IMPROVING THE PERFORMANCE OF THE RNA AMBER FORCE FIELD BY TUNING THE HYDROGEN-BONDING INTERACTIONS

Petra Kührová,[1,#] Vojtěch Mlýnský,[2,#] Marie Zgarbová,[1] Miroslav Krepl,[2] Giovanni Bussi,[3] Robert B. Best,[4] Michal Otyepka,[1] Jiří Šponer,[1,2] and Pavel Banáš[1,2*]*

[1] 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

[2] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic

[3] Scuola Internazionale Superiore di Studi Avanzati, SISSA, via Bonomea 265, 34136 Trieste, Italy

[4] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

[#] authors contributed equally

* corresponding authors email: sponer@ncbr.muni.cz and pavel.banas@upol.cz

ABSTRACT

Molecular dynamics (MD) simulations became a leading tool for investigation of structural dynamics of nucleic acids. Despite recent efforts to improve the empirical potentials (force fields, *ffs*), RNA *ffs* have persisting deficiencies, which hamper their utilization in quantitatively accurate simulations. Previous studies have shown that at least two salient problems contribute to difficulties in description of free-energy landscapes of small RNA motifs: (i) excessive stabilization of the unfolded single-stranded RNA ensemble by intramolecular base-phosphate and sugar-phosphate interactions, and (ii) destabilization of the native folded state by underestimation of stability of base pairing. Here, we introduce a general *ff* term (gHBfix) that can selectively fine-tune non-bonding interaction terms in RNA *ffs*, in particular the H-bonds. This potential affects the pair-wise interactions between all possible pairs of the specific atom types, while all other interactions remain intact, i.e., it is not a structure-based model. In order to probe the ability of gHBfix potential to refine the *ff* non-bonded terms, we performed an extensive set of folding simulations of RNA tetranucleotides and tetraloops. Based on these data we suggest particular gHBfix terms that can significantly improve the agreement between experimental data and the conformational ensembles estimated by the AMBER RNA *ff,* although the currently available version still remains far from being flawless. While attempts to tune the RNA *ffs* by conventional reparametrizations of dihedral potentials or non-bonded terms can lead to undesired side effects as we demonstrate for some recently published *ffs*, gHBfix has a clear promising potential to improve the *ff* performance while avoiding introduction of major new imbalances.

KEYWORDS

RNA, force field, MD simulation, enhanced sampling, folding, tetraloop.

INTRODUCTION

Molecular dynamics (MD) simulations have become a very important tool for studies of biomolecular systems such as nucleic acids with routine access to micro- or even millisecond timescales.[1-7] MD simulations are often instrumental for understanding and clarifying experimental results and for obtaining a more complete picture of their biological implications. Nevertheless, a sufficiently realistic description of biopolymers by the used empirical potentials (force fields, *ffs*) is essential for successful applications of MD.[4, 8] The polyanionic nature and high structural variability of ribonucleic acid (RNA) makes the development of RNA *ffs* an especially daunting task.[4] Despite huge efforts to fix problems that have emerged on ns-μs simulation timescales, RNA *ffs* still cause some behaviors in simulations which are not consistent with experiment.[4]

Limitations of the available RNA *ffs* have been reviewed in detail, with a suggestion that the currently available pair-additive RNA *ffs* are approaching the limits of their applicability.[4] Examples of such problems for some recently suggested *ff* versions are documented in the present study. A radical solution of the RNA *ff* problem could be the use of polarizable force fields,[9-10] but this would require a completely new and sophisticated parametrization, which is a very challenging goal. Another option is to augment the existing *ff* forms by some additional simple *ff* terms that could be used to tune the *ff* performance while minimizing adverse side effects. One such *ff* term is introduced in this study.

Compact folded RNA molecules are typically well-described by modern *ff*s on a sub-μs timescale when starting simulations from established experimental structures.[4] This has allowed many insightful studies on RNAs and protein-RNA complexes. However, the stability of folded RNAs on longer timescales is affected by the free-energy balance between folded, misfolded and unfolded states. Therefore, a common way to identify major problems in *ff*s is using enhanced sampling techniques, where temperature replica exchange molecular dynamics (T-REMD),[11] replica exchange with solute tempering (REST2),[12] and well-tempered metadynamics (MetaD)[13-14] are among the most popular. Replica exchange simulations profit from multiple loosely coupled simulations running in parallel over a range of different temperatures (T-REMD simulations) or Hamiltonians (e.g. REST2 simulations). Exchanges between replicas are attempted at regular time intervals and accepted conforming to a Metropolis-style algorithm. RE methods do not require any prior chemical insights regarding the folding landscapes. MetaD is a method based on a history-dependent biasing potential acting on a few degrees of freedom called collective variables (CVs).[13] The CVs exploit prior chemical information and their choice may have significant impact on the outcome of all CV-based computations. The benefits of enhanced sampling in *ff* analyses are crucial, because the unbiased MD simulations remain typically trapped in a free-energy basin corresponding to the starting structure, preventing identification of the global free-energy minimum and characterization of the free-energy balance between folded, misfolded and unfolded states. Obviously, even enhanced-sampling simulations are not a panacea and their capability to accelerate sampling has some limits; see Refs. [4] and [15] summarizing recent applications of enhanced sampling methods to RNA systems.

The structural dynamics of RNA tetranucleotides (TNs) represents one of the key benchmarks for testing RNA *ff*s.[16-26] TNs are ideal testing systems due to their small size and straightforward

comparison of their simulations with solution experiments. 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 it appears that contemporary simulation methods and hardware already allow to obtain sufficiently converged simulation ensembles for TNs.[19-22, 24-25, 27-28] TN simulations can specifically evaluate performance of *ff*s for several salient energy contributions, namely, (i) sugar-phosphate (SPh) and base-phosphate (BPh) interactions, (ii) base stacking interactions, (iii) backbone conformations and (iv) balance of these contributions with solvation. Experimental data revealed that TNs mostly populate A-form conformations[16-18] but MD simulations tend to significantly sample also non-native intercalated structures (or some other non-native structures, Figure 1) that are considered to be a *ff* artifact.[18-25, 29] Obviously, when using TNs as a benchmark for *ff* development, one has to be concerned about a possible over-fitting of the *ff* towards the canonical A-RNA conformation, which may have detrimental consequences for simulations of folded RNAs (see below).

Another key energy contribution that needs to be described by a *ff* are base-pair interactions. A large part of the thermodynamics stability of folded RNAs is due to formation of canonical A-RNA helices stabilized by canonical Adenine-Uracil (AU) and Guanine-Cytosine (GC) Watson-Crick base pairs and their stacking interactions.[30] Thus, their correct description is fundamental for folding studies of RNA molecules. Besides this, RNA molecules contain an astonishing variability of other base-pair patterns.[4, 31] Assessment of the capability of the *ffs* to describe these interactions requires analyses of more complicated systems like RNA tetraloops (TLs) and folded RNA motifs.[4] The GNRA and UNCG TLs (N and R stand for any and purine nucleotide, respectively) are the most abundant hairpin loops in RNA molecules.[32-33] These TLs contribute

to various biological functions including tertiary folding, RNA-RNA and protein-RNA interactions, ligand binding, and thus play key roles in transcription, translation and gene regulation.[34-40] Both above-noted TL types possess a clear dominant folded topology that is characterized by a set of signature molecular interactions that determine the consensus sequences (Figure 1).[33] The structure of isolated small TL motifs in solution is in dynamic temperature-dependent equilibrium between folded and unfolded conformations. The two main parts of TL's, i.e., their A-RNA stem and the structured loop itself, contribute differently to the folding free-energy landscape, and thus simulations of TLs allow us to simultaneously probe the capability of the *ffs* to describe A-RNA duplexes as well as some non-canonical interactions and backbone conformations.[41-43] In contrast to dimerization of A-form duplexes, sampling of the hairpin free-energy landscape is simplified due to unimolecular nature of its folding. TL's with longer stems formed by canonical base pairs are more stable and thus require higher temperature for melting (unfolding).[42-43] Therefore, minimal 8-nucleotide long (8-mers) TL motifs with just two base pairs are the preferred targets for computational studies due to their small size and relative 'metastability', which is further affected by the nature and orientation of both the closing and terminal base pairs.[42] Nevertheless, there is a clear experimental evidence that even these short oligomers should be dominantly in the folded state.[44] The native TL conformations thus represent a genuine and unambiguous benchmark for the simulation methodology.

