HnRNP A1/A2 Proteins Assemble onto 7SK snRNA via Context Dependent Interactions

Le Luo, Liang-Yuan Chiu, Andrew Sugarman, Paromita Gupta[b], Silvi Rouskin[b] and Blanton S. Tolbert

[a]Department of Chemistry, Case Western Reserve University
Cleveland, OH 44106

[b]Whitehead Institute for Biomedical Research, Cambridge, MA

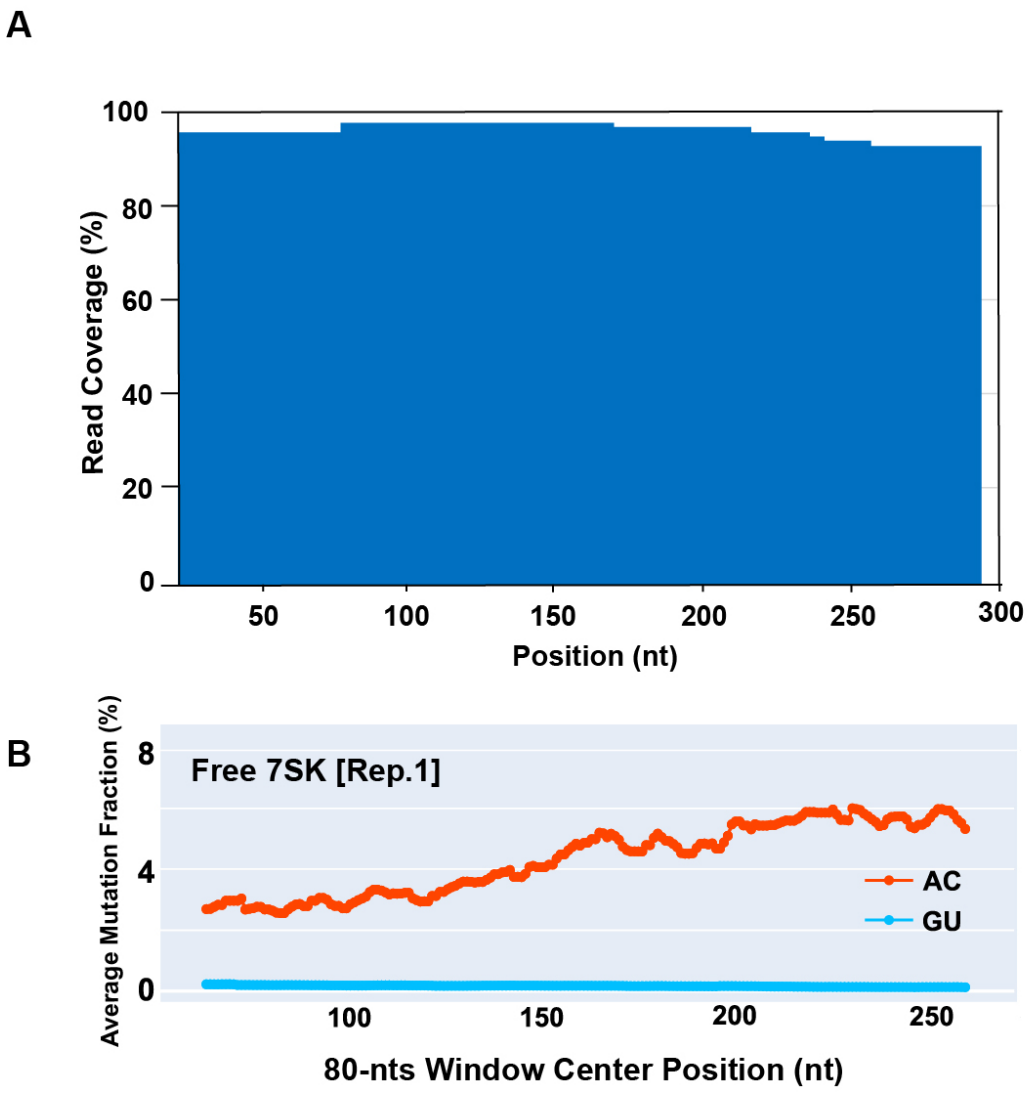

Figure S1. Quality Control of DMS MaPseq replicates of 7SK snRNA. (A) Read coverage of free 7SK snRNA DMS MaPseq (Replicate.1) shows a coverage of 96% or above throughout the target region (nts 20-295). The read coverage aligns all the sequencing reads in the dataset and directly counts the reads aligning to the target region of 7SK snRNA (nts 20-295). All datasets used in the paper have over 90% read coverage throughout the target region without any significant difference of coverage between 5' and 3' ends. (B) Average mutational fraction of free 7SK snRNA shows the moving average of DMS signal (AC, Red) and background mutation error (GU, Blue) with an averaging window of 80-nts (center position ranging from nts 59 to 255). Plotted is the average mutation rate at each center of the 80-nts window with a sliding step of 1-nt. The constant low GU background error of both datasets indicates an even read coverage and high fidelity of the method. The increase in AC mutation towards the 3' end is an intrinsic property of the RNA indicating a higher-degree of unpaired or weakly paired AC nts.

