Identification of weak non-canonical base pairs around riboswitch-ligand recognition sites by solid-state NMR exchange spectroscopy

Sha Zhao[1,2], Ziyang Wen[1], Xinming Li[3], Mengbing Zou[1,2], Ge Yu[6], Xiangyang Liu[6], Lixin Zhang[6], Yi Xue[3,*], Riqiang Fu[4,*], Shenlin Wang[1,2,5,6,7,*]

1. College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China
2. Beijing NMR Center, Peking University, Beijing, 100871, China
3. Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China
4. National High Magnet Field Lab, 1800 East Paul Dirac Drive, Tallahassee, FL, 32310, USA
5. Beijing National Laboratory for Molecular Sciences, Beijing, China
6. State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China
7. State Key Laboratory of Medicinal Chemical Biology, NanKai University, Tianjin 300071, China

*Corresponding author: Shenlin Wang, Yi Xue, Riqiang Fu
E-mail: wangshenlin@pku.edu.cn, yixue@mail.tsinghua.edu.cn, rfu@magnet.fsu.edu

Dedicated to the 100th anniversary of Chemistry at Nankai University.

Supplementary Materials

1. Experimental Procedures

1.1 Materials

The ^{15}N,^{13}C-labeled rNTP reagents were purchased from Cambridge Isotope Laboratories (Andover, MA, U.S.A.). The DNA templates for the *in vitro* transcription reactions were purchased from GENEWIZ company (Suzhou, China).

1.2 *In Vitro* Transcription Reaction and Purification of RNA

RiboA71 is the 71-nt aptamer domain of the *add* adenine riboswitch from *Vibrio vulnificus*, which specifically binds to adenine, leading to the regulation of *add* gene expression.[1] Three ^{15}N,^{13}C isotope-labeled wild-type riboA71 samples were prepared, including a uniformly ^{15}N,^{13}C-labeled riboA71 complex with adenine, a ^{15}N,^{13}C-uridine-labeled riboA71 without adenine and a ^{15}N,^{13}C-uridine-labeled G30A/U19A–riboA71 complex with adenine. The sequences of the wild-type and G30A/U19A mutant riboA71 are as follows.

Wild-type riboA71:

5′-GGGAAGAUAUAAUCCUAAUGAUAUGGUUUGGGAGUUUCUACCAAGAGCCUUAAACUC UUGAUUAUCUUCCC-3′

G30A/U19A-riboA71:

5′-GGGAAGAUAUAAUCCUAA<u>A</u>GAUAUGGUUU<u>A</u>GGAGUUUCUACCAAGAGCCUUAAACUC UUGAUUAUCUUCCC-3′

Wild-type riboA71 was prepared by *in vitro* transcription reactions according to a previously reported protocol.[2] The double-stranded DNA templates were prepared by annealing single-stranded DNA with the sequence

5′-TTAATACGACTCACTATAGGGAAGATATAATCCTAATGATATGGTTTGGGAGTTTCTACCAAG AGCCTTAAACTCTTGATTATCTTCCC-3′ and its complementary sequence at 95 °C for 10 min. The *in vitro* transcription reactions contained 5 mM ^{15}N,^{13}C-labeled rNTPs, 40 mM Tris-HCl (pH 7.0), 0.01% Triton X-100, 1 mM spermidine, 10 mM DTT, 45 mM $MgCl_2$, 0.85 μM DNA template and 0.12 mg/mL T7 RNA polymerase and were incubated at 37 °C for 24 h. To produce ^{15}N,^{13}C-uridine-labeled riboA71, the reactions were supplemented with ^{15}N,^{13}C-labeled rUTP and abundant natural rCTP, rGTP and rATP. The synthesized RNA was purified using 12% PAGE under denaturing conditions and eluted from the gel with buffer containing 20 mM Tris-HCl, 300 mM sodium acetate and 1 mM EDTA, pH 7.4. The final yield of the purified RNA sample was approximately 5 mg per 10 mL of reaction system. Purified riboA71 was buffer-exchanged into a buffer containing 10 mM KH_2PO_4, 30 mM KCl and 2 mM $MgCl_2$, pH 6.8, and concentrated to a final

concentration of 700 μM. To prepare the riboA71–adenine complex, an adenine stock solution was added to the riboA71 solution to a final concentration of 5 mM; the mixture was annealed at 95 °C for 5 min and incubated at 0 °C for 30 min.

G30A/U19A–riboA71 was prepared using a similar protocol as described for wild-type riboA71; the only difference was in the DNA template used in the *in vitro* transcription. The sequence of the G30A/U19A–riboA71 DNA template was 5′-TTAATACGACTCACTATAGGGAAGATATAATCCTAAAGATATGGTTTAGGAGTTTCTACCAAG AGCCTTAAACTCTTGATTATCTTCCC-3′ and its complementary strand. All other materials, reaction conditions and purifications step were identical to those described for wild-type riboA71.

1.3 Preparation of RNA Samples with Partial Deuteration

Ethanol-precipitated, partially deuterated RNA samples were prepared using a previously published protocol[2,3]. This scheme produces an RNA solid-state sample that can yield high-resolution SSNMR spectra while retaining a high level of hydration for studies of the water–RNA interaction.[2]

To produce the partially deuterated RNA samples, ^{15}N,^{13}C-labeled RNAs were annealed in a 75% D_2O/25% H_2O (*v/v*)-based buffer to achieve the desired level of H/D exchange of the imino and the amino protons. To obtain ethanol-precipitated, partially deuterated samples for SSNMR, the partially deuterated RNAs were mixed with a pre-chilled ethanol stock consisting of 75% C_2D_5OD/25% C_2H_5OH (*v/v*). The final concentration of ethanol in solution was 75%. The RNA pellets were collected by centrifugation at 9600 *g* for 3 min, followed by central packing into a 1.9 mm rotor for SSNMR studies.

1.4 Solid-state NMR Spectroscopy

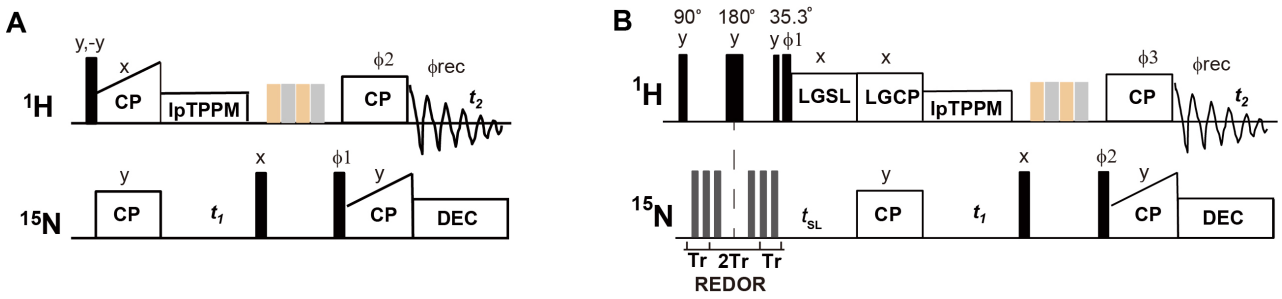
Approximately 4 mg of the ethanol-precipitated riboA71 was center-packed into a 1.9 mm rotor. All SSNMR experiments were performed on a 600 MHz Bruker Avance III spectrometer equipped with a 1.9 mm triple-resonance ^1H-X-Y magic-angle spinning (MAS) probe. All experiments were conducted at a MAS rate of 40 kHz. The typical π/2 pulse lengths were 2.5 μs for ^1H, 4.0 μs for ^{13}C and 4.5 μs for ^{15}N. All experiments were conducted at an effective temperature of 15 °C. The temperature was calibrated using the T_1 relaxation time of ^{79}Br in KBr powder.[4]

The 2D hNH Experiments. The dipolar-based 2D hNH experiments were performed using a previously reported pulse sequence.[5] The diagram of the pulse sequence is shown in Supplementary figure 1A. Both the ^1H-^{15}N and ^{15}N-^1H cross-polarization (CP) transfer were set at a constant field lock of ^{15}N at 60 kHz and an experimentally optimized proton power around n = 1 Hartmann–Hahn (HH) conditions (10% linear ramp). The contact times for ^1H-^{15}N and ^{15}N-^1H CP were 4 ms and 300 μs, respectively. The carrier frequencies were

set to 10 ppm for ^{1}H and 145 ppm for ^{15}N. The total acquisition times for the ^{1}H and ^{15}N dimensions were 20 ms and 15 ms, respectively. The spectral width of the ^{15}N dimension was set at 50 ppm. States–TPPI phase-sensitive detection was obtained in the indirect dimension (^{15}N) by incrementing the first $\pi/2$ pulse of the ^{1}H channel. In this scenario, the amide cross-peaks were folded in the spectra, with ^{15}N chemical shifts of 124 ppm for G(NH_2), 131 ppm for A(NH_2) and 147 ppm for C(NH_2). MISSISSIPPI pulse trains were used to suppress the solvent signals.[5] Low-power TPPM proton decoupling (nutation frequency ~10 kHz) and WALTZ-16 ^{15}N decoupling (nutation frequency ~10 kHz) were applied during the evolution of the chemical shifts of ^{15}N and ^{1}H, respectively. The recycle delay in the 2D hNH experiments was 2 s. The hNH experiments were recorded with 32 scans.

