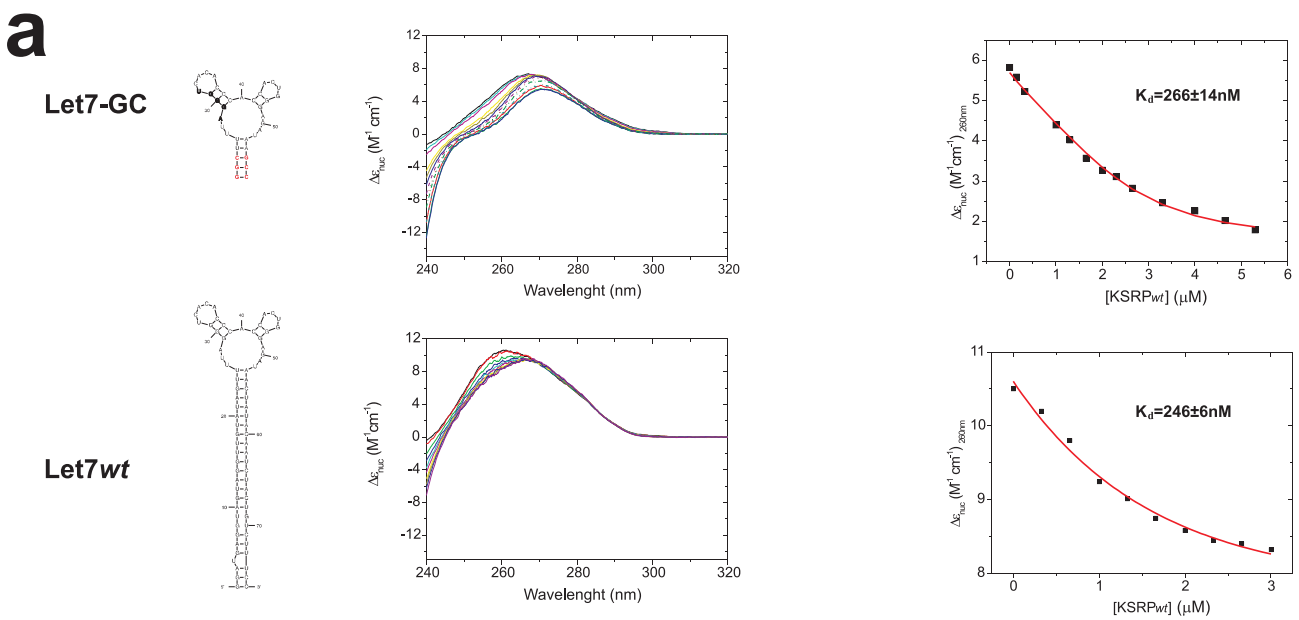
Supplementary Information

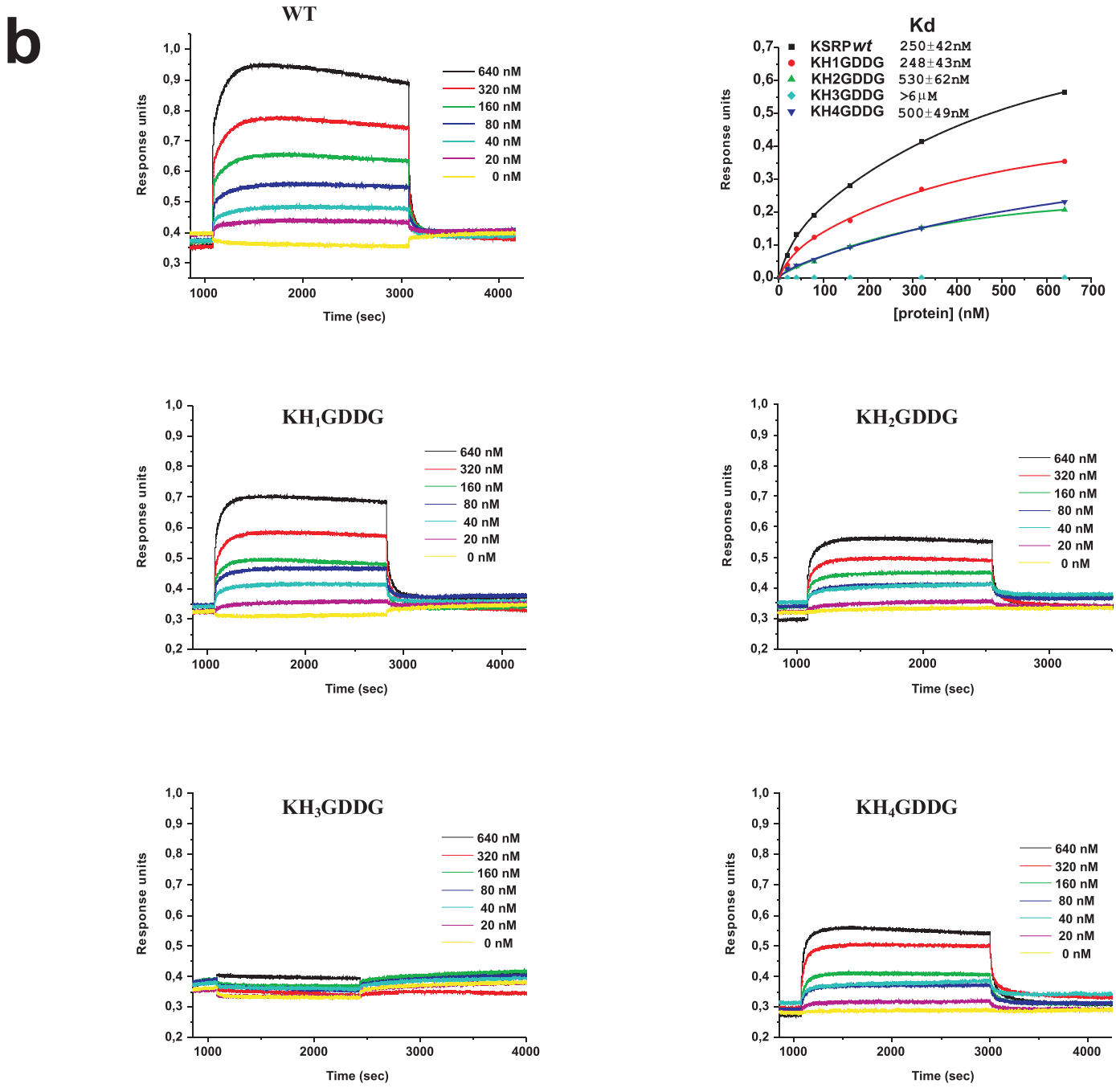
Title:
Non-canonical Guanine recognition mediates KSRP regulation of Let-7 biogenesis

Giuseppe Nicastro[1], María Flor García-Mayoral[1,5], David Hollingworth[1], Geoff Kelly[2], Stephen R. Martin[3], Paola Briata[4], Roberto Gherzi[4,6] and Andres Ramos[1,6]

[1]Molecular Structure Division, MRC National Institute for Medical Research, London, UK.
[2]MRC Biomedical NMR Centre, MRC National Institute for Medical Research, , London, UK.
[3]Physical Biochemistry Division, MRC National Institute for Medical Research, London, UK.
[4]Gene Expression Regulation Laboratory IRCCS AOU San Martino – IST, Genova, Italy.
[5] Departamento de Química Física Biológica, Instituto de Química Física Rocasolano, Consejo Superior de Investigaciones Científicas (CSIC), Madrid 28006, Spain.
[6,] Corresponding authors. aramos@nimr.mrc.ac.uk, rgherzi@ucsd.edu

Supplementary Information comprises:

