

# UNCG RNA tetraloop as a formidable force-field challenge for MD simulations

*Klaudia Mráziková,<sup>1,2#</sup> Vojtěch Mlýnský,<sup>1#\*</sup> Petra Kührová,<sup>3</sup> Pavlína Pokorná,<sup>1,2</sup> Holger Kruse,<sup>1</sup> Miroslav Krepl,<sup>1,3</sup> Michal Otyepka,<sup>1,3</sup> Pavel Banáš<sup>1,3\*</sup> and Jiří Šponer<sup>1,3\*</sup>*

<sup>1</sup> Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic

<sup>2</sup> National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

<sup>3</sup> Regional Centre of Advanced Technologies and Materials, Faculty of Science, Palacký University, Šlechtitelů 27, 783 71, Olomouc, Czech Republic

\* corresponding authors, emails: [mlynsky@ibp.cz](mailto:mlynsky@ibp.cz), [pavel.banas@upol.cz](mailto:pavel.banas@upol.cz), [sponer@ncbr.muni.cz](mailto:sponer@ncbr.muni.cz)

# authors contributed equally

## ABSTRACT

Explicit solvent atomistic molecular dynamics (MD) simulations represent an established technique to study structural dynamics of RNA molecules and an important complement for diverse experimental methods. However, performance of molecular mechanical (MM) force fields (*ffs*) remains far from satisfactory even after decades of development, as apparent from a problematic structural description of some important RNA motifs. Actually, some of the smallest RNA molecules belong to the most challenging systems for MD simulations and, among them, the UNCG tetraloop is saliently difficult. We report a detailed analysis of UNCG MD simulations, depicting the sequence of events leading to the loss of the UNCG native state during MD simulations. We identify molecular interactions, backbone conformations and substates that are involved in the process. The total amount of MD simulation data analyzed in this work is close to 1.3 millisecond. Then, we unravel specific *ff* deficiencies using diverse quantum mechanical/molecular mechanical (QM/MM) and QM calculations. Comparison between the MM and QM methods shows discrepancies in the description of the 5'-flanking phosphate moiety and both signature sugar-base interactions. Our work indicates that poor behavior of the UNCG tetraloop in simulations is a complex issue that cannot be attributed to one dominant and straightforwardly correctable factor. Instead, there is a concerted effect of multiple *ff* inaccuracies that are coupled and amplifying each other. We attempted to improve the simulation behavior by some carefully-tailored interventions but the results are still far from satisfactory, underlying the difficulties in development of accurate nucleic acids *ffs*.

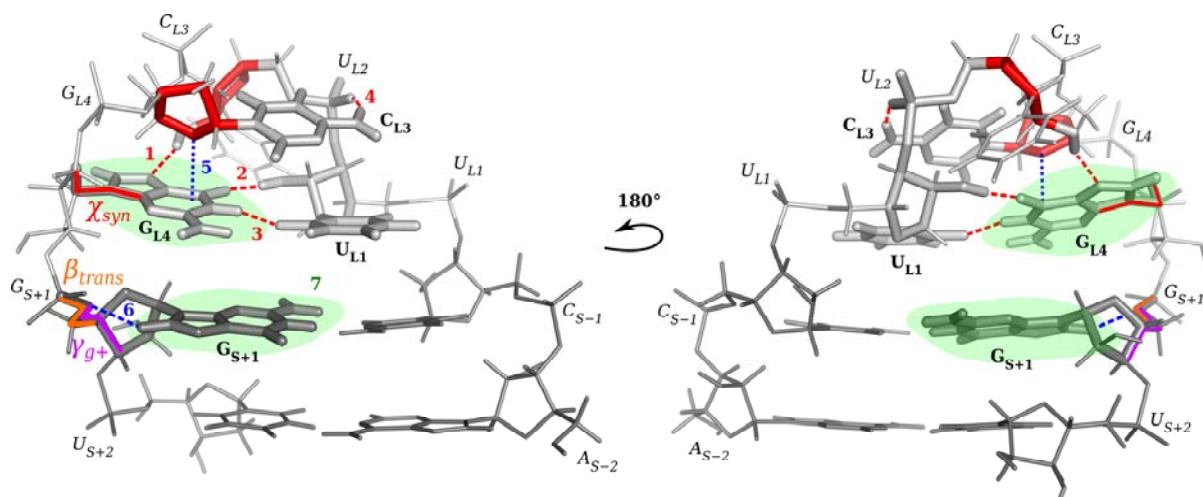
## INTRODUCTION

Explicit solvent atomistic molecular dynamics (MD) simulations represent an established tool to study structural dynamics of RNA molecules.<sup>1-3</sup> MD simulations can complement diverse experimental methods by, e.g., allowing to go beyond the typical static (ensemble-averaged) description of RNA molecules obtained by structural biology methods. However, the reliability of MD predictions depends on quality of the used molecular mechanical (MM) force fields (*ffs*). Unfortunately, performance of these *ffs* remains far from satisfactory despite ongoing development, and is also very system-dependent.<sup>4,5</sup> Many RNA molecules and protein/RNA complexes are quite well described in MD simulations, at least when starting the simulations from experimental structures. However, other systems progressively deteriorate upon extending the simulations. Surprisingly, even small RNA molecules may present persisting challenges for MD simulations and among them, the UNCG tetraloop is saliently difficult.<sup>6-11</sup>

RNA Tetraloops (TLs) are stem-loop RNA structures, which are common secondary structure elements in RNA. TLs participate in tertiary folding, RNA-RNA and protein-RNA interactions, ligand binding, and other important biological functions.<sup>12,13</sup> The UNCG TL (N stands for any nucleotide) is one of the two most abundant TLs.<sup>14-16</sup> It is thermodynamically very stable and possesses a clearly-defined dominant folded topology stabilized by a set of signature molecular interactions (**Figure 1**). The native state of UNCG TL remains difficult to be described by the current RNA *ffs*. Recent folding simulations of the UNCG TL indicated a large free-energy imbalance between native (folded) and other states (misfolded/unfolded conformations).<sup>6-10,17-20</sup> In addition, the characteristic UNCG native structure is lost in sufficiently long standard simulations. Two different effects were deemed to dominantly lead to the incorrect folded/unfolded free-energy balance, namely, (i) excessive stabilization of the unfolded ssRNA structure by intramolecular base-phosphate and sugar-phosphate interactions and (ii) destabilization of the native folded state by underestimation of the native H-bonds including the stem base pairing.<sup>8,9,21,22</sup> Recently, we introduced a general external potential tuning H-bond interactions (gHBfix),<sup>8,19</sup> which is used as an additional *ff* term to improve RNA simulations. Even with this modification, description of the UNCG TL remained imbalanced and we were not capable to fold the 8-mer TL. Despite occasional literature claims reporting success in UNCG simulations in the past (not unambiguously confirmed by other groups;<sup>8,9</sup> reviewed in ref. 4), we are not aware of any existing RNA *ff* able to properly describe UNCG TL with all its signature interactions.<sup>19,23</sup> Some success has been recently achieved with the DESRES *ff* with specific ion parameters using the UNCG 14-mer with longer A-RNA stem.<sup>24</sup> However, the same *ff* leads to a complete disruption (including secondary structure rearrangements) of some other important RNA systems even in standard simulations on sub-microsecond timescale.<sup>19,23</sup> Such side-effects in simulations of functional RNA molecules outweigh partial improvements for small models.<sup>19,23</sup> Thus, the UNCG *ff* issue remains unresolved.

In the present work, we analyze origins of the notoriously problematic description of the UNCG TL. First, we carry out a detailed analysis of our earlier and new MD simulations and obtain a detailed picture of the sequence of events that leads to loss of the native state. We identify molecular interactions, backbone substates and other conformational features that are involved in the process. Then, we try to unravel specific *ff* deficiencies using massive quantum mechanical/molecular mechanical (QM/MM)<sup>25</sup> and QM calculations. Obviously, as discussed elsewhere,<sup>26-28</sup> QM and QM/MM calculations cannot replace MD simulations due to the lack of Boltzmann sampling. There is generally a limited transferability between calculations of

potential-energy surfaces (gradient optimizations or energy scans) as done by QM methods and free energies that determine populations in MD simulations. Still, careful comparison of accurate-enough QM and QM/MM calculations with equivalent *ff* data provides substantial insights into the imbalances that are inherent to *ff*s. This is the goal of our QM/MM and QM computations. Finally, we try to improve the simulation behavior by some carefully-tailored interventions. Our work clearly indicates that poor behavior of the UNCG simulations is a complex issue and cannot be explained by one single dominant factor that would be straightforwardly correctable. Instead, there is a concerted effect of multiple *ff* inaccuracies which are coupled and magnifying each other's effect. Such accumulation of diverse inaccuracies in description of a short UNCG RNA sequence that in addition folds into a very stiff structure provides no conformational freedom to buffer the individual imbalances. It leads, even for moderate imbalances, to substantial structural strains.



**Figure 1.** Native structure of the UNCG TL. Nucleotides of the loop are labelled as  $U_{L1}$ ,  $U_{L2}$ ,  $C_{L3}$  and  $G_{L4}$ , and stem nucleotides are labeled as  $A_{S-2}$ ,  $C_{S-1}$ ,  $G_{S+1}$  and  $U_{S+2}$ . Phosphates and bases are identified by normal and bold font, respectively. We note that r(acUUCGgu) and r(gcUUCGgc) 8-mers were studied in this work. Signature features of UNCG TL are colored in red and include the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond (marked as “1” in the Figure),  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond (2),  $U_{L1}(O2)\dots G_{L4}(N1)$  H-bond (3),  $C_{L3}(N4H)\dots U_{L1}(pro-Rp)$  H-bond (7BPh interaction<sup>29</sup>) (4), *syn*-conformation of  $G_{L4}$  nucleotide ( $\chi_{syn}$ ) and C2'-*endo* pucker of  $U_{L2}$  and  $C_{L3}$  ribose. Other key structural features that were investigated in this work include: the  $C_{L3}/G_{L4}$  ribose-base stacking (5),  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  0BPh interaction (6), mutual position of  $G_{L4}$  and  $G_{S+1}$  nucleobases (7), and  $\beta_{trans}$  (orange) and  $\gamma_{g+}$  (violet) conformations of  $G_{S+1}$  nucleotide.

## METHODS

**Starting structures.** The initial coordinates were taken either from high-resolution NMR structure of the r(ggcacUUCGgugcc) 14-mer (PDB ID 2KOC<sup>30</sup>) or excised from the 2.8 Å-resolution X-ray structure (PDB ID 1F7Y<sup>31</sup>) as the r(gcUUCGgc) 8-mer. The starting topologies and coordinates for MD simulations were prepared using the tLEaP module of AMBER 16 program package.<sup>32</sup> Structure of the isolated small TL motif in solution is in dynamic

temperature-dependent equilibrium between folded and unfolded conformations. Most of the calculations were done using 8-mer oligonucleotides, resulting, upon correct folding, into a two-base-pair stem and the TL itself. As TLs with longer stems require higher temperature for melting (unfolding),<sup>33,34</sup> minimal 8-nucleotide long (8-mers) TL motifs with just two base pairs are commonly used in computational studies.<sup>6–10,18,19</sup> There is a clear experimental evidence that even these short oligomers should be dominantly in the folded state,<sup>35</sup> validating the 8-mer calculations.

**Classical MD simulations.** We used the standard *ff99bsc0* $\chi_{OL3}$ <sup>36–39</sup> AMBER RNA *ff* with the van der Waals (vdW) modification of phosphate oxygens developed by Steinbrecher et al.,<sup>40</sup> where the affected dihedrals were adjusted as described elsewhere.<sup>22</sup> This RNA *ff* is abbreviated as  $\chi_{OL3CP}$  henceforth and libraries can be found in the Supporting Information of ref. 8. In the simulations we in addition applied our gHBfix<sup>19</sup> H-bonding potential that was shown to improve overall performance of the RNA *ff*. In most cases we used its basic version from ref. 19, abbreviated subsequently as gHBfix19 in ref. 41. When any modified alternative is used, it is always specified in the text. Majority of simulations were solvated with the OPC<sup>42</sup> water model and in ~0.15 M KCl using the Joung–Cheatham<sup>43</sup> ion parameters. Specific tests involved TIP3P<sup>44</sup> and SPC/E<sup>45</sup> water models and were also performed in Na<sup>+</sup> net-neutral and 1.0 M KCl salt-excess ion environments. Other details about the MD protocols can be found in Section S1 in the Supporting Information. We have used also simulations from our preceding papers.<sup>19,23</sup>

**REST2 Settings.** The replica exchange solute tempering (REST2)<sup>46</sup> simulations of the r(gcUUCGgc) TL were performed at 298 K with either 12 or 16 replicas. Details about settings can be found elsewhere.<sup>8</sup> The scaling factor ( $\lambda$ ) values ranged from 1 to 0.59984 and from 1.0454 to 0.59984 for 12 and 16 replicas, respectively. Those values were chosen to maintain an exchange rate above 20%. The effective solute temperature ranged either from 298 K (12 replicas) or 285 K (16 replicas) to ~500 K. The hydrogen mass repartitioning<sup>47</sup> with a 4-fs integration time step was used.

