



HAL
open science

The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study

Elisa Frezza, Antoine Courban, Delphine M Allouche, Bruno Sargueil,
Samuela Pasquali

► To cite this version:

Elisa Frezza, Antoine Courban, Delphine M Allouche, Bruno Sargueil, Samuela Pasquali. The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study. *Methods*, 2019, 162-163, pp.108-127. 10.1016/j.ymeth.2019.05.021 . hal-02387476

HAL Id: hal-02387476

<https://hal.science/hal-02387476>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study

Elisa Frezza^{a,*}, Antoine Courban^a, Delphine Allouche^a, Bruno Sargueil^{a,*}, Samuela Pasquali^{a,*}

^a*Faculté de Pharmacie de Paris, Laboratoire de Cristallographie et RMN Biologiques, UMR 8015 - CNRS, Université Paris Descartes, 4 Avenue de l'Observatoire 75270 PARIS CEDEX 06, France.*

Abstract

Determination of the tridimensional structure of ribonucleic acid molecules is fundamental for understanding their function in the cell. A common method to investigate RNA structures of large molecules is the use of chemical probes such as SHAPE (2'-hydroxyl acylation analyzed by primer extension) reagents, DMS (dimethyl sulfate) and CMCT (1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfate), the reaction of which is dependent on the local structural properties of each nucleotide. In order to understand the interplay between local flexibility, sugar pucker, canonical pairing and chemical reactivity of the probes, we performed all-atom molecular dynamics simulations on a set of RNA molecules for which both tridimensional structure and chemical probing data are available and we analyzed the correlations between geometrical parameters and the chemical reactivity. Our study confirms that SHAPE reactivity is guided by the local flexibility of the different chemical moieties but suggests that a combination of multiple parameters is needed to better understand the implications of the reactivity at the molecular level. This is also the case for DMS and CMCT for which the reactivity appears to be more complex than commonly accepted.

Keywords: RNA, Chemical probing, SHAPE, flexibility, Molecular Dynamics simulations

1. Introduction

Ribonucleic acid (RNA) molecules are involved in most steps of the genetic expression including catalysis of central cellular functions [1, 2, 3]. RNA function depends crucially on the specific tridimensional folding of the molecule which in turns depends on the sequence and on

*Corresponding authors

Email addresses: elisa.frezza@parisdescartes.fr (Elisa Frezza),
bruno.sargueil@parisdescartes.fr (Bruno Sargueil), samuella.pasquali@parisdescartes.fr
(Samuela Pasquali)

Preprint submitted to Methods

May 22, 2019

the way the bases pair through hydrogen bonds (secondary structure) [4, 5]. High-resolution RNA structures can be determined experimentally by NMR spectroscopy [6], X-ray crystallography [7], and more recently using cryo-electron microscopy [8]. However, obtaining high-resolution tridimensional structures through these techniques is still a challenging task as shown by the relatively small number of structures deposited in the Nucleic Acid Data Bank (NDB) [9, 10]. One of the difficulties arise from the conformational flexibility of RNA molecules that allows them to adopt distinct conformations and results in rough energy landscapes [11]. Since high resolution structures of only a small fraction of RNA molecules are available, methods to predict RNA structure are essential.

Based on early experiments on RNA melting [12] and more recent analysis on large ribosomal molecules [13, 14], a hierarchical folding model has been proposed in which the secondary structure precedes long range interactions and tertiary folding [12, 15]. As a consequence, accurate modeling of the secondary structure is a crucial prerequisite to investigate tridimensional structures. Several bio-informatic secondary structure prediction methods have been developed and the integration of experimental data have been shown to greatly enhance their accuracy [16, 17, 18, 19, 20, 21].

Depending on the chemical reagents and their specificity to the RNA structure, distinct types of information about RNA structures can be obtained [22]. DMS (dimethyl sulfate) and CMCT (1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfate) are commonly used to investigate the reactivity of the nitrogenous bases [23, 24] (see Figure 1 for their structure). The former reacts with the nitrogen in position N1 of the adenine and in position N3 of the cytosine. The latter reacts in the position N3 of the uracil and N1 of guanine [23]. These chemical reactions compete with the Watson-Crick (WC) base pairing, giving a direct information on the paired or unpaired status of each base. More recently, the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) technology [25, 26, 16] has been developed to measure local nucleotide flexibility and dynamics. SHAPE reagents are small cyclic molecules which can acylate the 2'-hydroxyl group at flexible nucleotides through a nucleophilic attack. These modifications are subsequently detected as stops in primer extension or with mutational profiling (SHAPE-MaP) [27, 28, 29, 30]. SHAPE is independent of the nature of nucleotide and the reactivity seems to be mostly related to the conformation of the ribose [31]. Six different SHAPE reagents are reported in the literature (BzCN [32], 1M7 [33], 1M6 [34], NMIA [27], NAI and FAI [35]) that differ in chemical features and the half-time lifes, ranging from 0.25s for BzCN to 73 min for FAI at 37°C. The difference in reactivities obtained by two SHAPE reagents for the same sequence suggests that the interaction between the reagents and the RNA molecules can slightly change and provides complementary information [34, 36].

With the availability of high-throughput protocols, in the last decade, approaches using both SHAPE reagents [37, 38, 39] and DMS [40, 41] allowed studies of RNA in vivo [35, 40, 42] and the detection of regions involving RNA complexed to other biomolecules [43, 44]. The number of available quantitative data-sets obtained by chemical probing techniques is growing very rapidly and has already surpassed by two orders of magnitude the number of RNA 3D structures obtained with high-resolution techniques. Almost half of the chemical data-sets was obtained using SHAPE [45].

Several experimental studies have been carried out to understand the chemical origin of SHAPE reactivity and the relationship between SHAPE and local conformation of nucleotides [20, 34, 46, 47, 48]. SHAPE reagents are described to be more reactive with flexible nucleotides because these can sample multiple conformations, some of which enhance the reactivity of 2'-OH of the ribose [22, 49, 46]. Some recent studies suggested that there are three crucial structural parameters: i) the sugar pucker of the ribose; ii) the conformation of the adjacent phosphate group and iii) the presence of an RNA functional group within a short distance from the 2'-OH group [46, 31]. In particular, the nucleotides with C2'-*endo* sugar pucker appeared to be more reactive to SHAPE reagents, highlighting the important structural impact of the sugar pucker [46]. SHAPE reactivity seems mostly correlated with the base pairing (*cis*-Watson-Crick/Watson-Crick) and base stacking [20]. CMCT and DMS reactivity is essentially correlated with the WC base pairing status although fewer studies have been conducted.

In order to improve the incorporation of chemical probing data in secondary structure predictions, the link between the chemical reactivity and base-pairing has to be more accurately defined. All-atom molecular dynamics (MD) simulations are a suitable tool to investigate the flexibility of RNA molecules and attempt finding an atomistic, mechanical explanation for SHAPE, DMS and CMCT reactivities. Under the assumption that a crystal structure is representative of the major conformation adopted by the molecule in solution, and sampled with chemical probing, MD simulations from a known experimental structure can provide useful insights [47]. From a trajectory it is possible to compute local fluctuations, ribose pucker and detect WC pairing, to be compared to reactivities. Following this approach, in recent studies, SHAPE data-sets were correlated with structural fluctuations and parameters obtained by molecular dynamics simulations using an all-atom or a simplified representation for small (up to 78 nucleotides) RNA molecules [47, 50]. The best correlation between SHAPE reactivity and structural fluctuations was observed for the distance between consecutive nucleobases, and not for the 2'-OH groups or phosphate moieties, as one would have initially expected [47, 46]. The importance of C2'-*endo* sugar pucker in the SHAPE reactivity was also pointed out [48].

Although these studies have highlighted the role of some specific structural parameters in relation to SHAPE reactivity, some unexpected chemical reactivities remain to be understood and an exhaustive analysis of the interplay between local flexibility, structural features, WC pairing and SHAPE reactivity is still missing. Moreover, previous studies were limited in the sizes of the molecules considered, while the complexity of RNA structures increases dramatically with size and, with it, the difficulties to interpret chemical probing data and model RNA structure. Finally, to the best of our knowledge, no computational studies have been conducted to investigate the interplay between flexibility, dynamics of the hydrogen bonds and DMS/CMCT reactivity.

The aim of this work is to understand the interplay between flexibility, ribose pucker, canonical pairing and biochemical reactivity. We performed all-atom molecular dynamics on six RNA molecules ranging in size from 38 to 188 nucleotides, of known structures and chemical reactivities. In addition SHAPE, DMS and CMCT probing has been carried out on the complex RNA structure of the lariat capping ribozyme [51]. Among the different

SHAPE probes, we focused on 1M7 (see Figure 1 for its structure) as it is the most widely used in recent SHAPE experiments[52]. For each trajectory, we computed the fluctuations of a large set of geometrical parameters including distances, angles, helicoidal parameters and ribose pucker, the probability that a given nucleotide assumes a specific pucker family and the average number of hydrogen bonds at the WC edge of a nucleobase. To establish a correlation between all these metrics and the chemical reactivity, we performed an extensive statistical analysis.

Our study shows that the flexibility of individual chemical moieties is not enough to explain all the chemical reactivities and that the combination of multiple parameters is necessary. Moreover, we shed light on the correlation between DMS/CMCT reactivity and the dynamics of the hydrogen bonds.

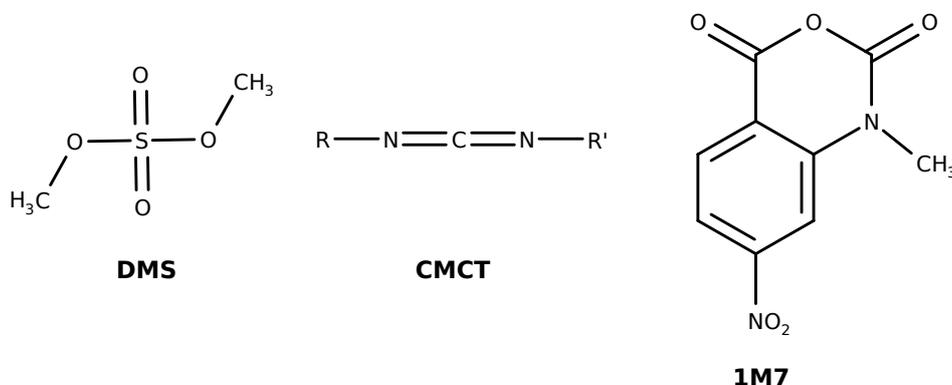


Figure 1: Chemical structure of the more common nitrogenous bases chemical probes (DMS and CMCT) and 1M7 SHAPE probe. DMS: dimethyl sulfate. CMCT: 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfate. 1M7: (1-methyl-7-nitroisatoic anhydride).

2. Methods

2.1. All-Atom Molecular Dynamics simulations

Molecular dynamics simulations were performed with the GROMACS 5 package [53, 54, 55, 56, 57] using the Amber ff99+parmbsc0 force field for RNA molecules [58]. To take in account the ligand in the binding site of the RNA molecule (like adenine in the adenine riboswitch), we performed DFT calculations to parameterize their atomic charges using Gaussian 16 [59]. We used the RESP procedure [60], which is part of the AMBER10 package to derive atomic charges [61]. Atomic coordinates and electrostatic potential of the ligand were obtained by DFT with the B3LYP functional and the 6-31G* basis set [59]. For the other parameters, we used the Generalized Amber Force Field [62].

The molecular systems were placed in cubic box and solvated with TIP4P-EW water molecules [63]. The distance between the solute and the box was set of at least 14 Å. The solute were neutralized with potassium cations and then K⁺Cl⁻ ion pairs were added to

reach the salt concentration of 0.1 M. We also added 0.02 M of MgCl_2 . We used the ion corrections of Joung et al. [64] as this force field has been shown to produce stable RNA structures [65]. The parameters for Mg^{2+} are taken from [66]. We opted for this force field for several reasons. First, this new set of parameters was developed to improve the kinetic properties of Mg^{2+} ions with water and with the phosphate ion and it was implemented in Amber99. Second, it also provided a better description of the structure of Mg-phosphate binding than previous sets [66]. Although Allner and co-workers have not tested the Mg^{2+} parameters with Cl^- ions, this force field is largely used in combination with the correction of Joung et al. [64] for KCl in the RNA community to study RNA flexibility and Mg^{2+} binding to RNA molecules [11, 67]. In our protocol some Mg^{2+} ions were placed at known binding sites while the remaining ions randomly replaced solvent molecules [67]. Long-range electrostatic interactions were treated using the particle mesh Ewald method [68, 69] with a real-space cutoff of 10 Å. The hydrogen bond lengths were restrained using P-LINCS [70, 56], allowing a time step of 2 fs [71]. Translational movement of the solute was removed every 1000 steps to avoid any kinetic energy build-up [72]. After energy minimization of the solvent and equilibration of the solvated system for 10 ns using a Berendsen thermostat ($\tau_T = 1$ ps) and Berendsen pressure coupling ($\tau_P = 1$ ps) [71], simulations were carried out in an NTP ensemble at a temperature of 300 K and a pressure of 1 bar using a Bussi velocity-rescaling thermostat [73] ($\tau_T = 1$ ps) and a Parrinello-Rahman barostat ($\tau_P = 1$ ps) [74]. During minimization and heating, the RNA backbone was kept fixed using positional restraints. These restraints were relaxed slowly during the equilibration. The length of the simulations were of 1000 ns. Table 1 summarises the number of ions and water molecules added in the box for each system under investigation.

2.2. Characterization of RNA flexibility

The flexibility of a biomolecule can be evaluated using several parameters that refer to different regions of the molecule. First, we calculated the atomistic Root Mean Square Fluctuations values and we averaged them to obtain the average fluctuations per residues that cannot differentiate the different contributions. In order to characterize the local flexibility, we considered distances, angles and rigid-body parameters and for each parameter we computed its fluctuations. The flexibility of a nucleotide is evaluated by the fluctuations exhibited by these parameters. Due to the time of our MD simulations, we can observe flexibility that occurs on rapid time scales (for example puckering changes). However we cannot observe base pair opening except on termini or where there are very weak non-WC base pairs. For example, non-WC base pair opening may not be fully captured in simulations on this timescale. These aspects go beyond the aim of this work.

Distance parameters. We computed the fluctuations for the following: i) the distance between consecutive C2 atoms (Figure 2); ii) the distance between consecutive C1' atoms; iii) the distance between consecutive P atoms and iv) the distance between O2' and O2 for the pyrimidine and between O2' and N3 for the purine, respectively. The first three distances are inter-base and the last one is intra-base (See Figure C.19 for all distances).

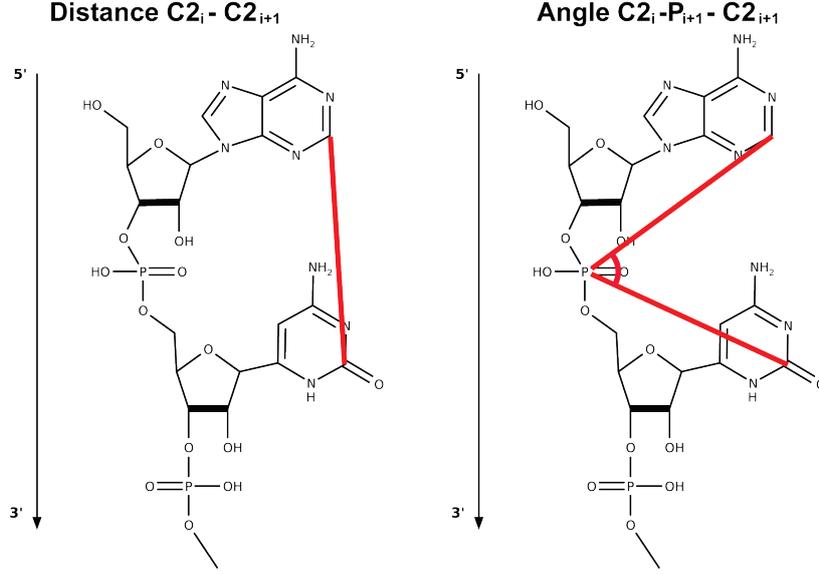


Figure 2: Left: Distance between consecutive C2 atoms ($C2_i-C2_{i+1}$). Right: Angle between consecutive C2, P and C2 atoms ($C2_i-P_{i+1}-C2_{i+1}$).

