Article

Progress toward SHAPE Constrained Computational Prediction of Tertiary Interactions in RNA Structure

Grégoire De Bisschop [1,2], Delphine Allouche [1,3], Elisa Frezza [1], Benoît Masquida [4], Yann Ponty [5], Sebastian Will [5] and Bruno Sargueil [1,*]

[1]   Université de Paris, CNRS, UMR 8038/CiTCoM, F-75006 Paris, France; gregoire.de.bisschop@ircm.qc.ca (G.D.B.); delphine.allouche@inserm.fr (D.A.); elisa.frezza@u-paris.fr (E.F.)
[2]   Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC H2W 1R7, Canada
[3]   Institut Necker-Enfants Malades (INEM), Inserm U1151, 156 rue de Vaugirard, CEDEX 15, 75015 Paris, France
[4]   Université de Strasbourg, CNRS UMR7156 GMGM, 67084 Strasbourg, France; b.masquida@unistra.fr
[5]   Ecole Polytechnique, CNRS UMR 7161, LIX, 91120 Palaiseau, France; yann.ponty@lix.polytechnique.fr (Y.P.); sebastian.will@lix.polytechnique.fr (S.W.)
*   Correspondence: bruno.sargueil@u-paris.fr

Abstract: As more sequencing data accumulate and novel puzzling genetic regulations are discovered, the need for accurate automated modeling of RNA structure increases. RNA structure modeling from chemical probing experiments has made tremendous progress, however accurately predicting large RNA structures is still challenging for several reasons: RNA are inherently flexible and often adopt many energetically similar structures, which are not reliably distinguished by the available, incomplete thermodynamic model. Moreover, computationally, the problem is aggravated by the relevance of pseudoknots and non-canonical base pairs, which are hardly predicted efficiently. To identify nucleotides involved in pseudoknots and non-canonical interactions, we scrutinized the SHAPE reactivity of each nucleotide of the 188 nt long lariat-capping ribozyme under multiple conditions. Reactivities analyzed in the light of the X-ray structure were shown to report accurately the nucleotide status. Those that seemed paradoxical were rationalized by the nucleotide behavior along molecular dynamic simulations. We show that valuable information on intricate interactions can be deduced from probing with different reagents, and in the presence or absence of Mg^{2+}. Furthermore, probing at increasing temperature was remarkably efficient at pointing to non-canonical interactions and pseudoknot pairings. The possibilities of following such strategies to inform structure modeling software are discussed.

Keywords: ribozyme; shape probing; RNA structure modeling

1. Introduction

RNA molecules are ubiquitous within the cell but are also present outside of the cell as in plant phloem [1]. Some of them have even been shown to be glycosylated and exposed on the cell surface [2]. They fulfil very diverse functions including coding and non-coding roles, are involved in most steps of the genetic expression regulation, and can specifically interact with small molecules, proteins, or other nucleic acids (DNA or RNA). Their sequence is indisputably critical for the coding role or the specific recognition of other DNA and RNA sequences. Beyond the importance of their sequence, they also adopt specific three-dimensional structures that are responsible among others for the catalytic activity of ribozymes (RNA enzymes), the strong and specific binding to metabolites, proteins and ribosome, the assembly of phase-separated membrane-less organelles [3,4], the availability of sequence signals (e.g., [5]) and the modulation of ribosomes progression. Structure can be considered as another layer of encoded information, not only in compactly folded RNA, such as rRNA, tRNA, ribozymes, riboswitches but also within mRNA [6]. This is particularly obvious within viruses' genomes in which information is very streamlined [7,8].

Given the central role of RNA and its structure, it is of great interest to be able to model RNA structure as accurately as possible. Physical methods that were very successful for determining protein structures suffer from limitations for RNA, for instance NMR can only be applied to short RNA and flexible RNA molecules are most of the time hardly amenable to crystallization for X-ray diffraction. In addition, those methods are very time consuming, while new RNA sequences are uncovered every day. The RNA structure modeling process most of the time firstly consists of predicting the secondary structure before inferring the three-dimensional architecture. However, identifying the correct secondary structure is complicated since the native structure is the result of indirect folding trajectories forming secondary and tertiary interactions. This is well exemplified by the formation of the P4P6 domain from the *Tetrahymena thermophila* group I intron. The secondary structure of the folding intermediate presents a P5abc domain with additional Watson–Crick base pairs as compared to the native form. The formation of tertiary interactions between P5abc and the P4 domain leads to the release of the additional base pairs that were yet required to direct the formation of the tertiary structure [9].

The accuracy of computational secondary structure modeling has been significantly improved due to the integration of experimental information from chemical structure probing in the form of soft constraints [10,11]. Two main types of chemical probes are used, those such as DMS (dimethyl Sulfate) or CMCT (cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate), that probe the availability of H-bond donor or acceptor sites located on the Watson–Crick face of the base [12], and SHAPE probes that report on the flexibility of the ribose [13–16]. Tremendous efforts and progress have been made over the last years to develop high throughput experimental [17–21] and computing pipelines [22,23].

