

Structural basis for the recognition of transiently structured AU-rich elements by Roquin

Oliver Binas^{1†}, Jan-Niklas Tants^{2†}, Stephen A. Peter³, Robert Janowski⁴, Elena Davydova⁴, Johannes Braun^{3#}, Dierk Niessing^{4,5}, Harald Schwalbe¹, Julia E. Weigand^{3,*} and Andreas Schlundt^{2,*}

¹ Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt and Center for Biomolecular Magnetic Resonance (BMRZ), 60438 Frankfurt, Germany

² Institute for Molecular Biosciences, Goethe University Frankfurt and Center for Biomolecular Magnetic Resonance (BMRZ), 60438 Frankfurt, Germany

³ Department of Biology, Technical University of Darmstadt, Darmstadt 64287, Germany

⁴ Institute of Structural Biology, Helmholtz-Zentrum München, 85764, Neuherberg, Germany

⁵ Institute of Pharmaceutical Biotechnology, Ulm University, 89081, Ulm, Germany

* To whom correspondence should be addressed. Tel: +49 69 798 29699; Fax: +49 69 798 29225; Email: schlundt@bio.uni-frankfurt.de

Correspondence may also be addressed to Tel: +49 6151 16 22005; Fax: +49 6151 16 22003;

Email: julia.weigand@tu-darmstadt.de

† Joint Authors

#Present Address: Center for Thrombosis and Homeostasis, Johannes Gutenberg University Medical Center, Mainz 55131, Germany

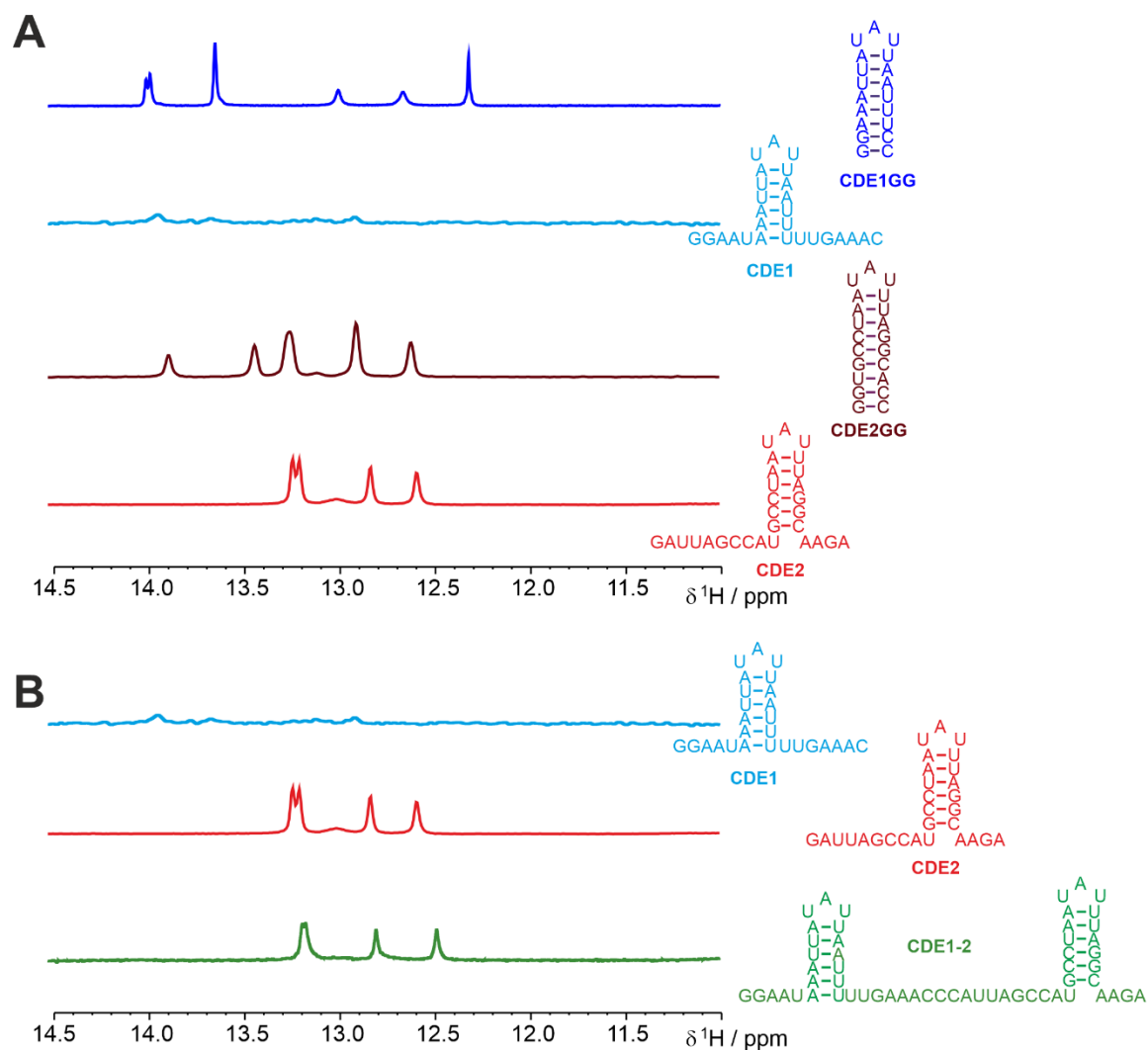
This Supplementary information comprises three tables, eight figures and references. Due to its size and complexity, Supplementary Table 3 will be provided and uploaded as a separate file.