2D ^{1}H-^{15}N WaterREXSY Experiments. Two-dimensional (2D) ^{1}H-^{15}N water–RNA exchange spectroscopy (WaterREXSY) experiments were performed using the pulse sequence described in Supplementary figure 1B. The pulse sequence was developed in a previous study[6]. Compared to the 2D hNH experiments, 2D WaterREXSY contains two additional blocks: a ^{1}H-^{15}N REDOR dephasing period with rotor-synchronized π pulses on the ^{15}N channel, and a Lee–Goldburg (LG) spin lock along the proton channel. ^{1}H-^{15}N REDOR dephasing was achieved by applying 16 rotor-synchronized π pulses to the ^{15}N channel to dephase the ^{1}H signals covalently attached to the ^{15}N atoms. The total REDOR dephasing time was 0.4 ms. A π pulse in the middle of the REDOR dephasing block was applied to the ^{1}H channel to refocus the ^{1}H chemical shift evolution. During the LG spin-lock period, a ^{1}H B_1 field of 53.2 kHz was applied at an offset of +37.6 kHz, resulting in an effective spin-locking field of 65.2 kHz along the magic angle. After the LG spin-lock period, the ^{1}H-^{15}N LG-CP transfer was set with the LG spin-lock on ^{1}H at the same parameters as those applied during the LG spin lock period, and on ^{15}N with the power experimentally optimized to achieve an efficient transfer. The contact time of ^{1}H-^{15}N CP was 300 μs. The ^{15}N-^{1}H CP and other experimental parameters were identical to those in the 2D hNH experiments. A series of 2D WaterREXSY experiments were performed using different mixing times in LG-spin lock period: 500 μs, 800 μs, 1200 μs, 2000 μs, 3000 μs, 4000 μs and 5000 μs. Each 2D WaterREXSY experiment was recorded with 80 scans, resulting in durations of approximately 19 hours per experiment.

All NMR data were processed using the TOPSPIN 3.2 program. The spectra were analyzed using CARA. Chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), using adamantane as a secondary standard.[7] The ^{1}H and ^{15}N chemical shifts were referenced indirectly using $\gamma^{13}C/\gamma^{1}H = 0.25145020$ and $\gamma^{15}N/\gamma^{13}C = 0.40297994$, respectively.



Supplementary Figure 1. Pulse sequences of two-dimensional (2D) proton-detected solid-state NMR (SSNMR) experiments, including (a) 2D hNH and (b) 2D ^{1}H-^{15}N water–RNA exchange spectroscopy (WaterREXSY) experiments. For both pulse sequences, the yellow and gray bars represent the water suppression period using MISSISSIPPI pulse trains.[5] The pulse phases are indicated on the pulse, with the exception of the following:: (a) $\phi_1 = (x, x, -x, -x)$, $\phi_2 = (y, y, y, y, -y, -y, -y, -y)$, and $\phi_{rec} = (y, -y, -y, y, -y, y, y, -y)$; (b) $\phi_1 = (y, -y)$, $\phi_2 =(x, x, -x, -x)$, $\phi_3 = (y, y, y, y, -y, -y, -y, -y)$, and $\phi_{rec} = (y, -y, -y, y, -y, y, y, -y)$. The phases of the ^{15}N pulses during the REDOR period are x, y, x, y.

1.5 Molecular Dynamics (MD) Simulation

For the MD simulation, the starting structure was taken from the crystal structure of the riboA71–adenine complex (PDB ID: 4TZX).[8] All crystallographic water molecules and Mg^{2+} ions were retained. For the adenine ligand, hydrogen atoms were added using Discovery Studio 3.5[9]. The partial charges were calculated using the antechamber module of the Amber 18 package[10] and the AM1-BCC method. The general AMBER force-field (GAFF) parameter was used for the adenine ligand, and the other parameters required for the ligand were generated using parmchk.[11] For riboA71 RNA, hydrogen atoms were added by using the tleap module of Amber 18. The force field used for the RNA was ff99bsc0XOL3.[12] The riboA71–adenine complex was solvated in a truncated octahedral box using the SPCE water model and a margin distance of 10 Å. Fifty-two potassium ions were added to neutralize the system.

The energy minimization and MD simulation were performed using the GPU accelerated version of pmemd in AMBER. The solvated system was minimized for 1000 steps with harmonic restraints (force constant = 500 kcal·mol^{-1}·Å^{-2}), followed by 1000 steps without restraints. The system was then heated for 20 ps from 0 to 288 K, and equilibrated for 1 ns at 288 K. The production stage of the MD simulation was conducted at 288 K using the NPT ensemble and a 2 fs integration step. During the simulations, all bonds involving hydrogen atoms were constrained using the SHAKE algorithm. The nonbonded cutoff was set to 9 Å. DSSR was used to analyze the formation of hydrogen bonds in the MD trajectory at an interval of 1 ps.[13]

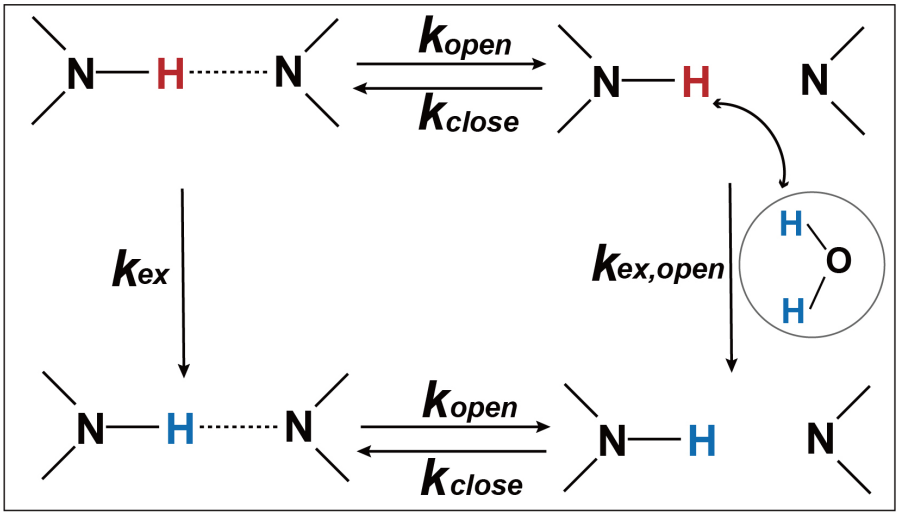
2. Results

2.1 “Open–Close” Model to Describe the Hydrogen Exchange Theory of the Base-Paired Imino Protons

The formalism of proton exchange has been extensively described using the “open–close” model.[14-17] Equilibrium exists between the open state, in which the H-bond breaks, and the closed state, in which the H-bond forms. The exchangeable protons can exchange with water only in the open state. In this model (Supplementary figure 2), k_{ex} is the overall rate constant of the water–RNA chemical exchange, which can be described using Eq. 1:

$$k_{ex} = \frac{kopen * kex,open}{kclose + kex,open} \quad (1)$$

where k_{open} and k_{close} are the rate constants for the opening and closing of the base pair, respectively, and $k_{ex,open}$ is the rate constant for proton exchange in the opening state.



Supplementary Figure 2. “Open-close” model to describe hydrogen exchange of base-paired imino protons.