However, the accurate prediction of the secondary structure remains a challenge for long RNA because of intrinsic RNA properties and technical problems. Firstly, most RNA are likely to adopt several conformations whether in vitro or in a cellular context, which may in some instances reflect experimental artefacts, but also reflect their potential to regulate their function through structural rearrangement. RNA structure probing experiments then yield averaged signals representing all the conformations and therefore do not properly inform the prediction software. Secondly, RNA structure includes a handful of non-canonical base pairings accounting for a significant part ($\approx$25%) of the total number of base pairs [24] that significantly contribute to the folding stability, often by forming specific structural motifs. The relevance of such observation is striking in the case of the P4P6 domain of the *Didymium iridis* lariat-capping ribozyme (DiLCrz), the ribozyme used in this study. The DiLCrz crystal structure [25] shows a P4P5 secondary structure quite different from the originally predicted one [26]. In the original secondary structure, a four base pairs core is followed by two more taking place after a bulging U, whereas the crystal structure reveals that only one non-canonical base pair takes place after the bulge. The loss of the two WC base pairs is energetically compensated by the formation of a trans WC base pair, which fosters A residues from J5/4 to interact with the shallow groove of the P4 stem to weave A-minor interactions. In other words, the loss of secondary structure interactions is compensated by formation of tertiary ones. Such examples showcase that modeling of non-canonical and [canonical and non-canonical) tertiary interactions is an essential part of meaningful RNA structure modeling, even if such complex interactions add fundamental challenges in comparison to computational or computer-assisted modeling of canonical RNA secondary alone. Note that thermodynamics models and calculations without such interactions (i.e., restricted to canonical, secondary interactions) are long established and highly efficient for computer-assisted secondary structure prediction [27–29]. We can distinguish several lines of computational structure modeling including non-canonical and tertiary interactions. First, there are approaches that model RNA 3D structure, e.g., by full-atom fragment-based sampling methods [30,31] or coarse-grained simulation [32], which then allows reading-off secondary and tertiary interactions from the 3D representation. Typically, such methods are computationally limited, e.g., it is reported that "problems

involving de novo building more than 80 nucleotides will be challenging for FARFAR2 conformational sampling" [30] whereas successful predictions are reported for up to 500 nt by RNAvista [33,34], which integrates fragment-based modeling with base pair annotation.

Other approaches like MC-Fold/MC-Sym [35], CycleFold [36], and RNAwolf [37] include non-canonical secondary interactions in their prediction of extended secondary structure. Such methods have the potential to predict larger structures, but (by principle) they do not predict tertiary interactions. The exact prediction of tertiary interactions has been well studied in the form of pseudoknot prediction. Pseudoknot motifs, which frequently occur in functional RNAs, are characterized by a short Watson–Crick helix, the pseudoknot, the strands of which are topologically separated: one of its two strands is intercalated between the strands of at least one other helix of the RNA. Because their prediction is computationally intractable in realistic energy models (NP-hard, inapproximable) [38,39], efficient algorithms were presented only for restricted classes of pseudoknots (e.g., [40,41]) and the most popular structure modeling methods ignore them completely. Nevertheless, pseudoknot prediction has been implemented in several faster algorithms [42,43] that first calculate pseudoknot free structures and then search, in a second round, for pairings between the still single-stranded bases, or which stochastically simulate the folding of the RNA chain in the course of virtual transcription [44]. Although quite successful, this strategy faces two pitfalls, first it generates several models that the user cannot discriminate, secondly, it relies on the accurate prediction of the secondary structure in the first round which is not certain for the reasons mentioned above. Although experimental probing data significantly improve structure prediction, it is not of much help for non-canonical interactions or pseudoknot pairings. Indeed, as they are involved in genuine Watson–Crick base pairs, pseudoknot positions show probing patterns similar to regular helices, while tertiary interactions are often mildly reactive as they are usually more dynamic than canonical pairings. Despite the above-mentioned difficulties, teams of experts gathering many competencies are able to generate impressively accurate tertiary structure models for compactly folded RNA [45–47].

Here we investigate several approaches in order to precisely identify nucleotides involved in tertiary structure elements, notably pseudoknots and kissing loops. To this end we considered two physicochemical properties of these structures: first, their formation is stabilized by the presence of magnesium and second, they have lower thermal stability than the core secondary structure. Our mid-term goal is to develop an experimental and computational workflow that allows any molecular biology laboratory to automatically model RNA secondary structure including kissing loops and pseudoknots and indicate the presence of other tertiary interaction. We take advantage of the well-established crystal structure of DiLCrz to serve as a benchmark for our study. This compactly folded RNA gathers in only 188 nucleotides many of the recurring structural motifs found in RNA including a pseudoknot, a kissing loop pairing, and tetraloop/receptor interaction.

2. Material and Methods

2.1. RNA Preparation

RNA was in vitro transcribed by run-off transcription using T7 polymerase in 40 mM Tris-HCl pH 8.0, 25 mM $MgCl_2$, 5 mM DTT, 5 mM rNTPs, 1 mM spermidine and 20 U RNAsin (Promega, Madison, WI, USA). Template DNA was then digested at 37 °C for 20 min with RQ1 DNAse (Promega) and RNA was precipitated by addition of 2.5 M lithium chloride and centrifugation at $16,100\times g$ for 30 min at 4 °C. RNA pellet was washed with 70% ethanol then resuspended in nuclease-free water. RNA was purified through G25 size exclusion chromatography and quantified spectrometrically. Its integrity as well as the absence of aberrant products were confirmed by gel electrophoresis.

2.2. Melting Curves

Six pmoles of RNA were diluted in 14 μL of H_2O and denatured for 2 min at 80 °C. 4 μL of pre-warmed 5X folding buffer (HEPES pH 7.5 40 mM, KCl 500 mM, $MgCl_2$ ranging

from 0 to 5 mM final concentrations) were added and the solution was allowed to cool to room temperature in 5 min. After the addition of 2 μL of RiboGreen (ThermoFisher, Waltham, MA, USA, final concentration: 300 nM), the samples were incubated for 10 min at 37 °C. The melting curves were obtained with a PikoReal RT PCR system (ThermoFisher) by heating the sample from 37 to 95 °C at 0.04 °C·s^{-1} and measuring the fluorescence emitted by the ribogreen. Raw fluorescence curves were normalized and derived. Derivatives from three replicates were averaged and smoothed. Data analysis was performed using Graphpad Prism v5.02.

