Supplementary Information

Refining RNA force field with Small-Angle X-ray Scattering of Helix-Junction-Helix RNA

Weiwei He,†‡ Nawavi Naleem,† Diego Kleiman,† and Serdal Kirmizialtin*,†

†Chemistry Program, Science Division, New York University Abu Dhabi
d‡Department of Chemistry, New York University

E-mail: serdal@nyu.edu

For the study, a two way junction (HJH) in different salt conditions and single stand RNA sequence of oligo-Us with various lengths are investigated using both brute force MD and well tempered metadynamics simulations (WTMD). Table S1 below summarizes the simulation set up for each system. The modified force field HB-CUFIX is available at [https://gitlab.com/KirmizialtinLab/hb_cufix](https://gitlab.com/KirmizialtinLab/hb_cufix)

Helix Junction Helix Simulations

The initial model of the RNA HJH construct, containing two 12 based-paired RNA duplexes joined by a junction of rU5 (Fig. 1a-b), was built using Nucleic Acid Builder (NAB) assuming A-form geometry. Then, the HJH was placed in a simulation box by aligning the long axis parallel to the z-axis. To avoid end effects, the simulation box along the z axis was extended by 20 Å at both ends of RNA HJH, resulted in an initial cubic box of side lengths of 13.0 nm. The RNA is represented by aimed force fields (AMBER99SB, DES-AMBER,
Table S1: Summary of simulations

| System  | Simulation type | Simulation length | Force field                          | Salt & Conc.          |
|---------|-----------------|-------------------|--------------------------------------|-----------------------|
| RNA HJH | WTMD            | ~250ns - ~1.5µs   | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM KCl             |
| RNA HJH | Brute force MD  | ~1.5µs            | AMBER99SB                           | 100mM KCl             |
| RNA HJH | WTMD            | ~2.2µs            | HB-CUFIX                            | 50mM KCl              |
| RNA HJH | WTMD            | ~1.5µs            | HB-CUFIX                            | 200mM KCl             |
| RNA HJH | WTMD            | ~400ns - ~1.4µs   | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM MgCl₂ + ~24mM KCl |
| ssRNA rU₄| Brute force MD  | 3.0µs             | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM KCl             |
| ssRNA rU₅| Brute force MD  | 3.0µs             | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM KCl             |
| ssRNA rU₇| Brute force MD  | 3.0µs             | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM KCl             |
| ssRNA rU₁₀| Brute force MD | 3.0µs             | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM KCl             |
| ssRNA rU₁₅| Brute force MD | 3.0µs             | AMBER99SB DES-AMBER H-AMBER99SB H-CUFIX HB-CUFIX | 100mM KCl             |
CUFIX, H-AMBER99SB, H-CUFIX, HB-CUFIX, respectively). DES-RES parameters were adopted from the Ref. and compiled in Gromacs topology format. All systems except DES-AMBER were solvated with TIP3P water. For DES-AMBER simulations we used TIP4P-D as suggested in Ref. Ions were added to match experimental conditions. For ions, we used Dang parameters in AMBER99SB/H-AMBER99SB and Cheatham parameters in DES-AMBER. Mg\(^{2+}\) parameters were adapted from Aqvist’s work and used in AMBER99SB, H-AMBER99SB and DES-AMBER. For H-CUFIX and HB-CUFIX we employed the same ion parameters as published in CUFIX. We study the RNA in 50, 100, and 200mM KCl. We also investigate the HJH RNA in 100mM MgCl\(_2\). Table S2 summarizes all the simulation setups for the HJH systems.

MD simulations and analysis were carried out using GROMACS 5.0.5 suite of programs. Periodic boundary conditions were applied in all directions during simulations. Long-range electrostatic interactions were computed using the particle mesh Ewald (PME) summation method with a grid spacing of 0.12 nm and an interpolation of order 4. Dispersion correction was employed for van der Waals interactions. We used a cutoff of 1.1 nm with non-bonded interactions. The neighbor search was treated with a cutoff radius of 1.1 nm and updated every 10 steps. The covalent bond lengths of the water and nucleic acids were constrained by SETTLE and LINCS algorithms, respectively. All simulations used a time step of 2 fs.

