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

Vojtěch Mlýnský,^{1} Petra Kührová,² Tomáš Kühř,³ Michal Otyepka,^{1,2} Giovanni Bussi,⁴ Pavel
Banáš,^{1,2} and Jiří Šponer^{1*}*

¹ Institute of Biophysics of the Czech Academy of Sciences, Kralovopolská 135, 612 65 Brno,
Czech Republic

² Regional Centre of Advanced Technologies and Materials, Department of Physical
Chemistry, Faculty of Science, Palacký University, tř. 17 listopadu 12, 771 46, Olomouc, Czech
Republic

³ Department of Computer Science, Faculty of Science, Palacký University, tř. 17 listopadu 12,
771 46, Olomouc, Czech Republic

⁴ Scuola Internazionale Superiore di Studi Avanzati, SISSA, via Bonomea 265,
34136 Trieste, Italy

* corresponding authors, email: mlynsky@ibp.cz and sponer@ncbr.muni.cz

ABSTRACT

Determination of RNA structural-dynamic properties is challenging for experimental methods. Thus atomistic molecular dynamics (MD) simulations represent a helpful technique complementary to experiments. However, contemporary MD methods still suffer from limitations of force fields (*ff*s), including imbalances in the non-bonded *ff* terms. We have recently demonstrated that some improvement of state-of-the-art AMBER RNA *ff* can be achieved by adding a new term for H-bonding called gHBfix, which increases tuning flexibility and reduces the risk of side-effects. Still, the first gHBfix version did not fully correct simulations of short RNA tetranucleotides (TNs). TNs are key benchmark systems due to availability of unique NMR data, although giving too much weight on improving TN simulations can easily lead to over-fitting to A-form RNA. Here we combine the gHBfix version with another term called tHBfix, which separately treats H-bond interactions formed by terminal nucleotides. This allows to refine simulations of RNA TNs without affecting simulations of other RNAs. The approach is in line with adopted strategy of current RNA *ff*s, where the terminal nucleotides possess different parameters for the terminal atoms than the internal nucleotides. The combination of gHBfix with tHBfix significantly improves the behavior of RNA TNs during well-converged enhanced-sampling simulations. TNs mostly populate canonical A-form like states while spurious intercalated structures are largely suppressed. Still, simulations of r(AAAA) and r(UUUU) TNs show some residual discrepancies with the primary NMR data which suggests that future tuning of some other *ff* terms might be useful.

KEYWORDS

RNA, force field, tetranucleotide, enhanced sampling, NMR, gHBfix

INTRODUCTION

Atomistic description of structural dynamics of RNA molecules is essential for understanding of some of key biomolecular processes including, e.g., proteosynthesis, splicing, ribosomal catalysis etc. Experimental techniques, such as X-ray crystallography, cryo-electron microscopy and nuclear magnetic resonance (NMR) spectroscopy provide essential structural information. However, the dynamics is obscured by ensemble averaging. Insights into biomolecular motions and structural dynamics can also be obtained by molecular dynamics (MD) simulations. However, currently used (state-of-the-art) empirical potentials (force fields, *ffs*) suffer from many inaccuracies which may even lead to a preferential sampling of spurious structures instead of the native states. Performance of the available pair-additive atomistic RNA *ffs* has been reviewed in detail,¹⁻⁴ with a suggestion that the currently used form of the pair-additive RNA *ffs* is approaching the limits of its applicability.⁴ In other words, it is becoming increasingly difficult to generally improve performance of the RNA simulations by refining parameters of the available *ff* terms without introducing some new unintended imbalances. Radical approach to overcome the limitations is the ongoing development of more sophisticated polarizable *ffs*.⁵⁻⁹ Alternatively, we have suggested that improvements could be achieved also by extending the pair-additive *ffs* by new simple terms that are uncoupled from the existing *ff* terms and thus allow tuning of simulations of problematic systems while minimizing undesirable side effects. We have introduced an additional *ff* term called HBfix (H-bond fix),¹⁰ that can be used to tune the description of key hydrogen-bonding interactions. HBfix is a simple short-range potential (typically acting in the range from 2 to 3 Å for H – acceptor distances) that either increases or decreases stability of target H-bonds (Figure 1). It has been first applied as a structure-specific potential which can improve folding of some RNA tetraloops¹⁰ and structural stability of protein-

RNA complexes.¹¹ Subsequently, we have suggested its generalized interaction-specific variant (gHBfix) which represents a true extension of the parent *ff*.¹² We have demonstrated that boosting of all --NH...N-- base–base interactions by 1.0 kcal/mol and destabilizing all sugar–phosphate (SPh) H-bonds by 0.5 kcal/mol significantly improves simulations of RNA tetranucleotides and GNRA tetraloop without deteriorating simulations of other systems (Figure 1). These parameters complement the common OL3 (*ff99bsc0* χ_{OL3})^{13–16} AMBER RNA *ff* version used together with modified phosphate van der Waals (vdW) parameters¹⁷ and OPC water model.¹⁸

Figure 1: Description of the gHBfix *ff* term used for either support (blue curve) or weakening (red curve) of H-bond interactions. The initial version of the gHBfix introduced in Ref. ¹² is tuning H-bond interactions in order to: (i) support base – base interactions (left panel, blue dashed lines for --NH...N-- interactions) and, simultaneously, (ii) destabilize SPh interactions (red dashed lines on the right panel); SPh interactions are those between 2'-OH groups and bridging (bO)/non-bridging (nbO) phosphate oxygens. The gHBfix potential is affecting all interactions of

the same type while an analogous term can be used to modulate only structure-specific (individual) H-bonds (abbreviated as HBfix^{10, 12}). In the present work, we suggest a variant separately targeting interactions involving terminal nucleotides (tHBfix term).

In this work we show that further improvement of RNA simulations can be achieved by tuning the SPh and base-phosphate (BPh) H-bonds involving the terminal nucleotides; this variant is henceforth abbreviated as tHBfix. This approach, although targeting only selected nucleotides can be considered as a generally transferable *ff* modification, as it can be applied to all simulated RNAs. Tuning of parameters of terminal nucleotides in RNA chains is fully justified as their properties can differ from internal nucleotides and their imbalanced description in simulations can have detrimental effects. In other words, modification of the interactions of terminal nucleotides can separately eliminate some simulation problems without affecting the remaining parts of the RNA molecules. This can reduce the risk of overfitting the *ff* when using simulations of small model RNA systems for its testing and training. The rationale is that terminal nucleotides have excessively large impact on simulations of small model systems and thus it is advisable to screen off their role as much as possible. Our ultimate goal is not to reproduce experimental data for these small model systems, but to study folded RNA structures and diverse protein-RNA complexes. Targeting the terminal residues of the *ff* may be considered both as an attempt to reflect potentially different electronic structures of these residues and a pragmatic means to increase tuning capability of the *ff*.