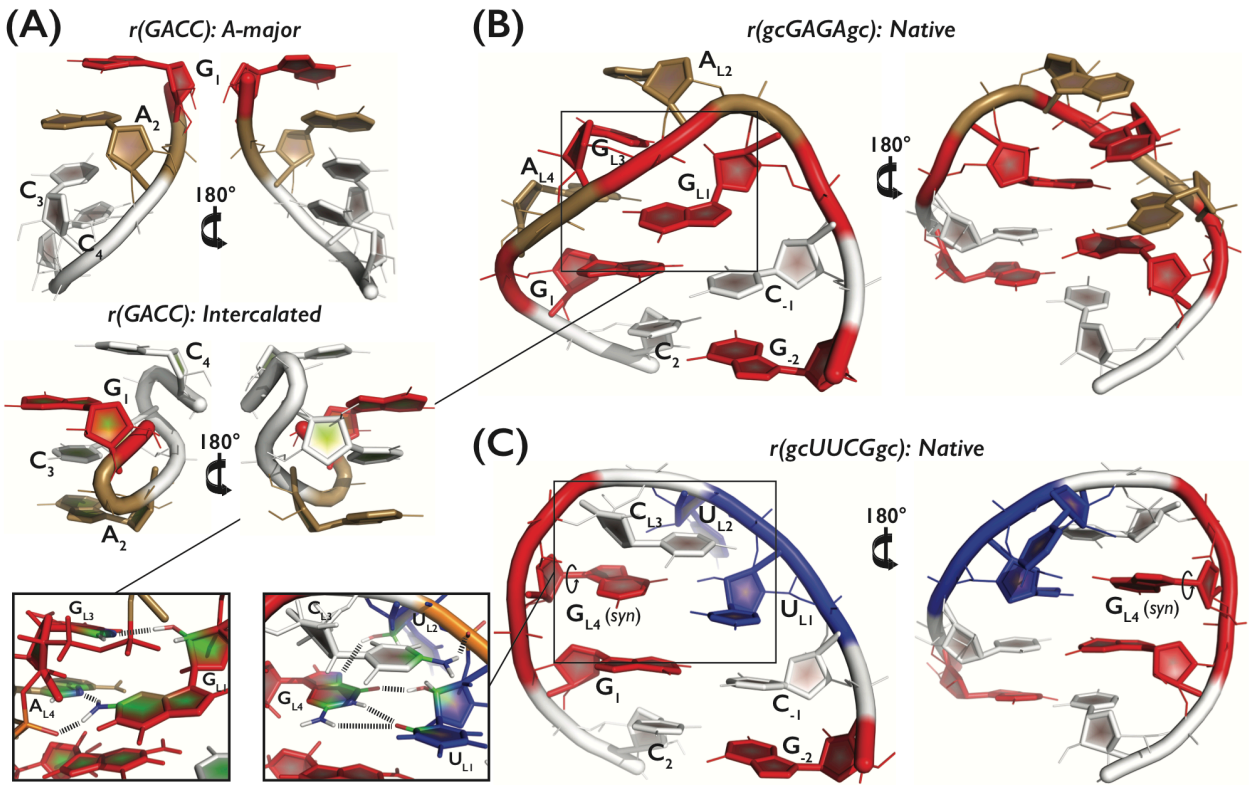

Figure 1. Tertiary structures and detail overview of the three systems, i.e., (A) r(GACC) TN in the dominant native A-major conformation (top) and the spurious intercalated structure (bottom), and (B) r(gcGAGAgc) and (C) r(gcUUCGgc) TLs in their native conformations. A, C, G and U nucleotides are colored in sand, white, red, and blue, respectively. The two insets (bottom left) highlight three and five signature H-bonds, i.e., G_{L1}(N2H)…A_{L4}(*pro*-R_P), G_{L1}(N2H)…A_{L4}(N7), and G_{L1}(2'-OH)…G_{L3}(N7) for GNRA TL (B) and U_{L1}(2'-OH)…G_{L4}(O6), U_{L2}(2'-OH)…G_{L4}(N7), C_{L3}(N4H)…U_{L2}(*pro*-R_P), and bifurcated G_{L4}(N1H/N2H)…U_{L1}(O2) for UNCG (C) TL, respectively.

Recent studies generated large conformational ensembles of TNs[19-22, 24-25, 27-28] and TLs[19-22, 24-25, 29, 45-48] in order to assess the performance of RNA *ff*s. They showed that the available RNA *ff*s have persisting deficiencies causing, e.g., (i) shifts of the backbone dihedral angles to nonnative

values, (ii) problems with the χ-dihedral angle distribution, and (iii) over-populated SPh and BPh hydrogen bond interactions.[16, 20, 47] Some of those studies also suggested potential directions for *ff* improvement, which included modifications of backbone dihedral terms, van der Waals radii and charges, RNA interaction with solvent/ions (adjustments of Lennard-Jones parameters typically accompanied by modifications of the Lennard-Jones combining rules via nonbonded fix, NBfix, to balance the RNA-solvent interaction) and enforced distributions from solution experiments.[19, 22-25, 45, 49-52] Computer folding of UNCG TLs appears to be much more challenging than folding of GNRA TLs and description of TNs,[20, 29, 45, 47, 53] as for the latter two systems some partial successes have been reported. Note that there have been repeated past claims in the literature about successful folding of RNA TLs in simulations. However, these were not confirmed by independent research groups, as extensively reviewed in Ref. [4]. The performance and possible shortcomings of another recently released *ff*[52] are discussed as part of this work.

Previously, we have shown that at least two different imbalances likely contribute to the (in)correct folded/unfolded free-energy balance of TNs and TLs. The first problem was excessive stabilization of the unfolded ssRNA structure by intramolecular BPh and SPh interactions.[54] The excessive binding of 2'-hydroxyl groups (2'-OH) towards phosphate nonbridging oxygens was reported earlier[18, 50] and could be partially reduced by using alternative phosphate oxygen parameters developed by Case et al.[55] in combination with the OPC[56] explicit solvent water model.[19] However, the OPC water model was shown to destabilize three-tetrad DNA quadruplex stems[57] while the modified phosphate parameters were not successful in correcting base-phosphate H-bonding in simulations of Neomycin-sensing riboswitch,[58] indicating that the modified phosphate parameters with OPC water model do not represent the

ultimate solution for tuning the phosphate-group interactions. The second problem was destabilization of the native folded state by underestimation of the native H-bonds including the stem base pairing. As a general correction for under- or overestimation of base-base and other H-bond interactions remains challenging,[4] we recently introduced a weak local structure-specific short-range biasing potential supporting the native H-bonds (HBfix). Its application led to a substantial improvement of GAGA TL folding[54] and stabilization of U1A protein-RNA interface.[59] The HBfix approach, however, does not provide a transferable *ff* that can be applied on systems, where the native structure is not known *a priori*.

In this work, we introduce a generalized formulation of the HBfix potential, in order to tune all interactions of the same kind, henceforth labelled as gHBfix. Most importantly, the gHBfix correction can be applied without previous knowledge of the native structure. gHBfix can be used for tunable modification of selected nonbonded terms, namely specified types of H-bond interactions, in order to improve the behavior of current state-of-the-art RNA *ff*. In the preceding applications, HBfix was used as a native-structure-centered *ff* correction to support known native interactions. In contrast, gHBfix is an interaction-specific *ff* correction, whose application depends only on the atom types. It is not biased in favor of any specific fold or RNA sequence. It can be easily applied and does not require any modification of standard simulation codes.[60-61] We used enhanced sampling methods and obtained a large amount of data (total simulations timescale more than 4 ms) during the testing phase. We mainly focused on structural dynamics and folding of TN and TL systems with the aim to generate ensembles with population of major conformers close to those reported by experimental datasets.[42-44] The gHBfix potential was subsequently tested on various important RNA structural motifs using unbiased MD simulations. Although the suggested *ff* modification certainly does not eliminate all the *ff* problems, as

discussed in detail, it brings valuable improvements so far without visible side effects. The primary goal of the present work, however, is to introduce the gHBfix term rather than to provide a fully refined force field. We suggest the methodology has substantial potential for further tuning.

METHODS

Starting structures and simulation setup. The starting structures of r(gcGAGAgc) and r(gcUUCGgc) (in unfolded states) and r(GACC) as well as the other four TNs were prepared using Nucleic Acid Builder of AmberTools14[62] as one strand of an A-form duplex. The topology and coordinates were prepared using the tLEaP module of AMBER 16 program package.[60, 63] Single strands were solvated using a rectangular box of OPC[56] water molecules with a minimum distance between box walls and solute of 10 Å, yielding ~2000 water molecules added and ~40×40×40 Å^3 box size for TN and ~7000 water molecules added and ~65×65×65 Å^3 box size for both TLs, respectively. Simulations were run at ~1 M KCl salt excess using the Joung-Cheatham ion parameters[64] (K^+: r = 1.705 Å, ε = 0.1937 kcal/mol, Cl^-: r = 2.513 Å, ε = 0.0356 kcal/mol). We used the *ff*99bsc0χ_{OL3}[65-68] basic RNA *ff* version with the van der Waals modification of phosphate oxygen developed by Case et al.[55]. These phosphate parameters for phosphorylated aminoacids were shown to improve the performance of RNA *ff*.[19, 54] All the affected dihedrals were adjusted as described elsewhere.[50] AMBER library file of this *ff* version can be found in Supporting Information (SI) of Ref. [54]. For the simplicity, the *ff* version is abbreviated as χ_{OL3} through the text.

REST2 Settings. The replica exchange solute tempering (REST2)[12] simulations of r(GACC) TN, r(gcGAGAgc), and r(gcUUCGgc) TLs were performed at 298 K with 8 (TN) and 12 (TLs)

replicas. Details about settings can be found elsewhere.[54] The scaling factor (λ) values ranged from 1 to 0.6017 (TN) or to 0.59984 (TLs) and were chosen to maintain and exchange rate above 20%. The effective solute temperature ranged from 298 to ~500 K. The hydrogen mass repartitioning[69] with a 4-fs integration time step was used. The length of REST2 simulations was 10 μs per replica (specific tests were terminated earlier; see Table S1 in Supporting Information for summary of all enhanced sampling simulations); the cumulative time of all REST2 simulations was ~4.2 ms.

gHBfix for support/weakening of specific interactions. The main aim of this work is extension of the previously introduced structure-specific HBfix potential to the generalized interaction-specific gHBfix potential (Figure 2), creating essentially new generally applicable RNA *ff* versions. Detailed description of the function form of the locally acting potential function can be found in the Supporting Information. Originally, the potential was used to support native H-bonds in a structure-specific manner that was sufficient to achieve folding of the r(gcGAGAgc) TL.[54] For the sake of completeness, we report equivalent structure-specific HBfix T-REMD folding simulation of r(gcUUCGgc) TL as part of this work. However, the main focus of this study was the development of the gHBfix potential, where all possible interactions between specific groups are affected. We have tested a number of variants and combinations of the gHBfix potential. The list of all RNA groups and atoms involved in the gHBfix is summarized in Tables 1 and 2 (see Table S1 for summary of performed REST2 simulations with particular gHBfix settings). For the clarity, each tested setting of the gHBfix potential is marked as gHBfix $\left(\begin{smallmatrix}\text{interaction pair}\\ \eta\end{smallmatrix}\right)$, where the upper label in the bracket indicates the specific interactions between RNA groups (Table 2) and the η parameter below defines total energy support or penalty for each H-bond interaction of this kind (in kcal/mol, Figure 2).