Supplementary Table 1: CDE sequences. Mutations with respect to the wild type are shown in red.

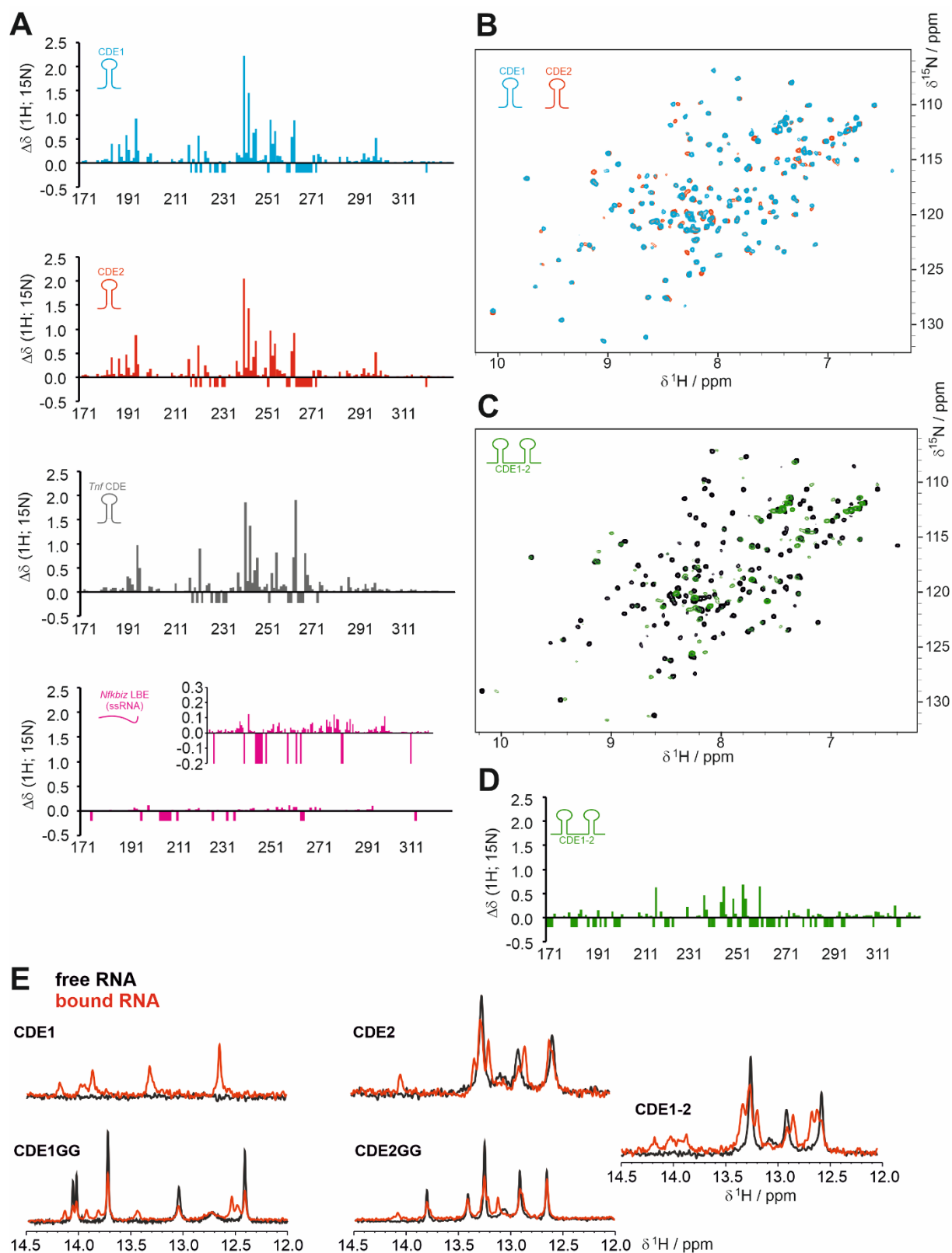
Construct	Sequence 5' → 3'
CDE1	GGAAUAAAUAUAUUAUUUUUGAAAC
CDE2	GAUUAGCCAUGCCUAAUAUUUAGGCAAGA
CDE1GG	GGAAAUUAUAUUAUUUCC
CDE2GG	GGUGCCUAAUAUUUAGGCACC
tandem CDE1-2	GGAAUAAAUAUAUUAUUUUUGAAACCAUAGCCAUGGCUAAUAUUUAGGCAAGA
Tnf CDE	ACAUGUUUUCUGUGAAAACGGAG
Luciferase reporter gene assay	
UCP3_wt	CAAGAUGGAAAAUAAAUAUAUUAUUUUUGAAACCAUAGGCAUGCCUAAUAUUUA GGCAAG
UCP3_NMR	CAAGAUGGAAAAUAAAUAUAUUAUUUUUGAAACCAUAGC CA UGCCUAAUAUUUA GGCAAG
RNA affinity purification	
UCP3_wt	GGCAAGAUGGAAAAUAAAUAUAUUAUUUUUGAAACCAUAGGCAUGCCUAAUAUU UAGGCAAG
UCP3_CDE1_mut	GGCAAGAUGGAAAAUA CCGGCGC UUAUUUUUGAAACCAUAGGCAUGCCUAAUAUU UAGGCAAG
UCP3_CDE2_mut	GGCAAGAUGGAAAAUAAAUAUAUUAUUUUUGAAACCAUAGGCAUGCCUAAUA GG G AGGCAAG
UCP3_double_mut	GGCAAGAUGGAAAAUA CCGGCGC UUAUUUUUGAAACCAUAGGCAUGCCUAAUA GG G AGGCAAG