The present study is based mainly on simulations of RNA tetranucleotides (TNs). Conformation ensembles of TNs provide one of the key benchmarks for testing RNA *ff*s due to their small size and straightforward comparison of their simulations with solution experiments.¹⁹⁻

³¹ Obviously, any quantitative *ff* assessment is critically dependent on the convergence of structural populations because only well-converged simulations can provide unambiguous benchmark datasets. Nevertheless, contemporary simulation methods and hardware already allow to obtain sufficiently converged simulation ensembles for TNs.^{12, 20, 30} TN simulations can specifically monitor performance of *ffs* for several salient energy contributions, namely, (i) SPh and BPh interactions, (ii) base stacking interactions, (iii) backbone conformations and (iv) balance of these contributions with solvation. Experimental data shows that TNs mostly populate A-form conformations while MD simulations tend to significantly sample also non-native intercalated structures (or some other non-native structures, Figure 2) that are considered to be *ff* artifacts.^{12, 20, 24, 30} Obviously, when using TNs as a benchmark for *ff* refinement, one has to be concerned about a possible over-fitting of the *ff* towards the canonical A-RNA conformation (a potential problem of some recently published *ffs*³²), which may lead to side-effects for simulations of folded RNAs.^{12, 33}

Here, we applied enhanced-sampling simulations to obtain conformational ensembles of five RNA tetranucleotides, i.e., r(AAAA), r(CAAU), r(CCCC), r(GACC), and r(UUUU). The basic gHBfix potential combined with additional fixes involving one or both terminal residues (tHBfix, Table 1) significantly increases agreement between predicted and experimental data. The obtained conformational ensembles revealed that r(CAAU), r(CCCC) and r(GACC) TNs are sampling more than 75% of time A-form-like conformations. The remaining r(AAAA) and r(UUUU) TNs displayed a higher complexity with a mixture of structures, which is not yet in a full agreement with the primary NMR data. Nevertheless, all the results are significantly improved in comparison with the previous work.¹² The presented approach is generally applicable and should further improve simulations of RNA molecules.

Figure 2: Tertiary structures of the five studied systems in the spurious intercalated structure involving the 5'-terminal nucleotide, which is often seen during MD simulations using common RNA AMBER *ffs*.^{12, 19-25, 27, 28, 30-32} A, C, G and U nucleotides are colored in sand, white, red, and blue, respectively.

METHODS

Starting structures and simulation setup. The initial coordinates of $r(AAAA)$, $r(CAAU)$, $r(CCCC)$, $r(GACC)$, and $r(UUUU)$ TNs were prepared using Nucleic Acid Builder of AmberTools14³⁴ as one strand of an A-form duplex. The starting topologies and coordinates for classical MD simulations of the Sarcin-Ricin loop and the T-loop RNA motifs (see below for further details) were prepared by using the tLEaP module of AMBER 16 program package.³⁵ Structures were solvated using a rectangular box with a minimum distance between box walls and solute of 12 Å, yielding ~2000 water molecules added and ~40×40×40 Å³ box size.

We used the standard OL3¹³⁻¹⁶ RNA *ff* with the vdW modification of phosphate oxygens developed by Steinbrecher et al.,¹⁷ where the affected dihedrals were adjusted as described elsewhere,^{10, 36} (see the Supporting Information of Ref. ¹⁰ for parameters) this version is abbreviated as χ_{OL3CP} henceforth. Additionally, we applied the external gHbfix potential¹² that

was shown to improve the overall *ff* performance. Most of simulations were carried out with the OPC¹⁸ water model and in ~0.15 M KCl salt using the Joung–Cheatham (JC)³⁷ ionic parameters. Specific tests involved TIP3P,³⁸ SPC/E,³⁹ and TIP4P-D⁴⁰ water models. One test simulation of the r(GACC) TN with the TIP4P-D water was combined with charmm22 ion parameters⁴¹ (see the Supporting Information for other details about the simulation protocol).

Enhanced sampling. The replica exchange solute tempering (REST2)⁴² simulations of all TNs were performed at $T = 298$ K with 8 replicas. Details about settings can be found in our previous work.¹⁰ The scaling factor (λ) values ranged from 1 to 0.601700871 and were chosen to maintain an exchange rate above 20%. The effective solute temperature ranged from 298 K to ~500 K. The hydrogen mass repartitioning⁴³ with a 4-fs integration time step was used. One test simulation of r(CAAU) TN was performed at temperature 275 K corresponding to the experimental conditions.^{22-24, 31} The same λ values were applied resulting in effective solute temperature range from 275 K to ~460 K.

HBfix-type potentials used in this work. We applied the previously introduced generalized (interaction-specific) gHBfix potential¹² that is improving the overall performance of the χ_{OL3CP} RNA *ff*. In the majority of the simulations, we have used the basic gHBfix potential with the same parameters indicated in the earlier study¹² as optimal (Figure 1, abbreviated here as gHBfix19, see also below). Besides that, in this work we have used several other variants of the HBfix-type refinements. We first tested additional gHBfix terms that penalized all BPh and SPh interactions in the studied systems. Later, we subtracted interactions that are established by groups from terminal nucleotides (Figure 3) and designed HBfix terms targeting only interactions that include the terminal nucleotides; these are termed as terminal-HBfix, i.e., tHBfix. The list of terminal groups and atoms from RNA nucleotides whose interactions were

modified by particular tHBfix potentials are listed in Table 1 and Table S1 in the Supporting Information.

Conformational analysis. The dominant conformations sampled during REST2 simulations were identified using a cluster analysis based on an algorithm introduced by Rodriguez and Laio⁴⁴ in combination with the ϵ RMSD⁴⁵ (see our previous works for the detailed description of the algorithm).^{10, 12} The clustering algorithm is based on identification of the cluster center (centroid). However, this approach does not allow separation of A-RNA major and A-RNA minor conformations. Due to the definition of the cluster points and a cluster hull representing a noise spread around a given cluster,¹⁰ these two conformations merge into one cluster, since the transition path between them was sampled by interconversions too frequently. Thus, we performed an additional ϵ RMSD analysis, where we used representative PDB's of the most important conformations and calculated populations of those states from MD trajectories based on simple ϵ RMSD cutoff representing a conservative definition of the conformational state (Table 2).

The convergence was checked by bootstrapping⁴⁶ using the recently introduced implementation.^{10, 12} Errors were estimated by resampling of the time blocks of the whole set of replicas and subsequent resampling of the coordinate-following replicas for the purpose of their demultiplexing to follow Hamiltonians in order to obtain a final resampled population at the reference unbiased Hamiltonian state (see Figure S1 in the Supporting Information of Ref. ¹² for more details). Figure S1 in the Supporting Information shows that the major conformers are uniformly represented within all replicas from the particular REST2 simulations.

Stacking analysis. We implemented an in-house code in order to probe the effect of stacking on the *syn/anti* balance of nucleobases. All heavy atoms of each nucleobase are fitted by a

plane,⁴⁷ projected into this plane and transformed to a convex planar polygon.⁴⁸ For each pair of polygons, the program quantifies their geometrical overlap as follows: A part of the second polygon in the stacking distance (from 3 to 4) is projected into the plane of the first polygon and the area of calculated intersection between both polygons is measured. The calculated values of the geometrical overlap for consecutive nucleobases lie between zero and one (in relative values), where zero means no overlap and one is the maximum overlap.

Comparison with the experiment. The conformational ensembles obtained from MD simulations were compared with solution experiments.²²⁻²⁴ We used and analyzed separately four NMR observables, i.e., (i) backbone ³J scalar couplings, (ii) sugar ³J scalar couplings, (iii) nuclear Overhauser effect intensities (NOEs), and (iv) the absence of specific peaks in NOE spectroscopy (uNOEs). Their combination provided the total χ^2 value for each REST2 simulation. The lower χ^2 value, the better agreement between the experimental data and the conformational ensemble from the particular simulation was achieved. Note that all χ^2 values below 1 indicate good agreement with the experiment and, considering the experimental error, it is not straightforward to decide among two simulations with $\chi^2 < 1$, which of them agrees better with the NMR datasets. However, the above-mentioned rule of thumb of $\chi^2 < 1$ can be applied primarily to the χ^2 contributions coming from ³J scalar couplings and NOEs as they rigorously follow statistical χ^2 distribution. In turn, the component coming from uNOEs rather qualitatively indicates violations of the NOEs data for particular contacts, so the interpretation of this χ^2 value is not straightforward. Thus, although simulations with *total* $\chi^2 < 1$ might be considered to be in agreement with the experimental data, one should always check also the *individual* χ^2 components corresponding to the particular NMR observables.