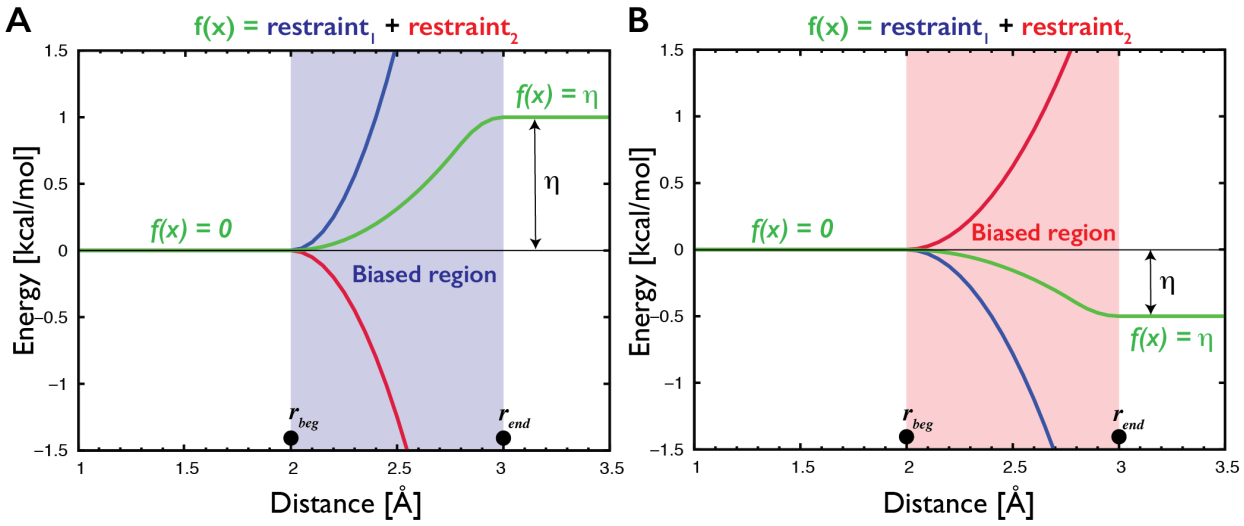

Figure 2. Description of the gHBfix potential (green curves) used for either support (A) or weakening (B) of H-bond interactions. In the present study, gHBfix potential is in all cases applied to the distance between hydrogen of the proton donor and proton acceptor heavy atom. The potential is constant (i.e., it provides zero forces) at all distances except for the narrow region between r_{beg} and r_{end} corresponding roughly to the expected location of the free energy barrier between bound and unbound states. For the used distance between hydrogen and proton acceptor heavy atom of the hydrogen bond we used here 2 Å and 3 Å for r_{beg} and r_{end}, respectively. The potential is formally composed of a combination of two flat-well restraints with opposite sign of curvature and linear extensions that cancel each other at distances above r_{end} (red and blue curves, see Supporting Information for detailed description of the function form). The η parameter defines total energy support (A) or penalty (B) for each H-bond interaction.

Conformational analysis. Native states of r(gcGAGAgc) and r(gcUUCGgc) TLs were determined based on the presence of all native H-bonds, i.e., those participating on the base pairing in the stem as well as signature interactions of the loop (presence of a H-bond was

inferred from a hydrogen-acceptor distance within cutoff 2.5 Å), in combination with the εRMSD metric.[70] The dominant conformations sampled in REST2 simulations were identified using a cluster analysis based on an algorithm introduced by Rodriguez and Laio[71] in combination with the εRMSD,[70] see Ref. [54] for details about implementation of the algorithm. The cluster analysis was performed for unscaled replicas (λ=1, T = 298 K). The error bars were estimated using bootstrapping[72] with resampling both over time- and replica-domains. In particular, the bootstrapping over the replica-domain can highlight the limited sampling in RE methods that is often not reported in literature. RE simulations are converged once all replicas sample the same ensemble. Such ultimate convergence occurs typically on a much longer time-scale than the time-scale needed to reach the steady-state population on the reference replica. Thus, seemingly converged steady-state population of the folded state as revealed by reference replica might still contain some statistical inaccuracy if the folding events were observed only in some replicas (see SI for details). Note that this is the case of the present REST2 simulations of TLs. More information about implementation of bootstrapping and about the mechanism of the cluster analysis can be found in Ref. [54].

Data analysis. All trajectories were analyzed with the PTRAJ module of the AMBER package[60] and the simulations were visualized using a molecular visualization program VMD[73] and PyMOL[74].

Additional simulations

Besides the REST2 simulations of different variants of our gHBfix potential, we report also a number of other simulations for various RNA systems, to check the performance of the modified *ff*. The cumulative time of all unbiased simulations was ~100 μs and for space reasons, methodological details of these simulations are given in Supporting Information.

Simulations with Shaw et al. RNA *ff*.

As part of our study, we have tested the recently published *ff* by D. E. Shaw and co-workers.[52] We have tested this *ff* in folding simulations of RNA TLs with short A-RNA stems as well as in standard simulations of selected RNA motifs, to document side effects caused by this *ff* version in folded RNAs. Methodological details are given in the Supporting Information of the article.

Table 1. The list of groups and atoms from RNA nucleotides whose interactions were modified by the gHBfix.

Donors	Nucleotides			
	A	C	G	U
NH (base)	N6(H61, H62)	N4(H41, H42)	N1(H1), N2(H21, H22)	N3(H3)
2-OH (sugar)	O2'(HO2')	O2'(HO2')	O2'(HO2')	O2'(HO2')
Acceptors				
N (base)	N1, N3, N7	N3	N3, N7	-
O (base)	-	O2	O6	O2, O4
O (sugar)	O4', O2'	O4', O2'	O4', O2'	O4', O2'
bO (phosphate)	O3', O5'	O3', O5'	O3', O5'	O3', O5'
nbO (phosphate)	*pro*-R_P, *pro*-S_P	*pro*-R_P, *pro*-S_P	*pro*-R_P, *pro*-S_P	*pro*-R_P, *pro*-S_P

Table 2. All possible combinations of groups used for various gHBfix settings. See Table 1 for atoms involved within each group. gHBfix combinations not tested in this work are in italics.

gHBfix	Donors	Acceptors
2-OH…nbO	2'-OH (sugar)	nbO (phosphate)
2-OH…bO	2'-OH (sugar)	bO (phosphate)
2-OH…O2	2'-OH (sugar)	O2' (sugar)
2-OH…O4	2'-OH (sugar)	O4' (sugar)
2-OH…N	2'-OH (sugar)	N (base)
2-OH…O	2'-OH (sugar)	O (base)
NH…N	NH (base)	N (base)
NH…O	NH (base)	O (base)
NH…O2	NH (base)	O2' (sugar)
NH…O4	*NH (base)*	*O4' (sugar)*
NH…bO	*NH (base)*	*bO (phosphate)*
NH…nbO	*NH (base)*	*nbO (phosphate)*

RESULTS AND DISCUSSION

In the present study, we have attempted to parameterize the gHBfix correction for the χ_{OL3}[65-66, 68] AMBER RNA *ff*. Our primary training systems were r(GACC) TN and two TL hairpins, namely r(gcGAGAgc) and r(gcUUCGgc) (see Figure 1 for structures). Our goal was to improve performance of the simulations for these systems while avoiding undesired side effects for other RNA systems. We have used extended REST2 simulations to generate converged conformational ensembles of the above small RNA systems while testing various combinations of the gHBfix potentials acting on specific types of H-bonds (see Methods and Table S1 in Supporting Information for overview of REST2 simulations). The obtained results were compared with the available experimental data.[16-18, 42-44] We have found a promising gHBfix correction that decisively improves behavior of the r(GACC) TN and r(gcGAGAgc) TL. We did not detect any side effects so far in standard simulations of a wide range of other RNA structures. This indicates that although the suggested modification is not robust enough to fold the r(gcUUCGgc) TL and is less convincing for some other TNs, it may provide an significant improvement of RNA simulations when added to the widely used χ_{OL3} RNA *ff*. We reiterate that the primary goal of the paper was introduction of the basic principles of the gHBfix potential and demonstration of its capability for improving RNA simulations.

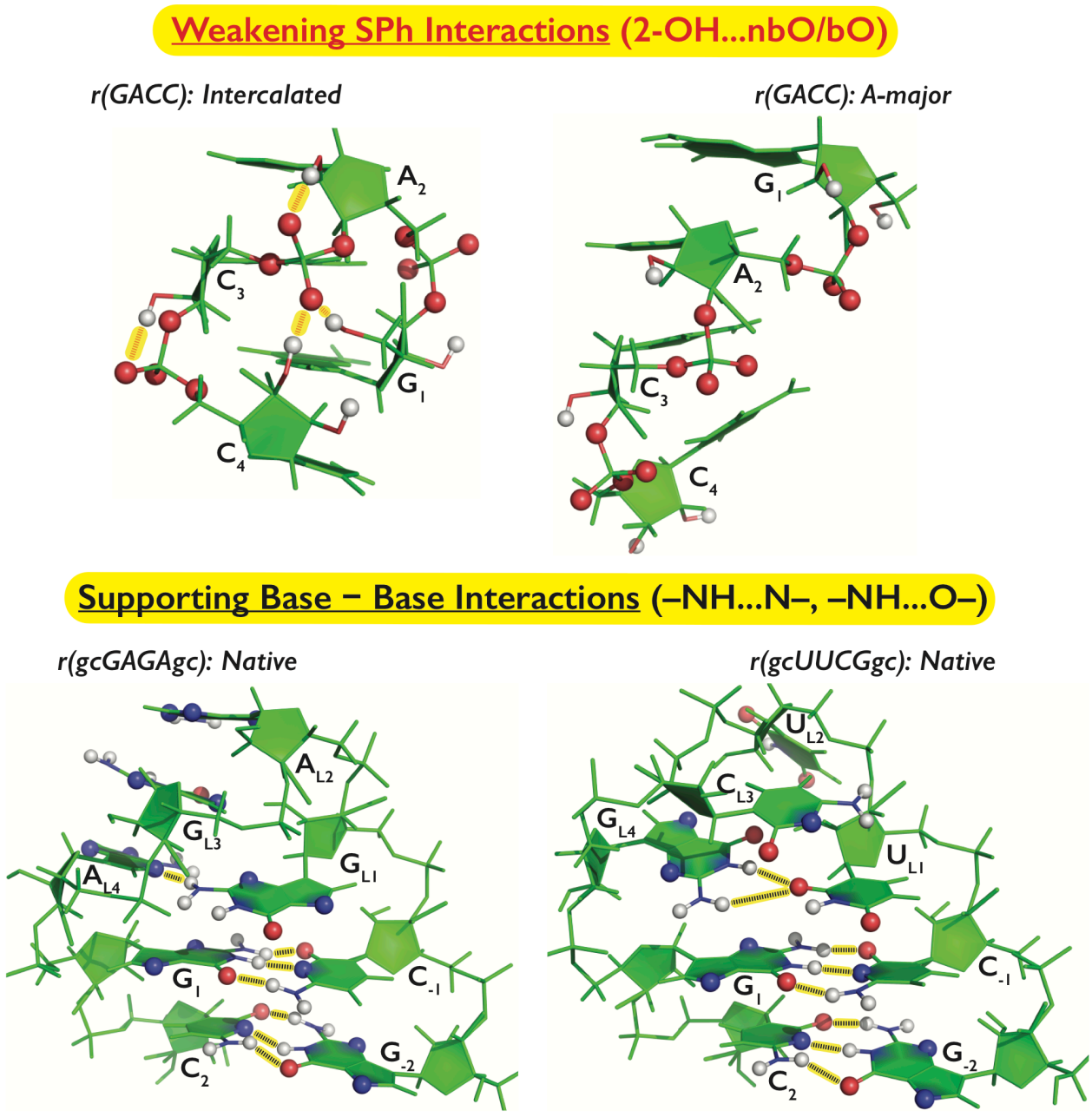

Figure 3. Application of the gHBfix potential and its impact on tertiary structures of the three main test systems, i.e., r(GACC) TN in the spurious intercalated structure (top left) and the dominant native A-major conformation (top right), and r(gcGAGAgc) (bottom left) and r(gcUUCGgc) (bottom right) TLs in their native conformations. Groups included in the gHBfix potential are highlighted in spheres (H, N, and O atoms in white, blue, and red, respectively). The particular version of the gHBfix depicted in the Figure is tuning H-bond interactions in order

to: (i) destabilize sugar – phosphate interactions (red dashed lines highlighted by yellow background on the top left panel) and, simultaneously, (ii) support base – base interactions (black dashed lines highlighted by yellow background in lower panels). Note that the gHBfix potential is affecting all interactions of the same type, i.e., not only those present in the depicted conformations.

