# Fine-tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides

Vojtěch Mlýnský, Petra Kührová, Tomáš Kühr, Michal Otyepka, Giovanni Bussi, Pavel Banáš, and Jiří Šponer

## **Table of Contents**

Additional description of the molecular dynamics protocol2
Reference temperature does not affect the agreement between computed and experimental results
Weakening of BPh interactions in the framework of gHBfix2
Weakening of BPh interactions in highly-structured motifs
Results from base stacking analysis
Application of gHBfix and tHBfix potentials with different solvent models
Note about the usage of reweighting procedures for the newly-designed gHBfix potentials5
Comment about usage of tHBfix in a form of NBfix correction
Supporting Tables
Supporting Figures
References

Additional description of the molecular dynamics protocol. The RNA molecule remained constrained during minimization and optimization of waters and ions. Subsequently, all RNA atoms were frozen and the solvent molecules with counter-ions were allowed to move during a 500-ps long molecular dynamics (MD) run under NpT conditions (p = 1 atm., T = 298.16 K) in order to relax the total density. After this, the RNA molecule was relaxed by several minimization runs, with decreasing force constant applied to the sugar-phosphate backbone atoms. Subsequently, the system was heated in two steps: the first step involved heating under NVT conditions for 100 ps, whereas the second step involved density equilibration under NpT conditions for and additional 100 ps. The particle mesh Ewald (PME) method for treating electrostatic interactions was used. The standard unbiased MD simulations were performed under periodic boundary conditions in the NpT ensemble at 298.16 K using weak-coupling Berendsen thermostat<sup>1</sup> with coupling time of a 1 ps. The SHAKE algorithm, with a tolerance of 10<sup>-5</sup> Å, was used to fix the positions of all hydrogen atoms, and a 10.0 Å cut-off was applied.

**Reference temperature does not affect the agreement between computed and experimental results.** The NMR datasets of TN motifs were measured at 275 K,<sup>2-4</sup> whereas we calculate the NMR observables from MD simulations using 298 K reference replica. Thus, we explicitly checked if and how the difference in temperature is affecting the comparison. We performed additional simulation using the newly designed gHBfix19 + tHBfix20 potential with the unbiased (reference) replica shifted to 275 K (see Methods section in the main text for the details). The outcome is comparable within limits of sampling to the 298 K simulation (Table 2 in the main text), which is in agreement with results from other studies.<sup>5-7</sup> Thus, we expect that TN conformational ensembles are equivalent at both temperatures and the results from experiments and computations can be compared despite the temperature difference.

Weakening of BPh interactions in the framework of gHBfix. We used the r(CAAU) TN as the initial testing system, as this TN has the intercalated structure more populated than the native A-form even with the gHBfix19; see Table 6 in Ref.<sup>8</sup>. We have prepared three new gHBfixes, i.e., gHBfix19v2 BPh, gHBfix19v2 BPhv2, and gHBfix19v3 BPh, where the gHBfix19 is combined with either 0.5 (BPh) or 1.0 (BPhv2) kcal/mol penalty of all possible BPh interactions (see Table S1 for full details; the description of SPh interactions in gHBfix $19_{y2}$  is marginally different from the gHBfix19 which should, however, have no practical effect on the simulations). In addition, in the gHB fix  $19_{v3 BPh}$  version, we further weakened (from 0.5 to 1.0 kcal/mol) the SPh interactions. We observed that weakening of BPh interactions is progressively improving the simulation outcome as the higher penalty to -NH...nbOcontacts decreased the population of intercalated structures more efficiently and, consequently, provided lower  $\chi^2$  values (Table 2 in the main text). Interestingly, weakening of BPh interactions allowed us to increase penalty to the SPh interactions. With the gHBfix19 setting the SPh interactions still appear to be overstabilized but their additional weakening (without penalizing BPh interactions) would eliminate occurrence of the A-RNA minor conformation from the structural ensemble.<sup>8</sup> A-RNA minor conformation is an auxiliary A-RNA conformation of the TN accompanying the major A-RNA form and, according to the experiments, it should be populated to a certain extent.<sup>4</sup> The best result among those three simulations was obtained with the gHBfix19 $_{v3_{BPh}}$  version, where all possible BPh interactions were penalized by 0.5 kcal/mol ( $\chi^2$  value of 2.50, Table 2 in the main text).