We further identified that the signal range for some NOEs, i.e., C2(H6)...C2(H5''), U2(H3')...U2(H5'), and U3(H3')...U3(H5'), was rarely populated during MD simulations. The experimental NOEs would thus indicate some rather atypical RNA backbone conformations which might also indicate some uncertainty in the experimental data. Those signals provided higher χ^2 values (for all simulations of the particular TN sequence) and thus, lowered the agreement between simulations and the experiment for r(CCCC) and r(UUUU) TNs (Table 2).

NOEs and uNOEs were obtained from MD simulations as averages over the N samples, i.e., $(u)NOE_{CALC} = \left(\frac{\sum_i^N r_i^{-6}}{N} \right)^{-1/6}$. The calculated NOEs were directly compared against experimental values. Each uNOE signal was considered as violated when the calculated value from the simulation was shorter than the assigned one from the experiment. 3J scalar couplings were obtained from the Karplus relationships (see Supporting Information of Ref. ³⁰ for the details and experimental datasets).

Figure 3: Tertiary structure of the r(CAAU) TN in the compact intercalated state. All groups responsible for the spurious H-bonds (black lines) forming BPh, SPh and sugar-base interactions are labelled and highlighted by spheres. Nucleotides are colored as in Figure 2.

RESULTS AND DISCUSSION

In our preceding work we demonstrated that performance of the basic OL3 AMBER RNA *ff*¹³,^{15-17, 42} is improved by using the $\text{gHBfix}\left(\begin{smallmatrix} 2-\text{OH}\dots\text{nbO/bO} \\ -0.5 \end{smallmatrix}\right)\left(\begin{smallmatrix} \text{NH}\dots\text{N} \\ +1.0 \end{smallmatrix}\right)$ variant,¹² abbreviated in the present paper as gHBfix19, see Table S1. This gHBfix version weakens all SPh H-bonds by 0.5 kcal/mol and supports all base-base $-\text{NH}\dots\text{N}-$ H-bonds by 1.0 kcal/mol (Figure 1). These parameters were derived to be combined with the OPC water model and the OL3 *ff* has been used with modified phosphate vdW parameters (χ_{OL3CP} , see Methods). In the present work we used the χ_{OL3CP} basic RNA *ff* as the starting version and tried to further improve its performance. We focused on five RNA TNs, i.e., r(AAAA), r(CAAU), r(CCCC), r(GACC), and r(UUUU), where highly accurate experimental data are available.²²⁻²⁴ The application of the basic gHBfix19 potential leads to visible but only partial improvement in structural description of three TNs.¹² More specifically, population of intercalated structures still remained significant for r(CAAU), r(AAAA), and r(CCCC) TNs.¹² Thus we tried additional means to improve the TN simulations.

Weakening of BPh interactions in the framework of gHBfix improves TN simulations. In our previous work,¹² we found out that sequences with higher propensity to form the intercalated structures, i.e., r(CAAU), r(AAAA), and r(CCCC), possess C or A as the 5'-terminal nucleotide. These nucleotides can form type-7 BPh (7BPh) interaction⁴⁹ in the intercalated state (Figure 4). BPh interactions were previously reported to be generally overpopulated in AMBER simulations

of unfolded states, contributing to spurious compaction of RNA single strands and thus hindering folding of RNA tetraloops.¹⁰ We have suggested that their weakening could suppress the spurious intercalated structures of TNs.¹² However, native BPh interactions are common in many folded RNA structures⁴⁹ and thus tuning of these interactions was not attempted in Ref. ¹².

In the present work 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. ¹². We have prepared three new gHBfixes, i.e., gHBfix19_{v2}_BPh, gHBfix19_{v2}_BPh_{v2}, and gHBfix19_{v3}_BPh, where the gHBfix19 is combined with either 0.5 (BPh) or 1.0 (BPh_{v2}) kcal/mol penalty of all possible BPh interactions (see Table S1 for full details; the description of SPh interactions in gHBfix19_{v2} is marginally different from the gHBfix19 which should, however, have no practical effect on the simulations). In addition, in the gHBfix19_{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...nbO– contacts decreased the population of intercalated structures more efficiently and, consequently, provided lower χ^2 values (Table 2). 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.¹² 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.²² 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 (χ^2 value of 2.50, Table 2). The small penalty sufficient for the

BPh interactions is promising since their large weakening could destabilize structures of folded RNAs, where BPh contacts are common and highly conserved.⁴⁹

Figure 4: The spurious intercalated state is highly populated during MD simulations of TNs containing C or A as the 5'-terminal nucleotide. (A) The 7BPh interaction⁴⁹ is formed by nucleotides that possess the amino group at the Hoogsteen edge, i.e., N6H and N4H for A and C, respectively (red circle). (B) The detail of the 7BPh interaction⁴⁹ in r(AAAA), r(CAAU), and r(CCCC) TNs. The spurious H-bond between the amino group and nbOs of the nucleotide at the 3'-end is highlighted (red line in yellow background).

In order to explicitly test the effect of weakening of BPh interactions for structural description of important RNA motifs, we used two systems with multiple conserved BPh interactions, namely the Sarcin-Ricin loop RNA motif (SRL, PDB ID 3DW4⁵⁰), a highly conserved RNA motif which was originally found in helix 95 of domain VI of the large ribosomal subunit,⁵¹ and the T-loop RNA motif, a frequently occurring five-nucleotide hairpin motif assuming U-turn-like structures.⁵² The ucuUGGAacaga 12-mer sequence was used as the T-loop model system for testing (extracted from the PDB ID 1JJ2,⁵³ 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 μ s-long timescales) using the χ_{OL3CP} RNA *ff* with the gHBfix19_{v3}_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 in the Supporting Information). However, we cannot rule out possible side effects on longer time scales and/or for other systems. Much broader testing considering weakening of all BPh interactions would be

required before introducing a *ff* version with all BPh contacts been weakened using the gHBfix *ff* term.

Weakening of specific interactions formed by terminal groups is comparable with tuning of all BPh contacts. Although BPh interactions appear to be over-stabilized within the current basic RNA AMBER *ff*, their general penalization could introduce side-effects. Despite that the above simulations on SRL and T-loop motifs were not affected, a complete testing of weakened BPh interactions would be a major challenge. Thus, we considered construction of terminal-nucleotide-specific HBfixes (tHBfixes), i.e. HBfix-type modifications where one or both interacting atoms belong to the terminal residues. This should allow to tune simulations of TNs while not introducing side effects for the other systems. The RNA sequences most prone to form intercalated structures, i.e., r(CAAU), r(AAAA), and r(CCCC), contain -NH_2 group at the Hoogsteen edge of the nucleotide (either C_{S1} or A_{S1}) at the 5'-end. This amino group is forming 7BPh interaction⁴⁹ (Figure 4) with nbOs of the other terminal nucleotide at the 3'-end. Thus, a penalty to this specific interaction between groups from nucleotides on 5'- and 3'-ends should not introduce any side effects for other RNA structures as it is changing interactions only between two terminal nucleotides. We designed the gHBfix19 + tHBfix(NH...nbO) variant (Table 1), which is composed of the basic gHBfix19¹² extended by tHBfix penalizing interactions between -NH_2 group from either C_{S1} / A_{S1} (5'-end) and both nbOs of the 3'-end nucleotide by 1.0 kcal/mol (Table S1). We firstly tested the effect on r(CAAU) TN and observed that the agreement with experimental data is improved (χ^2 value of 3.63 is slightly worse than χ^2 of 2.50 from the simulation with gHBfix19_{v3}_BPh applied to all BPh contacts, while the basic gHBfix19 value is 5.37, see Table 2). However, the population of intercalated structures is still

significant (~16%, Table 2). Thus, the introduced gHBfix19 + tHBfix(NH...nbO) potential leads to clear improvement, but the extra 1.0 kcal/mol penalty to the single H-bond contact is still not sufficient to fully eliminate the occurrence of the intercalated state of r(CAAU) TN.