A 5000-step energy minimization was performed for the solvated system to remove possible bad contacts that may arise during the random placement of water and ions. To prepare the structures for Well-tempered Metadynamics (WTMD) simulations, we used equilibration steps involving volume and solvent relaxation. For that, a 2-ns MD at a constant pressure of 1 atm and temperature of 300K served to determine the volume of the simulation box. Subsequently, we performed 200 ns constrained MD at canonical ensemble (NVT), where heavy atoms of RNA were restrained while water and ions were allowed to move. Coordinates of the last step of the simulations were saved and used as the starting structures for WTMD
|                | 50mM KCl     | 100mM KCl       | 200mM KCl       | 100mM MgCl₂      |
|----------------|--------------|-----------------|-----------------|------------------|
| RNA+water+ion  | 106 K⁺       | 172 K⁺          | 304 K⁺          | 148 Mg²⁺         |
| box            | 56 Cl⁻       | 122 Cl⁻         | 254 Cl⁻         | 40 K⁺            |
|                | 71527 H₂O    | 71856 H₂O       | 71592 H₂O       | 286 Cl⁻          |
|                |              |                 |                 | 71592 H₂O        |
| Water+ion      | 66 KCl       | 121 KCl         | 260 KCl         |                  |
| box            | 70605 H₂O    | 70499 H₂O       | 70220 H₂O       |                  |

**Well-tempered Metadynamics for Helix-Junction-Helix RNA**

To exhaustively explore the conformational space of HJH RNA, we performed Well-tempered Metadynamics (WTMD) implemented in PLUMED. We reduced the description of the dynamics of HJH to two trivial collective variables (CV): helix-helix distance \(d\) and customized azimuthal rotation \(\theta\) (Fig. 1a-b). During WTMD runs a harmonic restraint with a stiffness constant of 100 \(kJ \cdot mol^{-1} \cdot nm^{-2}\) were applied to two select phosphorus (P) atoms of the bottom helix such that its orientation remain fixed for ease of visualization. The upper helix is let free to move. Gaussians were deposited every 1 ps with a height starting from \(W=0.6 \text{ kJ/mol}\). We gradually decreased the heights by a bias factor of 6.0. The widths of the Gaussians were set to \(\sigma_d = 0.1 \text{ nm}\) for \(d\) and \(\sigma_\theta = 0.2 \text{ rad}\) for \(\theta\), respectively. To ensure convergence, we monitored the time evolution of the Gaussian heights deposited during the simulation (Fig. S1a-c, Fig. S3a-c, Fig. S7a-d). The simulations were considered converged when the heights of the deposited Gaussians in the space of the CVs decayed and converged to a threshold value < 0.005 kJ/mol. To further assess the convergence, the block analysis was employed (Fig. S4a-b). The convergence to a flat ordinate suggest the system under study reached ergodicity.
Data Analysis for Helix Junction Helix

Simulation data were recorded for every 2 ps used for further analysis. The saved conformations were used for further analysis. Namely, we compute the minimum distance between the two helices, the Small Angle X-Ray Scattering, orientational correlations of the bases, Förster resonance energy transfer (FRET) as well as computing duplex parameters. Details of our analysis methodology is summarized below.

Minimum Distance Analysis

The mindist plugin of GROMACS was implemented to analyze the minimum distance between the bottom helix (H1) and upper helix (H2). The closest atom pairs at each conformation were analyzed and the corresponding structures were fetched out from MD trajectory for further analysis if it is below a threshold value of 2.0 Å. Fig. S1d-f shows the time evolution of the H1-H2 minimum distance of AMBER99SB, DES-AMBER, and CUFIX during WTMD simulation. The same analysis was employed after the -HO correction and stacking correction (Fig. S3d-f).

Computing Small Angle X-ray Scattering profiles of HJH from MD

The conformations from WTMD were used to compute the average SAXS from simulations. About 2000 frames in equal intervals were extracted from simulations. For each frame, the SAXS profiles are computed.

To compute the SAXS profiles, we employed WAXSiS GROMACS program, the solvent/ion shell, excluded volume, and solvent degrees of freedom were considered explicitly. Following the theory of excess electron density, the buffer-subtracted intensity was measured by $I(q) = I_A(q) - I_B(q)$, where $I_A$ and $I_B$ denote the scattering intensities of sample solution (RNA+water+ion system, see Fig. 1c) and corresponding solvent background (water+ion system), respectively. From the RNA+water+ion system, we constructed the molecular envelope using a cutoff of 10 Å from the surface of HJH. The same envelope
was applied to the water+ion system. For buffer subtraction, a buffer solution with ionic
strength identical to the RNA+water+ion system was prepared by adding ion pairs and wa-
ter molecules in a periodic box with dimensions and concentrations matching those implied
in the RNA+water+ion system. We run 20-ns-long MD simulations in the NVT ensemble
to generate the bulk ensemble.