Angle parameters. We computed the fluctuations of eight angles in order to characterize the flexibility of the different chemical moieties. We defined the following angles: $C1'_i-C4'_i-P_{i+1}$, $C1'_i-P_{i+1}-C1'_{i+2}$, $C2_i-C1'_i-P_{i+1}$, $C2_i-C4'_i-P_{i+1}$, $C2_i-P_{i+1}-C2_{i+2}$ (Figure 2), $O2'_i-P_{i+1}-O5'_{i+1}$, $O3'_i-P_{i+1}-O2'_{i+1}$, $O5'_i-P_{i+1}-O2'_{i+1}$, where i represents the number of the nucleotide along the sequence. The angles are between either two or three consecutive bases (See Figure C.20 for all angles). Because of their periodicity, standard statistics cannot be used to compute the average and the fluctuations of angular values. For example, the arithmetic mean of 0° and 360° is 180° , which is misleading because for most purposes 360° is the same thing as 0° . Hence, we have to use the directional statistics instead [75]. The average of an angular value, α , can be defined as:

$$\langle \alpha \rangle = \begin{cases} \tan^{-1} \left(\frac{\langle C_p \rangle}{\langle S_p \rangle} \right) & \text{if } \langle C_p \rangle > 0, \langle S_p \rangle > 0 \\ \tan^{-1} \left(\frac{\langle C_p \rangle}{\langle S_p \rangle} \right) + \pi & \text{if } \langle C_p \rangle < 0 \\ \tan^{-1} \left(\frac{\langle C_p \rangle}{\langle S_p \rangle} \right) + 2\pi & \text{if } \langle C_p \rangle > 0, \langle S_p \rangle < 0 \end{cases} \quad (1)$$

where $\langle C_p \rangle = \frac{\sum_{i=1}^{N_s} \cos \alpha_i}{N_s}$ and $\langle S_p \rangle = \frac{\sum_{i=1}^{N_s} \sin \alpha_i}{N_s}$, α_i is the value of the angle at the time i and N_s is total number of snapshots of the trajectory. The angular fluctuation is defined as:

$$\sigma_\alpha = \sqrt{-2 \ln(\langle R_p \rangle)} \quad (2)$$

where $\langle R_p \rangle = \frac{\sqrt{\langle C_p \rangle^2 + \langle S_p \rangle^2}}{N_s}$.

Ribose conformations. The sugar ring is well described using pseudorotation parameters. Although there are four possible pseudorotation parameters for a five-membered ring [76], two in particular are useful to characterize the sugar conformation: the so-called phase (Pha) and amplitude (Amp). While the amplitude describes the degree of ring puckering, the phase describes which atoms are most out of the mean ring plane. We calculated these parameters using the expressions given below[77]:

$$\begin{aligned} \text{Amp} &= \sqrt{(a^2 + b^2)} \\ \text{Pha} &= \cos^{-1} \left(\frac{a}{\text{Amp}} \right) \end{aligned} \quad (3)$$

where $a = 0.4 \sum_{i=1}^5 \nu_i \cos[0.8\pi(i-1)]$ and $b = -0.4 \sum_{i=1}^5 \nu_i \sin[0.8\pi(i-1)]$ [77], with ν_i the ring dihedral i . This approach has the advantage of processing the ring dihedrals ν_1 (C1'-C2'-C3'-C4') to ν_5 (O4'-C1'-C2'-C3') in an equivalent manner.

Conventionally, sugar ring puckers are divided into 10 families described by the atom which is most displaced from the mean ring plane (C1', C2', C3', C4' or O4') and the direction of such displacement (*endo* for displacements on the side of the C5' atom and *exo* for displacements on the other side). Using the Curves+ program [78] for each simulation trajectory, for each nucleotide we computed the percentage of appearance for each family. In order to understand the interplay between the sugar conformation and the chemical reactivity, we grouped the sugar puckers into two large families. The sugar puckers C1'-*exo*, C2'-*endo*, C3'-*exo*, C4'-*endo* belong to the B-like family, while C1'-*endo*, C2'-*exo*, C3'-*endo*, C4'-*exo* belong to the A-like family.

Helicoidal parameters. In order to decompose the movement between two nucleotides, we adapted the definition of inter-base parameters, as described in Curves+ [78] and largely used for double-stranded nucleic acids, to the case of single-stranded RNA. The inter-nucleotide parameters consist of three independent translations (shift, slide, and rise) and three independent rotations (tilt, roll, and twist). We used eq.(1) and (2) to calculate averages and fluctuations of rotations.

2.3. Solvent accessibility

We explicitly tested the relationship between the accessibility of 2'-OH groups and SHAPE reactivity as well as the solvent accessibility of the nucleobases. We compared average accessibilities from 1 μ s-long MD simulation with experimental SHAPE reactivities. The solvent accessibility is computed by rolling a water probe sphere of radius 1.4 Å on the molecular surface.

2.4. Characterization of RNA base-pairing: hydrogen bonding analysis

Unlike double-stranded complementary molecules, in single stranded RNA the canonical Watson-Crick pairing is not the unique way to form a base-pair. A large variety of hydrogen bonds (HB) can be formed at the Watson-Crick edge, the Hoogsten edge or the sugar edge [79, 9, 10, 80]. Reaction of CMCT and DMS competes with the formation of HB at the WC

edge. In order to establish a relationship between the CMCT/DMS reactivity and hydrogen bonds, we distinguished between those hydrogen bonds that involve the WC edge and those that do not. For each nucleotide, we computed the average number of WC and non-WC hydrogen bonds and their fluctuations as well as the average of the total number of hydrogen bonds and its fluctuation for a given base.

At each frame of the trajectory we identified the HB based on cutoffs for the *Donor-H...Acceptor* distance and angle according to the Wernet-Nilsson criterion outlined in [81] as implemented in *mdtraj* [82] which considers a hydrogen bond formed if the distance between donor and acceptor heavy atoms is below a given distance cutoff dependent on the angle made by the hydrogen atom, donor, and acceptor atoms ("cone" criterion).

2.5. Statistical analysis

The degree of association between two independent variables, for example the fluctuations of a specific distance and the SHAPE reactivity, was first assessed through Pearson and Spearman correlation coefficients. The commonly used Pearson correlation coefficient (ρ_P) is a measure of the strength of a linear association between two variables while the Spearman correlation coefficient (ρ_S) is a non-parametric measure of rank correlation (statistical dependence between the rankings of two variables) and uses a simplified expression for the Pearson correlation coefficient. A correlation coefficient of .10 is thought to represent a weak or small association; a correlation coefficient of 0.30 is considered a moderate correlation; and a correlation coefficient of 0.50 or larger represents a strong or large correlation. A p -value of 0.05 is used as the cutoff for significance. If the p -value is less than 0.05, the correlation is significant because the *null* hypothesis of no difference between the means is rejected. If the p -value is larger than 0.05, we cannot conclude that a significant difference exists.

In order to quantify the strength of the association between N independent variables (predictor variables) and the dependent variable, we have to define the coefficient of multiple correlation, denoted R . The square of this coefficient is given by

$$R^2 = \mathbf{c}^T R_{xx}^{-1} \mathbf{c} \quad (4)$$

where $\mathbf{c} = (r_{x_1y}, r_{x_2y}, r_{x_Ny})^T$ with r_{x_Ny} the Pearson correlation coefficient between the variable x_N and the target value y and R_{xx}^{-1} the inverse of the correlation matrix between predictor variables, defined as:

$$R_{xx} = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_N} \\ r_{x_2x_1} & \ddots & & \vdots \\ \vdots & & \ddots & \\ r_{x_Nx_1} & \cdots & & r_{x_Nx_N} \end{pmatrix}. \quad (5)$$

For a set of parameters, we cannot use a simple Pearson correlation coefficient, because our data are discrete, hence we performed a test on the average values. We first discretized our data for the reactivity by grouping them in low reactivity, medium reactivity and high reactivity, and we built the contingency table of the modified experimental data and the

computed quantities (for example the average number of hydrogen bonds). Then, by applying the *null* hypothesis for a Chi-squared independence test, we performed the Chi-squared test to verify if our variables are independent or not. Significant contributions to Chi-squared were detected by looking at Pearson residuals: $\frac{X_{observed}-X_{expected}}{X_{expected}}$, which, for construction, follow a $N(0,1)$ distribution. The Chi-squared test is used to verify the independence between variables. It determines whether or not the variables are independent, but, when independence is rejected, it does not quantify the association between the variables.

To obtain a quantitative measure of the association, we considered the Cramer coefficient, Cramer’s V , defined by:

$$V = \sqrt{\frac{Chi^2}{Chi_{max}^2}} \quad (6)$$

where Chi^2 denotes the Chi-squared value obtained for the contingency table, and Chi_{max}^2 denotes the theoretical maximal Chi-squared of the contingency table, given by $Chi_{max}^2 = N \times \min(r - 1, C - 1)$ with N the total count of the table, r and c the table dimensions. Cramer’s V is, by construction, always between 0 (independence) and 1 (full dependence) and has a clear proportional interpretation: it is the fraction of the maximal deviation from independence that one would observe in case of full dependence [83].

Finally, in order to understand if there is a correlation between the chemical reactivity defined as low, medium and high, the non-parametric Kruskal-Wallis H test, which determines if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable, is used [83]. As for the other statistical approaches, a p -value lower the 0.05 indicates a significant correlation between the groups.

2.6. RNA structures and experimental data used as benchmark

In our study, we considered six RNA molecules of different size and function, for which both experimental reactivity data and 3D structures are available (see Figure 3). We performed all-atom molecular dynamics (MD) simulations for each RNA molecule starting from experimental structure and we computed the fluctuations of the parameters previously presented and their correlation coefficients with the chemical probing reactivities. The stability of RNA structures was assessed via Root-Mean-Square-Deviation and hydrogen bond analysis (see Appendix Appendix B for more details). SHAPE and CMCT/DMS data for all systems except the ribozyme (PDB ID: 4P8Z) were taken from available data-sets [84, 85, 86, 87, 88]. For the lariat capping ribozyme, the reactivity data were obtained for this study as previously described [89] (see Appendix A for details)[52, 90, 91, 52]. Table 1 summarizes the systems under investigation and the simulation setup.

For the comparison with simulation data we used normalized SHAPE, DMS and CMCT reactivities as explained in [84] and [86], respectively. For some nucleotides, the chemical reactivity could not be determined, and it is reported with either -999 or -10 in experiments. These nucleotides were excluded from our analysis.

PDB ID	Name	N_b	N_{H_2O}	N_{K^+}	N_{Cl^-}	$N_{Mg^{2+}}$	Exp.Data
2L1B	preQ1 Riboswitch	34	13366	57	35	5	S[84]
1Y26	<i>add</i> -riboswitch	72	34946	133	87	12	S[84],B[86]
1EHZ	Transfer Phe-RNA	77	51664	168	131	19	S[85],B[87, 88]
3OFC	5SrRNA	119	82272	267	209	30	S[84],B[86]
3PDR	M-box Riboswitch	161	96677	335	241	33	S[85]
4P8Z	LC ribozyme	188	77089	314	179	26	S*,B*

Table 1: RNA molecules under investigation. The following information is given: PDB ID code, name, number of nucleotides (N_b), number of water molecules (N_{H_2O}), K^+ , Cl^- and Mg^{2+} ions in the simulation box, and the type of probing experimental data available (S: SHAPE; B: CMCT and DMS).* See Appendix A.

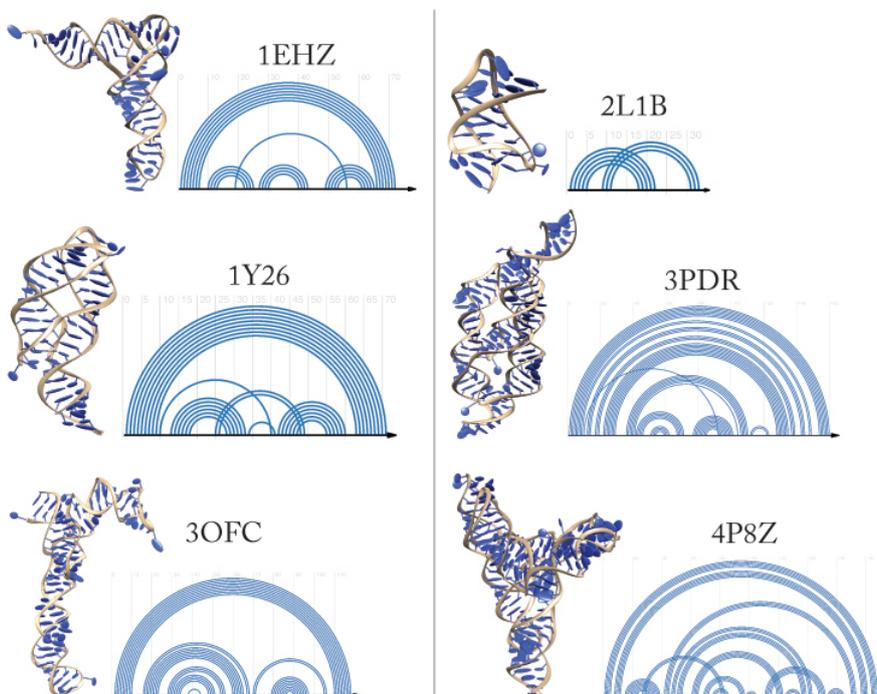


Figure 3: 3D structures of the RNA molecules under investigation and their arc diagram. 2L1B: PreQ1 Riboswitch. 1Y26: *add*-riboswitch. 1EHZ: Transfer Phe-RNA. 3OFC: 5SrRNA. 3PDR: M-box Riboswitch. 4P8Z: Lariat capping ribozyme.

3. Results and Discussion

In this section, we present and discuss the single and multiple correlations between SHAPE reactivity and fluctuations of the geometric parameters. We present the relationship between sugar pucker and 1M7 reactivity and the interplay between the former and canonical pairing, and finally, we discuss the relation between DMS/CMCT reactivity and geometrical parameters and the formation of hydrogen bonds at the WC edge.

3.1. Shape reactivity

We present here the statistical analysis of SHAPE reactivities in relation to the solvent accessibility of the 2'-hydroxyl position and the nucleobases, the fluctuation of nucleotides, to the fluctuations of distances, angles, and helicoidal parameters, and to the sugar pucker. We also investigate the relationship between sugar pucker and canonical pairing.

3.1.1. Solvent accessibility

First, we assessed whether SHAPE reactivity is correlated with the solvent accessibility of the 2'-hydroxyl position. As shown also by other studies, we find essentially no correlation between SHAPE reactivity and solvent accessibility at the 2'-hydroxyl group for five out of six RNA molecules (see Figure 4). However, a significant correlation between solvent accessibility of the nucleobases and SHAPE reactivity has been observed for all systems, and its strength spans from intermediate to strong for 5 out of 6 RNA molecules. We can speculate that the accessibility of the nucleobase to the solvent is typical of a configuration allowing also the stacking of the chemical probe, which has been shown to enhance the reactivity for 1M6 [34]. Further investigations are needed to better clarify this aspect by analyzing the reactivity mechanism. For 3OFC, the correlation is low possibly due to the fact that the 5SsRNA structure was extracted from the whole ribosome structure while our simulation does not take into account any of the existing interactions with ribosomal protein.

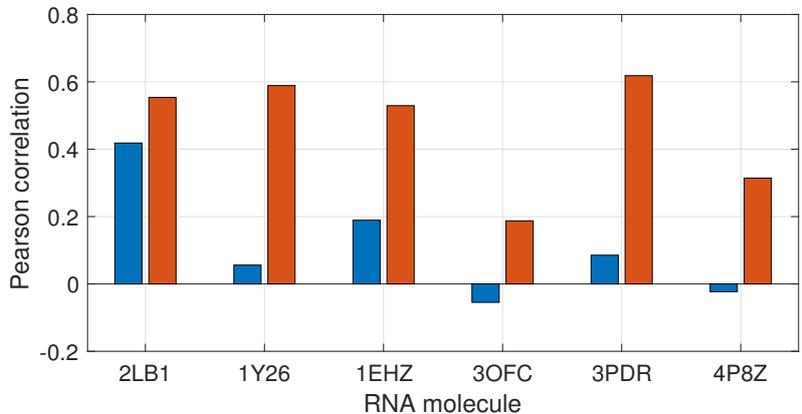


Figure 4: Pearson correlation computed for each RNA molecule between the solvent accessibility of 2'-hydroxyl position and the 1M7 reactivity (blue) and between the solvent accessibility of the nucleobase and the 1M7 reactivity (orange).

3.1.2. Root Mean Square Fluctuations (RMSF) and local fluctuation of distances, angles and helicoidal parameters

Previous results suggest that a key element for SHAPE reactivity is the flexibility of the nucleotide [46, 16, 47]. To further investigate this aspect, we computed the Pearson and Spearman correlation coefficients between the Root Mean Square Fluctuations (RMSF) and the experimental 1M7 reactivity (see Table 3.1.2). For four systems out of six, we obtained

a significant correlation between the two quantities while for 2L1B and 3OFC fluctuation profiles and reactivities exhibit important differences. In order to find some general rules linking structural properties with reactivities, these first results suggest that we need to focus on more local parameters to quantify the flexibility of the chemical moieties.

PDB ID	ρ_P	ρ_S
2L1B	0.10*	0.35
1Y26	0.48 ($2 \cdot 10^{-5}$)	0.24
1EHZ	0.75 ($1 \cdot 10^{-14}$)	0.62
3OFC	-0.01 (0.91)	0.02*
3PDR	0.84	0.13*
4P8Z	0.39	0.15*

Table 2: Pearson (ρ_P) and Spearman (ρ_S) correlation coefficients between the Root Mean Square Fluctuations (RMSF) and the experimental 1M7 reactivity. *: p -value higher than 0.05.

