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

Vojtěch Mlýnský,[1] Petra Kührová,[2] Tomáš Kühr,[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 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 terminal atoms than the internal nucleotides. Combination of gHBfix with tHBfix significantly improves the behavior of RNA TNs during well-converged enhanced-sampling simulations using replica exchange with solute tempering. 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 primary NMR data which suggests that future tuning of some other *ff* terms might be useful. Nevertheless, the tHBfix has a clear potential to improve modeling of key biochemical processes, where interactions of RNA single stranded ends are involved.

KEYWORDS
RNA, force field, tetranucleotide, enhanced sampling, NMR, HBfix

INTRODUCTION

Atomistic description of structural dynamics of RNA molecules is essential for understanding of some key biomolecular processes including, e.g., translation, splicing, and other RNA-regulated processes. 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. Unfortunately, currently used (state-of-the-art) empirical potentials (force fields, *ff*s) suffer from many inaccuracies which may even lead to a preferential sampling of spurious structures instead of native states. Performance of available pair-additive atomistic RNA *ff*s has been reviewed in detail,[1-4] with a suggestion that the currently used form of pair-additive RNA *ff*s is approaching limits of its applicability.[4] In other words, it is becoming increasingly difficult to generally improve performance of 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 *ff*s.[5-9] Alternatively, we have suggested that improvements could be achieved by extending pair-additive *ff*s 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 description of hydrogen-bonding (H-bond) 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 tetraloops[10] 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 gHBfix parameters complement the common OL3 (*ff*99bsc0χ_{OL3})[13-16] AMBER RNA *ff* version used together with modified phosphate van der Waals (vdW) parameters[17] and OPC water model.[18]