Supplementary Table 2: Roquin ROQ domain and AUF1 RRM domains as used in this study, i.e. after TEV cleavage from His₆-Thioredoxin-fusion proteins.

Construct	Sequence N → C
Roquin_ROQ	GAMALQHQNPPQLSSNLWAAVRARGCQFLGPAMQEEALKLVLLALEDGSAL SRKVLVLFVVQRLEPRFPQASKTSIGHVVQLLYRASCVKVTKRDEDSSLMQLK EEFRTYEALRREHDSQIVQIAMEAGLRIAPDQWSSLLYGDQSHKSHMQSIIDKL QT
AUF1_RRM2	GKTKEPVKKIFVGGLSPDTPEEKIREYFGGFGEVESIELPMDNKTNKRRG FCFITFKEEEPVKKIMEKKYHNVGLSKCEIKVAMSKE
AUF1_RRM1-2	GASKNEEDEGKMFIGGLSWDTTKDLKDYFSKFGEVVDCTLKLDPITGRSRGF GFVLFKESESVDKVMQKEHKLNGKVIDPKRAKAMKTKEPVKKIFVGGLSPDT PEEKIREYFGGFGEVESIELPMDNKTNKRRGFCFITFKEEEPVKKIMEKKYHNV GLSKCEIKVAMSKE