Weakening of BPh interactions in highly-structured motifs. In order to explicitly test the effect of weakening of BPh interactions for structural description of important RNA motifs, we performed standard MD simulations of Sarcin-Ricin loop (SRL) motif and T-loop motif. The SRL (PDB ID 3DW4<sup>9</sup>) is a highly conserved RNA motif which was originally found in helix 95 of domain VI of the large ribosomal subunit.<sup>10</sup> The T-loop RNA motif is a frequently five-nucleotide hairpin motif assuming U-turn-like structures.<sup>11</sup> The occurring ucuUGGAACaga 12-mer sequence was used as the T-loop model system for testing (extracted from the PDB ID 1JJ2,<sup>12</sup> residues 310 – 321). T-loops and especially the SRL motif are established testing systems for RNA ff validations because they contain an intricate network of non-canonical base pairs and BPh interactions, which is complemented by very complex backbone conformation. We performed standard MD simulations (few µs-long timescales) using the  $\chi_{OL3CP}$  RNA ff with the gHBfix19<sub>v3</sub>\_BPh version and observed that the partial weakening of BPh contacts is not *a priori* detrimental, because the overall fold and all key interactions within those motifs were maintained (Figures S2 and S3).

**Results from base stacking analysis.** We performed stacking overlap analysis with an inhouse program and probed the possible effect of stacking on the syn/anti balance of r(AAAA) TN by calculating and quantifying the geometrical overlap of the stacked nucleobases in their various syn/anti orientations, as sampled in the simulations (see Methods in the main text for details). Initially, we extracted A-form-like states from the conformational ensemble, i.e., states, where consecutive nucleotides are stacked on top of each other in the correct order (A<sub>S1</sub>-A<sub>S2</sub>-A<sub>S3</sub>-A<sub>S4</sub>). Among 16 possible combinations of *anti/syn* states for 4 nucleotides, all anti structures were indeed the most populated (~49%), followed by syn-anti-anti-anti (~14%), anti-anti-syn (~12%), syn-anti-syn-anti (~9%) and anti-anti-syn-anti (~7%) conformational states. Remaining combinations of glycosidic orientations were marginally populated (below 3%). Subsequently, we analyzed the geometrical overlap of consecutively stacked nucleobases, i.e., we separately analyzed A<sub>S1</sub>-A<sub>S2</sub> (5'-end), A<sub>S2</sub>-A<sub>S3</sub> (middle), and A<sub>S3</sub>-A<sub>s4</sub> (3'-end) stacking pairs, in their various *syn/anti* orientations. The largest stacking overlap was achieved with anti/anti combination at the 5'-end, but syn/syn and syn/anti orientations had the largest stacking overlap in the middle step and at the 3'end, respectively (Table S3). Based on this simple analysis, we suggest a possibility that overestimated stacking interaction might be pushing the A<sub>S3</sub> nucleotide to sample syn state more frequently during MD simulations. Nevertheless, the potentially excessive population of syn states in simulations of r(AAAA) may be a complex problem caused by a combination of different factors which requires further investigations.

Application of gHBfix and tHBfix potentials with different solvent models. We firstly tested the TIP3P water model<sup>13-14</sup> during REST2 simulations of r(GACC) with the gHBfix19 potential. The population of intercalated structures is significantly increased (~12%, Table 2 in the main text) and the  $\chi^2$  value indicating the disagreement with the experimental data rises up to 0.62. This result is similar to simulations with the OPC water without any gHBfix ( $\chi^2$  value of 0.75, Table 3 in Ref. <sup>8</sup>). In addition, we observed enhanced probability to form loop-like structures, where G<sub>S1</sub> established the Watson-Crick base pairs with C<sub>S3</sub> and C<sub>S4</sub> forming 1-3 loops (~10% population) and 1-4 loops (~3% population), respectively (Table 2 and Figure S6).