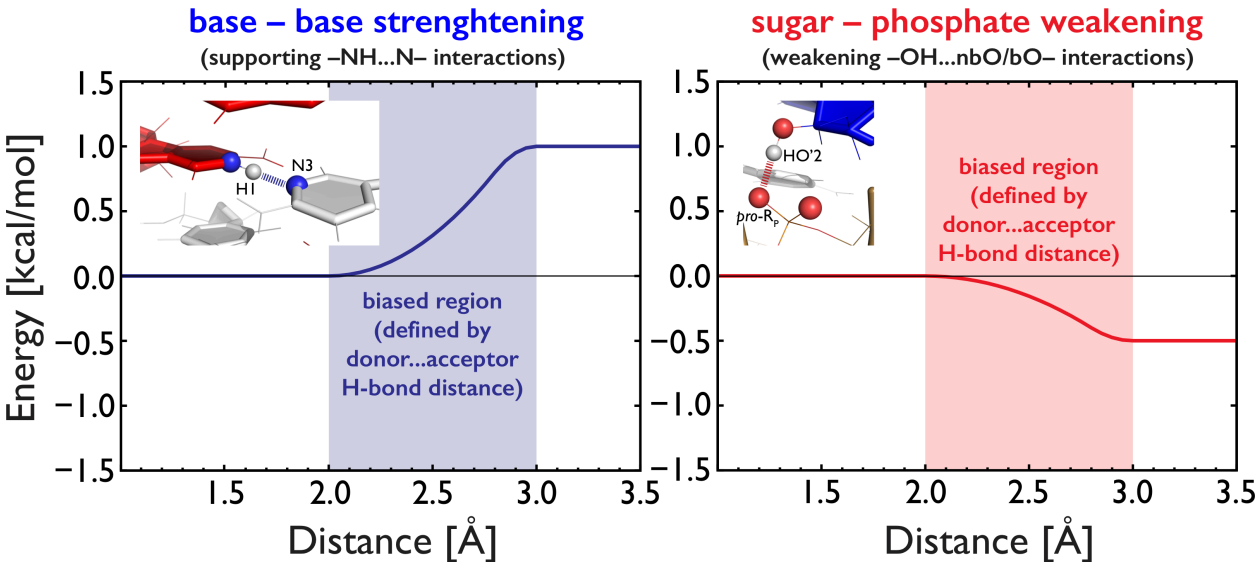

Figure 1: Description of the gHBfix *ff* term used for either support (blue curve) or weakening (red curve) of H-bond interactions. The initial version of gHBfix introduced in Ref. [12] is tuning H-bond interactions in order to: (i) support base – base –NH…N– interactions (left panel, blue dashed line) and, simultaneously, (ii) destabilize SPh interactions (red dashed line 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 SPh and base-phosphate (BPh) H-bonds involving 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 well-founded as their properties can differ from internal nucleotides and their imbalanced description in simulations can have detrimental effects. In other words, modification of interactions of terminal nucleotides can separately eliminate some simulation problems without affecting 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 final goal is not to reproduce experimental data for small model systems, but to provide parameters that can be used to study folded RNA structures and diverse protein-RNA complexes. Targeting terminal residues of the *ff* may be considered both as a pragmatic means to increase tuning capability of the *ff* without side effects and an attempt to reflect potentially different physical properties of these residues. Terminal nucleotides might indeed have different electronic structures. Moreover, terminal nucleotides are more exposed to solvent than internal nucleotides. As the strength of intramolecular H-bonds depends on the difference between those H-bonds and binding to the solvent, it is fully justified, when using empirical *ff*s, to treat interactions of terminal nucleotides differently from the internal ones. In

perspective, an improved description of terminal residues will allow better studies of important biochemical processes dependent on dynamics of RNA single stranded ends, e.g., interactions between 3'-single stranded ends of tRNAs and amino acids, splicing of the anticodon loop of tRNAs, RNAi and antisense systems using single stranded ends as therapeutics, etc.[19]

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.[20-32] 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, 21, 31] TN simulations are specifically sensitive to performance of *ff*s 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, 21, 25, 31] 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*[33]), which may lead to side-effects for simulations of folded RNAs.[12, 34]

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. Simulations of r(CAAU), r(CCCC) and r(GACC) TNs sampled 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 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.

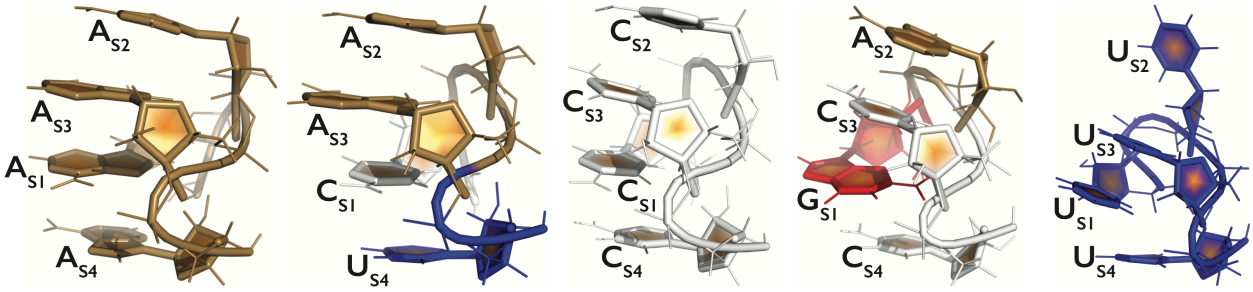

Figure 2: Tertiary structures of the five studied systems in the spurious intercalated structure involving the 5'-terminal nucleoside, which is often seen during MD simulations using common RNA AMBER *ff*s.[12, 20-26, 28-29, 31-33] A, C, G and U nucleotides are colored in sand, white, red, and blue, respectively.

METHODS

Starting structures and simulation setup. Initial coordinates of r(AAAA), r(CAAU), r(CCCC), r(GACC), and r(UUUU) TNs were prepared using Nucleic Acid Builder of AmberTools14[35] as one strand of an A-form duplex. Starting topologies and coordinates for 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.[36] 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 Å^3 box size.