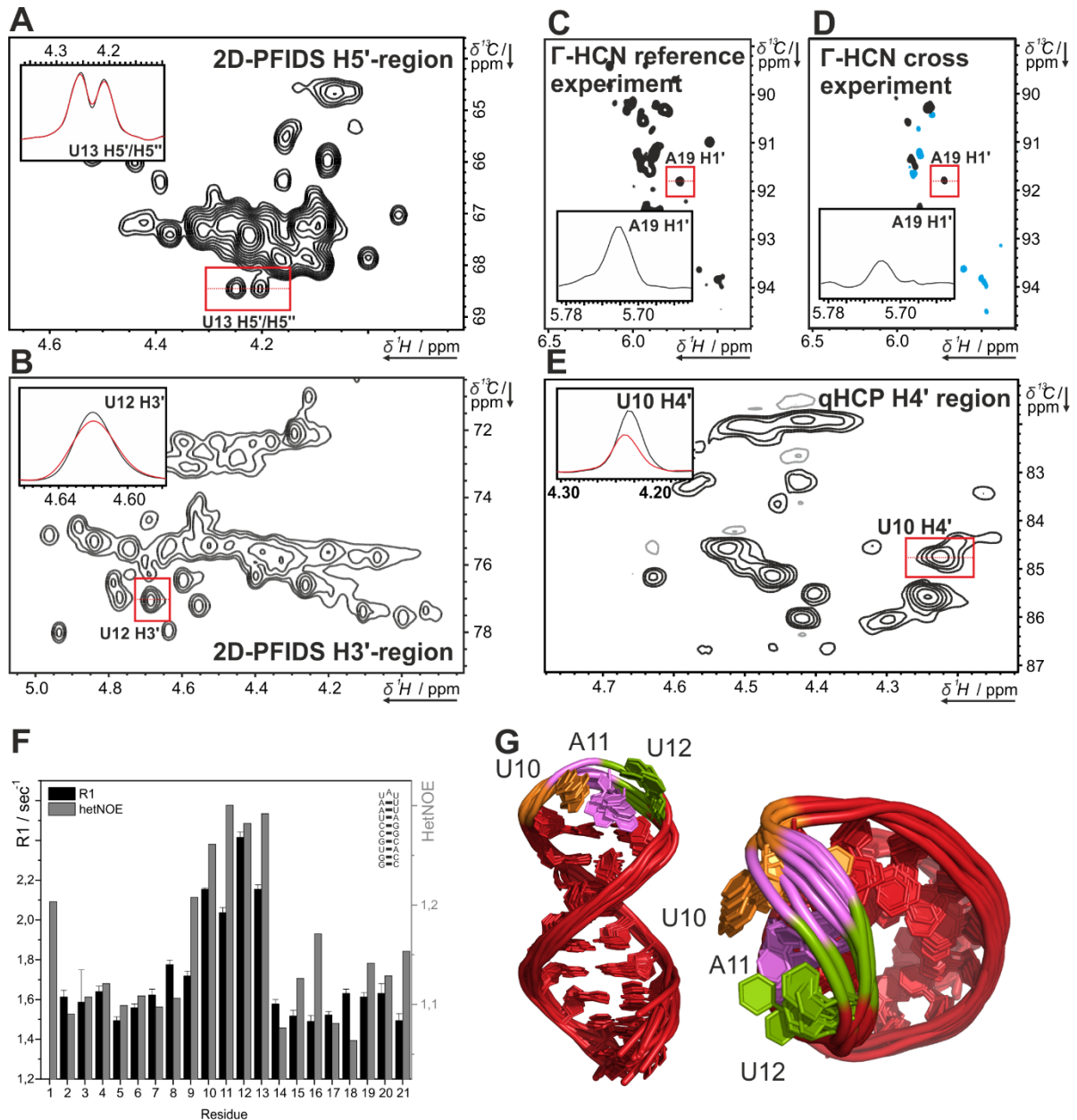


Supplementary Figure S1. Overview of all *UCP3* RNA variants used in this study. **(A)** ^1H NMR spectra highlighting the imino proton region of *UCP3* CDEs 1 and 2, either the wild type sequence or as stabilized (GG) form. The sequences and expected folds are shown next to the spectra and in **Figure 1B**. **(B)** Comparison of $1\text{D-}^1\text{H}$ NMR spectra highlighting the imino proton region of *UCP3* wild type CDEs 1 and 2 in the isolated form and the two of them when arranged as tandem CDE1-2 RNA (bottom panel).

