminus of the tandem RRM construct used for structure determination in this work (12-235). RRM1 in grey, RRM2 in pink, N- and C-terminal conserved extensions in red, dsRBD in green. The dsRBD-truncation mutant 1-235 used in our RIP assay ends after the extension of RRM2. Red coloring in sequence indicates high conservation as described in Fig. S2. Secondary structure elements as found in our structure are indicated above the sequence. The RRM-canonical RNA binding RNP sequences are underlined below. R98 in RNP1 that was mutated for the RIP assay is indicated with one asterix. The Ter truncation at R190 is indicated with a double asterix and 'Ter'. See also Figs. S1 and S2. **B) RNA Immunoprecipitation** from HEK293T cells transiently expressing FLAG-tagged DND1 or its mutants followed by qPCR using primers for published DND1 targets and a negative control (Table S1). Technical triplicate of a representative plot of three independent experiments is presented as relative enrichment over the input ($2^{-\Delta Ct}$). ΔCt is an average of (Ct [RIP] – (Ct [Input]) of technical triplicates with SD < 0.4. Error bars represent SD (ΔCt). Results from two other independent RIP experiments are shown in Fig. S3A. DND1 and mutants are well expressed in HEK293T cell culture (Fig. S3B).

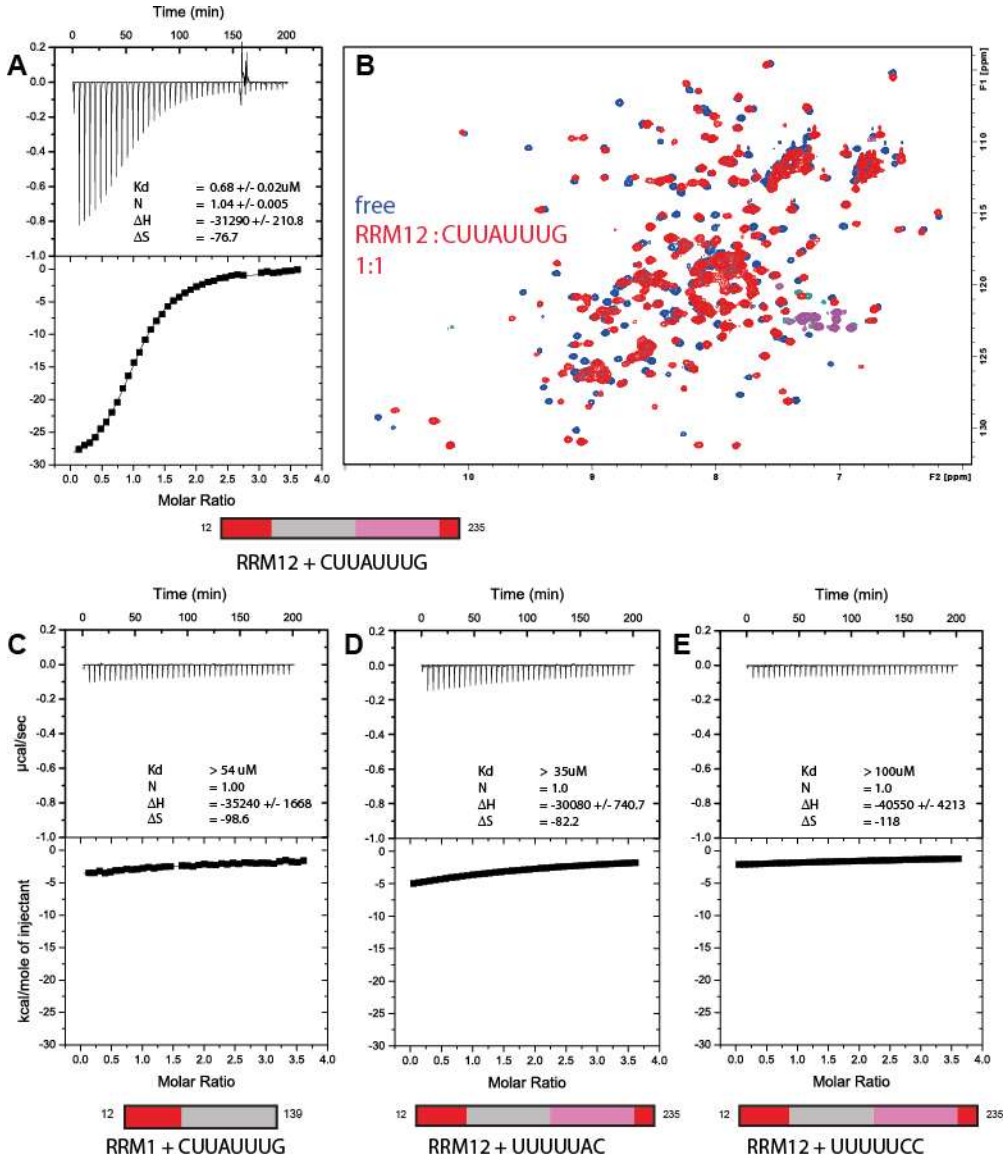

Figure 2. Cooperative binding of DND1's tandem RRMs to AU-rich RNA

A) ITC measurement of DND1's tandem RRMs domain titrated with CUUAUUUG RNA. N = 1.04 +/- 0.005; K_D = 0.68 +/- 0.02 µM; ΔH = -31.3+/- 0.2 kcal/mol; -TΔS = 22.9 kcal/mol. **B) overlay of two 1H-15N HSQC spectra** of the free DND1 tandem RRMs (in blue) and a 1:1 complex with CUUAUUUG RNA (in red) demonstrates cooperative RNA binding. NMR conditions: Protein concentration 0.2mM, 298K, 750mHz. NMR/ITC buffer: 100mM Potassium Phosphate pH 6.6, 1mM DTT **C) ITC measurement of DND1's extended RRM1** titrated with CUUAUUUG (N = 1 (set); K_D > 54 µM; ΔH = -35 +/- 2 kcal/mol; -TΔS = 29.4 kcal/mol) **D) ITC measurement of DND1's tandem RRMs** titrated with UUUUUAC (N = 1 (set); K_D > 35 µM; ΔH = -30.1 +/- 0.7 kcal/mol; -TΔS = 24.5 kcal/mol) and E) **DND1's tandem RRMs** domain titrated with UUUUUCC (N = 1 (set); K_D > 100 µM; ΔH = -41 +/- 4 kcal/mol; -TΔS = 35 kcal/mol). See also Fig. S4 and Table S2.