The agreement between the measured and simulated SAXS profiles was quantified using
the linear $\chi^2$ metric, the smaller the value, the better the fitness.

$$\chi^2 = \frac{1}{n - 1} \sum_{i=1}^{n} \left\{ \frac{[I_{exp}(q_i)] - [I_{com}(q_i)]}{\sigma(q_i)} \right\}^2,$$

where $I_{com}(q_i)$ and $I_{exp}(q_i)$ are the computed and experimental scattering intensities at $q_i$
respectively. $n$ is the number of $q$ points. The experimental error at $q_i$ is denoted as $\sigma(q_i)$.

**Orientation Correlation Function (OCF)**

To quantify the orientational correlation of the bases in the junction region of HJH, we
calculate the orientation correlation function (OCF):\(^{20}\)

$$OCF = \langle \cos \theta_{ij} \rangle = \langle \hat{r}_i \cdot \hat{r}_j \rangle$$

Here, $\hat{r}_i$ is the normalized bond vector between the $i^{th}$ and $(i + 1)^{th}$ phosphate in a given
chain. The average dot product among the bond vectors is calculated and expressed as a
function of separation distance ($|i - j|$). See details in Fig. S5a.

**Analysis of Duplex Parameters**

To monitor the effect of corrections on the duplex geometry, we compute average duplex
parameters from WTMD simulations using the 3-DNA program.\(^{21}\) We compare these pa-
rameters with average A-form duplex parameters from available structures in the protein
data bank. For groove geometries and P-P distances, we subtracted 5.8 Å from the obtained
values to take into account the vdw radii of the phosphate group. The same calculations were applied to 4 duplex sequences whose NMR structures with similar salt conditions were reported.

Computing FRET efficiency from MD simulations

To assess the performance of the force fields in divalent ions, we benchmarked the HJH against the FRET study. The simulations were done at \( \sim100\text{mM Mg}^{2+} \) and \( \sim24\text{mM K}^+ \) to mimic the experimental conditions. The conformational ensembles generated by WTMD were used to estimate the FRET efficiencies employing the free dye method.

Briefly, for each selected conformation the position of dyes whose configurations were apriori sampled by free MD simulations in water box was aligned to the designed label sites of RNA HJH (see the label sites and dye dynamics in Fig. 5e). Then the overlapping conformations (dye-dye or dye-RNA) were removed from the pool. The remaining combinations of the dye pair trajectories were used to compute the FRET efficiencies. This way, the explicit calculations of dye positions and orientations allow a higher accuracy in FRET calculations from simulations. The FRET calculations were handled using HandyFRET program that implemented free dye method. The dye molecules were aligned and attached to the sugar ring of each select residue by a linker as shown in Fig. S9. The details of how these dye molecules are modelled is explained in Ref.

Following the experimental work, the donor and acceptor used in our study are Alexa 488 and Cy5, respectively, and accordingly, for Förster radius, we took \( R_0 = 52 \AA \). Similar to SAXS calculations, computed values from each conformation were weighted by the Boltzmann factor obtained from the free energy mapped to CVs.

To compute average \( E_{FRET} \), we assume the rotational correlation time of the dyes \( \tau_{\text{rot}} \) to be faster than the fluorescence lifetime \( \tau_{\text{FRET}} \) (i.e. \( \tau_{\text{rot}} \ll \tau_{\text{FRET}} \)) leading to the dynamic averaging regime for \( E_{FRET} \) value as:
\[ E_{FRET} = \frac{\frac{3}{2} \langle \kappa^2 \rangle R_0^6}{\frac{3}{2} \langle \kappa^2 \rangle R_0^6 + r^6} \]  

(3)

where \( R_0 \) is the Förster radius.

