Understanding in-line probing experiments by modeling cleavage of nonreactive RNA nucleotides

VOJTĚCH MLÝNSKÝ and GIOVANNI BUSSI

Scuola Internazionale Superiore di Studi Avanzati (SISSA), 34136 Trieste, Italy

ABSTRACT

Ribonucleic acid (RNA) is involved in many regulatory and catalytic processes in the cell. The function of any RNA molecule is intimately related with its structure. In-line probing experiments provide valuable structural data sets for a variety of RNAs and are used to characterize conformational changes in riboswitches. However, the structural determinants that lead to differential reactivities in unpaired nucleotides have not been investigated yet. In this work, we used a combination of theoretical approaches, i.e., classical molecular dynamics simulations, multiscale quantum mechanical/molecular mechanical calculations, and enhanced sampling techniques in order to compute and interpret the differential reactivity of individual residues in several RNA motifs, including members of the most important GNRA and UNCG tetraloop families. Simulations on the multinanosecond timescale are required to converge the related free-energy landscapes. The results for uGAAAg and cUUCGg tetraloops and double helices are compared with available data from in-line probing experiments and show that the introduced technique is able to distinguish between nucleotides of the uGAAAg tetraloop based on their structural predispositions toward phosphodiester backbone cleavage. For the cUUCGg tetraloop, more advanced ab initio calculations would be required. This study is the first attempt to computationally classify chemical probing experiments and paves the way for an identification of tertiary structures based on the measured reactivity of nonreactive nucleotides.

Keywords: RNA; in-line probing; phosphodiester cleavage; QM/MM; metadynamics

INTRODUCTION

Ribonucleic acid (RNA) participates in several kinds of cellular processes, which involve the transmission of genetic information, the synthesis of proteins, cellular differentiation and development, the regulation of gene expression, and enzyme-like catalysis (Strulson et al. 2012; Kung et al. 2013; Sarkies and Miska 2014). Detailed information about RNA secondary structures is a preliminary step required for tertiary structure determination and, ultimately, for understanding RNA function (Walter 2009). Identification of specific small RNA motifs, like RNA tetraloops, is particularly important as they stabilize larger RNA structures and can be involved in RNA–RNA or RNA–protein interactions (Hall 2015). Indirect information about RNA secondary structure is usually obtained by chemical probing experiments (Xu and Culver 2009; Weeks 2010; Kubota et al. 2015). Among those, selective 2′-hydroxyl (2′-OH) acylation characterized by primer extension (SHAPE) (Merino et al. 2005) and in-line probing (Soukup and Breaker 1999) experiments provide sequence independent structural information on RNA at single-nucleotide resolution (Weeks 2010).

In-line probing characterizes backbone mobility by structural-dependent phosphodiester cleavage, which breaks RNA molecules into segments at distinct positions (Soukup and Breaker 1999). Unpaired nucleotides within single-stranded RNA regions are often unstable and degrade over time (Reynolds et al. 1996; Welch et al. 1997). The chemical reaction, termed as an internal transesterification, starts with a 2′-OH attack of neighboring phosphate moiety and results with 2′,3′-cyclic phosphate and 5′-hydroxyl termini (Soukup and Breaker 1999; Lilley 2003). The same RNA backbone cleavage (or RNA degradation) is performed by self-catalytic systems called RNA enzymes (ribozymes) (Doudna and Cech 2002; Scott 2007; Lilley and Eckstein 2008) and by ribonuclease A (RNase A) (Raines 1998), although in these cases with significantly higher rate constants. By comparing nucleotides from various RNA motifs, Soukup and Breaker (1999) observed a relation between the cleavage rate constant and the in-line attack angle of the scissile phosphate, i.e., the angle between O2′ oxygen, the adjacent phosphorus, and O5′ oxygen. Later, they defined the ability to

bring the active site toward the in-line attack conformation (the in-line attack angle close to 180°) as one of the catalytic strategies for the phosphodiester backbone cleavage used by ribozymes (Breaker et al. 2003; Emilsson et al. 2003). Since then, in-line probing is routinely used in studies of riboswitches, where the binding of a small molecule (ligand) results in a conformational change of the whole RNA molecule (Mandal and Breaker 2004; Montange and Batey 2008; Regulski and Breaker 2008; Garst et al. 2011). In general, chemical probing experiments are typically used to identify unpaired and flexible nucleotides, allowing one to choose among different predicted secondary structures. However, it must be noted that not all the unpaired nucleotides are usually reactive. The reactivity pattern of unpaired nucleotides could in principle provide a wealth of information that is usually discarded. To the best of our knowledge, the pattern of reactivity of specific motifs has never been analyzed in detail.

Computational techniques are an established tool for the investigation of structural and dynamical properties of RNA at an atomistic level (Schlick 2010; Cheatham and Case 2013; Šponer et al. 2013) and could in principle allow for an investigation and interpretation of reactivity patterns in RNA. In particular, quantum mechanical/molecular mechanical (QM/MM) (Warshel and Levitt 1976) approaches, where only the reactive portion of the system is described at the QM level, have been used to characterize cleavage reactions within catalytic RNA systems (Banáš et al. 2008; Lee et al. 2008; Nam et al. 2008a,b; Mlýnský et al. 2011, 2015; Rosta et al. 2011; Xu et al. 2012; Gu et al. 2013; Ganguly et al. 2014; Dubecký et al. 2015; Thaplyal et al. 2015; Zhang et al. 2015; Casalino et al. 2016). Single-point QM/MM calculations evaluate potential energy surfaces neglecting entropic contributions. The reconstruction of free-energy surfaces (FES) along the reaction pathway requires a combination of QM/MM calculations with molecular dynamics (MD) simulations and enhanced sampling techniques (Palermo et al. 2015). In this context, semi-empirical (SE) methods allow for a reasonable compromise between accuracy and efficiency (Christensen et al. 2016), allowing statistically converged FES to be computed. Two SE methods have been carefully parameterized for reactions involving the phosphate moiety (Nam et al. 2007; Yang et al. 2008).