Two kinetic regimes can characterize k_{ex}, depending on the relative magnitudes of the rate constants. In the regime where $k_{ex,open} >> k_{close}$, the exchange rate is limited by the formation of the exchange-accessible state (opening-limited exchange); thus, $k_{ex} = k_{open}$. This condition often occurs at high concentrations of a proton acceptor in the presence of a catalyst, such as NH_3.Another regime occurs at low concentrations of a proton acceptor, wherein the base pairs open and close many times before the exchange (pre-equilibrium exchange); thus, the exchange rate k_{ex} can be described as follows:

$$kex = \frac{kopen * kex,open}{kclose} = K\text{diss} * kex, open \quad (2)$$

where *K*diss (= k_{open}/k_{close}) is the equilibrium constant for base pair opening, which reflects the hydrogen stability of base pairs. Thus, the exchange rate constant is affected by the hydrogen stability and the proton exchange rate constant in the opening state.

In the riboA71–adenine complex, the solution buffer pH was 6.8, and no other catalyst was added. Therefore, the RNA exchange regime is also referred to as the pre-equilibrium exchange regime (Eq. 2). Here, the exchange rate between the hydrogen-bonded protons of the base pair and water is determined by *K*diss and the exchange rate between the protons in the open state and water. The latter rate is fast in the range of 10^6 s^{-1}, whereas the former is slow and is the rate-limiting step, which is determined by the stability of the hydrogen bonds within the base pairs.

2.2 Model of the 2D WaterREXSY Experiments

In the open–close model of the chemical exchange process between RNA and water (Supplementary figure 2), the intermediate N–H bond state cannot stably exist. In our experimental design, exchangeable protons, namely those in hydrogen bonds (N–H…N or N–H…O=C), are initially suppressed by the ^{1}H-^{15}N REDOR cycles and then repolarized by water through a chemical exchange. Thus, we can use a simplified exchange model (Supplementary figure 3) to describe our experiment: the proton in the hydrogen bond is in exchange with a pool of n_w water molecules at the exchange rate constant of k_{IM}. Solomon equations are then used to characterize the chemical exchange in this two-spin system (M: water; I: the hydrogen-bonded proton, i.e., the proton in N–H…N and the proton in N–H…O=C) under the irradiation of a Lee–Goldburg (LG) sequence.

$$\begin{cases} \frac{dM(t_{SL})}{dt_{SL}} = -\left(\frac{1}{T_{1}^{M}} + k_{IM}/n_w\right) M(t_{SL}) + k_{IM} I(t_{SL}) \\ \frac{dI(t_{SL})}{dt_{SL}} = k_{IM}/n_w \, M(t_{SL}) - \left(\frac{1}{T_{1}^{I}} + k_{IM}\right) I(t_{SL}) \end{cases} \quad (3)$$

where $M(t_{SL})$ and $I(t_{SL})$ are the M and I magnetizations along the magic angle at a given LG spin-lock time t_{SL}, respectively, and $\boldsymbol{T_{1}}$ represents their relaxation times under the t_{SL}. Because the M and I magnetizations are along the magic angle, the cross-relaxation among them is negligible in the above equations. k_{IM} represents the exchange rate constant between the proton in the N–H bond and the pool of n_w water molecules. In these differential equations, the factor of $1/n_w$ is taken into account to reflect that only one water molecule is in exchange with the I proton. The analytical solutions for these differential equations can be derived as follows. The eigenvalues for the differential equations in Eq. 3 are:

$$\lambda_{\pm} = -\lambda \pm D \quad , \quad (4)$$

where $\square = \frac{1}{2}\left[\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}} + \frac{1}{T_{1\square}^{I}} + \frac{1}{T_{1\square}^{M}}\right]$, $D = \left(\square^2 + \frac{1}{n_w}k_{\mathrm{IM}}\right)^{1/2}$ and $\square = \frac{1}{2}\left(\left(\frac{1}{n_w}-1\right)k_{\mathrm{IM}} + \frac{1}{T_{1\square}^{M}} - \frac{1}{T_{1\square}^{I}}\right)$.

As both I and M are spin-locked along the magic angle direction, it is reasonable to assume that $\frac{1}{T_{1\square}^{S}} = \frac{1}{T_{1\square}^{I}} \equiv$ $\frac{1}{T_{1\square}}$ for simplicity, leading to $\square = \frac{1}{2}\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}} + \frac{1}{T_{1\square}}$ and $D = \frac{1}{2}\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}}$. Thus, we obtain

$$\square_+ = -\square + D = -\frac{1}{T_{1\square}},$$

$$\square_- = -\square - D = -\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}} - \frac{1}{T_{1\square}^{I}}.$$

Consequently, the analytical solutions for the M and I spins are

$$M(t_{SL}) = a_+ exp(\square_+ t_{SL}) + a_- exp(\square_- t_{SL}) = \left\{a_+ + a_- exp\left[-\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}} t_{SL}\right]\right\} exp(-t_{SL}/T_{1\square}) \quad (5)$$

$$I(t_{SL}) = b_+ exp(\square_+ t_{SL}) + b_- exp(\square_- t_{SL}) = \left\{b_+ + b_- exp\left[-\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}} t_{SL}\right]\right\} exp(-t_{SL}/T_{1\square}) \quad (6)$$

Here, $a_- = -b_-$ and $a_+ = n_w b_+$.

By taking into account the initial boundary conditions: $M(t_{SL})|_{t_{SL}=0} = M(0) = a_+ + a_-$ and

$$I(t_{SL})|_{t_{SL}=0} = I(0) = b_+ + b_- = 0,$$

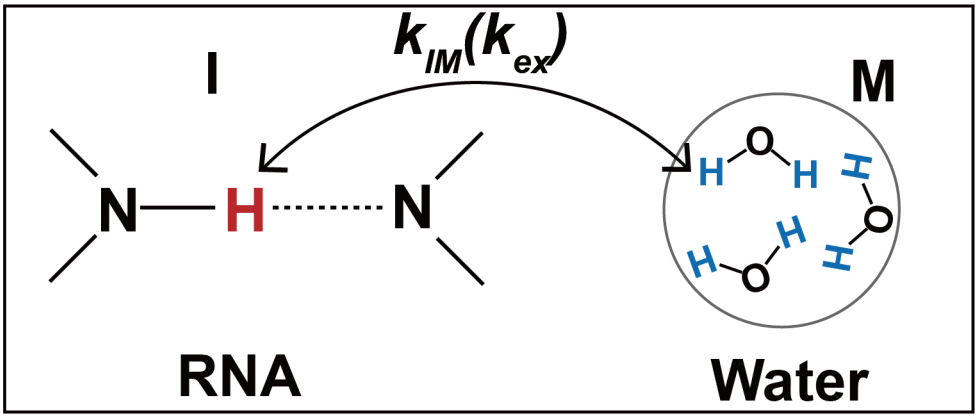
we can obtain the analytical solution as in the literature[6]:

$$I(t_{SL}) = pM(0)\left\{1 - exp\left[-\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}} t_{SL}\right]\right\} exp(-t_{SL}/T_{1\square})/\left(\frac{1}{n_w}+1\right) \quad (7)$$

Therefore, the build-up of the ^{1}H magnetization in the N–H…N or N–H…O=C hydrogen bond in the experiments is depicted in the pulse sequence. Supplementary figure 1B represents the exchange process between the pool of water molecules and the N–H**…**N hydrogen bond at an exchange rate constant of $\left(1+\frac{1}{n_w}\right)k_{\mathrm{IM}}$, which should be equivalent to k_{ex} in the open–close model for the chemical exchange process between RNA and water. Thus, we obtain:

$$I(t_{SL}) = \{I(0) + A[1 - \exp(-k_{ex}\, t_{SL})]\} exp(-t_{SL}/T_{1\square}) \quad (8)$$

Here, I(0) is the initial magnetization before the exchange-induced signal build-up, such as the contribution from incomplete REDOR dephasing and the CP transfer from ^{1}H to ^{15}N along the magic angle, whereas $\boldsymbol{A = M(0)/(n_w + 1)}$ is considered a constant coefficient that does not contribute to signal build-up. For the riboA71–adenine complex system, $\frac{1}{n_w}$ is $10^{-6.8}$ in a pH 6.8 solution. The $T_{1\rho}$ of the imino proton is 8.5 ms, as measured using the pulse sequence in Supplementary figure 1B with variable t_{SL} periods before ^{1}H-^{15}N CP. The k_{ex} was obtained by fitting the $I(t_{SL})$ against the t_{SL}.