Next, we performed gHBfix19 simulation of the r(GACC) TN with the SPC/E water model,<sup>1,</sup> <sup>15</sup> which is often used in simulations of protein-RNA complexes.<sup>16</sup> The simulation revealed higher total  $\chi^2$  value of 0.92, which was mostly caused by the difference between predicted

and observed NOE distances and by the fact that the MD simulation predicts signals that are not observed by the experiment (unobserved NOE signals, Table 2 in the main text). The increased  $\chi^2$  value of uNOEs (from 0.34 to 0.91 in OPC and SPC/E simulation, respectively) could be related to increase of populations of loop-like states from ~7% to ~32% in OPC and SPC/E simulation, respectively. Interestingly, the population of r(GACC) intercalated structures is low ( $\sim 2\%$ ), comparable to the OPC data (Table 2). Thus, we designed a new gHBfix19 NH-N<sub>+075</sub>, where we slightly weakened base-base interactions by decreasing the bias (from 1.0 kcal/mol to 0.75 kcal/mol) to all possible -NH...N- contacts (Table S1). Population of loop-like states dropped to  $\sim 7\%$  (from  $\sim 32\%$  in the simulation with the basic gHBfix19 potential, Table 2 in the main text). This appeared to us as a positive result. However, when we then tested effect of SPC/E water model in two simulations of r(CAAU) TN using combined gHBfix19 + tHBfix(NH...nbO) terms, the results were not optimistic. Population of intercalated states significantly increased (reaching more than 60%), which correlates with high total  $\chi^2$  values of 6.84 and 7.75 (Table 2 in the main text). Thus, SPC/E water model revealed comparable results with the OPC model for r(GACC) TN upon minor modification of gHBfix19 but results for r(CAAU) suggest that the SPC/E performance could be system-dependent. Thus we have terminated the SPC/E testing and suggest that finding suitable gHBfix + tHBfix parameters with SPC/E water model would require further work. Finally, we also tested the TIP4P-D water model, which was initially derived for simulations of disordered proteins<sup>17</sup> and was also recently suggested for RNA simulations.<sup>18</sup> Initially, we tested the TIP4P-D model during REST2 simulation of the r(GACC) TN using two different ion parameters, i.e., JC19 and charmm22.20 Results from both REST2 simulations of the r(GACC) TN revealed good agreement with the experiment (Table 2 in the main text) and are comparable to the simulation with the OPC water model. Subsequently, we tested the abovementioned combination of the basic gHBfix19, TIP4P-D water and JC ions on the other TN's from the set. We observed that TIP4P-D water model is providing comparable results with the OPC model considering the r(CAAU), r(CCCC), and r(UUUU) TNs (the  $\chi^2$  value is slightly lower for r(CAAU) TN and slightly higher for both r(CCCC) and r(UUUU) TNs, Table 2 in the main text). However, r(AAAA) REST2 simulations showed that the agreement with the experiment is much worse in comparison with the OPC water model and the amount of spurious intercalated states is significantly increased (from  $\sim 8\%$  to  $\sim 27\%$ , Table 2 in the main text). Thus, it appears that the performance of the TIP4P-D water model is typically on-par with the OPC water, but for some systems its behavior may deteriorate (as observed also for the SPC/E water model). The reason causing the difference between OPC and TIP4P-D water models for the r(AAAA) is not known and this point will require further research. Some role of the ions cannot be excluded<sup>8</sup> but full testing is beyond the scope of this work due to the high computer demands. We note that for the TIP4P-D water model we did not test tHBfix parameters though we assume that the effect of the tHBfix could be similar to the OPC water model.

In summary, our data support the view that the OPC model may be optimal for TN simulations, though transferability of its performance to other nucleic acids systems is not *a priori* guaranteed.<sup>21</sup> The results confirm sensitivity of nucleic acids simulations to water models.<sup>21-25</sup> This complicates tuning of nucleic acids *ff*s, especially when refining the non-bonded terms. Massive testing would be required to fully clarify the interrelation between the water models and the HBfix-type corrections, which is however out of scope of this work.