We used the standard OL3[13-16] 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, 37] (see the Supporting Information of Ref. [10] for parameters); this version is abbreviated as χ_{OL3CP} henceforth. Additionally, we applied the external gHBfix potential[12] that was shown to improve the overall *ff* performance. Most of simulations were carried out with the OPC[18] water model and in ~0.15 M KCl excess salt using the Joung−Cheatham (JC)[38] ionic parameters. Specific tests involved TIP3P,[39] SPC/E,[40] and TIP4P-D[41] water models. One test simulation of the r(GACC) TN with the TIP4P-D water was combined with charmm22 ion parameters[42] (see the Supporting Information for other details about the simulation protocol).

Enhanced sampling. Replica exchange solute tempering (REST2)[43] 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 repartitioning[44] 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.[23-25, 32] The same λ values were applied resulting in effective solute temperature range from 275 K to ~460 K. The 275 and 298 REST2 simulations revealed comparable results considering limits of sampling (see Supporting Information for details).

HBfix-type potentials used in this work. We applied the previously introduced generalized (interaction-specific) gHBfix potential[12] that is improving overall performance of the χ_{OL3CP} RNA *ff*. Most simulations used the basic gHBfix potential with the same parameters indicated earlier as optimal[12] (Figure 1, abbreviated as gHBfix19, see below). Besides that we have used several other variants of HBfix-type potentials. 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 5'-terminal nucleoside and 3'-terminal nucleotide (common abbreviation *terminal nucleotides* is used through the text, 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. Dominant conformations sampled during REST2 simulations were identified using a cluster analysis based on an algorithm introduced by Rodriguez and Laio[45] in combination with εRMSD,[46] discussed in details previously.[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 cluster points and a cluster hull representing a noise spread around a given cluster,[10] these two conformations merge into one cluster, since the transitions between them were sampled too frequently. Thus, as an additional descriptor, we applied simple εRMSD analysis, where we used representative PDB's of the most important conformations and calculated populations of those

states from MD trajectories based on an εRMSD cutoff representing a conservative definition of the conformational state (Table 2).

The convergence was verified by bootstrapping[47] using recently introduced implementation.[10, 12] Errors were estimated by resampling of time blocks of the whole set of replicas and subsequent resampling of the coordinate-following replicas. This allows their demultiplexing to follow Hamiltonians in order to obtain a final resampled population at the reference unbiased Hamiltonian state (see Figure S1 in 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 *syn/anti* balance of nucleobases. All heavy atoms of each nucleobase are fitted by a plane,[48] projected into this plane and transformed to a convex planar polygon.[49] 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 Å) is projected into the plane of the first polygon. The area of calculated intersection between both polygons is measured and expressed as a relative area with respect to area of the first polygon. Thus, calculated values of the geometrical overlap range 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.[23-25] We analyzed separately four NMR observables, i.e., (i) backbone ^{3}J scalar couplings, (ii) sugar ^{3}J scalar couplings, (iii) nuclear Overhauser effect intensities (NOEs), and (iv) the absence of specific peaks in NOE spectroscopy (uNOEs). Their combination calculated as weighted arithmetic mean provided the *total* χ^2 value for each REST2 simulation: $\chi^2_{total} = \frac{\sum_i^n \chi_i^2 n_i}{\sum_i^n n_i}$, where χ_i^2 and n_i are *individual* χ^2 values and the number of observables, respectively, corresponding to the particular NMR observables. The lower *total* χ^2 value, the better agreement between experimental data and 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 ^{3}J scalar couplings and NOE intensities 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 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 experimental data, one should always check also the *individual* χ^2 (χ_i^2) components corresponding to particular NMR observables.

We further identified that the signal range for some NOEs, i.e., C_{S2}(H6)…C_{S2}(H5’’), U_{S2}(H3’)…U_{S2}(H5’), and U_{S3}(H3’)…U_{S3}(H5’), was rarely populated during MD simulations, i.e., the measured distances were typically shifted from the expected range by more than 1 Å (the MD distance was longer than the NOE distance in ~99% and ~97% of r(CCCC) and r(UUUU) snapshots, respectively). The experimental NOEs would thus indicate some rather atypical RNA backbone conformations which might also indicate some uncertainty in the experimental data caused by, e.g., possible neglect of H5’ to H5’’ spin diffusion. Those signals provided higher χ^2 values (for all simulations of the particular TN sequence) and thus, lowered the agreement between simulations and 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, assuming Gaussian experimental errors as explained by Bottaro et al.[31]. The uNOEs were also defined as introduced in Ref. [31] with experimental details provided in their Supporting Information. ^{3}J scalar couplings were obtained from the Karplus relationships (see Supporting Information of Ref. [31] for the details and experimental datasets).