In this article, we combine QM/MM calculations and enhanced sampling methods to model phosphodiester backbone cleavage of nucleotides from three model systems: one tetraloop from each of the GNRA and UNCG families (R = purine and N = any nucleotide) and a double-stranded RNA (dsRNA). Regions undergoing the cleavage reaction were described by a DFTB3 SE method, which allowed us to perform simulations on tens-of-ns timescales and obtain converged free-energy landscapes. Our calculations required us to design a putative reaction pathway with a number of restraints in order to overcome persisting shortcomings within parameterization of the DFTB3. However, we were able to differentiate among nucleotides within simple motifs by comparing

their activation free-energy barriers. Computational results were validated against available experimental data from in-line probing measurements. To our knowledge, this represents the first attempt to design computations in order to understand and mimic in-line probing experiments and, more generally, to investigate the phosphodiester backbone cleavage within nonreactive RNA nucleotides, i.e., not considering the active centers of RNA catalytic motifs.

RESULTS

We performed combined QM/MM-MetaD simulations (see Materials and Methods) and reconstructed FES relative to the phosphodiester cleavage reaction for nucleotides within three simple RNA motifs: uGAAAg tetraloop, dsRNA (GC-duplex), and cUUCGg tetraloop. For the tetraloops, we computed the reactivity for the unpaired residues and for the closing base pairs. For the duplex, we chose three consecutive nonterminal residues from each strand (Supplemental Fig. S1). We thus analyzed relative differences in reactivities among 6 nt for each system by comparing $\Delta G^{\ddagger}$ cleavage barriers along reaction pathways designed to be equivalent for all nucleotides from different RNA motifs. Since our intention was not to give insight into the mechanism of phosphodiester backbone cleavage, possible involvement of other RNA groups, water molecules, and ions was omitted and any mechanistic interpretation deliberately neglected.

The computed FES profiles mapped ΔG changes along two tracked distance-based collective variables (CVs) (Fig. 1A) to describe the proton transfer and the phosphodiester cleavage. The most relevant information that we want to extract is the $\Delta G^{\ddagger}$ cleavage barrier, which was estimated as explained in Materials and Methods. The FES displays two energy minima containing reactant (R) and product (P) state geometries (Fig. 1B). The R state minimum is very broad because the simulation is allowed to sample all the possible geometries, including different orientations of the active 2′-OH group. This is necessary for the accurate estimation of $\Delta G^{\ddagger}$ barriers. On the contrary, we restrained the extensive separation of the RNA molecule after the cleavage reaction, leading to a narrower free-energy minimum associated to P state. The complete exploration of P state geometries would make convergence very difficult and is irrelevant for a proper estimation of the cleavage barrier.

The FES profiles for nucleotides within three different RNA motifs (uGAAAg and cUUCGg tetraloops, GC-duplex) are qualitatively very similar among each other (see Supplemental Fig. S2 for all computed FES). The R and transition (TS) states have slightly different locations for particular nucleotides within each RNA motif, but no relevant correlations. For instance, purine/pyrimidine or tetraloop/duplex differences were not detected. The computed $\Delta G^{\ddagger}$ barriers after 40-nsec-long QM/MM-MetaD simulations were in the range between 40.0 and 44.5 kcal/mol. Note that the reaction coordinates are affected by the applied
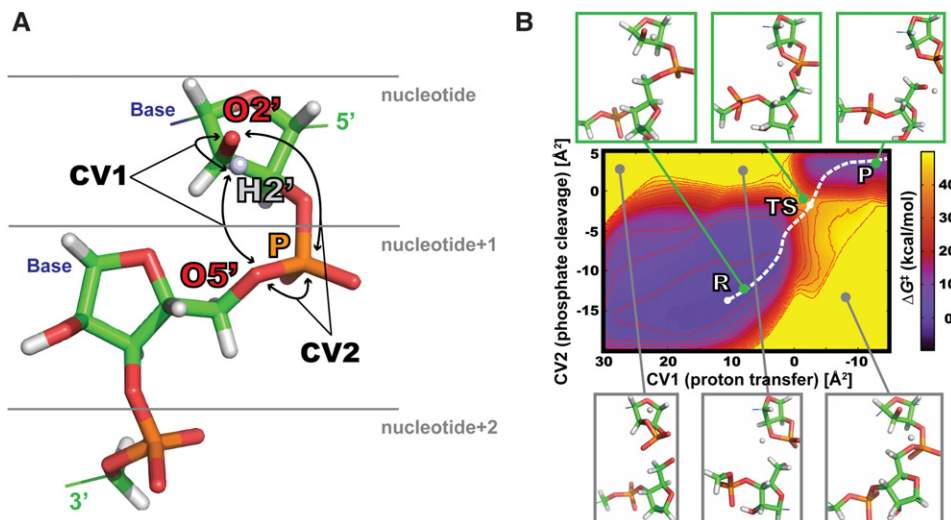
FIGURE 1. (*A*) Selection of QM region and definition of collective variables (CVs). QM region for a particular nucleotide contained ribose ring (reactive 2′-OH group), phosphate with ribose ring of nucleotide+1, and phosphate moiety of nucleotide+2. Bases were described at MM level of theory (AMBER *ff14*). FES of phosphate cleavage reaction was defined by two CVs. CV1 described the proton transfer (difference of squared [O5′…H2′] and [O2′…H2′] distances), and CV2 described the phosphate cleavage (difference of squared [O5′…P] and [O2′…P] distances). (*B*) FES of phosphodiester backbone cleavage for G4 nucleotide from uGAAAg tetraloop. Snapshots in boxes show different conformation and reaction states of QM atoms defined by specific values of CVs. The proton from 2′-OH is represented as unbound to be clearly visible. Green boxes show geometries close to R, TS, and P states, whereas gray boxes display structures energetically penalized by restraints (see Supplemental Material for details). The white dashed line marks the estimated reaction coordinate.

restraints and approximations, which forced the proton from the 2′-OH to be kept around the direct pathway (Fig. 1). Furthermore, the reported cleavage barriers are expected to be overestimated due to the complete exclusion of scenarios involving proton transfer through nonbridging oxygens (nbO) of the adjacent phosphate and/or through other proximal RNA groups. However, this allows for a consistent estimation of $\Delta G^{\ddagger}$ values and their comparison across nucleotides within different RNA motifs.