We also probed the effect of the extended gHBfix19 + tHBfix(NH...nbO) potential to all phosphates. Thus we penalized (by 1.0 kcal/mol) interactions between the $-NH_2$ group from the terminal C_{S1} (5'-end) and nbOs from all nucleotides of the r(CAAU) TN. We observed that the resulting conformational ensembles of these two different implementations of the gHBfix19 + tHBfix(NH...nbO) potential (i.e., biasing BPh interactions formed by nbOs of the phosphate from terminal nucleotide only and from all phosphates) are comparable within the limit of sampling (Table 2). Such a result was also confirmed when using the SPC/E water model instead of OPC (Table 2). We, however, tentatively propose that the second extended variant of the tHBfix(NH...nbO) potential (involving nbOs from all phosphates) might be useful for tuning simulations of larger RNA molecules, where different spurious (intercalated) states stabilized by BPh interactions involving the $-NH_2$ group from the 5'-end terminal nucleotide and nbOs from phosphates of different internal nucleotides might occur.

Table 1. The list of terminal groups and atoms from RNA nucleotides whose interactions were modified by the different tHBfix variants.^a

Label	Donors ^b				Acceptors ^c			
	5'-end Type	Res. # (atoms) ^d	Type	3'-end Res. # (atoms)	Nucleotide	Type	Res. # (atoms)	Nucleotide
tHBfix(NH...nbO)	base (-NH ₂)	S1 (H61, H62, H41, H42)			A, C	phosphate (nbO's)	S4 (<i>pro</i> -R _P , <i>pro</i> -S _P)	all
tHBfix(OH...nbO)			sugar (-OH)	S4 (HO'2, H3T)	all	phosphate (nbO's)	S3 (<i>pro</i> -R _P , <i>pro</i> -S _P)	all
	sugar (-OH)	S1 (H5T)			all	phosphate (nbO's)	S3 (<i>pro</i> -R _P , <i>pro</i> -S _P)	all
tHBfix(OH...O)	sugar (-OH)	S1 (H5T)			all	base (O)	S4 (O2)	U, C

tHBfix(OH...OH)			sugar (-OH)	S4 (HO'2, H3T)	all	sugar (OH)	S1 (O5')	all
tHBfix20 ^c	base (-NH ₂), sugar (-OH)	S1 (H61, H62, H41, H42), S1 (H5T)	sugar (-OH)	S4 (HO'2, H3T)	all	phosphate (nbO's), base (O), sugar (OH)	S4 (<i>pro</i> -R _P , <i>pro</i> -S _P), S3 (<i>pro</i> -R _P , <i>pro</i> -S _P), S4 (O2), S1 (O5')	all, U/C for S4 (O2)

^a see Tables S1 and S2 in the Supporting Information for the description of the particular tHBfix versions by using the detailed notation introduced in the original gHBfix paper¹² and for the full list of all specific interactions that were modified by the tHBfix20 potential, respectively.

^b H-bond donor (hydrogen of specific group considered).

^c H-bond acceptor (heavy atom (i.e., O or N) considered).

^d see Figure 2 for residue labelling.

^e combination of the four tHBfixes defined in the above rows, i.e., only those particular combinations of interactions listed in the tHBfix(NH...nbO), tHBfix(OH...nbO), tHBfix(OH...O), and tHBfix(OH...OH) rows are affected.

Intercalated states can be suppressed by penalizing SPh interactions formed by terminal –OH groups. Closer inspection of the intercalated state (Figures 2 and 3) revealed intricate network of SPh contacts, where the majority of them involve terminal –OH groups. Although these are already destabilized within the basic gHBfix19 by 0.5 kcal/mol, it appears that such destabilization is insufficient to fully correct the behavior of terminal –OH groups during MD simulations. Since the terminal –OH groups contain specific parameters within current *ffs*, e.g., they possess different electrostatic and vdW parameters compared to general –OH groups, it is entirely reasonable to treat their interaction separately from interactions established by the other –OH groups. Further, adding stronger destabilization for these specific contacts rather than penalizing all SPh interactions should be significantly less prone towards introducing undesirable side effects, e.g., elimination of A-RNA minor conformers, as discussed previously.¹² Thus, we designed the gHBfix19 + tHBfix20 variant for the CAAU sequence, where the additional tHBfix terms are destabilizing the following H-bonds by 1.0 kcal/mol: (i) the C_{S1}(N4H)...U_{S4}(*pro*-R_P/*pro*-S_P) BPh, (ii) the U_{S4}(2'-OH/3'-OH)...A_{S3}(*pro*-R_P/*pro*-S_P) SPh, (iii) the C_{S1}(5'-OH)...A_{S3}(*pro*-R_P/*pro*-S_P) SPh, (iv) the C_{S1}(5'-OH)...U_{S4}(O2) sugar-base H-bond, and (v) the

$U_{S4}(2'-OH/3'-OH)...C_{S1}(O5')$ sugar-sugar interactions (Table 1 and Table S1 in the Supporting Information). The REST2 simulation revealed the best agreement with the experiment for $r(CAAU)$ so far (the obtained χ^2 value of 2.05 is lowest so far, see Table 2). Thus, the structural dynamics of the $r(CAAU)$ during REST2 simulations is significantly improved and the population of intercalated structures is reasonably low ($\sim 5\%$, Table 2).

The gHBfix19 + tHBfix20 potential improves the structural dynamics of other TNs. We have carried out REST2 simulations with the gHBfix19 + tHBfix20 for the remaining four TNs. The newly designed potential significantly improves the structural dynamics of the $r(CCCC)$ TN (χ^2 value of 0.55, Table 2), since the agreement between all calculated and experimental observables, i.e., 3J -backbone couplings, 3J -sugar couplings, NOE and uNOE signals, is better (Table 2). The huge decrease of the χ^2 value coming from uNOE signals correlates with the decreased population of the spurious intercalated structure (from $\sim 37\%$ in simulation with the basic gHBfix19¹² to $\sim 6\%$). For the $r(GACC)$ TN, the calculated χ^2 value remains low (0.23) and is comparable with the value from the simulation with the basic gHBfix19¹² (χ^2 value of 0.40, Table 2). It is not possible to decide, which simulations agrees better with the experiment since all simulations with χ^2 value below 1 are satisfactory (see Methods).