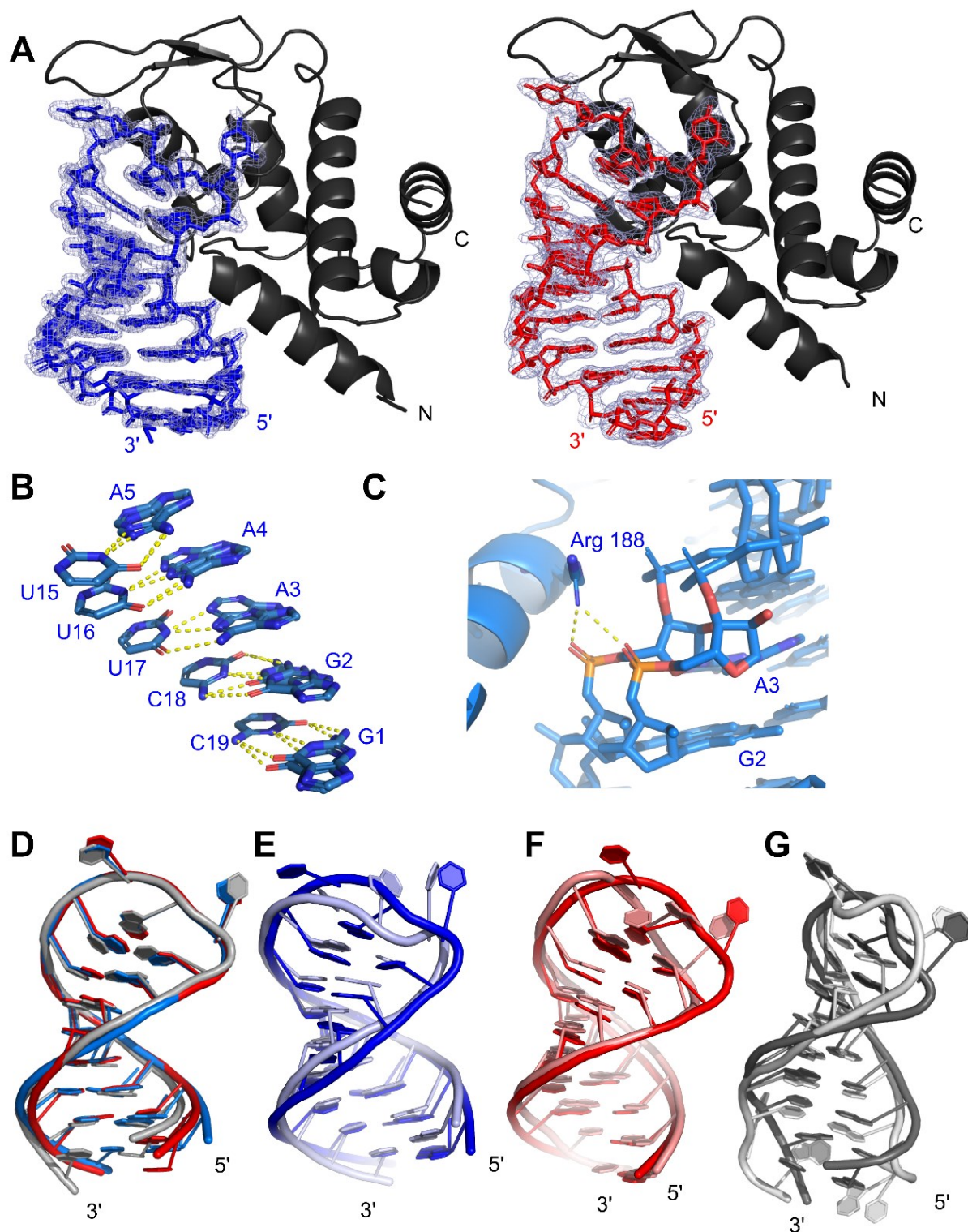


Supplementary Figure S2. The *UCP3* mRNA stem-loops are bound by the Roquin ROQ domain in a CDE-like manner. **(A)** Comparison of bar plots showing combined chemical shift perturbations (CSPs) against the ROQ domain sequence calculated from free vs. CDE-bound protein. The respective RNAs are shown as insets. The CSPs of the *UCP3* wild type CDE1 and CDE2 are derived from spectra shown in **Figures 2C and D**. The *Tnf* CDE CSP plot is adjusted from (1). For comparison, the panel also shows a CSP plot of the ROQ domain when titrated with the 15-mer ssRNA (LBE) found in the 3'-UTR of *Nfkbiz*

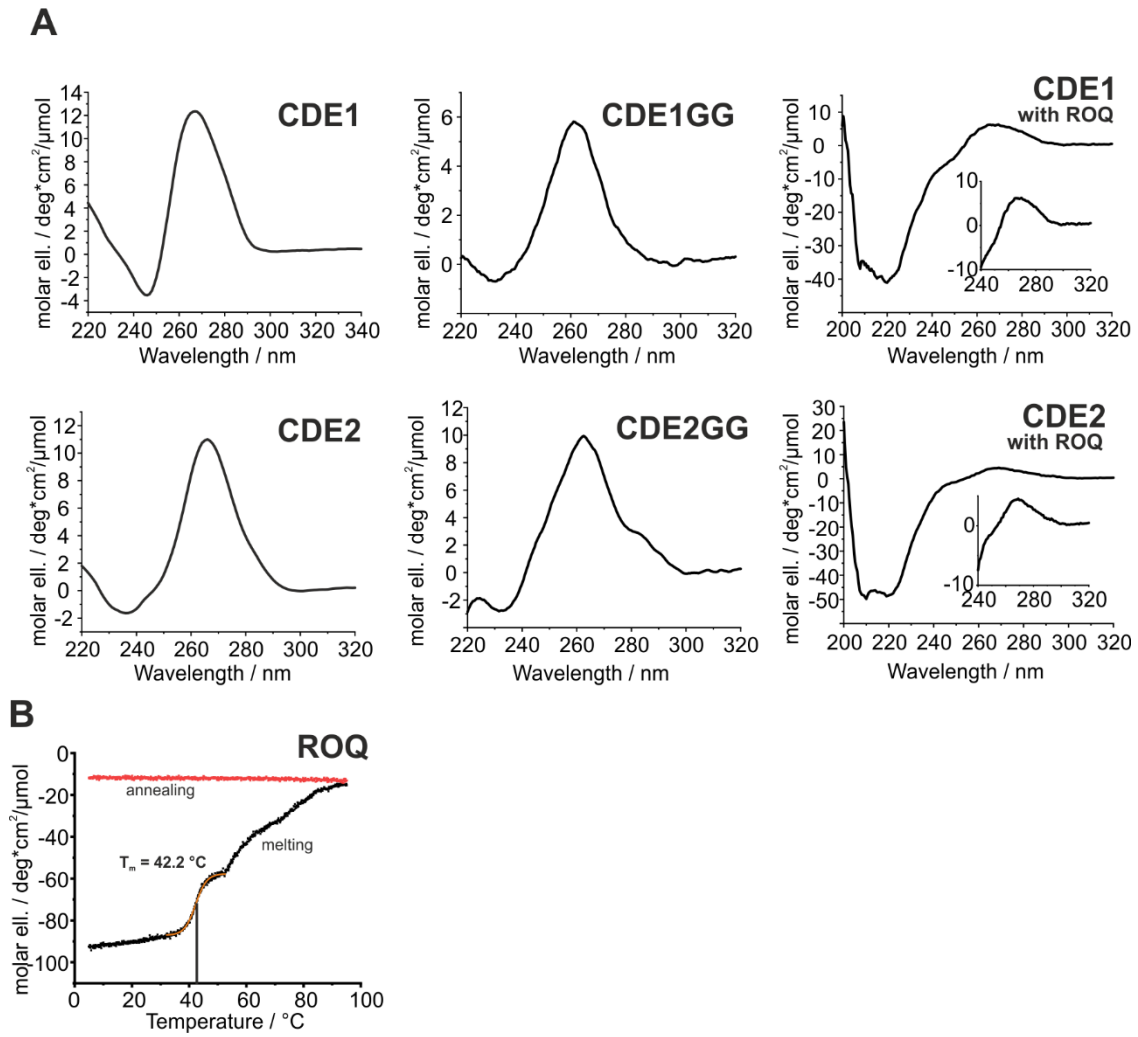
and is described with spectra in (2). Negative bars indicate residues with missing assignments in at least one of the two spectra or prolines. CSP patterns are highly similar. **(B)** Overlay of ^1H - ^{15}N -HSQC NMR spectra showing the Roquin ROQ domain in complex with either wild type *UCP3* CDE1 or CDE2 (in twofold excess, as shown in **Figures 2C and D**). Spectra of the final complexes show highly comparable chemical shifts indicating similar complexes. **(C)** Overlay of ^1H - ^{15}N -HSQC NMR spectra showing the Roquin ROQ domain alone (black) or in complex with the tandem *UCP3* CDE1-2 (green). **(D)** CSP plot of spectral overlay shown in **(C)**. **(E)** Overlay of ^1H -1D-NMR spectra zoomed into the imino proton region and showing the RNAs as denoted either alone (black) or in complex with equimolar amounts of ROQ domain protein (red).