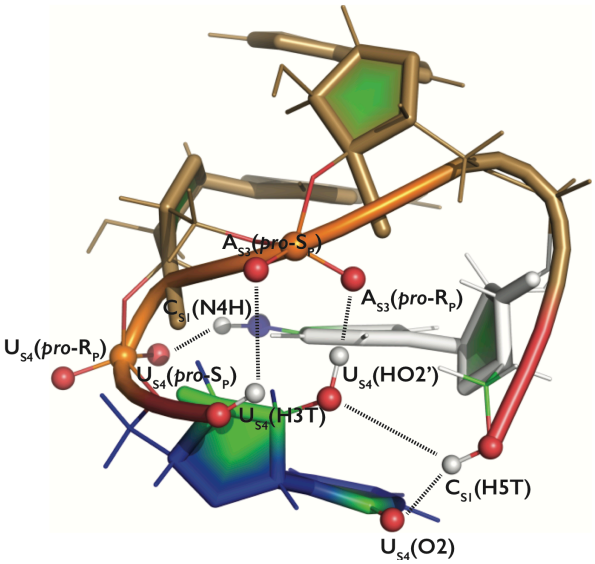

Figure 3: Tertiary structure of r(CAAU) TN in the compact intercalated state. All groups responsible for spurious H-bonds (black lines) forming BPh, SPh and sugar-base interactions are labelled and highlighted by spheres. Nucleotides are colored as in Figure 2.

RESULTS AND DISCUSSION

In our preceding work we demonstrated that performance of the basic OL3 AMBER RNA *ff*[13, 15-17, 43] is improved by using the gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ variant,[12] abbreviated in the present paper as gHBfix19, see Table S1. The gHBfix19 version weakens all SPh H-bonds by 0.5 kcal/mol and supports all base-base –NH…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.[23-25] The application of basic gHBfix19 potential leads to 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 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 intercalated structures, i.e., r(CAAU), r(AAAA), and r(CCCC), possess C or A as the 5′-terminal nucleoside. These residues can form type-7 BPh (7BPh) interaction[50] 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, because native BPh interactions are common in many folded RNA structures,[50] tuning of these interactions was not attempted in Ref. [12].

Here, we designed various gHBfixes in order to weaken BPh interactions and tested them in simulations of r(CAAU) TN (Table 2). The best result 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 BPh interactions is promising since their large weakening could destabilize structures of folded RNAs, where BPh contacts are common and highly conserved.[50] The results are in full details described in the Supporting Information.