**QM/MM, QM/COSMO, MM and MM/GB geometry optimizations of UNCG TL conformations.** We have carried out a series of gradient geometry optimizations in order to compare the MM (i.e., *ff*) description with advanced QM methods. The initial r(acUUCGgu) and r(gcUUCGgc) 8-mers were either taken from eight different MD snapshots (structures **1–8**) or excised from the NMR structure of the r(ggcacUUCGgugcc) 14-mer (PDB ID 2KOC, structure **9**). The counter-ions (K<sup>+</sup>) were added using the tLEaP module of AMBER 16 to neutralize the RNA. SPC/E water model sphere with radius ~40 Å from the geometrical center of the RNA was prepared using tLEaP for all the initial structures. Short water equilibrations (< 10 ps MD) of the added water molecules were performed before geometry optimizations for structures **1**, **2**, **3** and **4**. In contrast, five different water distributions were used for geometry optimizations of structures **5**, **6**, **7**, **8** and **9** obtained by 20 ps, 40 ps, 60 ps, 80 ps, and 100 ps-long solvent equilibrations. The resulting 25 starting structures are denoted as **5a–e**, **6a–e**, **7a–e**, **8a–e** and **9a–e**. Water distribution may influence the relaxed solute structures as shown in ref. 26. For the sake of simplicity, only results for structures after the 100 ps-long solvent equilibration (i.e., structures **5–9e**) are documented in the main text while the full set of data can be found in Supporting Information.

The additive QM/MM scheme with point-charge approximation for electrostatic embedding was used for all QM/MM calculations. The QM/MM module of AMBER 14<sup>48,49</sup> was coupled with TURBOMOLE V7.3<sup>50,51</sup> using an in-house modified version of Sander<sup>26,52,53</sup> from the AMBER 14 program package. For geometry optimization the limited-memory Broyden-

Fletcher-Goldfarb-Shanno quasi-Newton algorithm (L-BFGS)<sup>54</sup> was used with a convergence threshold of  $1 \text{ e}^{-4} \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-1}$  for the gradient norm. Turbomole was used to calculate energies and gradients of the QM region using the composite PBEh-3c<sup>55</sup> hybrid-DFT method with resolution-of-identity (RI) approximation for the Coulomb integrals.<sup>56</sup> The PBEh-3c method employs the double-zeta valence polarized def2-mSVP basis set, the empirical DFT-D3 dispersion correction<sup>57</sup> with Becke-Johnson (BJ) damping<sup>58</sup> and the geometrical counter-poise (gCP) correction for intra- and intermolecular basis set superposition error (BSSE).<sup>59</sup> The QM region included the whole RNA 8-mer while the MM region consisted of water molecules and counter-ions.

Similarly to the work by Pokorna et al.,<sup>26</sup> QM/MM optimizations were mirrored by equivalent MM optimizations from identical starting structures. For the sake of completeness we note that the *ff99bsc0* $\chi_{\text{OL3}}$  was used for structures **1-5** and **9** and  $\chi_{\text{OL3CP}}$  for structures **6-8**; the *ff* difference is, however, marginal for optimizations (see **Figure S1** in the Supporting Information). Geometrical parameters (backbone dihedrals, intramolecular distances and signature H-bonds) of the QM/MM- and MM-optimized TLs were analyzed using the CPPTRAJ<sup>60</sup> module of AMBER 16. Mutual orientation of nucleobase planes, i.e., angle between planes of the  $G_{L4}$  *syn* base and of the closing  $G_{S+1}$  base, and sugar-base stacking interaction were analyzed using the Geom\_util program ([https://github.com/hokru/geom\\_util](https://github.com/hokru/geom_util)).

Besides MM optimizations from the initial structures, we also performed MM re-optimizations starting from the QM/MM-optimized structures to reduce potential uncertainty due to comparison of too distinct local minima, i.e., a situation where the QM/MM and MM optimizations from the MD snapshot diverge to non-equivalent (too different) local minima.<sup>26</sup> The QM/MM to MM re-optimized geometry should typically stay close to the QM/MM-optimized geometry (see the Supporting Information). Therefore, comparing geometrical parameters between QM/MM-optimized and QM/MM to MM re-optimized geometries should conservatively show differences between QM and MM potential energy surfaces (PES), although the re-optimization may also mask MM imbalances by excessively attenuating the rearrangements, for detailed discussion see ref. 26.

RNA 8-mers were additionally optimized in implicit solvent. The Conductor-like screening model (COSMO)<sup>61</sup> and the Generalized Born model (GB)<sup>62,63</sup> were used for QM and MM, respectively, with dielectric constant  $\epsilon$  of 78.5.

**Relaxed  $U_{L1}(2'\text{-OH})$  dihedral scans of RNA TLs.** In order to analyze quality of the MM description of orientation of the key  $U_{L1}(2'\text{-OH})$  group, we compared QM and MM relaxed conformational energy scans of the  $U_{L1}(C1'\text{-}C2'\text{-}O2'\text{-}O2'\text{H})$  dihedral for structures **2**, **7**, **8** and **9**.

Firstly, the ONIOM scheme<sup>64</sup> was used for MM optimizations of the UNCG TL 8-mer (the high-layer) and solvent (the low-layer) using Gaussian 09.<sup>65</sup> As the same MM method ( $\chi_{\text{OL3CP}}$  RNA *ff*) was used for both layers the energy/gradient components of the high-layer calculations effectively cancel out. This approach results in MM geometry optimization with constraints using the microiterations approximation where water molecules are relaxed after each dihedral angle change. This is a technical workaround to enable efficient internal coordinate quasi-Newton optimizations necessary for constraints for an otherwise prohibitively large molecule. We mark this approach as MM/MM optimization throughout the paper.

To save computer time, the MM/MM-optimized geometries were used to initiate the corresponding QM/MM optimizations. The AMBER 14 coupled to Turbomole was used for QM/MM optimizations applying tight geometry restraints on the  $U_{L1}(C1'\text{-}C2'\text{-}O2'\text{-}O2'\text{H})$

dihedral (see the Supporting Information for details). The technical difference between using geometry constraints and restraints has no practical impact on the results.

**QM and MM calculations of small model systems.** We have also carried out calculations on various small model systems. We extracted specific structural moieties of the UNCG TL (see Section S2 in the Supporting Information for the list of studied small models) for comparative QM and MM analyses, i.e., local geometry optimizations, interaction energy scans and conformational energy scans. The *Xopt*<sup>66,67</sup> tool coupled with Turbomole was used for all the QM geometry optimizations performed at the PBEh-3c level using the RI approximation. Turbomole was also used for the rigid-monomer QM energy scans performed at the same level of theory. ORCA V4.1.0<sup>68,69</sup> was used for the reference (highest-quality) QM calculations performed at the DLPNO-CCSD(T)/TightPNO/CBS level using aug-cc-pV(T-Q)Z and corresponding auxiliary aug-cc-pVQZ/C basis sets, and with TightSCF setting.<sup>70-74</sup> MM geometry optimizations were performed using *Xopt* coupled with AMBER 14 or using Gaussian 09. QM and MM optimizations and single-point energy calculations of dinucleotide monophosphate models were performed with COSMO and GB implicit solvents ( $\epsilon = 78.5$ ). For rigid-monomers MM interaction energy scans, we used the development version of the *bff* program<sup>75</sup> using non-bonded energy terms of the AMBER *ff*. For the energy scans of small models, the MM method is denoted as AMBER *ff* and details are specified in the Supporting Information. Gaussian 09 was used for the MM relaxed conformational energy scan of the O3'-methylated-guanosine-monophosphate model. For additional computational details about calculations of the small models see Section S3 in the Supporting Information.

We used programs Molden,<sup>76</sup> VMD<sup>77</sup> and PyMOL<sup>78</sup> for visualizing and analyzing the structures.

## RESULTS AND DISCUSSION

We applied diverse computational techniques, namely standard MD simulations, enhanced sampling REST2 simulations, QM, MM and multiscale QM/MM calculations (see Methods) in order to understand limitations of current state-of-the-art AMBER RNA *ff*s in description of the UNCG TL. Firstly, we analyzed disruption (unfolding) paths from standard as well as enhanced sampling MD simulations. We identified common features that seem to be responsible for instability of the native topology, i.e., specific structural reconformations that precede the collapse of the loop. We then performed various QM/MM and MM calculations of the UNCG TL to analyze differences in backbone dihedrals, signature H-bonds and other structural features, as described by QM and MM, in order to assess potential *ff* problems. Furthermore, we applied QM and MM calculations to small model systems to study interactions that may be triggering structural changes within the loop. Finally, we proposed certain fixes, namely (i) modulation of the key 0BPh interaction, (ii) adjustment of the sugar-base signature H-bonds and (iii) reduction of the vdW radii of non-polar hydrogens, which were subsequently tested during series of enhanced sampling MD simulations. In the text we first describe in detail the loss of the TL's native structure in MD simulations and identify the individual potentially problematic energy terms. Then, we present the specific investigations visualizing and quantifying the *ff* limitations with aid of the QM methods. Finally, we describe modifications of the *ff* and their tests in simulations.

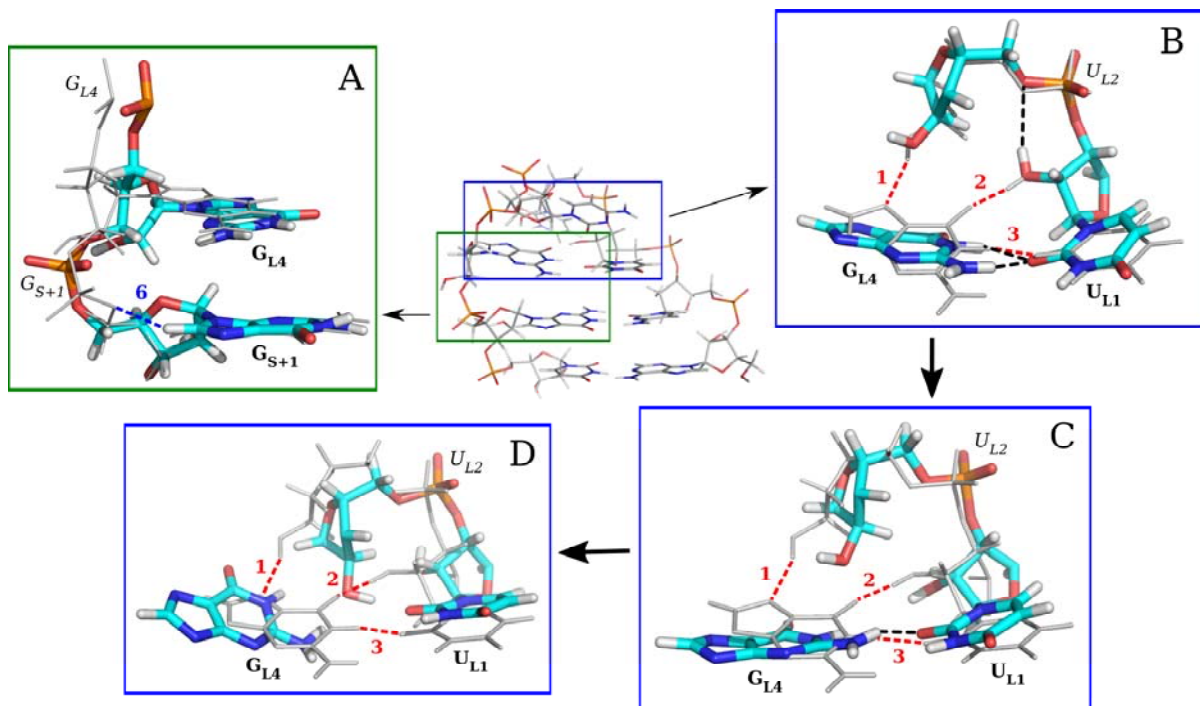
**Anatomy of disruption of the UNCG TL in simulations: a progressive loss of several interactions.** We inspected specific parts of trajectories from standard as well as enhanced sampling MD simulations of the UNCG TL from our previous work,<sup>19</sup> where signature interactions within the UNCG TL are lost and  $G_{L4}$  is repelled from its native conformation (**Figure 2**). Interestingly, disruption paths using multiple versions of the AMBER RNA *ff*<sup>19,36,38,39</sup> have common features and could be characterized as an interplay of several factors. Those factors involve interactions that are, apparently, wrongly described with current RNA *ffs*.