Weakening of the over-stabilized SPh interactions. According to the NMR data,[16-18] RNA TNs adopt mostly two different conformations, A-major and A-minor (Figure S1 in Supporting Information). However, recently published standard MD and replica exchange simulations[18-20, 22, 24-25, 27-28] using the χ_{OL3}[65-68] RNA *ff* reported unsatisfactory populations of canonical A-form conformations and, instead, significant population of artificial intercalated structures (Figures 1, 3, and S1 in Supporting Information) that are stabilized by SPh and BPh interactions.[18] Overstabilization of the BPh and especially SPh interactions was repeatedly suggested in our recent studies.[50, 54] Among all TNs for which benchmark experimental date is available,[16-18] i.e., r(GACC), r(CAAU), r(CCCC), r(UUUU), r(AAAA), the r(GACC) sequence was most frequently used in the preceding *ff* testing, therefore, we primarily focused on this system, while the other sequences were tested afterwards.

In an attempt to eliminate the unsatisfactory MD behavior, we first designed a negative gHBfix potential to probe the effect of destabilization of SPh interactions, i.e., H-bonds between all 2'-OH groups and phosphate nbOs (*pro*-R_P, *pro*-S_P), bOs (O3', O5') and sugar O4' oxygens (Figure 3 and Tables 1 and 2). Several variants of the negative gHBfix were tested using a set of REST2 simulations of r(GACC) TN used as the training system. The results from clustering analysis show that all introduced negative gHBfix potentials acting on SPh interactions increased

significantly the native RNA A-major population and essentially eliminated the presence of the intercalated structures (observed populations of artificial intercalated structures were marginal, typically ~1.0 %, Table 3). However, as a side effect, the canonical RNA A-minor conformation (characterized by the change of the α backbone dihedral at the 3'-end, allowing formation of the SPh contact, Figure S1 in Supporting Information) appeared to be destabilized and could even be eliminated when the repulsive gHBfix correction was too strong. For further tests, we took the gHBfix$\binom{2-\mathrm{OH}\cdots\mathrm{nbO/bO}}{-0.5}$ combination, which provides close agreement with experiments,[16] i.e., χ^2 value with respect to NMR observables of ~0.15 and populations of A-major/A-minor/Intercalated structures around 65%/18%/0%, respectively (Table 3). The chosen gHBfix$\binom{2-\mathrm{OH}\cdots\mathrm{nbO/bO}}{-0.5}$ setting consists of the negative gHBfix applied to all possible interactions between 2'-OH groups as proton donors and nbO and bO oxygens as acceptors with the destabilization constant η of –0.5 kcal/mol. Notice that comparable improvement over the control χ_{OL3} simulation was achieved by other tested potentials, especially those involving interactions between 2'-OH groups and nbO oxygens (Table 3), suggesting the possibility of further tuning of the gHBfix parameters for SPh interactions.

Table 3. Analysis of the most populated structural clusters (in %) obtained from r(GACC) REST2 simulations at the reference replica (T = 298 K) with five variants of the gHBfix parameters. [a]

Cluster	Exp.[b]	Ref. [c]	Population (%)				
			gHBfix $\binom{2-\mathrm{OH}\cdots\mathrm{nbO}}{-1.0}$	gHBfix $\binom{2-\mathrm{OH}\cdots\mathrm{nbO/bO}}{-1.0}$	gHBfix $\binom{2-\mathrm{OH}\cdots\mathrm{nbO/bO/O4}}{-1.0}$	gHBfix $\binom{2-\mathrm{OH}\cdots\mathrm{nbO}}{-0.5}$	gHBfix $\binom{2-\mathrm{OH}\cdots\mathrm{nbO/bO}}{-0.5}$
A-major	~65	36.9 ± 2.8	69.2 ± 0.5	76.5 ± 3.2	61.3 ± 4.3	65.1 ± 2.5	58.7 ± 2.1
A-minor	~18	14.4 ± 1.5	4.4 ± 0.5	3.7 ± 0.4	2.8 ± 0.5	7.7± 0.9	7.4 ±1.3
Intercalated [d]	0	~9	~0	~0	~0	~1	~1
χ^2 [e]	-	0.75	0.14	0.18	1.73	0.13	0.15

[a] Clustering was performed from the last 7 out of 10 μs-long trajectory.

[b] Experimental populations obtained by NMR refinement simulations[16] and by reweighting extensive MD simulations based on these NMR data.[27]
[c] Present control χ_{OL3} simulation without any external potential (see Methods for the detailed setup).
[d] See Figures 1, 2 and S1 in Supporting Information for examples of intercalated structures.
[e] χ^2 values were obtained as described elsewhere[27] by calculating and comparing backbone ^{3}J scalar couplings, sugar ^{3}J scalar couplings, nuclear Overhauser effect (NOE) intensities, and also using the absence of specific peaks in NOE spectroscopy data.

Stabilization of the base-base H-bonding interactions. Short TL motifs are ideal model systems for assessing the folding capability of *ff*s due to their small size, clearly defined native conformation and the possible variety of other competing conformations. We initially applied the gHBfix$\left(^{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}_{-0.5}\right)$ potential in r(gcGAGAgc) REST2 folding simulation but we did not obtain native states, i.e., neither the stem nor the loop sampled native conformations for the entire 10 μs of the REST2 simulation (Table 4). Nonetheless, the gHBfix$\left(^{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}_{-0.5}\right)$ was able to (at least partially) suppress condensed ('random-coil') states with excess of BPh and SPh interactions in favor of A-form single stranded structures (see Figure S1 in Supporting Information for geometries and populations of major clusters during r(gcGAGAgc) REST2 simulations). The persisting inability to sample the native state was not surprising because the problem with TL folding in χ_{OL3} RNA *ff* was deemed to be connected with two separate problems, namely, spurious stabilization of unfolded states by excessive BPh and SPh interactions and underestimation of the native H-bonds within the stem (canonical base-pairing interactions) and the loop (signature BPh and other interactions involving the 2'-OH groups, Figure 3).[47, 54] Recently, we have shown that addition of structure-specific HBfix potential supporting native H-bonds results in satisfactory GAGA TL folding.[54] This suggests that the base-pairing interactions should be additionally stabilized, which is also consistent with several recent studies on DNA and RNA guanine quadruplexes and their folding intermediates stabilized by GG Hoogsteen base pairs.[75-77] In addition, the usefulness of stabilization of base – base H-

bonds is also indicated by potentially excessive fraying of base-paired segments occasionally seen in MD simulations of folded RNAs.[4]

In an attempt to fix the base-base interactions in a general fashion, i.e., to support both canonical and non-canonical base pairing, we added additional terms to the gHBfix potential, which support H-bonds between any proton donor from nucleobases (–NH groups) and nucleobase proton acceptors (N/O– atoms, see Figures 2, 3 and Methods for details). We would like to reiterate that these gHBfix potentials are biased neither towards any specific type of base pairing[78] nor towards any specific pairs of nucleobases in the sequence, so they work irrespective to the base pairing in the native state. They are as general as all other *ff* terms and can be in principle transferred to other systems.

We have tested several variants of the base–base gHBfix function, by using a combination of separate terms for (–NH…N–) and (–NH…O–) H-bonds, each with three different values of the gHBfix bias constant η, namely 0.0, +0.5, and +1.0 kcal/mol. Thus, we performed nine r(gcGAGAgc) TL REST2 simulations with the gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}$ term introduced above for SPh interactions and various combinations of the gHBfix $\binom{\mathrm{NH\ldots N}}{\eta}$ and gHBfix $\binom{\mathrm{NH\ldots O}}{\eta}$ base-base terms (Table 4). Subsequently, we used clustering analysis to estimate the population of the native state (see Methods). It turned out that additional support of (–NH…N–) H-bond interactions (with η = +1.0 kcal/mol) is crucial in order to promote significant population of the folded stem and to stabilize the native arrangement of the TL (Figure 4 and Table 4). This is in agreement with the known *ff* imbalance, where N atoms tend to have too large van der Waals radii, forcing the (–NH…N–) H-bonds to fluctuate around larger distances with respect to Quantum Mechanical (QM) calculations and experimental datasets.[42-44, 79] We observed that a sufficiently strong support of (–NH…O–) H-bonds (with η equal to +1.0 kcal/mol) could also

stabilize folded stem and/or native stem/loop even in combination with weaker (–NH…N–) support. This likely compensates for the effect of weaker (–NH…N–) interactions; see the r(gcGAGAgc) REST2 simulation with the gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+0.5}\binom{\mathrm{NH\ldots O}}{+1.0}$ combination (Figure 4 and Table 4). Nonetheless, in such case the population of the native loop conformation is suboptimal and this rather heavy gHBfix combination leads already to some side effects in simulations of TNs (see below). It also leads to a significant population of the left-handed Z-form helix conformation[80] (stem guanines in *syn* orientation) instead of the dominant A-form in the r(gcGAGAgc) system (Figure 4). In general, it is advisable to keep the gHBfix *ff* corrections as mild as possible. In summary, we suggest to apply the strengthening on the (–NH…N–) H-bonds rather than on the (–NH…O–) H-bonds.