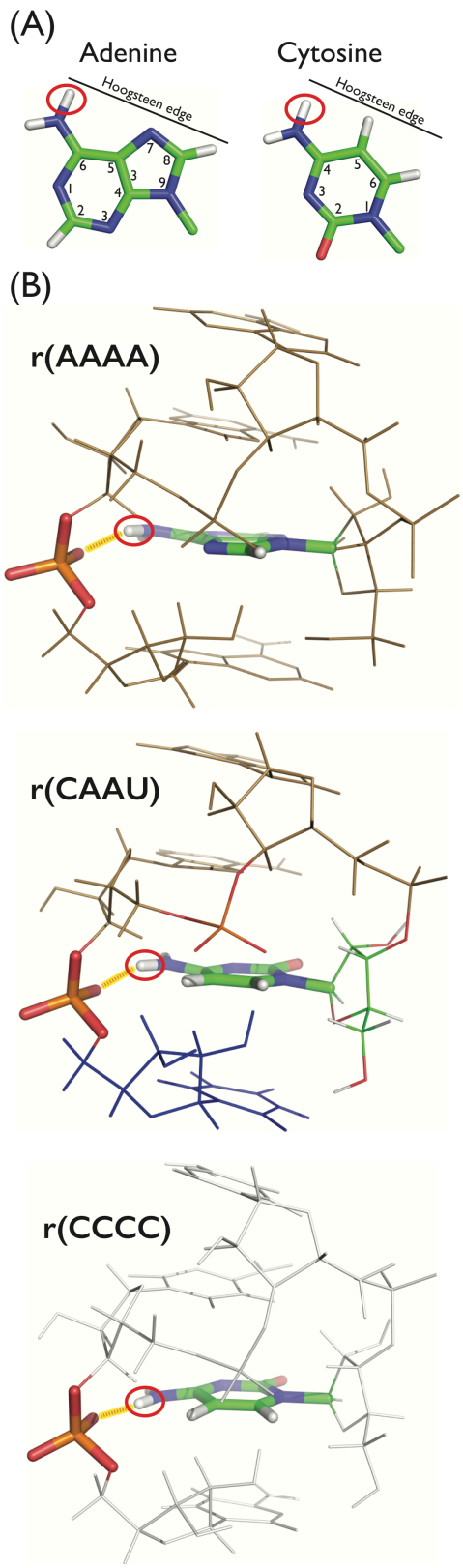

Figure 4: The spurious intercalated state is highly populated during MD simulations of TNs containing C or A 5′-terminal nucleoside. (A) The 7BPh interaction[50] is formed by 5'-terminal nucleosides that possess the amino group at the Hoogsteen edge, i.e., N6H and N4H for A and C,

respectively (red circle). (B) The detail of 7BPh interaction[50] in r(AAAA), r(CAAU), and r(CCCC) TNs. The spurious H-bond between amino group and nbOs of 3'-terminal nucleotide 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 performed standard simulations of Sarcin-Ricin loop motif and T-loop motif. These simulations (see the Supporting Information for full details) did not indicate any undesired weakening of BPh interactions. 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, where all BPh contacts are weakened by the gHBfix *ff* term.

Weakening of specific interactions formed by 5'- and 3'- 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-noted 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 terminal residues. This should allow to tune simulations of TNs while not introducing side effects for other systems. 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 5'-terminal nucleoside (either C_{S1} or A_{S1}). It forms 7BPh interaction[50] (Figure 4) with nbOs of 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 residues.

We designed the gHBfix19 + tHBfix(NH…nbO) variant (Table 1), which is composed of the basic gHBfix19[12] extended by tHBfix penalizing interactions between $-NH_2$ group from either C_{S1}/A_{S1} (5'-terminal nucleoside) and both nbOs of the 3'-terminal nucleotide by 1.0 kcal/mol (Table S1). We firstly tested its effect on r(CAAU) TN and observed that the agreement with experimental data is improved. The 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 (Table 2). However, population of intercalated structures is still significant (~16%, Table 2). Thus, the gHBfix19 + tHBfix(NH…nbO) potential leads to clear improvement, but the extra 1.0 kcal/mol penalty to single H-bond contact is still not sufficient to fully eliminate intercalation of r(CAAU) TN.

We also probed effect of extending the 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'-terminal nucleoside) and nbOs from all nucleotides of the r(CAAU) TN. 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 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'-terminal nucleoside and nbOs from phosphates of different internal nucleotides might occur.

Table 1. List of terminal groups and atoms from RNA nucleotides whose interactions were modified by different tHBfix variants.[a]

Label	Donors [b]					Acceptors [c]		
	5'-terminal nucleoside		3'-terminal nucleotide		Nucleotide			Nucleotide
	Type	Res. # (atoms) [d]	Type	Res. # (atoms)		Type	Res. # (atoms)	
tHBfix(NH…nbO)	base ($–NH_2$)	S1 (H61, H62, H41, H42)			A, C	phosphate (nbO's)	S4 (*pro*-R_P, *pro*-S_P) [e]	all
tHBfix(OH…nbO)			sugar (–OH)	S4 (HO'2, H3T)	all	phosphate (nbO's)	S3 (*pro*-R_P, *pro*-S_P)	all
	sugar (–OH)	S1 (H5T)			all	phosphate (nbO's)	S3 (*pro*-R_P, *pro*-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 [f]	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 (*pro*-R_P, *pro*-S_P), S3 (*pro*-R_P, *pro*-S_P), S4 (O2), S1 (O5')	all, U/C for S4 (O2)