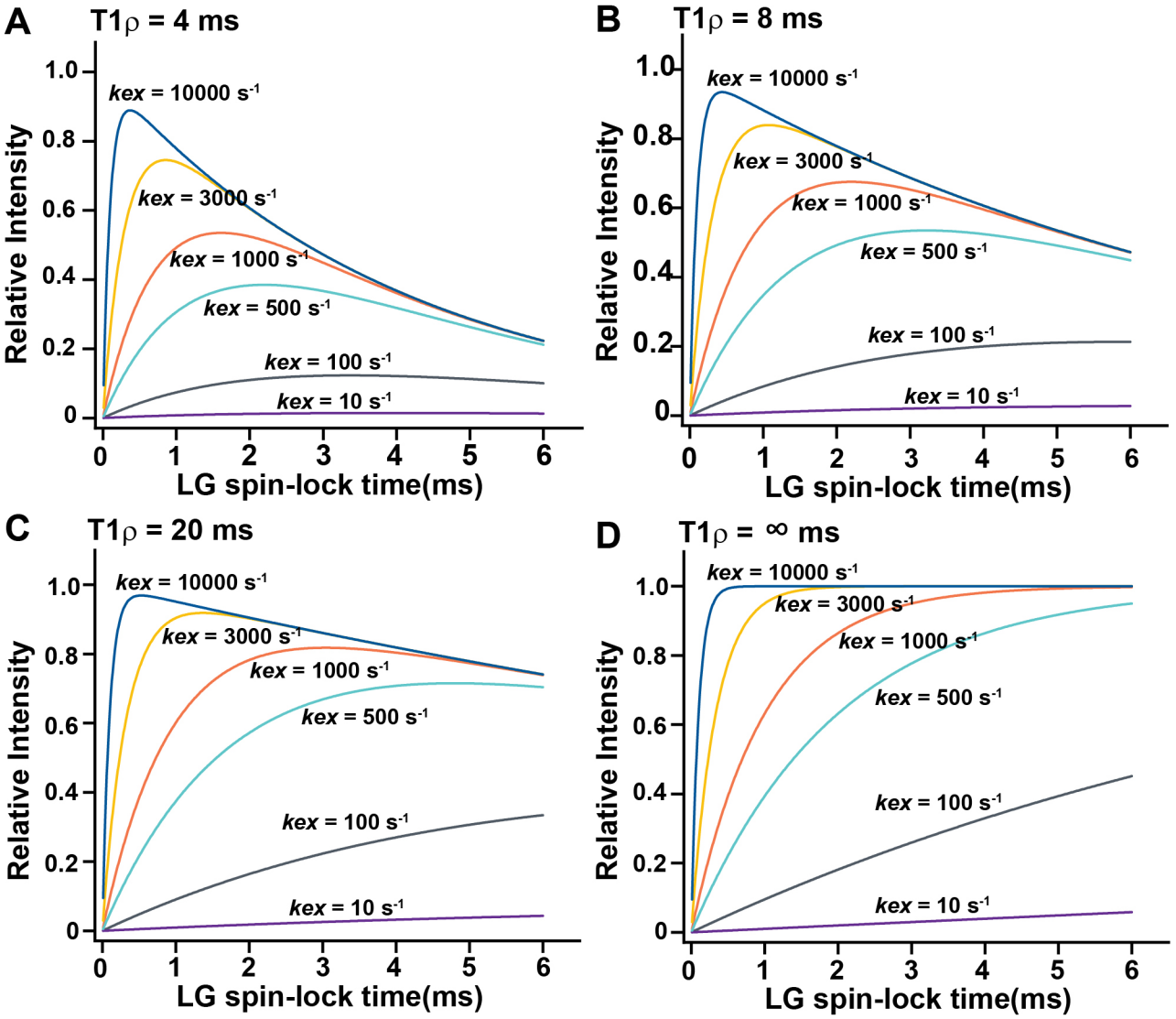


Supplementary Figure 3. Simplified water–RNA exchange model of the water–RNA exchange spectroscopy (WaterREXSY) experiment.

2.3 Expected Results of 2D WaterREXSY

In our 2D WaterREXSY scheme, the ^{1}H-^{15}N REDOR block suppresses the signals of the imino and amino protons that covalently attach to the ^{15}N atoms, whereas the water magnetization is less affected. During the t_{SL} period, a chemical exchange between the exchangeable protons and the excited water occurs, thus polarizing the exchangeable protons. Because the spin-diffusion among the protons is suppressed during the LG spin-lock on ^{1}H, the exchangeable protons can only be polarized by the chemical exchange process. This enables a characterization of the pure chemical exchange process. According to Eq. 8, the build-up of the cross-peak intensities of the exchangeable protons is determined by the exchange rate, k_{ex}, and the $T_{1\rho}$ relaxation time. The protons that experience a fast exchange with water (i.e., a k_{ex} on a sub-ms time scale) could yield a high optimal transfer efficiency, as a t_{SL} of a few ms is sufficient to reach equilibrium, and the magnetization loss due to relaxation is small. In contrast, a low transfer efficiency is expected for a slow exchange process because long mixing times are required to reach equilibrium, during which the relaxation loss greatly reduces the transfer efficiency. Supplementary figure 4 depicts the simulations using different combinations of $T_{1\rho}$ and k_{ex}. As shown in Supplementary figure 4B, at a $T_{1\rho}$ relaxation time of approximately 8 ms, the exchangeable protons with k_{ex} of 0.5 ms^{-1} can be observed effectively with a t_{SL} of 2 ms. In contrast,

protons with slow exchange rates, i.e., k_{ex} <10 s^{-1}, would be difficult to observe due to the low transfer efficiency.



Supplementary Figure 4. Simulations of build-up curves from the two-dimensional (2D) water–RNA exchange spectroscopy (WaterREXSY) experiment at different values of *T1ρ* and k_{ex}. The *T1ρ* times are listed above the simulations. The relative intensity reflects the I(t_{SL}) in Eq. 8, assuming that A = 1 for all simulations..

Accordingly, in RNA, the water–proton exchange rates of the imino or the amino protons within the base pair depend on the stability of the base pair. Empirically, the relative stabilities of the base pairs have a certain order, namely G–C > A–U > G•U > A•G > U•U > C•A,[18] and the stability of individual base pairs is also dependent on the sequence context. For very stable base pairs, namely the Watson–Crick canonical G–C and A–U base pairs, the k_{ex} rates of the imino protons of guanines and uridines are in the ranges of 1–10 s^{-1} and