Note about the usage of reweighting procedures for the newly-designed gHBfix potentials. When working on the gHBfix19 and tHBfix20 parameters, we usually tested the parameters on a grid containing energy values of -1.0, -0.5, +0.0, +0.5 or +1.0 kcal/mol and only optimized a limited number of parameters at the same time.<sup>8</sup> However, the simulations do indicate that in some cases finer optimization could have visible influence on the results. We made an initial attempt to refine simultaneously all the parameters of the gHBfix potential by using the reweighting scheme proposed in Ref<sup>26</sup>. We used set of fourteen available REST2 simulations of r(GACC) TN from our previous work, where different bias to base-base and SPh interactions was applied.<sup>8</sup> However, the outcome from reweighting showed us that the protocol using just one system (albeit with broad set of simulations with different parameters) with too many parameters is prone to suggest the optimized values of gHBfix parameters too far from the initial ff, which might be explained by overfitting to one particular system. For example, the reweighting scheme using only simulations of the r(GACC) TN motif suggested to significantly further penalize most BPh and sugar-base interactions. Thus, in order to use the reweighting scheme to simultaneously tune a larger number of parameters, one needs to also add data from converged simulations of other motifs, where these interactions are present in different contexts. This will be addressed in future studies.

**Comment about usage of tHBfix in a form of NBfix correction**. We note that to emulate the effect of the tHBfix20 potential by the NBfix terms, one needs to introduce new atom types that allow to differentiate between interactions formed by groups from terminal and other residues (sharing the same atom types) in order to apply different parameters. We subsequently tested the NBfix correction, which modified base-base, SPh and BPh interactions (Table S4), in a REST2 simulation of GACC TN. We observed that the NBfix simulation and the original simulation modified by the combined gHBfix19 + tHBfix20 potential are comparable. Both revealed similar  $\chi^2$  value of 0.23 and populations of major conformers within both ensembles were only marginally different (Table 2 in the main text). Due to the computer demands we did not make testing for the remaining TNs.

#### **Supporting Tables**

**Table S1:** Overview of tested variants and combinations of the gHBfix and tHBfix potentials.<sup>8</sup> Each tested setting of the HBfix potentials is marked as HBfix  $\binom{\text{interaction pair}}{\eta}$ , where the upper label in the parentheses defines the specific interactions between RNA groups and the lower parameter  $\eta$  defines the energy modification for each H-bond interaction of this kind (in kcal/mol).<sup>a</sup>

Labol	auDfy/tuDfy full notation b	Groups involved		
Laber	ghbiix/thbiix iuii ilotatioli	Donors	Acceptors	
gHBfix19	$gHBfix \binom{2 - OH \dots nbO/bO}{-0.5} \binom{NH \dots N}{+1.0}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	bO/nbO (phosphate), N (base)	
gHBfix19 <sub>v2</sub> °	$gHBfix \binom{2 - OH \dots nbO}{-0.5} \binom{NH \dots N}{+1.0}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	nbO (phosphate), N (base)	
gHBfix19 <sub>v3</sub>	$gHBfix \begin{pmatrix} 2 - OH \dots nbO \\ -1.0 \end{pmatrix} \begin{pmatrix} NH \dots N \\ +1.0 \end{pmatrix}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	nbO (phosphate), N (base)	
gHBfix19_NH-N+0.75	$gHBfix \binom{2 - OH \dots nbO/bO}{-0.5} \binom{NH \dots N}{+0.75}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	bO/nbO (phosphate), N (base)	
gHBfix19v2_BPh	$gHBfix \binom{2-OH \dots nbO}{-0.5} \binom{NH \dots N}{+1.0} \binom{NH \dots nbO}{-0.5}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	nbO (phosphate), N (base)	
$gHBfix19_{v2}\_BPh_{v2}$	$gHBfix \binom{2-OH \dots nbO}{-0.5} \binom{NH \dots N}{+1.0} \binom{NH \dots nbO}{-1.0}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	nbO (phosphate), N (base)	
gHBfix19 <sub>v3</sub> _BPh	$gHBfix \binom{2-OHnbO}{-1.0} \binom{NHN}{+1.0} \binom{NHnbO}{-0.5}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	nbO (phosphate), N (base)	
tHBfix(NHnbO) <sup>d</sup>	$\text{tHBfix}\begin{pmatrix} \text{NH} \dots \text{nbO} \\ -1 0 \end{pmatrix}$	NH (base)	nbO (phosphate)	
tHBfix(OHnbO) <sup>d</sup>	$tHBfix \begin{pmatrix} 2 - OH \dots nbO \\ -1.0 \end{pmatrix}$	2'-OH/3'-OH/ 5'-OH (sugar)	nbO (phosphate)	
tHBfix(OHO) <sup>d</sup>	$\text{tHBfix}\begin{pmatrix} 2 - \text{OH} \dots \text{O} \\ -1 & 0 \end{pmatrix}$	5'-OH (sugar)	O (base)	
tHBfix(OHOH) <sup>d</sup>	$tHBfix \begin{pmatrix} 2 - 0H \dots 02 \\ -1.0 \end{pmatrix}$	2'-OH/3'-OH (sugar)	5'-OH (sugar)	
tHBfix20 °	tHBfix $\binom{2 - 0H \dots nb0/0/02}{-1.0} \binom{NH \dots nb0}{-1.0}$	2'-OH/3'-OH/ 5'-OH (sugar), NH (base)	nbO (phosphate), N/O (base), 5'-OH (sugar)	