2.3. SHAPE Probing

SHAPE chemical probing was mostly performed as previously described in [48,49]. Briefly, 36 pmoles of DiLCrz RNA were diluted in 160 μL of water and denatured for 2 min at 80 °C. Then, 40 μL of pre-warmed folding buffer (HEPES pH 7.5 40 mM, KCl 500 mM, 5 mM MgCl$_2$) were added and the solution was allowed to cool down to room temperature for 5 min. The sample was then incubated at 37 °C. After 10 min, a 20 μL aliquot (corresponding to 6 pmoles) was removed and added to 2 μL of 40 mM 1M7 and allowed to react for 2 min, while a second 20 μL aliquot was mixed with neat DMSO. The temperature was then gradually (about 10 min between each step) increased to 53 °C, 65 °C, 74 °C and 85 °C and the same probing procedure was applied. Probing with NMIA was performed identically, BzCN probing was performed at a final concentration of 40 mM and HEPES was adjusted to 80 mM. Probed RNA were precipitated at −20 °C with 10% ammonium acetate 5 M, 2.5 vol. ethanol and 20 μg of glycogen as a carrier. The pellets were washed with ethanol 70%, vacuum-dried for 10 min and resuspended in 10 μL of water. For the reverse transcription, probed RNA were denatured for 3 min at 95 °C after the addition of 1 μL DMSO. 6 pmoles of fluorescent primers (WellRed D2 or D4 fluorophore, Sigma-Aldrich, Saint Louis, MO, USA) (5′-D2 or D4 CTG-TGA-ACT-AAT-GCT-GTC-CTT-TAA-TG-AAG-TAT-TTG-AGG-TGT-AGA-GTG-TTT-GAG-TAG-TAG-TGA-TTG-TCT-TGG-GAT-ACC-GGA-3′) were then added and allowed to hybridize by incubating at 65°C for 5 min then at 35 °C for 10 min. Finally, after the addition of 0.5 mM of each deoxynucleotide (guanine substituted with inosine) plus the MMLV reverse transcriptase (RNAse H (-)) and its buffer (Promega), the elongation step was achieved by incubating the samples for 2 min at 35 °C, 30 min at 42 °C and 5 min at 55 °C. The sequencing reaction was performed with 3 pmoles of untreated RNA, D2-labeled primer and 0.5 mM of dideoxyadenosine triphosphate during the elongation.

Fluorescently labeled cDNA were then precipitated at −20 °C with 10% sodium acetate pH 5.4, 20 μg of glycogen as a carrier and 2.5 vol. of ethanol. Pellets were washed with ethanol 70%, vacuum-dried for 10 min, resuspended in 40 μL of sample loading solution and run on a CEQ 8000 capillary electrophoresis sequencer (Sciex, Concord, ON, Canada). The resulting traces were analyzed using QuSHAPE [50].

Reactivities were the mean from three independent replicates. Within each triplicate, values more than 0.4 reactivity units apart from the other measurements were considered as outliers and discarded.

2.4. Comparison of Two Probing Profiles

Probing profiles were compared as described in [49]. In brief, reactivities averaged over three replicates in two conditions, R1 and R2, were used to calculate the absolute difference $\Delta R = |R_1 - R_2|$ and the relative change $\Delta R / (R_1 + R_2)$. The reactivity difference between two conditions is considered significant when both variables exceed a threshold of 0.2, and the *p*-value from a two-sided *t*-test is inferior to 0.05. Nucleotides with undetermined reactivity are excluded from the analysis. Such an approach excludes small reactivity changes between weakly reactive nucleotides as well as large yet unmeaning reactivity changes between highly reactive nucleotides.

2.5. Nucleotide Clustering

Nucleotides were clustered by the KMeans algorithm as vectors of reactivities obtained at 37, 53, 65, 74 and 85 °C. Clustering was done using scikit-learn and a custom python script available at https://github.com/gdebissc/DiLCrz, accessed on 3 November 2021. Data were log2-transformed and standardized (nucleotide-wise then temperature-wise). Undetermined reactivities were removed prior to clustering. Dimensionality reduction by principal component analysis reveals that nucleotides can be grouped in at least three or four clusters. The optimal number of clusters was further determined using the elbow method. Briefly, the within-cluster sum of squared errors, or distortion, was calculated for each cluster number between 1 and 10 (Figure S5B). The result graphically indicates that increasing the cluster number above four does not significantly reduce the distortion.

2.6. All-Atom Molecular Dynamics Simulations

MD simulations were carried out using GROMACS 5 package [51–54] using the Amber ff99+ parmbsc0 force field [55,56]. The molecular systems were placed in a cubic box and solvated with TIP4P-EW water molecules [57]. The distance between the solute and the box was set to at least 14 Å. The solute was neutralized with potassium cations and then K^+Cl^- ion pairs were added to reach the salt concentration of 0.15 M. We used the ion corrections of Joung et al. [58] as this force field has been shown to produce stable RNA structures [59]. The parameters for Mg^{2+} are taken from [60]. In our protocol, Mg^{2+} ions were first placed at known binding sites while the remaining ions randomly replaced solvent molecules to reach a concentration of 0.02 M [61]. Long-range electrostatic interactions were treated using the particle mesh Ewald method [62,63] with a real-space cut-off of 10 Å. The hydrogen bond lengths were restrained using P-LINCS [53,64], allowing a time step of 2 fs. The translational movement of the solute was removed every 1000 steps to avoid any kinetic energy build-up [65]. After energy minimization of the solvent and equilibration of the solvated system for 10 ns using a Berendsen thermostat (τ_T = 1 ps) and Berendsen pressure coupling (τ_P = 1 ps) [66], 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 [67] (τ_T = 1 ps) and a Parrinello–Rahman barostat (τ_P = 1 ps) [68]. During minimization and heating, all the heavy atoms of the solute were kept fixed using positional restraints. The length of the simulation was 1000 ns. To assess the stability of the trajectory, we verified that the root mean square deviation (RMSD) and the number of hydrogen bonds were stable along the trajectory.