In this perspective, we extensively analyzed the correlations between local fluctuations of the distances and angles, described in section 2.2, and SHAPE reactivity. Figure 5 shows the average over all structures of the Pearson and Spearman correlation coefficients between distance and angular fluctuations and reactivities. In Appendix D, we report all Pearson and Spearman correlation coefficients in details (see Table D.5-D.8). Recalling that Pearson correlation evaluates the linear relationship between two continuous variables while Spearman correlation evaluates the monotonic relationship, we observe that Pearson correlation is more adequate for treating these data-sets. Hence, SHAPE reactivity changes proportionally with the increasing of fluctuations.

Looking in more detail at the relationship between fluctuations of distances and reactivity, the largest correlation is found between 1M7 reactivity and the distance $C2_i-C2_{i+1}$ ($\rho_P = 0.6$). However, also the fluctuation of the distance between consecutive $C1'$ atoms has a correlation value above 0.5. For several angles under investigation, the Pearson correlation coefficient is above 0.5 and it is significant ($C2_i-C4'_i-P_{i+1}$, $C1'_i-P_{i+1}C1'_{i+1}$ and $C2_i-P_{i+1}C2_{i+1}$). Although for the other angles we obtained low/moderate correlations coefficients, p -values lower than 0.05 assure that these correlations are also significant.

To provide a full picture of the flexibility of the nucleotide, we also analyzed the helicoidal parameters. Figure 5 shows the average of the Pearson and Spearman correlation coefficients of helicoidal rotational parameters (Tilt, Roll, and Twist) and 1M7 reactivity. We neglected the correlation coefficients with the translation parameters (Shift, Slide, Rise) because the p -value was higher than 0.05. For all rotational parameters, we found a moderate Pearson correlation while we observe a more significant Spearman correlation between Roll, Twist and 1M7 reactivity. This suggests the existence of a monotonic relationship for helicoidal parameters, less stringent than the linear relationship observed for distances and angles, capturing the more global nature of helicoidal over simpler structural parameters. Finally, the correlation coefficients between Twist and SHAPE reactivity are the least dispersed. Table D.9 and D.10 summarize all these data.

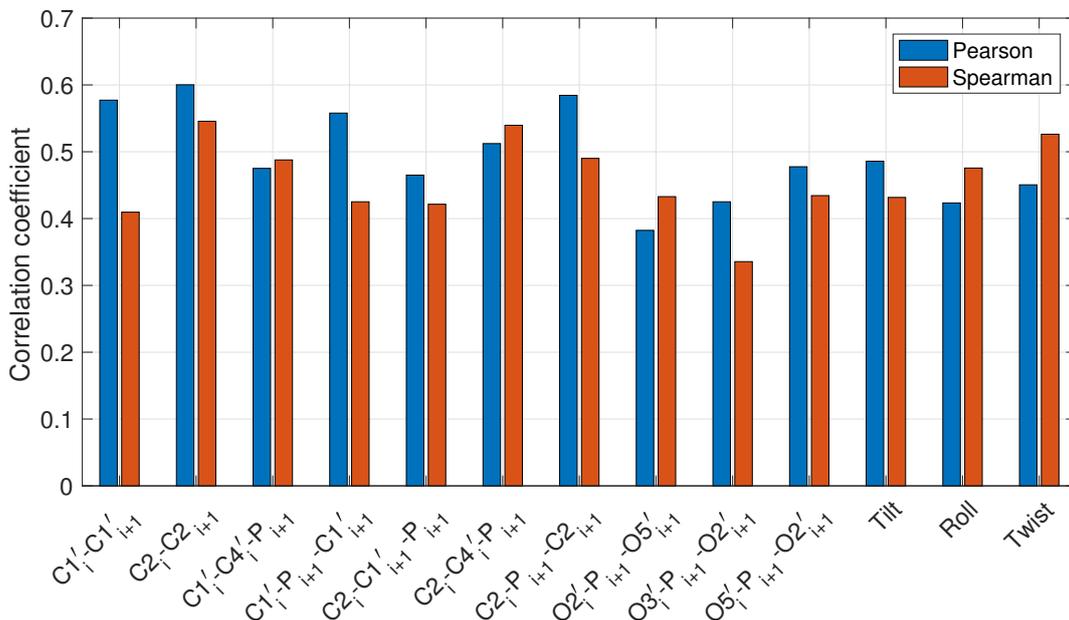


Figure 5: Average correlations coefficient between the 1M7 reactivity and different fluctuations of distances, angles and helicoidal parameters computed along the MD trajectory. From the left: fluctuations of distances $C1'_i-C1'_{i+1}$ and $C2_i-C2_{i+1}$; fluctuations of the angles $C1'_i-C4'_{i+1}-P_{i+1}$, $C1'_i-P_{i+1}-C1'_{i+1}$, $C2_i-C1'_{i+1}-P_{i+1}$, $C2_i-C4'_{i+1}-P_{i+1}$, $C2_i-P_{i+1}-C2_{i+1}$, $O2'_i-P_{i+1}-O5'_{i+1}$, $O3'_i-P_{i+1}-O2'_{i+1}$, $O5'_i-P_{i+1}-O2'_{i+1}$; fluctuations of tilt, roll, and twist. Blue: Pearson correlation. Orange: Spearman correlation.

The correlation coefficients of each parameter and 1M7 reactivity can vary significantly from one RNA structure to another (Table D.5-D.10). For example fluctuation of the distance between consecutive C2 atoms can range from 0.84 (strong correlation) to 0.26 (weak correlation). The weak correlation can be due to several reasons such as a sampling problem during the MD simulations (multiple conformations and size limit), a difference between the crystallographic structure and the RNA molecule in solution (moreover for 3OFC the structure were taken from the whole ribosome, so the presence of proteins and other RNA molecules can influence the structure of the native RNA molecule), or other structural parameters having an impact on the chemical reactivity, such as stacking or accessibility.

In order to better understand the meaning of correlation coefficients in this context, Figure 6 shows the RMSF, the fluctuation profile of the distance $C2_i-C2_{i+1}$ and the angle $C2_i-C4'_{i+1}-P_{i+1}$ and the 1M7 reactivity for the M-box Riboswitch RNA (PDB ID: 3PDR) and the 5SrRNA (PDB ID: 3OFC). For M-box Riboswitch RNA, the correlation between the SHAPE reactivity and the fluctuations profiles is strong because when the chemical reactivity is high, there is often a peak also in the fluctuations (high flexibility). For 5SrRNA, on the contrary, there is no correlation between the RMSF and the SHAPE reactivity, due a too broad RSMF profile that misses several peaks with respect to 1M7 profile. The low correlation between the fluctuation of the distance and 1M7 is mostly due to the absence of several peaks in the flexibility profile, that are however present in the angular fluctuation

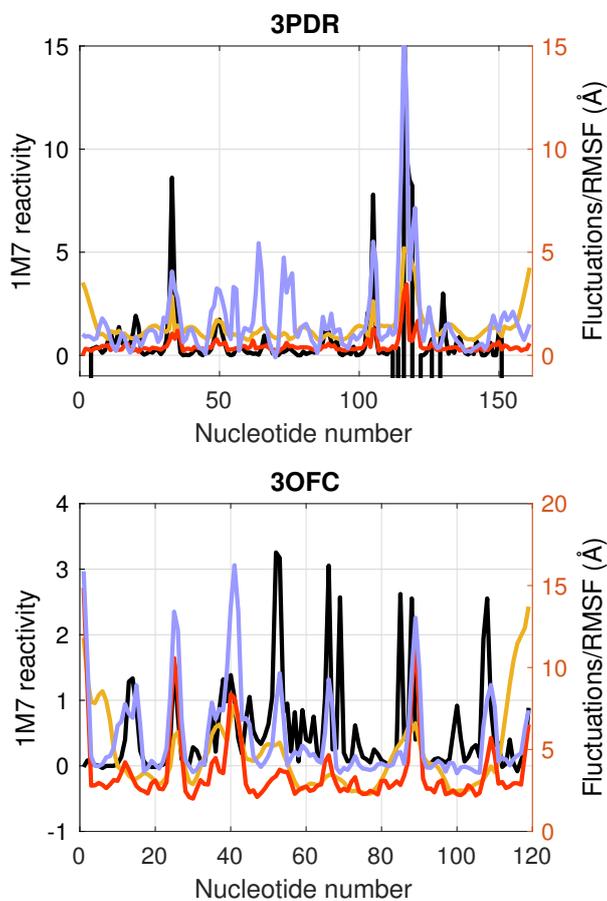


Figure 6: Comparison of the flexibility of the M-box Riboswitch RNA (PDB ID: 3PDR, upper panel) and the 5SrRNA (PDB ID: 3OFC lower panel). The SHAPE reactivities (black) are compared with the $C2_i-C2_{i+1}$ fluctuations (red), Root Mean Square Fluctuations, RMSF, (dark yellow) and $C2_i-C4'_{i+1}-P_{i+1}$ fluctuations (blue). Pearson correlation coefficients for 3PDR: 0.84 (RMSF), 0.85 ($C2_i-C2_{i+1}$), 0.79 ($C2-C4'-P$). Pearson correlation coefficient for 3OFC: -0.01(RMSF), 0.26 ($C2_i-C2_{i+1}$), 0.40 ($C2_i-C4'_{i+1}-P_{i+1}$).

(moderate correlation).

In this analysis, one problem can arise from the difference in normalization of experimental reactivities and of fluctuations from simulations which can bias the correlation coefficient. However, to develop a predictive approach for secondary and tridimensional structures, we need to establish a precise correspondence between the region of small, medium and high reactivity and their equivalent in the flexibility, or structural features, computed from molecular modeling. In this perspective, as a proof of concept, we analyzed the data according to ranks. We considered all values of a specific parameter, without making distinctions between individual structures and we defined ranks for the experimental and computational data classifying the SHAPE data into low reactivity (<0.3), medium reactivity (between 0.3 and 0.7) and high reactivity (>0.7). For simulation data, we normalized values by means of

the highest 10% and then we created ranks of uniform spacing up to a threshold value. By using box plots and the Kruskal-Wallis H test [83] we verified that there is a significant difference between 1M7 reactivity and the geometric parameters under investigation. In order to quantify the degree of association we calculated the Cramer’s V from the contingency table. Figure 7 shows the box plots for the fluctuations of the angle $C2_i-C4'_i-P_{i+1}$ and the twist for all systems. Based on the Kruskal-Wallis H test, there is a significant relationship between these variables, therefore, if the fluctuation of the parameter for a given nucleotide increases, its reactivity increases as well. However, based on the Cramer’s V parameter this association is moderate (between 0.3 and 0.4 for the fluctuations parameters). Although more qualitative than what previously presented, these rank analysis allow to state with more certainty the existence of relationship between reactivity and flexibility, which could be misinterpreted by more simple statistics.

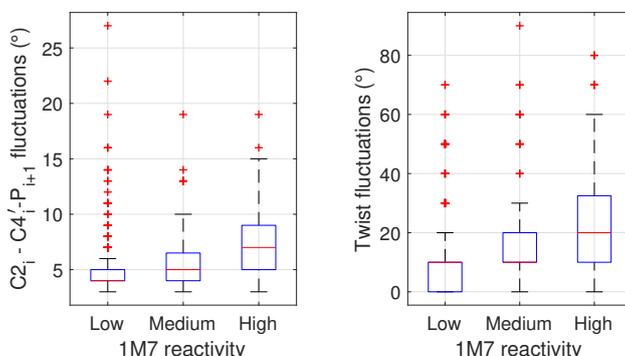


Figure 7: Box plot of the fluctuation of the angle $C2_i-C4'_i-P_{i+1}$ (left) and of the twist (right) as a function of the intensity of the reactivity (low, medium or high).

3.1.3. Multiple correlation approach

The previous correlation analysis suggests that several parameters play a role in the chemical reactivity and multiple parameters at once might influence the interaction with the probe. We therefore considered several sets of parameters for which we computed multiple correlation coefficient with SHAPE reactivity using eq.(4). Table 3 provides multiple correlation coefficients for five sets of parameters that combine the fluctuations of distances, angles and helicoidal parameters. Parameter sets are defined in increasing complexity as: i) M_1 fluctuations of the distance $C2_i-C2_{i+1}$ and the angle $C1'_i-P_{i+1}-C1'_{i+1}$; ii) M_2 fluctuations of distance $C2_i-C2_{i+1}$ and of the angle $C1'_i-C4'_i-P'_{i+1}$; iii) M_3 fluctuations of distance $C2_i-C2_{i+1}$ and of the angles $C1'_i-C4'_i-P'_{i+1}$, $C1'_i-P_{i+1}-C1'_{i+1}$, $C2_i-C1'_i-P_{i+1}$, $C2_i-C4'_i-P_{i+1}$, $C2_i-P_{i+1}-C2_{i+1}$; iv) M_4 fluctuations of distance $C2_i-C2_{i+1}$, of the angle $C1'_i-C4'_i-P'_{i+1}$ and the rotational helicoidal parameters; v) M_5 the fluctuations of distance $C2_i-C2_{i+1}$ and $C1'_i-C1'_{i+1}$, of the angle $C1'_i-C4'_i-P'_{i+1}$ and the rotational helicoidal parameters.

With respect to the single correlation, the combination of several parameters in sets increases the correlation between the flexibility and the experimental data. For example, in the case of 3OFC, single correlation coefficients are weak (see Table D.5-D.10) while

multiple correlation coefficients range from weak to moderate (strong) with the increasing of the diversity of parameters included in the set of the multiple correlation. This suggests that the SHAPE reactivity depends on the local flexibility of several chemical moieties in agreement of what already shown in the profiles reported in Figure 6. We can observe that for systems for which single parameter profiles mainly agree with one another, the different multiple correlations sets give all similar correlation results, as expected, while for systems where single sets exhibit significant discrepancies, combining them in multiple sets improves correlations. These results are in agreement with previous attempts to understand the origin of the SHAPE reactivity where several simple rules were proposed based on the geometry of the ribose [46].

PDB ID	M_1	M_2	M_3	M_4	M_5
2L1B	0.62	0.66	0.71	0.83	0.86
1Y26	0.78	0.75	0.86	0.74	0.84
1EHZ	0.56	0.56	0.58	0.51	0.63
3OFC	0.29	0.45	0.56	0.48	0.59
3PDR	0.85	0.86	0.86	0.76	0.87
4P8Z	0.35	0.36	0.45	0.44	0.41

Table 3: Multiple correlation coefficient between a set of parameters computed along the MD simulations and the 1M7 reactivity calculated by using eq.(4) . M_1 : fluctuations of distance $C2_i-C2_{i+1}$ and the angle $C1'_i-P_{i+1}-C1'_{i+1}$. M_2 : fluctuations of distance $C2_i-C2_{i+1}$ and of the angle $C1'_i-C4'_i-P'_{i+1}$. M_3 : fluctuations of distance $C2_i-C2_{i+1}$ and of the angles $C1'_i-C4'_i-P'_{i+1}$, $C1'_i-P_{i+1}-C1'_{i+1}$, $C2_i-C1'_i-P_{i+1}$, $C2_i-C4'_i-P_{i+1}$, $C2_i-P_{i+1}-C2_{i+1}$. M_4 : fluctuations of distance $C2_i-C2_{i+1}$, of the angle $C1'_i-C4'_i-P'_{i+1}$ and the rotational helicoidal parameters (roll, tilt and twist). M_6 : fluctuations of distances $C2_i-C2_{i+1}$ and $C1'_i-C1'_{i+1}$, of the angle $C1'_i-C4'_i-P'_{i+1}$ and the rotational helicoidal parameters (roll, tilt and twist).

3.1.4. Ribose flexibility

Another important feature of the nucleotide is the pucker of the ribose and its fluctuations. From molecular dynamics simulations we computed for each nucleotide the probability that a given pucker family is present as well as the average and the fluctuation of the pucker phase and pucker amplitude (see eq.(3)). Figure 8 shows a stacked plot of the pucker probabilities for 1Y26. The stacked plots for the other molecular systems are presented in Appendix E. As expected, the $C3'$ -endo is the most common state and the nucleotide can also assume the rarer $C2'$ -endo conformation [46]. However, the ribose does not only assume the $C2'$ -endo conformation, but also explores other transient states (for example the $C1'$ -exo) that are less stable. To better understand the stability and the kinetics of these conformations, we also computed the matrix of transition probabilities for the puckering conformations. Hence, we computed the probability of moving from state i to state j and to stay in the same state. In this matrix, each row sums to 1. Figure 9 shows the matrix of transition probabilities obtained in this study. This matrix can allow us to analyze the stability of each puckering conformation with respect to the transition states. For example, the $C3'$ -endo has a higher probability to move to either $C2'$ -exo or $C4'$ -exo. As expected, these

intermediate states are not stable and their transition probabilities is comparable with the probability to stay at the same state. In the case of *C2'-endo*, the most probable transition states are *C1'-exo* and *C3'-exo*. We also observed that the exchange rate is faster for the *C3'-endo* conformations than *C2'-endo* ones.