The orientations of dyes evaluated by the \( \kappa^2 \) assumed an isotropic distributions of orientations of the donor and acceptor giving rise to a probability distribution:

\[
P(\kappa^2) = \begin{cases} 
\frac{1}{2\sqrt{3}\kappa^2} \ln(2 + \sqrt{3}) & 0 \leq \kappa^2 \leq 1 \\
\frac{1}{2\sqrt{3}\kappa^2} \ln\left(\frac{2+\sqrt{3}}{\sqrt{\kappa^2} + \sqrt{\kappa^2 - 1}}\right) & 1 \leq \kappa^2 \leq 4
\end{cases}
\]

(4)

where \( \kappa \) is the orientation factor determined by the positioning of donor and acceptor.\textsuperscript{24,25}

**Single Strand RNA simulations**

We study the single strand RNAs (ssRNA) of uracil oligomers of varying lengths, rU\(_4\), rU\(_5\), rU\(_7\), rU\(_{10}\), and rU\(_{15}\). Initial structures of ssRNAs were generated with Avogadro software.\textsuperscript{26} The RNAs were modeled by each of the force fields under study (AMBER99SB, DES-AMBER, CUFIX, HB-CUFIX). The GROMACS 5.0.5 suite of programs was used to generate the minimized solvated structures and topologies. The simulation box size was 6 nm x 6 nm x 6 nm for all systems, except rU\(_{15}\). For the rU\(_{15}\), a simulation box of 8 nm x 8 nm x 8 nm was used to ensure enough space for extended states. The water was modeled by TIP3P\textsuperscript{3} for AMBER99SB, CUFIX, and HB-CUFIX. For DES-AMBER, we used TIP4P-D\textsuperscript{4}. To mimic experimental conditions K\(^+\) and Cl\(^-\) were added such that the bulk concentration was set to 100 mM. For that, we used the aforementioned compatible Amber99SB, DES-AMBER, and CUFIX ion parameters. Table S3 summarizes the setups for each ssRNA system.

The OpenMM 7.4.2\textsuperscript{27} was used to perform MD simulations, except for DES-AMBER. We employed DES-AMBER parameters with a compatible water model which includes virtual sites. The GROMACS starting files with virtual sites are not supported yet by OpenMM.
Therefore, we used GROMACS software to simulate DES-AMBER setups. For OpenMM the solvated systems were minimized with OpenMM using the steepest descend algorithm with default parameters of tolerance of 1 kJ/mol. The minimization was performed by keeping all solute atoms and solvent atoms flexible. The minimized structures were equilibrated for 100 ns, this time with all bond constraints of the solute and the solvent invoked. In the production runs however, only bond lengths with H atoms of the solute were constrained, using SHAKE\textsuperscript{28} algorithm. Similar to the HJH simulations we used SETTLE\textsuperscript{12,29} to constrain the water. The nonbonded interactions were calculated with cutoff values of 10.0 Å where particle mesh Ewald method (PME)\textsuperscript{30} was used to treat electrostatics. The equilibration and production runs were performed in the NPT ensemble, where the temperature was set at 300 K and pressure to 1 atm. The equations of motion were solved using the Langevin Integrator\textsuperscript{31} with input parameters of the friction coefficient of 1.0 picoseconds\textsuperscript{−1} and timestep of 2 femtoseconds. The pressure was maintained by Monte Carlo barostat, where input parameters were set to 1 atm at 300 K, while attempting pressure updates with a frequency of every 25 steps. Three microseconds of simulation time were used to sample conformations. In the production runs, the coordinates were saved for every 10 ps interval for data analysis. In DES-AMBER simulations we used similar settings to OpenMM parameters except for the solute bond constraining we used LINCS\textsuperscript{13} and for temperature and pressure coupling we employed velocity rescaling\textsuperscript{32} and Berendsen\textsuperscript{33} methods, respectively.

| rU\textsubscript{N} | N\textsubscript{H\textsubscript{2}O} | N\textsubscript{K\textsuperscript{+}} | N\textsubscript{Cl\textsuperscript{−}} |
|-------------------|---------------------|-------------------|-------------------|
| 4                 | 6941                | 16                | 13                |
| 5                 | 6940                | 17                | 13                |
| 7                 | 6929                | 19                | 13                |
| 10                | 6908                | 22                | 13                |
| 15                | 16743               | 45                | 31                |
Data Analysis of ssRNA

We study the local and global structural parameters of ssRNAs. We compute the radius of gyration, end to end distance, $^3J$ couplings, stacking and intercalation ratios, as well as persistence lengths.

Computing the Free Energy Surface of $rU_4$

The end-to-end distance, $R$, and radius of gyration, $R_g$ of RNA conformations were computed using MDTraj.\textsuperscript{34} Using the simulation data, we estimate the free energy, from the probability density $P(R, R_g)$, $F(R, R_g) = -k_BT \ln(P(R, R_g))$, where $k_BT$ represents the thermal energy.