2.7. Definition of the Parameters

To get insights into the relationship between the structure and the reactivity, we computed several quantities along the MD simulation. First, we computed the hydrogen bonds and the stacking for each frame using the software MC-Annotate [69–71]. The hydrogen bonds were classified using the Leontis/Westhof classification [24,72] and we detailed if the HBs are formed between the O2′ and an atom in the Watson–Crick (W), Hoogsten (H) or the B side (the position is between the W and H side) or between OP and an atom in these sides [70].

Second, we computed the multiplets (triplets, quadruplet, etc.) and the secondary structure for each frame using the software DSSR (dissecting the spatial structure of RNA) v.2 [73]. For each hydrogen bond and multiplet, we computed the percentage of the occurrence.

To take into account the ribose conformation, we described the sugar ring using pseudorotation parameters. Although there are four possible pseudorotation parameters for a five-membered ring [74], 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 [75]:

$$Amp = \sqrt{(a^2 + b^2)} \; ; \; Pha = \left(\frac{a}{Amp} \right)$$

where $a = 0.4 \sum_{i=1}^{5} v_i cos[0.8\pi(i-1)]$ and $b = -0.4 \sum_{i=1}^{5} v_i sin[0.8\pi(i-1)]$, with v_i the ring dihedral angle i. This approach has the advantage of processing the ring dihedrals v_1 (C1′-C2′-C3′-C4′) to v_5 (O4′-C1′-C2′-C3′) in an equivalent manner [76]. 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 [76] 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.

Finally, as in our previous work [14], we computed a set of distances, valence angles and dihedral angles to capture the flexibility of DiLCrz RNA structure. For each quantity, we computed the average value and its standard deviation.