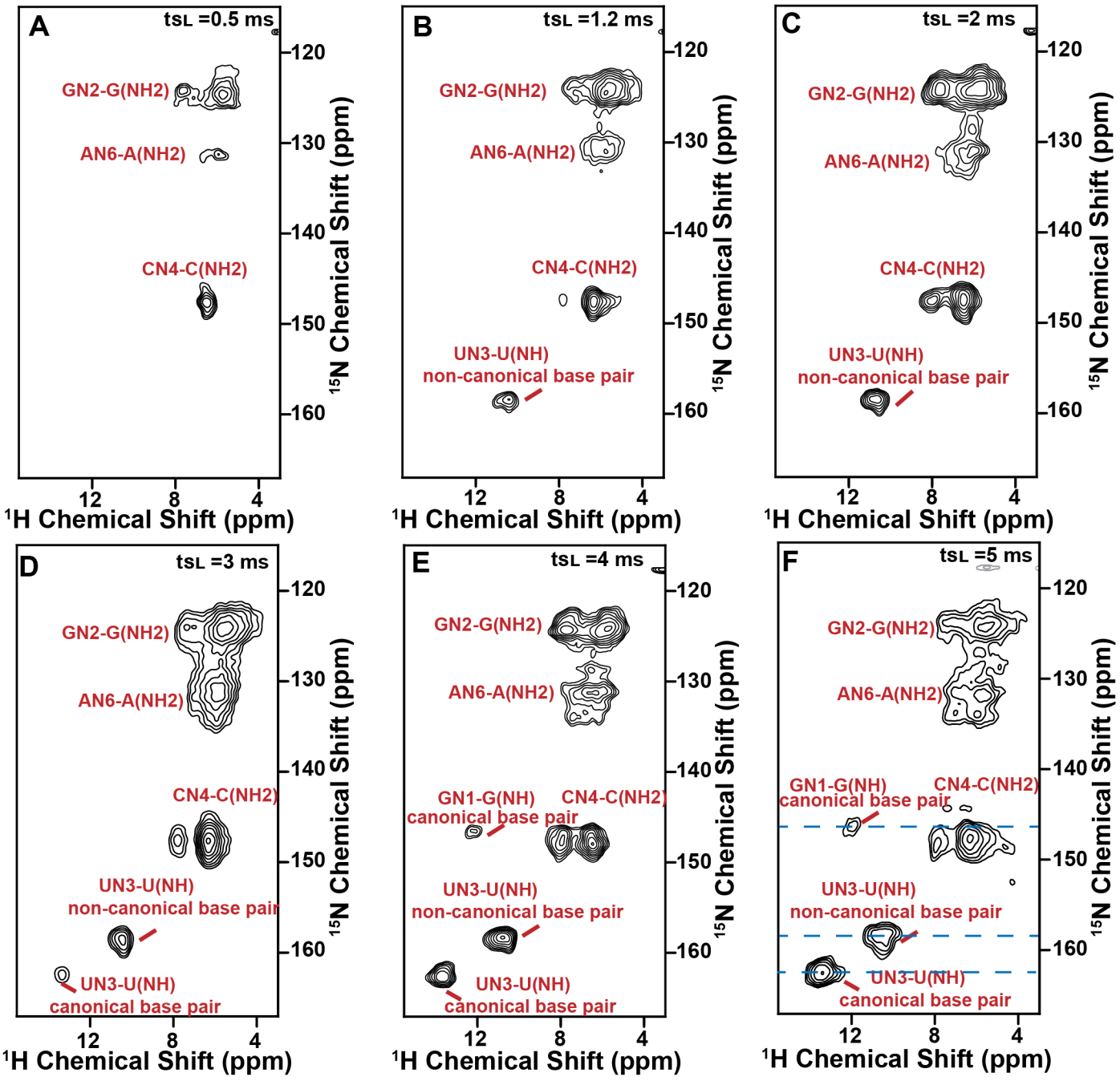
20–100 s^{-1}, respectively. The non-canonical base pairs, namely G•U, A•G, U•U and C•A, are relatively weak and have k_{ex} rates in the range of 100–1000 s^{-1} or faster.

Based on the physicochemical properties of RNA and the simulations in Supplementary figure 4, we expect to observe the hydrogen-bonded protons within the weak base pairs on the 2D WaterREXSY spectra within a few ms of t_{SL}, provided that the k_{ex} of the exchange is approximately 500 s^{-1}. In contrast, protons in very stable base pairs are difficult to observe during 2D WaterREXSY experiments. Because both strong and weak base pairs are detectable in 2D hNH experiments, a comparison of the 2D hNH and 2D WaterREXSY spectra would be expected to reveal the weak base pairs in RNA.

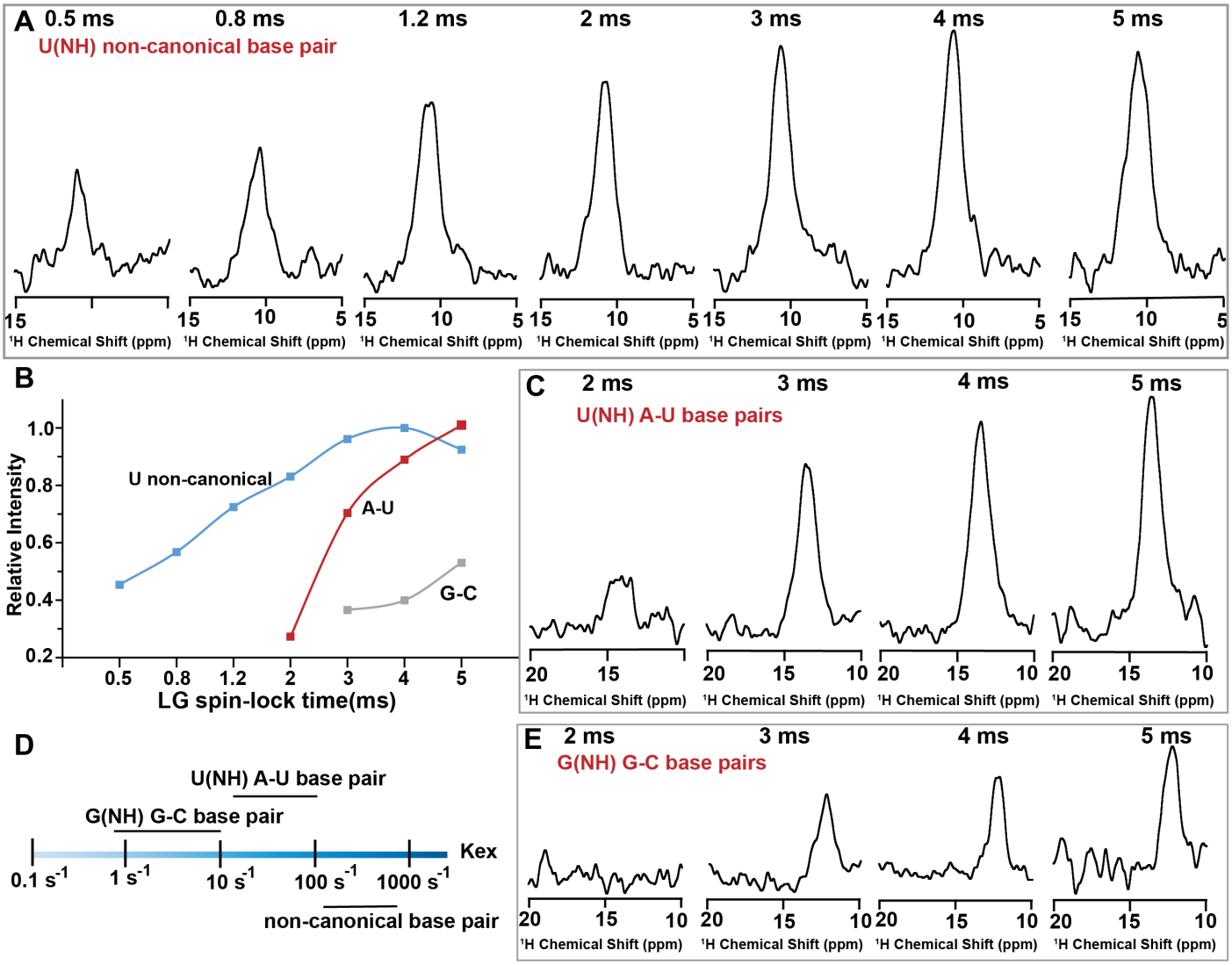
2.4 RNA–Water Exchange Rates for Different Types of Imino and Amino Groups.

The 2D WaterREXSY experiment includes a ^{1}H-^{15}N REDOR block to dephase the imino signals. Without the t_{SL} period, imino groups that undergo a slow chemical exchange with water would not be observed on the 2D WaterREXSY spectra. This would serve as a control experiment to evaluate the completeness of REDOR dephasing. Supplementary figure 5 shows a series of 2D WaterREXSY experiments with different t_{SL} periods and a 0.4 ms REDOR dephasing time. At a t_{SL} of 500 μs, the imino groups of both guanines within canonical G–C base pairs and uridines within canonical A–U base-pairs were not observed (Supplementary figure 5A), indicating the effectiveness of the REDOR dephasing. These imino groups exhibited a very slow build-up and first appeared on the 2D WaterREXSY spectra at a t_{SL} of 2 ms and 3 ms, respectively. 1D ^{1}H signals against t_{SL} of imino of uridines in non-canonical base pair, A-U base pairs and guanines in G-C base pairs are shown in Supplementary figure 6. This is consistent with the expectation of stable canonical base pairs. The imino groups of uridines within the non-canonical base pairs appear first at a t_{SL} of 0.5 ms and reach the reflection point at a mixing time of 4 ms. Fitting yields an exchange rate of 379 ± 57 s^{-1} at 15 °C,

The amino groups experience faster exchange rates than the imino groups. The bulk chemical exchange rates between the amino groups and water were fitted as 430 ± 91 s^{-1}, 899 ± 98 s^{-1} and 288 ± 136 s^{-1} for $C(NH_2)$, $A(NH_2)$ and $G(NH_2)$, respectively, consistent with the fast exchange rates of NH_2 groups (Supplementary figure 7).