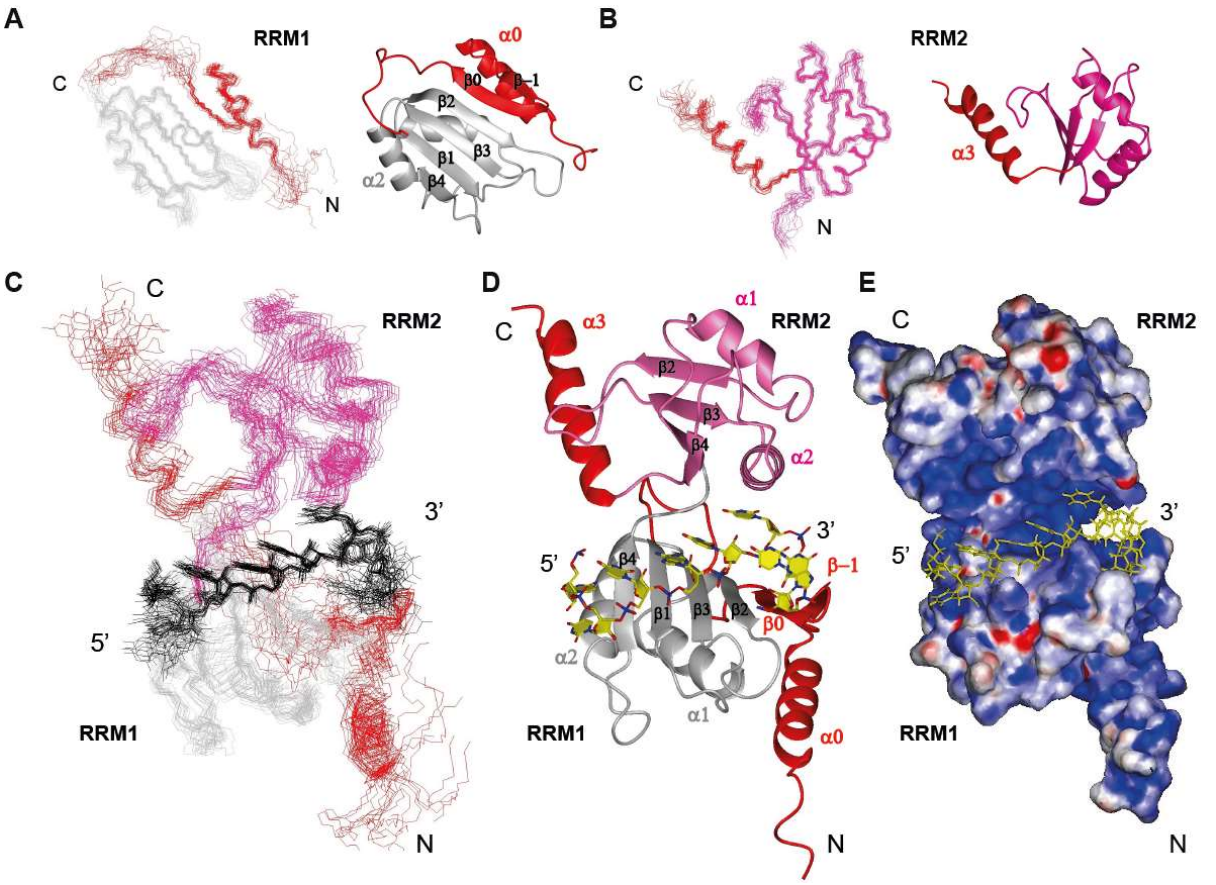

Figure 3. Solution structures of DND1's tandem RRMs bound to CUUAUUUG RNA

A) Ensemble of 20 superimposed lowest energy conformers. RRM1 in grey, RRM2 in pink, conserved N- and C-terminal extensions in red. RNA in black. **B) Superimposed 20 lowest energy conformers and representative structures for RRM1** and **C) RRM2** within the tandem RRM-RNA complex. **D) Representative structure**, color coding as in A, RNA in yellow **E) Electrostatic surface potential plot** shows the RNA is buried in a positively charged cleft between the RRMs. See also Fig. S5 and Table S3.

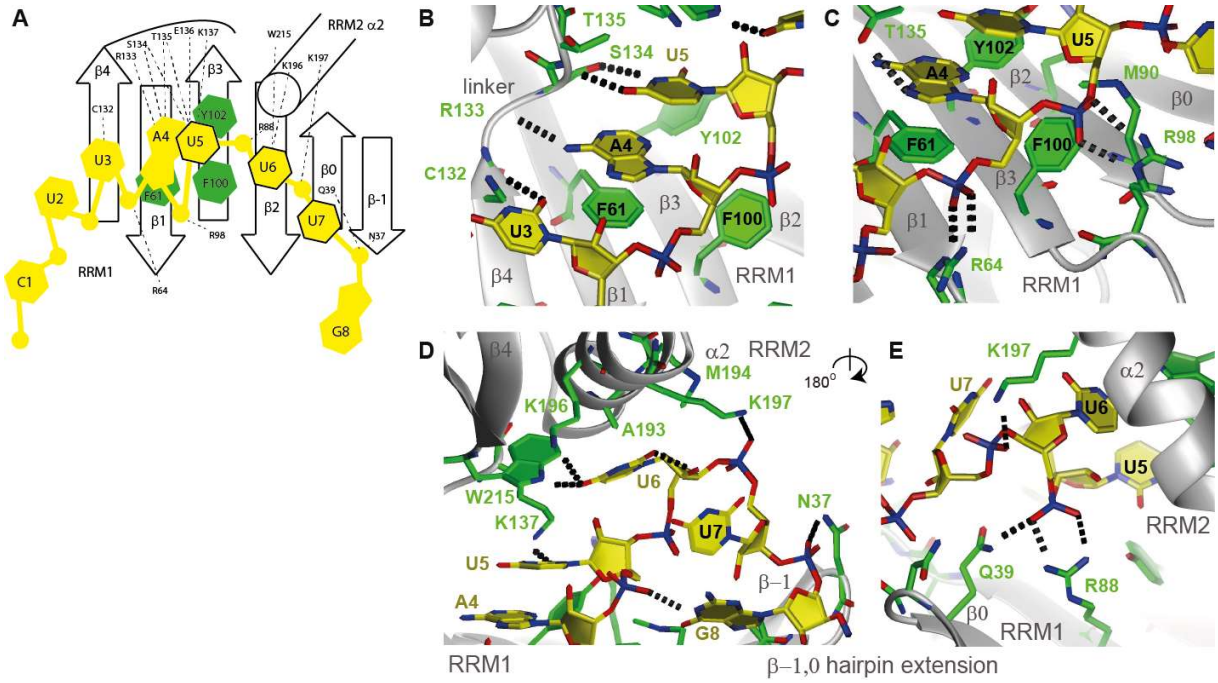

Figure 4. Intermolecular contacts between the DND1 tandem RRMs and the CUUAUUUG RNA A) schematic view of protein-RNA interactions B) U_3, A_4 and U_5 base moieties C) A_4 and U_5 backbone D) U_5, U_6, and U_7 binding to the interdomain linker and the RRM2 binding pocket. G_8 binding to the eRRM1 β-hairpin extension E) Cooperative binding by RRM1 and 2 of the U_6 and U_7 phosphate backbone, seen from the back of the molecule. Protein sidechains in green, RNA in yellow. Hydrogen-bonds in dots. See also Fig. S7.