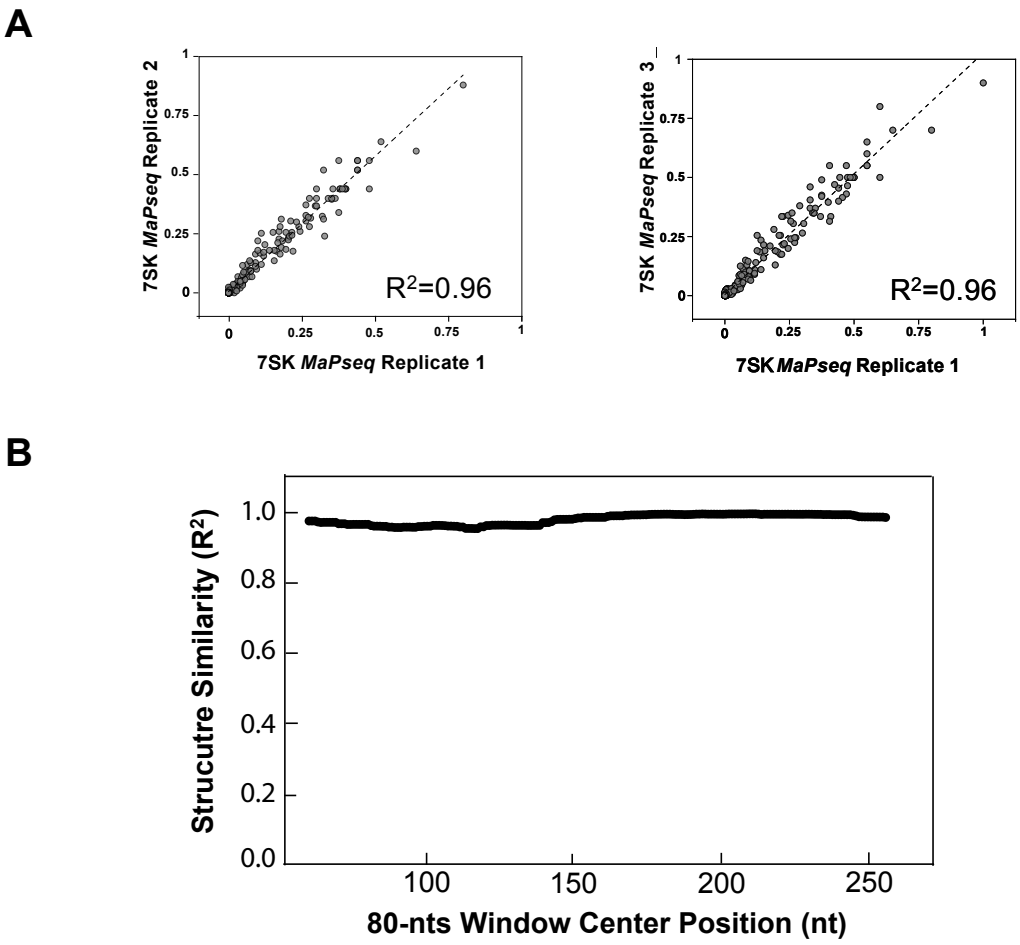

Figure S2. Correlation between Free 7SK snRNA DMS-MaPseq replicates. (A) Linear regression of DMS reactivities of Replicate 2 (Left, Y-axis) and Replicate 3 (Right, Y-axis) to the Replicate 1 shows a high degree of reproducibility with overall R^2 of 0.96. The replicates were performed with 7SK snRNA prepared individually in the same buffer condition (No Mg^{2+}) to ensure reproducibility. (B) Structure similarity (R^2) plot (center position ranging from nts 59 to 255) shows a high degree of agreement between Replicate 2 and Replicate 1, throughout the target region. The structure similarity (R^2) is measured by taking the average linear regression of the DMS reactivity per base with the averaging window of 80-nts and sliding step of 1-nt. The value of each nucleotide represents the average R^2 values of the 80-nt window that the nucleotide is centered. For example, the value of C100 is the average linear regression value of the 80-nt window that spans from G61 to U140. As the DMS reactivity data of 7SK snRNA were collected from nts 20 to 295 (target region, excluding the primer regions of nts 1-19 and nts 296-320), the structure similarity plot ranges from nts 59 to 255.

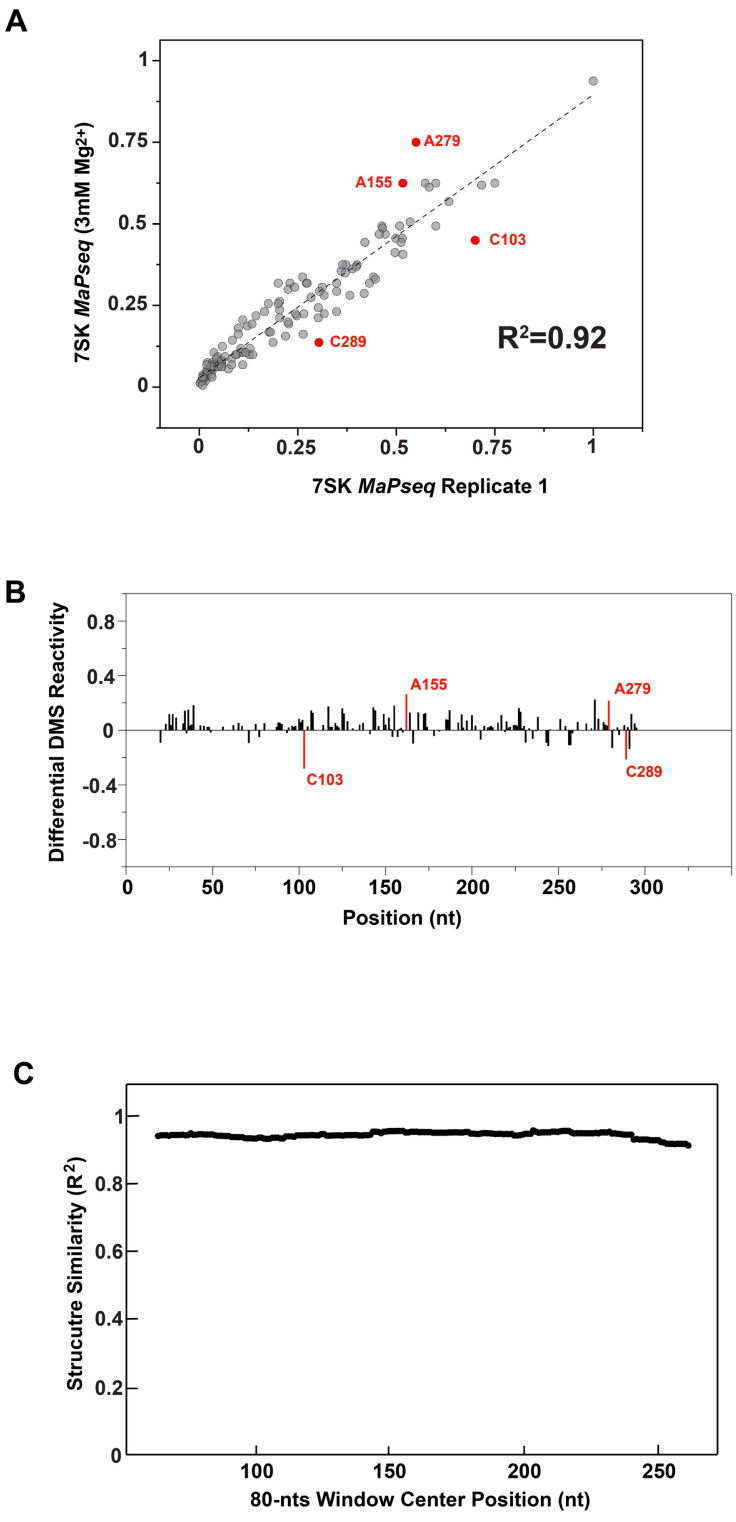

Figure S3. Correlation between Free 7SK DMS-MaPseq with and without Mg^{2+}. (A) Linear regression correlation between DMS reactivities of 7SK snRNA with 3 mM Mg^{2+} (Y-axis) versus the sample without Mg^{2+} (X-axis, Replicate 1) shows a high degree of agreement between the two replicates (R^2=0.92). The data points labeled in red

correspond to the nucleotides with significant differences between the two experiments. Notably, the outliers are not found within a particular region of 7SK snRNA indicating non-specific effects of Mg^{2+}. (B) Differential DMS reactivity (Mg^{2+} minus Replicate 1) of 7SK snRNA with 3mM Mg^{2+} addition shows DMS reactivity changes of each nucleotide of the target region (nts 20-295). The nucleotides labeled in red correspond to the data points labeled in red in Figure S2A. (C). Structure similarity (R^2) of 7SK snRNA +/- 3mM Mg^{2+} shows a high degree of agreement between the two experimental conditions throughout the target region. The value of each nucleotide represents the average R^2 values of the 80-nts window that the nucleotide is centered (ranging from nts 59 to 255).