Computing the $^3J$ Couplings and base stacking of $rU_4$

The Barnaba\textsuperscript{35} package was used to calculate the $^3J$ couplings and base stacking. The first 300 ns of simulations were discarded to avoid initial bias. We compute H5'-P, H5''–P, H4’–H5’, H4’–H5”, H1’–H2’, H2’–H3’, and H3’–H4’ $^3J$ couplings for each frame and averaged from time series. For comparison, the experimental data from earlier works\textsuperscript{2,36} were used. To quantify the similarity between computed, and experimental, $^3J$ couplings, we used the following equation:

$$\chi = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left( \frac{^3J_{\text{exp}} - ^3J_{\text{sim}}}{\sigma_{\text{exp}}} \right)^2}$$

where $M$ is the number of frames, $^3J_{\text{sim}}$ and $^3J_{\text{exp}}$ are calculated and experimental $^3J$ coupling constants respectively, and $\sigma_{\text{exp}}$ is the standard deviation in the experimental data, that is taken to be 1.5 Hz as employed in earlier studies.\textsuperscript{36,37} Based on Eq. 5, the smaller the $\chi$ the better the agreement between experiment and simulation. The summary of the $\chi$ values of the $^3J$ coupling for each residue and averages are displayed in Table 3. The results of the stacking analysis are reported in Table 4 for the whole trajectory and in Table S6 for the most dominant cluster.
Computing the Persistence Length of ssRNAs

To compute the persistence length, $L_P$, we solved the following equation:

$$R_g^2 = \left(\frac{L_P^3}{3}\right) - L_P^2 + \left(\frac{2L_P^3}{l}\right) - \left(\frac{2L_P^4}{l^2}\right) (1 - e^{-l/L_P})$$

(6)

where $l = (N - 2)a$ is the contour length with $N$ is the number of bonds and $a$ is the average monomer length estimated from adjacent P-P distances.
Figure S1: Time evolution of the collective variable $d$ (black) and Gaussian hills (gray) during metadynamics simulations of a) AMBER99SB, b) DES-AMBER and c) CUFIX. The change in the minimum distance between helices is monitored in each force field: d) AMBER99SB, e) DES-AMBER and f) CUFIX. The insets of d-f show a snapshot from collapsed states.
Figure S2: Time evolution of the minimum distance between H1 and H2 from brute force MD simulations for AMBER99SB.
Figure S3: Time evolution of the collective variable $d$ (black) and Gaussian hills (gray) during metadynamics simulations of a) H-AMBER99SB, b) H-CUFIX and c) HB-CUFIX. The minimum distance between helices is monitored for e) H-AMBER99SB, f) H-CUFIX and g) HB-CUFIX, respectively.
Figure S4: To check the convergence of the free energy surface in metadynamics simulations we monitor the average error on the free energy as a function of the size of blocks a) H-CUFIX, b) HB-CUFIX.
Figure S5: The orientational correlation function (OCF) computed for the junction region of HJH. a) Shows the description of OCF. b) OCF of force field under study as a function of base pair separation. c) A snapshot from each pool. Conformations were selected from the pool of the lowest free energy region of each force field by clustering based on the junction region.
Figure S6: Comparison of the experimental SAXS data\textsuperscript{[38]} with computed SAXS profiles for HB-CUFIX. a) The Kratky plots of HJH in 50mM KCl, giving rise to $\chi^2 = 1.69$. b) The Kratky plots of HJH in 200mM KCl, giving rise to $\chi^2 = 1.47$. 
Figure S7: Time evolution of the collective variable $d$ (black) and Gaussian hills (gray) for HJH in MgCl$_2$ for each forcefield under study. a) AMBER99SB, b) DES-AMBER, c) CUFIX, and d) HB-CUFIX. The minimum distance between helices is monitored during sampling for e) AMBER99SB, f) DES-AMBER, g) CUFIX and g) HB-CUFIX, respectively. The inset of g) shows the partial fraying of HJH during metadynamics simulation in CUFIX.
Figure S8: The comparison of dye to dye distance distributions and FRET efficiency distributions in Mg\(^{2+}\) salt condition for each forcefield, a-d) are distance distributions for AMBER99SB, DES-AMBER, CUFIX, and HB-CUFIX, respectively, e-h) are FRET efficiencies. The insets between the graphs show the dynamics of dye molecules attached to the representative conformation of each pool.
Figure S9: The implementation of free dye method for FRET calculations. a) Fluorescent label sites are U34 and U49 for acceptor Cy5 and donor Alexa488, respectively. b) The chromophore pairs with linker were attached to the sugar ring of select residue, U34 and U49.
Figure S10: The time evolution of radius of gyration, $R_g$, of sequence rU$_4$ from brute force MD. Different colors represent the force fields under study, AMBER99SB (green), DES-AMBER (magenta), CUFIX (yellow), and HB-CUFIX (cyan). The dashed lines display the average values.
Table S4: The comparison of experiment and simulations in $^3J$ coupling parameters for $rU_4$ all the snapshots in the most dominant cluster referred in Fig. 6.