The results for the remaining two TNs, i.e., $r(AAAA)$ and $r(UUUU)$, are ambiguous. Importantly, gHBfix19 + tHBfix20 eliminates population of intercalated structures within conformational ensembles of both TNs (to 1.2% and 1.9% for $r(AAAA)$ and $r(UUUU)$, respectively). However, total χ^2 values remain still rather high (1.08 and 1.81 for $r(AAAA)$ and $r(UUUU)$, respectively) indicating that the populations of states during MD simulations are not fully consistent with the experimental ensemble (Table 2). However, it is difficult to exactly

pinpoint the source of the difference. The χ^2 value of 1.81 for r(UUUU) TN is even marginally higher than the one from the simulation with the basic gHBfix19 (1.63).¹² Conformational ensembles of r(AAAA) and r(UUUU) REST2 simulations show that the major clusters are different from the canonical A-RNA major/minor conformers. The r(UUUU) TN is expected to be highly dynamical and to sample diverse conformations. We identified that 1-3/2-4 stacked structures are preferred during the REST2 simulation (clusters 1, 4 and 5 with total ~25% populations, with ~35% conformations unassigned; see Figure S4 in the Supporting Information). These states (typically with *syn* conformation of at least one U) appear to cause errors in NOEs and uNOEs data.

r(AAAA) TN preferred the four-stack state during simulations, but we noticed that two out of four nucleotides often favored *syn* conformation of χ dihedral (Figure S5 in the Supporting Information). The A_{S1} nucleotide at the 5'-end revealed higher propensity to sample *syn* conformation. Although it can be explained by the *syn*-specific A_{S1}(5'-OH)...A_{S1}(N3) H-bond,⁵⁴ it is possible that the *syn* state is overpopulated. Interestingly, in Ref. ²⁷ the combination of MD simulations with NMR data for nucleotides and dinucleotides suggested the need for increasing stabilization of the *anti* conformation for A nucleotides in the *ff*. Similar indication was obtained from another fitting based on the same experimental data as used here⁵⁵ (see <https://github.com/bussilab/ff-fitting-tools/blob/master/Analysis.ipynb>). In other words, the large tendency of the A nucleotide to sample *syn* conformation could indicate that its *syn/anti* balance is not fully correct in the χ_{OL3CP} RNA *ff*, suggesting a possible modification of the dihedral potential. However, the likely spurious preference for *syn* orientation in r(AAAA) simulations might be also caused by not enough accurate description of the stacking interactions. Stacking is

suspected to be overestimated within current RNA *ffs*,⁵⁶⁻⁵⁸ albeit the OPC water model might partly reduce such overestimation.^{19, 59}

We performed stacking overlap analysis with an in-house 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 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_{S1}-A_{S2}-A_{S3}-A_{S4}). 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-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 the consecutively stacked nucleobases, i.e., we separately analyzed A_{S1}-A_{S2} (5'-end), A_{S2}-A_{S3} (middle), and A_{S3}-A_{S4} (3'-end) stacking pairs, in their various *syn/anti* orientations. The largest stacking overlap is achieved with *anti/anti* combination at the 5'-end, but *syn/syn* and *syn/anti* orientations have the largest stacking overlap in the middle step and at the 3'end, respectively (Table S3 in the Supporting Information). Based on this simple analysis, we suggest a possibility that overestimated stacking interaction is pushing the A_{S3} 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.

gHBfix and tHBfix combined with different solvent models. The OPC water model¹⁸ is nowadays widely used for enhanced-sampling simulations of simple RNA motifs^{10, 12, 26-28, 30, 60} because it somewhat reduces the excessive formation of SPh contacts,¹⁹ especially in combination with the revised phosphate parameters.¹⁷ On the other hand, the OPC water is suboptimal for G-quadruplex simulations.^{61, 62} It indicates that it may not represent a universal solution for MD simulations of all nucleic acids, since desirable weakening of spurious interactions can sometimes be accompanied by undesirable destabilization of native interactions. Thus, we investigated effects of the gHBfix correction to the RNA $\ddot{f}$ in combination with different solvent models.

We firstly tested the TIP3P water model^{38, 63} during REST2 simulations of r(GACC) with the gHBfix19 potential.¹² The population of intercalated structures is significantly increased (~12%, Table 2) and the χ^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 (χ^2 value of 0.75, Table 3 in Ref. ¹²). In addition, we observed enhanced probability to form loop-like structures, where G_{S1} established the Watson-Crick base pairs with C_{S3} and C_{S4} forming 1-3 loops (~10% population) and 1-4 loops (~3% population), respectively (Table 2, Figure S6 in the Supporting Information).

Next, we performed gHBfix19 simulation of the r(GACC) TN with the SPC/E water model,^{39, 64} which is often used in simulations of protein-RNA complexes.¹¹ The simulation revealed higher total χ^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). The increased χ^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 (~2%), comparable to the OPC data (Table 2). Thus, we designed a new gHBfix19_NH-N_{+0.75}, 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 ~7% (from ~32% in the simulation with the basic gHBfix19 potential, Table 2). Then we tested the effect of SPC/E water model in two simulations of r(CAAU) TN using the combined gHBfix19 + tHBfix(NH...nbO) terms. Unfortunately, population of intercalated states significantly increased (reaching more than 60%), which correlates with high total χ^2 values of 6.84 and 7.75 (Table 2). 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. Finding of suitable gHBfix + tHBfix parameters with the 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 proteins⁴⁰ and was also recently suggested for RNA simulations.³² Initially, we tested the TIP4P-D model during REST2 simulation of the r(GACC) TN using two different ion parameters, i.e., JC³⁷ and charmm22.⁴¹ Results from both REST2 simulations of the r(GACC) TN revealed good agreement with the experiment (Table 2) 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 χ^2 value is slightly lower for r(CAAU) TN and slightly higher for both r(CCCC) and r(UUUU) TNs, Table 2). 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 ~8% to ~27%, Table 2). 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 model). The reason causing the difference between SPC/E 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 excluded³³ but full testing is beyond the scope of this work due to the high computer demands.

In conclusion, although our results are far from providing an extensive testing of the water models it seems that usage of three-point water models, e.g., TIP3P and SPC/E, instead of OPC in TN simulations increases the population of loop-like and/or intercalated states. However, it cannot be ruled out that it would be possible to find gHBfix and tHBfix settings which provide better results also with these water models. The gHBfix settings, however, would have to be different from those used with OPC. The simulation outcome with the TIP4P-D water was generally comparable with the OPC water model, except of the rAAAA TN. We note that for this water model we did not test the tHBfix parameters though we assume that the effect of the tHBfix could be similar to the OPC water model. The 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.⁶¹ The results confirm sensitivity of nucleic acids simulations to water models.^{4, 19, 61, 65, 66} This complicates tuning of nucleic acids *ffs*, 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.

Reference temperature does not affect the agreement between computed and experimental results. The NMR datasets of TN motifs were measured at 275 K,²²⁻²⁴ 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 for the details). The outcome is comparable within limits of sampling to the 298 K simulation (Table 2). 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.

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.¹² 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 ⁵⁵. We used the 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.¹² 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 β , 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.

The possibility to reproduce the effect of the combined gHBfix19 and tHBfix20 potential by the NBfix approach. In principle, similar effect as obtained by the HBfix-type of potentials could be achieved by modification of the pairwise vdW parameters via breakage of the combination (mixing) rules, i.e., the so-called nonbonded fix (NBfix) approach; see recent review⁶⁷ for details. However, as we discussed earlier,¹² use of the HBfix-type approach is more straightforward and brings some advantages. To tune all possible H-bonds by the NBfix approach is a demanding work with huge dimensionality, as it is necessary to create new atom types to differentiate atoms on different residues with same atom types and different partial charges. In addition, as discussed by us in Ref. ¹², the NBfix approach affects both interaction energy and optimal geometry, so its applicability may be limited by geometrical constraints. Thus, the HBfix terms (e.g., gHBfix, tHBfix) are more flexible for fine-tuning of the non-bonded interactions with respect to the NBfix approach in pair-specific treatment of H-bonds and interatomic contacts. In the preceding study we have nevertheless published a set of NBfix parameters that have similar effect as the gHBfix19 potential.¹² In practice, the NBfix parameters could be derived a posteriori after finding the suitable gHBfix version, since the gHBfix allows more direct tuning of the H-bonds. Nevertheless, it is probable that both approaches will be combined in future for tuning of *ffs*.