Supplementary Figure 5. Two-dimensional (2D) WaterREXSY spectra of riboA71-adenine complex with different t_{SL}. (A-F) Collected 2D WaterREXSY spectra with t_{SL} of 500 μs (A), 800 μs (B), 1.2 ms (C), 2 ms (D), 3 ms (E), 4 ms (F) and 5 ms (G). The nucleotide-type specific assignments are highlighted for the corresponding resonances. The gray dashed lines indicate the ^{15}N chemical shifts of the 1D extracted slices shown in Supplementary figure 6. The GN1–G(NH) cross peaks at (D) and the UN3–U(NH) cross peaks at (C) are not visible at the current contour level.

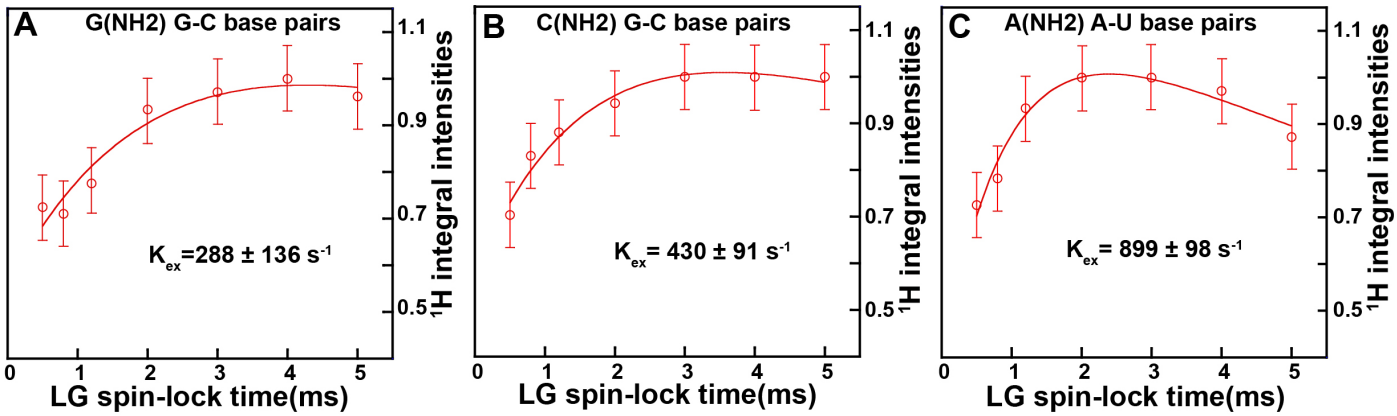


278

279 **Supplementary Figure 6**. One-dimensional (1D) ^{1}H slices with ^{15}N chemical shift of 158.6 ppm (A),
280 162.7 ppm (C) and 146.5 ppm (E). (B) In the plots of ^{1}H signals vs. t_{SL} of uridine imino protons
281 depict non-canonical base pairs (blue), A–U base pairs (red) and guanines in G–C base pairs (gray).
282 (D) Exchange rate constant of different imino groups.

283

284

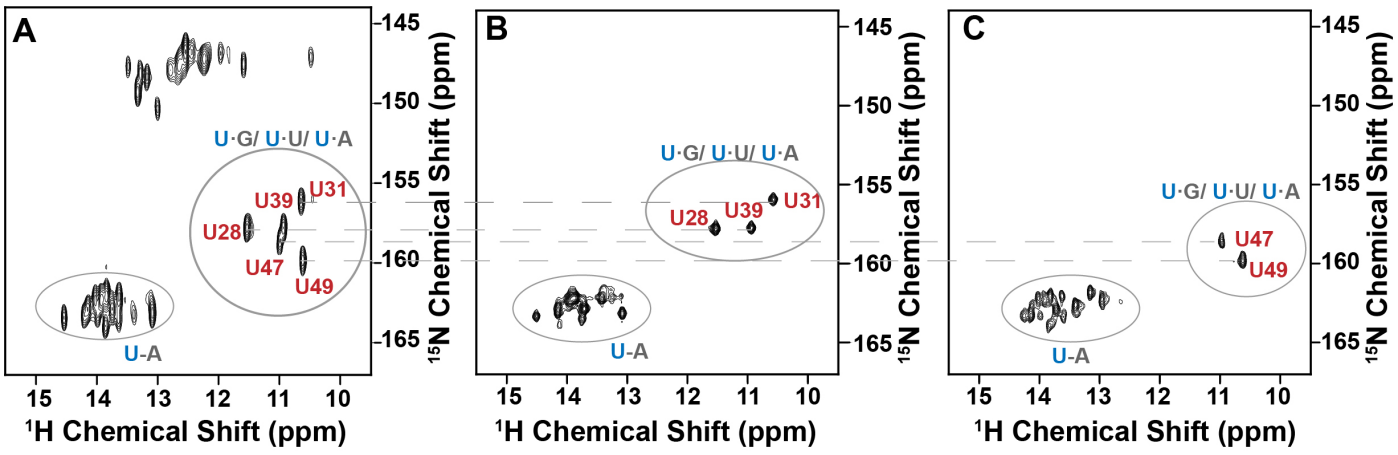


Supplementary Figure 7. The intensity of amino protons with different t_{SL}. The red line represents the fitting curve.

2.5 Assignments of the RiboA71–Adenine Complex in Solution.

The 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the riboA71–adenine complex in solution was collected (Supplementary figure 8A). The well-resolved imino peaks of uridines within the non-canonical base pairs were unambiguously assigned according to previous reports,[19,20] including U28, U39 and U31 in the P2 stem and U47 and U49 in J2–3 (Supplementary figure 8A). Supplementary figures 8B and 8C show the 2D ^{1}H-^{15}N HSQC spectra of the ^{15}N,^{13}C-uridine-labeled riboA71 without the adenine ligand and the ^{15}N,^{13}C-uridine-labeled G30A/U19A riboA71–adenine complex.

For free wild-type riboA71, only three imino groups corresponding to U28, U31 and U39 in the stems were observed in the region of uridine within non-canonical base pairs on ^{1}H-^{15}N HSQC. U47 and U49 were not detected in the free wild-type riboA71, suggesting the structural heterogeneity of that region in adenine-free riboA71. For the G30A/U19A riboA71–adenine complex, the U31–U39 and U28–G42 non-canonical base pairs were changed to U28–A42 and A31–U39 canonical base pairs, thus shifting those cross peaks to the region of U–A canonical base pairs. Thus, only two non-canonical uridine imino peaks, U47 and U49, were observed on the HSQC spectrum of the ^{15}N,^{13}C-uridine-labeled G30A/U19A riboA71–adenine complex. The corresponding assignments were conducted by comparing the spectra of these constructs with that of the wild-type riboA71–adenine complex.



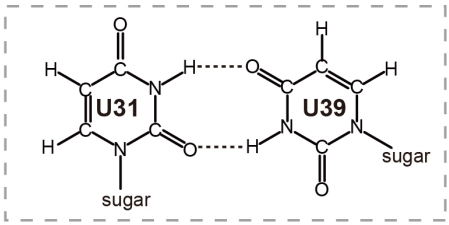
Supplementary Figure 8. The solution NMR spectra of riboA71-adenine complex, free riboA71 and mutant riboA71-adenine complex. Two-dimensional (2D) ^{1}H-^{15}N HSQC spectra of ^{15}N,^{13}C-labeled riboA71-adenine complex(A), ^{15}N,^{13}C-uridine-labeled wild-type riboA71 without adenine ligand (B) and ^{15}N,^{13}C-uridine-labeled G30A/U19A riboA71-adenine complex(C). The gray dashed lines highlight the assignments of the uredine imino cross peaks.