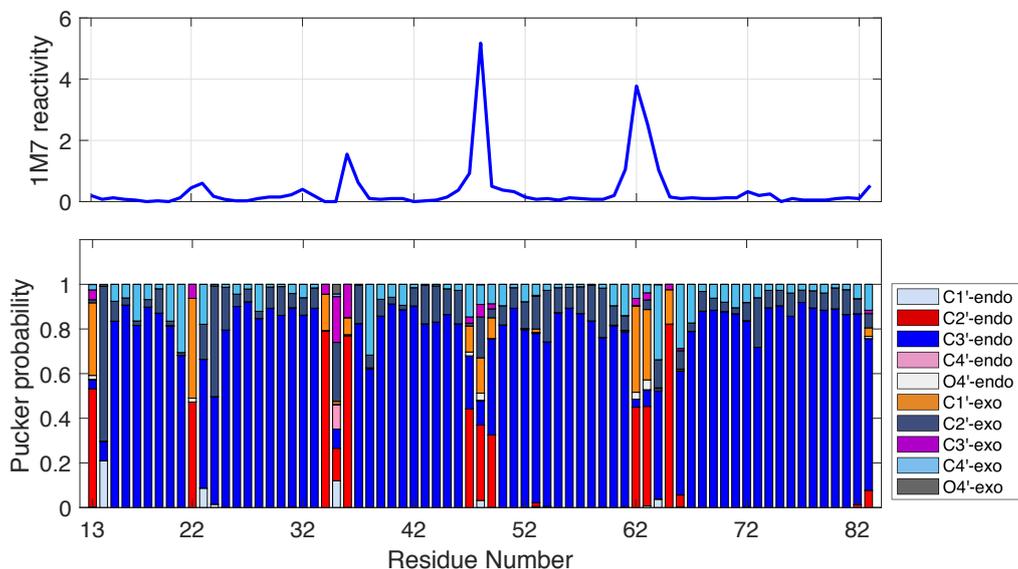


Figure 8: 1M7 reactivity profile (top) and stacked bar plot of the pucker probability along MD simulation (bottom) for 1Y26. Red: *C2'-endo*, Pink: *C4'-endo*, Orange red: *C1'-exo*, Magenta: *C3'-exo*. Blue: *C3'-endo*, light blue: *C1'-endo*, dark blue: *C2'-exo*, cyan: *C4'-exo*. Light grey: *O4'-endo*, dark grey: *O4'-exo*. Red scale: B-family. Blue scale: A-family.

With the aim of understanding the relationship between ribose conformations and flexibility of the nucleotide, we first computed the Spearman correlation coefficient between the fluctuations of the pucker phase and the geometric parameters. These correlations are always significant and from moderate to strong indicating that the fluctuations of the geometric parameters reflect quite well the flexibility of the ribose conformation without distinction between the specific pucker families (see Table D.11-D.13). The highest correlation coefficients obtained for the three geometric classes (distance, angle and helicoidal parameters) were found using the fluctuation of the distance $O2'_i - b_i$ (0.76), the angle $C1'_i - C4'_i - P_{i+1}$ and roll (0.72), respectively. In order to establish a direct correlation with the ribose pucker, we also computed the Spearman correlation coefficient between the probability that the ribose conformation belongs to the B-like family (*C2'-endo*, *C4'-endo*, *C1'-exo* or *C3'-exo*) and the geometric fluctuations. For the three geometric classes, the highest correlation was 0.63 for fluctuations of the distance $O2' - N3/O2$, the angle $C1'_i - P_{i+1} - C1'_{i+1}$ and roll, respectively. Weeks and co-workers showed that when this distance ($O2' - N3/O2$) is short and the ribose assumes *2'-endo* conformations the reactivity can be enhanced [31]. Our results indicate that *C2'-endo*-like conformations are correlated with a larger variability of this distance allowing the system to explore short-distance conformations known to enhance reactivity. Our findings provide a possible structural explanation for the higher reactivity and highlight that

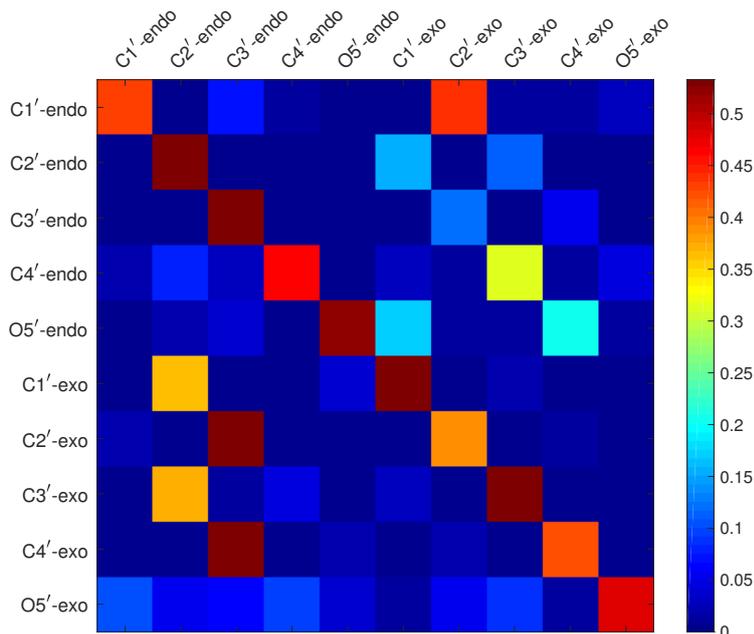


Figure 9: The matrix of transition probabilities between the ribose pucker i and j . Pucker under investigation: $C1'-endo$, $C2'-endo$, $C3'-endo$, $C4'-endo$, $O4'-endo$, $C1'-exo$, $C2'-exo$, $C3'-exo$, $C4'-exo$, $O4'-exo$.

the $C2'-endo$ conformation and the cognate transient states better reflect the local flexibility of a nucleotide, than only the $C2'-endo$ as previously suggested [31, 46].

We then analyzed the correlation between the ribose pucker, its fluctuation and the 1M7 reactivity. It is well known that $C2'-endo$ can enhance the nucleophilicity of the 2'-hydroxyl by allowing non-standard conformations of the nucleotide [46]. Figure 8 compares the 1M7 reactivity with the pucker probability for each nucleotide for 1Y26: the $C2'-endo$ conformation together with the $C2'-endo$ transient states are the most probable among hyper-reactive nucleotides, as previously observed by Weeks and co-workers [46] Figure 10 shows the box plot for the probability for a given nucleotide to assume a $C3'-endo$ pucker and its fluctuation as a function of the 1M7 reactivity. Based on the Kruskal-Wallis H test, the $C3'-endo$ pucker and the 1M7 reactivity are anti-correlated: high reactivity is related to either the $C2'-endo$ pucker or the other transient conformations between the $C2'-endo$ and $C3'-endo$ pucker. To establish the degree of association, we computed the Cramer's V between the two sets. We obtained a value of 0.31, meaning that there is a moderate correlation between 1M7 reactivity and ribose pucker. In Figure 10 (middle box) we can observe that although the ribose has assumed a rare conformation, the 1M7 reactivity is low, and this low reactivity is associated with a high flexibility of the ribose. We can then speculate that other factors beside the sugar pucker might influence the reactivity. For example, low reactivity might also be the consequence of a low accessibility caused by

stacking interactions or of a modification in the distance between the base and the ribose. Despite this, 72% of medium and high reactivities are obtained in the cases of C2'-endo conformation and its transient states, hence we can also conclude that this states are highly over-represented among hyper-reactive nucleotides.

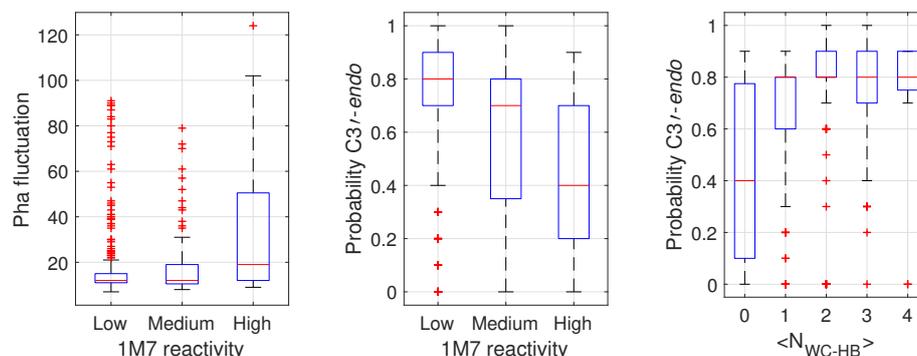


Figure 10: Box plot of the fluctuation of the sugar pucker as a function of the intensity of the reactivity (low, medium or high) (left), of the probability for a given nucleotide to assume a C3'-endo pucker as a function of the intensity of the reactivity (low, medium or high) (centre) and the average number of hydrogen bonds formed at the Watson-Crick edge, $\langle N_{WC-HB} \rangle$ (right).

In order to predict secondary structures taking SHAPE data into account, the common assumption included in the algorithms [92, 16, 93, 94] is that reactive nucleotides correspond to unpaired bases. Hence, understanding if there is a direct correlation between the sugar pucker and the formation of hydrogen bonds at the Watson-Crick edge is fundamental. Before displaying the results of this analysis, we would like to point out that unbiased all-atom MD simulations do not allow to investigate secondary structures rearrangements (like folding), but only local changes. As previously mentioned in section 2.1, we tested the stability of our simulations by analyzing the conservation of some specific hydrogen bonds and multiplets. We also checked that the convergence of the average number of WC hydrogen bonds ($\langle N_{WC-HB} \rangle$) was reached for each molecular system (see Figure B.17). Along the MD trajectory, $\langle N_{WC-HB} \rangle$ is stable and fluctuates around an average value for each molecular system under investigation. To better characterize the WC-HB, we also computed their lifetimes, using the following auto-correlation function:

$$C_{HB}(t) = \frac{\langle h(0)h(t) \rangle}{\langle h \rangle} \quad (7)$$

where $h(t)$ is equal to 1 if the H-bond exist from time 0 to time t . Figure B.18 shows some examples of the auto-correlation function $C_{HB}(t)$ obtained for our molecular systems. The lifetime of WC-HB within loops, internal bulges or at helix's termini is around 250-300 ps. As expected, their lifetime is longer in the helical regions where it spans from 500 ps up to values larger than 5 ns.

Since simulations converged with respect to base pairing, we can analyze the relationship between the sugar pucker and the average number of hydrogen bonds formed by the Watson-Crick edge. By analyzing the box plot (see Figure 10), we can conclude that the number of

hydrogen bonds at the Watson-Crick edge increases with the probability that a nucleotide assumes a C3'-*endo* pucker. However, nucleotides with a C2'-*endo* pucker can also be involved in canonical pairing. This observation is also confirmed by a low Cramer’V value (0.28) and a Spearman correlation coefficient of 0.42.

3.2. DMS/CMCT reactivity

Although several studies were carried out to understand the molecular implications of SHAPE reactivity, to best of our knowledge no computational studies have been carried out to understand the relationship between nucleotide flexibility and chemical reactivity for DMS or CMCT. We computed Spearman correlation coefficient for each individual structure (see Table D.14) and for all the data together between the average number of hydrogen bonds (HB) formed at WC edge and the DMS/CMCT reactivity. For both reagents, we obtained a significant anti-correlation with a correlation coefficient of -0.59 and -0.42 for DMS and CMCT respectively. Calculating the Cramer’V, we also verified that there is no bias in the correlations. The Cramer’V confirms a relatively strong and moderate correlation as shown by Spearman coefficient. Figure 11 summarizes these correlations through box plots. We observe instances in which the base is reactive while in the crystal structure it forms a canonical pair as well as instances in which the base is non-reactive but unpaired in the crystal. We can speculate that such discrepancies because crystallization stabilizes or select for some interactions that are poorly stable in the chemical probing conditions in solution. It is well known that some local structural differences can be observed between RNA molecules in solution and in the crystal. Indeed, Vincens and co-workers showed that local conformational changes that take place during crystallization can be identified by comparing SHAPE data obtained in solution, crystallization solution and crystal [95].

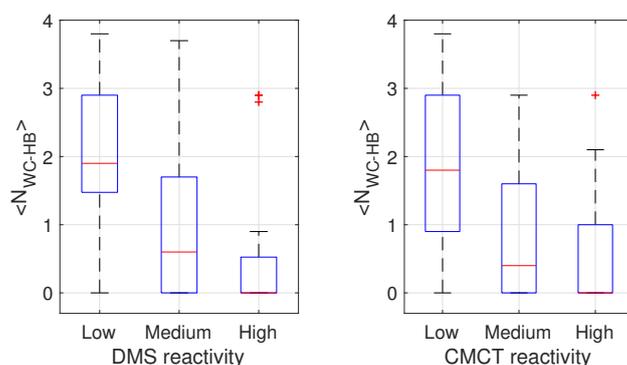


Figure 11: Box plot of the average number of hydrogen bonds formed at the Watson-Crick edge, N_{WC-HB} , as function of the DMS (left) and CMCT (right) reactivity (low, medium or high).

We also computed the correlation coefficient between the DMS/CMCT reactivity and the fluctuations of the same geometric parameters investigated for SHAPE reactivity. For all the RNA molecules, we did not find any geometric parameter with a significant correlation with DMS reactivity. In contrast, for CMCT reactivity, we observed a significant positive

correlation with the fluctuation of distance $C1'_i-C1'_{i+1}$ ($\langle\rho\rangle_P = 0.52$) and of the angle $C1'_i-P'_{i+1}-C1'_{i+1}$ ($\langle\rho\rangle_P = 0.57$). Significant positive correlations were also obtained with all helicoidal parameters (shift, slide, rise, tilt, roll, and twist, see Table Appendix D for all the details). The highest correlation coefficient ($\langle\rho\rangle_P = 0.57$) is between the CMCT reactivity and the twist. The presence of a significant correlation between some structural parameters and CMCT suggests that the microscopic interpretation for CMCT is more complex than for DMS and it may explain the weaker correlation between its reactivity and canonical pairing.

4. Conclusions

In this work we investigated the relation between RNA flexibility, described through the fluctuations of distances, angles, helicoidal parameters and ribose pucker, and the reactivity towards SHAPE, DMS and CMCT probes. For six RNA molecules of different sizes and shape, we performed all-atom molecular dynamics simulations reaching the timescale of one microsecond. We performed a statistical analysis to characterize the relationship between SHAPE/DMS/CMCT reactivity and several geometrical parameters, including distances, sugar pucker and hydrogen bonds. Our statistical approach allows us to study the trend between chemical reactivity and structural parameters and to gather some mechanical understanding of the information from chemical probing.

First, we verified that there is no significant correlation between chemical reactivity and the accessibility of 2-OH'. In contrast, a significant correlation was observed between chemical reactivity and nucleobase solvent accessibility. We analyzed distances, angles and inter-base helicoidal parameters and we computed their fluctuations from the MD simulations. When we focus on one single structural parameter, we observe moderate correlations with 1M7 reactivity while combining two or more geometrical parameters significantly improves correlation. Hence we conclude that the flexibility of a single parameter does not allow to fully describe SHAPE reactivity. In addition we observed for some nucleotides that the base flexibility and therefore its pairing status is not always correlated with the ribose flexibility. Another important feature for the SHAPE reactivity is the ribose pucker. As expected, the C3'-*endo* conformation is the most common. However, the ribose can assume not only the C2'-*endo*, but also the other intermediate conformations. We also observed that the C3'-*endo* conformation changes faster than the C2'-*endo*. Based on our statistical analysis, we observed that the fluctuations of geometrical parameters are only moderately correlated to the pucker families and their flexibility. Our simulations confirm the general assumption that a high SHAPE reactivity is mainly associated to high sugar flexibility, but show a complex scenario on how this can happen. Despite this, the C2'-*endo* and its transient states are highly overrepresented among hyper-reactive nucleotides. In addition, our results suggest the possibility for a nucleotide to react with 1M7 even though it is rigid, highlighting that our understanding of SHAPE chemical reactivity is still incomplete and will necessitate further chemical studies. Although current secondary structure prediction models consider a SHAPE-reactive nucleotide as unpaired, our study suggest that there is not a strict correlation between canonical pairing and sugar pucker flexibility.

We also analyzed the interplay between the geometrical parameters, the canonical pairing (i.e. hydrogen bonds at the WC edge) and the DMS/CMCT reactivity. Interestingly, we found a moderate significant correlation between CMCT reactivity and the fluctuation of distance $C1'_i-C1'_{i+1}$, of the angle $C1'_i-P'_{i+1}-C1'_{i+1}$ and all helicoidal parameters. However no significant correlation between the structural parameters and DMS reactivity was observed. These findings suggest that CMCT reactivity is sensitive also to the local flexibility of the nucleotide, and not only to its WC paired status, while DMS is less sensitive to the local environment of the base. For both reagents, we obtained a significant anti-correlation between their reactivity and the average number of hydrogen bonds (HB) formed at WC edge. Their correlation are stronger than those obtained between structural parameters and 1M7 reactivity, with the correlation between canonical pairing and DMS reactivity being the strongest. Looking at what bases are found paired in the experimental structures, we would have expected these correlations to be even higher than what we obtained, but we detect instances in which bases that are paired in the experimental structure, and stable in the simulations, exhibit a high reactivity, indicating an unpaired base from chemical probing. This result suggests the need for further investigation at experimental and computational level of the effect of the chemical reaction on the nucleotide and its environment. Despite these anomalies, our study highlights the intrinsic relationship between DMS/CMCT reactivity and canonical pairing, which should therefore be included into 2D and 3D predictions whenever possible.