In the present work we also derived the NBfix parameters to reproduce the interaction energy curves obtained with the gHBfix19 and tHBfix20 terms in the same way as it is discussed in Ref.¹². Details are provided in the Supporting Information together with the NBfix parameters which can be used in simulations as an alternative to the gHBfix19 + tHBfix20 parameter set. 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 the Supporting Information), 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 χ^2 value of 0.23 and populations of major conformers within both ensembles were only marginally different (Table 2). Due to the computer demands we did not make testing for the remaining TNs.

CONCLUDING REMARKS

Performance of pair-additive biomolecular *ffs* is determined by a wide range of parameters. Although the most common approach to refine biomolecular *ffs* is tuning of the dihedral potentials, it is becoming clear that refinement of RNA *ffs* requires also tuning of the nonbonded terms.^{10, 12, 19-21, 24, 28, 30, 32, 55} Correcting errors in H-bonding and stacking interactions exclusively via formally intramolecular dihedral potentials is certainly a non-optimal approach. It has also been suggested that improvements of the existing pair-additive biomolecular *ffs* may profit from increase of flexibility of the parametrization, which can be achieved for example by breaking of

the Lennard-Jones term combination rules (i.e., the NBfix approach) or by adding new simple terms to the *ff* form (the gHBfix approach, see Figure 1).^{4, 12}

Unique primary solution NMR data are available for a set of RNA TNs and contemporary enhanced-sampling simulation methods allow to obtain well-converged ensembles of these model systems. Therefore, RNA TNs can be straightforwardly used to tune the RNA *ffs* by comparing conformational ensembles from simulations with NMR experimental data, although one should always keep in mind that there is a high risk of overfitting the *ff* towards RNA A-form. Recently, we have used the gHBfix approach to improve performance of the common χ_{OL3CP} AMBER RNA *ff* version by strengthening the –NH...N– base-base H-bonds and weakening the SPh H-bonds; this *ff* version is abbreviated as gHBfix19 in the present work.¹²

The gHBfix19 version visibly improved the RNA simulations without so far any detected side-effects.¹² However, this *ff* version was still not fully satisfactory in reproducing the experimental data for RNA TNs. Here we further modify the gHBfix19 potential by tuning stability of H-bond interactions formed by the terminal nucleotides, namely the SPh and BPh interactions. The approach is abbreviated as tHBfix term. The idea to treat H-bonds formed by groups from terminal residues differently from H-bonds formed by equivalent groups of other nucleotides can be justified by several reasons. First, these groups have different parameters compared to the internal nucleotides in the parent RNA *ffs*. Second, the terminal nucleotides in reality may have somewhat different electronic-structure properties from the internal nucleotides. Finally, MD simulations of TNs are very significantly affected by interactions involving the terminal nucleotides while such interactions usually have insignificant effect on simulations of larger RNAs. Thus, considering the empirical nature of the *ff*, it is meaningful to not sacrifice

transferability of the parameters of the internal nucleotides in order to perfect simulations of the short TNs.

We applied enhanced-sampling methods on the testing set containing five RNA TNs and compared the structural ensembles from simulations with accurate experimental datasets. The total raw amount of our simulations is over two milliseconds. Our results show that r(CAAU), r(CCCC) and r(GACC) TNs are mostly sampling A-form-like conformations in agreement, within the resolution limits, with the experiments. The remaining r(AAAA) and r(UUUU) TNs display a higher complexity with a mixture of structures. Comparison with the primary NMR data shows that their MD ensembles are likely different, to a detectable extent, from those populated in the experiment. However, it was not possible to unambiguously identify the exact origin of these residual discrepancies.

As expected based on the fact that we increased the number of parameters and that we used the NMR data in training them, the results with the newly designed χ_{OL3CP} + gHBfix19 + tHBfix20 potential are significantly improved in comparison with our previous work with gHBfix19.¹² Although we did not make any extensive testing we suggest that side effects from adding the tHBfix20 parameters should be marginal since only terminal nucleotides are affected. Due to the profound sensitivity of TN simulations to the water models, the gHBfix19 + tHBfix20 parameter set is rather specific for the OPC water model. We also present set of NBfix parameters that should provide similar results as the gHBfix19 + tHBfix20 refinement.

In conclusion, we show that targeting the terminal residues of the *ff* may be considered both as an attempt to reflect potentially different electronic structures of these residues and a pragmatic means to increase tuning capability of the *ff*. Changing behavior of terminal nucleotides should not have side effects when simulating larger RNA motifs. Additionally, we propose that specific

modifications of H-bonds formed by terminal nucleotides might potentially improve other *ff* features like spurious dynamics of RNA termini, description of base-pair fraying and simulations of protein-RNA complexes, where RNA self-interactions can create problems.⁴ In overall, the tHBfix potential is a generally transferable *ff* modification and can be applied to generic RNA motifs, although future refinements of the parameters are to be expected. This is because tuning of the non-bonded terms, paring them with water models and subsequent testing of all different *ff* versions on diverse RNAs represents a major computational challenge which cannot be fully accomplished within the framework of one computational study.

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge. Detailed description of MD protocol, Supporting Tables and Figures (PDF).

AUTHOR INFORMATION

Corresponding Authors

mlynsky@ibp.cz

sponer@ncbr.muni.cz

ACKNOWLEDGMENT

Scott Kennedy and Douglas Turner are acknowledged for providing the NMR dataset for the r(CAAU) TN. This work was supported by the Operational Programme Research, Development and Education–European Regional Development Fund, Project No.

CZ.02.1.01/0.0/0.0/16_019/0000754 (P.B., P.K., M.O.), Czech Science Foundation (18-25349S to P.B., P.K. and 20-16554S to V.M. and J.S.) and by Praemium Academiae (J.S.). This work was also supported by the project SYMBIT reg. number: CZ.02.1.01/0.0/0.0/15_003/0000477 financed by the ERDF (J.S.)

Table 2: REST2 simulations of RNA TNs using the common χ_{OL3CP} RNA *ff* combined with diverse gHBfix and tHBfix variants.