3. Results

3.1. Probing DiLCrz Structure in Presence or Absence of Mg^{2+} Ions

DiLCrz RNA structure was first probed with different SHAPE reagents, namely NMIA, 1M7 and BzCN in native-like conditions including 5 mM of magnesium ions (Figure 1A). The reactivity maps obtained are discussed in the light of the X-ray structure previously reported [25]. We use the following nomenclature: Px are helices numbered as in Figure 1, Lx are the corresponding terminal loops, Jx/y are the nucleotides joining helix x to helix y following a 5′ to 3′ orientation, and 3WJ is the P2/P2.1/P10 junction. The exact nature of all the interactions evidenced in the 3D structure is summarized in the Supplementary Table S1. Overall, the three reactivity patterns fit remarkably well with the three-dimensional structure. The crystal structure was accompanied by SAXS data that already showed the agreement of the conformation in solution with the crystal structure. All these remarks indicate that the dominant conformation of DiLCrz in solution corresponds to the reported three-dimensional structure. Indeed, most nucleotides predicted to be involved in cis Watson–Crick base pairing are not reactive. Exceptions to this trend correspond to the region U_{47}-G_{49}/A_{143}, and to few positions at the edge of helices (G_{12} (P2) or U_{128}-A_{132} (P8)), which appear to react with NMIA, 1M7 or BzCN although they are located within a supposedly stable stem (Figure 1A). Very reactive nucleotides are located in loops, while moderately reactive positions are mostly involved in non-canonical or long-range interactions. However, the reciprocal is not true, for instance, not all single stranded nucleotides are reactive (see for instance G_{74}, G_{87} or A_{101}), and the reactivity of many nucleotides involved in tertiary contacts is indistinguishable from the reactivity of the nucleotides involved in the secondary structure. Notably, most nucleotides involved in pseudoknots are not reactive, which is deceptive for modeling algorithms that cannot predict such type of interaction at the first intention. As previously described for NMIA and 1M6 with various RNA [77,78], the reactivity profiles are very similar for the three probes, although some nucleotides appear to have different susceptibility to the different reagents (see Figures 1A and S1). Yet, as discussed below, no systematic rule informative for modeling algorithms could be deduced from comparing the different profiles.
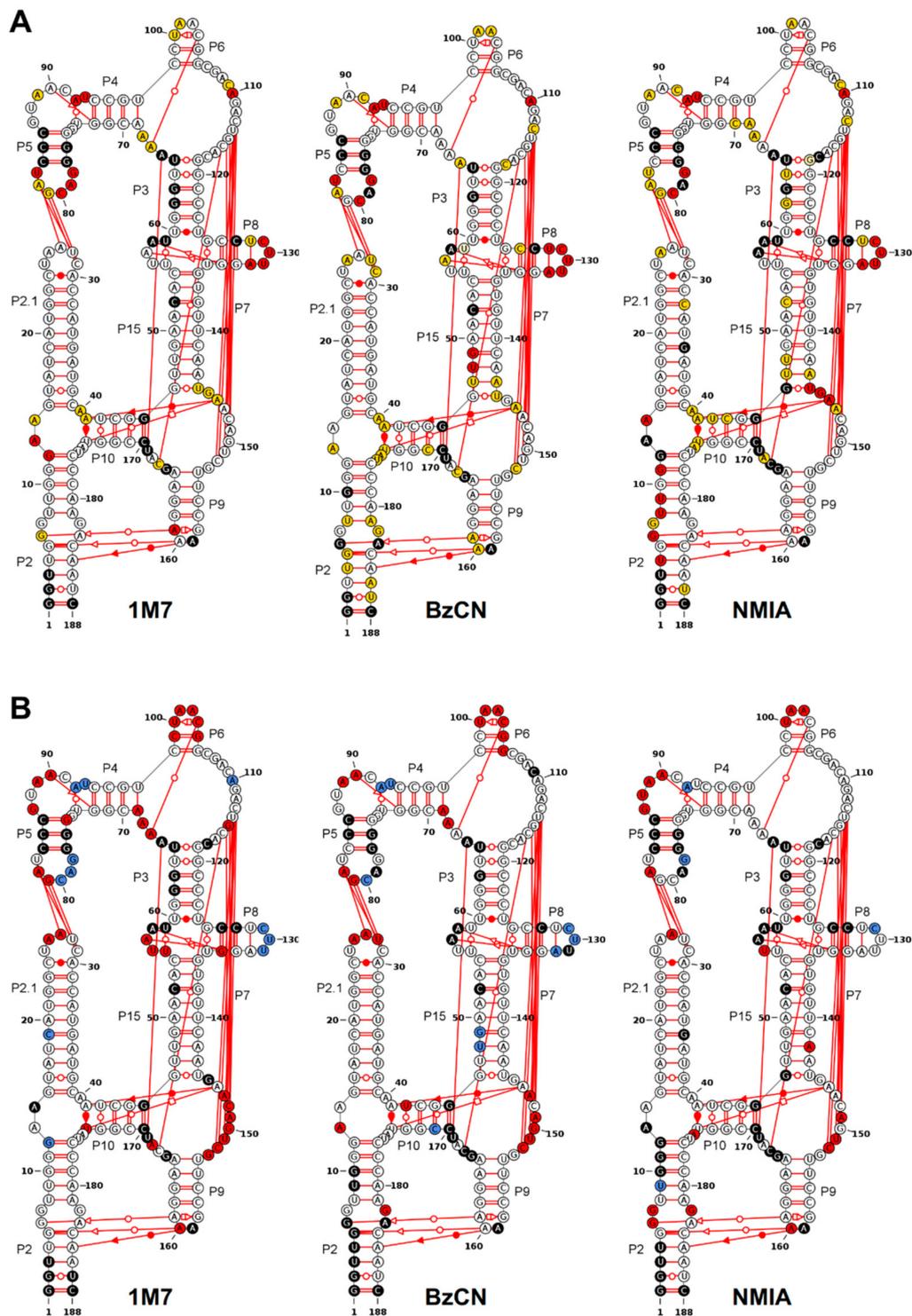
Figure 1. DiLCrz reactivity profiles with 1M7, BzCN and NMIA. 1M7, BzCN or NMIA were used to probe the structure in the presence or absence of magnesium. Probing was conducted at 37 °C in 40 mM HEPES pH 7.5, 500 mM KCl with or without 5 mM MgCl$_2$. (**A**) Color coded reactivities obtained in the presence of magnesium pasted over DiLCrz secondary structure. White: low reactivity (0 to 0.4), yellow: moderate reactivity (0.4 to 0.7), red: high reactivity (over 0.7). (**B**) 1M7 reactivities obtained in the presence and absence of magnesium were compared as described in Material and Methods. Red: more reactive in the absence of magnesium. Blue: more reactive in the presence of magnesium. White: no significant difference. Nucleotides in black denote undetermined reactivity.

Magnesium ions are known to be a key element for RNA tertiary structure stability and compaction [79–82]. In order to capture the magnesium-dependent folding of DiLCrz the RNA was probed with the three SHAPE reagents in the absence of $MgCl_2$. Here again, although not identical, the reactivity patterns obtained are very similar. Many of the nucleotides showing a reactivity change upon Mg^{2+} addition are identical for the three different probes (Figures 1B and S2). Most importantly, the vast majority of the less reactive nucleotides (27 out of 32, i.e., 84% for 1M7) in the presence of Mg^{2+} are either single-stranded, involved in the pseudoknot, the kissing loop or a non-canonical interaction according to the crystal structure (see Supplementary Table S1). Among the exceptions are nucleotides involved in the 2 bp P6 which would be predicted to be very unstable on its own and is stabilized by the tertiary structure. A_{69} or G_{75} involved in helix closing base pairs also appear more sensitive to the probe in the absence of magnesium ions as well as G_{134} which is also involved in a tertiary interaction. Interestingly enough, few positions appear more reactive in the presence of Mg^{2+}, among which nucleotides in L5 and L8 have been observed to be the most mobile in the crystal structure. Forty-four nucleotides are only involved in tertiary contacts, i.e., participating in pseudoknot-type interactions or in non-canonical base pairs but not at the same time in a WC base pair (triple interaction). Amongst these, 19 are significantly less reactive in presence of Mg^{2+} towards one probe or another (15 with 1M7). In contrast the reactivity of only 7 (for the three probes) out of the 102 nucleotides only involved in regular helices are sensitive to the presence of Mg^{2+}. However, if probing in the absence of magnesium fairly clearly highlights nucleotides involved in the three-dimensional folding, there is one notable exception. Indeed we noted that the reactivity of the 5′ strand of the P7 pseudoknot (G_{111}-G_{115}) remains essentially unaffected by the absence of magnesium while P7-3′ becomes very reactive. This suggests that the pseudoknot is destabilized in absence of Mg^{2+} and that under such conditions P7-5′ sequence adopts an alternative folding. Interestingly, unconstrained predictions using RNAfold [83] or RNAstructure [84] consistently yield models in which A_{112}-G_{115} are base paired with C_{70}-U_{74} (Figure S3).