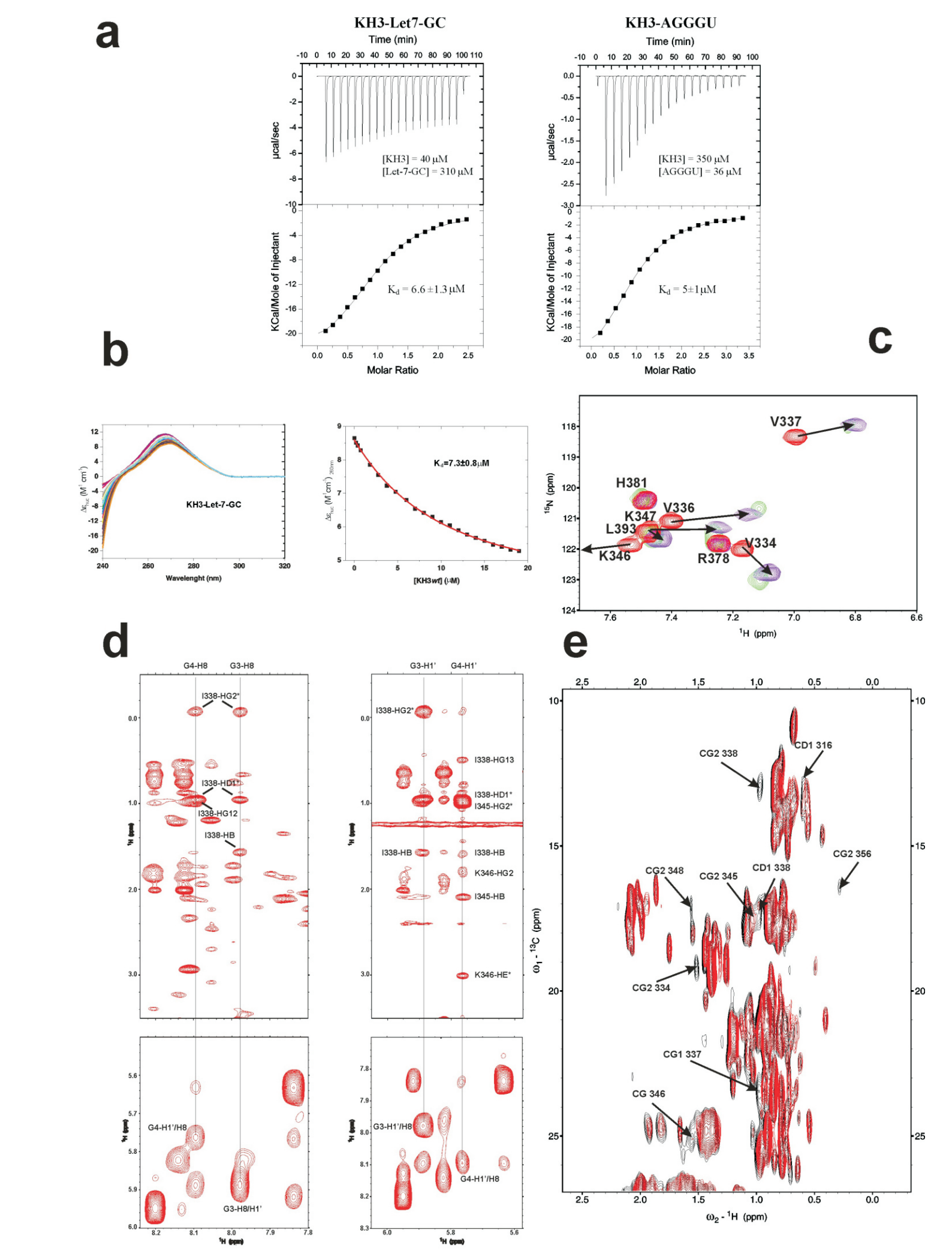




Supplementary Figure 1

Supplementary Figure 1

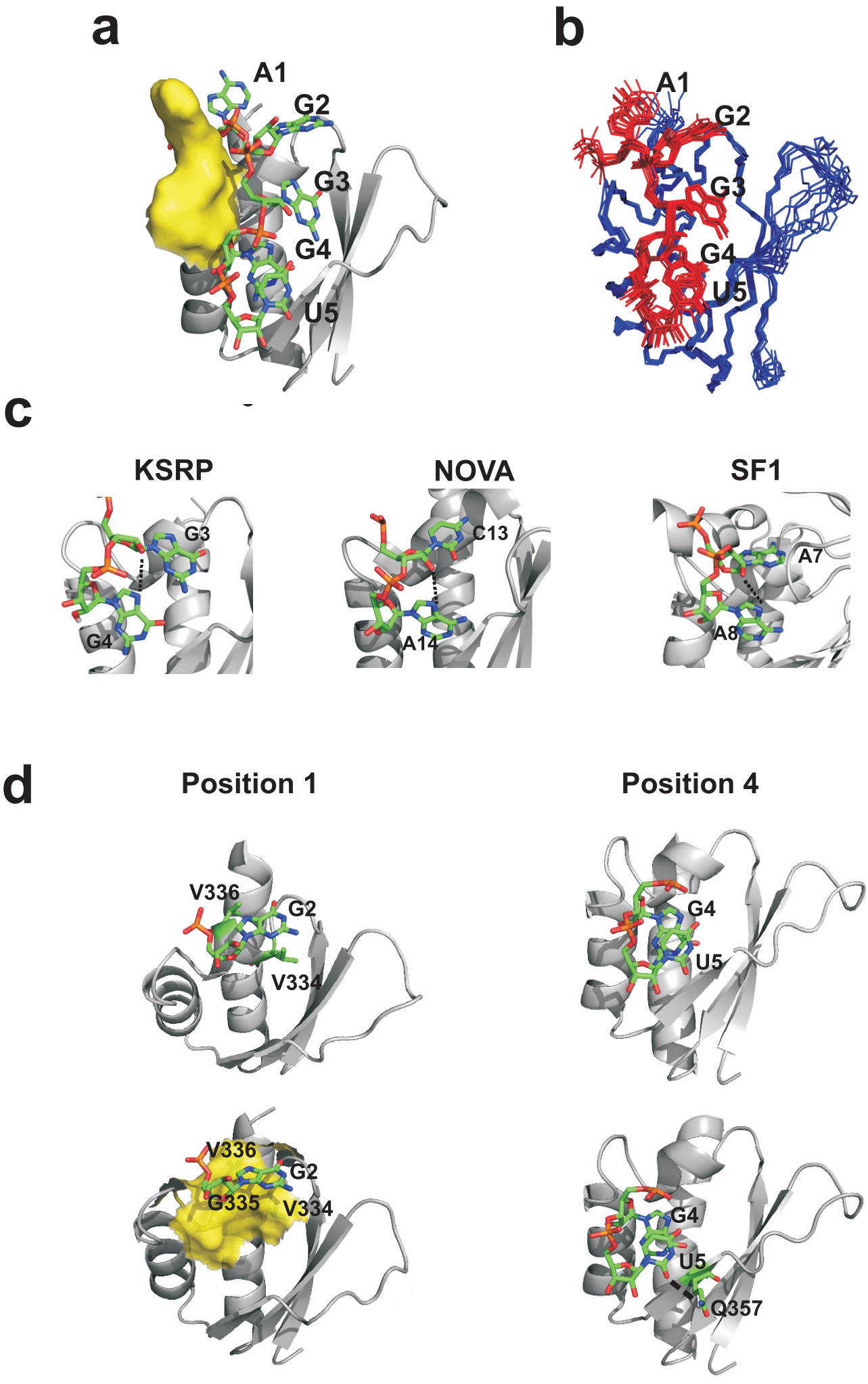
KSRP-pre-Let-7a interaction.
a) The binding affinities of the terminal loop in the context of the wild type pre-Let-7a stem and of a shorter GC-rich stem (Let-7-GC) were compared using CD spectroscopy. Left - The two RNA constructs. Middle - The CD spectra recorded during the titrations of the two RNAs with KSRP. Right - Fitting of the CD signal at 260 nm as a function of protein concentration. Experimental values are represented by black squares, while the fit is displayed as a red line.
b) Four KSRP mutants, each with abolished RNA binding capability in a different KH domain, are used to define the domains' contribution to KSRP / Let-7-GC interaction. In our BLI experiments the biotinylated RNA is immobilised on a streptavidin-coated sensor and the BLI signal is recorded at different protein concentrations for *wt* KSRP and the four GDDG mutant proteins. Top right panel - The RU equilibrium values from these measurements are plotted against protein concentrations. The fit is displayed as a continuous line and values of K_ds ± SD are reported. Other panels – BLI data for the KSRP wild type and four mutants are displayed and protein concentration reported.