All simulations used 8 replicas and were run for 10 μs (the last 7 μs were used for data analysis).^a

Motif	gHBfix label ^a	Solvent ^b	χ^2	Clustering [%] [A-form / Loop / Intercalated]	# of clusters	ϵRMSD populations [%]
			[³ J-backbone, ³ J-sugar, NOE, uNOE (# of violations) / Total] ^c			[A-form (major, minor) / Loop (1-3, 1-4) / Intercalated]
r(GACC) ^d	gHBfix19	OPC	0.39, 0.50, 1.13, 0.34 (3) / 0.40	~74/~7/~4	5	48.0, 9.8 / 0.8, 7.0 / 2.1
r(GACC)	gHBfix19	TIP3P	0.58, 0.48, 1.31, 0.58 (5) / 0.62	~57/~17/~18	9	32.5, 6.8 / 10.3, 2.7 / 11.7
r(GACC)	gHBfix19	SPC/E	0.57, 0.66, 1.44, 0.91 (6) / 0.92	~60/~32/~4	7	36.6, 7.0 / 9.7, 14.4 / 1.5
r(GACC)	gHBfix19_NH-N _{+0.75}	SPC/E	0.36, 0.41, 1.00, 0.13 (4) / 0.20	~84/~7/~2	7	53.3, 10.0 / 1.7, 4.0 / 2.6
r(GACC)	gHBfix19	TIP4P-D	0.29, 0.42, 1.09, 0.00 (0) / 0.10	~86/~0/~3	5	67.8, 18.0 / 0.0, 0.0 / 1.1
r(GACC)	gHBfix19	TIP4P-D ^e	0.32, 0.37, 1.00, 0.22 (4) / 0.28	~74/~8/~6	8	62.0, 7.8 / 0.0, 6.5 / 1.9
r(GACC)	gHBfix19 + tHBfix20	OPC	0.34, 0.37, 1.11, 0.15 (3) / 0.23	~81/~5/~0	4	60.3, 7.2 / 0.5, 4.7 / 1.0
r(GACC)	NBfix ^f	OPC	1.11, 0.55, 1.08, 0.11 (5) / 0.23	~76/~4/~0	7	46.6, 13.2 / 1.9, 0.4 / 2.9
r(CAAU) ^d	gHBfix19	OPC	0.73, 1.14, 1.81, 6.07 (28) / 5.37	~33/~0/~51	5	13.2, 4.6 / 0.0, 0.0 / 48.1
r(CAAU)	gHBfix19	TIP4P-D	0.52, 1.14, 1.74, 5.04 (16) / 4.48	~42/~0/~51	4	19.1, 19.3 / 0.0, 0.0 / 47.8
r(CAAU)	gHBfix19 _{v2} _BPh	OPC	0.70, 0.97, 1.64, 5.18 (34) / 4.59	~58/~0/~22	8	18.5, 5.3 / 0.0, 0.0 / 19.9
r(CAAU)	gHBfix19 _{v2} _BPh _{v2}	OPC	0.69, 0.95, 1.63, 4.18 (33) / 3.74	~55/~0/~17	8	19.5, 6.4 / 0.0, 0.0 / 13.6
r(CAAU)	gHBfix19 _{v3} _BPh	OPC	0.64, 0.81, 1.69, 2.72 (25) / 2.50	~64/~0/~8	8	25.8, 5.1 / 0.0, 0.0 / 6.8
r(CAAU)	gHBfix19 + tHBfix(NH...nbO)	SPC/E	0.74, 1.26, 3.03, 7.67 (40) / 6.84	~8/~3/~69	5	3.7, 0.9 / 0.0, 0.0 / 40.3
r(CAAU)	gHBfix19 + tHBfix(NH...nbO) ^g	SPC/E	1.17, 1.09, 5.70, 8.45 (36) / 7.75	~0/~0/~88	3	0.8, 0.7 / 0.0, 0.0 / 62.5
r(CAAU)	gHBfix19 + tHBfix(NH...nbO)	OPC	0.56, 1.01, 1.33, 4.09 (33) / 3.63	~63/~0/~19	6	19.9, 8.5 / 0.0, 0.0 / 16.0
r(CAAU)	gHBfix19 + tHBfix(NH...nbO) ^g	OPC	0.54, 0.96, 1.25, 3.82 (34) / 3.40	~61/~0/~17	9	22.3, 8.4 / 0.0, 0.0 / 13.9
r(CAAU)	gHBfix19 + tHBfix20	OPC	0.49, 0.80, 1.40, 2.22 (29) / 2.05	~74/~0/~7	8	29.1, 6.8 / 0.0, 0.0 / 5.1
r(CAAU)	gHBfix19 + tHBfix20 ^h	OPC	0.49, 0.88, 1.37, 2.56 (32) / 2.34	~68/~0/~3	8	31.1, 5.5 / 0.0, 0.0 / 2.3
r(AAAA) ^d	gHBfix19	OPC	0.66, 1.23, 1.08, 1.14 (12) / 1.11	~30/~0/~10	17	12.1, 4.4 / 0.0, 0.0 / 8.1

r(AAAA)	gHBfix19	TIP4P-D	0.69, 1.12, 1.03, 2.22 (11) / 1.96	~29/~0/~32	13	14.6, 6.8 / 0.0, 0.0 / 26.8
r(AAAA)	gHBfix19 + tHBfix(NH...nbO)	OPC	0.67, 1.38, 1.20, 1.02 (11) / 1.04	~38/~0/~5	16	12.1, 4.5 / 0.0, 0.0 / 1.7
r(AAAA)	gHBfix19 + tHBfix20	OPC	0.68, 1.48, 1.28, 1.06 (11) / 1.08	~32/~0/~1	18	12.5, 4.5 / 0.0, 0.0 / 1.2
r(CCCC) ^d	gHBfix19	OPC	0.44, 0.68, 1.54, 3.22 (11) / 2.83	~55/~0/~40	2	39.1, 5.5 / 0.0, 0.0 / 37.1
r(CCCC)	gHBfix19	TIP4P-D	0.45, 0.53, 1.73, 3.72 (10) / 3.26	~49/~0/~48	3	41.0, 6.4 / 0.0, 0.0 / 46.6
r(CCCC)	gHBfix19 + tHBfix(NH...nbO)	OPC	0.31, 0.46, 1.14, 1.30 (9) / 1.20	~74/~0/~20	5	52.2, 7.8 / 0.0, 0.0 / 26.4
r(CCCC)	gHBfix19 + tHBfix20	OPC	0.25, 0.38, 1.22, 0.50 (6) / 0.55	~85/~0/~8	5	63.3, 7.1 / 0.0, 0.0 / 6.1
r(UUUU) ^d	gHBfix19	OPC	0.53, 0.87, 3.53, 1.65 (19) / 1.63	~5/~0/~3	10	6.6, 1.7 / 0.0, 0.0 / 3.8
r(UUUU)	gHBfix19	TIP4P-D	0.52, 0.72, 3.85, 1.52 (16) / 1.51	~25/~0/~4	13	14.0, 6.1 / 0.0, 0.0 / 3.8
r(UUUU)	gHBfix19 + tHBfix20	OPC	0.52, 0.97, 3.46, 1.86 (17) / 1.81	~14/~0/~0	10	6.4, 1.5 / 0.0, 0.0 / 1.9

^a see Table S1 in the SI for the detailed definition of the tested variants of the gHBfix and tHBfix potentials.

^b 0.15 M KCl, JC ion parameters.³⁷

^c χ^2 values were obtained by comparing calculated and experimental backbone ³J scalar couplings, sugar ³J scalar couplings, nuclear Overhauser effect intensities (NOEs), and the absence of specific peaks in NOE spectroscopy (uNOEs, number of violations is also reported). Note that we identified some potential uncertainty in the experimental datasets, which resulted in higher χ^2 values and lowered the agreement between simulations and the experiment for r(CCCC) and r(UUUU) TNs (see Methods for details).

^d data taken from the preceding gHBfix paper.¹²

^e simulation run with charmm22 ion parameters.⁴¹

^f NBfix *ff* reparameterization was prepared in a way to be comparable with the combined gHBfix19 + tHBfix20 potential (see Table S4 in the Supporting Information for details).

^g extended variant of tHBfix(NH...nbO) between –NH₂ group from Cs₁ (5'-end) and nbO atoms of *all phosphates*, i.e., not only nbOs from the terminal Us₄ nucleotide.

^h REST2 simulation run with unbiased replica shifted from 298 K to 275 K (to mimic the experimental temperature).