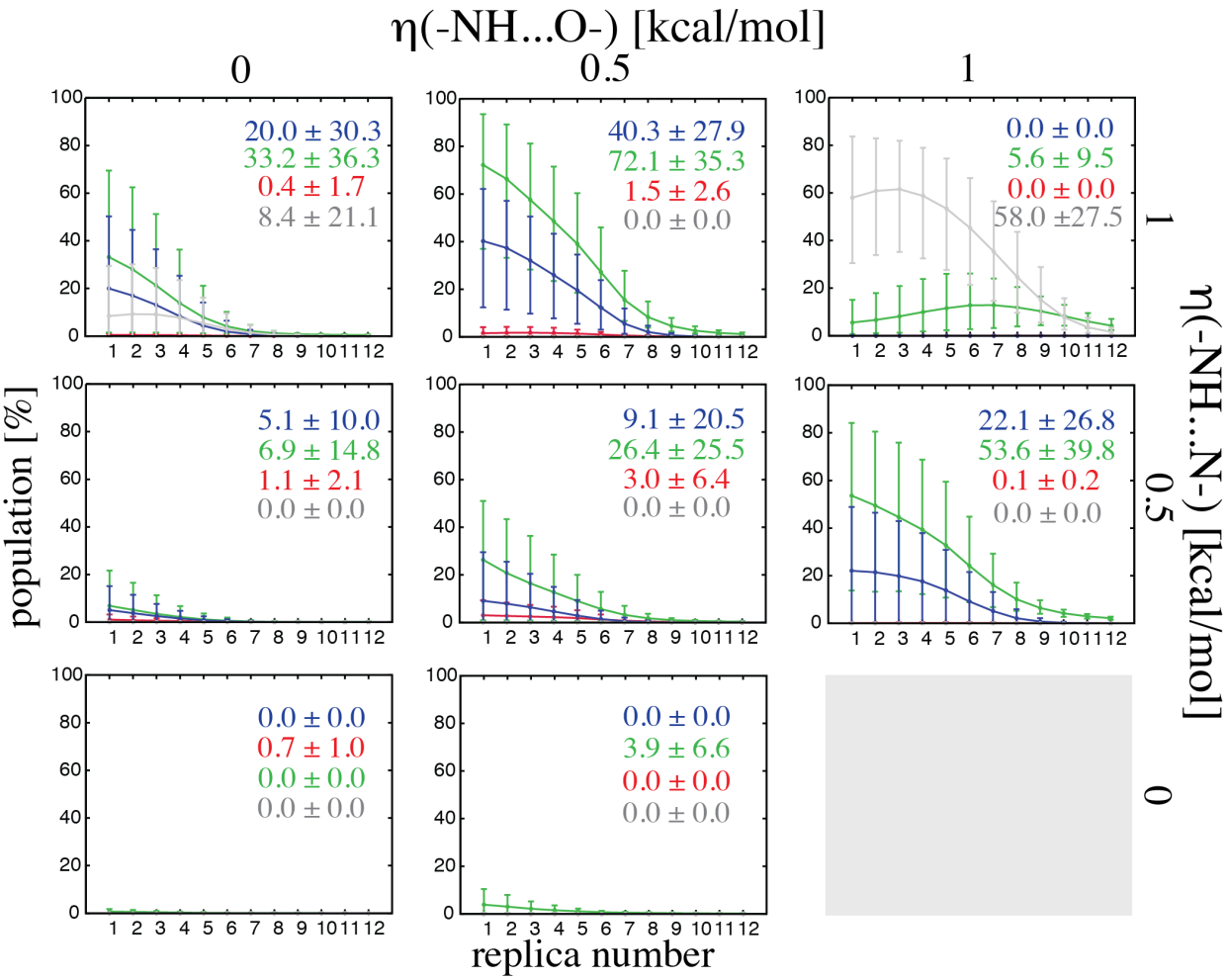

Figure 4. Populations (%) of the most important types of structures in r(gcGAGAgc) REST2 simulations with various gHBfix potentials for all twelve ladder replicas with errors estimated using bootstrapping with resampling both over time- and replica-domains (see Methods for details). Population of the native state is shown in blue. The remaining populations represent correctly folded A-form stem with any loop conformation (in green), correctly folded apical loop with stem not folded (red), and structures with left-handed Z-form stem with any loop conformation, i.e., including structures with properly structured apical loop accompanied with Z-form stem (gray). Displayed numbers highlight the final population in the unbiased replica 1 (T=298 K). Note that the gHBfix$\left(^{2-\mathrm{OH\ldots nbO/bO}}_{-0.5}\right)$ potential was applied in all simulations.

Next, we analyzed the effect of the above-introduced base-base gHBfix biases on the r(GACC) TN conformational dynamics. We performed a similar set of nine REST2 simulations of r(GACC) as in the case of r(gcGAGAgc) TL folding differing in base-base gHBfix terms, all with $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}$ weakening SPh interactions (Table 5). We identified some spurious side effects of certain combinations of base-base gHBfix, namely, in the case of a high +1.0 kcal/mol value of gHBfix bias constant η applied to (–NH…O–) interactions. In such cases, the population of the native RNA A-major conformation was significantly reduced (Figure 5). For example, $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}\binom{\mathrm{NH\ldots O}}{+1.0}$ potential reduced the population of RNA A-major conformation to just 15.5% which is even below what we obtained during the standard simulation without any gHBfix potential (Table 3). Although we did not detect intercalated structures during any r(GACC) REST2 simulation with gHBfix potentials, an excessive support applied to (–NH…O–) interactions resulted in loop-like structures closed by spurious base-base interactions (Figure S1 in Supporting Information). Hence, REST2 simulations with both r(GACC) TN and r(gcGAGAgc) TL show that the $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ potential represents currently the best compromise (among those variants tested) for tuning H-bonding interactions with a correct balance between different conformers according to the experiments.[16, 18, 42]

Table 4. Tuning of the gHBfix potential for base pairing and its effects on the folding of r(gcGAGAgc) TL. Populations (in %) of two major conformations, i.e., the native state (with properly folded stem and loop) and all states with folded stem (independent of the loop conformation), are displayed for each gHBfix combination at the reference REST2 replica

(T = 298 K). See Figure S1 for the examples of most populated clusters from REST2 simulations. [a]

(–NH…N–) interaction bias (kcal/mol)	(–NH…O–) interaction bias (kcal/mol)		
	$\binom{\mathrm{NH...O}}{+0.0}$	$\binom{\mathrm{NH...O}}{+0.5}$	$\binom{\mathrm{NH...O}}{+1.0}$
$\binom{\mathrm{NH...N}}{+1.0}$	20.0/33.2 [b]	40.3/72.1	0.0/5.6
$\binom{\mathrm{NH...N}}{+0.5}$	5.1/6.9	9.1/26.4	22.1/53.6
$\binom{\mathrm{NH...N}}{+0.0}$	0.0/0.0 [c]	0.0/3.9	--/-- [d]

[a] Cluster analysis performed for the last 7 out of 10 μs-long trajectory at 298 K; the gHBfix$\binom{2-\mathrm{OH...nbO/bO}}{-0.5}$ potential applied in all simulations.
[b] The numbers before and after "/" report population of native state with all signature interactions and all states with folded stem, respectively. The population of native state is thus inherently included within the population of all possible structures with folded stem.
[c] REST2 simulation with the gHBfix$\binom{2-\mathrm{OH...nbO/bO}}{-0.5}$ potential affecting only SPh interactions.
[d] REST2 simulation was not performed.

Table 5. The effect of gHBfix potential for base pairing on structural dynamics of r(GACC) TN. Populations (in %) of two major conformations, i.e., RNA A-major (the first number) and A-minor (the second number), are displayed for each gHBfix combination at the reference replica (T = 298 K). See Figure S1 in Supporting Information for examples of the other populated conformations from the REST2 simulations. The third number in the bracket displays the χ^2, which further validates simulations against the data from experiments. [a,b]

(–NH…N–) interaction bias (kcal/mol)	(–NH…O–) interaction bias (kcal/mol)		
	$\binom{\mathrm{NH...O}}{+0.0}$	$\binom{\mathrm{NH...O}}{+0.5}$	$\binom{\mathrm{NH...O}}{+1.0}$
$\binom{\mathrm{NH...N}}{+1.0}$	46.2/9.7 (0.32 [c])	35.1/6.0 (0.78)	13.5/2.4 (1.98)
$\binom{\mathrm{NH...N}}{+0.5}$	56.8/5.6 (0.11)	55.9/9.0 (0.23)	33.2/5.7 (0.95)
$\binom{\mathrm{NH...N}}{+0.0}$	58.7/7.4 (0.15) [d]	57.6/9.8 (0.15)	54.1/9.3 (0.42)

[a] gHBfix$\binom{2-\mathrm{OH...nbO/bO}}{-0.5}$ weakening SPh interactions is applied in all simulations.
[b] Cluster analysis performed for the last 7 out of 10 μs-long trajectory at 298 K.
[c] χ^2 values were obtained as described elsewhere[27] by calculating and comparing backbone ^{3}J scalar couplings, sugar ^{3}J scalar couplings, nuclear Overhauser effect (NOE) intensities, and also using the absence of specific peaks in NOE spectroscopy data.
[d] REST2 simulation with the gHBfix$\binom{2-\mathrm{OH...nbO/bO}}{-0.5}$ potential only (Table 3, last column).