Overall, our study suggests a complex picture of the chemical probing process where the reactivity is influenced simultaneously by several local parameters and of which we still do not have a full understanding. Further studies are necessary to draw a connection of the interplay of data sharing common features, based on the flexibility propensity of the different chemical moieties, the ribose pucker, the nucleotide accessibility and other geometrical properties of the environment of the base, with the chemical reactivity, and from this, draw some general rules to be exploited in secondary structure predictions. Even if several questions on the relationship between reactivity and RNA structures remain open, at this stage, our work has highlighted the structural features with a strong correlation with the chemical reactivity. This information can already be harnessed in 3D molecular modelling by introducing soft-constraints on local structural features (for example on the sugar pucker) directly biasing simulations in agreement with the chemical probing profile. For the 2D prediction algorithms, having a physical interpretation of the probe interaction with the RNA, will allow a better evaluation of the pseudo-energies and pairing probabilities. Moreover, further investigation of the relationship between local flexibility, base pairing as well as stacking geometries, and chemical reactivity will allow to identify specific motifs and nucleotide local geometries, thus improving the existing scoring functions of 2D algorithms. This strategy can be used for *ab initio* folding as well as for refinement of experimental structures when the molecule undergoes changes in the environment inducing rearrangements in its secondary structure, enhancing our understanding of RNA structural plasticity.

Acknowledgements

We are grateful to the GENCI France supercomputers for generous allocation of CPU time (DARI A0030710273 and DARI A0040710357 allocation grants). We thank Dr. Lavery for proving us the new version of CURVES+ not yet published and Dr. Martin for helpful discussions about statistical analysis.

Abbreviations

1M7: 1-methyl-7 nitrosatoic anhydride; CMCT: 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfate; DMS: dimethyl sulfate; HB: hydrogen bond; MD: Molecular dynamics; ρ_P : Pearson correlation coefficient; ρ_S : Spearman correlation coefficient; RMSF: Root Mean Square Deviation; RNA: Ribonucleic acid; SHAPE: 2'-hydroxyl acylation analyzed by primer extension; WC: Watson-Crick.

Author Contributions

EF: designed the work, performed Gaussian calculations, ran MD simulations, analyzed the results, wrote the manuscript. AC: analyzed the results. DA: performed and analyzed the SHAPE, DMS and CMCT experiments for 4P8Z, wrote the appendix. BS: designed the work and the probing experiments for 4P8Z, contributed to manuscript writing. SP: designed the work, wrote the manuscript.

Appendix A. Lariat capping ribozyme from *Didymium iridis* probing and secondary structure prediction

SHAPE. Lariat capping ribozyme from *Didymium iridis* secondary structure was predicted with SHAPE probe 1M7 (1-methyl-7 nitrosatoic anhydride). 6 pmol of RNA were resuspended in 18 μL of water, denatured at 80 °C for 2 min and cooled down at room temperature for 10 min in a probing buffer (40 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl_2). Then, after incubation at 25 or 37 °C for 10 min, RNA was treated with 2 mM of 1M7 or DMSO (negative control) and incubated for 5 min at 25 or 37 °C. The modified or unmodified RNA was purified by ethanol precipitation and pellets were resuspended in 10 μL of water. Modification was revealed by reverse transcription using 5' fluorescent primers (D2 or D4 WellRED, Sigma Aldrich) and M-MLV RNase H-reverse transcriptase (Promega). To read the whole sequence of lariat capping ribozyme we used this following primer: 258NP_rev: 5' CTG-TGA-ACT-AAT-GCT-GTC-CTT-TAA 3'. Briefly, RNA treated was denatured for 3 min at 95 °C with 1 μL of DMSO and cooled in ice for 3 min. Three μL of primer were added and samples were incubated for 5 min at 65 °C and for 10 min at 35 °C and subsequently cooled on ice. Reverse transcription was performed in several steps: 2 min at 35 °C, 30 min at 42 °C and 5 min at 55 °C. cDNA were separate and detected by capillary electrophoresis (Beckman Coulter, Ceq8000). Data were analyzed using software QuSHAPE[96]. RNA probing was performed in triplicate for each *Didymium* with distinct RNA preparations.

DMS. 6 pmol of RNA were added to the same buffer with MgCl_2 used in the SHAPE experiments. Then, RNA was denatured at 80°C for 2 min and $1\ \mu\text{L}$ of DMS was then added (DMS:ethanol dilution of 1:12). After 5 min of incubation at 37°C , the reaction was stopped and 400 mM of Tris were added at pH 7.5. The samples were then put in ice before cleaning them via precipitation. The pellets were taken from $10\ \mu\text{L}$ of water.

CMCT. 6 pmol of RNA were denatured at 80°C for 2 min. Then, we added it to the 50 mM of potassium borate at pH 8 and 5 mM of MgCl_2 , the solution was incubated at room temperature for 5 min. Then, $10\ \mu\text{L}$ of CMCT ($42\ \text{mg}\cdot\text{mL}^{-1}$) were added and then incubated for 10 mins at 37°C . The samples were put in ice to stop the reaction. The modified or unmodified RNA was purified by ethanol precipitation and pellets were resuspended in $10\ \mu\text{L}$ of water.

Appendix B. Analysis of stability of RNA structures along all-atom molecular dynamics simulations

In this appendix, we present supplementary data to assess the stability of RNA molecules along all-atom MD simulations. To this end, we first computed Root-Mean-Square-Deviation (RMSD) along the trajectories by using the starting structure as reference (see Figure B.12-B.14 and the RMSD between the average structure of the last 50 ns and the experimental structure (Table B.4). For 4 systems out of 6, the RMSD values are equal or below $3.1\ \text{\AA}$. However, we would like to point out that RMSD is not always the best metric to assess the quality of 3D RNA structures as shown in the RNA-Puzzles Rounds [97, 3]. Therefore, we also compared the number and the type of hydrogen bonds along the MD simulations with the ones present in the crystallographic structure and finally we analyzed if the multiplets are conserved or not along the MD simulations.

In the two cases where the RMSD value is higher than $3.5\ \text{\AA}$ (4P8Z and 3OFC), the larger RMSD is due to inter-domains relative movements that do not affect individual domains structure. Chemical reactivity is influenced only by local conformations and should not be affected by such conformational changes. For 4P8Z, a relative inter-domains motion is responsible of the larger RMSD. Figure B.15 shows the distance time series for some relevant hydrogen bonds involved in the multiplets for 4P8Z. Thanks to this analysis, we can better assess the stability of 3D structure. For 3OFC, we observed that both an open (0-300 ns) and a close (400-1000 ns) conformation are present. We would like to point out that for 3OFC the structure were taken from the whole ribosome while the chemical probing has been performed on the naked RNA. Despite that, hydrogen bonds are mostly all conserved along the MD simulation and the triplets are stable. Figure B.16 shows the distance time series for some relevant hydrogen bonds involved in the triplets in 3OFC. We can observe that for the triplet C31-G51-G53, the triplet is mostly stable along the simulations, however the system investigates for 20 ns another unstable state and then returns to the native triplet.

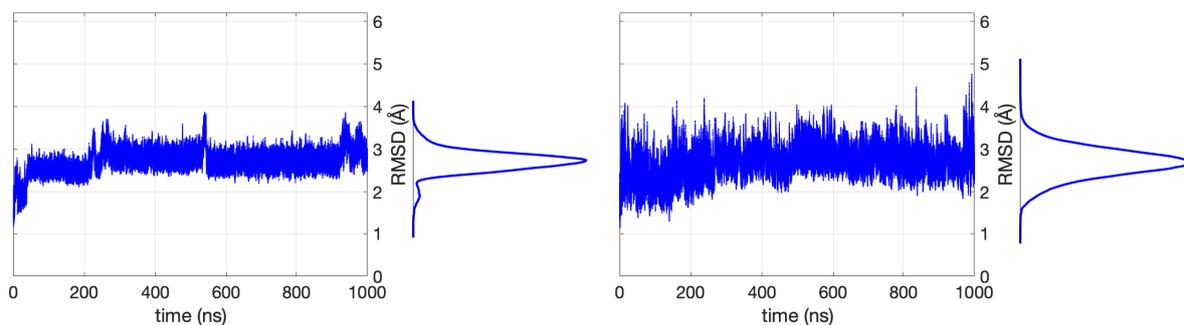


Figure B.12: Root-Mean-Square-Deviation (RMSD) time series and its marginal histogram of 2L1B (left) and 1Y26 (right).

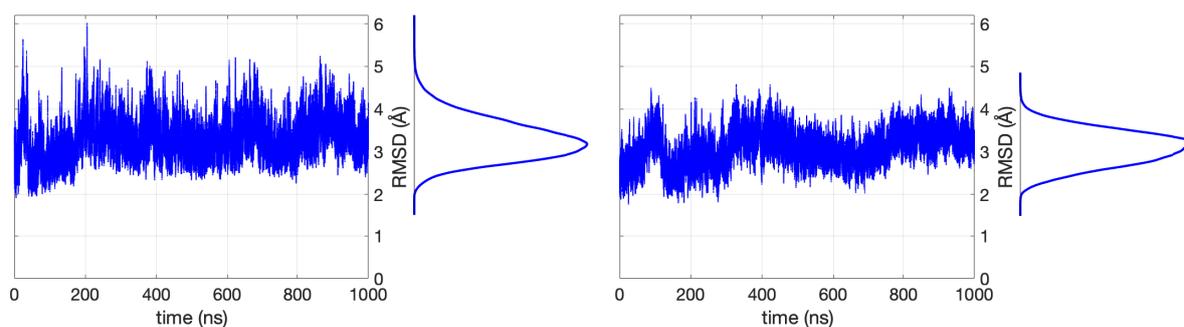


Figure B.13: Root-Mean-Square-Deviation (RMSD) time series and its marginal histogram of 1EHZ (left) and 3PDR (right).

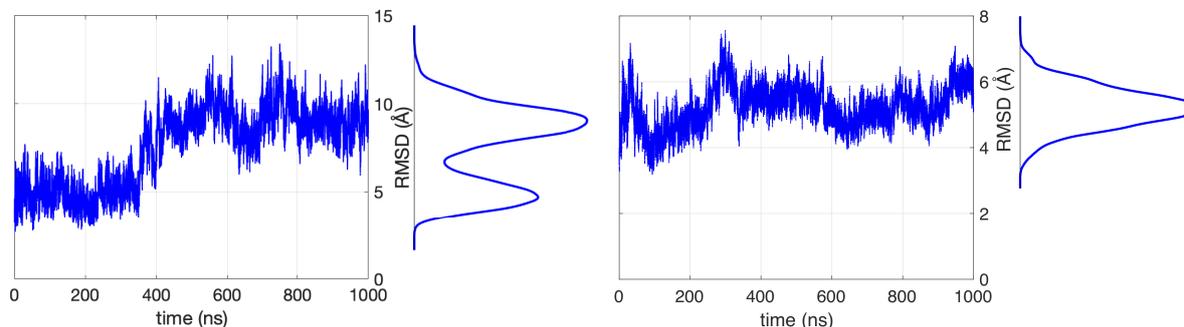


Figure B.14: Root-Mean-Square-Deviation (RMSD) time series and its marginal histogram of 3OFC (left) and 4P8Z (right).

PDB ID	RMSD(\AA)
2L1B	2.8
1Y26	2.2
1EHZ	2.9
3OFC	8.7
3PDR	3.1
4P8Z	5.7

Table B.4: RMSD between the average structure obtained of the last 50 ns and the experimental structure.

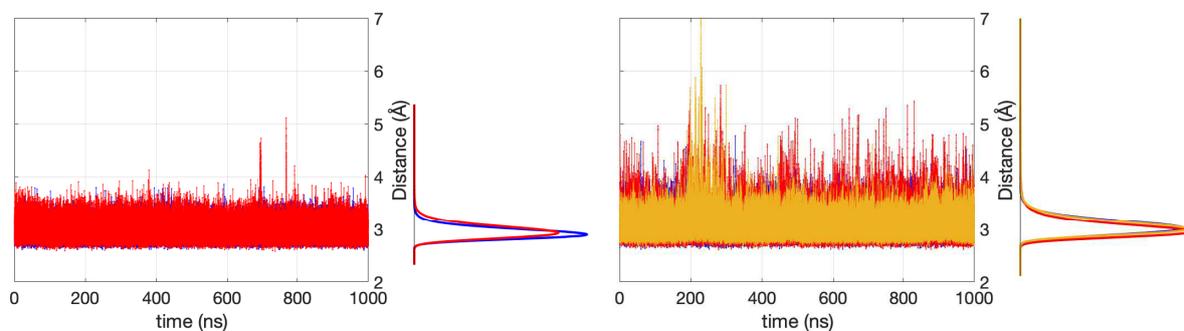


Figure B.15: Left: Time series of the distance A146:N7-U174:N3 (blue) and A41:N3-A146:N6 (red) to assess the stability of the triplet A41-A146-U174 in 4P8Z. Right: Time series of the distance G158:N2-A161:N7 (blue), G6:N2-A161:N1 (red) and G6:N1-A183:N1 (dark yellow) to assess the multiplet G6-G158-A161-A183 in 4P8Z.

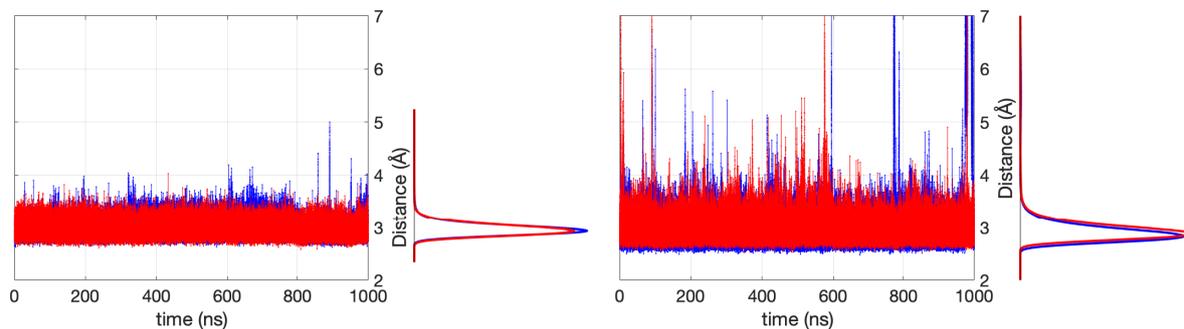


Figure B.16: Left: Time series of the distance C38:N3-G44:N1 (blue) and G44:N2-C47:N3 (red) to assess the stability of the triplet C38-G44-C47 in 3OFC. Right: Time series of the distance C31:O2-A53:N6 (blue) and C31:N4-G51:O6 (red) to assess the stability of the triplet C31-G51-A53 in 3OFC.

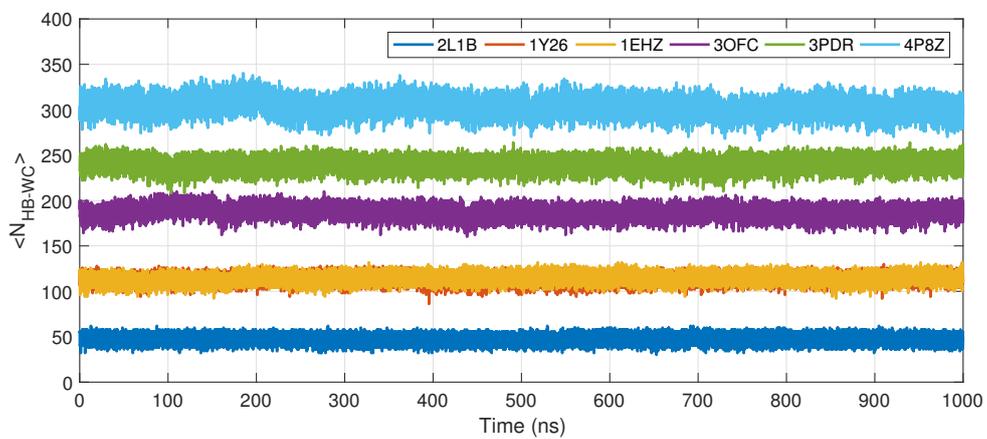


Figure B.17: Average number of WC-HB ($\langle N_{\text{WC-HB}} \rangle$) time series for the six RNA molecules under investigation.

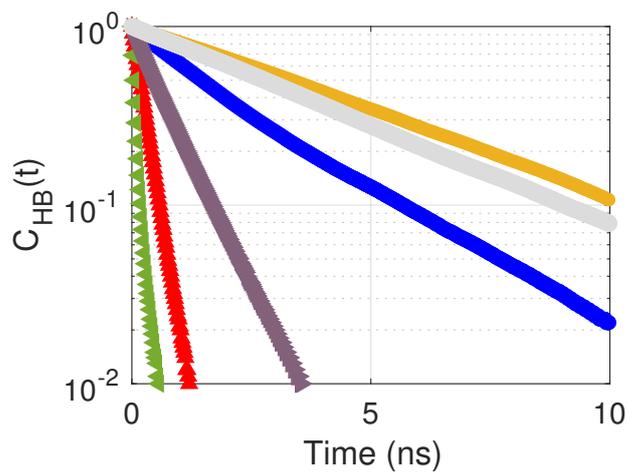


Figure B.18: Example of some hydrogen-bond correlation functions (C_{HB}) obtained for some specific WC hydrogen bonds in our MD simulations.