First important factor is conformation of the  $G_{S+1}$  phosphate that is forming (in its native state) a 0BPh interaction<sup>29</sup> (**Figure 2A**). The sugar-phosphate backbone around the  $G_{S+1}$  phosphate has a specific native conformation (see below) that is characterized by  $\beta$  and  $\gamma$  dihedrals in *trans* and *gauche+* (*g+*) orientations, respectively (**Figure 3**). The *ffs* rather prefer  $G_{S+1}$  phosphate in an alternative, i.e., flipped state, characterized by  $\beta_{g+}$  and  $\gamma_{trans}$  dihedrals. However, counterintuitively, disruptions of the UNCG TL are initiated from the native state of the  $G_{S+1}$  phosphate and not from the alternative (likely spurious) flipped conformation. We suggest that occurrence of the alternative backbone state is a consequence of structural strain in the native structure (in the *ff* description) which is reduced either by adopting the flipped backbone conformation, or by onset of the overall TL structure disruption.

The other structural factors are the underestimated stability and subsequent loss of both signature sugar-base H-bonds within the loop, i.e., the  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  and  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bonds. The 2'-OH groups in both H-bonds form alternative transient H-bonds during standard MD simulations. The former H-bond is reversibly lost due to the flip of  $U_{L1}(2'-OH)$  towards the  $U_{L2}$  phosphate,<sup>21</sup> so that a transient  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond is established (**Figure 2B**). The  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond is also reversibly lost and the  $U_{L2}(2'-OH)$  is then either pointing towards the solvent or alternative  $U_{L2}(2'-OH)\dots G_{L4}(O6)$  H-bond is formed. Importantly, formation of the alternative  $U_{L2}(2'-OH)\dots G_{L4}(O6)$  H-bond is not in disagreement with the experiment (see below).<sup>30</sup> We further realized that loss of the  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond is accompanied with structural rearrangement of the  $G_{L4}$  nucleobase.  $G_{L4}$  nucleotide (in its *syn* state) forms the GU wobble base pair with the  $U_{L1}$  nucleotide. Once the  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond is lost,  $G_{L4}$  makes a partial twist within the loop, where the signature base-base  $G_{L4}(N1H)\dots U_{L1}(O2)$  H-bond is weakened and a bifurcated  $G_{L4}(N1H/N2H)\dots U_{L1}(O2)$  H-bond is established instead (**Figure 2B**). This transient state is frequently observed in simulations though the signature  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond is typically reformed, followed by return of the  $G_{L4}$  nucleobase back to its canonical conformation. However, when the second sugar-base H-bond is lost simultaneously, i.e., neither the signature  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond nor the alternative  $U_{L2}(2'-OH)\dots G_{L4}(O6)$  H-bond are formed,  $G_{L4}$  has an increased tendency to flip out of the loop (**Figure 2C**). Consequently, another structural feature, i.e., sugar-base stacking between the  $G_{L4}$  nucleotide and the O4' atom of the  $C_{L3}$  ribose, is lost. In other words, the simultaneous loss of the signature interactions formed by both  $U_{L1}(2'-OH)$  and  $U_{L2}(2'-OH)$  groups opens the path for further destabilization of the structure, i.e., flip out of the  $G_{L4}$ . As mentioned above, the scenario when  $G_{L4}$  flips out of the loop requires the  $G_{S+1}$  phosphate in its native state.

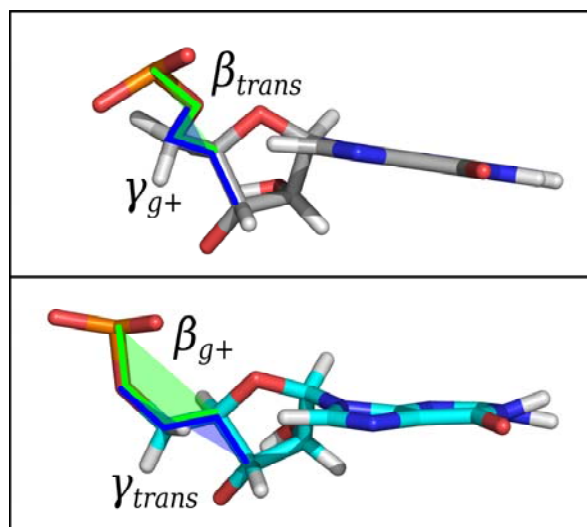
Once the  $G_{L4}$  is oriented towards the solvent (bulge-out state,<sup>6,79</sup> **Figure 2D**) it typically remains there, i.e., the disruption pathway appears to be irreversible within typical timescales of standard MD simulations (several  $\mu$ s). Nonetheless, recently published 10  $\mu$ s-long standard MD simulations with the  $\chi_{OL3CP} + gHBfix19$  *ff* version revealed occasional return of the  $G_{L4}$  nucleotide back from the bulge-out state to its native conformation within the loop.<sup>19</sup> We

identified the same flip-out flip-in pathway also in simulations with the DESRES<sup>24</sup> RNA *ff*.<sup>23</sup> Thus, the results indicate existence of a rather narrow conformational corridor (bottleneck) for expulsion and insertion of the G<sub>L4</sub>. It is worth noting that the tendency of expelling G<sub>L4</sub> out of the loop is higher when another signature interaction within the loop, i.e., the C<sub>L3</sub>(N4H)...U<sub>L1</sub>(*pro-R*<sub>P</sub>) base-phosphate (BPh) interaction type 7 (7BPh)<sup>29</sup> is weakened.



**Figure 2.** Disruption of the UNCG TL. The native structure of the 8-mer r(acUUCGgu) TL is shown in the middle with two important regions: (i) G<sub>L4</sub>-G<sub>S+1</sub> dinucleotide moiety (green) and (ii) tetranucleotide loop (blue). Insets **A-D** show only the specific fragments, where details about loss of important interactions or key conformational characteristics are highlighted. The overlaid native structure is depicted by grey lines in each of the insets. The important interactions 1-3 and 6 of the native UNCG TL are marked in a similar way as in **Figure 1** for each of the insets. The disruption pathway of the UNCG TL is depicted by insets **B**, **C** and **D** and its flow is indicated by arrows. All the interactions and nucleotides shown by the insets are labelled similarly to **Figure 1**. (**A**) G<sub>L4</sub>-G<sub>S+1</sub> dinucleotide with a flipped G<sub>S+1</sub> phosphate accompanied by loss of the OBPh interaction (6). (**B**) Part of the loop region, where the U<sub>L1</sub>(2'-OH) group is oriented towards the U<sub>L2</sub> phosphate forming the spurious U<sub>L1</sub>(2'-OH)...U<sub>L2</sub>(O5') H-bond. The partial twist of the G<sub>L4</sub> nucleobase leads to formation of the bifurcated G<sub>L4</sub>(N1H/N2H)...U<sub>L1</sub>(O2) H-bond. Signature H-bonds 1 and 2 (see **Figure 1**) are lost; the U<sub>L1</sub>(2'-OH)...U<sub>L2</sub>(O5') and G<sub>L4</sub>(N1H/N2H)...U<sub>L1</sub>(O2) H-bonds are depicted by black dashed lines. (**C**) Signature H-bonds (1-3) are lost and G<sub>L4</sub> is forming only the spurious G<sub>L4</sub>(N2H)...U<sub>L1</sub>(O2) H-bond (black dashed line). (**D**) G<sub>L4</sub> is further repelled from the loop and is not stabilized by any H-bond (bulge-out state). The common disruption pathway seen in simulation is documented as a movie in the Supporting Information.





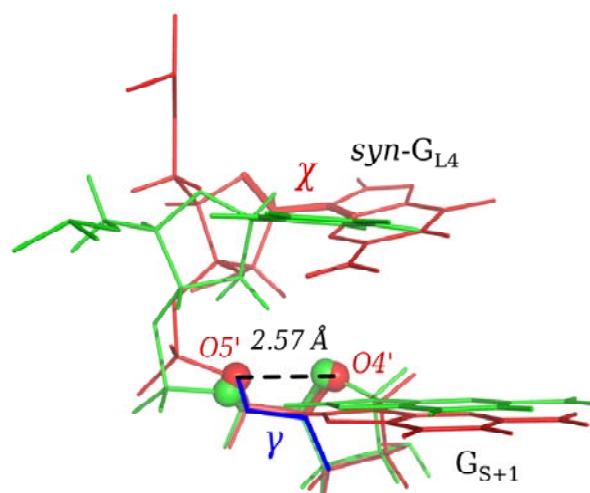
**Figure 3.**  $G_{S+1}$  phosphate in its native topology characterized by  $\beta$  and  $\gamma$  dihedrals in *trans* and *gauche+* ( $g+$ ) orientations, respectively (top), and in the flipped state characterized by  $\beta$  and  $\gamma$  dihedrals in  $g+$  and *trans* orientations, respectively (bottom).

**Relationship between the  $G_{S+1}$  phosphate conformation and the OBPh interaction.** As noted above, in simulations of folded UNCG TL the  $G_{S+1}$  phosphate fluctuates frequently between two substates. The  $G_{S+1}$  phosphate flipping is mainly characterized by the shift of  $\beta_{trans}/\gamma_{g+}$  dihedrals towards alternative  $\beta_{g+}/\gamma_{trans}$  conformations (**Figure 3**). The flips can be monitored by the RNA suite name code<sup>80</sup> as they change backbone conformations (RNA suites) from the canonical **1a** towards unusual **1e**, **1c**, or **unidentified** RNA suites. Before the loss of native UNCG TL topology the  $G_{S+1}$  phosphate flipping is entirely reversible. In these initial parts of simulations the alternative flipped state is prevalent (~70% population, **Figure S2** and **Figure S3** in the Supporting Information)<sup>19</sup> and is accompanied by loss of the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  OBPh interaction (**Figure 2A**).

OBPh is a weak C-H...O H-bond established by the C6-H6/C8-H8 groups of the pyrimidine/purine nucleotides which in most RNAs typically involves bridging O5' phosphate oxygen from the same residue).<sup>29</sup> In **1a** suite conformation of canonical A-RNA duplexes, O5' oxygens are often located below planes of the nucleobases. The unique structural context of UNCG TL characterized by the presence of neighboring  $G_{L4}$  nucleotide in *syn* conformation is causing a local structural change (deformation with respect to canonical **1a** suite of A-RNA duplexes) of the sugar-phosphate backbone around the  $G_{S+1}$  nucleotide (**Figure 4**). According to the sugar-phosphate backbone nomenclature,<sup>81</sup> the typical values of dihedrals around the  $G_{S+1}$  phosphate in its native UNCG conformation are on the edge of canonical **1a** suite definition (A-RNA like) with the  $\gamma$  dihedral shifted towards higher values of  $\sim 100^\circ$ , similarly to **1e** suite. The UNCG-specific conformation of the  $G_{L4}$ - $G_{S+1}$  sugar-phosphate moiety is under potential strain of two steric clashes of  $G_{S+1}(O5')$  oxygen with (i)  $G_{S+1}(C8-H8)$ , i.e., the above mentioned OBPh interaction, and (ii)  $G_{S+1}(O4')$  of ribose. The suspected steric clash of OBPh interaction can be easily relaxed in structural context of A-RNA duplex by slight elongation of C8H8/C6H6...O5' distances (**Figure S4** in the Supporting Information). However, in the specific structural context of UNCG TL caused by  $G_{L4}$  *syn* orientation, the suspected steric clash of OBPh interaction of  $G_{S+1}$  phosphate can be relaxed only at the expense of intensifying the other steric clash between

$G_{S+1}(O5')$  and  $G_{S+1}(O4')$  oxygens. As a consequence, the  $G_{S+1}$  phosphate is prone to flip towards the alternative (noncanonical)  $\beta_{g+}/\gamma_{trans}$  state, where the 0BPh  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  interaction is lost.