Supplementary Figure 2

Supplementary Figure 2

The KH3-AGGGU interaction recapitulates the interaction with the larger pre-let-7 terminal loop.
a) ITC measurements show that KH3 has the same affinity for Let-7-GC (left) and the AGGGU (right) RNAs. ITC titration of KH3 with Let-7-GC RNAs. Baseline corrected calorimetric titration data (top) and the area of each peak after each injection (bottom) are displayed. The fit to a 1:1 binding model is shown as a continuous line. The protein and RNA concentration and the dissociation constants ± SD are reported.
b) Near-UV CD spectra recorded during a titration of Let-7-GC with KSRP KH3 (left). Fitting of the CD signal at 260 nm as a function of protein concentration (right). Experimental values are represented by black squares, while the fit is displayed as a red line. The K_d of the complex is 7.3 μM.
c) Superposition of a representative region of ^{15}N-correlation spectra of ^{15}N-labeled KH3 free (red), in complex with AGGGU (blue) and in complex with Let-7-GC (green). Analogous chemical shift changes are observed for residues of the KH3 domain upon titration with the two RNAs, indicating that the interaction with AGGGU recapitulates the one with the pre-Let-7 terminal loop.
d) Representative inter-molecular NOEs between KH3 and the AGGGU RNA. Displayed is a section of a homonuclear ^{1}H-^{1}H NOESY spectrum of the KH3-AGGGU complex (mixing time 250 μs). Representative intermolecular NOE cross-peaks are labelled.
e) Superposition of methyl region of ^{13}C-correlation spectra of KSRP *wt* alone (black) and in complex with the AGGGU RNA (red). The resonances of KH3 that disappear due to binding and to the connected exchange broadening are labelled.