| Residue | AMBER99SB | DES-AMBER | CUFIX | HB-CUFIX |
|---------|-----------|-----------|-------|----------|
| H5' - P | 2         | 2.05      | 1.08  | 1.91     | 1.34     |
|         | 3         | 1.61      | 2.12  | 2.17     | 1.56     |
|         | 4         | 0.74      | 1.17  | 0.91     | 2.03     |
| ($\beta$) | 3         | 3.93      | 0.62  | 3.90     | 0.64     |
|         | 4         | 1.60      | 1.56  | 1.66     | 1.26     |
| H5'' - P | 2         | 0.61      | 0.63  | 0.69     | 0.78     |
|         | 3         | 3.93      | 0.62  | 3.90     | 0.64     |
|         | 4         | 1.60      | 1.56  | 1.66     | 1.26     |
| H4' - H5' | 1         | 1.61      | 1.79  | 2.11     | 2.35     |
| ($\gamma$) | 2         | 1.33      | 1.18  | 1.36     | 0.86     |
|         | 3         | 1.06      | 1.19  | 1.59     | 0.80     |
| H4' - H5'' | 1        | 1.03      | 0.84  | 1.70     | 1.63     |
| ($\gamma$) | 2        | 0.94      | 0.70  | 0.94     | 0.88     |
|         | 3        | 0.77      | 0.68  | 0.72     | 0.93     |
| H3' - P | 1         | 1.92      | 1.60  | 1.65     | 2.01     |
| ($\epsilon$) | 2        | 1.48      | 1.28  | 1.72     | 1.74     |
|         | 3        | 0.89      | 1.53  | 1.03     | 1.81     |
| H1' - H2' | 1         | 4.26      | 1.85  | 4.34     | 2.38     |
| ($\nu_1$) | 2         | 3.65      | 3.84  | 3.72     | 2.84     |
|         | 3         | 3.56      | 3.61  | 3.72     | 2.47     |
|         | 4         | 3.66      | 4.32  | 3.27     | 4.33     |
| H2' - H3' | 1         | 0.78      | 0.83  | 0.71     | 0.87     |
| ($\nu_2$) | 2         | 0.75      | 0.65  | 0.72     | 0.74     |
|         | 3         | 0.82      | 1.16  | 0.75     | 0.73     |
| H3' - H4' | 1         | 2.67      | 3.67  | 2.74     | 3.87     |
| ($\nu_3$) | 2         | 3.17      | 2.83  | 3.05     | 3.65     |
|         | 3         | 2.66      | 1.85  | 2.60     | 2.60     |
| Average backbone |         | 1.44      | 1.20  | 1.60     | 1.37     |
| Average ribose ring |     | 2.60      | 2.46  | 2.56     | 2.53     |
| Average Total |         | 1.90      | 1.70  | 1.99     | 1.84     |

The difference between experiment and simulation is measured by $\chi$ described in $SI$ (equation 5) for all the snapshots in the most dominant cluster referred to Fig. 6. The $^3J$ coupling data for experiments were taken from Ref.26 and an experimental error of 1.5 Hz was employed as in Ref.26,37.
Table S5: The comparison of experiment and simulations in $^3J$ coupling (Hz) parameters and base stacking for $rU_4$ the cluster center snapshots of the most dominant cluster which shown in Fig. 6e-h.