Unambiguous characterization of the  $G_{L4}$ - $G_{S+1}$  sugar-phosphate moiety is difficult even for experiments as there are only few nuclear Overhauser effect intensities (NOEs) and no  $^3J$ -scalar couplings available for detailed classification of this structural part.<sup>30</sup> We realized that some NOEs, namely  $G_{S+1}(H8) - G_{S+1}(H4')$  and  $G_{S+1}(H8) - G_{L4}(H3')$  are violated even for the NMR starting structure. Other violations (considering  $H5'$  and  $H5''$  protons) can be explained by possible neglect of spin diffusion in the NMR experiment. In the MD ensemble those parts of trajectories with native  $G_{S+1}$  phosphate conformation have better agreement with the experiment (for ten NOE signals around the  $G_{L4}$ - $G_{S+1}$  sugar-phosphate moiety) than those with flipped phosphate states except of two signals: (i)  $G_{L4}(H8)\dots C_{L3}(H5'')$  and (ii)  $G_{S+1}(H8)\dots G_{S+1}(H5'')$ . Nevertheless, although the preference of RNA *ffs* for the alternative  $G_{S+1}$  phosphate conformation is likely a *ff* artifact, the experimental uncertainties around the  $G_{L4}$ - $G_{S+1}$  moiety may indicate some real dynamics in this key part of the UNCG TL.



**Figure 4.** Local deformation of the sugar-phosphate backbone around the  $G_{S+1}$  phosphate caused by the *syn* conformation of  $G_{L4}$  nucleotide (red) and comparison with the canonical **1a** suite of A-RNA duplex (green). Both superimposed fragments are taken from the starting structure of the r(ggcacUUCGgugcc) 14-mer (PDB ID 2KOC<sup>30</sup>). The UNCG-specific conformation of sugar-phosphate backbone (in red) contains two close contacts of the  $G_{S+1}(O5')$  oxygen with (i)  $G_{S+1}(O4')$  and (ii)  $G_{S+1}(C8H8)$  groups, which might cause steric clashes under the *ff* description and then lead to the  $G_{S+1}$  phosphate flip. The  $\chi$  dihedral of the *syn*  $G_{L4}$  nucleotide,  $\gamma$  dihedral of the  $G_{S+1}$  nucleotide and the average experimental distance<sup>30</sup> between  $O5'$  and  $O4'$  atoms are highlighted.

**The 0BPh interaction is incorrectly described by the *ff*.** We analyzed the 0BPh interaction using electronic structure calculations. We compared QM/MM and MM behavior of two distances that characterize the 0BPh interaction, namely the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  and  $G_{S+1}(O4') - G_{S+1}(O5')$  interactions. We took the UNCG 8-mer TLs, where the  $G_{S+1}$  phosphate maintained its native geometry (starting structures **6a-e**, **7a-e** and **9a-e**, see Methods). Fifteen gradient geometry optimizations carried out from these starting structures showed that the RNA *ff* prefers longer distances in comparison with QM by  $\sim 0.13$  Å and  $\sim 0.05$  Å for the

$G_{S+1}(C8H8)\dots G_{S+1}(O5')$  and  $G_{S+1}(O4') - G_{S+1}(O5')$  interactions, respectively (**Table 1**, **Table S1** and **Table S2** in the Supporting Information). Additionally, we used dinucleotide monophosphate model (**Figure S5** in the Supporting Information) possessing the *native*  $G_{S+1}$  phosphate conformation and analyzed its QM and MM-optimized geometries. Again, both the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  and  $G_{S+1}(O4') - G_{S+1}(O5')$  distances are longer (by  $\sim 0.26$  Å and  $\sim 0.10$  Å) in MM compared to the QM description (**Table 1** and **Table S3** in the Supporting Information).

We then performed MM interaction- and conformational-energy scans of the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  distance using two models: (i) a smaller dimethyl-phosphate – methyl-guanine molecular complex lacking the  $G_{S+1}(O4') - G_{S+1}(O5')$  contact, and (ii) a larger O3'-methylated-guanosine-monophosphate intramolecular model, which included the  $G_{S+1}$  ribose with its O4' oxygen atom and thus takes into account the possible repulsion between  $G_{S+1}(O4')$  and  $G_{S+1}(O5')$  atoms (**Figures S6A** and **S6B** in the Supporting Information). The smaller model showed prolonged MM  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  distance (by 0.11 Å) in comparison with QM (**Figure 5**). The difference doubles to 0.22 Å for the larger model (**Figure S7** in the Supporting Information). We suggest that the increased difference (by 0.11 Å) obtained for the larger model is due to the  $G_{S+1}(O4') - G_{S+1}(O5')$  repulsion. In addition, the small-model calculations reveal that the PBEh-3c QM method used for the large-scale QM/MM calculations overestimates the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  H-bond distance by 0.08 Å with respect to the QM reference (**Figure 5**). In other words, the identified overestimation of short-range by the MM is even larger than indicated by the above comparison between PBEh-3c QM/MM and MM data.

In summary, the 0BPh interaction is not properly described by the AMBER RNA *ffs*. Both  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  and  $G_{S+1}(O4') - G_{S+1}(O5')$  contacts contribute to the imbalance. The optimal interatomic distance and mainly the short-range repulsion (**Figure 5**) are severely overestimated. We propose that there is enough structural flexibility to overcome this inaccuracy for the canonical A-RNA duplex conformation characterized by the **1a** suite. However, the discrepancy becomes unmasked in the restricted conformational space within the UNCG TL formed by the  $G_{L4}$  *syn* nucleotide /  $G_{S+1}$  phosphate motif. The overestimated repulsion in the MM approximation forces the  $G_{S+1}$  phosphate into the flipped states seen during the MD simulations.

**Table 1.** Comparison of the QM and MM description of the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  and  $G_{S+1}(O4') - G_{S+1}(O5')$  distances (in Å) within UNCG 8-mers and dinucleotide monophosphate models.<sup>a</sup>

UNCG TL	$G_{S+1}(C8H)\dots G_{S+1}(O5')$			$G_{S+1}(O4') - G_{S+1}(O5')$		
	initial value	$\Delta d_{QM/MM}^b$	$\Delta d_{MM}^c$	initial value	$\Delta d_{QM/MM}^b$	$\Delta d_{MM}^c$
Structure <b>6e</b>	2.18	-0.09	0.10	2.74	0.04	0.09
Structure <b>7e</b>	2.39	0.00	0.09	2.70	0.05	0.13
Structure <b>9e</b>	2.15	0.06	0.15	2.57	0.05	0.11

Dinucleotide monophosphate	$G_{S+1}(C8H)\dots G_{S+1}(O5')$			$G_{S+1}(O4') - G_{S+1}(O5')$		
	initial value	$\Delta d_{QM/COSMO}^d$	$\Delta d_{MM/GB}^e$	initial value	$\Delta d_{QM/COSMO}^d$	$\Delta d_{MM/GB}^e$
<i>planar-native</i>	2.37	-0.14	0.13	2.76	-0.05	0.11
<i>tilted-native</i>	2.18	0.04	0.26	2.74	0.00	0.06

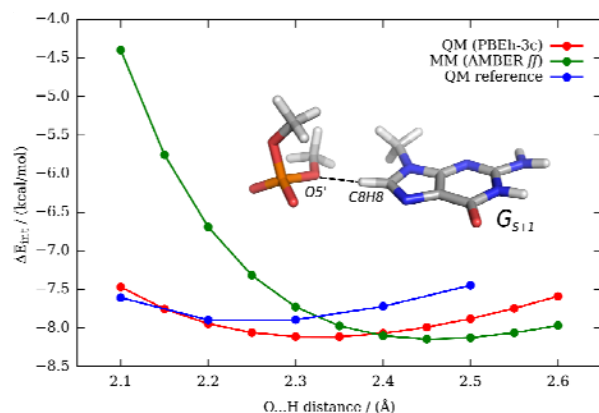
<sup>a</sup> See Methods for description of structures. UNCG TLs (structures **6**, **7** and **9**) and dinucleotide monophosphate models (*planar-native*, *tilted-native*) possessing the native conformation of  $G_{S+1}$  phosphate were chosen for the analysis. UNCG TLs were optimized by QM/MM or MM while the dinucleotide monophosphate models were optimized with geometry restraints by QM/COSMO or MM/GB. For UNCG 8-mer, values are shown only for structures **6e**, **7e**, and **9e**. See Section S3 in the Supporting Information for the explanation and for complete data, i.e., for 8-mer structures with different water distributions (**a-e**) and for dinucleotide monophosphate models optimized without geometry restraints (**Tables S1, S2 and S3**).

<sup>b</sup>  $\Delta d_{\text{QM/MM}} = \text{QM/MM value} - \text{initial value}$ ; i.e., change of the distance upon optimization

<sup>c</sup>  $\Delta d_{\text{MM}} = \text{MM value} - \text{initial value}$

<sup>d</sup>  $\Delta d_{\text{QM/COSMO}} = \text{QM/COSMO value} - \text{initial value}$

<sup>e</sup>  $\Delta d_{\text{MM/GB}} = \text{MM/GB value} - \text{initial value}$



**Figure 5.** Comparison of the QM (PBEh-3c; red), MM (AMBER *ff* – see Methods; green) and QM reference (DLPNO-CCSD(T)/CBS; blue) interaction energy profiles along the  $G_{S+1}(\text{C8H8})\dots G_{S+1}(\text{O5}')$  H...O distance of the dimethyl-phosphate – methyl-guanine model. Interaction energy scan was initiated from the  $G_{S+1}$  conformation found in the UNCG TL. Optimal distances of 2.32 Å, 2.46 Å and 2.24 Å for QM, MM and QM reference, respectively, show overestimation of the optimal  $G_{S+1}(\text{C8H8})\dots G_{S+1}(\text{O5}')$  distance by MM. Note that the MM description not only overestimates the optimum H-bond distance, but also sharply exaggerates the steepness of the short-range repulsion, which is a hallmark artifact of the 6-12 Lennard-Jones potentials when describing close interatomic contacts and steric clashes.<sup>82</sup> Since the computations are done in the gas phase and the system is non-neutral, the magnitude of the overall attraction is overestimated for all three methods compared to the same interaction in its complete environment. However, the relative positions of the curves are expected to be unaffected by this simplification.

**Relationship between the  $G_{S+1}$  phosphate conformation and positioning of the  $G_{L4}$  nucleotide.** In MD simulations, the angle between planes of the  $G_{L4}$  *syn* base and of the closing  $G_{S+1}C_{S-1}$  base pair fluctuates in the range of  $\sim 10^\circ$  to  $\sim 30^\circ$ . We call the arrangements at the edges of the range as *planar* and *tilted* positions of  $G_{L4}$ . Both positions appear consistent with the NMR data. We have carried out QM analyses of the positioning of the  $G_{L4}$  *syn* base, which indicate that the *native* state of the  $G_{S+1}$  phosphate could excessively force the  $G_{L4}$  nucleotide to the *planar* conformation during MD simulations. This conformation is prone to weakening of the signature  $U_{L2}(2'-\text{OH})\dots G_{L4}(\text{N7})$  sugar-base H-bond, which may contribute to problems in the MD description. However, we were not able to further quantify this potential MM imbalance due

to the uncertainties inherent to the used computations. The full description is available in the Supporting Information.

**The signature  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond is imperfectly described by *ff*s.** The  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  sugar-base H-bond is another player in the overall UNCG balance. According to the structural data, it could alternate with the  $U_{L2}(2'-OH)\dots G_{L4}(O6)$  H-bond.<sup>30</sup> Melting experiments with various substitutions at the  $U_{L2}(2'-OH)$  position (preventing the H-bonding) suggested its only negligible effect on the thermodynamic stability of the loop.<sup>83</sup> Analysis of 10  $\mu$ s-long standard MD simulations<sup>19</sup> revealed that the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond is not very stable, since the  $U_{L2}(2'-OH)$  is pointing either towards the solvent or is too far from both  $G_{L4}(N7)$  and  $G_{L4}(O6)$  acceptors in  $\sim 25\%$  of all snapshots with the native TL arrangement. When the  $U_{L2}(2'-OH)$  is interacting with  $G_{L4}$  nucleobase, the native  $G_{L4}(N7)$  acceptor is preferred over the  $G_{L4}(O6)$  ( $\sim 75\%$  and  $\sim 25\%$ , respectively). In addition, even when formed, the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond shows poor directionality, with the O-H group pointing above the G aromatic moiety rather than directly to the N7 atom. This may be sign of a local imbalance of the *ff* description and understabilization of the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  interaction. Importantly, loss of the signature  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond allows reformation of the  $G_{L4}$  nucleobase within the loop and formation of the bifurcated  $G_{L4}(N1H/N2H)\dots U_{L1}(O2)$  H-bond. It is worth mentioning that the bifurcated  $G_{L4}(N1H/N2H)\dots U_{L1}(O2)$  H-bond is in agreement with some X-ray structures,<sup>21</sup> however, in simulations its formation is typically followed by flip of the  $G_{L4}$  nucleobase out of the loop and subsequent disruption of the native state (**Figure 2B**). The  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  dynamics is thus directly kinetically related to loss of the native state.