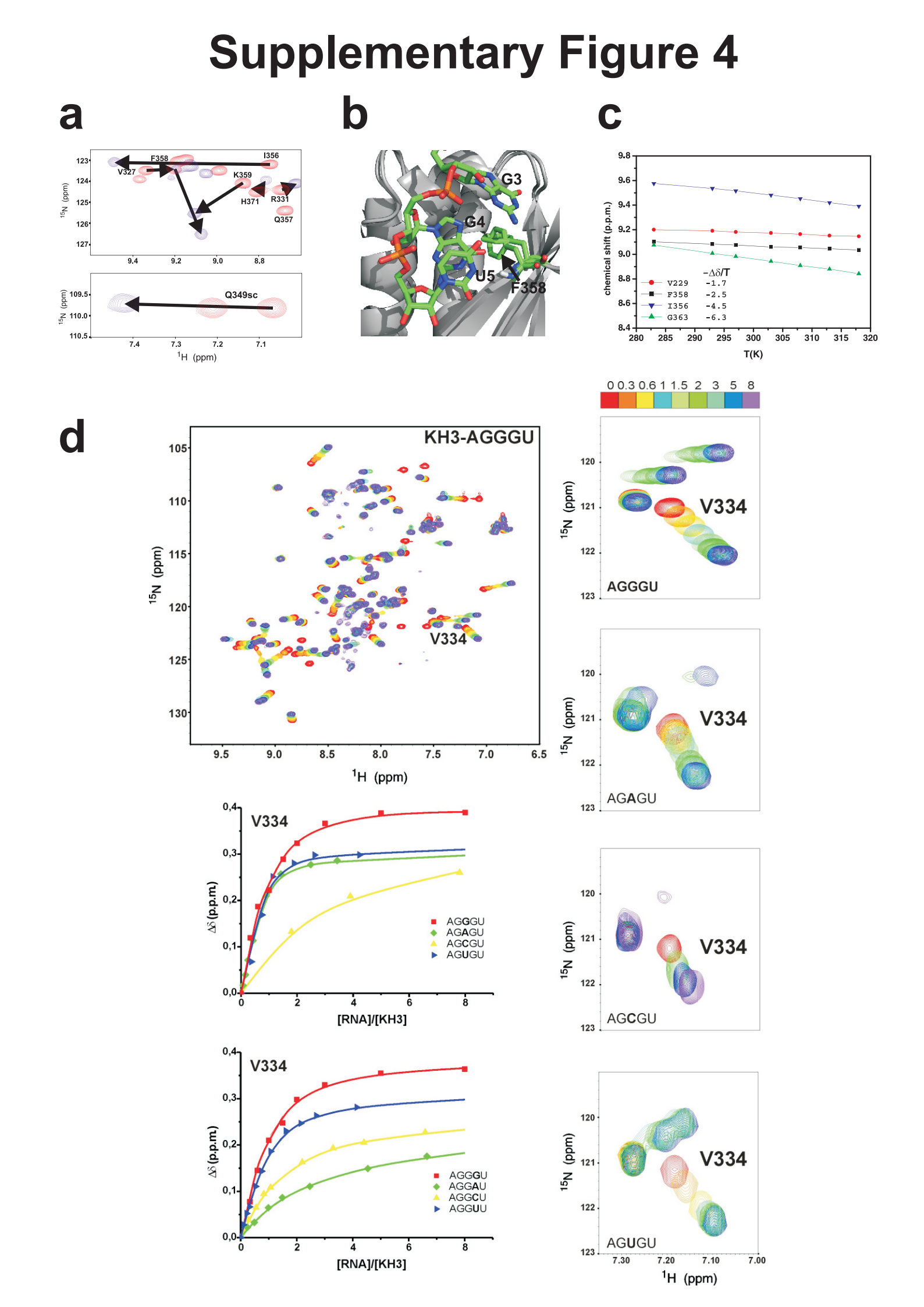


Supplementary Figure 3

Supplementary Figure 3

Structure of the KSRP KH3-AGGGU complex.
a) The protein (lowest energy conformer) is represented as a cartoon (grey), the surface of the GxxG loop is displayed in yellow and the RNA is represented as a stick model. The contact between the RNA phosphate moieties and the GxxG loop orients the nucleobases towards the hydrophobic groove, allowing the specific recognition of the Watson-Crick edges of the bases.
b) Superimposition of the 10 lowest energy structures. The protein backbone is in blue, the RNA heavy atoms are in red.
c) An intra-molecular hydrogen bond is observed in a number of KH-RNA structures between the 2'OH of the nucleotide in position 2 and the N9 of the base in position 3. In this image the H-bond is displayed on the structures of the KSRP KH3, Nova KH3 and SF1 complexes. In KSRP KH3 the presence of the H-bond is supported by a 10-fold decrease in affinity upon mutation of the 2'OH to H (Supplementary Table 1).
d) Interactions of G2 and U5 in the KH3-AGGGU complex. The position of G2 is well defined in our structure. However, contrarily to what observed for G3 and G4, the Watson-Crick edge of the base is not engaged in H-bonds with protein moieties (top left) but rather sits on a hydrophobic platform formed by the V334, G335 and V336 amino acids in KH3 α1 (left). Such an interaction is consistent with the lack of sequence specificity in position 1 reported by our SIA analysis (García-Mayoral, M.F., Díaz-Moreno, I., Hollingworth, D. & Ramos, A. The sequence selectivity of KSRP explains its flexibility in the recognition of the RNA targets. *Nucleic Acids Res.* **36**, 5290-5296 (2008))
The position of the U5 aromatic is defined mainly by its stacking interaction with the nucleobase of G4, an interaction which is present in all of the reported KH-NA structures. The interaction of U5 with the protein is limited to a small number of contacts with solvent exposed side chains. Although in about 30% of the structures we observe an H-bond between the U5 O2 and the side chain of Q357, similar to the contact observed for a cytosine in the Nova2-RNA complex or for an uracile in the NusA-RNA complex, the presence of a stable H-bond is not backed up by direct experimental evidence. Our SIA data indicate that U and G are favoured in this position, but the degree of selectivity is small. In general, the flexibility of the interacting side chains makes it very difficult to define unambiguous contacts between U5 and the protein.