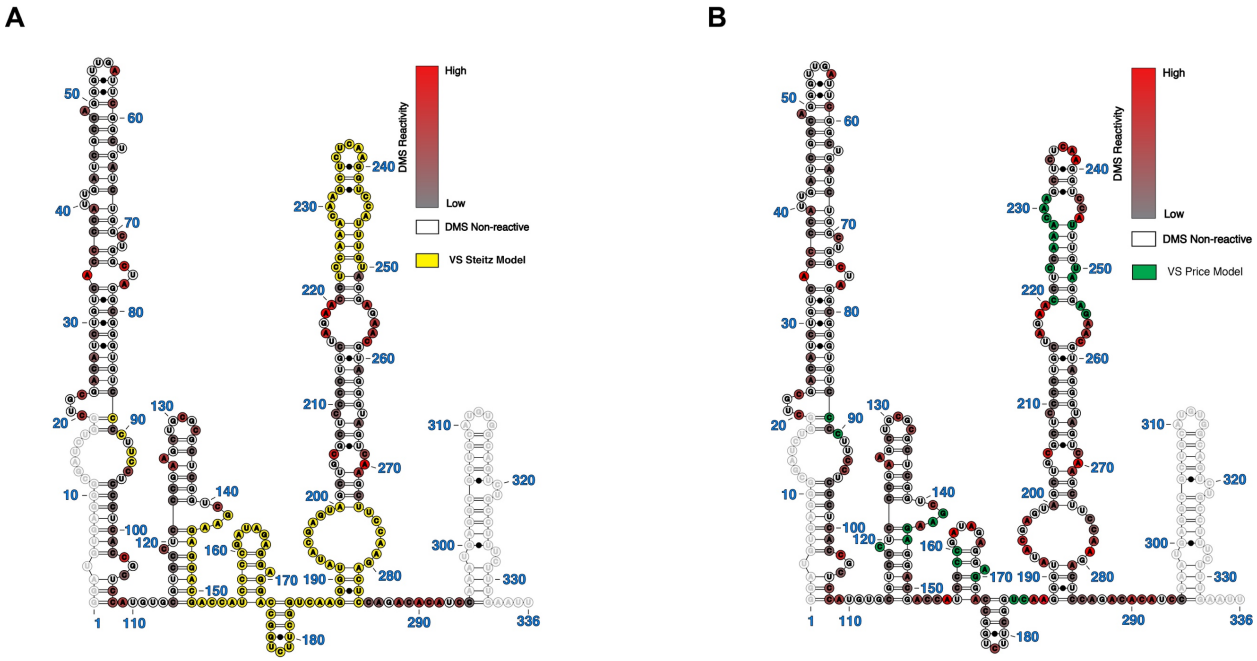

Figure S4. DMS model comparison to the previous secondary structural models. (A) The full-length secondary structure of 7SK snRNA (331-nts, expressed *in vivo*) was determined using chemo-enzymatic probing by the Steitz group. The differences of 7SK snRNA full-length secondary structures between the Steitz model and the DMS model determined here are highlighted in yellow. The full-length secondary structure of 7SK snRNA (332-nts, expressed *in vivo*) was determined using SHAPE-based chemistry by the Price group. The differences of 7SK snRNA full-length secondary structure between the Price model and the DMS model were highlighted in green.

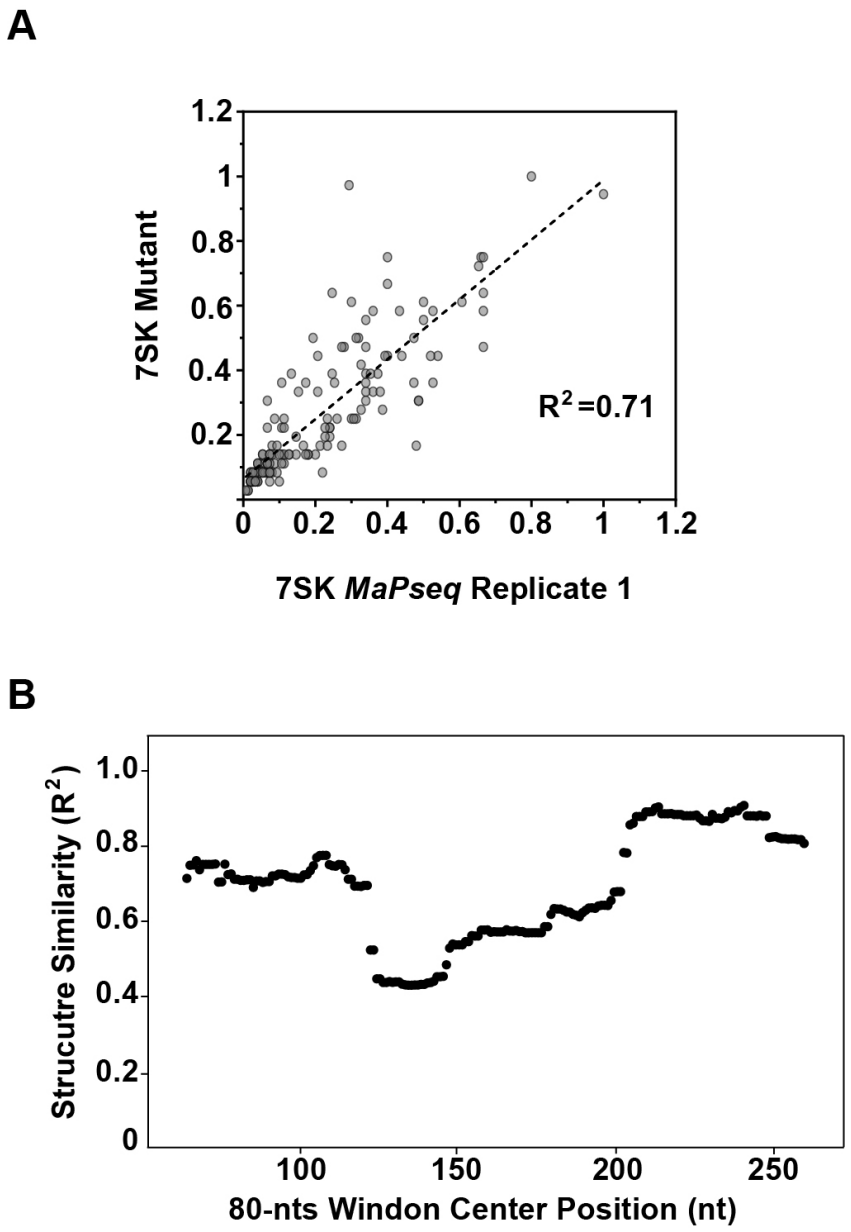

Figure S5, Correlation between wildtype 7SK snRNA and the GAGA mutant construct. (A). Linear regression of DMS reactivities between the GAGA mutant (Y axis) and the wildtype (X axis, Replicate 1) shows a moderately lower level of agreement (R^2 = 0.7). (B) Structure similarity (R^2) plot shows a lower level of agreement between the GAGA mutant and the wildtype 7SK snRNA (Replicate 1) . The value of each nucleotide represents the average R^2 values of the 80-nts window that the nucleotide is centered (ranging from nts 59 to 255).

.