Appendix C. Definition of geometrical parameters

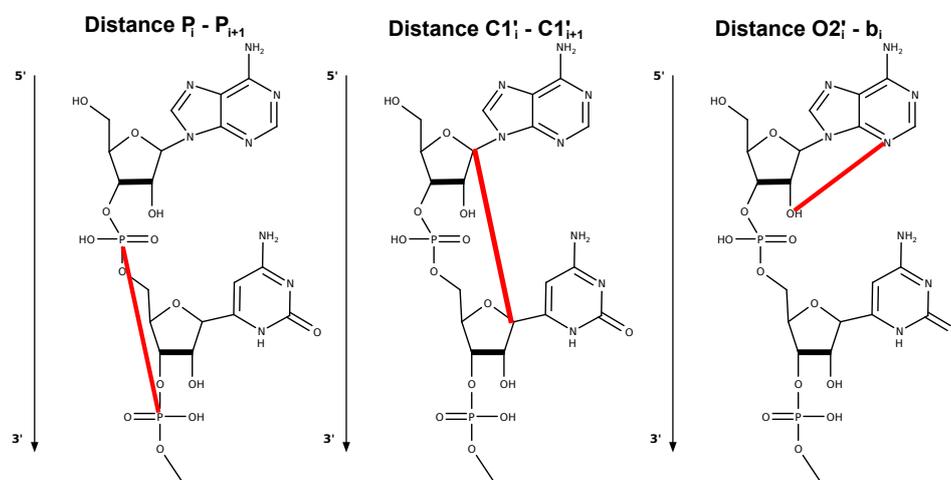


Figure C.19: Definition of the distances. Left: $P_i - P_{i+1}$. Centre: $C1'_i - C1'_{i+1}$. Right: $O2'_i - b_i$ with $b_i = N3$ for A,G and $b_i = O2$ for U,C. The index i represents the number of the nucleotide.

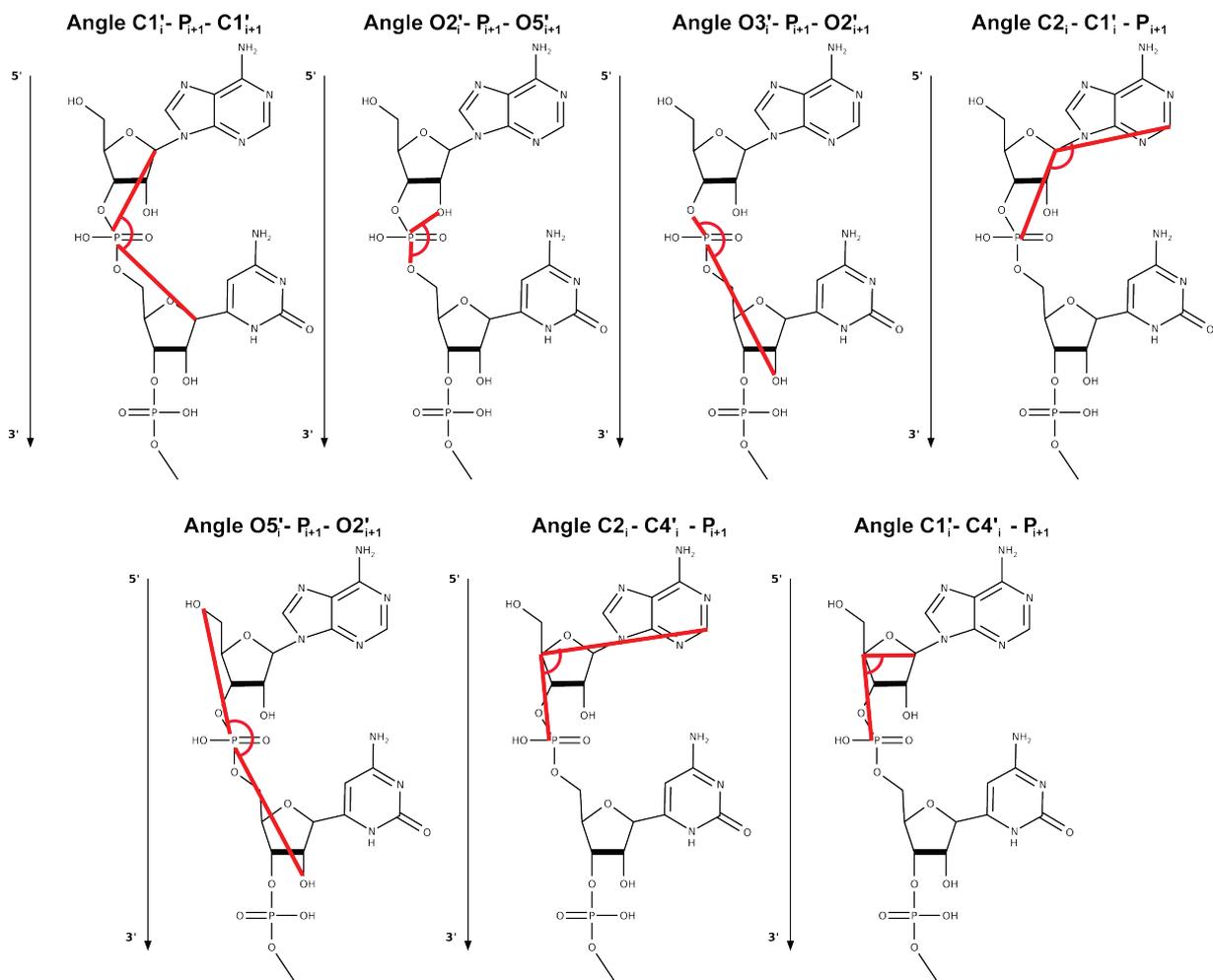


Figure C.20: Definition of the angles under investigation. From the left to the right: $C1'_i - C4'_i - P_{i+1}$, $C1'_i - P_{i+1} - C1'_{i+1}$, $C2_i - C1'_i - P_{i+1}$, $C2_i - C4'_i - P_{i+1}$, $O2'_i - P_{i+1} - O5'_{i+1}$, $O3'_i - P_{i+1} - O2'_{i+1}$, $O5'_i - P_{i+1} - O2'_{i+1}$, where i represents the number of the nucleotide.

Appendix D. Correlation coefficients

PDB ID	d_1	d_2	d_3	d_4
2L1B	0.66	0.62	0.65	0.31
1Y26	0.78	0.75	0.69	0.49
1EHZ	0.38	0.54	0.19*	0.20*
3OFC	0.21	0.26	0.33	0.16*
3PDR	0.82	0.85	0.65	0.48
4P8Z	0.32	0.33	0.17	-0.04*

Table D.5: Pearson correlation coefficient between 1M7 reactivity and the fluctuation of the distances $C1'_i-C1'_{i+1}$ (d_1), $C2_i-C2_{i+1}$ (d_2), $O2'_i-b_i$ with $b_i=N3$ for A,G and $b_i=O2$ for U,C (d_3), and P_i-P_{i+1} (d_4). *: p -value higher than 0.05.

PDB ID	d_1	d_2	d_3	d_4
2L1B	0.70	0.73	0.51	0.07*
1Y26	0.37	0.46	0.23	0.33
1EHZ	0.59	0.71	0.56	0.44
3OFC	0.30	0.47	0.25	0.09*
3PDR	0.12*	0.43	0.09*	-0.01*
4P8Z	0.23	0.36	0.29	-0.10*

Table D.6: Spearman correlation coefficient between 1M7 reactivity and the fluctuation of the distances $C1'_i-C1'_{i+1}$ (d_1), $C2_i-C2_{i+1}$ (d_2), $O2'_i-b_i$ with $b_i=N3$ for A,G and $b_i=O2$ for U,C (d_3) and P_i-P_{i+1} (d_4). *: p -value higher than 0.05.

PDB ID	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
2L1B	0.66	0.54	0.43	0.45	0.66	0.31*	0.27*	0.32*
1Y26	0.43	0.77	0.60	0.59	0.69	0.40	0.69	0.64
1EHZ	0.34	0.35	0.38	0.46	0.44	0.45	0.27	0.39
3OFC	0.44	0.29	0.24	0.40	0.25	0.26	0.39	0.31
3PDR	0.59	0.81	0.76	0.79	0.85	0.55	0.61	0.81
4P8Z	0.32	0.34	0.23	0.23	0.37	0.30	0.22	0.18

Table D.7: Pearson correlation coefficient between 1M7 reactivity and the fluctuation of the angles $C1'_i-C4'_{i+1}$ (a_1), $C1'_i-P_{i+1}-C1'_{i+1}$ (a_2), $C2_i-C1'_{i+1}-P_{i+1}$ (a_3), $C2_i-C4'_{i+1}-P_{i+1}$ (a_4), $C2_i-P_{i+1}-C2_{i+1}$ (a_5), $O2'_i-P_{i+1}-O5'_{i+1}$ (a_6), $O3'_i-P_{i+1}-O2'_{i+1}$ (a_7), $O5'_i-P_{i+1}-O2'_{i+1}$ (a_8). *: p -value higher than 0.05.

PDB ID	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
2L1B	0.51	0.48	0.42	0.55	0.44	0.50	0.33	0.48
1Y26	0.32	0.30	0.26	0.33	0.31	0.32	0.27	0.32
1EHZ	0.41	0.40	0.48	0.54	0.48	0.42	0.34	0.45
3OFC	0.35	0.24	0.22	0.32	0.29	0.24	0.24	0.18
3PDR	0.20	0.13	0.16	0.21	0.27	0.09*	0.09*	0.17
4P8Z	0.25	0.20	0.16	0.27	0.28	0.21	0.16	0.16

Table D.8: Spearman correlation coefficient between 1M7 reactivity and the fluctuation of the angles C1'_i-C4'_i-P_{i+1} (a₁), C1'_i-P_{i+1}-C1'_{i+1} (a₂), C2_i-C1'_{i+1}-P_{i+1} (a₃), C2_i-C4'_{i+1}-P_{i+1} (a₄), C2_i-P_{i+1}-C2_{i+1} (a₅), O2'_i-P_{i+1}-O5'_{i+1} (a₆), O3'_i-P_{i+1}-O2'_{i+1} (a₇), O5'_i-P_{i+1}-O2'_{i+1} (a₈). *: *p*-value higher than 0.05.

PDB ID	Shift	Slide	Rise	Tilt	Roll	Twist
2L1B	0.24*	0.44	0.21*	0.65	0.43	0.35
1Y26	0.45	0.49	0.58	0.55	0.47	0.54
1EHZ	0.28	0.34	0.21*	0.24	0.28	0.34
3OFC	0.29	0.38	0.29	0.29	0.27	0.31
3PDR	0.62	0.60	0.41	0.69	0.67	0.71
4P8Z	0.38	0.29	0.39	0.37	0.34	0.37

Table D.9: Pearson correlation coefficient between 1M7 reactivity and the fluctuation of the translation (Shift, Slide and Rise) and rotational helicoidal parameters (Tilt, Roll, Twist). *: *p*-value higher than 0.05.

PDB ID	Shift	Slide	Rise	Tilt	Roll	Twist
2L1B	0.57	0.59	0.39	0.52	0.70	0.61
1Y26	0.52	0.50	0.38	0.36	0.35	0.59
1EHZ	0.66	0.66	0.54	0.58	0.62	0.65
3OFC	0.35	0.40	0.38	0.41	0.36	0.42
3PDR	0.30	0.23	0.30	0.21	0.25	0.41
4P8Z	0.41	0.41	0.45	0.46	0.48	0.43

Table D.10: Spearman correlation coefficient between 1M7 reactivity and the fluctuation of the translation (Shift, Slide and Rise) and rotational helicoidal parameters (Tilt, Roll, Twist) and 1M7 reactivity. *: *p*-value higher than 0.05.

PDB ID	d_1	d_2	d_3	d_4
2L1B	0.57	0.49	0.65	0.32
1Y26	0.52	0.41	0.74	0.30
1EHZ	0.56	0.60	0.74	0.25
3OFC	0.73	0.78	0.68	0.26
3PDR	0.70	0.71	0.79	0.47
4P8Z	0.75	0.65	0.77	0.46

Table D.11: Spearman correlation coefficient between the fluctuation of pucker phase and the fluctuation of the distances $C1'_i-C1'_{i+1}$ (d_1), $C2_i-C2_{i+1}$ (d_2), $O2'_i-b_i$ with $b_i=N3$ for A,G and $b_i=O2$ for U,C (d_3) and P_i-P_{i+1} (d_4).*: p -value higher than 0.05.

PDB ID	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
2L1B	0.75	0.66	0.60	0.70	0.64	0.77	0.66	0.75
1Y26	0.58	0.61	0.66	0.66	0.49	0.64	0.54	0.63
1EHZ	0.80	0.73	0.72	0.76	0.65	0.75	0.57	0.75
3OFC	0.64	0.76	0.60	0.67	0.72	0.80	0.62	0.74
3PDR	0.74	0.75	0.74	0.71	0.73	0.75	0.72	0.77
4P8Z	0.71	0.82	0.78	0.76	0.64	0.77	0.64	0.74

Table D.12: Spearman correlation coefficient between the fluctuation of pucker phase and the fluctuation of the angles $C1'_i-C4'_i-P_{i+1}$ (a_1), $C1'_i-P_{i+1}-C1'_{i+1}$ (a_2), $C2_i-C1'_{i+1}-P_{i+1}$ (a_3), $C2_i-C4'_{i+1}-P_{i+1}$ (a_4), $C2_i-P_{i+1}-C2_{i+1}$ (a_5), $O2'_i-P_{i+1}-O5'_{i+1}$ (a_6), $O3'_i-P_{i+1}-O2'_{i+1}$ (a_7), $O5'_i-P_{i+1}-O2'_{i+1}$ (a_8). *: p -value higher than 0.05.

PDB ID	Shift	Slide	Rise	Tilt	Roll	Twist
2L1B	0.57	0.74	0.62	0.60	0.61	0.70
1Y26	0.56	0.67	0.53	0.62	0.58	0.54
1EHZ	0.69	0.76	0.70	0.73	0.76	0.71
3OFC	0.68	0.71	0.58	0.69	0.75	0.71
3PDR	0.69	0.74	0.63	0.69	0.72	0.69
4P8Z	0.66	0.76	0.67	0.74	0.75	0.68

Table D.13: Spearman correlation coefficient between the fluctuation of pucker phase and the fluctuation of the translation (Shift, Slide and Rise) and rotational helicoidal parameters (Tilt, Roll, Twist) and 1M7 reactivity. *: p -value higher than 0.05.

PDB ID	DMS	CMCT
1Y26	-0.63 ($1 \cdot 10^{-4}$)	-0.45 ($4 \cdot 10^{-3}$)
1EHZ	-0.74 ($2 \cdot 10^{-7}$)	-0.49 ($1 \cdot 10^{-3}$)
3OFC	-0.58 ($1 \cdot 10^{-6}$)	-0.35 ($6 \cdot 10^{-3}$)
4P8Z	-0.51 ($2 \cdot 10^{-6}$)	-0.44 ($2 \cdot 10^{-4}$)

Table D.14: Spearman correlation coefficient between the average number of HB formed at WC edge, N_{WC-HB} , and DMS and CMCT reactivity, respectively. In the bracket, we reported the p -value.

PDB ID	$C1'_i-C1'_{i+1}$	$C1'_i-P_{i+1}C1'_{i+1}$	Shift	Slide	Rise	Tilt	Roll	Twist
1Y26	0.49	0.52	0.34	0.45	0.41	0.45	0.49	0.46
1EHZ	0.46	0.44	0.41	0.43	0.49	0.53	0.53	0.45
3OFC	0.27	0.28	0.36	0.35	0.45	0.38	0.36	0.52
4P8Z	0.77	0.85	0.69	0.75	0.72	0.73	0.61	0.78

Table D.15: Pearson correlation coefficient between CMCT reactivity and the fluctuations of the distance $C1'_i-C1'_{i+1}$, the angle $C1'_i-P_{i+1}C1'_{i+1}$, Shift, Slide, Rise, Tilt, Roll and Twist.