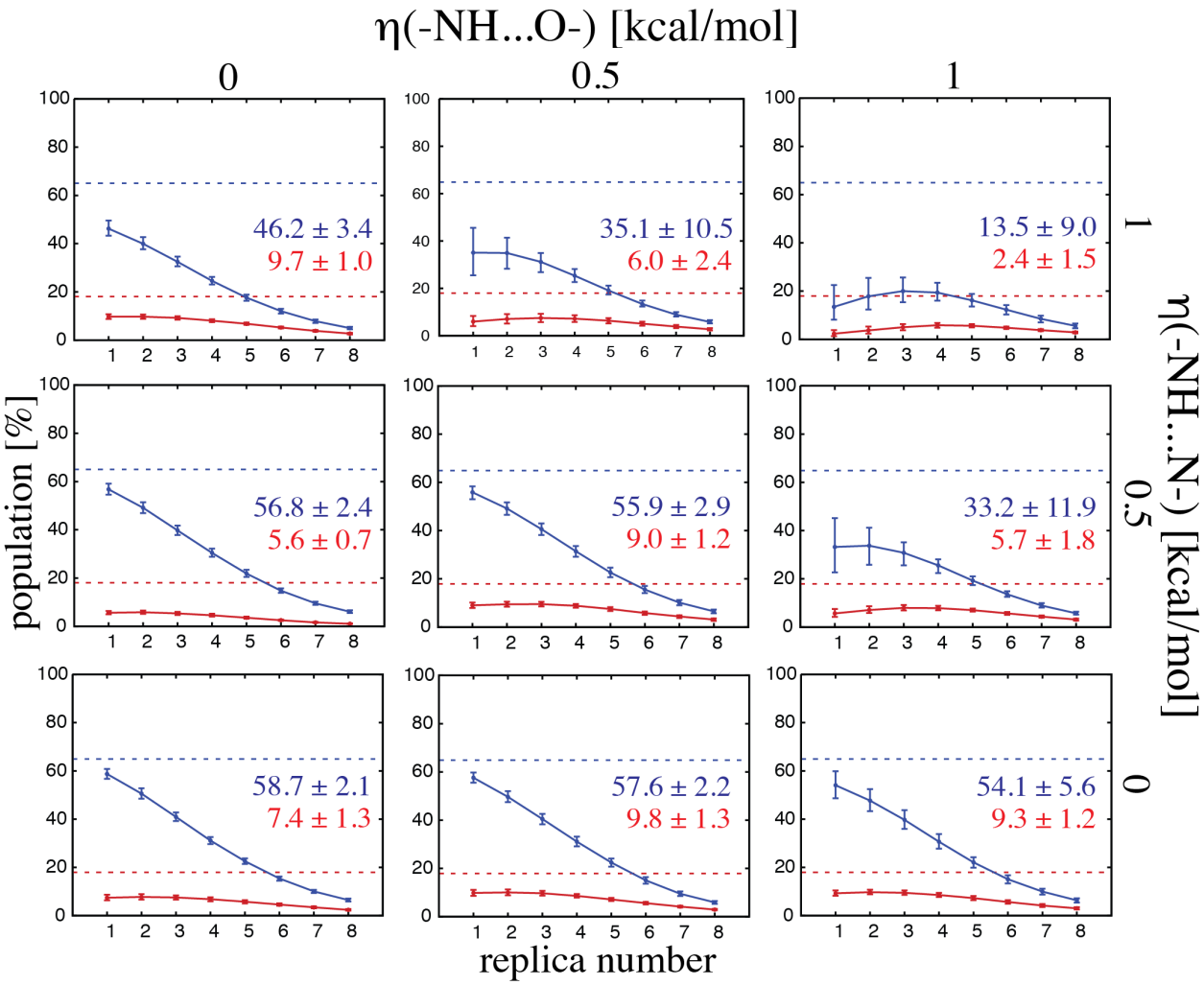

Figure 5. Population analysis of r(GACC) REST2 simulation with various gHBfix potentials. Occurrences (in %) of major conformers, i.e., RNA A-major (in blue) and A-minor (in red) is shown for each eight replicas in the ladder. Displayed numbers indicate the final population in the unbiased replica 1 (T=298 K). Dashed horizontal lines indicate populations suggested by experiments.[16] Note that the gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}$ potential was applied in all simulations.

Present combination of gHBfix parameters is not sufficient to entirely eliminate the intercalated state in other TN sequences. Besides the r(GACC) TN sequence, four other TNs, i.e., r(CAAU), r(CCCC), r(AAAA), and r(UUUU), are commonly used for *ff* validation[17-20, 22, 24-

[25, 27] due to availability of the corresponding benchmark NMR data. We thus performed additional REST2 simulations of those TN sequences in order to explore the effects of the gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ potential on them. REST2 simulations were initiated with the same setup as for the r(GACC) TN (see Methods). We found that in contrast to r(GACC) TN, the application of the gHBfix *ff* to these sequences resulted in a visible albeit still limited improvement over the χ_{OL3}. Namely, the population of native A-form was increased in all four sequences while the population of artificial intercalated structure was significantly reduced only in two of them, r(CCCC) and r(AAAA) (see Table 6). Nonetheless, the population of intercalated structure still remained unsatisfactory in r(CAAU), r(CCCC), and r(AAAA) sequences, and notably, the r(CAAU) sequence revealed intercalated structure to be still more populated than the native A-form despite the gHBfix correction (Table 6). This suggests that the gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ variant is not yet robust enough to entirely eliminate all artificial intercalated states in TNs and further *ff* modifications would be vital. It is worth to note that with gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ correction, all sequences having still unsatisfactory population of the intercalated structure involve C or A as 5'-terminal nucleotide, i.e., nucleotide, which can form 7BPh interaction[63] in the intercalated state. This indicates that weakening of BPh interactions that were previously reported to be overpopulated in unfolded states[54] might be used to eliminate the artificial intercalated structure. Work is in progress in our laboratories to find appropriate gHBfix potentials and parameters to refine simulations of all TNs simultaneously. However, although we already have promising results, finding the best solution is beyond the scope of the present study. It requires testing of a large number of parameter combinations, including subsequent verification of absence of side effects for a broad set of other RNA systems.

Table 6. The population in % of native RNA A-form (including both A-major and A-minor states) and artificial intercalated structures as obtained in standard χ_{OL3} and with the suggested gHBfix modification.

sequence	A-form		Intercalated	
	χ_{OL3}	gHBfix	χ_{OL3}	gHBfix
r(CAAU)	<~2% [a]	~21% [d]	~50% [a,b]	~51% [d]
r(CCCC)	~20% [b,c]	~54% [d]	~56% [b,c]	~39% [d]
r(AAAA)	~21% [a,b]	~25% [d]	~23% [a,b]	~11% [d]
r(UUUU)	~4% [a,b]	~5% [d]	~18% [b]	~3% [d]

[a] Ref. [18]
[b] Ref. [27]
[c] Ref. [19]
[d] this study

Comment on base – phosphate interactions. When tuning H-bond interactions we need to take into account that under or over-estimation of a given H-bond is likely not uniform across diverse RNA structures. The apparent *ff* deficiency for a given interaction reflected by its population can be substantially context-dependent, as it is also affected by contributions from many other terms such as base stacking, backbone conformations, etc. Thus, one has to be always concerned with potential of over-corrections, which could improve some structures but deteriorate others. This is the reason why we so far did not apply gHBfix for the BPh interactions. Tuning of BPh interactions may have conflicting consequences in simulations of TNs, where they should be avoided and TLs, where they do form significant signature interactions. As BPh interactions are widespread in folded RNAs,[63] their uniform weakening or strengthening in the *ff* could lead to ambiguous results. We nevertheless work on finding suitable gHBfix modifications of the BPh interactions that would improve RNA simulations without introducing side effects.

UNCG TL as a challenging system. As noted in the Introduction, earlier studies indicated that the UNCG TL is a considerably more difficult system than the TNs and GNRA TLs. In order to prove that the native state population of the r(gcUUCGgc) TL might be improved by modification of non-bonded terms, we initially applied the structure-specific HBfix potential to all ten native interactions of the r(gcUUCGgc) TL motif, i.e., to the six H-bonds of the two canonical GC base pairs of the stem and the five signature H-bonds within the loop (Figure 1). Each native H-bond was biased by +1.0 kcal/mol in favor of the bound state (see Methods for details). Similar structure-based HBfix successfully folds the GNRA TL.[54] We obtained converged results, with the population of r(gcUUCGgc) TL native state (27.1 ± 9.6% at 298 K) in a reasonable agreement with experiments[42-44] (see Supporting Information for details). With aid of the structure-based HBfix, the UUCG TL was readily able to significantly populate native fold including *syn*-orientation of the G_{L4} nucleotide (Figure 1), without any specific tuning of the guanosine dihedral potential. However, we recall that these corrections are structure-specific and cannot thus be used for general RNA sequences.

Next, we tried gHBfix potentials that revealed promising results for structural description and folding of both r(GACC) TN and r(gcGAGAgc) TL. Namely, we performed r(gcUUCGgc) REST2 folding simulation with $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}\binom{\mathrm{NH\ldots O}}{+1.0}$, $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$, and $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+0.5}$ potentials (Table S1 in Supporting Information). Unfortunately, we did not detect the presence of the native state. The application of $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}\binom{\mathrm{NH\ldots O}}{+1.0}$ showed rapid stem folding but the loop fluctuated between several non-native conformers (Figure S3 in Supporting Information). In contrast, the $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ and $\mathrm{gHBfix}\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+0.5}$ potentials led to either zero or marginal stem folding while misfolded states dominated the simulations (Figure S3

in Supporting Information). The fact that gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}$ potential did not lead to any folding of the native state in r(gcUUCGgc) TL, while it was sufficient to reveal significant folding in r(gcGAGAgc) TL confirms that the inherent stability of the UUCG loop region is underestimated by the *ff* significantly more than in the case of GNRA TL.[45] This conclusion is further supported by the observation that only misfolded loop states are populated when the stem is forced to fold by gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}\binom{\mathrm{NH}\ldots\mathrm{O}}{+1.0}$ potential, which already overstabilizes the base-base interactions (see above).

The analysis of misfolded conformers obtained by the r(gcUUCGgc) REST2 simulations with the presently available gHBfix potentials showed several noncanonical interactions in the loop region dominating over the native ones. Based on this analysis, we decided to probe effect of either stabilization or destabilization of few other kinds of H-bonds. In particular, we tested the effect of the (i) stabilization of all possible interactions between 2'-OH groups and H-bond acceptors of nucleobases with bias energy η equaling to either +0.5 kcal/mol or +1.0 kcal/mol (simulations denoted as gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}\binom{\mathrm{NH}\ldots\mathrm{O}}{+1.0}\binom{2-\mathrm{OH}\ldots\mathrm{N/O}}{\eta}$), and (ii) the destabilization of interactions between all nucleobase proton donors (–NH) and all O2' oxygens, and between 2'-OH groups and O2'/O4' oxygens, both with η equaling to –0.5 kcal/mol (simulation denoted as gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}\binom{\mathrm{NH}\ldots\mathrm{O}}{+1.0}\binom{\mathrm{NH}\ldots\mathrm{O2}}{-0.5}\binom{2-\mathrm{OH}\ldots\mathrm{O2/O4}}{-0.5}$, see Table S1 in Supporting Information for overview of the REST2 simulations). In both cases, we observed that the population of structures with correctly folded A-form stem slightly increased (up to ~70%) but the loop still sampled only non-native misfolded states similarly to those obtained with the gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}\binom{\mathrm{NH}\ldots\mathrm{O}}{+1.0}$, see Figure S3 in Supporting Information.