Thus, albeit very informative, the comparison of SHAPE reactivity profiles in the presence and absence of magnesium proves, in this case, to be insufficient to pinpoint the nucleotides involved in kissing-loops and pseudoknots.

3.2. Structure Characterisation by Thermal Denaturation

In order to evaluate the thermal stability of the various three-dimensional structural elements and highlight the involved nucleotides, we followed DiLCrz thermal unfolding. We first set up a melting experiment using the RiboGreen fluorescent dye as described [85]. Ribogreen fluoresces upon specific binding to single strand RNA. Upon temperature increase the different structural elements unfold, more single strand RNA becomes available for ribogreen binding and consequently increases the fluorescence signal. However, in the meantime, RiboGreen binding to RNA is destabilized as the temperature increases. The fluorescence curve in function of temperature therefore results from a non-specific constant fluorescence decay inflected or even inverted by the local increase of RNA single strand availability. As for UV thermal melting results, the curve inflections are determined by plotting the derivative of the relative fluorescence units (RFU) over temperature (-dRFU/dt). In such a plot, maxima represent the melting temperature transitions. Using this technology, we followed DiLCrz thermal denaturation at increasing $MgCl_2$ concentrations ranging from 0 to 5 mM (see Figures 2A and S4). In absence of Mg^{2+}, fluorescence regularly decreases until 62 °C with only a slight inflection at 50 °C, followed by a significant increase peaking at 75 °C. The 50 °C inflection becomes more pronounced as [$MgCl_2$] increases, shifts to 53 °C in presence of 2.5 mM and is finally individualized as a 57 °C peak at 5 mM $MgCl_2$. In the meantime, the 75 °C peak gradually shifts to a higher temperature to reach 81 °C at 5 mM $MgCl_2$ where a slight inflection at 70 °C can also be observed. As the 57 °C peak builds up with the presence of magnesium ions, we interpret it as the cooperative unfolding of the tertiary structure while the large peak at the higher temperature would

represent the cumulative melting of the multiple secondary structure elements. To confirm this hypothesis, we repeated the experiment using three variants designed to disrupt the tertiary structure, $A_{90}G$, $G_{111}C$ and $A_{168}G$. Mutation $A_{90}G$ disrupts the triple interaction A_{90}/G_{72}-C_{94} that stabilizes P5-P4-P6 stacking, $G_{111}C$ destabilizes the P7 pseudoknot, and finally, A_{168} interacts with A_{110} and is in the heart of the catalytic center since the branching reaction occurs between C_{167} and U_{169} [86]. As can be observed in Figure 2B, in the presence of magnesium the 57 °C peak is almost or even completely abolished for the three mutants while the 80 °C peak is present for all three mutants and essentially similar to what is observed with the WT. In absence of magnesium, the melting profiles of the WT and the three variant constructs are almost identical, showing a single major peak at 75 °C (Figure 2C). This strongly suggests that the 75–80 °C represents the secondary structure, stable in absence of magnesium ions, while the 57 °C peak can be attributed to the cooperative melting of the tertiary structure which is formed only in the presence of divalent ions.
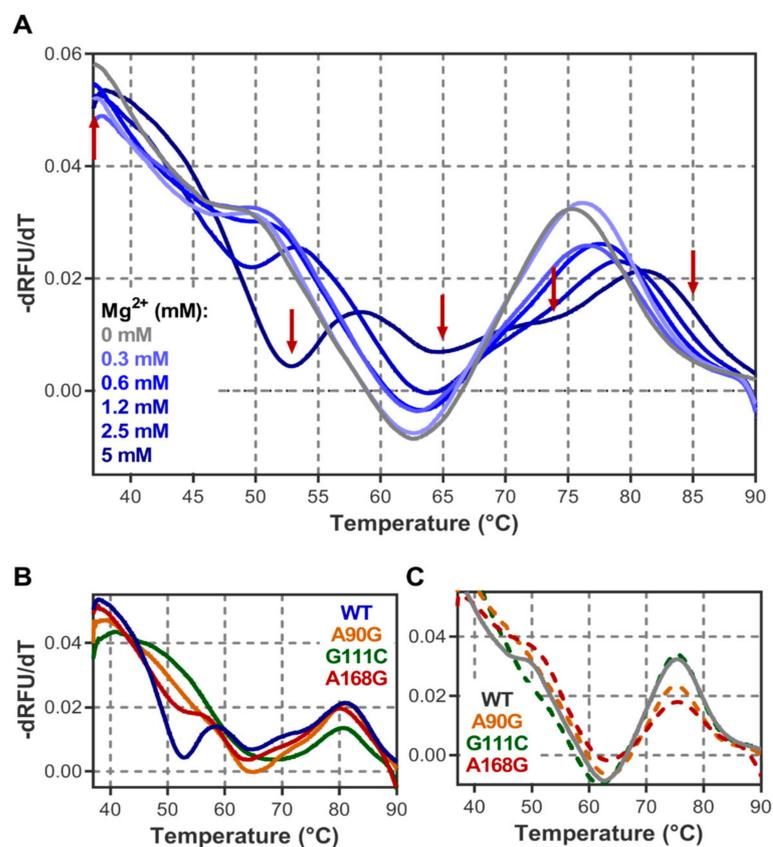