REFERENCES

1. Vangaveti, S.; Ranganathan, S. V.; Chen, A. A., Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdiscip Rev RNA* **2017**, 8 (2).
2. Smith, L. G.; Zhao, J.; Mathews, D. H.; Turner, D. H., Physics-based all-atom modeling of RNA energetics and structure. *Wiley Interdiscip Rev RNA* **2017**, 8 (5).
3. Nerenberg, P. S.; Head-Gordon, T., New developments in force fields for biomolecular simulations. *Curr Opin Struct Biol* **2018**, 49, 129-138.
4. Spomer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; Walter, N. G.; Otyepka, M., RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* **2018**, 118 (8), 4177-4338.
5. Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J. P.; Ponder, J. W.; Ren, P., AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J Chem Theory Comput* **2018**, 14 (4), 2084-2108.
6. Lemkul, J. A.; MacKerell, A. D., Jr., Polarizable force field for RNA based on the classical drude oscillator. *J Comput Chem* **2018**, 39 (32), 2624-2646.
7. Lagardere, L.; Jolly, L. H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J. P., Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem Sci* **2018**, 9 (4), 956-972.
8. Liu, C.; Piquemal, J. P.; Ren, P., AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J Chem Theory Comput* **2019**, 15 (7), 4122-4139.
9. Baker, C. M., Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdiscip Rev Comput Mol Sci* **2015**, 5 (2), 241-254.
10. Kuhrova, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Spomer, J.; Otyepka, M.; Banas, P., Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J Chem Theory Comput* **2016**, 12 (9), 4534-48.
11. Spomer, 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 (3).
12. Kuhrova, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Spomer, J.; Banas, P., Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J Chem Theory Comput* **2019**, 15 (5), 3288-3305.
13. 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 (9), 2309-2309.
14. 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 (12), 1049-1074.
15. Perez, A.; Marchan, I.; Svozil, D.; Spomer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J* **2007**, 92 (11), 3817-3829.

16. 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* (9), 2886-2902.
17. Steinbrecher, T.; Latzer, J.; Case, D. A., Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput* **2012**, *8* (11), 4405-4412.
18. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J Phys Chem Lett* **2014**, *5* (21), 3863-3871.
19. Bergonzo, C.; Cheatham, T. E., Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J Chem Theory Comput* **2015**, *11* (9), 3969-3972.
20. 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* (9), 1578-1590.
21. 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* (2), 900-915.
22. 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* (29), 9261-70.
23. 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* (6), 996-1010.
24. 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* (6), 2729-2742.
25. Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E., 3rd, Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J Chem Theory Comput* **2014**, *10* (1), 492-499.
26. Gil-Ley, A.; Bottaro, S.; Bussi, G., Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J Chem Theory Comput* **2016**, *12* (6), 2790-8.
27. Cesari, A.; Gil-Ley, A.; Bussi, G., Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J Chem Theory Comput* **2016**, *12* (12), 6192-6200.
28. Yang, C.; Lim, M.; Kim, E.; Pak, Y., Predicting RNA Structures via a Simple van der Waals Correction to an All-Atom Force Field. *J Chem Theory Comput* **2017**, *13* (2), 395-399.
29. Schrod, M. V.; Andrews, C. T.; Elcock, A. H., Large-Scale Analysis of 48 DNA and 48 RNA Tetranucleotides Studied by 1 μ s Explicit-Solvent Molecular Dynamics Simulations. *J Chem Theory Comput* **2015**, *11* (12), 5906-17.
30. 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* (5), eaar8521.
31. 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**.

32. 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* (7), E1346-E1355.
33. Kuhrova, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Sponer, J.; Banas, P., Correction to "Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions". *J Chem Theory Comput* **2019**.
34. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J Comput Chem* **2005**, *26* (16), 1668-88.
35. Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; P.A., K., *AMBER 2016*. San Francisco, 2016.
36. Mlynsky, V.; Kuhrova, P.; Zgarbova, M.; Jurecka, P.; Walter, N. G.; Otyepka, M.; Sponer, J.; Banas, P., Reactive conformation of the active site in the hairpin ribozyme achieved by molecular dynamics simulations with epsilon/zeta force field reparametrizations. *J Phys Chem B* **2015**, *119* (11), 4220-9.
37. 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* (30), 9020-9041.
38. 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* (2), 926-935.
39. 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* (8), 3684-3690.
40. 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* (16), 5113-23.
41. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998**, *102* (18), 3586-616.
42. Wang, L.; Friesner, R. A.; Berne, B. J., Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2) (vol 115, pg 9431, 2011). *J Phys Chem B* **2011**, *115* (38), 11305-11305.
43. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **2015**, *11* (4), 1864-1874.
44. Rodriguez, A.; Laio, A., Clustering by fast search and find of density peaks. *Science* **2014**, *344* (6191), 1492-1496.

45. Bottaro, S.; Di Palma, F.; Bussi, G., The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res* **2014**, *42* (21), 13306-14.
46. Hall, P.; Horowitz, J., A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions. *Ann Stat* **2013**, *41* (4), 1892-1921.
47. https://www.ilikebigbits.com/2015_03_04_plane_from_points.html.
48. Jarvis, R. A., On the identification of the convex hull of a finite set of points in the plane. *Inf Process Lett* **1973**, *2* (1), 18-21.
49. Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B., Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res* **2009**, *37* (15), 4898-918.
50. Olieric, V.; Rieder, U.; Lang, K.; Serganov, A.; Schulze-Briesse, C.; Micura, R.; Dumas, P.; Ennifar, E., A fast selenium derivatization strategy for crystallization and phasing of RNA structures. *Rna* **2009**, *15* (4), 707-715.
51. 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.
52. Nagaswamy, U.; Fox, G. E., Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *Rna* **2002**, *8* (9), 1112-9.
53. 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* (15), 4214-21.
54. Cang, X.; Sponer, J.; Cheatham, T. E., 3rd, Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res* **2011**, *39* (10), 4499-512.
55. 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* (6), 3425-3431.
56. Sponer, J.; Riley, K. E.; Hobza, P., Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys Chem Chem Phys* **2008**, *10* (19), 2595-610.
57. Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P., Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys Chem Chem Phys* **2010**, *12* (35), 10476-93.
58. Kruse, H.; Banas, P.; Sponer, J., Investigations of Stacked DNA Base-Pair Steps: Highly Accurate Stacking Interaction Energies, Energy Decomposition, and Many-Body Stacking Effects. *J Chem Theory Comput* **2019**, *15* (1), 95-115.
59. Hase, F.; Zacharias, M., Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Res* **2016**, *44* (15), 7100-8.
60. Kuhrova, P.; Banas, P.; Best, R. B.; Sponer, J.; Otyepka, M., Computer Folding of RNA Tetraloops? Are We There Yet? *J Chem Theory Comput* **2013**, *9* (4), 2115-2125.
61. 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* (8), 3911-3926.
62. Islam, B.; Stadlbauer, P.; Vorlíčková, M.; Mergny, J.-L.; Otyepka, M.; Šponer, J., Stability of Two-quartet G-quadruplexes and Their Dimers in Atomistic Simulations. *bioRxiv* **2019**, 820852.
63. Neria, E.; Fischer, S.; Karplus, M., Simulation of activation free energies in molecular systems. *J Chem Phys* **1996**, *105*, 1902-1921.

64. 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 (5-6), 429-436.
65. 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 (7), 2506-2520.
66. 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 (9), 2963-2980.
67. Yoo, J.; Aksimentiev, A., New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys Chem Chem Phys* **2018**, 20 (13), 8432-8449.