Subsequently, we probed the simultaneous effect of both gHBfix biases introduced in the previous paragraph, i.e., additional stabilization of sugar-base H-bonds and destabilization of interactions involving O2' and O4' oxygens. In contrast to simulations discussed in the previous paragraph, we also modified gHBfix bias energies η for base-base interactions, so that the additional simulations probed the gHBfix$\binom{\text{2-OH...nbO/bO}}{-0.5}\binom{\text{NH...N}}{\eta}\binom{\text{NH...O}}{\eta}\binom{\text{2-OH...N/O}}{+0.5}\binom{\text{NH...O2}}{-0.5}\binom{\text{2-OH...O2/O4}}{-0.5}$ function, in which η values equaled either to +0.5 or +1.0 kcal/mol (see Table S1 in Supporting Information). Despite all attempts, the folded native state appeared only in simulation using the gHBfix$\binom{\text{2-OH...nbO/bO}}{-0.5}\binom{\text{NH...N}}{+0.5}\binom{\text{NH...O}}{+0.5}\binom{\text{2-OH...N/O}}{+0.5}\binom{\text{NH...O2}}{-0.5}\binom{\text{2-OH...O2/O4}}{-0.5}$ potential, with a rather marginal population of ~6 %. It was accompanied with ~14% population of misfolded states involving properly folded A-form stem but non-native loop conformations (Figure S3 in Supporting Information). In summary, additional biases to SPh and base-base interactions were not sufficient to correct the free-energy imbalance between native folded and misfolded states of the r(gcUUCGgc) TL in χ_{OL3}. Additional modifications involving sugar-base interactions and hydrogen bonding between 2'-OH groups and O2/O4' oxygens were required in order to detect at least a small fraction of native states during folding simulations of r(gcUUCGgc) motif. Thus it seems that the correct description of the r(gcUUCGgc) folding might suffer also from some other *ff* inaccuracies that go beyond fine-tuning of non-bonded terms. In contrast to TNs which should be relatively easily correctable (see above), we presently have no clues how to decisively improve folding of the UNCG TLs.

Application of gHBfix to a diverse set of folded RNAs shows no side effects. Structural dynamics and folding simulations of both TN and TL motifs revealed that the combination of the

gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}$ potential with χ_{OL3} significantly improves their structural behavior. Although further work is required, especially considering folding of the gcUNCGgc TL, we decided to test the performance of the χ_{OL3}/gHBfix$\binom{2-\mathrm{OH}\ldots\mathrm{nbO/bO}}{-0.5}\binom{\mathrm{NH}\ldots\mathrm{N}}{+1.0}$ combination for other important (and much larger) RNA motifs and systems. Namely, we performed a set of standard MD simulations of Sarcin-Ricin loop, ribosomal L1-stalk, Kissing-loop complex, Hairpin ribozyme, preQ and Neomycin-sensing riboswitches, Kink-turns, RNA duplex and G-quadruplex (see Table S2 for overview of unbiased MD simulations). Importantly, the gHBfix potential did not cause any side effects and unexpected rearrangements in comparison with the unmodified widely used χ_{OL3} RNA *ff* (see Supporting Information for details). Therefore, gHBfix appears to significantly improve simulation performance for at least some difficult RNA systems while not causing any undesired side effects in standard simulations of folded RNAs.

Comment on other RNA *ff*s. As noted in the Introduction, despite occasional optimistic claims in the literature, none of the other *ff*s available in the literature as of the end of 2017 was proven to satisfactorily simulate the RNA TLs and TNs without undesired side effects, as reviewed in Ref. [4]. Another RNA *ff* modification (abbreviated as DESRES here) has been published recently,[52] reparametrizing the non-bonded as well as dihedral terms of the AMBER RNA *ff* and complementing the resulting parametrization with a specific water model.[81] We have tested this *ff* (see Supporting Information for full details) with the following results.

The DESRES parameters, as implemented by us (see Supporting Information for the parameters), lead to serious side effects for some important folded RNA molecules. For example, structures of RNA Kink-turns and ribosomal L1-stalk RNA segment were entirely and reproducibly disrupted (Figures S15 and S16 in Supporting Information), suggesting that the

DESRES *ff* may be excessively biased in favor of the A-form RNA while also some non-canonical base pairs may be destabilized by the introduced modifications of the non-bonded terms (Figures S11-S14 in Supporting Information). Such unstable behavior has never been observed in common AMBER *ff*s (including the version used and tested here modified by the gHBfix potential) which work well for these systems.

Further, using extensive REST2 simulations, we were not capable to fold the r(gcGAGAgc) and r(gcUUCGgc) TLs, despite the fact that even these short constructs should have a non-negligible folded population based on experimental data.[44] In other words, the DESRES potential[52] with our protocol/implementation did not bring any observable benefit over standard χ_{OL3}[17-18, 42-44] for the description of small 8-mer TLs. Note that in the original paper, the DESRES *ff* was reported to fold TLs with very long stems. Thus, the folding events may result from the increased propensity to form A-form double helix. Interestingly, we have carried out series of unbiased DESRES simulations of the UUCG TL with a long stem starting from the folded states and we have observed loss of UUCG signature interactions on a time scale around 5-10 μs in all of them (Figure S9 in Supporting Information). We thus suggest that the DESRES *ff* would require more tests before being used to simulate general RNA motifs.

CONCLUSIONS

Several recent studies suggested that the nucleic acid *ff*s suffer from non-optimal description of the non-bonded interactions.[3, 19-20, 24-25, 45, 49, 52, 54] Here we introduce a general pair potential, a generalized HBfix (gHBfix), that could be used to selectively fine-tune specific non-bonded terms, in particular specific types of hydrogen-bonding interactions. The gHBfix potential could be easily combined with any existing *ff*s to probe the effect of the modification of its non-bonded

terms. Most importantly, unlike in case of vdW or charge modifications, it affects only the selected type of the interactions, while the others, including, e.g., interactions with the solvent and ions, remain intact. Thus, it reduces the likelihood of introducing undesirable side effects that are frequently associated with other ways of modifying non-bonded terms. We emphasize that the gHBfix term is an entirely legitimate way of modifying of the pair-additive NA *ff* since, in fact, it is not more empirical than the other terms. The common MM parameters such as atomic charges and radii are in fact not QM observables, so it seems genuine to start straightforwardly target properties of key molecular interactions instead of trying to refine properties of MM atoms. Thus, the gHBfix term is as general as all the other *ff* terms and can be physically justified, as it may compensate, e.g., for the lack of polarization terms in H-bonding.[4, 54] A similar effect can be achieved also by modification of van der Waals parameters via NBfix,[25, 46, 49, 82] although to our opinion in significantly less controllable manner due to the long range of van der Waals interactions, with higher risk of imposing some spurious artifacts. It is possible that in future that NBfix and gHBfix corrections could be applied simultaneously, in a synergy.

We tested the gHBfix potential for fine-tuning the solute-solute H-bond interactions in RNA folding of small TLs and TN RNAs; in particular by strengthening of the base-base interactions and weakening the SPh interactions that were both suggested to, e.g., interfere with correct description of the TLs folding. We found out that the negative gHBfix applied to all possible interactions between 2'-OH groups as proton donors and nbO and bO oxygens as acceptors (i.e., weakening of the SPh interactions) is able to eliminate (or reduce) some *ff* artifacts, e.g., the presence of intercalated structures of TNs. Furthermore, the additional support of (–NH…N–) H-bonds (i.e, positive gHBfix strengthening the base pairing interactions) significantly promotes

population of the TL folded stem and the native arrangement of the GNRA loop. The necessity of additional support of H-bonds between groups with N atoms is probably connected with their outsized van der Waals radii in *ff*s as suggested in some recent studies.[42-44, 79] In summary, the presented gHBfix$\binom{2-\mathrm{OH\ldots nbO/bO}}{-0.5}\binom{\mathrm{NH\ldots N}}{+1.0}$ potential eliminates or destabilizes spurious conformations such as intercalated structures of the r(GACC) TN. The conformational ensembles generated by the *ff* complemented by the gHBfix reveal better agreement with available experimental data than the corresponding ensemble generated by the original unmodified AMBER χ_{OL3} RNA *ff*; in particular a significant improvement was observed in population of both canonical A-major and A-minor single stranded structures in r(GACC) TN folding simulations and the population of native state in the simulation of small 8-mer GNRA TL. Most importantly, we have tested the suggested gHBfix tuning on a wide portfolio of other RNA structures without noticing any undesired side effects in standard simulations. We suggest that especially the improved stabilization of base pairs may be profitable for many other systems in long simulations.

Although the suggested gHBfix provides significant improvement to the standard RNA *ff*, it is not yet sufficient to eliminate intercalated structures for r(CAAU), r(CCCC), r(AAAA) TNs and to correct the free-energy imbalance between folded and misfolded states observed for the challenging r(gcUUCGgc) TL. Thus, further tuning of the gHBfix potential is required and it seems that UUCG TL might require additional corrections that go beyond the fine-tuning of non-bonded terms, e.g., by reparametrization of the dihedral terms. Considering the variability and complexity of RNA structures, and the simplicity of the functional form of current *ff*s, it is probably naive to suppose that one would be able to introduce a perfect RNA *ff* with the ability to provide correct behavior of all RNA systems by tuning just some non-bonded terms within the

framework of pair-additive *ff*s.[4] On the other hand, the introduced gHBfix potential increases the flexibility of the *ff* in a controlled way by adding a small number of independent parameters and clearly shows that even the behavior of the pair-wise non-polarizable *ff*s could still be significantly improved. In general, *ff* improvements may be aided by using additional new terms that go beyond the basic conventional *ff* form and that further increase the flexibility of the *ff*. gHBfix is an example of such modifications.