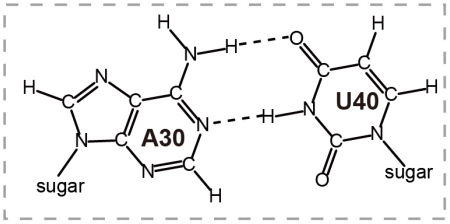
2.6 Crystal Structure of the RiboA71–Adenine Complex

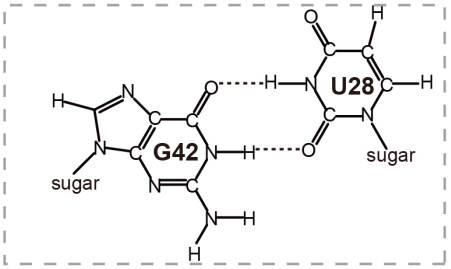
The structure of the riboA71–adenine complex was determined previously[21]. The complex adopts a tuning fork-like overall fold (Supplementary figure 9) containing three stems, P1–P3. Of them, P1 forms the handle, and P2 and P3 are aligned parallel to each other.[21] The hairpin loops, L2 and L3, anchor together to form a kissing loop interaction. J1–2, J2–3 and J3–1 are three junction-connecting segments that constitute the ligand-binding pocket. Five layers of base triplexes form around the ligand binding pocket. The planes of these five triplexes are in nearly parallel alignment and are stabilized by π–π packing. Adenine recognition occurs through the formation of a U47•U51•(adenine-U74) tetrad that involves residues in J2–3 and J3–1 in the center of the five layers. The two triplexes A23•(G46-C53) and water-mediated A73•(A52-U22), which involve residues from J1–2, J2–3 and J3–1, are located above the adenine binding plane. Another two triplexes, C50•(U75-A21) and U49•(A76-U20), which involve residues in J2–3 and the junction base pairs of stem P1, are below the adenine-binding plane. As shown in the structure, U47•U51•(adenine-U74) is sandwiched between the water-mediated A73•(A52-U22) and C50•(U75-A21) base triplexes such that the bound adenine is completely surrounded by RNA both along its periphery and above and below its base plane.

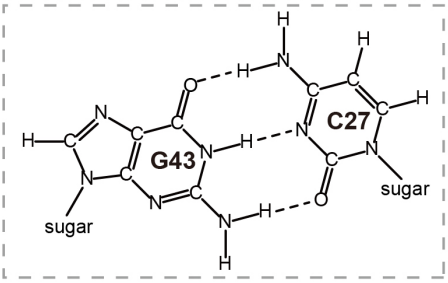
Supplementary figure 9 shows a structure model of the riboA71–adenine complex. Two non-canonical base pairs involving uridines form within the P2 stem, including a U31•U39 base pair and a U28•G42 base pair. Correspondingly, the imino groups of U31, U39 and U28 were detected on the 2D ^{1}H-^{15}N HSQC spectra, and all have ^{1}H chemical shifts typical for non-canonical base pairs. The imino groups of these three uridines are also observed on the 2D ^{1}H-^{15}N HSQC spectra of free riboA71 and demonstrate that these two base pairs

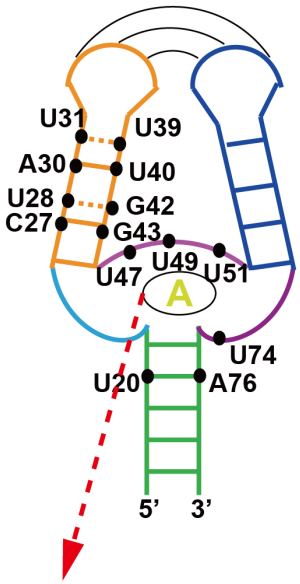
329 form in free riboA71, consistent with earlier research. In SSNMR, these uridines were observed on the 2D
330 hNH spectrum of free riboA71. Adenine binding includes multiple hydrogen bonds involving uridines,
331 including U47, U51 and U74. In addition to the hydrogen bonds between adenine and the stabilizing uridines,
332 a hydrogen bond between U47 and U51 is indicated by the crystal structure. Because of these bonds, the
333 imino groups of U47 and U51 were observed in the 2D ^{1}H-^{15}N HSQC spectrum of the riboA71–complex.[19,20]
334 In this work, both U47 and U51 were detected on the 2D hNH spectrum of the G30A/U19A riboA71–adenine
335 complex.