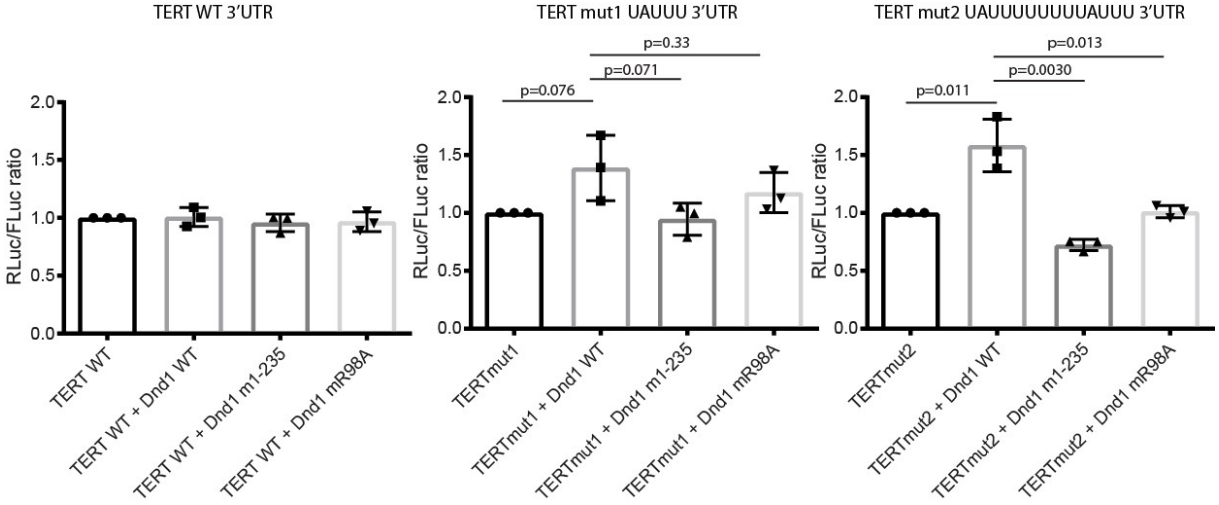

Figure 5 Introduction of an AU-rich motif into a UTR is sufficient, RNA binding by the tandem RRMs is not enough for target rescue by DND1

A wild-type TERT-UTR psiCHECK2 luciferase construct or the same construct with introduction of a single UAUUU or double UAUUUUUUUUAUUU DND1 tandem RRM target site, and either wild type or mutant FLAG-tagged DND1 were transfected into HEK293T. Relative luciferase activity is the ratio between Renilla control and firefly luciferases, adjusted to 1 for 100%. The results are represented as means and SD from three independent experiments. p-values from two-tailed Welch's t-test. Immunostaining with anti-FLAG antibody in Fig. S3B shows that DND1 and all mutants are well expressed in HEK293T cells. See also Fig. S8.

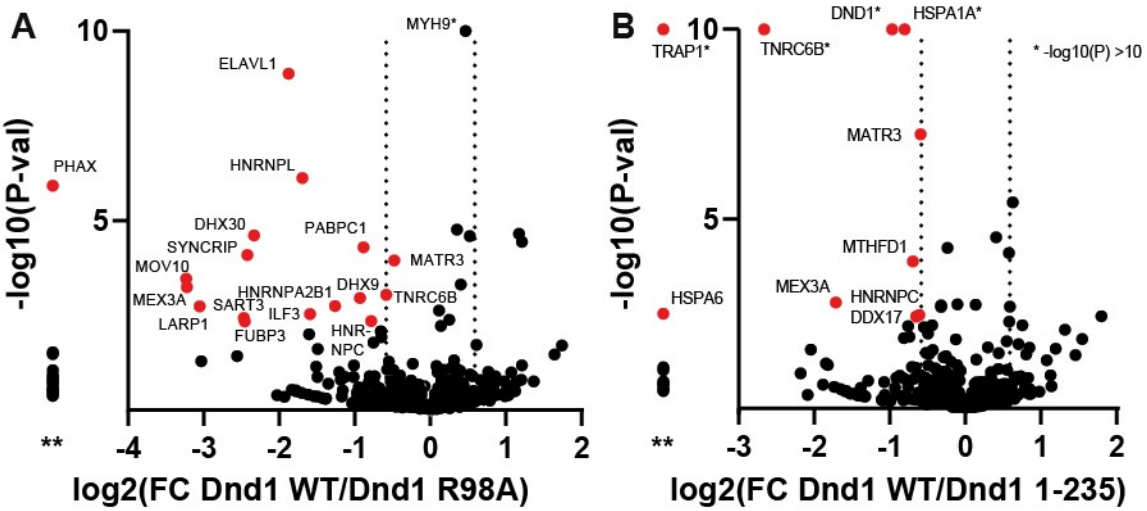

Figure 6. Proteomics analysis revealed differential interaction networks of targets of mutant versus WT DND1. In total, 769 proteins were quantified and used for *t*-testing to generate the volcano plot indicating the $\log_2$ fold change (FC Dnd1WT/Dnd1mut) ratio on the *X*-axis and −log *p* value on the *Y*-axis for each protein. * shows proteins with -log(Pval) >10. 17 proteins (in red) were present at at least 1.5 FC lower levels in the R98A RNA binding mutant as compared to DND1 WT, while 9 proteins were present at at least 1.5 fold change lower levels in the 1-235 dsRBD (a *t*-test was performed with FDR = 10%, see table S4). ** The column left of the y-axis shows proteins present in WT but completely missing in the mutants.