We employed a set of QM calculations to investigate this signature sugar-base H-bond. Initially, we compared donor-acceptor distances and directionality of the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond in QM/MM- and MM-optimized UNCG 8-mers (**Tables S5** and **S6** in the Supporting Information). We observed that although the QM and MM donor-acceptor distances are similar (**Table S5**) the O-H...N angle is smaller (by  $\sim 18^\circ$  in average) in MM compared to QM (**Table 2**, and **Table S6** and **Figure S9** in the Supporting Information). Thus, we designed a small model (composed of  $G_{L4}$  base and  $U_{L2}$  ribose; **Figures S6C** and **S6D** in the Supporting Information) and performed interaction energy scans of the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond. We used two different values of the O-H...N angle, i.e. (i)  $\sim 170^\circ$  (almost ideal angle for the H-bond), and (ii)  $\sim 143^\circ$  (mimicking the H-bond often seen during MD simulations). MM minimum of the interaction energy profiles is slightly shifted to shorter donor-acceptor distances (by  $\sim 0.15$  Å) and is slightly narrower in comparison with the broad QM minimum (**Figure 6**). Thus, any enforced prolongation of the O...N distance by e.g. structural dynamics of other moieties within the loop would result in excessive weakening of the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond under MM description. In addition, the *ff* description obviously neglects the stabilizing effect of X-H bond stretching; this point is not analyzed by our calculations since the QM scan was done with rigid monomers, but it has been discussed elsewhere.<sup>26,53</sup>

More importantly, comparison of QM and MM PES's revealed that the effect of loss of directionality of this H-bond, i.e. comparing the almost ideal H-bond angle ( $\sim 170^\circ$ ) with the lowered one ( $\sim 143^\circ$ ) is underestimated (by  $\sim 0.7$  kcal/mol) by the MM method (**Figure 6**). Although MM still prefers (in the isolated small system) a H-bond with good directionality, the penalty for making the H-bond less directional is lower at the MM level compared to the QM level. Therefore, if some other energy terms in the UNCG TL profit from non-planarity of this

H-bond due to the overall conflict among some other energy terms, the MM description may be excessively submissive to such strains.

Next, we used the  $U_{L2}$  ribose model and analyzed behavior of the C2'-O2'-O2'H angle and C1'-C2'-O2'-O2'H dihedral by QM and MM optimizations (**Figure S10** in the Supporting Information). We performed two conformation energy scans using the model: (i) C2'-O2'-O2'H angle bending and (ii) rotation of the C1'-C2'-O2'-O2'H dihedral. Both QM and MM conformational scans were comparable for the bending of the C2'-O2'-O2'H angle (**Figure S11** in the Supporting Information) and free optimizations showed a slightly shifted minimum (by 2.8°) between the QM and MM levels of theory (**Table S7** in the Supporting Information). In contrast, conformational energy scans of the C1'-C2'-O2'-O2'H dihedral revealed differences especially around the region of interest, where the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond could be established (conformations with dihedral values between ~150° and ~210°, **Figure 7**). We identified a very flat QM minimum (~190°) that is completely absent on the MM potential energy surface (PES). Instead, the MM curve is steeply descending in this region until it finds a deeper minimum around 260° (**Figure 7** and **Table S7** in the Supporting Information). This result is in agreement with analysis published by Mladek and coworkers<sup>84</sup> using larger sugar-to-sugar models (see **Section S4** in the Supporting Information for more details).

We note that after the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  or the alternative  $U_{L2}(2'-OH)\dots G_{L4}(O6)$  H-bond is lost in MD simulations, the  $U_{L2}(2'-OH)$  group binds to another acceptor within the loop. Thus, it is meaningless to analyze a shift of the C1'-C2'-O2'-O2'H dihedral once the H-bond to  $G_{L4}$  is lost. However, QM/MM and MM optimizations of UNCG TL 8-mers revealed that the C1'-C2'-O2'-O2'H dihedral is larger in MM-optimized geometries (**Table S8** in the Supporting Information), which is possibly associated with non-planarity of the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond in MD simulations. Thus, all calculations indicate that both directionality and strength of the signature  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond are not optimally described by the *ff*, which may contribute to the common loss of this H-bond during MD simulations of UNCG TL.

Any large imbalance in description of the 2'-OH...N interactions could have major impact on RNA simulations, as these sugar-base H-bonds are abundant.<sup>85</sup> Further, correction of such imbalance would be difficult, due to lack of specific *ff* terms to affect directionality of H-bonds. Thus, we performed a comparable analysis for the two 2'-OH...N1(A) H-bonds in an RNA kink-turn. However, in contrast to the UNCG TL case, these H-bonds behave stably (see **Section S5**, **Table S9** and **Figure S12** in the Supporting Information for details). We thus suggest that the accuracy of the description of the 2'-OH...N interactions is generally acceptable for simulations of majority of RNAs and the moderate imbalance towards non-planarity is unmasked for the UNCG TL due to its uniquely tight conformation.

**Table 2.** Description of directionality of the signature  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond (O-H...N angle in °) in QM/MM and MM optimizations of the whole TL 8-mers.<sup>a</sup>

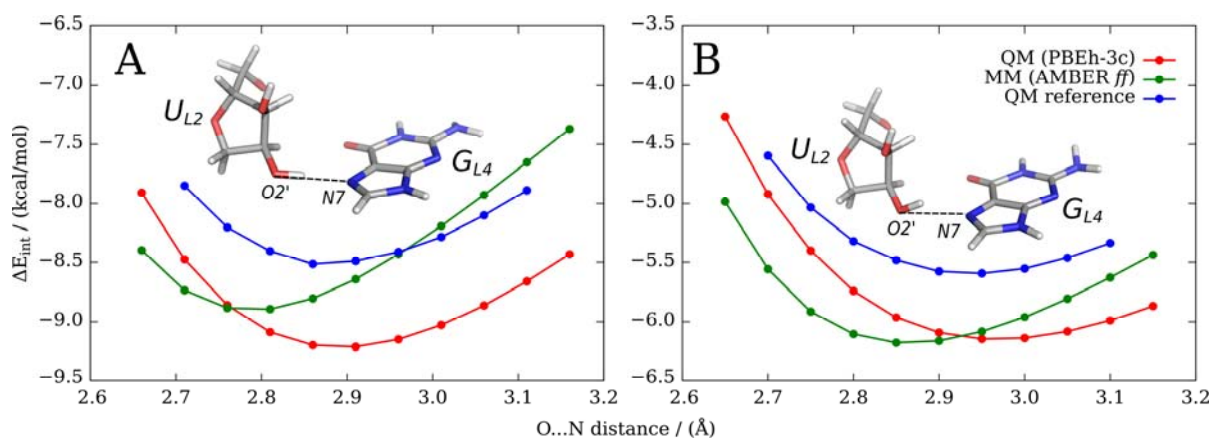
structure	initial angle	$\Delta\text{angle}_{\text{QM/MM}}^b$	$\Delta\text{angle}_{\text{MM}}^c$
<b>1</b>	143.4	16.1	5.2
<b>2</b>	118.1	45.2	20.8
<b>4</b>	140.8	14.4	-12.6
<b>5e</b>	161.7	9.3	2.5
<b>6e</b>	157.0	5.5	-7.0
<b>9e</b>	143.5	25.6	-0.9

<sup>a</sup> structure **3** was excluded because the  $U_{L2}(2'-OH)$  forms an alternative H-bond with  $G_{L4}(O6)$  instead of  $G_{L4}(N7)$  atom. Structures **7** and **8** were excluded because they contain spurious state of the  $U_{L1}(2'-OH)$  group, which may influence position of the  $G_{L4}$  nucleobase and subsequently the  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond. Values are

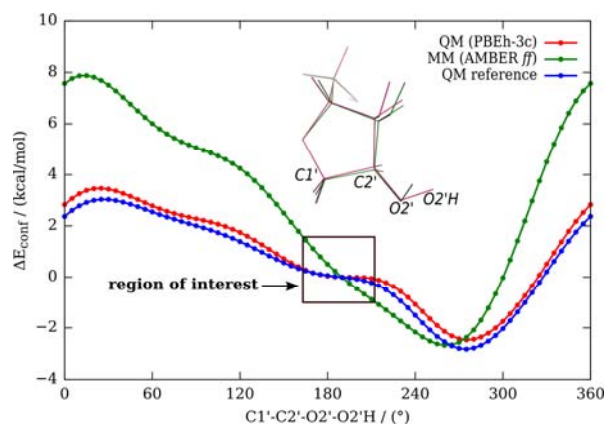
shown only for structures **5e**, **6e**, and **9e** (see Table S6 in the Supporting Information for complete data, i.e., for 8-mer structures with different water distributions **a-e**).

<sup>b</sup>  $\Delta\text{angle}_{\text{QM/MM}} = \text{angle}_{\text{QM/MM}} - \text{angle}_{\text{initial}}$ ; i.e., change of the angle upon optimization

<sup>c</sup>  $\Delta\text{angle}_{\text{MM}} = \text{angle}_{\text{MM}} - \text{angle}_{\text{initial}}$



**Figure 6.** Interaction energy scans of the  $U_{L2}(2'-\text{OH})\dots G_{L4}(N7)$  H-bond using QM (PBEh-3c; red), MM (AMBER *ff*; green) and QM reference (DLPNO-CCSD(T)/CBS; blue) methods. Ribose is shifted along the O...N vector in a distance range from 2.65 Å to 3.15 Å with the step length of 0.05 Å. All remaining degrees of freedom were frozen (see the Supporting Information for details). **(A)** The O-H...N angle is 170° at start of the scan. **(B)** The O-H...N angle is initially 143°.



**Figure 7.** Conformation energy profile of the  $C2'-\text{endo}$   $U_{L2}$  ribose  $C1'-C2'-O2'-O2'H$  dihedral using QM (PBEh-3c; red), MM (AMBER *ff*; green) and QM reference (DLPNO-CCSD(T)/CBS; blue) methods. Zero energy is set for a dihedral value of 190° (**Figure S10** in the Supporting Information). Region of interest where the  $U_{L2}(2'-\text{OH})\dots G_{L4}(N7)$  H-bond could be established is highlighted by the black rectangle. While QM finds flat region  $\sim 190^\circ$  MM continues steeply to the global minimum at  $\sim 260^\circ$ , where the intramolecular  $O2'H\dots O3'$  interaction is formed (**Table S7** in the Supporting Information). See Section S3 in the Supporting Information for computational details and Section S4 and ref. <sup>84</sup> for comparison with the earlier scans of the 2'-OH group.

**The  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  signature sugar-base H-bond is weakened by flips of the  $U_{L1}(2'-OH)$  towards the  $U_{L2}$  phosphate.** The signature  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  sugar-base H-bond is crucial for stability of the UNCG TL. It should be firmly established according to the structural data<sup>30</sup> and any substitution at the  $U_{L1}(2'-OH)$  position leads to a significant drop of the thermodynamic stability.<sup>83</sup> MD simulations show reversible loss of the  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond and transient formation of an alternative  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond, a known *ff* artifact that was firstly classified in ref. 21. We performed several standard MD simulations to understand this flip and its dependency on other factors. Although the signature  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond is favored, there is a 10-30% population of the competing  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond. The latter H-bond is typically short-living (~hundreds of ps) and the flips are entirely reversible. The flips were more frequent in SPC/E solvent in comparison with TIP3P and OPC water models (**Table S10** in the Supporting Information). Moreover, the flips were more common in 1.0 M KCl salt-excess ionic conditions than with 0.15 M KCl ion-strength and  $K^+$  net-neutral ionic conditions. The modified vdW parameters for phosphate oxygens<sup>40</sup> did not bring any improvement (**Table S10** in the Supporting Information).

We then investigated competition between the signature  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  and alternative  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond using QM methods. Firstly, we analyzed the QM/MM- and MM-optimized 8-mer TLs with native  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond and observed comparable QM and MM description of this H-bond (structures **1-5**, **6a-e** and **9a-e**). Then we analyzed structures **7a-e** and **8a-e** where the  $U_{L1}(2'-OH)$  group is flipped towards the  $U_{L2}$  phosphate. For the structure **7** having the *native* state of the  $G_{S+1}$  phosphate, we observed renewal of the signature  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond in all five (**a-e**) QM/MM optimizations, whereas in equivalent MM optimizations the  $U_{L1}(2'-OH)$  group remained flipped towards the  $U_{L2}(O5')$  (see the Supporting Information for starting and final structures). Note that the starting structures **a-e** used for optimizations differ only in solvent equilibration while the RNA configuration is identical (see Methods).