We tested carefully the statistical convergence of computed phosphodiester cleavage barriers. The analysis showed that $\Delta G^{\ddagger}$ barriers of six investigated nucleotides within the uGAAAg tetraloop fluctuate within a few kcal/mol over the time of the simulation (shaded lines, Fig. 2A), making it difficult to differentiate among nucleotides. All those barriers were estimated by using the final bias potential. In order to increase the accuracy of the method, we then calculated $\Delta G^{\ddagger}$ from time-averaged bias potentials. This latter approach gives a smoother convergence, which enables nucleotides to be clearly distinguished. The resulting $\Delta G^{\ddagger}$ barriers of uGAAAg nucleotides are clearly converged after 40 nsec of cumulative simulated time (Fig. 2A). Note that even initial estimated averages (∼7 nsec of total simulated time) show clear differences within $\Delta G^{\ddagger}$ among tested nucleotides. The GC-duplex was used as a control simulation because the computed $\Delta G^{\ddagger}$ barriers are expected to be identical for three equivalent nucleotides. Our approach shows that the computed $\Delta\Delta G^{\ddagger}$ differences of equivalent G and C nucleotides are negligible after 40 nsec, i.e., up to 0.4 and 0.5 kcal/mol, respectively (Supplemental Fig. S3A). We also analyzed the

location of TS during different stages of the simulations because the estimation of $\Delta G^{\ddagger}$ depends on the position of TS and R states on the FES. The TS positions of all nucleotides within the uGAAAg tetraloop are located within a small region in the CV space (Fig. 2B). The variance in positions of R states was even smaller (data not shown). The same trend was observed for the nucleotides within the GC-duplex (Supplemental Fig. S3B), whereas differences in TS positions on FES were slightly larger for nucleotides within the cUUCGg tetraloop (Supplemental Fig. S4B).

The phosphodiester backbone cleavage $\Delta G^{\ddagger}$ barriers are generally between 41 and 42 kcal/mol, but some nucleotides showed intrinsic differences. G4 within the uGAAAg tetraloop revealed the lowest barrier among all the nucleotides (40.0 kcal/mol, Supplemental Fig. S2). The following A5 showed a significantly higher barrier of 43.1 kcal/mol, whereas the $\Delta G^{\ddagger}$ barriers for the remaining nucleotides were comparable (∼41.5 kcal/mol). All the nucleotides within the GC-duplex revealed comparable barriers between 40.9 and 41.5 kcal/mol. Among cUUCGg nucleotides, U5 showed the lowest barriers (40.6 kcal/mol), whereas U4 and C6 provided significantly higher barriers of 44.5 and 43.5 kcal/mol, respectively (Supplemental Figs. S2, S4A).

We then compared the computed $\Delta G^{\ddagger}$ barriers against data from the in-line probing measurements. The experimental reactivities (pseudo free energies) were derived from available polyacrylamide gel electrophoresis (PAGE) data sets, quantified and normalized separately according to the scheme described in Materials and Methods. We observed good agreement between computed and experimental
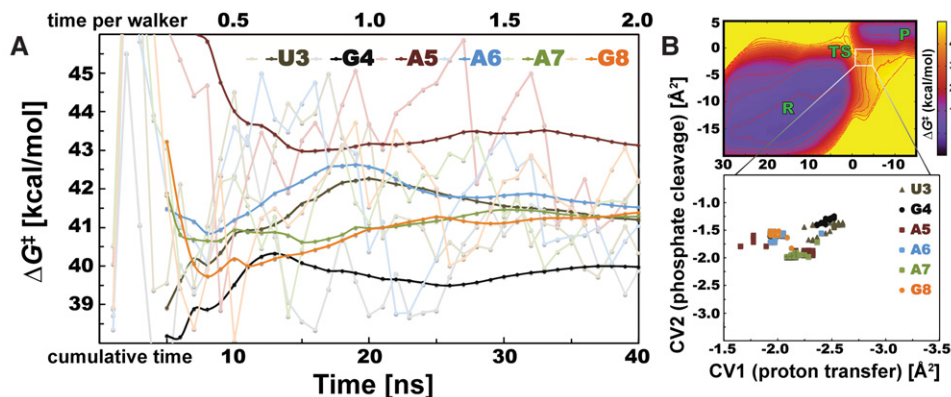
FIGURE 2. Convergence of cleavage barriers and distribution of TS positions of nucleotides from uGAAAg tetraloop. (*A*) Instantaneous estimates of $\Delta G^{\ddagger}$ barrier (shaded lines) fluctuate, whereas barriers calculated from time-averaged bias potentials (see Materials and Methods) are converged (bold lines). Each uGAAAg nucleotide is displayed in a specific color with two (shaded and bold) tones. *Horizontal* axis has two scales, showing timescale of single simulation (one walker, *upper* scale) or cumulative time considering 20 concurrent QM/MM-MetaD simulations (*lower* scale). The $\Delta G^{\ddagger}$ values were analyzed every 1 nsec in total timescale. (*B*) Position of TS states on FES is not changing significantly during QM/MM-MetaD simulations. TS states of all tested nucleotides from uGAAAg tetraloop are located *within* the same area with minimal changes of CV2 (describing the proton transfer). Colors for each nucleotide and the interval for analysis correspond with those on panel *A*.

reactivities for nucleotides within the uGAAAg tetraloop. Our computations overestimated the $\Delta G^{\ddagger}$ barriers for U3 and A5 nucleotides, but led to the correct trend overall (Fig. 3A). However, a similar comparison for the nucleotides within the cUUCGg tetraloop revealed some differences. U5, C6, and G7 were identified as reactive nucleotides according to the experiment (Strauss et al. 2012), but the computed $\Delta G^{\ddagger}$ barrier for C6 was significantly higher, suggesting that the nucleotide is nonreactive (Fig. 3C). Note that the in-line probing data for a uniform GC-duplex are not available, but that reactivity of paired residues is typically lower than reactivity for unpaired residues (Soukup and Breaker 1999; Kulshina et al. 2009; Erion and Strobel 2011; Strauss et al. 2012; Nelson et al. 2013; Hickey and Hammond 2014; Furukawa et al. 2015).