| Residue       | NMR | AMBER99SB | DES-AMBER | CUFIX | HB-CUFIX |
|---------------|-----|-----------|-----------|-------|----------|
| H5' - P       | 2   | 2.0       | 3.34      | 1.36  | 2.62     | 2.04     |
| (β)           | 3   | 2.0       | 1.24      | 2.96  | 1.01     | 2.35     |
| 4             | 2.0 | 2.70      | 4.35      | 1.03  | 2.22     |
| H5'' - P      | 2   | 2.0       | 1.63      | 3.85  | 2.15     | 2.75     |
| (β)           | 3   | 2.0       | 7.03      | 1.88  | 5.21     | 2.40     |
| 4             | 2.0 | 2.08      | 1.18      | 6.15  | 2.53     |
| H4' - H5'     | 1   | 3.7       | 1.22      | 2.06  | 2.37     | 1.42     |
| (γ)           | 2   | 2.5       | 0.24      | 0.52  | 0.07     | 1.77     |
| 3             | 2.5 | 1.01      | 2.66      | 1.96  | 1.73     |
| H4' - H5''    | 1   | 2.8       | 1.86      | 1.05  | 0.82     | 1.64     |
| (γ)           | 2   | 2.5       | 3.42      | 2.84  | 4.63     | 1.30     |
| 3             | 2.5 | 2.11      | 0.63      | 1.13  | 1.33     |
| H3' - P       | 1   | 8.2       | 8.01      | 7.87  | 8.21     | 5.23     |
| (ε)           | 2   | 7.8       | 8.40      | 10.6  | 10.76    | 3.66     |
| 3             | 7.8 | 7.76      | 6.11      | 6.05  | 5.91     |
| H1' - H2'     | 1   | 4.5       | 9.18      | 1.07  | 9.75     | 0.29     |
| (ν1)          | 2   | 5.2       | 10.63     | 9.99  | 9.98     | 0.16     |
| 3             | 5.2 | 10.98     | 10.21     | 11.18 | 1.45     |
| 4             | 3.7 | 1.81      | 11.68     | 0.55  | 10.64    |
| H2' - H3'     | 1   | 4.7       | 4.59      | 3.45  | 4.68     | 4.27     |
| (ν2)          | 2   | 4.6       | 3.80      | 5.43  | 5.15     | 2.72     |
| 3             | 4.8 | 4.20      | 6.19      | 4.88  | 5.95     |
| H3' - H4'     | 1   | 5.3       | 0.63      | 11.44 | 0.26     | 11.23    |
| (ν3)          | 2   | 5.5       | 0.21      | 1.20  | 0.60     | 11.19    |
| 3             | 5.4 | 0.44      | 4.00      | 1.80  | 10.36    |
| 1-2 stack     | none| −         | −         | −     | −        | −        |
| 2-3 stack     | none| −         | √         | −     | −        | −        |
| 3-4 stack     | none| −         | −         | −     | −        | −        |
| 1-3 stack     | none| √         | −         | √     | √        | −        |
| 1-4 stack     | none| −         | −         | −     | −        | −        |
| 2-4 stack     | none| √         | −         | √     | −        | −        |

The $^3J$ coupling (Hz) and stacking calculated for the cluster center conformation of the most dominant cluster shown in the main text Fig. 6e-h. The $^3J$ coupling (Hz) and stacking data for experiments were taken from Ref. The Barnabas software was used to calculate the stacking interactions. In the $rU_4$ chain, the nucleotides are numbered 1 to 4 starting with the 5' end. As an example, 1–2 stack represents the bases of nucleotides 1 and 2 stacked on each other. Also, it is possible to have multiple stacking interactions in the same MD snapshot.
Table S6: The comparison of experiment and simulations of stacking for rU₄ for all snapshots in the most dominant cluster referred to Fig. 6.

|        | NMR | AMBER99SB | DES-AMBER | CUFIX | HB-CUFIX |
|--------|-----|-----------|-----------|-------|----------|
| 1-2 stack | none | 0%        | 0%        | 0%    | 51.9%    |
| 2-3 stack | none | 0%        | 0%        | 0%    | 47.2%    |
| 3-4 stack | none | 0%        | 0%        | 0%    | 1.0%     |
| 1-3 stack | none | 59.5%     | 77.7%     | 61.4% | 0%       |
| 1-4 stack | none | 0%        | 22.2%     | 0%    | 0%       |
| 2-4 stack | none | 40.4%     | 0%        | 38.6% | 0%       |

The experimental NMR data were taken from Ref.36 and the Barnaba35 software was used to calculate the stacking interactions in all snapshots in the most dominant cluster. In the rU₄ chain, the nucleotides are numbered 1 to 4 starting with the 5’ end. As an example, 1–2 stack represents the bases of nucleotides 1 and 2 stacked on each other. Also, it is possible to have multiple stacking interactions in the same MD snapshot.
Figure S11: The orientational correlation function (OCF) of single strand RNA of sequence rU$_{15}$ as a function of P-P $|i - j|$ separation (see Fig. S5a for details). Different colors represent different force fields under study, AMBER99SB (green), DES-AMBER (magenta), CUFIX (yellow), and HB-CUFIX (cyan).