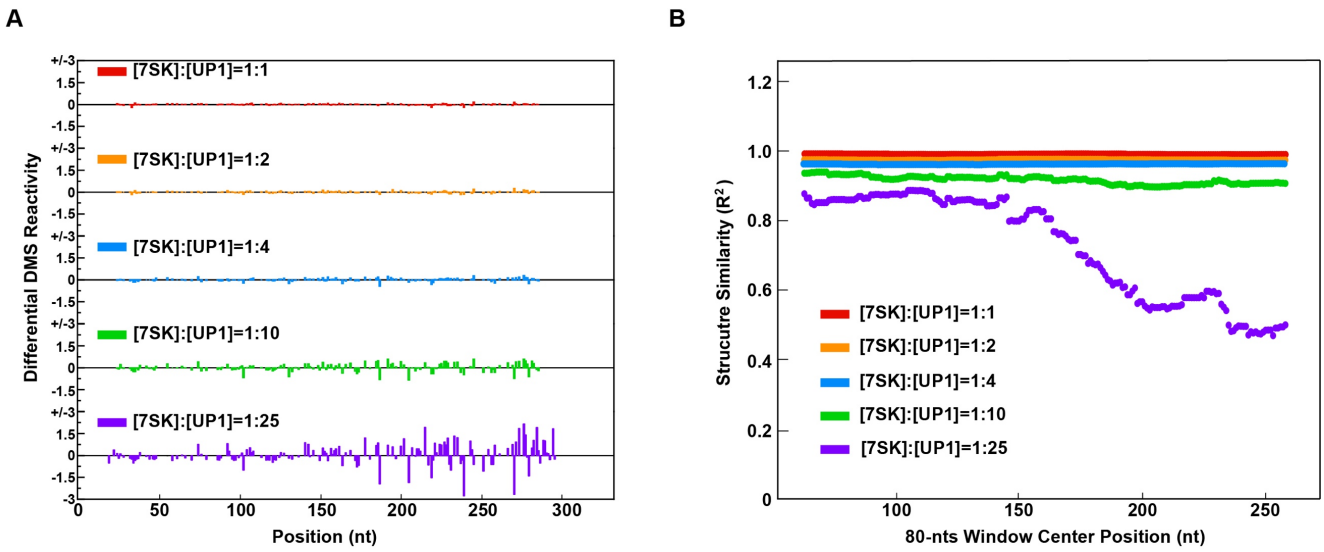

Figure S6. The DMS reactivity change of 7SK snRNA upon UP1 titration. (A). Differential DMS reactivity (7SK-UP1 complexes minus Replicate 1) of 7SK snRNA upon UP1 titration shows the DMS reactivity changes of each nucleotide throughout the target region (nts 20-295) in different [7SK]:[UP1] conditions. As shown, the stepwise increases of the differential DMS reactivity indices of the full length 7SK snRNA (concentration fixed at 100nM) upon UP1 protein titrated in from 1:1 (Red), 1:2 (Orange), 1:4 (Blue), 1:10 (Green) to 1:25 (Purple). (B). Structure similarity (R^2) of each titration condition shows the stepwise changes of 7SK secondary structure upon UP1 titration. The value of each nucleotide represents the average R^2 values of the 80-nts window that the nucleotide is centered (ranging from nts 59 to 255).

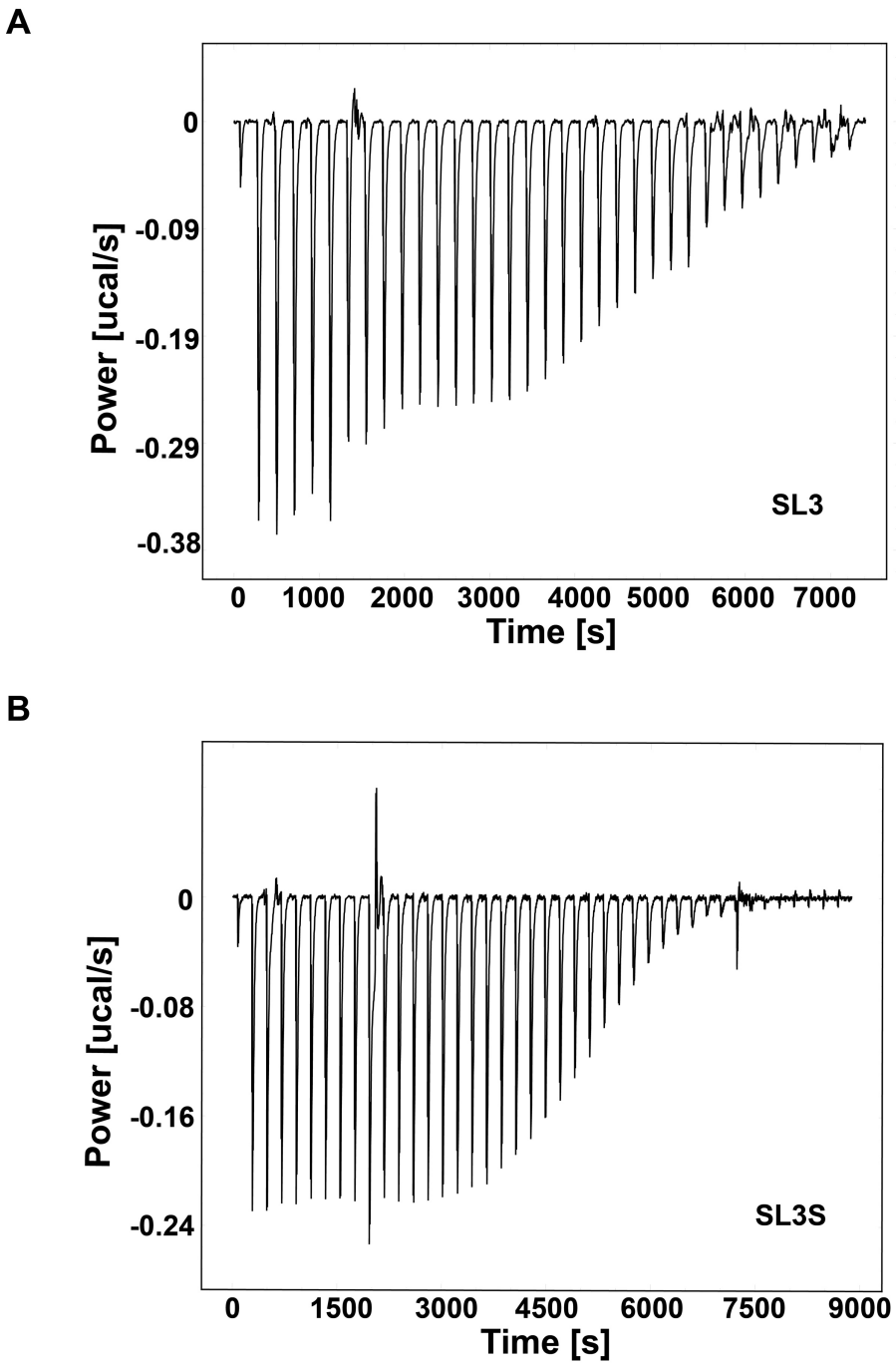

Figure S7. Baseline corrected calorimetric signals for SL3 and SL3S titrated with UP1 protein. (A) Non-integrated calorimetric signals of the SL3-UP1 titration. (B) Non-integrated calorimetric signals of the SL3S-UP1 titration. Data were collected in 10 mM K_2HPO_4, 120 mM KCl, 1 mM TCEP, 0.5 mM of EDTA, pH 6.5 at 298K. The UP1 protein (~80 μM) was titrated into 1.4 ml of 5 μM RNA over 36 injections of 8 μL each.