Figure 2. DiLCrz melting curves. DiLCrz equilibrated at 37 °C in 40 mM HEPES pH 7.5, 500 mM KCl and the indicated concentrations of magnesium was heated at 0.04 °C/s up to 95 °C, and melting curves were acquired using the RiboGreen dye fluorescence. (**A**) DiLCrz melting profiles in the function of Mg concentration. Gray: absence of magnesium; light blue to dark blue: 0.3, 0.6, 1.2, 2.5 and 5 mM $MgCl_2$, respectively. Red arrows indicate the temperatures at which probing was conducted. (**B**,**C**) Effects of $A_{90}G$, $G_{111}C$ and $A_{168}G$ mutations on the melting curves in the presence (**B**) or absence (**C**) of 5 mM $MgCl_2$. Fluorescence derivatives were averaged from three independent experiments and smoothed.

These results indicate that the thermal stability of DiLCrz secondary and tertiary structures are in two distinct temperature ranges under our experimental conditions. This led us to measure the 1M7 SHAPE reactivity profiles along the thermal denaturation. We thus established the reactivity profiles at five different temperatures from the folded structure (37 °C) to the essentially denatured sequence (85 °C) going through the local

minima 53 °C, 65 °C and 74 °C (Figure S5). As expected for a progressive unfolding process, the average reactivity gradually increases while the reactivity value dispersion decreases with temperature. This is consistent with all nucleotides becoming equally very reactive. Note that some nucleotides still appear poorly reactive at 85 °C. This is due to the reactivity calculation/normalization process and means that they are less reactive than the average, but not that they are unreactive. In order to get a clearer view of the behavior of each nucleotide along the temperature curve, we clustered their reactivity profile using a k-means algorithm (see material and methods). Four clusters define the following categories. The first (grey in Figure 3A) gathers nucleotides that are very reactive at 37 °C, and remain very reactive (although their absolute reactivity value goes down because of the reactivity normalization process [50]). The second cluster regroups nucleotides which become highly susceptible to 1M7 modification over 37 °C (red in Figure 3A), while the third cluster is constituted by nucleotides which become increasingly reactive with temperature starting at 53 °C (green in Figure 3A). Finally, the last category clusters positions that become increasingly reactive only over 65 °C (blue in Figure 3A). Annotating the DiLCrz structure scheme makes visible that essentially the four categories also coincide with the different types of structural elements (Figure 3A,B). Indeed, most nucleotides of the same structural element fall in one of the four categories which can roughly be defined as single strand nucleotides (grey), nucleotides involved in tertiary contacts (red), secondary structure (green), and finally the highly stable helix P2.1 (in blue). Note that although highly unstable the U_{128}-A_{132} base pair is attributed to the blue cluster, this is an artefact and limitation of our clustering methods. As general trends, tertiary structure opens up cooperatively between 53 and 65 °C while secondary structure starts melting at 65 °C and some elements appear to be stable over 80 °C. Notable exceptions are nucleotides in P10 (U_{42}, G_{171}, G_{172}), the 5' part of P3 (G_{61}) and the bottom of P15 (C_{142}-A_{143}) that cluster with the tertiary contact nucleotides (red in Figure 3A). Indeed, those nucleotides appear to be involved in peculiar or dynamic pairings (see below). As nucleotides clustered on the basis of a clear reactivity increase between 37 and 53 °C appear to be mostly involved in tertiary structure, we carried out a statistical analysis to identify the positions with a significantly increased reactivity over this temperature transition (see Figure 3C and material and methods). Interestingly this approach is more conservative than the clustering, highlighting most nucleotides involved in the tertiary structure (non-canonical base-pairs and long distance pairing involved in pseudoknots). Notably, nucleotides involved in both sides of the P7 pseudoknot and the P5-P2.1 kissing loop are emphasized by this strategy. In addition to information about the tertiary structure, SHAPE reveals that the different helices open up at different temperatures, but not always in the order predicted by thermodynamic calculations on isolated helices. For example, P15 ($\Delta G° = -8.1$ kcal), which is predicted to be the second most stable secondary structure element after P2.1 ($\Delta G° = -10.7$ kcal), is essentially accessible at 65 °C while nucleotides within P2.1 are not reactive before 85 °C. Even more surprising, P9 ($\Delta G° = -2.8$ kcal) or even the two base pair helix P6, which is predicted to be hardly stable when isolated, are not highly reactive below 85 °C. This is a mere example that thermodynamic calculation on isolated duplexes poorly describes helices that are part of a larger structure.
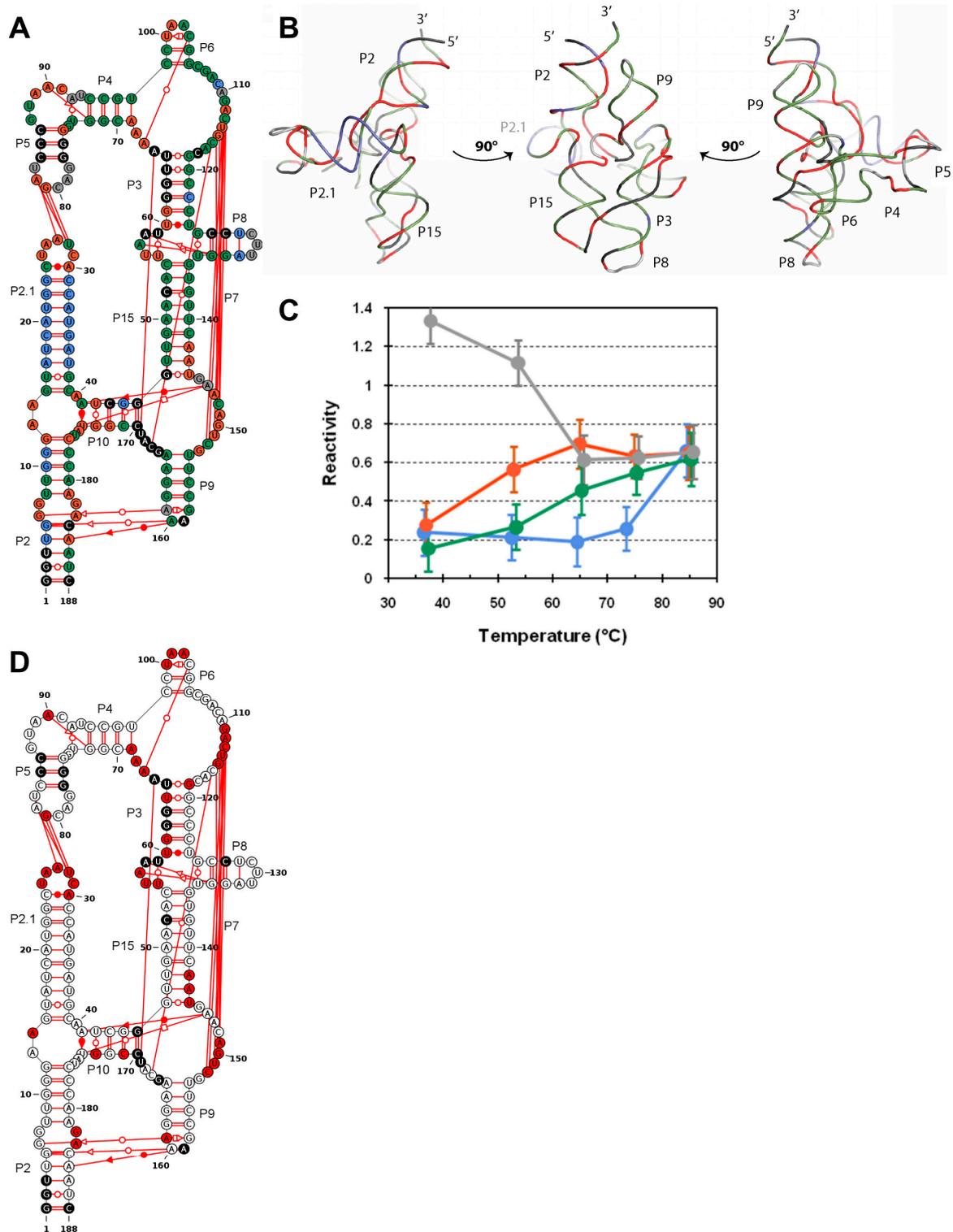
Figure 3. DiLCrz reactivity profiles at different temperatures. DiLCrz structure was probed with 1M7 at 37, 53, 65, 74 and 85 °C in 40 mM HEPES pH 7.5, 500 mM KCl and 5 mM MgCl$_2$. (**A**) Clustering of the nucleotides according to their reactivity change. KMeans clustering was performed on 1M7 reactivities. DiLCrz secondary structure coloring refers to the nucleotides' assigned clusters. (**C**) Previously reported DiLCrz X-ray structure [25] colored according to the clustering. (**B**) Evolution of the average reactivity within each of the four clusters. Error bars refer to the reactivity mean absolute deviation within each cluster. Points are horizontally jittered for clarity. (**D**) Comparison of 1M7 reactivities between 37 and 53 °C. Significant reactivity differences are highlighted in red. The comparison was performed as described in Material and Methods. Nucleotides in black denote undetermined reactivity.