DISCUSSION

In this article, we used QM/MM-MetaD calculations to classify in-line probing experiments characterized by the phosphodiester backbone cleavage reaction. To this aim, we calculated the free-energy profiles modeling the RNA backbone cleavage for 18 nucleotides (nt) within two RNA tetraloops and a dsRNA. The computed $\Delta G^{\ddagger}$ barrier obtained from the QM/MM-MetaD calculation is expected to be related to the reactivity of the particular nucleotide as observed in in-line probing experiments. The aim of our study was not to predict absolute reactivities but to explain differential reactivities observed among nucleotides within the same or from different RNA motifs. To minimize the error in differential estimates, we forced the system to explore a similar reaction pathway for each nucleotide and prolonged simulations in the tens-of-ns timescale. The approach provides converged and consistent results for all the nucleotides. Results were

then assessed by comparing $\Delta G^{\ddagger}$ barriers of identical nucleotides within a GC-duplex motif and by analyzing reference in-line probing reactivities from PAGE gels.

We included a number of artificial restraints, which were required in order to automatize the computational protocol, i.e., to allow the straightforward comparison of various nucleotides within distinct RNA motifs. Restraints improved stability of the simulations namely by preventing spurious rupture of bonds and by excluding several unphysical geometries detected during phosphodiester cleavage reaction. However, all the backbone dihedrals, sugar puckers, glycosidic bonds, as well as base-pairing and stacking interactions were left free to rearrange. We actually observed significant dynamics during our QM/MM-MetaD simulations. This is important since the essence of in-line probing experiments is to quantify the effects of RNA structural fluctuations on the phosphodiester cleavage rate. The introduction of restraints is nevertheless expected to affect the computed barriers, and thus the respective cleavage rates, in two specific ways. First, restraints forced the cleavage reaction to proceed through the designed reaction pathway that is similar for all nucleotides and likely different from the validated in-line attack reaction pathway. The possible contributions of other reaction pathways, which could be different from one nucleotide to the other, were thus ignored. As a result, the reactivity trends estimated as differences of $\Delta G^{\ddagger}$ barriers could in principle be compressed. Second, all the computed $\Delta G^{\ddagger}$ barriers are expected to be overestimated by excluding the scenario where the proton from the 2′-OH group is transferred through nbO atoms. Considering the typical timescale of in-line probing experiments (hours/days), the $\Delta G^{\ddagger}$ barriers derived from the estimated rate constants using the Eyring equation are expected to be in the range from ∼26 to ∼32 kcal/mol (Soukup and Breaker 1999), i.e., by ∼7 to ∼13
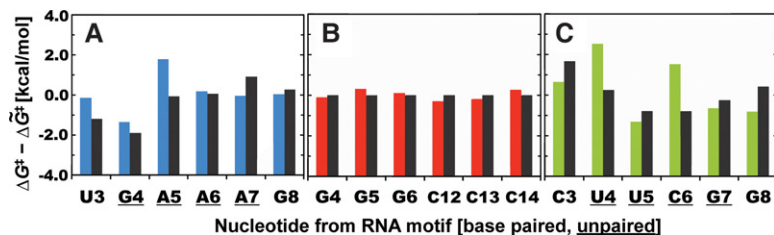
FIGURE 3. Comparison between computed and experimentally derived cleavage barriers for nucleotides from three different motifs. Correlation between computed $\Delta G^{\ddagger}$ barriers (with specific color for each motif) and barriers derived from experiments (black, pseudo free energies, see Materials and Methods for details) for nucleotides from uGAAAg tetraloop (*A*), GC-duplex (*B*), and cUUCGg tetraloop (*C*). Both computed and experimentally derived $\Delta G^{\ddagger}$ barriers are displayed as deviations from the corresponding median value. Note that in-line probing data for a uniform GC-duplex are not available, but equivalent G–C base pairs are expected to have similar $\Delta G^{\ddagger}$ barriers (deviations from the median are set to zero in the figure). Unpaired nucleotides from tetraloops are underlined (see Supplemental Fig. S1 for nucleotide labeling).

kcal/mol lower than the herein reported $\Delta G^{\ddagger}$ barriers. Both those issues are very difficult to tackle since they depend on intrinsic deficiencies of the DFTB3 parameterizations. Despite the fact that DFTB and other SE methods improved significantly during the last decade (Huang et al. 2014; Christensen et al. 2016), their general application toward chemical reactions remains challenging. In particular, recent studies carefully assessed the performance and revealed limitations of DFTB methods in description of phosphoranes and phosphoryl transfers (Gaus et al. 2014; Mlýnský et al. 2014; Huang et al. 2015; Lu et al. 2016). In this study, we still opted for the usage of DFTB3 because we did not aim for an accurate description of states along phosphodiester cleavage reaction. We rather focused on relative differences of $\Delta G^{\ddagger}$ barriers among different nucleotides, forcing the reaction to proceed through the same pathway for all the analyzed nucleotides. We expect such an approach to be more robust and less sensitive to the applied QM method. QM/MM-MetaD simulations, at least on a several-ns timescale, are required to converge these FES computations to a level allowing for the discrimination of reactivity patterns, ruling out more accurate QM methods such as DFT or ab initio.

Here, we ranked different RNA nucleotides, i.e., base paired/unpaired, purine/pyrimidine, from duplex and tetraloops by their tendency to undergo phosphodiester backbone cleavage. Nucleotide reactivities reported by in-line probing experiments were used as a reference. A number of experimental data sets for specific motifs from different RNA systems are available (Supplemental Fig. S5; Soukup and Breaker 1999; Kulshina et al. 2009; Erion and Strobel 2011; Strauss et al. 2012; Furukawa et al. 2015), but their quantitative estimation is often affected by unclear signals (Erion and Strobel 2011; Hickey and Hammond 2014; Furukawa et al. 2015) and/or participation in a tertiary interaction within a complex RNA structure (Soukup and Breaker 1999; Kulshina et al. 2009; Erion and Strobel 2011). Hence, we used a single specific experiment providing distinct signals for all nucleotides as a reference of each of the tetraloop motifs