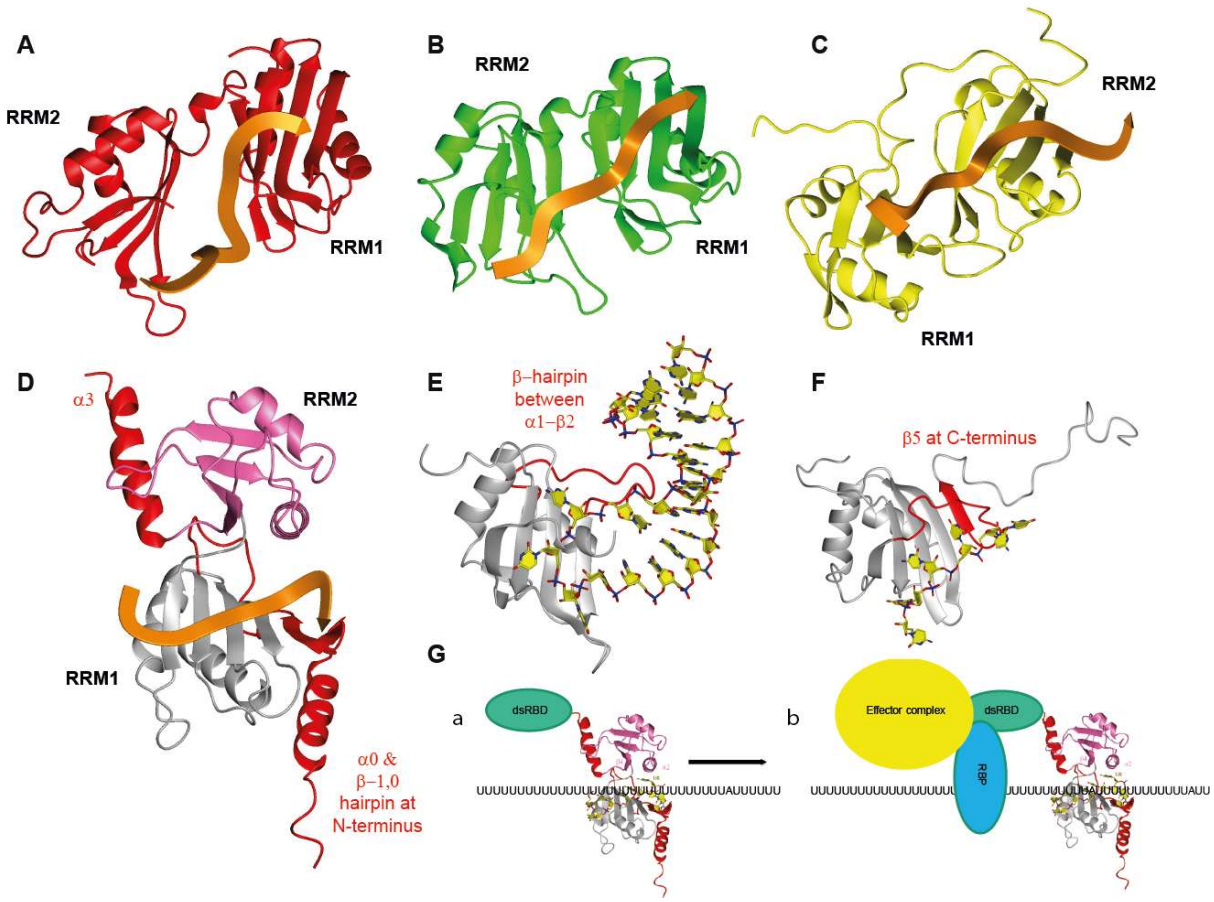

Figure 7. Structural comparison of tandem RRM-RNA complexes (A-D), RRMs using structural extensions to the canonical RRM fold to increase affinity to their RNA targets (D-F) and model of DND1 action mechanism

A) The N-terminal two RRMs of HuD bound to AU-rich RNA B) The N-terminal two RRMs of PABP bound to poly(A) RNA C) The RRMs of TDP-43 bound to UG-rich RNA D) The RRMs of DND1 bound to AU-rich RNA with extensions to the canonical RRM fold shown in red E) FUS RRM bound to a stem-loop from hnRNPA2/B1 pre-mRNA F) RRM2 of PTB bound to CU-rich RNA G) model of AU-rich mRNA target regulation by DND1.

Main table and legend

Table 1: NMR statistics

Dnd1(12-235):CUUAUUUG		
NMR Restraints		
Distance Restraints		5200
Protein	intramolecular	4947
	intraresidual	942
	sequential (|i-j|=1)	1313
	medium range (1<|i-j|<5	1190
	long range (|i-j|>=5	1423
	hydrogen bonds[a]	79
RNA	intramolecular	150
	intraresidual	116
	sequential (|i-j|=1)	34
	medium range (1<|i-j|<5)	0
	long range (|i-j|>=5)	0
	hydrogen bonds	0
Complex	intermolecular	103
	long range (|i-j|>=5	103
	hydrogen bonds	0
Torsion Angles[b]		
Protein	backbone	176
RNA	sugar pucker (DELTA)	8
RDCs		
Protein	amide NH	127
Energy Statistics[c]		
Average distance constraint violations		
0.3-0.4 Å		10.3 +/- 2.5
>0.4 Å		5.5 +/- 2.0
Maximal (A)		0.66 +/- 0.14
Average angle constraint violations		
>5 degree		3.0 +/- 1.2
Maximal (degree)		9.0 +/- 4.5
Average RDC violations		
>5 Hz		5.2 +/- 1.2
Maximal (Hz)		6.1 +/- 0.9
RDC correlation coefficient		0.96 +/- 0.01
Q = rms(Dcalc-Dobs)/rms(Dobs) (%)		30.7 +/- 2.2
Q normalized by tensor		17.9 +/- 1.5
Mean AMBER Violation Energy		
Constraint (kcal mol^{-1})		297.5 +/- 21.9
Distance (kcal mol^{-1})		168.9 +/- 18.5
Torsion (kcal mol^{-1})		10.7 +/- 2.4
Alignment (kcal mol^{-1})		104.4 +/- 9.3
Mean AMBER Energy (kcal mol^{-1})		-7568.9 +/- 18.5
Mean Deviation from ideal covalent geometry		
Bond Length (A)		0.0041 +/- 0.0000
Bond Angle (degrees)		1.511 +/- 0.009
Ramachandran plot Statistics[c,d]		
Residues in most favoured regions (%)		82.1 +/- 1.9
Residues in additionally allowed regions (%)		15.3 +/- 2.1
Residues in generously allowed regions (%)		2.0 +/- 0.8
Residues in disallowed regions (%)		0.6 +/- 0.5
RMSD to mean structure		
Protein Dnd1 RRM1 16-133		
Backbone atoms		0.95 +/- 0.23