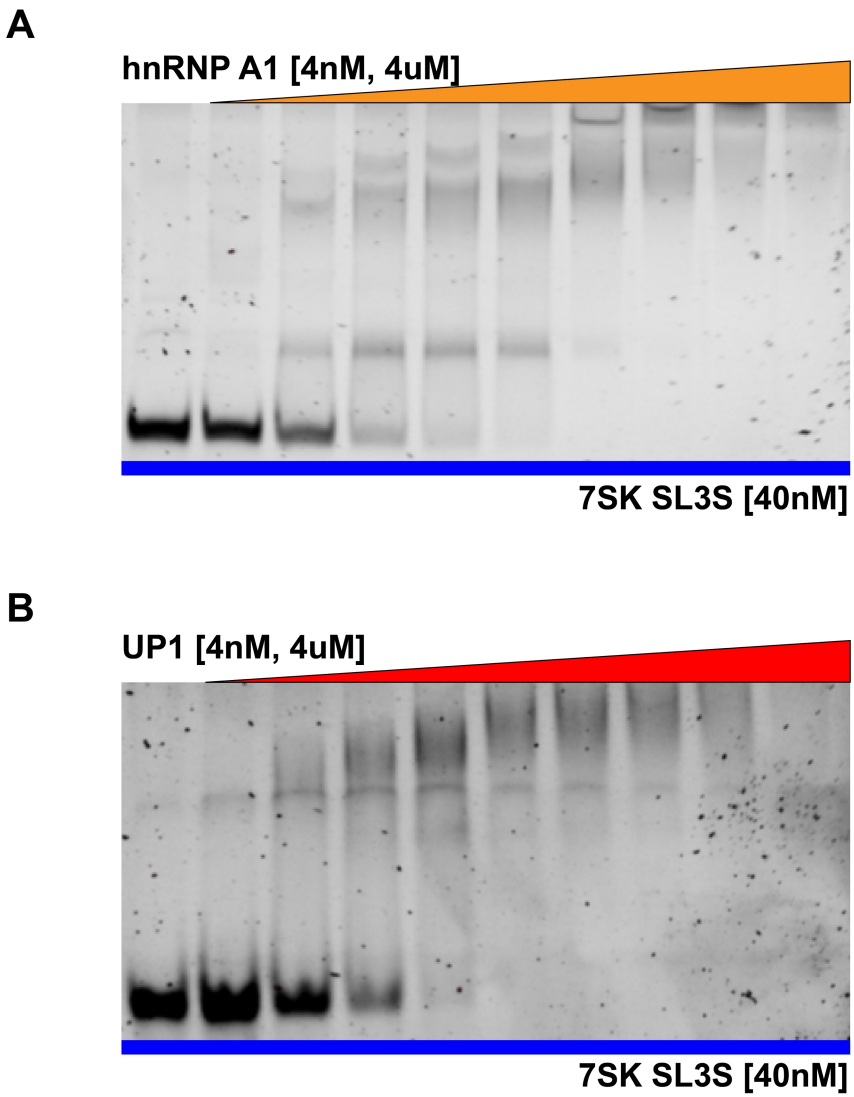

Figure S8. EMSA of HnRNP A1 and UP1 Titration into 7SK snRNA SL3S. (A) HnRNP A1 titrated into 7SK SL3S RNA. (B) UP1 titrated into 7SK SL3S RNA. The concentration of the 7SK SL3S was set constant at 40 nM and the concentration of the protein in each lane are: 0, 4 nM, 40 nM, 80 nM, 160 nM, 320 nM, 640 nM,1 μM, 2 μM, 4 μM.

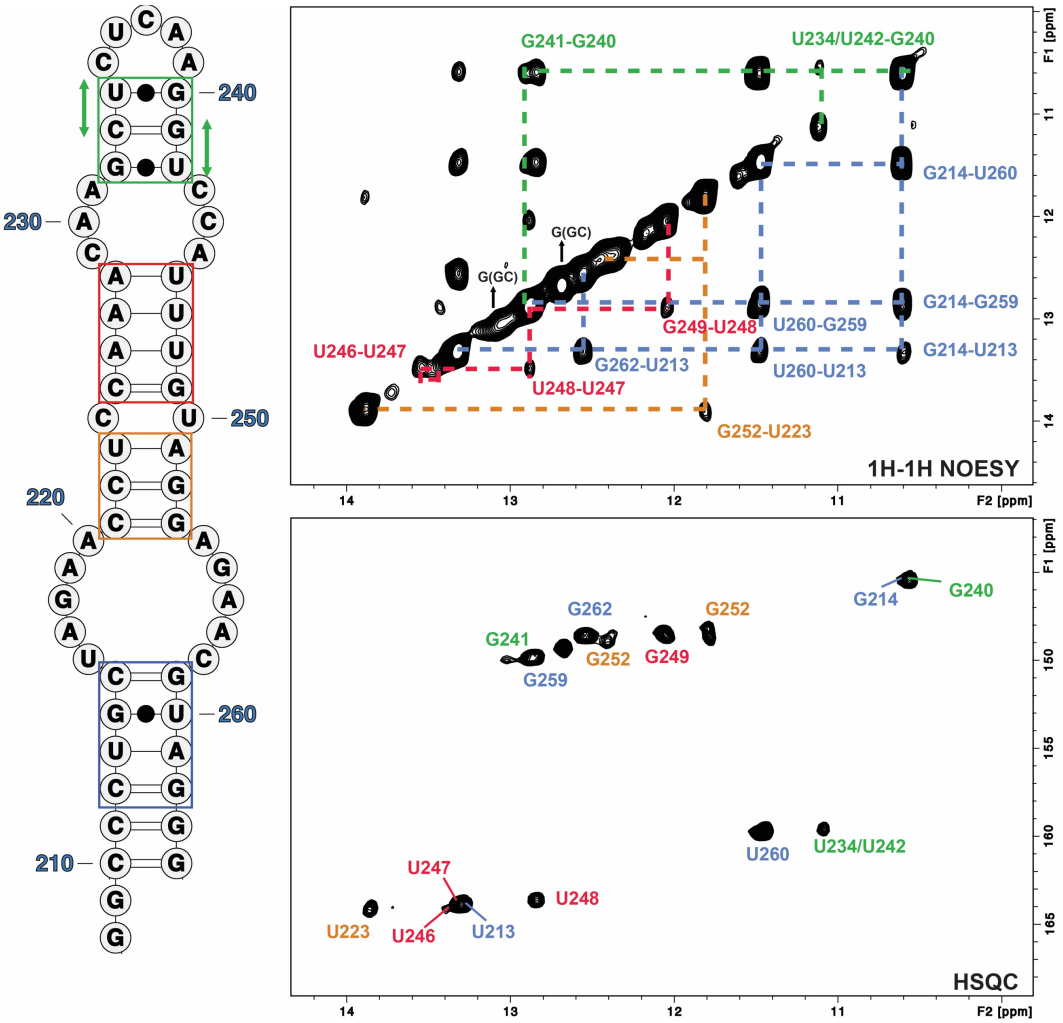

Figure S9. 7SK snRNA SL3S Imino Assignments. Left, secondary structure of the isolated SL3S construct (57-nts) used to further assist assignment of SL3M. Right upper, ^{1}H-^{1}H H_2O NOESY spectrum of SL3S collected in 10 mM of K_2HPO_4, 50 mM of KCl, pH 5.50, and 10% D_2O at 283K. The vertical and horizontal dashed lines trace the sequential NOE stacking pattern of each stable helical region, demonstrating that SL3S is an independently folded domain. Right lower, ^{1}H-^{15}N HSQC spectrum of SL3S collected in the buffer condition of 10 mM of K_2HPO_4, 50 mM of KCl, pH 5.50, and 10% D_2O at 283 K confirms the ^{1}H-^{1}H H_2O NOESY assignments.