Subsequently, we performed relaxed MM/MM and QM/MM scans (see Methods for details) of the UNCG TL 8-mers along the  $C1'-C2'-O2'-O2'H$  dihedral to quantify the potential energy curves between the two conformations (**Figure 8**). We used MD snapshots with different states of the  $G_{S+1}$  phosphate and different orientations of the  $U_{L1}(2'-OH)$  group as starting points, namely: (i) *native* state of the  $G_{S+1}$  phosphate and *native* orientation of the  $U_{L1}(2'-OH)$  group (structure **9**), (ii) *native*  $G_{S+1}$  phosphate and *flipped*  $U_{L1}(2'-OH)$  group (structure **7**), (iii) *flipped* state of the  $G_{S+1}$  phosphate and *native* orientation of the  $U_{L1}(2'-OH)$  group (structure **2**), and (iv) *flipped*  $G_{S+1}$  phosphate and *flipped*  $U_{L1}(2'-OH)$  group (structure **8**).

For all four scans (**Figure 8**), the MM/MM calculation (the green curves) overstabilizes the structure with alternative  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond (the  $C1'-C2'-O2'-O2'H$  dihedral  $\sim 180^\circ$ ) with respect to the native  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond (the  $C1'-C2'-O2'-O2'H$  dihedral  $\sim 80^\circ$ ), relatively to the QM/MM data (red curves).

For UNCG TLs with native  $G_{S+1}$  phosphate conformation we observed that going from the native H-bond ( $\sim 80^\circ$ ) towards the alternative H-bond ( $\sim 180^\circ$ ), the QM approach did not find any second minimum for  $\sim 180^\circ$  (**Figure 8A**). In contrast, the alternative orientation of the  $U_{L1}(2'-OH)$  group towards  $U_{L1}(O5')$  was recognized as a local minimum by MM (**Figure 8A**). Starting from the MD snapshot with the alternative H-bond, both QM/MM and MM/MM approaches identified the alternative state as a minimum. However, while for the QM/MM scan the native orientation of the  $U_{L1}(2'-OH)$  group was the global minimum, the alternative orientation was the global minimum of the MM/MM scan (**Figure 8B**).

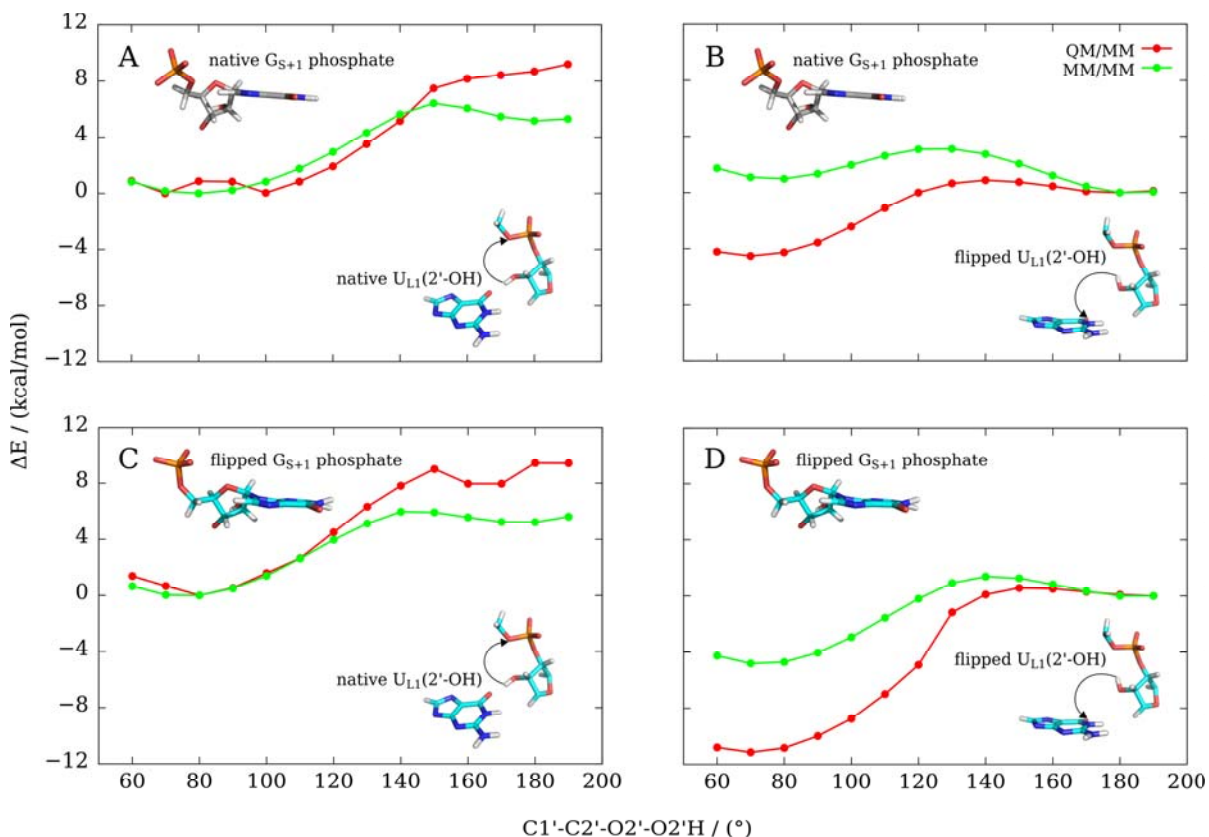


Energy scans with the  $G_{S+1}$  phosphate in the flipped conformation (shift of  $\beta/\gamma$  dihedrals and absence of the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  OBPh interaction) revealed the following picture. Starting from the native orientation of the  $U_{L1}(2'-OH)$  group both QM/MM and MM/MM approaches indicated that the global minimum is located near  $\sim 80^\circ$  and the alternative orientation has no local minimum (**Figure 8C**). However, the MM/MM curve clearly underestimates the energy difference. Starting from the alternative state of the  $U_{L1}(2'-OH)$  group, both QM/MM and MM/MM indicated the native orientation as the global minimum of the scan. However, the MM/MM curve is again substantially shifted in favor of the  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond relatively to the QM/MM curve (**Figure 8D**).

The MM/MM energy scans were performed with the modified vdW parameters for phosphate oxygens (CP). We thus performed the MM/MM energy scan for structure **8** also without the CP modification. The difference was negligible (**Figure S13** in the Supporting Information).

In summary, all PES scans showed that formation of the alternative  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond is overstabilized by the MM potential. In addition, in the simulations, the spurious and reversible flipping of the  $U_{L1}(2'-OH)$  group ( $\sim 180^\circ$ ) is connected with dynamics of the  $G_{L4-syn}$  nucleotide by preferring the planar  $G_{L4}$  conformation. It has, subsequently, a higher tendency to depart from the loop and collapsing it (see the paragraph above about the disruption of the UNCG TL).

We note that the energy scans should always be considered within the limits of sampling inherent to mere structure optimizations.<sup>26</sup> Detailed inspection of **Figure 8** shows that there is an apparent inconsistency of (especially) the MM curves between panels A – B and C – D, i.e., between scans with the same  $G_{S+1}$  phosphate conformation. This indicates that other (still unidentified) structural factors are probably involved. We explicitly inspected those mentioned in our earlier study,<sup>21</sup> i.e.,  $\epsilon(U_{L1})$ ,  $\epsilon(C_{S-1})$ ,  $\zeta(C_{S-1})$ , and  $\zeta(U_{L1})$  dihedrals as well as sugar pucker of the  $U_{L1}$  nucleotide but we did not find any clear structural changes along the PES curves. Thus, it appears that the flipped state of the  $U_{L1}(2'-OH)$  group might cause some structural adjustments distributed across many parameters that are difficult to identify (see the Supporting Information for initial and final structures from PES scans for comparison).



**Figure 8.** Comparison of the conformational energy profiles of the  $U_{L1}$  ribose along the  $C1'-C2'-O2'-O2'H$  dihedral for QM/MM (red) and MM/MM (green) methods. Initial structural orientations of the  $G_{S+1}$  phosphate and the  $U_{L1}(2'-OH)$  group are highlighted as structural insets and directions of scans are shown as black arrows.

**The  $C_{L3}(O4')$ ... $G_{L4}$  sugar-base stacking is slightly over-repulsive in AMBER *ff*.** Sugar-base stacking between  $C_{L3}(O4')$  and  $G_{L4}$  nucleobase belongs to the key characteristics of the ‘Z-like’ motif.<sup>86</sup> This contact represents a moderately attractive vdW interaction<sup>87-89</sup> which, however, appears to be rather imposed by the overall structural context than being formed on its own.<sup>89</sup> As it includes an exceptionally short distance between the  $O4'$  atom and the nucleobase ring, its MM description may be challenging, which may lead to destabilization of the position of the  $C_{L3}$  ribose above the  $G_{L4}$  nucleobase in MD simulations. Indeed, loss of the sugar-base stacking during MD simulations is associated with flip of the  $G_{L4}$  nucleobase out of the loop. Thus, we analyzed distance of the  $C_{L3}(O4')$  atom to the  $G_{L4}$  *syn* base plane in our QM/MM- and MM-optimized UNCG 8-mers. MM overestimates the  $O4'$  – guanine distance by  $\sim 0.04$  Å compared to QM/MM (Table S11 in the Supporting Information). Additional calculations using a small ribose-guanine sugar-base stacking model (Figure S6E in the Supporting Information) showed that MM prefers a larger optimal distance (by  $\sim 0.13$  Å; Figure S15) compared to QM (PBEh-3c). In summary, short-range repulsion of the sugar-base stacking is overestimated by MM, which likely also contributes to disruption of the UNCG loop during MD simulations, especially when assuming that the sugar-base stacking is a consequence of the overall complex topology of UNCG TL.

**Large-scale implicit solvent QM calculations are consistent with the QM/MM data.** In addition to QM/MM and MM calculations performed in the explicit solvent, we also tested the effect of implicit solvation for QM and MM geometry optimizations of the UNCG TL. We used the COSMO implicit solvent for the QM calculations which are compared with equivalent MM/GB calculations. The results obtained with implicit solvation are consistent with the above QM/MM and MM data in explicit solvent and are discussed in Section S3 of the Supporting Information.

**Attempts to improve the UNCG TL simulations.** We made several attempts to improve the UNCG simulations by simple corrections based on the above investigations. We tested REST2 folding simulations of the r(gcUUCGgc) 8-mer and standard simulations of the r(ggcacUUCGgugcc) 14-mer. The former test includes the complete folding landscape, but the simulations could be affected also by stability of the two-base-pair stem, which may be underestimated by the *ff*. For the 14-mer, we monitored transitions between the fully native conformation and  $G_{L4}$  bulged-out conformation.<sup>6,23,79</sup>

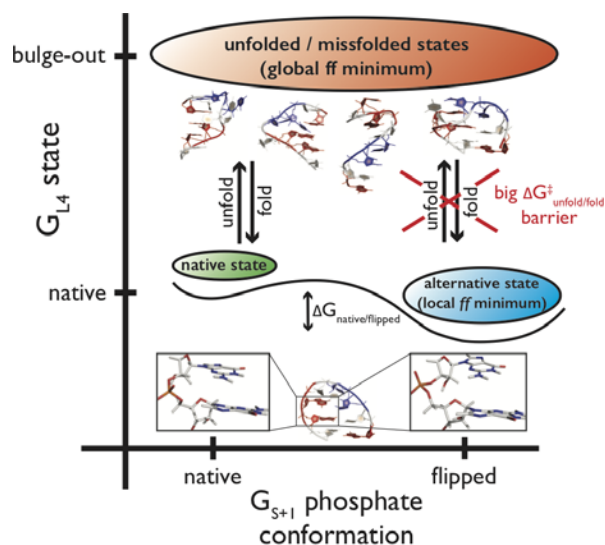
**NBfix correction of the 0BPh interaction maintains the  $G_{S+1}$  native phosphate conformation.** We first attempted to eliminate the  $G_{S+1}$  phosphate flipping. We modified the pairwise vdW parameters via breakage of the combination (mixing) rules by using the so-called nonbonded fix (NBfix).<sup>90</sup> We changed the minimum-energy distance of Lennard-Jones potential (i.e.,  $R_{ij}$  parameter) for the  $-H8...O5'$  pair (i.e., between H5 and OR atom types) to 2.8808 Å. The depth of the potential well ( $e_{ij}$  parameter) was kept at its default value of 0.0505 kcal/mol. In other words, we decreased the sum of vdW radii between H8 atoms of all purine bases and  $O5'$  oxygens of phosphates by 0.25 Å.