Heavy atoms	1.23 +/- 0.20
Protein Dnd1 RRM2 134-216	
Backbone atoms	0.57 +/- 0.12
Heavy atoms	0.92 +/- 0.12
Protein Dnd1 RRM12 16-216	
Backbone atoms	1.05 +/- 0.14
Heavy atoms	1.36 +/- 0.13
RNA UAUUU	
Backbone atoms	0.43 +/- 0.13
Heavy atoms	0.60 +/- 0.18
Complex RRM12 16-216;UAUUU	
Backbone atoms	1.04 +/- 0.14
Heavy atoms	1.26 +/- 0.13

RMSD, root-mean-square deviation
a Hydrogen bond constraints were identified from slow exchanging amide and imino protons in D_2O
b Torsion angle based on TALOS+ predictions; sugar puckers based on homonuclear TOCSY
c Dnd1 RRM12, 16-216 Chain ID: A (Sequence Range:12-235); RNA CUUAUUUG: 3-7, Chain ID: B (Sequence Range:1-8)
d Ramachandran plot, as defined by the program Procheck (Laskowski et al. 1996).

Methods

REAGENTS & RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
FLAG-M2-HRP	Sigma	A8592
HuR rabbit polyclonal antibody	ProteinTech	11910-1-AP
Bacterial and Virus Strains		
Top10 competent cells	This study	
BL21(DE3) competent cells	This study / Novagen	
Biological Samples		
DMEM	Sigma	D6429
FBS	Sigma	F7524
OptiMEM	GIBCO	31985047
Chemicals, Peptides, and Recombinant Proteins		
TEV protease	In-house produced	
Critical Commercial Assays		
Clarify ™ Western ECL substrate	BioRad	170-5061
Dual-Glow Luciferase Assay System	Promega	E2920
Deposited Data		
Chemical Shifts	BioMagResBank	xxx
Structural Ensemble	Protein Data Bank	xxx
Experimental Models: Cell Lines		
HEK293T cell line	ATCC	CRL-3216™
Oligonucleotides		
CUUAUUUG RNA	Dharmacon	
selectively ribose-13C labeled RNA	This study	
Oligonucleotides for cloning & site-directed mutagenesis	Microsynth	Table S1
miRIDIAN miR-221-3p miRNA mimic	Dharmacon	MIMAT0000278
Scrambled miRNA mimic	Dharmacon	CN-001000-01-05
Recombinant DNA		
pET-M11 vector	EMBL	
pET-His-TEV	EMBL	
psiCHECK2 vector	Promega	
pCMV-SPORT6-hsDnd1 source plasmid	Source BioScience	IRATp970F0747D
pcFLAG_DNA3.1	This study / Invitrogen	
Software and Algorithms		
NMR data acquisition and processing	Topspin (Bruker)	Version 2.1
NMR data processing and relaxation analysis	NMRPipe	
Spectral analysis	NMRFAM-Sparky	Version 1.1412
Structure refinement	AMBER 12	Ff12SB force field
Structure calculation	CYANA	Version 3.98
Peak picking, structure calculation	ATNOS-CANDID	Version 3.1
Statistical analyses	PRISM	Version 7.04

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Frédéric Allain (allain@bc.biol.ethz.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Sigma) containing 10% fetal bovine serum (FBS, Sigma) including antibiotics (0.05 mg/mL of streptomycin, and 50 U/mL of penicillin (Sigma)) in a humidified incubator (Thermo Scientific Heraeus Series 6000 Incubators, Thermo Scientific) with 5% CO_2 at 37ºC.

METHOD DETAILS

Protein expression and purification

DNA fragments encoding human Dnd1 RRM1 (12-139), RRM2 (136-227) or the tandem RRM12 (12-235) were PCR amplified from the source plasmid pCMV-SPORT6-hsDnd1, an IMAGE cDNA clone (clone ID MGC:34750; IMAGE: 5172595) purchased from Source BioScience (Nottingham UK) with the primers listed in Table S1. They were cloned into the pET-M11 vector (EMBL) with an N-terminal TEV-cleavable 6xHis-tag between the NcoI and Acc65I restriction sites, using BbsI instead of NcoI to cut the insert to circumvent insert-internal restriction sites. Protein mutants were obtained by PCR-based site-directed mutagenesis with the pET-M11-RRM12 (12-235) plasmid as a template according to the QuikChange protocol (Stratagene) and the primers listed in Table S1. All protein constructs were expressed in *E. coli* BL21(DE3) cells (Novagen) in Studier-medium P-5052 supplemented with $^{15}NH_4Cl$ or P-50501 supplemented with $^{15}NH_4Cl$ and ^{13}C-glycerol (CIL). Precultures were grown in PA-0.5G medium (Studier 2005; Tyler et al. 2005). Random fractionally deuterated protein for recording of triple-resonance spectra for backbone assignment was expressed in 100% D2O (CIL) in which the media components were directly dissolved. Protein was expressed for 60h at 15°C in the presence of 100 µg/mL Kanamycin. Cells were harvested by centrifugation at 4 °C, 15 min at 2,600*g*, and the cell pellet was resuspended in lysis buffer (20 mM Tris, pH 8, 1 M NaCl, 0.2% Triton-x-100 (w/v), 10 mM imidazole, and 2 mM 2-mercaptoethanol). Cells were lysed with two freeze-thaw cycles and three passes through the Emulsiflex cell cracker (Avestin). Before

lysis 0.5mg/ml lysozyme, 25ug/ml DNAseI and 1mM Pefabloc SC (Sigma-Aldrich) was added. After centrifugation at 4 °C for 20 min at 43,000*g*, the cleared supernatant was sterile-filtered and loaded onto 2mL Ni-NTA beads (Qiagen), equilibrated with lysis buffer, per liter of bacterial culture. The column was washed with 10 column volumes of lysis buffer, 20 columns of lysis buffer without Triton and 5 column volumes of the same buffer with 30mM Imidazole, before the protein was eluted with elution buffer (lysis buffer without Triton and with 330 mM imidazole). For cleavage of the His_6 tag, the pooled fractions were dialyzed against lysis buffer (1M NaCl and no imidazole) in the presence of in-house purified TEV protease (1:100 w/w TEV:protein) at 4 °C overnight. Next day the TEV cleavage reaction was reloaded three times over a fresh Ni-NTA column to remove the His_6-TEV protease, the His_6-tag fusion and contaminating proteins. The proteins were concentrated with Vivaspin 20-mL centrifugal devices with 5,000 or 10,000 MWCO (Sartorius) and buffer-exchanged into NMR buffer over PD-10 gel-filtration columns (GE-healthcare).

RNA samples.