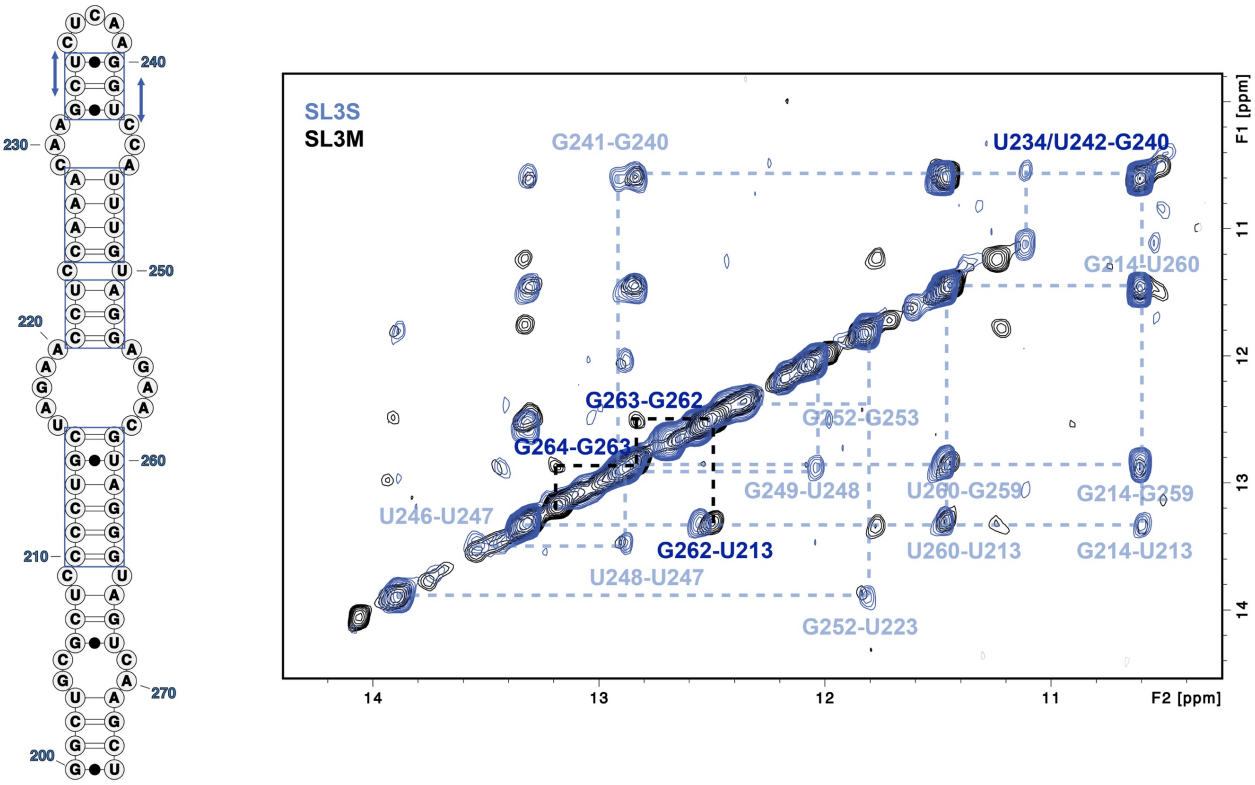

Figure S10. Overlay of ^{1}H-^{1}H-NOESY spectra of SL3M and SL3S. NMR data were collected in 10 mM K_2HPO_4, 50 mM KCl, pH 5.5, and 10% D_2O at 283 K. The majority of the NOE cross peaks of SL3M and SL3S overlay well on the ^{1}H-^{1}H-NOESY, revealing the upper apical loop of SL3 is an independently folded domain.