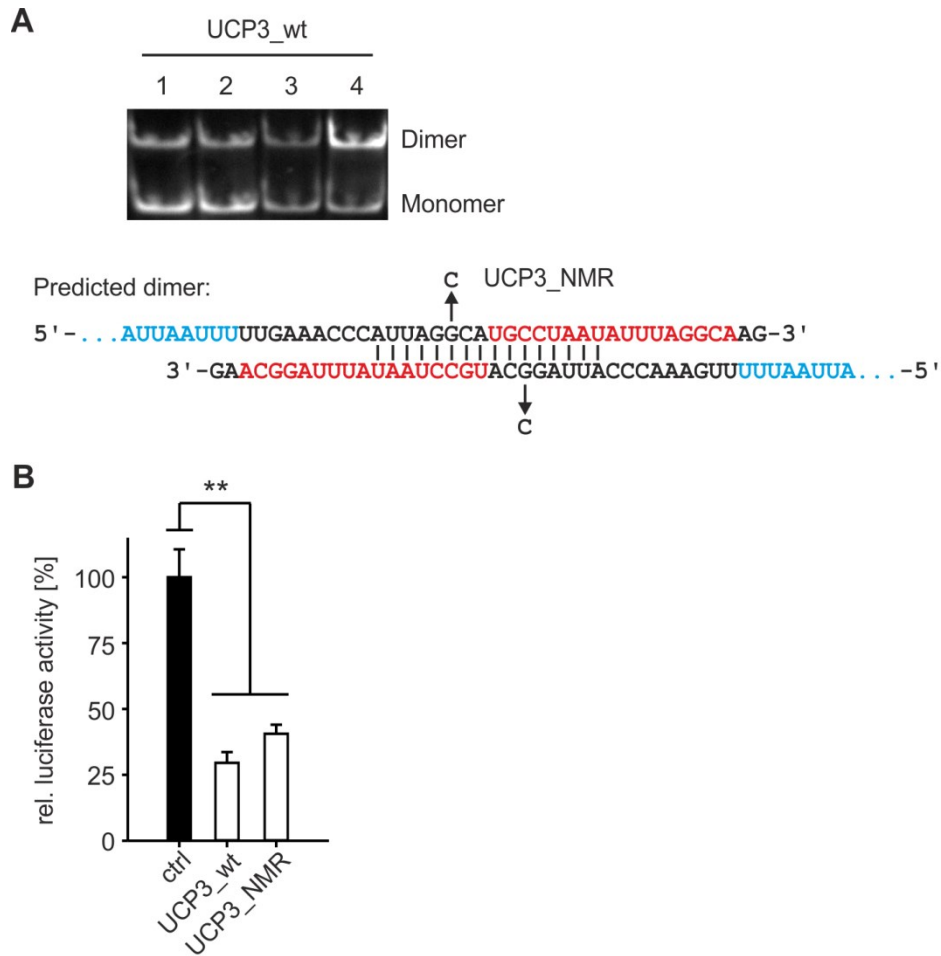
Supplementary Figure S4. NMR experiments for structure determination (dihedral restraints) and dynamics of CDE2GG. **(A and B)** H5' **(A)** and H3' **(B)** regions 2D-PFIDS experiment for determination of H-P coupling constants. A 1D-slice demonstrating the difference between decoupled and coupled spectra is shown in the insert. **(C and D)** Γ -HCN experiment for determination of the glycosidic bond angles via C-H dipolar cross-correlated relaxation rates. Reference **(C)** and cross experiment **(D)** are shown with the insert highlighting the intensity difference in 1D-slices (see reference (3)). **(E)** H4' region of the 2D quantitative HCP experiment for the determination of C-P coupling constants. The insert shows the difference in intensity between the coupled and uncoupled experiment. **(F)** Diagram of spin-lattice relaxation rate and hetNOE of CDE2GG against residue number. Data was measured on sugar C1' nuclei (R1) and aromatic protons (hetNOE) of each residue. **(G)** Unrefined NMR structure of CDE2GG (PDB: 6XXB) Tri-loop residues are labeled in sequential order (orange, magenta and green respectively) to highlight differences in loop geometry.