Unlabeled RNA oligonucleotides were purchased from Dharmacon, deprotected according to the manufacturer's instructions, lyophilized and resuspended twice in water for large-scale protein-RNA complex production or NMR buffer for titrations or ITC (20mM MES pH 6.6, 100 mM NaCl). For the solid phase synthesis of selectively ribose-labeled oligos 2′-*O*-TOM protected ribonucleoside phosphoramidites and solid supports containing [$^{13}C_5$]-labeled ribose moieties were synthesized as described, followed by their sequence-specific introduction into the CUUAUUUG oligonucleotide (Wenter et al. 2006).

NMR sample preparation of Protein-RNA complexes

Final protein was analyzed for nucleic acid contamination using A_{260nm}/A_{280nm} and concentration was estimated using A_{280nm} and a theoretical extinction coefficient of 18140 M^{-1} cm^{-1} for RRM1, 5930 M^{-1} cm^{-1} for RRM2 and 23470 M^{-1} cm^{-1} for RRM12. RNA concentrations were estimated using OligoCalc (Kibbe 2007). In the final buffer exchange step the RRM constructs were added dropwise to a 10% molar excess of RNA in the presence of 10ul of SuperaseIn RNase inhibitor (Ambion) per sample, concentrated and further purified by size exclusion chromatography on a Sephadex 75 10/30 column

(GE healthcare) in 100 mM $KHPO_4$/ KH_2PO_4 pH 6.6, 1 mM DTT. The fractions containing the protein-RNA complex were concentrated to 400-700 uM with Vivaspin 5-mL centrifugal devices with 10,000 MWCO (Sartorius). Before the measurements a 10% molar excess of RNA was added to saturate the protein as well as 10% v/v D_2O. Complexes were lyophilized before resuspending in D_2O for NMR experiments that are conducted in deuterated solvent.

Plasmids for cell culture assays

Total RNA was extracted from cultured human fibroblasts (GM03814, Coriell Institute for Medical Research, USA). 1ug was then used for reverse transcription reaction using Oligo$(dT)_{18}$ and M-MuLV Reverse Transcriptase RNaseH$^-$ (Finnzymes). The 3'UTR of TERT and fragments corresponding to positions 183–282 (according to Ensembl transcript ENST00000228872) of the 3'UTR of *p27* including the predicted *miR-221* binding sites were amplified from the cDNA templates using the appropriate primers in table S1 introducing XhoI and NotI restriction sites. The Dnd1 binding site was introduced into the TERT 3'UTR insert using PCR overlap extension using the primers in table S1. 3'UTR PCR products were directionally cloned downstream of the Renilla luciferase open reading frame (ORF) of the psiCHECK2 vector (Promega) that also contains a constitutively expressed firefly luciferase gene, which was used to normalize transfections. Dnd1 fragments encoding the full-length human protein (1-353) or a dsRBD truncation (1-235) were amplified as described for the protein expression plasmids and cloned BamHI/EcoRI into an in-house modified pcDNA3.1+ plasmid (Invitrogen) with an N-terminal FLAG tag cloned NheI/HindIII. All plasmids were confirmed by sequencing and deposited at AddGene (www.addgene.org/Frederic_Allain). Plasmids for transfections were prepared using the Nucleobond Xtra midiprep kit (Macherey-Nagel) according to the manufacturer's protocol.

Transfections and Dual luciferase activity analysis

HEK293T cells were transfected with Lipofectamine 2000 Reagent (Invitrogen) after seeding them 16 hours prior at 70`000 cells per well (24-well plate) or $2.8x10^6$ cells per 10cm dish for immunoblotting analysis and RNA immunoprecipitation. For transfections in 24-well plates, Lipofectamine 2000 Reagent was diluted in serum free medium (OptiMEM, GIBCO) and incubated for 5 min at room temperature.

Plasmid DNA (0.5 µg per plasmid as indicated) and/or 50 nM final miR-221-3p miRNA mimic (miRIDIAN, Dharmacon) or control mimic was then added, vortexed, and incubated for 20 min at room temperature while cell culture media was exchanged to DMEM containing 10% FBS without antibiotics. Finally, the transfection complexes were added to the cell culture vessel in a drop wise manner while swirling. Transfection media were changed 6 h later to regular culture media. Luciferase activity was measured 48 hours after transfection using the Dual-Glow Luciferase Assay System (E2920 Promega, USA) on a GloMax® Discover Multimode Microplate Reader (Promega, USA). The results are represented as means and standard deviation (SD) from three independent experiments.

Immunoblotting analysis of protein expression and antibodies

Total cellular protein was extracted from $6x10^5$ HEK293T cells using a RIPA buffer (1% NP-40, 150 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.5% Sodium deoxycholate, 0.5% SDS) complemented with EDTA-free protease inhibitor cocktail (Roche) followed by brief sonication. Protein concentrations were determined by DC Assay (Bio-Rad). For each sample, 14 µg of total cellular protein was separated on 12% SDS-PAGE gels and transferred on PVDF membranes. The following antibody was used: FLAG-M2-HRP (SIGMA, A8592). Immunoblots were developed using the Clarify ™ Western ECL substrate (BioRad) kit and were detected using an imaging system (ChemiDoc™ MP – BioRad). All membranes were stained using a coomassie blue staining solution to ensure equal loading. The analysis was performed in triplicate.

RNA immunoprecipitation

The RNA-immunoprecipitation (RIP) procedure was adapted from Vogt and Taylor (Vogt and Taylor 2013). Briefly, subconfluent cells from one 10 cm dish were harvested 48 hours after transfection, washed in PBS1X, and cross-linked with 1% formaldehyde. Glycine (0.125 M final) was added to quench the formaldehyde. Cells were pelleted by centrifugation and washed with PBS1X. Immunoprecipitation (IP) lysis buffer (50 mM HEPES at pH 7.5, 0.4 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT, 0.5% Triton X-100, 10% glycerol) containing 1 mM PMSF, protease inhibitors (Roche), and RNase inhibitor (Thermo Scientific) was added to the cell pellet. After sonication (Bioruptor, Diagenode), cell lysates were precleared with IP Lysis buffer containing 1% BSA. 40 µl of magnetic FLAG-M2 beads

(SIGMA, M8823) were added to precleared cell lysate and incubated on a rotary wheel at 4 °C overnight. FLAG-M2 beads were washed with IP lysis buffer five times and pelleted by centrifugation. RIP buffer (50 mM HEPES at pH 7.5, 0.1 M NaCl, 5 mM EDTA, 10 mM DTT, 0.5% Triton X-100, 10% glycerol, 1% SDS) containing RNase inhibitor was added to the pellet and incubated 1 hour at 70 °C to reverse the cross-links. After centrifugation, the supernatant was used for RNA extraction using TRIzol reagent (Life technologies) followed by a DNase I treatment and a subsequent reverse transcription with oligo $d[T]_{18}$ using the GoScript RT kit (Promega). One-step RT-qPCR was performed using the SYBR FAST Mix optimized for LightCycler 480 (KAPA, KK4611) with primers listed in table S1. The results are presented as relative enrichment over the input ($2^{-\Delta Ct}$). ΔCt is an average of (Ct [RIP] – (Ct [Input]) of technical triplicates with SD < 0.4. Three independent RIP experiments were performed.