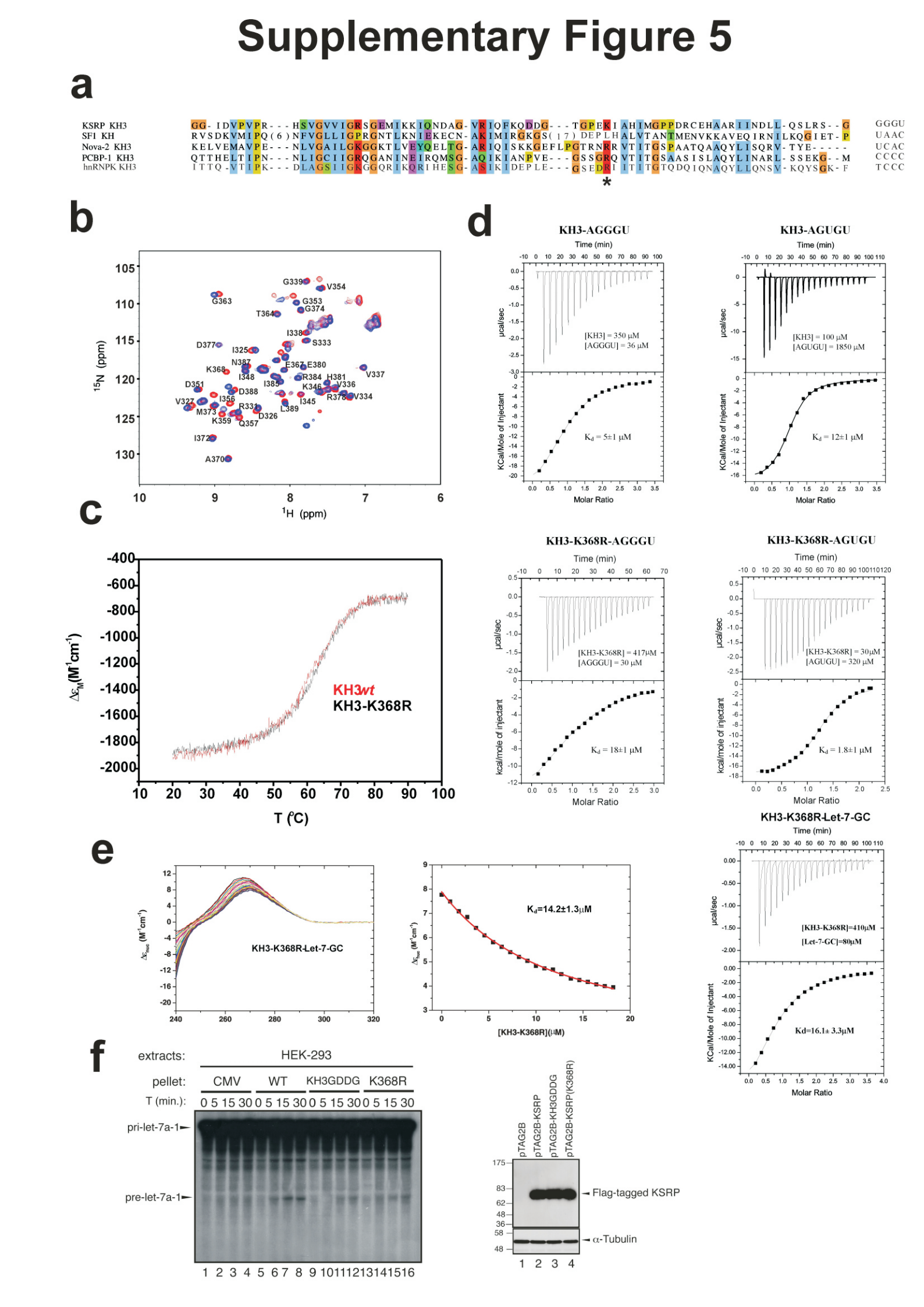
Supplementary Figure 4



Supplementary Figure 4

Specificity of KH3-RNA interactions.
a) Large downfield chemical shift changes support the engagement of the I356 backbone amide and Q349 sidechain NH2 in hydrogen bonds upon RNA binding.
b) Superimposition of the free and AGGGU-bound KH3 structure. The F358 aromatic ring rotates ~40° upon RNA binding resulting in a shielding of the F358 amide and a -1.2ppm shift of its resonance can be estimate with good approximation (Perkin SJ and Dwek RA. Comparison of ring-current shifts calculated from the crystal structure of egg white lysozyme of hen with the proton nuclear magnetic resonance spectrum of lysozyme in solution, *Biochemistry,* **19**, 245-258, (1980)). The bound ring position is indicated by an arrow.
c) The temperature dependence of chemical shifts of amide protons is consistent with the observed H-bonding. We compare here the temperature dependent chemical shift changes of V229, which is engaged in intra-molecular β-strand-β-strand H bonding, of F358, which is engaged in inter-molecular H-bonding, of I356, which is H-bonded by the side chain of Q349 in the complex and of G363, which is in a flexible protein loop.
d) KH3-RNA binding by NMR. Top left - Superimposition of ^{15}N-HSQC correlation spectra recorded during a titration of KSRP KH3 with increasing concentration of AGGGU RNA. Middle and bottom left - Nucleobase specificity in positions 2 and 3. The amide chemical shift changes (Δδ) of a representative resonance (V334) during titrations of KH3 with eight different RNAs are plotted against RNA:protein ratios. The calculated binding isotherm is displayed as a continuous line. Right, the shift of V334 is displayed for the four permutations of the nucleobase in position 2. To more easily to compare the trends observed in the different titrations, the same colour coding (displayed on top) was chosen to report on the relative RNA:protein ratios.

Supplementary Figure 5



Supplementary Figure 5

Design of a KH3 mutant with conserved fold and stability but different RNA binding specificity.