(Strauss et al. 2012; Nelson et al. 2013). We observed that nucleotides within the cUUCGg tetraloop revealed some differences between predicted and calculated reactivities (Fig. 3C). Calculated $\Delta G^{\ddagger}$ barriers of U4 and C6 nucleotides are significantly overestimated, resulting in a poor correlation between theory and experiment ($R = 0.17$, Fig. 4A). We notice that (i) the FES profiles for cUUCGg nucleotides are statistically converged (Supplemental Fig. S4A), (ii) experimental reactivities for nucleotides within the cUUCGg motif reveal similar trends among different systems (Supplemental Fig. S5B), and (iii) the procedure of quantification of experimental reactivities, i.e., using digitalized images (see Materials and Methods) provide almost identical profiles against the raw count data (a comparison for the uGAAAg tetraloop is reported, Supplemental Fig. S6). Thus, the poor correlation between computed $\Delta G^{\ddagger}$ barriers and experimental reactivities for cUUCGg suggests possible limitations of our approach. We carefully inspected the structures along the cleavage reaction and found that the reactive 2′-OH of U5 and, especially U4, established H-bonds with other RNA groups outside the QM region (described by the empirical force field, Supplemental Fig. S7). Such interactions could result in overstabilization of R states, leading to the overestimation of the computed $\Delta G^{\ddagger}$ barriers by the current approach. Furthermore, the cleavage site of C6 favored rare conformations with high in-line attack angles. Such geometry is not favorable for the mechanism enforced herein and would require us to explore the in-line attack reaction pathway, where nbO atoms (and/or external RNA groups, water molecules) are involved in the proton transfer. This was not possible due to deficiencies within DFTB3 parameterization (see Supplemental Material for details). One possible way to improve the results for cUUCGg would be to increase the number of atoms within the QM region (described by DFTB3). However, we identified that RNA groups forming those interactions are typically belonging to nucleotides located further away along the sugar–phosphate backbone, making those calculations unfeasible.

On the other hand, the agreement between theory and experiment is satisfactory for both GC-duplex and uGAAAg tetraloop. In the former case, nucleotides revealed very similar $\Delta G^{\ddagger}$ barriers, which are expected for three consecutive equivalent G and C within dsRNA. The fact that each of these barriers was obtained with a totally independent simulation further confirms the low statistical error and hence the reproducibility of our computational approach. The possible differences between purine and pyrimidine nucleotides were negligible for this motif (within the error of our approach, Supplemental Fig. S3A). However, we observed that those $\Delta G^{\ddagger}$ barriers are comparable with cleavage barriers of several
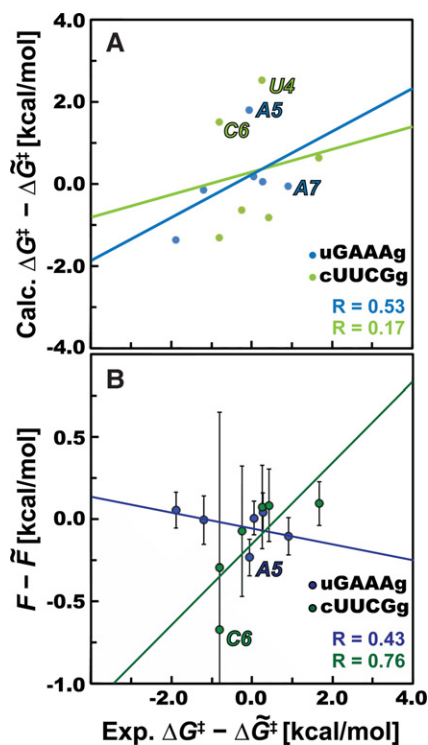
FIGURE 4. Direct correlation between theory and experiments for nucleotides within two tetraloops. (*A*) The correlation between calculated $\Delta G^{\ddagger}$ and barriers derived from experiments for nucleotides within uGAAAg (blue) and cUUCGg (green) tetraloops. Both computed and experimental barriers are displayed as deviations from the corresponding median value. (*B*) A similar correlation plot obtained using the logarithm of in-line fitness $F = -RT\ln(\text{Fitness})$ (Soukup and Breaker 1999), instead of the here calculated $\Delta G^{\ddagger}$. We displayed averaged values of fitness from the set of X-ray structures (considering 41 uGAAAg and 754 cUUCGg tetraloops, respectively) from the PDB database with bars showing their standard deviations. *F*-values are displayed the same way as $\Delta G^{\ddagger}$ barriers subtracting medians. Points corresponding to specific residues that are worsening the correlation (outliers) are labeled.

unpaired residues within tetraloop motifs. This is unexpected, since the reactivity of paired nucleotides is generally lower (Reynolds et al. 1996; Welch et al. 1997; Soukup and Breaker 1999). We speculate that such behavior is caused by the number of restraints used in our computations, although it could also be linked to the approximations in the DFTB3 method. In the uGAAAg tetraloop, QM/MM-MetaD simulations can clearly predict that the unpaired G4 is significantly more reactive than the other nucleotides. Simulations are long enough to consider this difference statistically significant. Other nucleotides have higher $\Delta G^{\ddagger}$ barriers, in agreement with the lower reactivity observed in experiments (Fig. 3A). Plotting computational and experimental reactivities against each other revealed that the two unpaired nucleotides (A5 and A7) are worsening the correlation ($R = 0.53$) due to slightly overestimated (A5) and underestimated (A7) $\Delta G^{\ddagger}$ barriers (Fig. 4A). It is worth noting that the overall stability and flexibility of the tetraloop motifs can be affected by the nature of the closing base pair (Proctor et al. 2002; Blose

et al. 2009). For this reason, we explicitly replaced the G–C base pair observed in the crystal structure with the wobble G–U pair found in the sequence used in the reference in-line probing experiments.

Our results can also be compared against predictions using the approach introduced by Soukup and Breaker (1999), where a correlation between geometrical parameter (in-line fitness) and cleavage rates was proposed. The fitness parameter combines the in-line attack angle and the distance between O2′ and the adjacent phosphorus. Interestingly, the fluctuations of the in-line attack angle were proposed as a proxy for the chemical reactivity of individual nucleotides (Kirmizialtin et al. 2015). We searched among high-resolution X-ray structures (≤ 3.5 Å) of RNA molecules from the RCSB Protein Data Bank (PDB) (Berman et al. 2000) and used the baRNAba tool (Bottaro et al. 2014) to extract representative uGAAAg tetraloops. The average fitness values for those nucleotides are anti-correlated with the experimental reactivities derived from gels (Fig. 4B), showing that the in-line fitness formula (Soukup and Breaker 1999) derived from static X-ray structures cannot reproduce the experimental reactivity pattern for this motif. This is not surprising, since the in-line fitness was not designed for differentiating among random nucleotides, but rather for the specific identification of highly reactive nucleotides within catalytic centers of ribozymes (see Fig. 8 in the original paper, Soukup and Breaker 1999).