We would like to point out that the primary purpose of this paper is introduction of the basic methodology of gHBfix potentials and demonstration of its efficiency in tuning the performance of the RNA *ff*. We do not attempt to release a new version of RNA AMBER *ff*, although the version presented and tested in this work does provide some improvements while not showing any undesired side effect for the tested set containing various important RNA systems. It can thus be quite safely used in RNA simulations. Work is in progress to balance some additional H-bond interactions using the gHBfix potentials as well as to optimize the biasing parameters η and to merge the gHBfix potentials with adjustments of some of the core *ff* terms. We suggest that future refinements of the pair-additive RNA *ffs* will likely require adding additional *ff* terms (such as the gHBfix) to the basic functional form to increase flexibility of the parametrization. This together with testing on a broad set of RNA systems will help to avoid over-fitting of the *ff* in favor of one type of target RNA structures, such as, e.g., the TNs or A-form RNA, as demonstrated in our study. Due to the huge dimensionality of the parameter space and mutual inter-dependence of the *ff* terms, finding an optimal *ff* will be most likely a long-term process, which would profit from collaborative efforts within the RNA simulation community.

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge. Description of the functional form of the gHBfix potential, details about convergence of enhanced sampling simulations, structure preparation and other details about simulation protocols, description of the UNCG folding simulation with structure specific HBfix, description of REST2 and standard MD simulations with DESRES potential, description of standard MD simulations with the gHBfix potential, Supporting Tables and Figures (PDF). C++ code to generate restrains for AMBER input files for simulation with the external gHBfix (gHBfix.pdf). AMBER input files containing DESRES parameters (desres.zip).

AUTHOR INFORMATION

Corresponding Authors

sponer@ncbr.muni.cz, pavel.banas@upol.cz

ACKNOWLEDGMENT

We thank Petr Jurečka and Petr Stadlbauer for useful discussions. Sandro Bottaro is acknowledged for critically reading the manuscript and providing comments. This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic [LO1305 to M.O., P.B., P.K.]; and Czech Science Foundation [18-25349S to P.B., P.K.]. J.S. acknowledges support by Praemium Academiae.

REFERENCES

(1) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in. *Acc Chem Res* **2010,** *43*, 40-47.
(2) Cheatham, T. E.; Case, D. A. Twenty-Five Years of Nucleic Acid Simulations. *Proc Natl Acad Sci U S A* **2013,** *99*, 969-977.
(3) Sponer, J.; Banas, P.; Jurecka, P.; Zgarbova, M.; Kuhrova, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J Phys Chem Lett* **2014,** *5*, 1771-1782.
(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) Maffeo, C.; Yoo, J.; Comer, J.; Wells, D. B.; Luan, B.; Aksimentiev, A. Close encounters with DNA. *J Phys Condens Matter* **2014,** *26*, 413101.
(6) 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*.
(7) 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*.
(8) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr Opin Struct Biol* **2018,** *49*, 129-138.
(9) 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.
(10) Huang, J.; Lemkul, J. A.; Eastman, P. K.; MacKerell, A. D., Jr. Molecular dynamics simulations using the drude polarizable force field on GPUs with OpenMM: Implementation, validation, and benchmarks. *J Comput Chem* **2018**.
(11) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999,** *314*, 141-151.
(12) 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.
(13) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci U S A* **2002,** *99*, 12562-12566.
(14) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* **2008,** *100*.
(15) Mlynsky, V.; Bussi, G. Exploring RNA structure and dynamics through enhanced sampling simulations. *Curr Opin Struct Biol* **2018,** *49*, 63-71.
(16) 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.
(17) 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* **2013,** *52*, 996-1010.
(18) 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.

(19) 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.
(20) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., 3rd Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015,** *21*, 1578-1590.
(21) 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*, 492-499.
(22) 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.
(23) 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.
(24) 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.
(25) 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.
(26) 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.
(27) 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.
(28) Bottaro, S.; Gil-Ley, A.; Bussi, G. RNA folding pathways in stop motion. *Nucleic Acids Res* **2016,** *44*, 5883-5891.
(29) Haldar, S.; Kuhrova, P.; Banas, P.; Spiwok, V.; Sponer, J.; Hobza, P.; Otyepka, M. Insights into Stability and Folding of GNRA and UNCG Tetra loops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics. *J Chem Theory Comput* **2015,** *11*, 3866-3877.
(30) Mathews, D. H.; Turner, D. H. Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol* **2006,** *16*, 270-278.
(31) Sweeney, B. A.; Roy, P.; Leontis, N. B. An Introduction to Recurrent Nucleotide Interactions in RNA. *Wiley Interdiscip Rev RNA* **2015,** *6*, 17-45.
(32) Hall, K. B. Mighty tiny. *RNA* **2015,** *21*, 630-631.
(33) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. Single nucleotide RNA choreography. *Nucleic Acids Res* **2006,** *34*, 1481-1491.
(34) Varani, G. Exceptionally Stable Nucleic-Acid Hairpins. *Annu Rev Biophys Biom* **1995,** *24*, 379-404.
(35) Uhlenbeck, O. C. Nucleic-Acid Structure - Tetraloops and Rna Folding. *Nature* **1990,** *346*, 613-614.
(36) Pley, H. W.; Flaherty, K. M.; Mckay, D. B. Model for an Rna Tertiary Interaction Front the Structure of an Intermolecular Complex between a Gaaa Tetraloop and an Rna Helix. *Nature* **1994,** *372*, 111-113.
(37) Xin, Y. R.; Laing, C.; Leontis, N. B.; Schlick, T. Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA* **2008,** *14*, 2465-2477.

(38) Chauhan, S.; Woodson, S. A. Tertiary interactions determine the accuracy of RNA folding. *J Amer Chem Soc* **2008,** *130*, 1296-1303.
(39) Marino, J. P.; Gregorian, R. S.; Csankovszki, G.; Crothers, D. M. Bent Helix Formation between Rna Hairpins with Complementary Loops. *Science* **1995,** *268*, 1448-1454.
(40) Brion, P.; Westhof, E. Hierarchy and dynamics of RNA folding. *Annu Rev Biophys* **1997,** *26*, 113-137.
(41) Antao, V. P.; Tinoco, I. Thermodynamic Parameters for Loop Formation in Rna and DNA Hairpin Tetraloops. *Nucleic Acids Res* **1992,** *20*, 819-824.
(42) Sheehy, J. P.; Davis, A. R.; Znosko, B. M. Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* **2010,** *16*, 417-429.
(43) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **2004,** *101*, 7287-7292.
(44) Abdelkafi, M.; Ghomi, M.; Turpin, P. Y.; Baumruk, V.; Herve du Penhoat, C.; Lampire, O.; Bouchemal-Chibani, N.; Goyer, P.; Namane, A.; Gouyette, C., et al. Common structural features of UUCG and UACG tetraloops in very short hairpins determined by UV absorption, Raman, IR and NMR spectroscopies. *J Biomol Struct Dyn* **1997,** *14*, 579-593.
(45) Bottaro, S.; Banas, P.; Sponer, J.; Bussi, G. Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J Phys Chem Lett* **2016,** *7*, 4032-4038.
(46) Chen, A. A.; Garcia, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci U S A* **2013,** *110*, 16820-16825.
(47) 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.
(48) Smith, L. G.; Tan, Z.; Spasic, A.; Dutta, D.; Salas-Estrada, L. A.; Grossfield, A.; Mathews, D. H. Chemically Accurate Relative Folding Stability of RNA Hairpins from Molecular Simulations. *bioRxiv* **2018,** *354332*.
(49) Yoo, J.; Aksimentiev, A. Improved Parameterization of Amine-Carboxylate and Amine-Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J Chem Theory Comput* **2016,** *12*, 430-443.
(50) 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.
(51) Zgarbova, M.; Jurecka, P.; Banas, P.; Havrila, M.; Sponer, J.; Otyepka, M. Noncanonical alpha/gamma Backbone Conformations in RNA and the Accuracy of Their Description by the AMBER Force Field. *J Phys Chem B* **2017,** *121*, 2420-2433.
(52) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci U S A* **2018,** *115*, E1346-E1355.
(53) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J Chem Theory Comput* **2010,** *6*, 3836-3849.
(54) 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.

(55) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J Chem Theory Comput* **2012,** *8*, 4405-4412.
(56) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J Phys Chem Lett* **2014,** *5*, 3863-3871.
(57) 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.
(58) Krepl, M.; Vogele, J.; Kruse, H.; Duchardt-Ferner, E.; Wohnert, J.; Sponer, J. An intricate balance of hydrogen bonding, ion atmosphere and dynamics facilitates a seamless uracil to cytosine substitution in the U-turn of the neomycin-sensing riboswitch. *Nucleic Acids Res* **2018,** *46*, 6528-6543.
(59) 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*.
(60) 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.
(61) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
(62) 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.
(63) 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.
(64) 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.
(65) 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.
(66) 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.
(67) 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. *Biophysical Journal* **2007,** *92*, 3817-3829.
(68) 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.
(69) 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.
(70) Bottaro, S.; Di Palma, F.; Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res* **2014,** *42*, 13306-13314.

(71) Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014,** *344*, 1492-1496.
(72) Hall, P.; Horowitz, J. L.; Jing, B. Y. On Blocking Rules for the Bootstrap with Dependent Data. *Biometrika* **1995,** *82*, 561-574.
(73) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph Model* **1996,** *14*, 33-38.
(74) *The PyMOL Molecular Graphics System, Version 2.0*
(75) Sponer, J.; Bussi, G.; Stadlbauer, P.; Kuhrova, P.; Banas, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of guanine quadruplex molecules-funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochim Biophys Acta* **2017,** *1861*, 1246-1263.
(76) Yang, C.; Kulkarni, M.; Lim, M.; Pak, Y. Insilico direct folding of thrombin-binding aptamer G-quadruplex at all-atom level. *Nucleic Acids Res* **2017,** *45*, 12648-12656.
(77) Havrila, M.; Stadlbauer, P.; Kuhrova, P.; Banas, P.; Mergny, J. L.; Otyepka, M.; Sponer, J. Structural dynamics of propeller loop: towards folding of RNA G-quadruplex. *Nucleic Acids Res* **2018**.
(78) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001,** *7*, 499-512.
(79) Banas, P.; Mladek, A.; Otyepka, M.; Zgarbova, M.; Jurecka, P.; Svozil, D.; Lankas, F.; Sponer, J. Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J Chem Theory Comput* **2012,** *8*, 2448-2460.
(80) Hall, K.; Cruz, P.; Tinoco, I., Jr.; Jovin, T. M.; van de Sande, J. H. 'Z-RNA'--a left-handed RNA double helix. *Nature* **1984,** *311*, 584-586.
(81) 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.
(82) 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.