Immunoprecipitation and shotgun LC-MS/MS analysis of DND1 interactome

$1.2x10^6$ HEK293T cells were seeded the night prior to transfection per 10cm dish. 24µg of DNA per sample was transfected using 50µl of Lipofectamine 2000. Cells were harvested, cell pellet washed with 1xPBS and resuspended in 1ml IP-Lysis buffer (50mM Tris-HCl, pH 7.4, 150mM NaCl, 1mM EDTA, 0.5% Triton X-100) supplemented with PMSF and complete (Roche) protease inhibitors. Samples were incubated on ice for 10 min and sonicated. Lysates were centrifuged at 14000g at 4°C. Protein was quantified using the DC protein assay (BioRad) and normalized to 1mg/ml total protein. 40µl of FLAG-M2 magnetic beads (SIGMA) (50% slurry) per IP were washed 3x using 750µl IP-Lysis buffer. Lysate was added and incubated over night at 4°C.

For western blotting, after removal of supernatant, the beads were washed gently twice in IP-Lysis buffer, followed by 5 times in 1x TBS (pH 7.4). Protein was eluted by adding 100µl 0.1 M Glycine pH 3.0 to the beads followed by shaking for 6 minutes at 23°C and 600rpm. 10µl 0.5M Tris-HCl, pH 7.4 was added to the supernatant and 10µl of elution was used for WB analysis.

For shotgun LC-MS/MS analysis the beads were washed twice with 100µl 50%ACN/NH4CO3 and once with 100µl ACN. 45 µl of digestion buffer (10 mM Tris/2 mM CaCl2, pH 8.2) and 5 µl trypsin (100 ng/ µl in 10 mM HCl) were added. Digestion for 30 min at 60°C was assisted by microwave. Supernatant was collected and peptides were extracted from the beads with 150µl of 0.1%TFA/50% acetonitrile. Supernatants were combined and dried, dissolved in 20 µl 0.1% formic acid, further diluted 1:10 in 0.1% FA of which 2 µl were injected. Database searches were performed using the Mascot search engine

against SwissProt. Scaffold (4.8.9) was used to filter sequence assignments with a protein threshold set to 1% false discovery rate (FDR), minimum number of peptides 2 and peptide threshold set to 0.1% FDR. Enrichment of pulled-down proteins was calculated by dividing the number of peptides found for that protein by the total number of peptides in that sample. To compare enrichment for WT DND1 to the mutants, multiple t-tests were performed on the median enrichment values of three independent pulldowns using GraphPad Prism 7.04. Discoveries were determined using the Two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, with Q/FDR = 10%. Only proteins with a median of peptides present in the DND1 WT pulldown larger than two were compared.

NMR data collection and assignments

All NMR spectra were recorded at 298K on Bruker AVIII600 MHz, AVIII700 MHz, and Avance 900 MHz spectrometers equipped with cryoprobes and a Bruker AVIII750MHz spectrometer using standard NMR experiments if not mentioned otherwise (Cavanagh et al. 2007). The data were processed using Topspin 3.1 (Bruker) and NMR Pipe (Delaglio et al. 1995) and analyzed with NMR-FAM-SPARKY (Lee et al. 2015). Sequence-specific backbone assignments were 93% complete for non-proline residues and were obtained from 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{13}C-HSQC, 3D ^{1}H-^{15}N-NOESY (t_{mix} = 120 ms) and a suite of 3D TROSY-based backbone experiments (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB and HN(CO)CACB)(Salzmann et al. 2002b, 2002a) run on a random fractionally deuterated ^{13}C,^{15}N-labeled (1:1.1) tandem RRM-CUUAUUUG complex. Sidechain protons were assigned to 80% completeness using 3D ^{1}H-^{15}N-NOESY (t_{mix} = 120 ms), 3D ^{1}H-^{13}C-HMQC-NOESY (t_{mix} = 70 ms), 3D ^{1}H-^{13}C-HSQC-aromatic-NOESY (t_{mix} = 80 ms), 3D (H)CCH- (t_{mix} = 21.7 ms) and HC(C)H-TOCSY (t_{mix} = 23 ms) and 3D H(C)CH-COSY on a sample of fully protonated ^{13}C,^{15}N-labeled (1:1.1) tandem RRM-CUUAUUUG complex. Sidechains in the free ^{13}C,^{15}N-labeled RRM2 were assigned using H(C)(CCCO)NH-TOCSY (t_{mix} = 17.75 ms) and (H)C(CCCO)NH-TOCSY (t_{mix} = 17.75 ms) and transferred to the RRM12-CUUAUUUG complex where this was possible. RNA was assigned using the following set of spectra: 2D TOCSY (t_{mix} = 60 ms), 2D NOESY (t_{mix} = 150 ms) recorded on a 1:1 complex of unlabeled tandem RRM-CUUAUUUG complex. 2D ^{1}H-^{13}C-HSQC recorded on 1:1 complexes between unlabeled tandem RRMs and selectively ^{13}C ribose-labeled C*UU*AU*UU*G or CU*UA*UU*UG* RNA where an asterix after the nucleotide represents a ^{13}C labeled ribose moiety. 2D F_2 filtered NOESY (t_{mix} = 120 ms) and 2D F_1 filtered, F_2 filtered NOESY recorded on ^{13}C,^{15}N-labeled tandem RRMs in 1:1 complex with 1)

unlabeled CUUAUUUG RNA 2) selectively ^{13}C ribose-labeled C*UU*AU*UU*G RNA 3) selectively ^{13}C ribose-labeled CU*UA*UU*UG* RNA. Sugar puckers in the complex were identified from 2D ^{1}H-^{1}H-TOCSY (t_{mix} = 60 ms). Strong H1'-H2' and weak H3'-H4' cross-peaks defined all puckers as C2'-endo. All χ dihedral angles were restrained to anti conformations based on lack of strong intraresidue H1'-H6/H8 NOEs. Intermolecular NOEs were identified using 2D ^{13}C F_2 filtered 2D NOESY (t_{mix} = 60ms) and 3D F_3 filtered, F_2 edited ^{13}C HMQC-NOESY (t_{mix} = 70 ms) in D_2O. Intramolecular NOEs of RNA were identified using 2D ^{13}C F_1 filtered, F_2 filtered NOESY (t_{mix} = 150ms) in D_2O.