Inspection of the starting structure used for uGAAAg computations (PDB ID 4QLM) (Ren and Patel 2014) revealed accidental *syn* orientation of the unpaired A5 and A7, which surprisingly improved the correlation between in-line fitness and experimental reactivity ($R = 0.85$, Supplemental Fig. S8A). We recall that considering the structures extracted from the PDB as well as solution experiments (Heus and Pardi 1991; Jucker et al. 1996; Bottaro et al. 2016), the *syn* conformation is rare and unexpected for nucleotides within the uGAAAg tetraloop. Since we did not find any apparent crystallographic contact that may invoke those reorientations, it appears likely that the higher flexibility of unpaired bases affected the refinement and resulted in poor electron densities for unpaired nucleotides located away from the important (binding) centers of the *ydaO* riboswitch. *Syn/anti* flips of A5, A6, and A7 nucleotides also occurred during classical MD simulation used for the system equilibration. Remarkably, all nucleotides revealed the correct *anti* conformation after 100 nsec of the simulation time, i.e., in the structure used for subsequent QM/MM-MetaD calculations, indicating that the MM force field used was able to recover the expected native structure. We notice that the *syn/anti* flips of all unpaired nucleotides from the uGAAAg tetraloop were also occasionally detected during QM/MM-MetaD simulations (in timescale of tens to hundreds of ps), despite the fact that the starting structure contained all bases in correct *anti* conformation. This may suggest that possible *anti/syn* reorientation might induce the phosphodiester backbone cleavage by enabling a

more favorable ribose pucker state (C2′-endo) for the initial nucleophilic attack and/or different conformations of the adjacent phosphate. Interestingly, functional nucleotides within catalytic centers of RNAs are frequently found in *syn* conformation (Sokoloski et al. 2011).

In conclusion, we presented an approach to characterize the reactivity in RNA motifs. We used QM/MM calculations with semi-empirical methods, in combination with multiple-walkers metadynamics, to compute the free-energy barriers associated with phosphodiester backbone cleavage in generic, noncatalytic nucleotides. The computational protocol is fast and robust, though limited by the currently available parameters for the DFTB3 method. Remarkably, our procedure was able to reproduce and explain differential reactivities in a common RNA tetraloop (uGAAAg). However, reactivities in another tetraloop (cUUCGg) were more difficult to classify. Our results suggest that better DFTB3 parameters would be required for appropriate modeling of phosphodiester cleavage reactions of this system and our protocol may serve as a benchmark for the further improvements of the semi-empirical method. To the best of our knowledge, this study represents the first computational approach for the interpretation and classification of chemical probing experiments. The introduced procedure could be applied to more complex RNA motifs, providing the initial step for fast and cheap distinguishing among several experimentally suggested RNA structures.

MATERIALS AND METHODS

Initial structures of RNA motifs were taken from crystal structures, i.e., PDB ID 4QLM (uGAAAg tetraloop) (Ren and Patel 2014), 1QCU (dsRNA) (Klosterman et al. 1999), and 4JF2 (cUUCGg tetraloop) (Liberman et al. 2013). Tetraloop motifs contain 10 nt and the dsRNA duplex consists of eight G–C base pairs (see Supplemental Fig. S1 and Supplemental Methods section for structures and details). The QM region included two ribose rings and two phosphates (Fig. 1A). We used the DFTB3 method (Gaus et al. 2011), as implemented in GROMACS 5.0 (Abraham et al. 2015; Kubař et al. 2015), in combination with PLUMED (Tribello et al. 2014). Recent corrections (Huang et al. 2014) that improve the description of ribose rings (sugar-puckers) were additionally applied using PLUMED. Bases were described at the MM level by AMBER *ff*14 (Cornell et al. 1995; Wang et al. 2000; Pérez et al. 2007; Zgarbová et al. 2011) in order to handle all of them at the same level of theory. We explicitly tested the performance of all available DFTB3 parameter sets, i.e., MIO (Gaus et al. 2011), 3OB (Gaus et al. 2014), and 3OB-OPhyd (Gaus et al. 2014). After accurate validations we opted for the MIO set, and all results reported herein were calculated by that setup. To avoid spurious reactions and unphysical geometries, we had to enforce specific reaction pathways with a number of artificial restraints to disallow the rupture of bonds not involved in the cleavage reaction. These restraints might penalize the reactive in-line attack geometry, and the enforced pathway is likely to be different from the reaction monitored by in-line probing experiments (see Supplemental Material for details).

Well-tempered metadynamics (MetaD) (Laio and Parrinello 2002; Barducci et al. 2008) under the multiple-walker algorithm (Raiteri et al. 2006) was used to accelerate the phosphodiester cleavage and to estimate the associated FES. Two CVs were used (Fig. 1A): one to describe the proton transfer and the other to describe the phosphodiester cleavage. FES was computed either considering the final bias potential or considering time-averages of the bias potential (Micheletti et al. 2004), and convergence was monitored during different stages of the simulation. Further discussion and justification for the time-averaging procedure can be found in the Methods section of the Supplemental Material. The activation free-energy ($\Delta G^{\ddagger}$) barrier of the phosphodiester backbone cleavage for the particular nucleotide was extracted from computed FES by localizing the saddle point (TS) on the reaction coordinate, i.e., the most convenient path (requiring the lowest energies) connecting two areas with minimal energies on the FES, corresponding to R and P state geometries (Fig. 1).

Experimental reactivities for specific nucleotides within uGAAAg and cUUCGg tetraloops were quantified by analyzing PAGE data. We took digitalized images extracted from the original papers (Strauss et al. 2012; Nelson et al. 2013) and we integrated the color density present in the area of the image corresponding to each nucleotide. Subsequently, pseudo-free-energy reactivities were derived using an approach similar to the one developed for SHAPE experiments (Low and Weeks 2010): $\Delta G_{Exp} = -m\ln s$, where s is the signal intensity from gels and $m = 2.6$ kcal/mol. Intensities were normalized by shifting the medians of experimental reactivities to match those of the calculated barriers for each tetraloop motif. We note that all the considered systems were studied using identical settings and analysis procedures so as to allow for an unbiased comparison.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

REFERENCES

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2:** 19–25.

Banáš P, Rulíšek L, Hánošová V, Svozil D, Walter NG, Šponer J, Otyepka M. 2008. General base catalysis for cleavage by the active-site cytosine of the hepatitis delta virus ribozyme: QM/MM calculations establish chemical feasibility. *J Phys Chem B* **112:** 11177–11187.