a) Sequence alignment of a selection of KH domains for which protein-NA complexes are available, with the bound NA sequence on the right. K368 is indicated by an asterisk. The aligment has been performed in ClustalX and is displayed in the same program.
b) Superimposition of ^{15}N- correlation spectra of the wild type KH3 (blue) and the K368R mutant (red) indicate that the mutations do not significantly change the protein fold.
c) Superimposition of the 220 nm CD signals recorded during the thermal denaturation of wild type and K368R KSRP KH3. The two melting curves are superimposable.
d) ITC titration of KH3 wild type and K368R mutant with AGGGU, AGUGU RNAs and of K368R mutant with Let-7-GC RNAs. Baseline corrected calorimetric titration data (top) and the area of each peak after each injection (bottom) are displayed. Protein and RNA concentrations and K_ds are reported. The solid line represents the best fit of the data to a single binding site.
e) Near-UV CD spectra recorded during a titration of Let7-GC with KH3 K368R. Fitting of the CD signal at 260 nm as a function of protein concentration (right). Experimental values are represented by black squares, while the fit is displayed as a red line.
f) (right) *In vitro* pri-Let-7a-1 processing assays performed using total cell extracts from HEK-293 cells preincubated with anti-Flag immunoprecipitated extracts from pTAG2B- (lanes 1-4), pTAG2B-KSRP- (lanes 5-8), pTAG2BKH3GDDG- (lanes 9-12) and pTAG2BKSRPK368R- (lanes 13-16) transfected HEK-293 cells. Internally ^{32}P-labeled RNA substrate was added and used to monitor pri-Let-7a-1 processing. (left) Representative immunoblots from total cell extracts from HEK-293 cells transiently transfected with either empty Flag-tagged pTAG2B vector, or pTAG2B-KSRP, pTAG2BKH3GDDG, pTAG2BKSRPK368R hybridized with anti-Flag and anti-alpha-tubulin antibodies.

Supplementary Table 1- Interaction between KSRP KH3 and RNA

RNA sequences	Wild typeKH3	K368R KH3
AGGGU	5 ± 1 (5.9 ± 1)	19 ± 2 (18 ± 1)
AGAGU	8 ± 1	22 ± 3
AGCGU	151 ± 20	18 ± 1
AGUGU	6 ± 1 (12 ± 1)	0.4 ± 0.2 (1.8 ± 1)
AGGAU	100 ± 10	
AGGCU	50 ± 10	
AGGUU	30 ± 1	
AGdGGU	62 ± 3	

K_{d}s, (μM) calculated by NMR and (in parenthesis) ITC.