Appendix E. Ribose pucker analysis

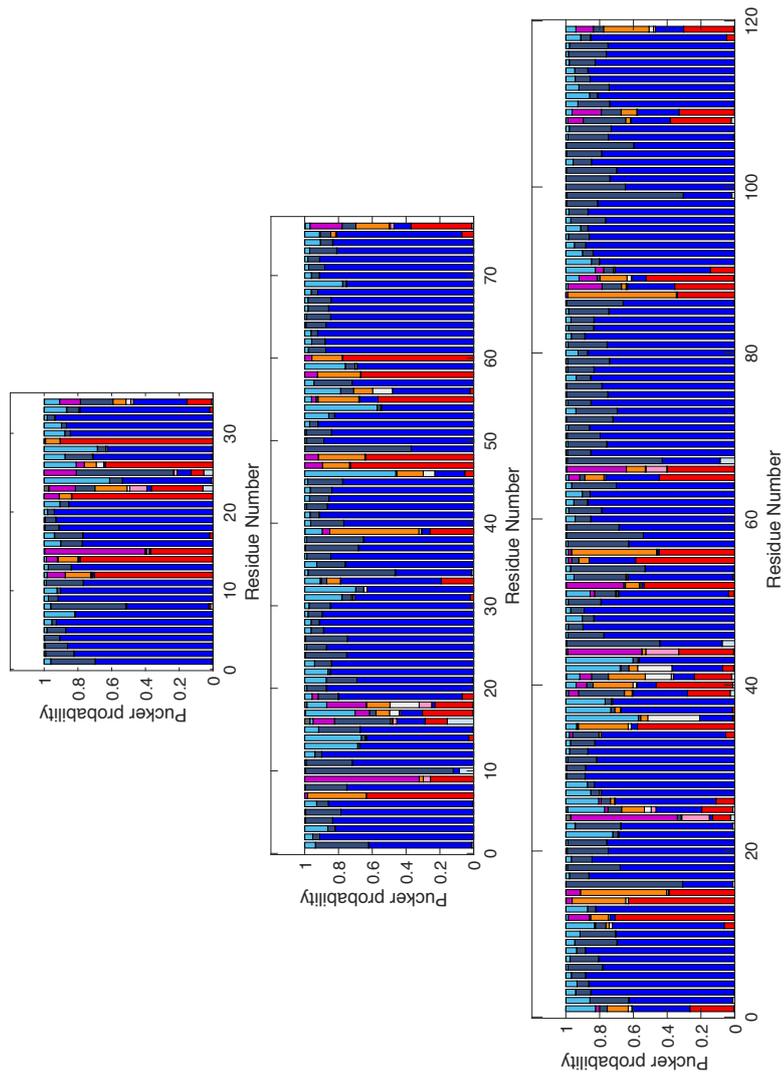


Figure E.21: Stacked bar plot of the pucker probability along MD simulation for 2LB1 (top), 1EHZ (middle) and 3OFC (bottom) .Red: C2'-endo, Pink: C4'-exo, Orange red: C1'-exo, Magenta: C3'-endo, light blue: C1'-endo, dark blue: C2'-exo, cyan: C4'-exo. Light grey: O4'-endo, dark grey: O4'-exo.

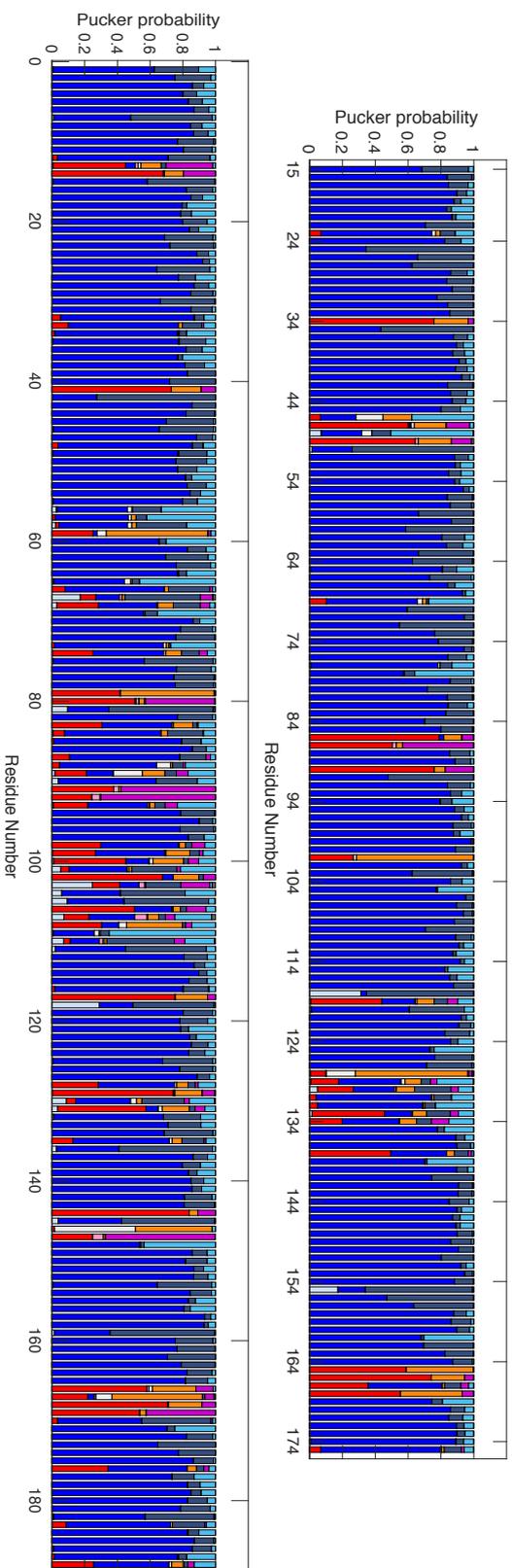


Figure E.22: Stacked bar plot of the pucker probability along MD simulation for 3PDR (top) and 4P8Z (bottom). Red: C2'-endo, Pink: C4'-endo, Orange red: C1'-exo, Magenta: C3'-exo. Blue: C3'-endo, light blue: C1'-endo, dark blue: C2'-exo, cyan: C4'-exo. Light grey: O4'-endo, dark grey: O4'-exo.

- [1] P. T. Li, J. Viereg, I. Tinoco, How rna unfolds and refolds, *Annual Review of Biochemistry* 77 (1) (2008) 77–100, pMID: 18518818. arXiv:<https://doi.org/10.1146/annurev.biochem.77.061206.174353>, doi:10.1146/annurev.biochem.77.061206.174353. URL <https://doi.org/10.1146/annurev.biochem.77.061206.174353>
- [2] K. V. Morris, J. S. Mattick, The rise of regulatory rna, *Nature Reviews Genetics* 15 (2014) 423 EP –. URL <http://dx.doi.org/10.1038/nrg3722>
- [3] Z. Miao, R. W. Adamiak, M. Antczak, R. T. Batey, A. J. Becka, M. Biesiada, M. J. Boniecki, J. M. Bujnicki, S.-J. Chen, C. Y. Cheng, et al., Rna-puzzles round iii: 3d rna structure prediction of five riboswitches and one ribozyme, *RNA* 23 (5) (2017) 655–672. doi:10.1261/rna.060368.116. URL <http://dx.doi.org/10.1261/rna.060368.116>
- [4] D. M. Lilley, How rna acts as a nuclease: some mechanistic comparisons in the nucleolytic ribozymes, *Biochemical Society Transactions* 45 (3) (2017) 683–691. arXiv:<http://www.biochemsoctrans.org/content/45/3/683.full.pdf>, doi:10.1042/BST20160158. URL <http://www.biochemsoctrans.org/content/45/3/683>
- [5] D. Incarnato, S. Oliviero, The rna epistructurome: Uncovering rna function by studying structure and post-transcriptional modifications, *Trends in Biotechnology* 35 (4) (2017) 318 – 333. doi:<https://doi.org/10.1016/j.tibtech.2016.11.002>. URL <http://www.sciencedirect.com/science/article/pii/S0167779916302037>
- [6] F. Allain, G. Varani, How accurately and precisely can rna structure be determined by nmr?, *Journal of molecular biology* 267 (2) (1997) 338–351. doi:10.1006/jmbi.1996.0855. URL <https://doi.org/10.1006/jmbi.1996.0855>
- [7] K. S. Keating, A. M. Pyle, Semiautomated model building for rna crystallography using a directed rotameric approach, *Proceedings of the National Academy of Sciences* 107 (18) (2010) 8177–8182. arXiv:<http://www.pnas.org/content/107/18/8177.full.pdf>, doi:10.1073/pnas.0911888107. URL <http://www.pnas.org/content/107/18/8177>
- [8] R. F. Garmann, A. Gopal, S. S. Athavale, C. M. Knobler, W. M. Gelbart, S. C. Harvey, Visualizing the global secondary structure of a viral rna genome with cryo-electron microscopy, *RNA* 21 (5) (2015) 877–886. doi:10.1261/rna.047506.114. URL <http://dx.doi.org/10.1261/rna.047506.114>
- [9] H. Berman, W. Olson, D. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. Hsieh, A. Srinivasan, B. Schneider, The nucleic acid database. a comprehensive relational database of three-dimensional structures of nucleic acids, *Biophysical Journal* 63 (3) (1992) 751–759. doi:10.1016/s0006-3495(92)81649-1. URL [http://dx.doi.org/10.1016/S0006-3495\(92\)81649-1](http://dx.doi.org/10.1016/S0006-3495(92)81649-1)
- [10] B. Coimbatore Narayanan, J. Westbrook, S. Ghosh, A. I. Petrov, B. Sweeney, C. L. Zirbel, N. B. Leontis, H. M. Berman, The nucleic acid database: new features and capabilities, *Nucleic Acids Research* 42 (D1) (2014) D114–D122. doi:10.1093/nar/gkt980.
- [11] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter, M. Otyepka, Rna structural dynamics as captured by molecular simulations: A comprehensive overview, *Chemical Reviews* 118 (8) (2018) 4177–4338, pMID: 29297679. arXiv:<https://doi.org/10.1021/acs.chemrev.7b00427>, doi:10.1021/acs.chemrev.7b00427. URL <https://doi.org/10.1021/acs.chemrev.7b00427>
- [12] I. Tinoco, C. Bustamante, How rna folds, *Journal of Molecular Biology* 293 (2) (1999) 271 – 281. doi:<https://doi.org/10.1006/jmbi.1999.3001>. URL <http://www.sciencedirect.com/science/article/pii/S0022283699930012>
- [13] S. R. Holbrook, Rna structure: the long and the short of it, *Current Opinion in Structural Biology* 15 (3) (2005) 302 – 308, sequences and topology/Nucleic acids. doi:<https://doi.org/10.1016/j.sbi.2005.04.005>. URL <http://www.sciencedirect.com/science/article/pii/S0959440X05000849>
- [14] N. J. Reiter, C. W. Chan, A. Mondragón, Emerging structural themes in large rna molecules, *Current Opinion in Structural Biology* 21 (3) (2011) 319 – 326. doi:<https://doi.org/10.1016/j.sbi.2011.03.003>. URL <http://www.sciencedirect.com/science/article/pii/S0959440X1100039X>
- [15] P. Brion, E. Westhof, Hierarchy and dynamics of rna folding, *Annual Review*

- of Biophysics and Biomolecular Structure 26 (1) (1997) 113–137, pMID: 9241415. arXiv:<https://doi.org/10.1146/annurev.biophys.26.1.113>, doi:10.1146/annurev.biophys.26.1.113. URL <https://doi.org/10.1146/annurev.biophys.26.1.113>
- [16] K. E. Deigan, T. W. Li, D. H. Mathews, K. M. Weeks, Accurate shape-directed rna structure determination, *Proceedings of the National Academy of Sciences* 106 (1) (2008) 97–102. doi:10.1073/pnas.0806929106. URL <http://dx.doi.org/10.1073/pnas.0806929106>
- [17] R. Lorenz, M. T. Wolfinger, A. Tanzer, I. L. Hofacker, Predicting rna secondary structures from sequence and probing data, *Methods* 103 (2016) 86 – 98, *advances in RNA Structure Determination*. doi:<https://doi.org/10.1016/j.ymeth.2016.04.004>. URL <http://www.sciencedirect.com/science/article/pii/S1046202316300743>
- [18] A. Krokhotin, A. M. Mustoe, K. M. Weeks, N. V. Dokholyan, Direct identification of base-paired rna nucleotides by correlated chemical probing, *RNA* 23 (1) (2017) 6–13. arXiv:<http://rnajournal.cshlp.org/content/23/1/6.full.pdf+html>, doi:10.1261/rna.058586.116. URL <http://rnajournal.cshlp.org/content/23/1/6.abstract>
- [19] K. M. Kutchko, A. Laederach, Transcending the prediction paradigm: novel applications of shape to rna function and evolution, *Wiley Interdisciplinary Reviews: RNA* 8 (1) (2016) e1374. doi:10.1002/wrna.1374. URL <http://dx.doi.org/10.1002/wrna.1374>
- [20] E. Bindewald, M. Wendeler, M. Legiewicz, M. K. Bona, Y. Wang, M. J. Pritt, S. F. Le Grice, B. A. Shapiro, Correlating shape signatures with three-dimensional rna structures, *RNA* arXiv:<http://rnajournal.cshlp.org/content/early/2011/07/13/rna.2640111.full.pdf+html>, doi:10.1261/rna.2640111. URL <http://rnajournal.cshlp.org/content/early/2011/07/13/rna.2640111.abstract>
- [21] S. Kirmizialtin, S. P. Hennelly, A. Schug, J. N. Onuchic, K. Y. Sanbonmatsu, Integrating molecular dynamics simulations with chemical probing experiments using shape-fit, *Computational Methods for Understanding Riboswitches* (2015) 215–234doi:10.1016/bs.mie.2014.10.061. URL <http://dx.doi.org/10.1016/bs.mie.2014.10.061>
- [22] K. M. Weeks, Advances in rna structure analysis by chemical probing, *Current Opinion in Structural Biology* 20 (3) (2010) 295–304. doi:10.1016/j.sbi.2010.04.001. URL <http://dx.doi.org/10.1016/j.sbi.2010.04.001>
- [23] C. Ehresmann, F. Baudin, M. Mougel, P. Romby, J.-P. Ebel, B. Ehresmann, Probing the structure of rnas in solution, *Nucleic Acids Research* 15 (22) (1987) 9109–9128. doi:10.1093/nar/15.22.9109. URL <http://dx.doi.org/10.1093/nar/15.22.9109>
- [24] R. Lavery, A. Pullman, A new theoretical index of biochemical reactivity combining steric and electrostatic factors: An application to yeast trn^{phe}, *Biophysical Chemistry* 19 (2) (1984) 171 – 181. doi:[https://doi.org/10.1016/0301-4622\(84\)85017-6](https://doi.org/10.1016/0301-4622(84)85017-6). URL <http://www.sciencedirect.com/science/article/pii/0301462284850176>
- [25] E. J. Merino, K. A. Wilkinson, J. L. Coughlan, K. M. Weeks, Rna structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (shape), *Journal of the American Chemical Society* 127 (12) (2005) 4223–4231. arXiv:<https://doi.org/10.1021/ja043822v>, doi:10.1021/ja043822v. URL <https://doi.org/10.1021/ja043822v>
- [26] K. A. Wilkinson, E. J. Merino, K. M. Weeks, Rna shape chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in trn^{asp} transcripts, *Journal of the American Chemical Society* 127 (13) (2005) 4659–4667, pMID: 15796531. arXiv:<https://doi.org/10.1021/ja0436749>, doi:10.1021/ja0436749. URL <https://doi.org/10.1021/ja0436749>
- [27] K. A. Wilkinson, E. J. Merino, K. M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension (shape): quantitative rna structure analysis at single nucleotide resolution, *Nature Protocols* 1 (3) (2006) 1610–1616. doi:10.1038/nprot.2006.249.