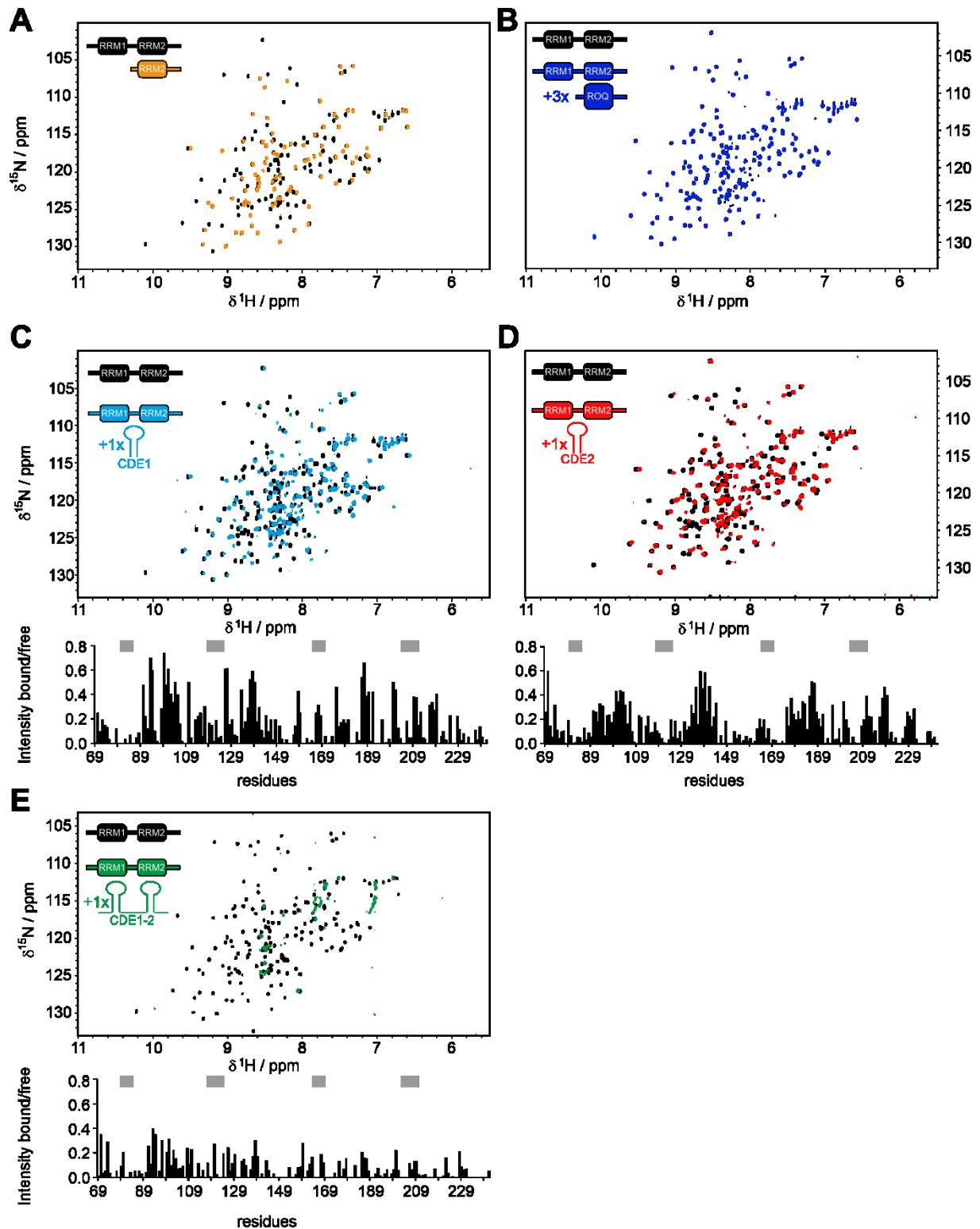
Supplementary Figure S5. Structures of *UCP3* CDE-AREs alone and in complex with ROQ. **(A)** Crystal structures of the ROQ domain in complex with CDE1GG (left) or CDE2GG (right). The RNAs are shown with their 2Fo-Fc electron density maps contoured at +1 sigma. **(B)** H-bond pattern between complementary bases as shown within the two bound conformations of CDE1GG. **(C)** Polar interaction between the side chain of Roquin Arg 188 and the CDE1GG RNA backbone at A3, the latter of which is shown in its two conformers. **(D)** Superimposition of the *UCP3* CDE1GG, CDE2GG and the CDE of *Tnfα* in their ROQ-bound form, respectively. **(E - G)** Superimposition of the free (lighter colour) and ROQ-bound forms (darker colour) of *UCP3* CDE1GG **(E)**, CDE2GG **(F)** and *Tnf* CDE **(G)**.