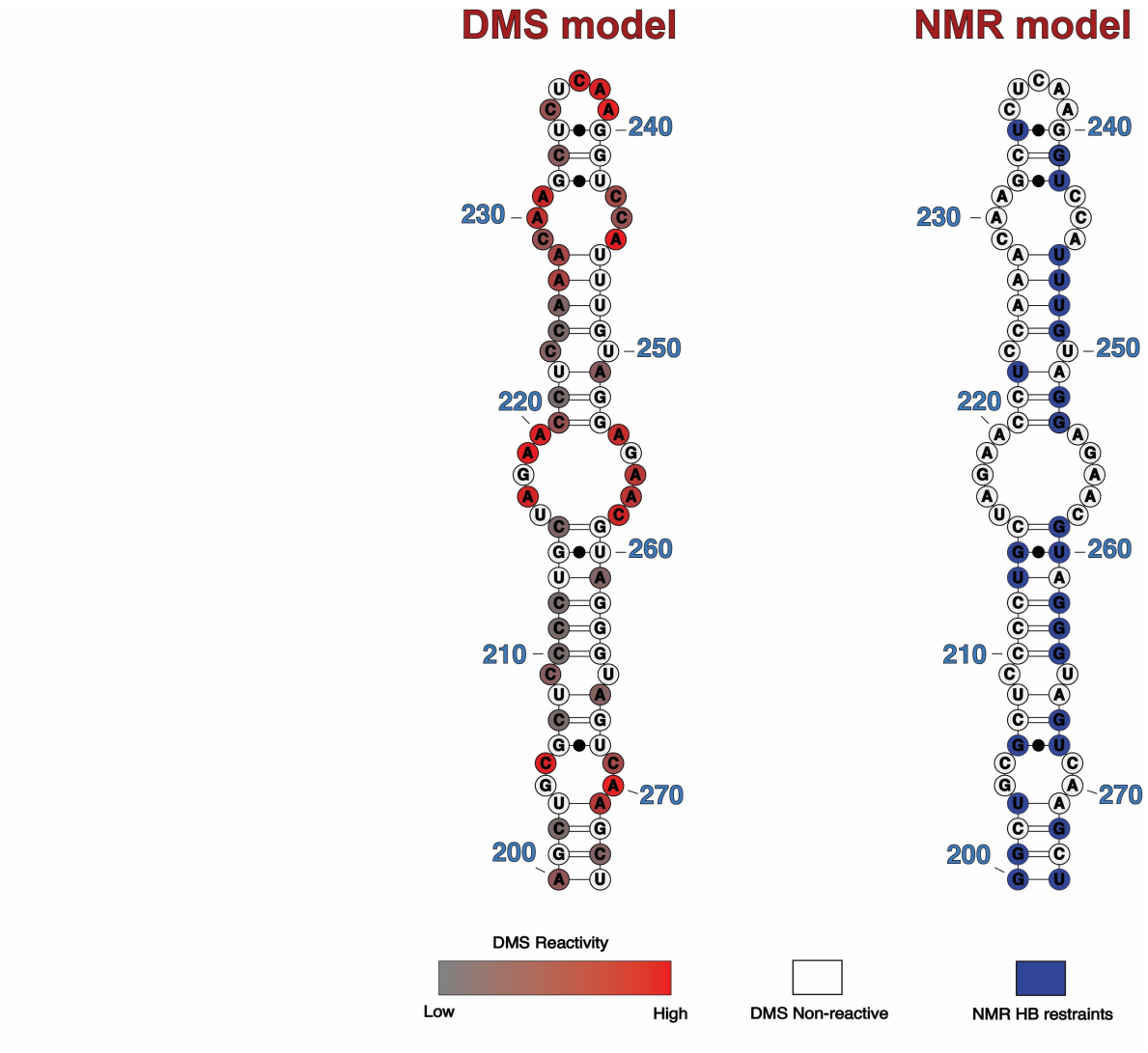

Figure S11. Comparison of Secondary Structure Model of 7SK SL3M Determined by DMS-MaPseq and NMR. (A). Secondary structure of SL3M in the context of the full-length RNA determined by DMS-MaPseq. (B). Hydrogen bond evidence derived from NMR that the 7SK SL3M folds with a structure consistent with the DMS calculated model.

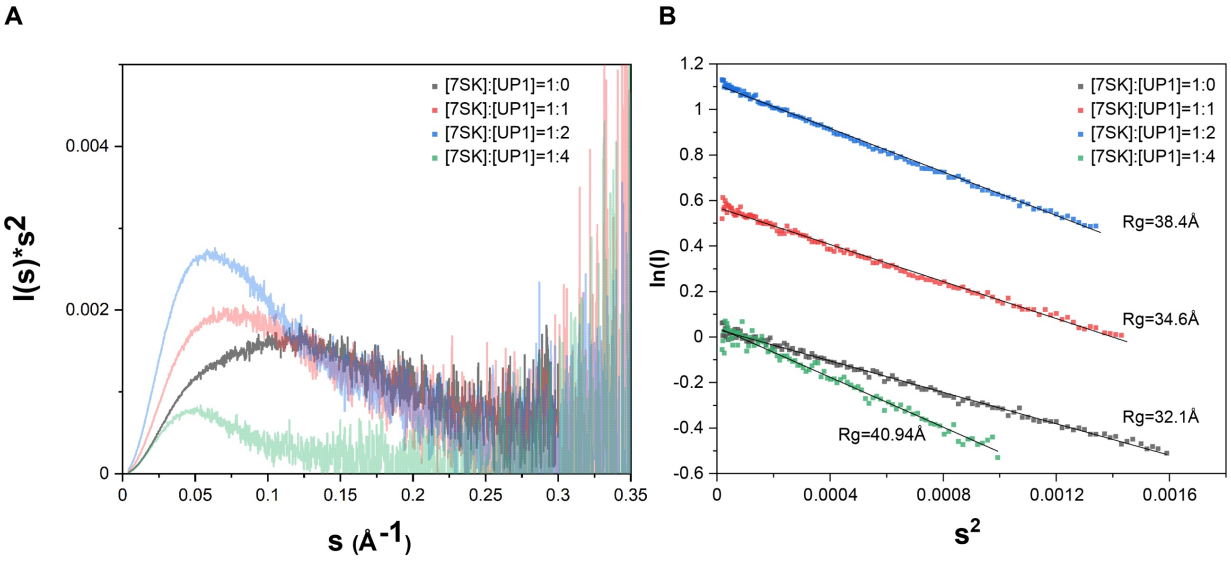

Figure. S12 SAXS analysis of 7SK SL3M with Kratky plot and Guinier plot. (A). Kratky plot of 7SK snRNA SL3M with increasing concentrations of UP1. [7SK]:[UP1] from 1:0 (Gray), 1:1(Red), 1:2 (Blue) to 1:4 (Green). The inverted parabolic shape of the plots indicate the overall folded nature of free 7SK SL3M and when bound to UP1. Data were collected in 50 mM KCl and 5 mM MES, pH 6.5 at room temperature. (B) Guinier plot of 7SK snRNA SL3M with increasing concentration of UP1. [7SK]:[UP1]= 1:0 (Gray), 1:1(Red), 1:2 (Blue) to 1:4 (Green). A gradual increase in the Rg can be observed for the different UP1-SL3M stoichiometric complexes.

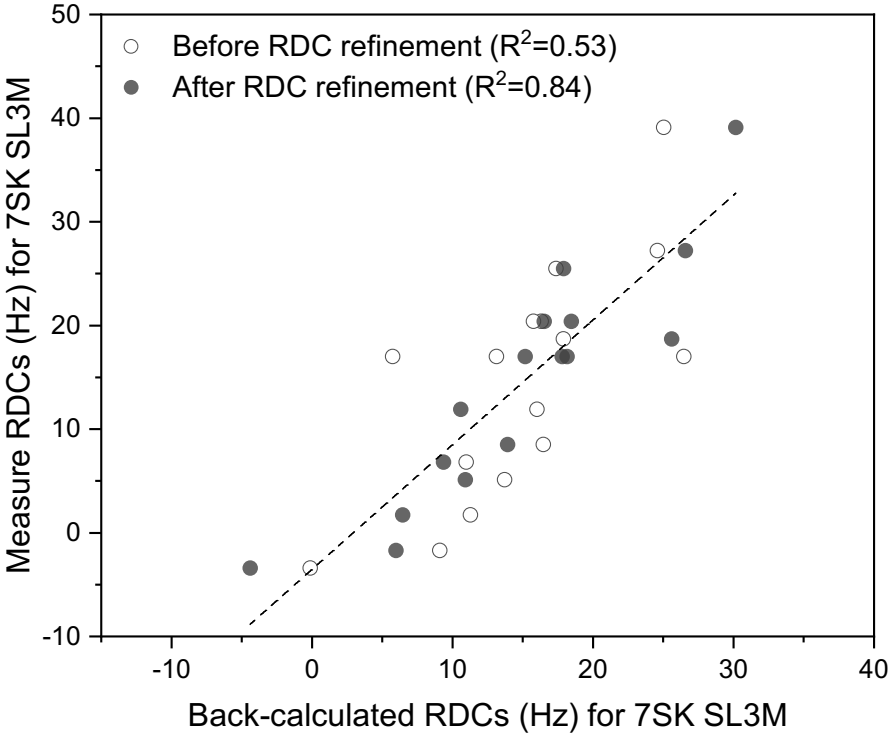

Figure S13. Correlation plot between measured and back-calculated RDCs. Comparison of the measured and back-calculated RDCs before and after RDC refinement from a representative lowest energy ROSETTA 7SK snRNA SL3M model.