Barducci A, Bussi G, Parrinello M. 2008. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett* **100:** 020603.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. *Nucleic Acids Res* **28:** 235–242.

Blose JM, Proctor DJ, Veeraraghavan N, Misra VK, Bevilacqua PC. 2009. Contribution of the closing base pair to exceptional stability in RNA tetraloops: roles for molecular mimicry and electrostatic factors. *J Am Chem Soc* **131:** 8474–8484.

Bottaro S, Di Palma F, Bussi G. 2014. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res* **42:** 13306–13314.

Bottaro S, Gil-Ley A, Bussi G. 2016. RNA folding pathways in stop motion. *Nucleic Acids Res* **44:** 5883–5891.

Breaker RR, Emilsson GM, Lazarev D, Nakamura S, Puskarz IJ, Roth A, Sudarsan N. 2003. A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA* **9:** 949–957.

Casalino L, Palermo G, Rothlisberger U, Magistrato A. 2016. Who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns. *J Am Chem Soc* **138:** 10374–10377.

Cheatham TE, Case DA. 2013. Twenty-five years of nucleic acid simulations. *Biopolymers* **99:** 969–977.

Christensen AS, Kubař T, Cui Q, Elstner M. 2016. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chem Rev* **116:** 5301–5337.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **117:** 5179–5197.

Doudna JA, Cech TR. 2002. The chemical repertoire of natural ribozymes. *Nature* **418:** 222–228.

Dubecký M, Walter NG, Šponer J, Otyepka M, Banáš P. 2015. Chemical feasibility of the general acid/base mechanism of *glmS* ribozyme self-cleavage. *Biopolymers* **103:** 550–562.

Emilsson GM, Nakamura S, Roth A, Breaker RR. 2003. Ribozyme speed limits. *RNA* **9:** 907–918.

Erion TV, Strobel SA. 2011. Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch. *RNA* **17:** 74–84.

Furukawa K, Ramesh A, Zhou Z, Weinberg Z, Vallery T, Winkler WC, Breaker RR. 2015. Bacterial riboswitches cooperatively bind Ni^{2+} or Co^{2+} ions and control expression of heavy metal transporters. *Mol Cell* **57:** 1088–1098.

Ganguly A, Thaplyal P, Rosta E, Bevilacqua PC, Hammes-Schiffer S. 2014. Quantum mechanical/molecular mechanical free energy simulations of the self-cleavage reaction in the hepatitis delta virus ribozyme. *J Am Chem Soc* **136:** 1483–1496.

Garst AD, Edwards AL, Batey RT. 2011. Riboswitches: structures and mechanisms. *Cold Spring Harbor Perspect Biol* **3:** a003533.

Gaus M, Cui Q, Elstner M. 2011. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theory Comput* **7:** 931–948.

Gaus M, Lu X, Elstner M, Cui Q. 2014. Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. *J Chem Theory Comput* **10:** 1518–1537.

Gu H, Zhang S, Wong K-Y, Radak BK, Dissanayake T, Kellerman DL, Dai Q, Miyagi M, Anderson VE, York DM, et al. 2013. Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2'-O-transphosphorylation. *Proc Natl Acad Sci* **110:** 13002–13007.

Hall KB. 2015. Mighty tiny. *RNA* **21:** 630–631.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253:** 191–194.

Hickey SF, Hammond MC. 2014. Structure-guided design of fluorescent *s-adenosylmethionine* analogs for a high-throughput screen to target SAM-I riboswitch RNAs. *Chem Biol* **21:** 345–356.

Huang M, Giese TJ, Lee T-S, York DM. 2014. Improvement of DNA and RNA sugar pucker profiles from semiempirical quantum methods. *J Chem Theory Comput* **10:** 1538–1545.

Huang M, Giese TJ, York DM. 2015. Nucleic acid reactivity: challenges for next-generation semiempirical quantum models. *J Comput Chem* **36:** 1370–1389.

Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264:** 968–980.

Kirmizialtin S, Hennelly SP, Schug A, Onuchic JN, Sanbonmatsu KY. 2015. Integrating molecular dynamics simulations with chemical probing experiments using SHAPE-FIT. *Methods Enzymol* **553:** 215–234.

Klosterman PS, Shah SA, Steitz TA. 1999. Crystal structures of two plasmid copy control related RNA duplexes: an 18 base pair duplex at 1.20 Å resolution and a 19 base pair duplex at 1.55 Å resolution. *Biochemistry* **38:** 14784–14792.

Kubař T, Welke K, Groenhof G. 2015. New QM/MM implementation of the DFTB3 method in the gromacs package. *J Comput Chem* **36:** 1978–1989.

Kubota M, Tran C, Spitale RC. 2015. Progress and challenges for chemical probing of RNA structure inside living cells. *Nat Chem Biol* **11:** 933–941.

Kulshina N, Baird NJ, Ferré-D'Amaré AR. 2009. Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat Struct Mol Biol* **16:** 1212–1217.

Kung JT, Colognori D, Lee JT. 2013. Long noncoding RNAs: past, present, and future. *Genetics* **193:** 651–669.

Laio A, Parrinello M. 2002. Escaping free-energy minima. *Proc Natl Acad Sci* **99:** 12562–12566.

Lee T-S, López CS, Giambasu GM, Martick M, Scott WG, York DM. 2008. Role of Mg^{2+} in hammerhead ribozyme catalysis from molecular simulation. *J Am Chem Soc* **130:** 3053–3064.

Liberman JA, Salim M, Krucinska J, Wedekind JE. 2013. Structure of a class II preQ$_1$ riboswitch reveals ligand recognition by a new fold. *Nat Chem Biol* **9:** 353–355.

Lilley DMJ. 2003. The origins of RNA catalysis in ribozymes. *Trends Biochem Sci* **28:** 495–501.

Lilley DMJ, Eckstein F. 2008. *Ribozymes and RNA catalysis*. Royal Society of Chemistry, UK.

Low JT, Weeks KM. 2010. SHAPE-directed RNA secondary structure prediction. *Methods* **52:** 150–158.

Lu X, Fang D, Ito S, Okamoto Y, Ovchinnikov V, Cui Q. 2016. QM/MM free energy simulations: recent progress and challenges. *Mol Simul* **42:** 1056–1078.