NMR titrations

NMR titrations were performed by adding unlabeled concentrated RNA (1-5mM) to ^{15}N-labeled protein (0.1-0.2mM) in NMR buffer (20mM MES pH 6.6, 100 mM NaCl) and monitored by ^{1}H-^{15}N-HSQC. To monitor the chemical shift perturbations of the tandem RRM mutants upon addition of RNA, 1:1 complexes were directly prepared in NMR buffer (100 mM $KHPO_4$/ KH_2PO_4 pH 6.6, 1 mM DTT) as described under 'NMR sample preparation'.

NMR relaxation and RDC measurements

Backbone dynamics data of the tandem RRMs in complex with CUUAUUUG were recorded on a 1:1.1 complex of random fractionally deuterated ^{13}C,^{15}N-labeled protein with unlabeled RNA on a Bruker AVANCE 750 MHz spectrometer at 298 K. The heteronuclear ^{1}H-^{15}N values were measured with reference and NOE experiments recorded in interleaved fashion, employing water flip-back pulses. Heteronuclear NOE values are reported as the ratio of peak heights in paired spectra collected with and without an initial period (4 s) of proton saturation during the 5-s recycle delay. ^{15}N T1 and T2 values were measured using TROSY-based pseudo-3D experiments employing flip-back pulses and gradient selection (Zhu et al. 2000). T1 spectra were acquired with delays, T = 40, 150, 300, 500, 900, 1500, 2200 and 3000 ms, T2 spectra were acquired with CPMG delays, T = 17, 34, 51, 68, 103, 137, 188, and 239 ms. ^{15}N T1 and T2 values were extracted by plotting the decay of HN volumes and fitting the curves with standard exponential equations using the nlinLS lineshape fitting program within the NMRPipe package (Delaglio et al. 1995). Residual dipolar coupling (RDC) restraints were extracted using ^{1}H-^{15}N TROSY run on a fully protonated ^{15}N-labeled (1:1.1) tandem RRM-CUUAUUUG complex in NMR buffer

(isotropic dataset) and NMR buffer mixed with 4.2% C12E5 polyethylene glycol / hexanol medium (Rückert and Otting 2000) (anisotropic dataset). RDCs were derived by subtracting the isotropic from anisotropic ^{1}H chemical shift differences between TROSY and anti-TROSY spectra recorded in an interleaved manner. Only un-overlapped peaks were analyzed and RDC restraints were employed only for structured residues with ^{15}N het-NOE values larger than 0.6. The RDC rhombicity and anisotropy components were determined in CYANA by grid-search using an initial protein structure and further refined in subsequent structure calculations.

Structure calculation and refinement

Intramolecular protein distance restraints were derived from 3D ^{1}H-^{15}N NOESY (t_{mix} = 80ms) and 3D ^{1}H-^{13}C HMQC-NOESY (t_{mix} = 70 ms), 3D ^{1}H-^{13}C HSQC-aroNOESY (t_{mix} = 80 ms) and 2D NOESY (t_{mix} = 80 ms). The protein resonance assignments of the tandem RRM-CUUAUUUG complex and a list of manually assigned protein core NOEs were used as input for automatic peak picking and NOESY assignment using ATNOSCANDID (Herrmann et al. 2002b, 2002a) in a two-step procedure. First intra-RRM NOEs were assigned by including only resonance assignments for one individual RRM in two separate runs. Second an ATNOSCANDID NOE assignment was performed using all resonance assignments. In this run a list of upper limit distance restraints combining the restraints obtained in the runs performed with assignments for the individual RRMs was included. This procedure was found to be necessary to obtain the correct global topology for the two RRMs. The resulting peak lists were then checked and supplemented manually with additional picked peaks and several critical manual NOE assignments. The optimized NOESY peak lists from this procedure were re-assigned with the NOEASSIGN module of CYANA 3.96 (Güntert and Buchner 2015) while iteratively adjusting and keeping key manual assignments fixed during iterative refinement of the structure. Intra-protein hydrogen bonds were identified for HN resonances which were protected during hydrogen-deuterium exchange by reference to intermediate structures and added as restraints in further rounds of structure calculation. Following the determination of the protein structure in the bound state the structure of the complex was determined. Intra-RNA and intermolecular NOESY peaks were picked and assigned manually and calibrated using known distances of H5-H6 cross-peaks of pyrimidines. In structure calculations including the RNA, unambiguous intermolecular NOEs were included first for initial

positioning of the nucleotides. Intermolecular NOEs with ambiguous assignments were then included as ambiguous restraints in CYANA and assigned unambiguously based on preliminary calculations. To further confirm the intermolecular restraints, we back-calculated short intermolecular distances from our final structures and inspected the spectra for completeness of intermolecular NOEs. Final structure calculations in CYANA included intra-protein, intra-RNA and intermolecular NOEs, protein dihedral backbone restraints, intra protein hydrogen bond restraints, and restraints for sugar pucker and syn or anti conformations identified from NOE patterns of H6 or H8 resonances. Protein dihedral backbone restraints derived from TALOS+ (Shen et al. 2009) and additional manually defined β-hairpin turn restraints were used for the N-terminal β-hairpin extension. In the final structure calculation 500 structures were calculated with CYANA and the 50 lowest energy structures were selected for refinement with the SANDER module of AMBER12 (Case et al. 2012) using the ff12SB force field with implicit solvent and 20 were selected based on the criteria of lowest amber energy and lowest intermolecular restraint violations.

Isothermal Titration Calorimetry

ITC experiments were performed on a VP-ITC microcalorimeter (Microcal). Protein was dialyzed in ITC buffer 20mM MES pH 6.6, 100 mM NaCl, 1mM 2-mercaptoethanol. RNA (100-400 μM) was dissolved in NMR buffer and titrated into protein (3.5-11 μM) in 2 μL followed by 8 μL (RRM12) or 10 μL (RRM1) every 300 s at 25 ^{o}C with a stirring rate of 307 rpm. Raw data was analyzed in Origin 7.0.

Data deposition

The coordinates for the structural models of DND1-RRM12:CUUAUUUG have been deposited in the Protein Data Bank under ID code PDB xxx, and the assignments have been deposited at BMRB under ID code BMRB: xxx.