[a] see Supporting Information for description of particular tHBfix versions by using the detailed notation introduced in the original gHBfix paper (Table S1)[12] and for one by one list of all specific interactions that were modified by the tHBfix20 potential (Table S2).

[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] alternative version can be used that includes all phosphates

[f] combination of the four tHBfixes defined in the above rows, i.e., only those particular combinations of interactions listed in 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 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 behavior of terminal –OH groups during MD simulations. Since the terminal –OH groups contain specific parameters within current *ff*s, 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 (an auxiliary A-RNA conformation accompanying the major A-RNA form)[23], as discussed previously.[12] Thus, we designed the gHBfix19 + tHBfix20 variant for the r(CAAU) sequence, where the tHBfix20 potential is 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 lowest r(CAAU) χ^2 value so far (2.05, see Table 2). Thus, structural dynamics of the r(CAAU) during REST2 simulations is significantly improved and population of intercalated structures is reasonably low (~5%, Table 2).

The gHBfix19 + tHBfix20 potential improves the structural dynamics of other TNs. We have carried out REST2 simulations with the gHBfix19 + tHBfix20 potential for the remaining four TNs. The newly designed potential significantly improves structural dynamics of the r(CCCC) TN (χ^2 value of 0.55, Table 2), since agreement between all calculated and experimental observables, i.e., ^{3}J-backbone couplings, ^{3}J-sugar couplings, NOE and uNOE signals, is better (Table 2). The huge decrease of χ^2 value coming from uNOE signals correlates with the decreased population of spurious intercalated structure (from ~37% in simulation with the basic gHBfix19[12] to ~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 basic gHBfix19[12] (0.40, Table 2). It is not possible to decide, which simulation agrees better with the experiment since all simulations with χ^2 value below 1 are satisfactory (see Methods).

The results for 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 identify source of the difference. The χ^2 value of 1.81 for r(UUUU) TN is even marginally higher than the one from simulation with basic gHBfix19 (1.63).[12] Conformational ensembles of r(AAAA) and r(UUUU) REST2 simulations show that the major clusters are different from canonical A-RNA major/minor conformers. The r(UUUU) TN is expected to be highly dynamical and to sample diverse conformations far from canonical A-RNA.[25] 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 adopted *syn* conformation of χ dihedral (Figure S5 in the Supporting Information). The 5'-terminal A_{S1} nucleotide 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,[51] it is possible that the *syn* state is overpopulated. Interestingly, in Ref. [28] combination of MD simulations with NMR data for nucleotides and dinucleotides suggested the need for increasing stabilization of *anti* conformation for A nucleotides in the *ff*. Similar indication was obtained from another fitting based on the same experimental data as used here[52] (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 imbalanced description of stacking interactions. Stacking is suspected to

be overestimated within current RNA *ff*s,[53-55] albeit the OPC water model should partly reduce such overestimation.[20, 56] This second possibility is potentially supported by base stacking analysis which is presented in the Supporting Information.

gHBfix and tHBfix combined with different solvent models. The OPC water model[18] is widely used for enhanced-sampling simulations of simple RNA motifs[10, 12, 27-29, 31, 57] because it somewhat reduces the excessive formation of SPh contacts,[20] especially in combination with the revised phosphate parameters.[17] On the other hand, the OPC water is suboptimal for G-quadruplex simulations.[58-59] It indicates that OPC solvent may not represent a universal solution for simulations of nucleic acids, since desirable weakening of spurious interactions can be accompanied by undesirable destabilization of native interactions. Thus, we have investigated effects of the gHBfix correction to the RNA *ff* in combination with TIP3P[39, 60], SPC/E[40, 61], and TIP4P-D[41] solvent models (Table 2). The results (see Supporting Information) suggest that three-point water models increase population of loop-like and/or intercalated states. 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 settings, however, would have to be different from those used with OPC.