<sup>a</sup> see the original paper<sup>8</sup> for the detailed definition of the gHBfix potential and for the complete list of atoms from each RNA nucleotide whose interactions are modified by the gHBfix. All listed potentials are applied between hydrogen of the proton donor and proton acceptor heavy atom. In the present study we applied the HBfix terms in all cases using the interval between 2 Å and 3 Å, see Figure 1 in the main text.

<sup>b</sup> as used in the original paper introducing the gHBfix potential.<sup>8</sup>

<sup>c</sup> This variant is marginally different from the basic gHBfix19 version by not modulating interactions between the 2'-OH groups and bridging phosphate oxygens (bO). We assume the difference between gHBfix19 and gHBfix19<sub>v2</sub> versions has minimal impact on the simulations, due to the marginal role of the interactions involving the bridging oxygens.

<sup>d</sup> see Table 1 in the main text for the detailed list of atoms from RNA nucleotides whose interactions were modified.

<sup>e</sup> containing the combination of all above-introduced tHBfix terms. See the Table S2 with enumeration of all the interactions.

TN	H-bonds		
	C <sub>S4</sub> (2'-OH/3'-OH)C <sub>S3</sub> ( <i>pro</i> -R <sub>P</sub> / <i>pro</i> -S <sub>P</sub> )		
r(GACC)	$G_{S1}(5'-OH)C_{S3}(pro-R_P/pro-S_P)$		
	$G_{S1}(5'-OH)C_{S4}(O2)$		
	Cs4(2'-OH/3'-OH)Gs1(O5')		
	$C_1(N4H)U_{S4}(pro-R_P/pro-S_P)$		
	$U_{S4}(2'-OH/3'-OH)A_{S3}(pro-R_P/pro-S_P)$		
r(CAAU)	$C_{S1}(5'-OH)A_{S3}(pro-R_P/pro-S_P)$		
	$C_{S1}(5'-OH)U_{S4}(O2)$		
	$U_{s4}(2'-OH/3'-OH)C_{s1}(O5')$		
$r(\Lambda \Lambda \Lambda \Lambda)$	$A_1(N6H)A_{S4}(pro-R_P/pro-S_P)$		
	$A_{S4}(2'-OH/3'-OH)A_{S3}(pro-R_P/pro-S_P)$		
1(11111)	$A_{S1}(5'-OH)A_{S3}(pro-R_P/pro-S_P)$		
	As4(2'-OH/3'-OH)As1(O5')		
	$C_1(N4H)C_{S4}(pro-R_P/pro-S_P)$		
	$C_{S4}(2'-OH/3'-OH)C_{S3}(pro-R_P/pro-S_P)$		
r(CCCC)	$C_{S1}(5'-OH)C_{S3}(pro-R_P/pro-S_P)$		
	$C_{S1}(5'-OH)C_{S4}(O2)$		
	Cs4(2'-OH/3'-OH)Cs1(O5')		
	U <sub>S4</sub> (2'-OH/3'-OH)U <sub>S3</sub> ( <i>pro</i> -R <sub>P</sub> / <i>pro</i> -S <sub>P</sub> )		
r(UUUU)	$U_{S1}(5'-OH)U_{S3}(pro-R_P/pro-S_P)$		
	Us1(5'-OH)Us4(O2)		
	Us4(2'-OH/3'-OH)Us1(O5')		

**Table S2:** List of interactions that were modified by the tHBfix20 potential for the particular TN motif.

**Table S3:** The stacking overlap (in relative values) between consecutive nucleobases from the simulation of the r(AAAA) TN and its dependency on the *syn/anti* combination of nucleobases (see Methods in the main text for details).