Mandal M, Breaker RR. 2004. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5:** 451–463.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127:** 4223–4231.

Micheletti C, Laio A, Parrinello M. 2004. Reconstructing the density of states by history-dependent metadynamics. *Phys Rev Lett* **92:** 170601.

Mlýnský V, Banáš P, Walter NG, Šponer J, Otyepka M. 2011. QM/MM studies of hairpin ribozyme self-cleavage suggest the feasibility of multiple competing reaction mechanisms. *J Phys Chem B* **115:** 13911–13924.

Mlýnský V, Banáš P, Šponer J, van der Kamp MW, Mulholland AJ, Otyepka M. 2014. Comparison of *ab initio*, DFT, and semiempirical QM/MM approaches for description of catalytic mechanism of hairpin ribozyme. *J Chem Theory Comput* **10:** 1608–1622.

Mlýnský V, Walter NG, Šponer J, Otyepka M, Banáš P. 2015. The role of an active site Mg^{2+} in HDV ribozyme self-cleavage: insights from QM/MM calculations. *Phys Chem Chem Phys* **17:** 670–679.

Montange RK, Batey RT. 2008. Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **37:** 117–133.

Nam K, Cui Q, Gao J, York DM. 2007. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J Chem Theory Comput* **3:** 486–504.

Nam K, Gao J, York DM. 2008a. Electrostatic interactions in the hairpin ribozyme account for the majority of the rate acceleration without chemical participation by nucleobases. *RNA* **14:** 1501–1507.

Nam K, Gao J, York DM. 2008b. Quantum mechanical/molecular mechanical simulation study of the mechanism of hairpin ribozyme catalysis. *J Am Chem Soc* **130:** 4680–4691.

Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. 2013. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat Chem Biol* **9:** 834–839.

Palermo G, Cavalli A, Klein ML, Alfonso-Prieto M, Dal Peraro M, De Vivo M. 2015. Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc Chem Res* **48:** 220–228.

Pérez A, Marchán I, Svozil D, Šponer J, Cheatham TE, Laughton CA, Orozco M. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys J* **92:** 3817–3829.

Proctor DJ, Schaak JE, Bevilacqua JM, Falzone CJ, Bevilacqua PC. 2002. Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry* **41:** 12062–12075.

Raines RT. 1998. Ribonuclease A. *Chem Rev* **98:** 1045–1066.

Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M. 2006. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J Phys Chem B* **110:** 3533–3539.

Regulski EE, Breaker RR. 2008. In-line probing analysis of riboswitches. *Methods Mol Biol* **419:** 53–67.

Ren A, Patel DJ. 2014. c-di-AMP binds the *ydao* riboswitch in two pseudo-symmetry-related pockets. *Nat Chem Biol* **10:** 780–786.

Reynolds MA, Beck TA, Say PB, Schwartz DA, Dwyer BP, Daily WJ, Vaghefi MM, Metzler MD, Klem RE, Arnold LJ. 1996. Antisense oligonucleotides containing an internal, non-nucleotide-based linker promote site-specific cleavage of RNA. *Nucleic Acids Res* **24:** 760–765.

Rosta E, Nowotny M, Yang W, Hummer G. 2011. Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J Am Chem Soc* **133:** 8934–8941.

Sarkies P, Miska EA. 2014. Small RNAs break out: the molecular cell biology of mobile small RNAs. *Nat Rev Mol Cell Biol* **15:** 525–535.

Schlick T. 2010. *Molecular modeling and simulation: an interdisciplinary guide.* Springer Science & Business Media, NY.

Scott WG. 2007. Ribozymes. *Curr Opin Struct Biol* **17:** 280–286.

Sokoloski JE, Godfrey SA, Dombrowski SE, Bevilacqua PC. 2011. Prevalence of *syn* nucleobases in the active sites of functional RNAs. *RNA* **17:** 1775–1787.

Soukup GA, Breaker RR. 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5:** 1308–1325.

Šponer J, Šponer JE, Mládek A, Banáš P, Jurečka P, Otyepka M. 2013. How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists. *Methods* **64:** 3–11.

Strauss B, Nierth A, Singer M, Jäschke A. 2012. Direct structural analysis of modified RNA by fluorescent in-line probing. *Nucleic Acids Res* **40:** 861–870.

Strulson CA, Molden RC, Keating CD, Bevilacqua PC. 2012. RNA catalysis through compartmentalization. *Nat Chem* **4:** 941–946.

Thaplyal P, Ganguly A, Hammes-Schiffer S, Bevilacqua PC. 2015. Inverse thio effects in the hepatitis delta virus ribozyme reveal that the reaction pathway is controlled by metal ion charge density. *Biochemistry* **54:** 2160–2175.

Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. 2014. PLUMED 2: new feathers for an old bird. *Comput Phys Commun* **185:** 604–613.

Walter NG. 2009. The blessing and curse of RNA dynamics: past, present, and future. *Methods* **49:** 85–86.

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

Warshel A, Levitt M. 1976. 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.

Weeks KM. 2010. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **20:** 295–304.

Welch M, Majerfeld I, Yarus M. 1997. 23S rRNA similarity from selection for peptidyl transferase mimicry. *Biochemistry* **36:** 6614–6623.

Xu Z, Culver GM. 2009. Chemical probing of RNA and RNA/protein complexes. *Methods Enzymol* **468:** 147–165.

Xu J, Zhang JZ, Xiang Y. 2012. *Ab initio* QM/MM free energy simulations of peptide bond formation in the ribosome support an eight-membered ring reaction mechanism. *J Am Chem Soc* **134:** 16424–16429.

Yang Y, Yu H, York D, Elstner M, Cui Q. 2008. Description of phosphate hydrolysis reactions with the self-consistent-charge density-functional-tight-binding (SCC-DFTB) theory. 1. Parameterization. *J Chem Theory Comput* **4:** 2067–2084.

Zgarbová M, Otyepka M, Šponer J, Mládek A, Banáš P, Cheatham TE III, Jurečka P. 2011. 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.

Zhang S, Ganguly A, Goyal P, Bingaman JL, Bevilacqua PC, Hammes-Schiffer S. 2015. Role of the active site guanine in the *glms* ribozyme self-cleavage mechanism: quantum mechanical/molecular mechanical free energy simulations. *J Am Chem Soc* **137:** 784–798.