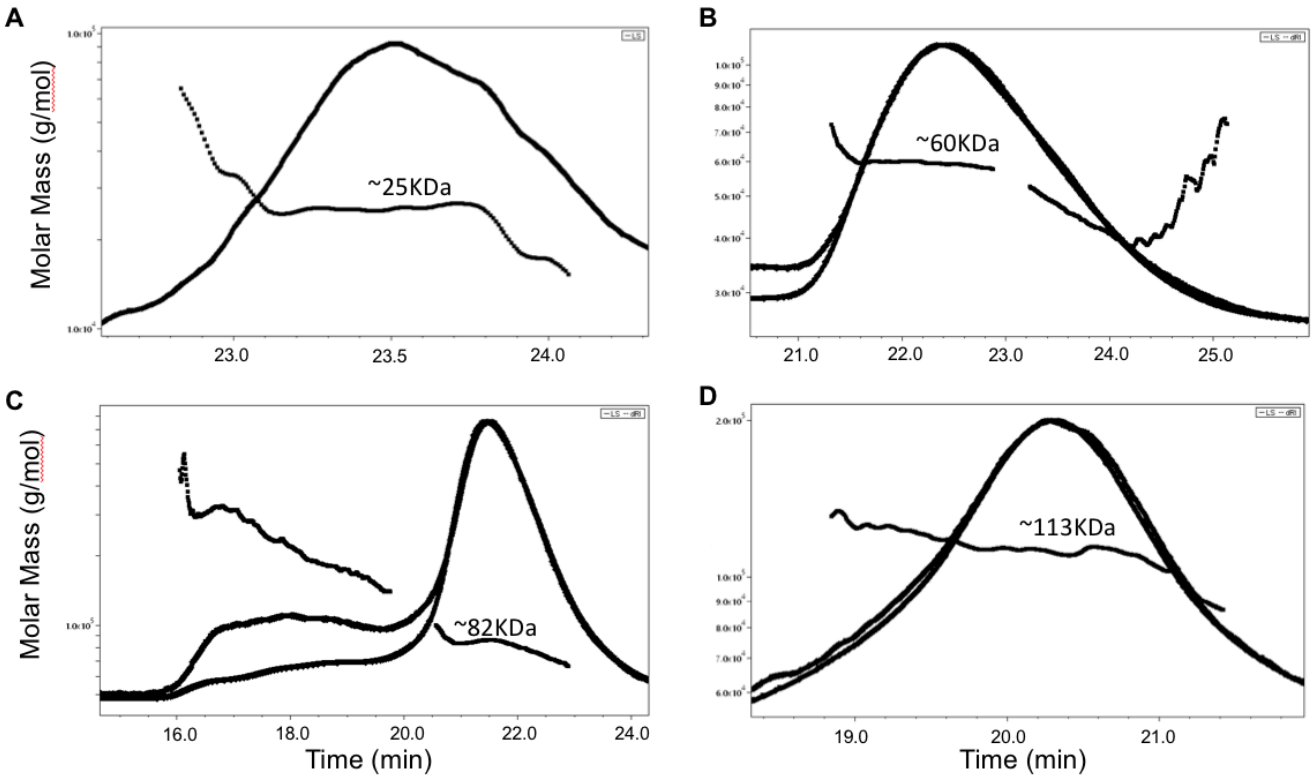

Figure S14. Stoichiometric complexes of UP1-SL3M resolved by SEC-MALS. Free 7SK snRNA SL3M and 1:1, 1:2 and 1:4 of 7SK snRNA SL3M-UP1 complexes were characterized by MALS in line with the SEC column. The theoretical molecular weights for the free 7SK SL3M is 23.99 kDa and 47.5, 71, and 118 kDa for the 1:1, 1:2 and 1:4 SL3M-UP1 complex. Data were collected in 50 mM KCl and 5 mM MES, pH 6.5 bufffer condition at room temperature.

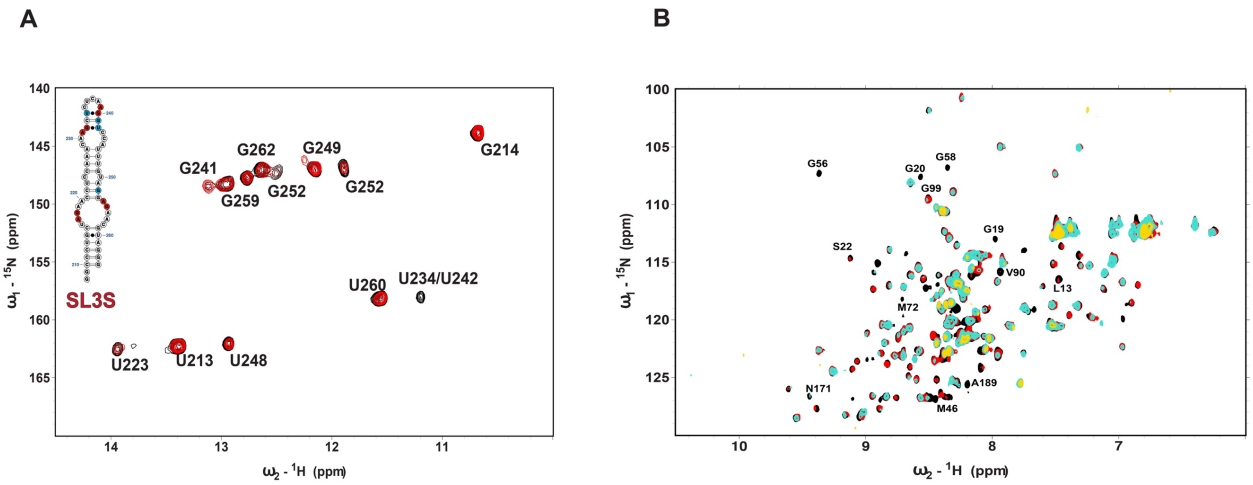

Figure S15. HnRNP A1/A2 proteins bind to the upper surface of SL3 to form a specific complex by NMR. (A) The ^{15}N-(GU) selectively-labeled SL3S RNA sample was held constant at 150 μM throughout the titration. Free SL3S (black) and the SL3S-UP1 complex at a 1:1.25 molar ratio (red) were collected for ^{1}H-^{15}N HSQC spectrum on the imino proton range. The assignment was done with reference to ^{1}H-^{1}H NOESY of SL3S. The residues with attenuated signals are shown in blue on the secondary structure of SL3S indicating their proximity to the exposed (red) AG sites. NMR data were collected in 10 mM K_2HPO_4, 50 mM KCl,10% D_2O pH 5.5 buffer condition at 288K on 900MHz NMR. (B)The selective ^{15}N labeling UP1 protein titrated with 7SK snRNA SL3S in ^{1}H-^{15}N HSQC spectrum with different UP1 to 7SK snRNA SL3S ratio: 1:0 (black), 1:0.3 (red), 1:0.5 (cyan) and 1:2 (yellow). Data were collected in 10mM K_2HPO_4,120mM KCl, 280mM NaCl, and pH 6.5 buffer condition at 303K on 800MHz NMR.