χ-dihedral	5'-end	middle	3'-end
major states	$(A_{s1}-A_{s2})$	$(A_{s2}-A_{s3})$	$(A_{S3}-A_{S4})$
anti/anti	0.31	0.30	0.30
anti/syn	0.21	0.25	0.25
syn/anti	0.20	0.30	0.41
syn/syn	0.14	0.38	0.29

**Table S4:** The modified Lennard-Jones combining rules (NBfix) obtained as an attempt to mimic modulation of the interaction-energy curves introduced by the tHBfix20 potential. The NBfix parameters mimicking modulation of the gHBfix19 potential (included thus also in the combined gHBfix19 + tHBfix20 version) and other details can be found in the Supporting Information of Ref. <sup>8</sup>.

			R (Å) <sup>d</sup>		$\varepsilon$ (kcal/mol) <sup>e</sup>	
tHBfix20 <sup>a</sup>	η <sup>b</sup> (kcal/mol)	Atom pairs modified °		NBfix		NBfix
			$\chi$ ol3cp <sup>f</sup>	mimicking 2	$\chi$ ol3cp <sup>f</sup>	mimicking
				tHBfix20 <sup>g</sup>		tHBfix20 <sup>g</sup>
NHnbO	-1.0	HN(H)OX(OP)	2.3493	2.4993	0.0574	0.1449
2-OHbO	-1.0	HT(HO)OT(OR)	1.7718 <sup>h</sup>	2.3918	0	0.17
2-OHnbO	-1.0	HT(HO)OX(OP)	1.7493 <sup>h</sup>	2.1693	0	0.1889
2-OHO	-1.0	HT(HO)OU(O)	1.6612 <sup> h</sup>	2.1912	0	0.1889

<sup>a</sup> interactions modified by the tHBfix20 potential

<sup>b</sup> total potential energy modulation for each H-bond interaction by the tHBfix potential function (see Methods in the main text for details and Figure 1 in the main text for definition of the  $\eta$  parameter; its negative value means that the interaction is penalized).

<sup>c</sup> new atom types were designed in order to differentiate among interactions formed by groups from terminal and internal nucleotides (original atom types from the  $\chi_{OL3CP}$  *ff* are in parenthesis)

<sup>d</sup> distance, where the Lennard-Jones potential for the interaction of atoms *i* and *j*,  $R_{i,j}$ , is exactly zero. Parameters the particular Lennard-Jones pairs are derived as:  $R_{i,j} = R_i + R_j$ 

<sup>e</sup> depth of the potential well for the interaction of atoms *i* and *j*,  $\varepsilon_{ij}$ . Parameters for the particular Lennard-Jones

pairs are derived as:  $\varepsilon_{i,j} = \sqrt{(\epsilon_i + \varepsilon_j)}$ 

<sup>f</sup> ff99bsc0 $\chi_{OL3}$ <sup>27-30</sup> RNA ff version with the vdW modification of phosphate oxygens developed by Steinbrecher et al.<sup>31</sup>

<sup>g</sup> NBfix *ff* reparameterization was prepared in a way to be comparable with the  $tHBfix \begin{pmatrix} 2-OH...nbO/O/O2 \\ -1.0 \end{pmatrix} (NH...nbO) potential (i.e. tHBfix20, see Table S1).$ <sup>h</sup> in the original *ff* these values are formally equal to radii of the bO, nbO and carbonyl group oxygens,

<sup>h</sup> in the original *ff* these values are formally equal to radii of the bO, nbO and carbonyl group oxygens, respectively, as the radius of polar hydrogen is equal to zero. However, they are essentially irrelevant due to zero  $\varepsilon$  value.