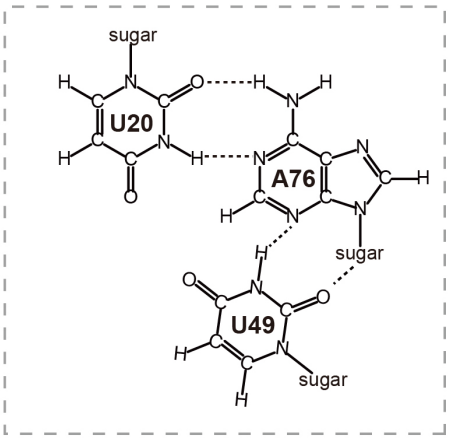
C53−G46 ∘ A23

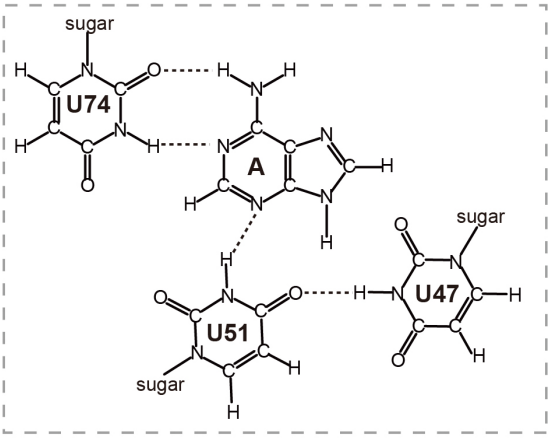
U22−A52 w A73

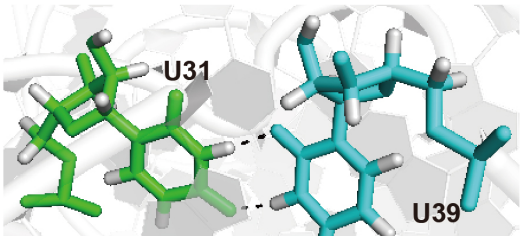
U47 ∘ U51 ∘ A−U74

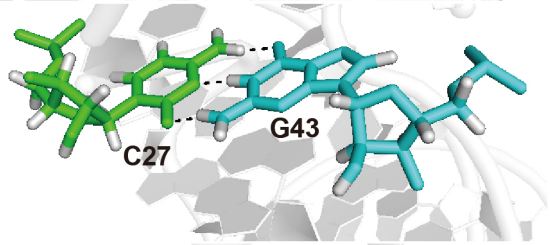
C50 ∘ U75−A21

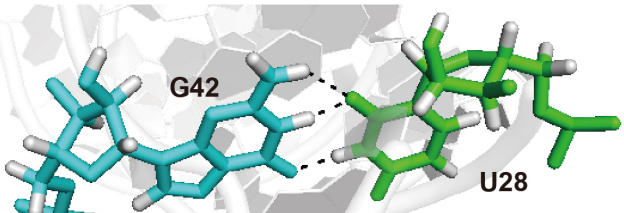
U49 ∘ A76−U20

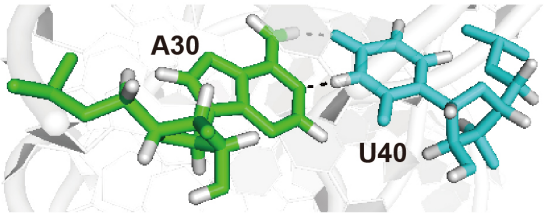


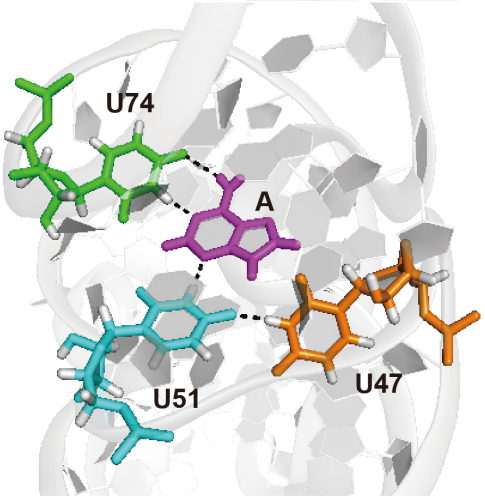


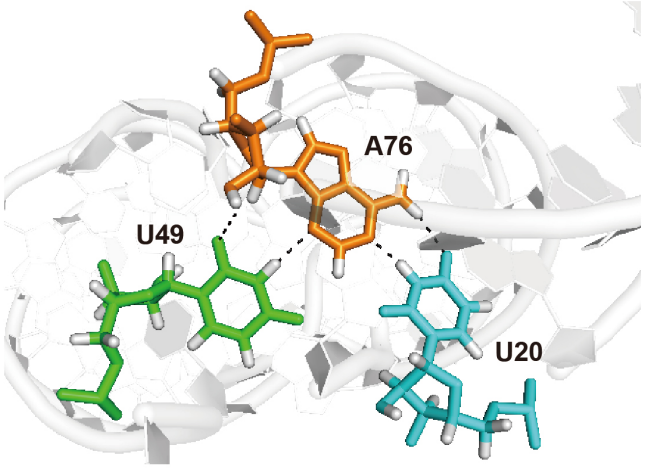








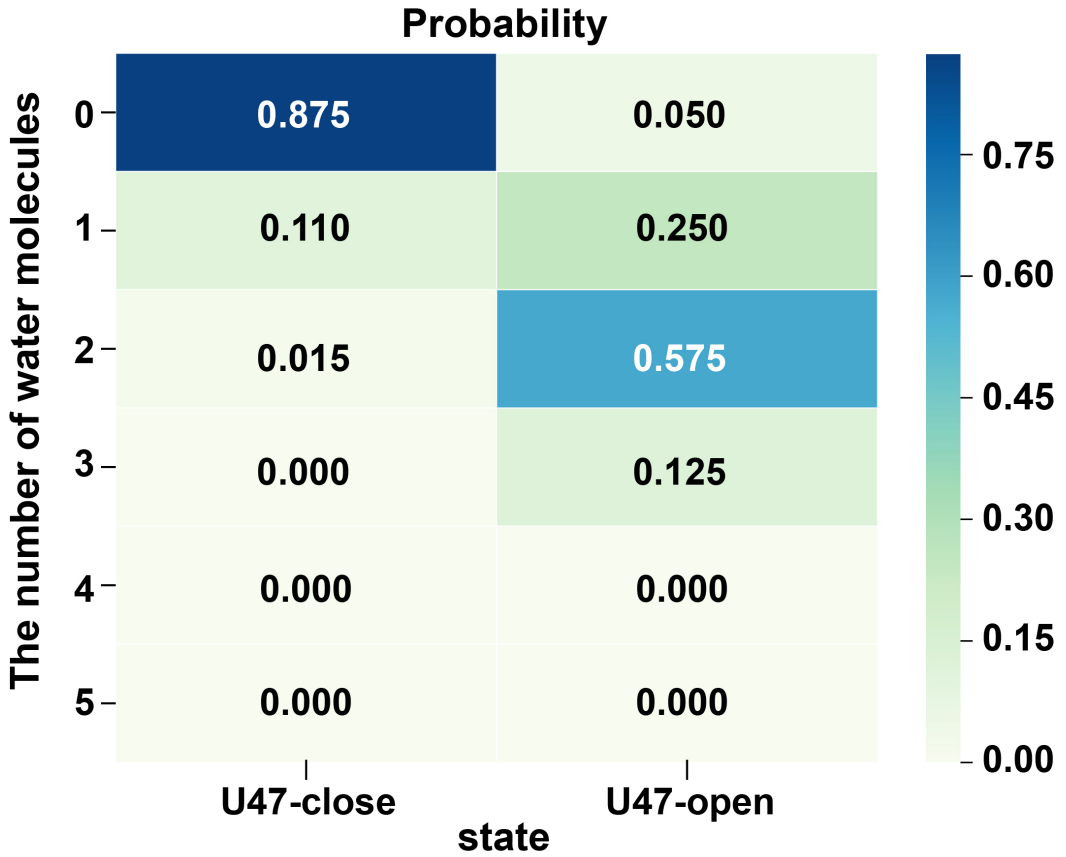




Supplementary Figure 9. Topological view of riboA71-adenine complex. A30-U40 and G43-C27 are shown as the representative of the canonical base pairs. The U31•U30, U28•G42, U49•A76-U20 and U47•U51•(adenine-U74) are shown to highlight the non-canonical base pairs

2.7 Water accessibility of U47 in open state and close state

In order to evaluate the influence of solvent accessibility of U47 in open state and close state, which may make contribution to the exchange rate, probability of water molecules within 3 Å of imino proton of U47 in the 1 μs MD simulation is calculated (**Supplementary figure** 10). The probabilities of water molecules close to the U47 are 12.5% and 95% respectively. Statistically, the number of the frames detecting the imino proton of U47 without any water molecules in close proximity is 875 out of 1000 frames in the close state. In contrast, this number is only 50 out of 1000 frames. Therefore, the probability value of no water around imino proton of U47 in the close state can be calculated to be 87.5%. Similarly, the probabilities of observing one or more water molecules are higher in open state than in close state. These results indicate that the water accessibilities of U47 in open state is higher than close state.



Supplementary Figure 10. Probability distribution of water molecules within 3 Å of imino proton of U47 in open and close state in 1 μs MD simulation.

Uncategorized References

1 Stagno, J. R. *et al.* Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242-246, doi:10.1038/nature20599 (2017).

2 Zhao, S. *et al.* High-resolution solid-state NMR spectroscopy of hydrated non-crystallized RNA. *Chem Commun (Camb)* **55**, 13991-13994, doi:10.1039/c9cc06552k (2019).

3 Yang, Y. *et al.* Proton-detected solid-state NMR detects the inter-nucleotide correlations and architecture of dimeric RNA in microcrystals. *Chem Commun (Camb)* **53**, 12886-12889, doi:10.1039/c7cc07483b (2017).

4 Thurber, K. R. & Tycko, R. Measurement of sample temperatures under magic-angle spinning from the chemical shift and spin-lattice relaxation rate of 79Br in KBr powder. *J Magn Reson* **196**, 84-87, doi:10.1016/j.jmr.2008.09.019 (2009).

5 Zhou, D. H. & Rienstra, C. M. High-performance solvent suppression for proton detected solid-state NMR. *J Magn Reson* **192**, 167-172, doi:10.1016/j.jmr.2008.01.012 (2008).

6 Fu, R., Miao, Y., Qin, H. & Cross, T. A. Probing Hydronium Ion Histidine NH Exchange Rate Constants in the M2 Channel via Indirect Observation of Dipolar-Dephased (15)N Signals in Magic-Angle-Spinning NMR. *J Am Chem Soc* **138**, 15801-15804, doi:10.1021/jacs.6b08376 (2016).

7 Morcombe, C. R. & Zilm, K. W. Chemical shift referencing in MAS solid state NMR. *J Magn Reson* **162**, 479-486 (2003).

8 Zhang, J. & Ferre-D'Amare, A. R. Dramatic improvement of crystals of large RNAs by cation replacement and dehydration. *Structure* **22**, 1363-1371, doi:10.1016/j.str.2014.07.011 (2014).

9 BIOVIA, D. S. Discovery Studio, 3.5. *San Diego: Dassault Systèmes* (2012).

10 Case., D. A. *et al.* AMBER 2018. *University of California, San Francisco* (2018).

11 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).

12 Zgarbova, M. *et al.* Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput* **7**, 2886-2902, doi:10.1021/ct200162x (2011).

13 Lu, X. J., Bussemaker, H. J. & Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* **43**, e142, doi:10.1093/nar/gkv716 (2015).

14 Kim, W. *et al.* Base-pair opening dynamics of primary miR156a using NMR elucidates structural determinants important for its processing level and leaf number phenotype in Arabidopsis. *Nucleic Acids Res* **45**, 875-885, doi:10.1093/nar/gkw747 (2017).

15 Choi, S. R. *et al.* Base-pair Opening Dynamics of Nucleic Acids in Relation to Their Biological Function. *Comput Struct Biotechnol J* **17**, 797-804, doi:10.1016/j.csbj.2019.06.008 (2019).

16 Snoussi, K. & Leroy, J. L. Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry* **40**, 8898-8904, doi:10.1021/bi010385d (2001).

17 Szulik, M. W., Voehler, M. & Stone, M. P. NMR analysis of base-pair opening kinetics in DNA. *Curr Protoc Nucleic Acid Chem* **59**, 7 20 21-18, doi:10.1002/0471142700.nc0720s59 (2014).

18 Kierzek, R., Burkard, M. E. & Turner, D. H. Thermodynamics of single mismatches in RNA duplexes. *Biochemistry* **38**, 14214-14223, doi:10.1021/bi991186l (1999).

19 Warhaut, S. *et al.* Ligand-modulated folding of the full-length adenine riboswitch probed by NMR and single-molecule FRET spectroscopy. *Nucleic Acids Res* **45**, 5512-5522, doi:10.1093/nar/gkx110 (2017).

20 Reining, A. *et al.* Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499**, 355-359, doi:10.1038/nature12378 (2013).

21 Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* **11**, 1729-1741, doi:10.1016/j.chembiol.2004.11.018 (2004).