- URL <http://dx.doi.org/10.1038/nprot.2006.249>
- [28] M. J. Smola, G. M. Rice, S. Busan, N. A. Siegfried, K. M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (shape-map) for direct, versatile and accurate rna structure analysis, *Nature Protocols* 10 (11) (2015) 1643–1669. doi:10.1038/nprot.2015.103. URL <http://dx.doi.org/10.1038/nprot.2015.103>
- [29] S. Tian, P. Cordero, W. Kladwang, R. Das, High-throughput mutate-map-rescue evaluates shape-directed rna structure and uncovers excited states, *RNA* 20 (11) (2014) 1815–1826. doi:10.1261/rna.044321.114. URL <http://dx.doi.org/10.1261/rna.044321.114>
- [30] J. T. Low, K. M. Weeks, Shape-directed rna secondary structure prediction, *Methods* 52 (2) (2010) 150–158, rNA: From Sequence to Structure and Dynamics. doi:<https://doi.org/10.1016/j.ymeth.2010.06.007>. URL <http://www.sciencedirect.com/science/article/pii/S1046202310001611>
- [31] J. L. McGinnis, K. M. Weeks, Ribosome rna assembly intermediates visualized in living cells, *Biochemistry* 53 (19) (2014) 3237–3247. doi:10.1021/bi500198b. URL <http://dx.doi.org/10.1021/bi500198b>
- [32] S. A. Mortimer, K. M. Weeks, Time-resolved rna shape chemistry: quantitative rna structure analysis in one-second snapshots and at single-nucleotide resolution, *Nature Protocols* 4 (10) (2009) 1413–1421. doi:10.1038/nprot.2009.126. URL <http://dx.doi.org/10.1038/nprot.2009.126>
- [33] S. A. Mortimer, K. M. Weeks, A fast-acting reagent for accurate analysis of rna secondary and tertiary structure by shape chemistry, *Journal of the American Chemical Society* 129 (14) (2007) 4144–4145, PMID: 17367143. arXiv:<https://doi.org/10.1021/ja0704028>, doi:10.1021/ja0704028. URL <https://doi.org/10.1021/ja0704028>
- [34] K.-A. Steen, G. M. Rice, K. M. Weeks, Fingerprinting noncanonical and tertiary rna structures by differential shape reactivity, *Journal of the American Chemical Society* 134 (32) (2012) 13160–13163. doi:10.1021/ja304027m. URL <http://dx.doi.org/10.1021/ja304027m>
- [35] R. C. Spitale, P. Crisalli, R. A. Flynn, E. A. Torre, E. T. Kool, H. Y. Chang, Rna shape analysis in living cells, *Nature Chemical Biology* 9 (2012) 18 EP –. URL <https://doi.org/10.1038/nchembio.1131>
- [36] C. M. Gherghe, S. A. Mortimer, J. M. Krahn, N. L. Thompson, K. M. Weeks, Slow conformational dynamics at c2'-endo nucleotides in rna, *Journal of the American Chemical Society* 130 (28) (2008) 8884–8885, PMID: 18558680. arXiv:<https://doi.org/10.1021/ja802691e>, doi:10.1021/ja802691e. URL <https://doi.org/10.1021/ja802691e>
- [37] D. Loughrey, K. E. Watters, A. H. Settle, J. B. Lucks, Shape-seq 2.0: systematic optimization and extension of high-throughput chemical probing of rna secondary structure with next generation sequencing, *Nucleic Acids Research* 42 (21) (2014) e165–e165. doi:10.1093/nar/gku909. URL <http://dx.doi.org/10.1093/nar/gku909>
- [38] S. M. Vasa, N. Guex, K. A. Wilkinson, K. M. Weeks, M. C. Giddings, Shapefinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis, *RNA* 14 (10) (2008) 1979–1990. doi:10.1261/rna.1166808. URL <http://dx.doi.org/10.1261/rna.1166808>
- [39] K. A. Wilkinson, R. J. Gorelick, S. M. Vasa, N. Guex, A. Rein, D. H. Mathews, M. C. Giddings, K. M. Weeks, High-throughput shape analysis reveals structures in hiv-1 genomic rna strongly conserved across distinct biological states, *PLoS Biology* 6 (4) (2008) e96. doi:10.1371/journal.pbio.0060096. URL <http://dx.doi.org/10.1371/journal.pbio.0060096>
- [40] M. Zubradt, P. Gupta, S. Persad, A. M. Lambowitz, J. S. Weissman, S. Rouskin, Dms-mapseq for genome-wide or targeted rna structure probing in vivo, *Nature Methods* 14 (1) (2016) 75–82. doi:10.1038/nmeth.4057. URL <http://dx.doi.org/10.1038/nmeth.4057>
- [41] S. Rouskin, M. Zubradt, S. Washietl, M. Kellis, J. S. Weissman, Genome-wide probing of rna structure

- reveals active unfolding of mrna structures in vivo, *Nature* 505 (2013) 701 EP –.
 URL <https://doi.org/10.1038/nature12894>
- [42] K. E. Watters, A. M. Yu, E. J. Strobel, A. H. Settle, J. B. Lucks, Characterizing rna structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (shape-seq), *Methods* 103 (2016) 34 – 48, *advances in RNA Structure Determination*. doi:<https://doi.org/10.1016/j.ymeth.2016.04.002>.
 URL <http://www.sciencedirect.com/science/article/pii/S104620231630072X>
- [43] M. J. Smola, J. M. Calabrese, K. M. Weeks, Detection of rna-protein interactions in living cells with shape, *Biochemistry* 54 (46) (2015) 6867–6875. doi:10.1021/acs.biochem.5b00977.
 URL <http://dx.doi.org/10.1021/acs.biochem.5b00977>
- [44] M. Smola, T. Christy, K. Inoue, C. Nicholson, M. Friedersdorf, J. Keene, D. Lee, J. Calabrese, K. Weeks, Shape reveals transcript-wide interactions, complex structural domains, and protein interactions across the xist lncrna in living cells, *PNAS; Proceedings of the National Academy of Sciences* 113 (37) (2016) 10322–10327.
- [45] P. Cordero, J. B. Lucks, R. Das, An rna mapping database for curating rna structure mapping experiments, *Bioinformatics* 28 (22) (2012) 3006–3008. doi:10.1093/bioinformatics/bts554.
 URL <http://dx.doi.org/10.1093/bioinformatics/bts554>
- [46] J. L. McGinnis, J. A. Dunkle, J. H. D. Cate, K. M. Weeks, The mechanisms of rna shape chemistry, *Journal of the American Chemical Society* 134 (15) (2012) 6617–6624. doi:10.1021/ja2104075.
 URL <http://dx.doi.org/10.1021/ja2104075>
- [47] G. Pinamonti, S. Bottaro, C. Micheletti, G. Bussi, Elastic network models for rna: a comparative assessment with molecular dynamics and shape experiments, *Nucleic Acids Research* 43 (15) (2015) 7260–7269. doi:10.1093/nar/gkv708.
 URL <http://dx.doi.org/10.1093/nar/gkv708>
- [48] V. Mlýnský, G. Bussi, Molecular dynamics simulations reveal an interplay between shape reagent binding and rna flexibility, *The Journal of Physical Chemistry Letters* 9 (2) (2018) 313–318. doi:10.1021/acs.jpcclett.7b02921.
- [49] K. M. Weeks, D. M. Mauer, Exploring rna structural codes with shape chemistry, *Accounts of Chemical Research* 44 (12) (2011) 1280–1291. doi:10.1021/ar200051h.
 URL <http://dx.doi.org/10.1021/ar200051h>
- [50] T. Hurst, X. Xu, P. Zhao, S.-J. Chen, Quantitative understanding of shape mechanism from rna structure and dynamics analysis, *The Journal of Physical Chemistry B* 122 (18) (2018) 4771–4783. doi:10.1021/acs.jpcc.8b00575.
 URL <http://dx.doi.org/10.1021/acs.jpcc.8b00575>
- [51] M. Meyer, H. Nielsen, V. Oliéric, P. Roblin, S. D. Johansen, E. Westhof, B. Masquida, Specification of a group i intron into a lariat capping ribozyme, *Proceedings of the National Academy of Sciences* 111 (21) (2014) 7659–7664. arXiv:<http://www.pnas.org/content/111/21/7659.full.pdf>, doi:10.1073/pnas.1322248111.
 URL <http://www.pnas.org/content/111/21/7659>
- [52] J. Deforges, S. de Breyne, M. Ameer, N. Ulryck, N. Chamond, A. Saaidi, Y. Ponty, T. Ohlmann, B. Sargueil, Two ribosome recruitment sites direct multiple translation events within hiv1 gag open reading frame, *Nucleic Acids Research* 45 (12) (2017) 7382–7400. doi:10.1093/nar/gkx303.
- [53] H. J. Berendsen, D. van der Spoel, R. van Drunen, Gromacs: a message-passing parallel molecular dynamics implementation, *Computer Physics Communications* 91 (1) (1995) 43–56.
- [54] E. Lindahl, B. Hess, D. Van Der Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, *Molecular modeling annual* 7 (8) (2001) 306–317.
- [55] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, Gromacs: fast, flexible, and free, *Journal of computational chemistry* 26 (16) (2005) 1701–1718.
- [56] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, *Journal of chemical theory and computation* 4 (3) (2008) 435–447.

- [57] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit., *Bioinformatics* 29 (7) (2013) 845–54. doi:10.1093/bioinformatics/btt055.
- [58] A. Pérez, I. Marchán, D. Svozil, J. Sponer, I. Cheatham, Thomas E., C. A. Loughton, M. Orozco, Refinement of the amber force field for nucleic acids: Improving the description of α/γ conformers, *Biophysical Journal* 92 (11) (2007) 3817–3829. doi:10.1529/biophysj.106.097782. URL <https://doi.org/10.1529/biophysj.106.097782>
- [59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian16 Revision B.01, gaussian Inc. Wallingford CT (2016).
- [60] J. Wang, P. Cieplak, P. A. Kollman, How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules?, *Journal of computational chemistry* 21 (12) (2000) 1049–1074.
- [61] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, Amber 10, Tech. rep., University of California (2008).
- [62] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field, *Journal of Computational Chemistry* 25 (9) (2004) 1157–1174. doi:10.1002/jcc.20035. URL <http://dx.doi.org/10.1002/jcc.20035>
- [63] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, T. Head-Gordon, Development of an improved four-site water model for biomolecular simulations: Tip4p-ew, *The Journal of Chemical Physics* 120 (20) (2004) 9665–9678. arXiv:<https://doi.org/10.1063/1.1683075>, doi:10.1063/1.1683075. URL <https://doi.org/10.1063/1.1683075>
- [64] I. S. Joung, T. E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *The Journal of Physical Chemistry B* 112 (30) (2008) 9020–9041. doi:10.1021/jp8001614. URL <http://dx.doi.org/10.1021/jp8001614>
- [65] I. Bešková, M. Otyepka, K. Réblová, J. Šponer, Dependence of a-rna simulations on the choice of the force field and salt strength, *Physical Chemistry Chemical Physics* 11 (45) (2009) 10701. doi:10.1039/b911169g. URL <http://dx.doi.org/10.1039/b911169g>
- [66] O. Allnér, L. Nilsson, A. Villa, Magnesium ion–water coordination and exchange in biomolecular simulations, *Journal of Chemical Theory and Computation* 8 (4) (2012) 1493–1502.
- [67] R. A. Cunha, G. Bussi, Unraveling mg2+–rna binding with atomistic molecular dynamics, *RNA* 23 (5) (2017) 628–638. doi:10.1261/rna.060079.116. URL <http://dx.doi.org/10.1261/rna.060079.116>
- [68] T. Darden, D. York, L. Pedersen, Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems, *The Journal of chemical physics* 98 (12) (1993) 10089–10092.
- [69] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh ewald method, *The Journal of chemical physics* 103 (19) (1995) 8577–8593.
- [70] B. Hess, H. Bekker, H. J. Berendsen, J. G. Fraaije, Lincs: a linear constraint solver for molecular simulations, *Journal of computational chemistry* 18 (12) (1997) 1463–1472.

- [71] H. J. Berendsen, J. P. M. v. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, *The Journal of chemical physics* 81 (8) (1984) 3684–3690.
- [72] S. C. Harvey, R. K.-Z. . Z. Tan, T. E. Cheatham, The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition, *Journal of Computational Chemistry* 19 (7) (1998) 726–740.
- [73] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *The Journal of chemical physics* 126 (1) (2007) 014101.
- [74] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied physics* 52 (12) (1981) 7182–7190.
- [75] N. I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, 1996.
URL <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0521568900>
- [76] C. J. Marzec, L. A. Day, An exact description of five-membered ring configurations. i. parameterization via an amplitude s , an angle γ , the pseudorotation amplitude q and phase angle p , and the bond lengths, *Journal of Biomolecular Structure and Dynamics* 10 (6) (1993) 1091–1123, pMID: 8357542. doi:10.1080/07391102.1993.10508697.
- [77] E. Westhof, M. Sundaralingam, A method for the analysis of puckering disorder in five-membered rings: the relative mobilities of furanose and proline rings and their effects on polynucleotide and polypeptide backbone flexibility, *Journal of the American Chemical Society* 105 (4) (1983) 970–976. arXiv:<https://doi.org/10.1021/ja00342a054>, doi:10.1021/ja00342a054.
URL <https://doi.org/10.1021/ja00342a054>
- [78] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, Conformational analysis of nucleic acids revisited: Curves+, *Nucleic Acids Research* 37 (17) (2009) 5917–5929. doi:10.1093/nar/gkp608.
URL <http://dx.doi.org/10.1093/nar/gkp608>
- [79] N. B. Leontis, E. Westhof, Geometric nomenclature and classification of rna base pairs., *RNA* 7 (4) (2001) 499–512. arXiv:<http://rnajournal.cshlp.org/content/7/4/499.full.pdf+html>.
URL <http://rnajournal.cshlp.org/content/7/4/499.abstract>
- [80] D. Bhattacharyya, S. C. Koripella, A. Mitra, V. B. Rajendran, B. Sinha, Theoretical analysis of noncanonical base pairing interactions in rna molecules, *Journal of Biosciences* 32 (1) (2007) 809–825. doi:10.1007/s12038-007-0082-4.
URL <https://doi.org/10.1007/s12038-007-0082-4>
- [81] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odellius, H. Ogasawara, L. Näslund, T. Hirsch, L. Ojamäe, P. Glatzel, L. Pettersson, A. Nilsson, The structure of the first coordination shell in liquid water, *Science* 304 (5673) (2004) 995–999.
- [82] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, Mdtraj: A modern open library for the analysis of molecular dynamics trajectories, *Biophysical Journal* 109 (8) (2015) 1528–1532. doi:10.1016/j.bpj.2015.08.015.
URL <http://dx.doi.org/10.1016/j.bpj.2015.08.015>
- [83] H. Cramér, *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1946.
- [84] C. E. Hajdin, S. Bellaousov, W. Huggins, C. W. Leonard, D. H. Mathews, K. M. Weeks, Accurate shape-directed rna secondary structure modeling, including pseudoknots, *Proceedings of the National Academy of Sciences* 110 (14) (2013) 5498–5503. arXiv:<http://www.pnas.org/content/110/14/5498.full.pdf>, doi:10.1073/pnas.1219988110.
URL <http://www.pnas.org/content/110/14/5498>
- [85] G. M. Rice, C. W. Leonard, K. M. Weeks, Rna secondary structure modeling at consistent high accuracy using differential shape, *RNA* 20 (6) (2014) 846–854. doi:10.1261/rna.043323.113.
URL <http://dx.doi.org/10.1261/rna.043323.113>
- [86] P. Cordero, W. Kladwang, C. C. VanLang, R. Das, Quantitative dimethyl sulfate mapping for automated rna secondary structure inference, *Biochemistry* 51 (36) (2012) 7037–7039, pMID: 22913637. arXiv:<https://doi.org/10.1021/bi3008802>, doi:10.1021/bi3008802.

- URL <https://doi.org/10.1021/bi3008802>
- [87] W. Kladwang, A. Becka, T. Mann, Wipapat kladwang, alex becka, tom mann;pmid:none. tr-naph_dms.0005.
- [88] W. Kladwang, A. Becka, T. Mann, Wipapat kladwang, alex becka, tom mann;pmid:none. tr-naph_cmc.0005.
- [89] D. Allouche, Développement d'une stratégie innovante de modélisation de la structure des ARN Etude des mécanismes moléculaires des IRES de type III, Ph.D. thesis, Ecole doctorale Médicament, Toxicologie, Chimie Imagerie, Université Paris Descartes (2018).
- [90] L. Weill, L. James, N. Ulryck, N. Chamond, C. H. Herbreteau, T. Ohlmann, B. Sargueil, A new type of ires within gag coding region recruits three initiation complexes on hiv-2 genomic rna, *Nucleic Acids Research* 38 (4) (2009) 1367–1381. doi:10.1093/nar/gkp1109.
- [91] L. James, B. Sargueil, Rna secondary structure of the feline immunodeficiency virus 5utr and gag coding region, *Nucleic Acids Research* 36 (14) (2008) 4653–4666. doi:10.1093/nar/gkn447.
- [92] R. Lorenz, D. Luntzer, I. L. Hofacker, P. F. Stadler, M. T. Wolfinger, Shape directed rna folding, *Bioinformatics* 32 (1) (2016) 145–147. doi:10.1093/bioinformatics/btv523.
- [93] S. Washietl, I. L. Hofacker, P. F. Stadler, M. Kellis, Rna folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction, *Nucleic Acids Research* 40 (10) (2012) 4261–4272. doi:10.1093/nar/gks009.
- [94] K. Zarrinhalam, M. M. Meyer, I. Dotu, J. H. Chuang, P. Clote, Integrating chemical footprinting data into rna secondary structure prediction, *PLOS ONE* 7 (10) (2012) 1–13. doi:10.1371/journal.pone.0045160.
URL <https://doi.org/10.1371/journal.pone.0045160>
- [95] Q. Vicens, A. R. Gooding, A. Laederach, T. R. Cech, Local rna structural changes induced by crystallization are revealed by shape, *RNA* 13 (4) (2007) 536–548. doi:10.1261/rna.400207.
URL <http://dx.doi.org/10.1261/rna.400207>
- [96] F. Karabiber, J. L. McGinnis, O. V. Favorov, K. M. Weeks, Qushape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis, *RNA* 19 (1) (2013) 63–73. arXiv:<http://rnajournal.cshlp.org/content/19/1/63.full.pdf+html>, doi:10.1261/rna.036327.112.
URL <http://rnajournal.cshlp.org/content/19/1/63.abstract>
- [97] J. A. Cruz, M.-F. Blanchet, M. Boniecki, J. M. Bujnicki, S.-J. Chen, S. Cao, R. Das, F. Ding, N. V. Dokholyan, S. C. Flores, et al., Rna-puzzles: A casp-like evaluation of rna three-dimensional structure prediction, *RNA* 18 (4) (2012) 610–625. doi:10.1261/rna.031054.111.
URL <http://dx.doi.org/10.1261/rna.031054.111>