### **Supporting Figures**



**Figure S1:** A representative example of conformational sampling and convergence of one REST2 simulation, namely the REST2 simulation of r(CCCC) TN with combined gHBfix19 + tHBfix(NH...nbO) potential. Time evolution of major conformers, i.e., A-RNA major (green), A-RNA minor (blue), and the spurious intercalated structure (red) in all eight replicas analyzed from continuous (demultiplexed) trajectories.



**Figure S2:** Structural analysis of three standard MD simulations of Sarcin-Ricin RNA loop (SRL) motif. See Figure 1 in Ref. <sup>32</sup> for the structure. Plots in the first, second, and third column display fluctuations of distances indicating signature BPh interactions, i.e., G7(N2)...A17(*pro*-R<sub>P</sub>), G16(N2)...U8(*pro*-R<sub>P</sub>), and A6(N6)...C18(*pro*-R<sub>P</sub>) distances,

respectively. We compared the behavior in three different RNA *ffs* during 1  $\mu$ s long MD simulations, i.e., the standard  $\chi_{OL3CP}^{27-30}$  (OL3CP, plots at the top),  $\chi_{OL3CP}$  with the gHBfix19 potential<sup>8</sup> (gHBfix, plots in the middle), and  $\chi_{OL3CP}$  with the gHBfix19 combined with -0.5 kcal/mol penalty of all possible BPh interactions and with increased weakening of all the SPh interactions to -1.0 kcal/mol (gHBfix19<sub>v3</sub>\_BPh, plots at the bottom).



**Figure S3:** Structural analysis of three unbiased MD simulations of T-loop RNA motif. See Figure 1 in Ref. <sup>33</sup> for the structure. Plots in the first, second, and third column display fluctuations of distances indicating signature H-bonds, i.e., G5(2'-OH)...A7(N7),  $G5(N1)...A8(pro-R_P)$ , and G5(N2)...A8(O5') distances, respectively. We compared the

behavior in three different RNA *ffs* during 2 µs long MD simulations, i.e., the standard  $\chi_{OL3CP}^{27-30}$  (OL3CP, plots at the top),  $\chi_{OL3CP}$  with the external gHBfix19 potential<sup>8</sup> (gHBfix, plots in the middle), and  $\chi_{OL3CP}$  with the gHBfix19 combined with –0.5 kcal/mol penalty of all possible BPh interactions and with increased weakening of all the SPh interactions to –1.0 kcal/mol (gHBfix19<sub>v3</sub>\_BPh, plots at the bottom).



**Figure S4:** Tertiary structures of the most populated clusters from r(UUUU) REST2 simulation with the  $\chi_{OL3CP}$  RNA *ff* modified by combination of gHBfix19<sup>8</sup> + tHBfix20 (present work, Table S1) potentials. H-atoms are not shown for clarity. Population of unassigned structures is ~32%.



**Figure S5:** Tertiary structures of the most populated clusters from r(AAAA) REST2 simulation with the  $\chi_{OL3CP}$  RNA *ff* modified by combination of gHBfix19<sup>8</sup> + tHBfix20 (Table S1) potentials. H-atoms are not shown for clarity.



**Figure S6:** Tertiary structures of the loop-like clusters of the r(GACC) TN, which are commonly occurring during REST2 simulations (especially with three-point charge water models). A, C, and G nucleotides are colored in sand, white, and red, respectively. H-atoms are not shown for clarity. The loop-like clusters indicate overestimation of base pairing.

## References

(1) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, *81*, 3684-3690.

(2) 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.

(3) Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H. The Nuclear Magnetic Resonance of CCCC RNA Reveals a Right-Handed Helix, and Revised Parameters for AMBER Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry-Us* **2013**, *52*, 996-1010.

(4) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. Benchmarking AMBER Force Fields for RNA: Comparisons to NMR Spectra for Single-Stranded r(GACC) Are Improved by Revised chi Torsions. *J Phys Chem B* **2011**, *115*, 9261-9270.