References

(1) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. Journal of computational chemistry 2005, 26, 1668–1688.

(2) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. Proc. Natl. Acad. Sci. U.S.A. 2018, 115, E1346–E1355.

(3) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics 1983, 79, 926–935.
(4) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B* 2015, 119, 5113–5123.

(5) Dang, L. X. Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: a molecular dynamics study. *Journal of the American Chemical Society* 1995, 117, 6954–6960.

(6) Joung, I. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The journal of physical chemistry B* 2008, 112, 9020–9041.

(7) Aqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* 1990, 94, 8021–8024.

(8) Yoo, J.; Aksimentiev, A. Improved parameterization of amine–carboxylate and amine–phosphate interactions for molecular dynamics simulations using the CHARMM and AMBER force fields. *Journal of chemical theory and computation* 2016, 12, 430–443.

(9) Yoo, J.; Aksimentiev, A. Refined parameterization of nonbonded interactions improves conformational sampling and kinetics of protein folding simulations. *The journal of physical chemistry letters* 2016, 7, 3812–3818.

(10) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* 2008, 4, 435–447.

(11) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of chemical physics* 1993, 98, 10089–10092.

(12) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE
algorithm for rigid water models. *Journal of computational chemistry* **1992**, *13*, 952–962.

(13) Hess, B.; Bekker, H.;! Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463–1472.

(14) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008**, *100*.

(15) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A., et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180*, 1961–1972.

(16) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604–613.

(17) Knight, C. J.; Hub, J. S. WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic acids research* **2015**, *43*, W225–W230.

(18) Park, S.; Bardhan, J. P.; Roux, B.; Makowski, L. Simulated x-ray scattering of protein solutions using explicit-solvent models. *The Journal of chemical physics* **2009**, *130*, 04B607.

(19) Chen, P.-c.; Hub, J. S. Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophysical journal* **2014**, *107*, 435–447.

(20) Plumridge, A.; Andresen, K.; Pollack, L. Visualizing disordered single-stranded RNA: connecting sequence, structure, and electrostatics. *Journal of the American Chemical Society* **2019**, *142*, 109–119.
(21) Lu, X.-J.; Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nature protocols 2008, 3, 1213.

(22) Sutton, J. L.; Pollack, L. Tuning RNA flexibility with helix length and junction sequence. Biophysical journal 2015, 109, 2644–2653.

(23) Walczewska-Szewc, K.; Corry, B. Accounting for dye diffusion and orientation when relating FRET measurements to distances: three simple computational methods. Physical Chemistry Chemical Physics 2014, 16, 12317–12326.

(24) Van Der Meer, B. W.; Coker, G.; Chen, S.-Y. S. Resonance energy transfer: theory and data; VCH publishers, 1994.

(25) Dale, R. E.; Eisinger, J.; Blumberg, W. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. Biophysical journal 1979, 26, 161–193.

(26) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchinson, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminf. 2012, 4, 1–17.

(27) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol. 2017, 13, e1005659.

(28) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 1977, 23, 327–341.
(29) Andersen, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **1983**, *52*, 24–34.

(30) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(31) Izaguirre, J. A.; Sweet, C. R.; Pande, V. S. Multiscale dynamics of macromolecules using Normal Mode Langevin. *Pacific Symposium on Biocomputing* **2010**, *15*, 240–251.

(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101.

(33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(34) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MD-Traj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528 – 1532.

(35) Bottaro, S.; Bussi, G.; Pinamonti, G.; Reißer, S.; Boomsma, W.; Lindorff-Larsen, K. Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* **2019**, *25*, 219–231.

(36) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2729–2742.

(37) Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Sci. Adv.* **2018**, *4*, eaar8521.
(38) Chen, Y.-L.; Lee, T.; Elber, R.; Pollack, L. Conformations of an RNA helix-junction-helix construct revealed by SAXS refinement of MD simulations. *Biophysical journal* 2019, *116*, 19–30.