The effect of the NBfix correction was significant for both the interaction energy scan of the dimethyl-phosphate – methyl-guanine model (see paragraph The 0BPh interaction is incorrectly described by the *ff*) and MD simulations of the r(ggcacUUCGgugcc) 14-mer. For former case, we changed the  $G_{S+1}$   $-H8...O5'$  pair of a small model and it resulted in significant improvement of MM interaction energies with shift of the optimal distance towards QM level description (**Figure S16**). For the later, the NBfix correction was applied for all six purine bases of the r(ggcacUUCGgugcc) 14-mer. The  $G_{S+1}$  phosphate favored its native geometry compared to the flipped one (~80%/~20%, **Figure S17** in the Supporting Information). The NBfix correction shifts the  $G_{S+1}(C8H8)...G_{S+1}(O5')$  distance histograms towards shorter values (**Figure S18** in the Supporting Information). Thus, elimination of the clash stabilizes the 0BPh interaction and the native phosphate moiety position, as intended. However, it does not bring any overall help to the TL simulations. The disruption of the loop in standard simulations is even accelerated once the native  $G_{S+1}$  phosphate conformation is stabilized; in one of the three simulations  $G_{L4}$  left its binding pocket just after ~35 ns (**Figure S17** in the Supporting Information). Hence, elimination of the  $G_{S+1}$  phosphate flipping, which is partially easing the apparent conformational stress within the UNCG loop, is actually speeding up loss of the TL structure initiated by the  $G_{L4}$  *syn* nucleotide departure. As discussed above, the local flipped backbone conformation “sequesters”  $G_{L4}$  in the pocket away from the global unfolding pathway. This is a typical example of compensation of errors, in this case affecting lifetime of the folded state. Thus, despite the lack of improvement in simulations, we consider the modification of 0BPh interaction as correct *per se*.

**gHBfix alone is insufficient to fold the gcUUCGgc TL.** gHBfix potentials are general  $ff$  terms added to the basic  $ff$  for tuning stability of H-bonds. In our previous work we have shown that combining the basic  $\chi_{OL3CP}$   $ff$  with the gHBfix19 improves RNA simulations without side-effects, though it was still insufficient to fold the UNCG TL.<sup>19</sup>

We have thus tried several additional gHBfix settings to specifically improve folding of the r(gcUUCGgc) 8-mer. The most complex version strengthened base – base H-bonds and weakened sugar – phosphate interactions comparably to the gHBfix19 potential. In addition it also strengthened the sugar donor – base acceptor H-bonds (as they represent key signature H-bonds in the UNCG TL) and weakened base donor – sugar acceptor and sugar – sugar H-bonds. This version is abbreviated as gHBfix<sub>UNCG19</sub> (see Supporting Information for details). With gHBfix<sub>UNCG19</sub> we have been capable to see occasional UNCG folding events but the overall population of the folded TL 8-mer remained negligible. Full methods and results are summarized in the Supporting Information.

**Proposed pathway of the G<sub>L4</sub> departure and insertion.** Although the gHBfix<sub>UNCG19</sub> variant is not sufficient to correctly fold the r(gcUUCGgc) 8-mer, it gives interesting insights. We analyzed all available REST2 simulations of the r(gcUUCGgc) 8-mer as well as the r(ggcacUUCGgugcc) 14-mer standard simulations showing some capability of G<sub>L4</sub> to return back to the native position after its initial departure, i.e., from the bulge-out state<sup>6,79</sup> back to the native state.<sup>19,23</sup> We have found that the folding (and refolding) pathway requires a conformation, where the G<sub>S+1</sub> phosphate is in the native state. Thus, the identified  $ff$  imbalance between native and flipped conformation of the G<sub>S+1</sub> phosphate is kinetically stabilizing the broader native basin (including the flipped G<sub>S+1</sub> phosphate) once reached, as discussed in the previous section. However, it also hinders folding of the TL as the folding pathway goes through a bottleneck with the native G<sub>S+1</sub>, which is enthalpically disfavored by the  $ff$  due to steric clash associated with the OBPh interaction (**Figure 9**). This explains why attempts to tune solely the stability of H-bonds were insufficient to correct the free-energy imbalance between native and misfolded states of the r(gcUUCGgc) TL.



**Figure 9.** Proposed scheme describing folding and unfolding pathways of the r(gcUUCGgc) TL in MD simulations based on key conformational states of the G<sub>S+1</sub> phosphate and G<sub>L4</sub> nucleotide. The main conformational states are highlighted: (i) native state with both G<sub>L4</sub> and

$G_{S+1}$  phosphate in native conformation, (ii) alternative state (local minimum in  $ff$ ) with  $G_{L4}$  in native and  $G_{S+1}$  phosphate in flipped conformation, and (iii) global  $ff$  minimum including all misfolded and unfolded states. Both folding and unfolding pathways require the  $G_{S+1}$  phosphate in the native conformation. The diagram explains why eliminating just the flipped  $G_{S+1}$  phosphate structures does not help to improve the simulations. Representative native structure of the r(gcUUCGgc) TL is shown at bottom and both native and flipped  $G_{S+1}$  phosphate conformations are highlighted as insets. Some examples of misfolded structures often found in MD ensembles are also shown (top). G, C, and U nucleotides are in red, white and blue, respectively.

**Final set of REST2 simulations.** As an additional effort, we took the following  $ff$  version: the standard  $\chi_{OL3CP}$  RNA  $ff$  combined with the gHBfix<sub>UNCG19</sub> correction for H-bonds and NBfix correction for the  $-H8...O5'$ -atom-pair clash. In addition, we reduced the vdW radii of all non-polar H atoms (H1, H4, H5 and HA atoms) universally to 1.2 Å. We have tried this additional vdW modification as it seemed to us that these hydrogens may cause some steric conflicts in the native UUCG conformation, similarly to the OBPh interaction. The depths of the Lennard-Jones potentials were not modified. All these fixes were aimed to stabilize the OBPh interaction and further ease some possible – yet unidentified – steric clashes within the UNCG loop. We performed 20  $\mu$ s-long REST2 folding simulation of the r(gcUUCGgc) 8-mer (all replicas initiated from the completely unfolded single stranded structure). There was an increased probability of folding (**Figure S20** in the Supporting Information) though the overall population of the native state within the reference replica was still low. We obtained ~33% population of structures with correct stems and ~7% population of entirely correct native structure. This means that population of the correctly structured loop in the sub-ensemble with correct stem formation was ~20%. Thus, the introduced corrections increased the propensity of folding but were still not able to fully stabilize the native state within the reference replica. Nevertheless, it is probably the best folding simulation that has been achieved for the 8-mer UNCG TL system by any currently available  $ff$ .

As the last attempt, we carried out another REST2 simulation starting now from the native state (in all replicas), where we added to the above potential a restraint (see the Supporting Information for details) to the  $U_{L1}(C1'-C2'-O2'-O2'H)$  dihedral to prevent flipping of the  $U_{L1}(2'-OH)$  group. It aimed to stabilize the signature  $U_{L1}(2'-OH)...G_{L4}(O6)$  H-bond instead of the alternative  $U_{L1}(2'-OH)...U_{L2}(O5')$  H-bond. In this simulation, the native state was lost in all replicas after ~7.5  $\mu$ s (**Figure S21** in the Supporting Information) with no tendency for refolding, which may indicate that the dihedral restraint caused some undesired side-effects.

## CONCLUDING REMARKS

The compact and well-determined native conformation of the UNCG TL has served as a hallmark reminder of persisting problems of RNA  $ff$ s for the structural description of RNA molecules.<sup>4,6,8–11,19,21,79,91</sup> In this work we clarify why this RNA motif is so difficult for MD simulations.

We first inspected in detail the disruption (unfolding) paths of the UUCG motif during standard as well as enhanced sampling MD simulations and identified common rearrangements that precede disruption of the UUCG native topology. Collapse of the loop is characterized by a

gradual loss of several key structural features: (i)  $G_{L4}$ - $G_{S+1}$  phosphate moiety conformation, (ii)  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  0BPh interaction, (iii)  $C_{L3}$ - $G_{L4}$  sugar-base stacking interaction, and (iv) both  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  and  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  sugar-base H-bonds.

We then performed numerous QM/MM and high-level QM calculations that reveal diverse limitations of the empirical *ff* potential to properly describe the complex network of energy contributions shaping up the UNCG TL. Namely, we identified that both directionality and strength of the signature  $U_{L2}(2'-OH)\dots G_{L4}(N7)$  H-bond are not well described by the *ff*. The second signature  $U_{L1}(2'-OH)\dots G_{L4}(O6)$  H-bond is weakened by formation of the alternative  $U_{L1}(2'-OH)\dots U_{L2}(O5')$  H-bond, which is overstabilized by the *ff*. Flipping of the  $U_{L1}(2'-OH)$  group (from base to phosphate) is connected with moderate relocation of the  $G_{L4}$ -*syn* nucleotide within the binding pocket. Subsequently,  $G_{L4}$  has higher tendency to depart from the loop. Further tests indicate that the sugar-base stacking interaction also suffers from a problematic MM description.

We realized that collapse of the UUCG TL is connected with intrinsic dynamics of the  $G_{L4}$ - $G_{S+1}$  sugar-phosphate moiety. Experimental characterization of this region is ambiguous as there are only few direct experimental observables<sup>30</sup> and some NOEs are violated even in the NMR-based models. MD simulations show that there are two states of  $G_{S+1}$  phosphate, i.e., native and flipped one, characterized by the shift of  $\beta_{trans}/\gamma_{g+}$  dihedrals towards alternative  $\beta_{g+}/\gamma_{trans}$  conformations. Thus, it appears that there is some intrinsic dynamics within this key part of the UNCG TL, which could be an analogy to recent indication of dynamic behavior of the  $G_{L4}$ -*syn* nucleotide.<sup>79</sup> Nevertheless, preference of RNA *ffs* for the alternative  $G_{S+1}$  phosphate conformation is clearly a *ff* artifact caused by incorrect description of the  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  0BPh interaction. Its optimal interatomic distance and short-range repulsion are severely overestimated in *ffs*, forcing the  $G_{S+1}$  phosphate into the flipped state seen during the MD simulations.

Although being spurious, population of the flipped  $G_{S+1}$  substate kinetically decelerates the unfolding. This is because folding (and refolding) pathways transit through a conformation having the  $G_{S+1}$  phosphate in the native state. Hence, also folding pathways reaching the (artificial) local minimum involving the flipped  $G_{S+1}$  phosphate have to pass through the native state conformation, which is enthalpically disfavored by the overestimated  $G_{S+1}(C8H8)\dots G_{S+1}(O5')$  steric clash in AMBER RNA *ffs*. This is probably the reason why all our previous attempts of tuning solely the H-bond interactions were insufficient to correct the free-energy imbalance between native and misfolded states of the r(gcUUCGgc) TL.<sup>19</sup>

We attempted to fix the 0BPh clash by the NBfix correction, in order to stabilize the native  $G_{S+1}$  phosphate conformation. This, however, accelerated disruption of the loop in standard simulations. Nevertheless, coupling this correction with the gHBfix<sub>UNCG19</sub> potential specifically designed to support H-bonds important for the UNCG TL improved REST2 folding simulations of the r(gcUUCGgc) 8-mer. We achieved ~7% total population of the fully correct (both loop and stem) native structure while population of the correctly structured loop upon stem formation was ~20%. It is fair to admit that this result is still not fully satisfactory and was achieved by extended *ff* modifications that might cause side-effects for other RNA systems, which we did not test.

In conclusion, the UNCG TL remains a considerable challenge for atomistic simulations with classical *ffs*. We have identified a number of individual energy imbalances without claiming to achieve a complete coverage of the problems. Our data could serve as an important guidance for tuning of pair-additive as well as polarizable *ffs*.<sup>92,93</sup> However, when trying to design *ff*

corrections based on this knowledge, we still could not get a decisive improvement. There could be several reasons for this. For example, there could be additional imbalances at the stem/loop interface that remain to be identified. Additionally, the extent of how simulations are affected by the uncorrected sugar-base stacking interaction is unknown. Lastly, a more fundamental problem could be the simplicity of the *ff per se*. Even when we correctly identify the imbalances, it could be challenging to suggest their correction with simple MM terms that fail to mimic the true electronic structure effects. As the key result, our work reveals that the poor behavior of UNCG simulations cannot be explained by a singular dominant factor that would be straightforwardly correctable. There is a concerted effect of multiple and mutually coupled *ff* inaccuracies, acting in the context of the uniquely tight conformation of the UNCG TL, where the RNA chain does not provide a sufficient conformational flexibility to absorb or circumvent the individual imbalances. In addition, tuning the simulation performance of the UUCG TL is complicated due to the enormous computer demands of the folding simulations and difficulty to reach their quantitative convergence.