(5) Aytenfisu, A. H.; Spasic, A.; Grossfield, A.; Stern, H. A.; Mathews, D. H. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J Chem Theory Comput* **2017**, *13*, 900-915.

(6) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J Chem Theory Comput* **2014**, *10*, 492-499.

(7) Zhao, J.; Kennedy, S. C.; Berge, K. D.; Turner, D. H. Nuclear Magnetic Resonance of Single Stranded RNAs and DNAs of CAAU and UCAAUC as Benchmarks for Molecular Dynamics Simulations. *J Chem Theory Comput* **2020**, *16*, 1968-1984.

(8) Kuhrova, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Sponer, J.; Banas, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J Chem Theory Comput* **2019**, *15*, 3288-3305.

(9) Olieric, V.; Rieder, U.; Lang, K.; Serganov, A.; Schulze-Briese, C.; Micura, R.; Dumas, P.; Ennifar, E. A fast selenium derivatization strategy for crystallization and phasing of RNA structures. *RNA* **2009**, *15*, 707-715.

(10) Gutell, R. R.; Schnare, M. N.; Gray, M. W. A Compilation of Large Subunit (23s-Like and 23s-Like) Ribosomal-Rna Structures. *Nucleic Acids Res.* **1992**, *20*, 2095-2109.

(11) Nagaswamy, U.; Fox, G. E. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **2002**, *8*, 1112-1119.

(12) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *Embo J.* **2001**, *20*, 4214-4221.

(13) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79*, 926-935.

(14) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J Chem Phys* **1996**, *105*, 1902-1921.

(15) Berweger, C. D.; van Gunsteren, W. F.; Müller-Plathe, F. Force field parametrization by weak coupling. Re-engineering SPC water. *Chem Phys Lett* **1995**, *232*, 429-436.

(16) Sponer, J.; Krepl, M.; Banas, P.; Kuhrova, P.; Zgarbova, M.; Jurecka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes? *Wiley Interdiscip Rev RNA* **2017**, *8*.

(17) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* **2015**, *119*, 5113-5123.

(18) 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. USA* **2018**, *115*, E1346-E1355.

(19) Joung, I. S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* **2008**, *112*, 9020-9041.

(20) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998**, *102*, 3586-3616.

(21) Havrila, M.; Stadlbauer, P.; Islam, B.; Otyepka, M.; Sponer, J. Effect of Monovalent Ion Parameters on Molecular Dynamics Simulations of G-Quadruplexes. *J Chem Theory Comput* **2017**, *13*, 3911-3926.

(22) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578-1590.

(23) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banas, P.; Koca, J.; Cheatham, T. E., 3rd; Jurecka, P.; Sponer, J. Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J Chem Theory Comput* **2012**, *8*, 2506-2520.

(24) Sklenovsky, P.; Florova, P.; Banas, P.; Reblova, K.; Lankas, F.; Otyepka, M.; Sponer, J. Understanding RNA flexibility using explicit solvent simulations: the ribosomal and group I intron reverse kink-turn motifs. *J Chem Theory Comput* **2011**, *7*, 2963-2980.

(25) Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P., et al. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* **2018**, *118*, 4177-4338.

(26) Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banas, P.; Sponer, J.; Bussi, G. Fitting Corrections to an RNA Force Field Using Experimental Data. *J Chem Theory Comput* **2019**, *15*, 3425-3431.

(27) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Amer Chem Soc* **1996**, *118*, 2309-2309.

(28) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J* **2007**, *92*, 3817-3829.

(29) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Chem Theory Comput* **2000**, *21*, 1049-1074.

(30) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput* **2011**, *7*, 2886-2902.

(31) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput* **2012**, *8*, 4405-4412.

(32) Havrila, M.; Reblova, K.; Zirbel, C. L.; Leontis, N. B.; Sponer, J. Isosteric and Nonisosteric Base Pairs in RNA Motifs: Molecular Dynamics and Bioinformatics Study of the Sarcin Ricin Internal Loop. *J Phys Chem B* **2013**, *117*, 14302-14319.

(33) Chan, C. W.; Chetnani, B.; Mondragon, A. Structure and function of the T-loop structural motif in noncoding RNAs. *Wiley Interdiscip Rev RNA* **2013**, *4*, 507-522.