Possibility to reproduce the effect of combined gHBfix19 and tHBfix20 potential by NBfix approach. In principle, similar effect as obtained by HBfix-type of potentials could be achieved by modification of pairwise vdW parameters via breakage of the combination (mixing) rules, i.e., the so-called nonbonded fix (NBfix) approach.[62] However, as we discussed earlier,[12] finding HBfix-type parameters is much more straightforward and brings some advantages. To tune all possible H-bonds by NBfix 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,[12] the NBfix approach affects both interaction energy and optimal geometry, so its applicability may be limited by undesired geometrical changes. Thus, HBfix terms (e.g., gHBfix, tHBfix) are more flexible for fine-tuning with respect to the NBfix approach in pair-specific treatment of H-bonds and interatomic contacts. In practice, NBfix parameters could be derived *a posteriori* after finding a suitable gHBfix version. In the preceding study we have provided NBfix parameters that have similar effect as the gHBfix19 potential.[12] In the present work we also derive NBfix parameters to reproduce the interaction energy curves obtained with the gHBfix19 and tHBfix20 terms in the same way as in Ref. [12]. Details are provided in the Supporting Information together with NBfix parameters which can be used in simulations as an alternative to the gHBfix19 + tHBfix20 parameter set.

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 dihedral potentials, improvement of RNA simulations requires also adjustments of nonbonded terms.[10, 12, 20-22, 25, 29, 31, 33, 52] 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 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. Therefore, RNA TNs can be straightforwardly used to tune RNA *ffs* by comparing conformational ensembles from simulations with NMR experimental data, although one should always keep in mind that there is a 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.[12]

The gHBfix19 visibly improved 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 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 *ff*s. The terminal nucleotides are more exposed to solvent than internal nucleotides. Because the strength of intramolecular H-bonds depends on the difference between direct H-bonds and hydration, it is genuine to treat interactions of terminal nucleotides differently from internal ones. Finally, MD simulations of TNs are significantly affected by interactions involving terminal nucleotides while such interactions often have insignificant effect on simulations of larger RNAs. Thus, considering the empirical nature of the *ff*, it is meaningful to not sacrifice transferability of parameters of internal nucleotides in order to perfect simulations of the short TNs.

We applied enhanced-sampling methods on testing set containing five RNA TNs and compared 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 resolution limits, with the experiments. The r(AAAA) and r(UUUU) TNs display a higher complexity with a mixture of structures. 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 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 profound sensitivity of TN simulations to water models, the gHBfix19 + tHBfix20 parameter set is rather specific for the OPC water model. It might be possible to find in future better gHBfix19 and tHBfix20 parameters using finer parameter optimization by reweighting[52]). However, reweighting would most likely require adding additional different molecules into the training set, as we discuss in more detail in Supporting Information. We also present set of NBfix parameters that should provide similar results as the gHBfix19 + tHBfix20 refinement.

In conclusion, we show that targeting 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*. Tuning properties of terminal nucleotides should not have side effects when simulating larger RNA motifs. We also propose that specific modifications of H-bonds formed by terminal nucleotides should 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] Ultimately, it should improve modeling of those biomacromolecular processes where single stranded ends of RNAs are involved. 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 non-bonded terms, pairing 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 framework of one computational study.

ASSOCIATED CONTENT

Supporting Information. The following file is available free of charge and contains detailed description of MD protocol, comparison of REST2 simulations at different temperature, details about weakening of BPh interactions, results from base stacking analysis, details about usage of gHBfix and tHBfix potentials with various solvent models, comments about reweighting procedure and application of NBfix terms, 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 Czech Science Foundation (by 20-16554S to V.M. and J.S. at IBP and 18-25349S to P.B. and P.K. at UPOL) and 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.).