#### ASSOCIATED CONTENT

**Supporting Information.** The following files are available free of charge <https://pubs.acs.org>. Details about MD protocol, list of small models, additional computational details of small models, additional details about conformation scans of the 2'-OH group, computational results of the kink-turn Kt-7, details about large scale QM calculations in implicit solvent, relationship between  $G_{S+1}$  phosphate conformation and positioning of the  $G_{L4}$  nucleotide, additional results from folding of the gcUUCGgc TL using the gHBfix potential, description of the  $U_{L1}(C1'-C2'-O2'-O2'H)$  dihedral restraint, supporting tables and figures (PDF). Starting and final structures from QM/MM, MM optimizations and PES scans (PDF). Video showing the common UNCG disruption pathway during MD simulations (MP4).

#### AUTHOR INFORMATION

##### Corresponding Authors

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

#### ACKNOWLEDGMENT

This work was supported by Financial support from the Operational Programme Research, Development and Education, European Regional Development Fund project no. CZ.02.1.01/0.0/0.0/16\_019/0000754 [to M.O., P.B., P.K.] of the Ministry of Education, Youth and Sports of the Czech Republic, by Praemium Academiae (J.S.), Czech Science Foundation (20-16554S to M.K., V.M. and J.S., 18-25349S to P.B., P.K.), and by the project SYMBIT reg. number: CZ.02.1.01/0.0/0.0/15\_003/0000477 financed by the ERDF (H.K., J.S., P.P.).

#### ABBREVIATIONS

RNA, UNCG, tetraloop, molecular dynamics, force field, QM/MM

#### 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.* **43**, 40–47 (2010).
2. Cheatham, T. E. & Case, D. A. Twenty-five years of nucleic acid simulations. *Biopolymers* **99**, 969–977 (2013).
  3. Sponer, J. *et al.* Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *Journal of Physical Chemistry Letters* vol. 5 1771–1782 (2014).
  4. Sponer, J. *et al.* RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chem. Rev.* **118**, 4177–4338 (2018).
  5. Nerenberg, P. S. & Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **49**, 129–138 (2018).
  6. Kuhrova, P., Banas, P., Best, R. B., Sponer, J. & Otyepka, M. Computer folding of RNA tetraloops? Are we there yet? *J. Chem. Theory Comput.* **9**, 2115–2125 (2013).
  7. Haldar, S. *et al.* Insights into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics. *J. Chem. Theory Comput.* **11**, 3866–3877 (2015).
  8. Kuhrova, P. *et al.* Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **12**, 4534–4548 (2016).
  9. 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* **21**, 1578–1590 (2015).
  10. Bottaro, S., Banas, P., Sponer, J. & Bussi, G. Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J. Phys. Chem. Lett.* **7**, 4032–4038 (2016).
  11. 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.* **13**, 900–915 (2017).
  12. Varani, G. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379–404 (1995).
  13. Uhlenbeck, O. C. Tetraloops and RNA folding. *Nature* **346**, 613–614 (1990).
  14. Tuerk, C. *et al.* CUUCGG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1364–1368 (1988).
  15. Woese, C. R., Winker, S. & Gutell, R. R. Architecture of ribosomal RNA: Constraints on the sequence of ‘tetra-loops’. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8467–8471 (1990).
  16. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A. & Williams, L. D. Single nucleotide RNA choreography. *Nucleic Acids Res.* **34**, 1481–1491 (2006).
  17. Sorin, E. J., Engelhardt, M. A., Herschlag, D. & Pande, V. S. RNA simulations: Probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* **317**, 493–506 (2002).
  18. Chen, A. A. & Garcia, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **110**, 16820–16825 (2013).
  19. Kuhrova, P. *et al.* Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **15**, 3288–3305 (2019).
  20. Cesari, A. *et al.* Fitting Corrections to an RNA Force Field Using Experimental Data. *J. Chem. Theory Comput.* **15**, 3425–3431 (2019).
  21. Banas, P. *et al.* Performance of molecular mechanics force fields for RNA simulations:



- Stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849 (2010).
22. Mlynsky, V. *et al.* 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* **119**, 4220–4229 (2015).
  23. Kuhrova, P. *et al.* Correction to: Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **16**, 818–819 (2020).
  24. 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.* **115**, E1346–E1355 (2018).
  25. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).
  26. Pokorna, P., Kruse, H., Krepl, M. & Sponer, J. QM/MM Calculations on Protein-RNA Complexes: Understanding Limitations of Classical MD Simulations and Search for Reliable Cost-Effective QM Methods. *J. Chem. Theory Comput.* **14**, 5419–5433 (2018).
  27. Gkionis, K., Kruse, H. & Sponer, J. Derivation of Reliable Geometries in QM Calculations of DNA Structures: Explicit Solvent QM/MM and Restrained Implicit Solvent QM Optimizations of G-Quadruplexes. *J. Chem. Theory Comput.* **12**, 2000–2016 (2016).
  28. Ryde, U. How Many Conformations Need to Be Sampled to Obtain Converged QM/MM Energies? the Curse of Exponential Averaging. *J. Chem. Theory Comput.* **13**, 5745–5752 (2017).
  29. 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.* **37**, 4898–4918 (2009).
  30. Nozinovic, S., Fürtig, B., Jonker, H. R. A., Richter, C. & Schwalbe, H. High-resolution NMR structure of an RNA model system: The 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* **38**, 683–694 (2009).
  31. Ennifar, E. *et al.* The crystal structure of UUCG tetraloop. *J. Mol. Biol.* **304**, 35–42 (2000).
  32. Case, D. A. *et al.* Amber 2016. *Univ. California, San Fr.* (2016) doi:10.1021/ct200909j.
  33. Sheehy, J. P., Davis, A. R. & Znosko, B. M. Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* **16**, 417–429 (2010).
  34. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7287–7292 (2004).
  35. Abdelkafi, M. *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.* **14**, 579–593 (1997).
  36. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
  37. Wang, J., 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. Comput. Chem.* **21**, 1049 (2000).
  38. Perez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the

- description of  $\alpha/\gamma$  conformers. *Biophys. J.* **92**, 3817–3829 (2007).
39. Zgarbová, M. *et al.* Refinement of the Cornell *et al.* Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
  40. Steinbrecher, T., Latzer, J. & Case, D. A. Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **8**, 4405–4412 (2012).
  41. Mlynsky, V. *et al.* Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides. *J. Chem. Theory Comput.* **16**, 3936–3946 (2020).
  42. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building water models: A different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
  43. 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* **112**, 9020–9041 (2008).
  44. 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.* **79**, 926–935 (1983).
  45. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
  46. Wang, L., Friesner, R. A. & Berne, B. J. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115**, 9431–9438 (2011).
  47. 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.* **11**, 1864–1874 (2015).
  48. Case, D. A. *et al.* Amber14. *University of California, San Francisco* (2014) doi:10.1007/s13398-014-0173-7.2.
  49. Götz, A. W., Clark, M. A. & Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **35**, 95–108 (2014).
  50. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
  51. TURBOMOLE. V7.1, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>. (2018).
  52. Krepl, M. *et al.* 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.* **46**, 6528–6543 (2018).
  53. Pokorna, P., Krepl, M., Kruse, H. & Sponer, J. MD and QM/MM Study of the Quaternary HutP Homo-hexameric Complex with mRNA, 1-Histidine Ligand, and Mg<sup>2+</sup>. *J. Chem. Theory Comput.* **13**, 5658–5670 (2017).
  54. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**, 503–528 (1989).
  55. Grimme, S., Brandenburg, J. G., Bannwarth, C. & Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **143**, (2015).

56. Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **4**, 4285–4291 (2002).
57. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
58. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
59. Kruse, H. & Grimme, S. A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems. *J. Chem. Phys.* **136**, 154101 (2012).
60. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
61. Klamt, A. & Schuurmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* 799–805 (1993) doi:10.1039/P29930000799.
62. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **246**, 122–129 (1995).
63. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824–19839 (1996).
64. Dapprich, S., Komaromi, I., Byun, K. S., Morokuma, K. & Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.* **461–462**, 1–21 (1999).
65. Frisch, M. J. *et al.* Gaussian 09, Revision D.01. *Gaussian Inc.* Wallingford CT (2013).
66. Kruse, H. & Spöner, J. Towards biochemically relevant QM computations on nucleic acids: controlled electronic structure geometry optimization of nucleic acid structural motifs using penalty restraint functions. *Phys. Chem. Chem. Phys.* **17**, 1399–1410 (2015).
67. Kruse, H. Xopt. <https://github.com/hokru/xopt> (development version) (2017).
68. Neese, F. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 73–78 (2012).
69. Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, 4–9 (2018).
70. Riplinger, C., Pinski, P., Becker, U., Valeev, E. F. & Neese, F. Sparse maps - A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **144**, 024109 (2016).
71. Guo, Y. *et al.* Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **148**, (2018).
72. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).
73. Kendall, R. A., Dunning, T. H. & Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **96**, 6796–6806 (1992).
74. Weigend, F., Kohn, A. & Hattig, C. Efficient use of the correlation consistent basis sets in

- resolution of the identity MP2 calculations. *J. Chem. Phys.* **116**, 3175–3183 (2002).
75. Kruse, H. BrnoFF. <https://github.com/hokru/BrnoFF> (2017).
  76. Schaftenaar, G. & Noordik, J. H. Molden: A pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided. Mol. Des.* **14**, 123–134 (2000).
  77. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
  78. DeLano, W. L. The PyMol Molecular Graphics System, Version 2.0. *Schrodinger, Inc.: New York* (2017).
  79. Bottaro, S., Nichols, P. J., Vögeli, B., Parrinello, M. & Lindorff-Larsen, K. Integrating NMR and Simulations Reveals Motions in the UUCG Tetraloop. *Nucleic Acids Res.* 1–18 (2020) doi:10.1093/nar/gkaa399.
  80. Leontis, N. B. *et al.* The RNA Ontology Consortium: An open invitation to the RNA community. *RNA* **12**, 533–541 (2006).
  81. Richardson, J. S. *et al.* RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **14**, 465–81 (2008).
  82. Morgado, C. A., Jurecka, P., Svozil, D., Hobza, P. & Sponer, J. Balance of attraction and repulsion in nucleic-acid base stacking: CCSD(T)/complete-basis-set-limit calculations on uracil dimer and a comparison with the force-field description. *J. Chem. Theory Comput.* **5**, 1524–1544 (2009).
  83. Williams, J. D., Boots, J. L. & Hall, K. B. Thermodynamics of 2'-ribose substitutions in UUCG tetraloops. *RNA* **7**, 44–53 (2001).
  84. Mladek, A. *et al.* Energies and 2'-hydroxyl group orientations of RNA backbone conformations. Benchmark CCSD(T)/CBS database, electronic analysis, and assessment of DFT methods and MD simulations. *J. Chem. Theory Comput.* **10**, 463–480 (2014).
  85. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
  86. D'Ascenzo, L., Leonarski, F., Vicens, Q. & Auffinger, P. 'Z-DNA like' fragments in RNA: A recurring structural motif with implications for folding, RNA/protein recognition and immune response. *Nucleic Acids Res.* **44**, 5944–5956 (2016).
  87. Sponer, J., Gabb, H. A., Leszczynski, J. & Hobza, P. Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys. J.* **73**, 76–87 (1997).
  88. Chawla, M. *et al.* Occurrence and stability of lone pair- $\pi$  stacking interactions between ribose and nucleobases in functional RNAs. *Nucleic Acids Res.* **45**, 11019–11032 (2017).
  89. Kruse, H., Mrazikova, K., D'Ascenzo, L., Sponer, J. & Auffinger, P. Short but weak! The Z-DNA lone-pair... $\pi$  conundrum challenges standard carbon van der Waals radii. *Angew. Chemie Int. Ed.* (2020) doi:10.1002/anie.202004201.
  90. 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.* **12**, 430–443 (2016).
  91. 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.* **13**, 395–399 (2017).
  92. Zhang, C. *et al.* AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **14**, 2084–2108 (2018).

93. Lemkul, J. A. & MacKerell, A. D. Polarizable force field for RNA based on the classical drude oscillator. *J. Comput. Chem.* **39**, 2624–2646 (2018).