Supplementary Figure S6. Stabilization of *UCP3* CDEs by the ROQ domain. **(A)** Raw CD spectra of CDE1 and CDE2 with and without addition of ROQ, as well as CDE1GG and CDE2GG alone. Shown is the molar ellipticity as a function of wavelength. deg=degree **(B)** CD melting curve of ROQ showing the melting and annealing curve. Shown is the molar ellipticity as a function of temperature. deg=degree. No annealing is observable due to precipitation of the protein. T_m was obtained from sigmoidal fitting in the region marked orange.



Supplementary Figure S7. Verification of the NMR tandem CDE construct. **(A)** The *UCP3* wild type (*UCP3_wt*) RNA dimerizes at higher concentrations. Native PAGE of 200 ng *UCP3_wt* folded in structure buffer (10 mM Tris-HCl pH 7.0, 10 mM MgCl₂, 100 mM KCl). 1-4) Different folding conditions: 1) 5 min at 95°C, snap-cooling, 2) 5 min at 95°C, slow cooling, 3) 5 min at 65°C, snap-cooling, 4) 5 min at 65°C, slow cooling. The predicted interaction between two *UCP3_wt* RNAs is shown. A G-to-C conversion in the NMR-adjusted tandem CDE (*UCP3_NMR*) inhibits dimerization. **(B)** Repression mediated by the *UCP3* wild type and NMR-adjusted tandem CDE. Wild type tandem CDE (*UCP3_wt*) and the mutant used in NMR experiments (*UCP3_NMR*) were fused to firefly luciferase. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'-UTR sequences. $n = 3$. (**) P -value < 0.01 (Student's t -test, two-tailed, paired).



Supplementary Figure S8. AUF1 RRM1-2 NMR experiments. **(A)** The AUF1 RRMs are independent in solution: Overlay of ^1H - ^{15}N -HSQC spectra of ^{15}N -labelled AUF1 RRM1-2 tandem alone and with ^{15}N -RRM2 alone. **(B)** The AUF1 RRMs do not interact with the Roquin ROQ domain: Overlay of ^1H - ^{15}N -HSQC spectra of ^{15}N -AUF1 RRM1-2 tandem alone and with threefold stoichiometric excess of ROQ domain (unlabelled). **(C-E)** *UCP3* CDE-binding of AUF1: Overlay of ^1H - ^{15}N -HSQC spectra of ^{15}N -labelled AUF1 RRM1-2 tandem alone and with wild type *UCP3* CDE1 **(C)**, CDE2 **(D)** or the tandem CDE1-2 **(E)** at equimolar amounts. The RNAs are unlabelled. For quantification, the intensity ratios of bound over free tandem RRM amide peaks are plotted versus primary sequence below the respective

spectrum. The grey bars indicate the RNP motifs that are by default responsible for RNA-recognition in a canonical RRM and well cover the dips in the respective plots.

Supplementary references

1. Schlundt, A., Heinz, G.A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D. and Sattler, M. (2014) Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nat Struct Mol Biol*, **21**, 671-678.
2. Essig, K., Kronbeck, N., Guimaraes, J.C., Lohs, C., Schlundt, A., Hoffmann, A., Behrens, G., Brenner, S., Kowalska, J., Lopez-Rodriguez, C. *et al.* (2018) Roquin targets mRNAs in a 3'-UTR-specific manner by different modes of regulation. *Nat Commun*, **9**, 3810.
3. Rinnenthal, J., Richter, C., Ferner, J., Duchardt, E. and Schwalbe, H. (2007) Quantitative gamma-HCNCH: determination of the glycosidic torsion angle chi in RNA oligonucleotides from the analysis of CH dipolar cross-correlated relaxation by solution NMR spectroscopy. *J Biomol NMR*, **39**, 17-29.