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 [^{3}J-backbone, ^{3}J-sugar, NOE, uNOE (# of violations) / **Total**] [c]	Clustering [%] [A-form / Loop / Intercalated]	# of clusters	εRMSD populations [%] [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 Supporting Information for the detailed definition of the tested variants of the gHBfix and tHBfix potentials and Table S2 for one by one list of interactions modified by tHBfix20.

[b] 0.15 M KCl, JC ion parameters.[38]

[c] χ^2 values were obtained by comparing calculated and experimental backbone ^{3}J scalar couplings, sugar ^{3}J 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.[42]

[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 –NH_2 group from C_{S1} (5'-terminal nucleoside) and nbO atoms of *all phosphates*, i.e., not only nbOs from the terminal U_{S4} 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) 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*.
(3) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr Opin Struct Biol* **2018,** *49*, 129-138.
(4) 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.
(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*, 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*, 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., et al. 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*, 956-972.
(8) Liu, C.; Piquemal, J. P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J Chem Theory Comput* **2019,** *15*, 4122-4139.
(9) Baker, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdiscip Rev Comput Mol Sci* **2015,** *5*, 241-254.
(10) Kuhrova, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banas, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J Chem Theory Comput* **2016,** *12*, 4534-4548.
(11) 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*.
(12) 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.
(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*, 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*, 1049-1074.
(15) 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.
(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*, 2886-2902.
(17) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput* **2012,** *8*, 4405-4412.
(18) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J Phys Chem Lett* **2014,** *5*, 3863-3871.
(19) Morris, K. V.; Mattick, J. S. The rise of regulatory RNA. *Nat Rev Genet* **2014,** *15*, 423-437.
(20) Bergonzo, C.; Cheatham, T. E. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J Chem Theory Comput* **2015,** *11*, 3969-3972.
(21) 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.
(22) 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.
(23) 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.
(24) 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.
(25) 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.
(26) 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.
(27) Gil-Ley, A.; Bottaro, S.; Bussi, G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J Chem Theory Comput* **2016,** *12*, 2790-2798.
(28) 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*, 6192-6200.
(29) 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*, 395-399.
(30) Schrodt, M. V.; Andrews, C. T.; Elcock, A. H. Large-Scale Analysis of 48 DNA and 48 RNA Tetranucleotides Studied by 1 mus Explicit-Solvent Molecular Dynamics Simulations. *J Chem Theory Comput* **2015,** *11*, 5906-5917.
(31) Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Sci Adv* **2018,** *4*, eaar8521.
(32) 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.
(33) 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.

(34) 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**.
(35) 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*, 1668-1688.
(36) 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., et al. *AMBER 2016*. San Francisco, 2016.
(37) 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*, 4220-4229.
(38) 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.
(39) 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.
(40) 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.
(41) 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.
(42) 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.
(43) 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*, 11305-11305.
(44) 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*, 1864-1874.
(45) Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014,** *344*, 1492-1496.
(46) Bottaro, S.; Di Palma, F.; Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res*. **2014,** *42*, 13306-13314.
(47) Hall, P.; Horowitz, J. A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions. *Ann Stat* **2013,** *41*, 1892-1921.
(48) https://www.ilikebigbits.com/2015_03_04_plane_from_points.html.
(49) Jarvis, R. A. On the identification of the convex hull of a finite set of points in the plane. *Inf Process Lett* **1973,** *2*, 18-21.
(50) 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*, 4898-4918.
(51) 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*, 4499-4512.

(52) 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.
(53) Sponer, J.; Riley, K. E.; Hobza, P. Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys Chem Chem Phys* **2008,** *10*, 2595-2610.
(54) 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*, 10476-10493.
(55) 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*, 95-115.
(56) Hase, F.; Zacharias, M. Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Res*. **2016,** *44*, 7100-7108.
(57) 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*, 2115-2125.
(58) 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.
(59) 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. *J Chem Theory Comput* **2020,** *asap*.
(60) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J Chem Phys* **1996,** *105*, 1902-1921.
(61) 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.
(62) 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*, 8432-8449.