3.3. Molecular Dynamics Analysis

The vast majority of the high SHAPE reactivities observed in the study described above reveal single-stranded positions, while nucleotides involved in the secondary structure are essentially unreactive. However, we were puzzled by the reactivity of some nucleotides involved in stable Watson–Crick pairings which also appear to be differentially susceptible to the different SHAPE probes, or to be unexpectedly reactive at 53 °C or in the absence of divalent cations. In order to get insight into these unexpected SHAPE reactivities and better interpret probing results, we resort to molecular dynamics (MD). We combined the different quantities computed along the MD simulations to shed light on some particular regions of DiLCrz structure. First, we focused on the nucleotides that for each pair of probes show a significant difference in chemical reactivity. The nucleotides U_{83} (the 3′ residue from L5) and A_{161} (the 3′ residue from L9) are both involved in non-canonical hydrogen bonds and at some point of the MD simulation they can form a triplet (26% of the time) and a quadruplet (97% of the time), respectively. Figure 4A–C shows some distinct conformations for the pairing of the nucleotides U_{83} and A_{161}. The ribose of U_{83} not only adopts several conformations, in particular the C2′-endo and C3′-endo, but also some pseudo-rotation intermediates, which is a sign of medium to high reactivity [14]. Moreover, the global flexibility of the nucleotide can explain the high reactivity for the three probes. On the contrary, in the case of A_{161}, the ribose puckering changes from the stable C3′-endo conformation to an intermediate one (C2′-exo), and is less flexible than the previous ones along the simulation.

Second, we analyzed the nucleotide G_{49} (P15) that has a low reactivity to NMIA and 1M7 and a high reactivity toward BzCN. This result suggests that this nucleotide assumes a transient state before reaching a final and stable conformation. This hypothesis is supported by the observation that along the MD simulation, on the one hand, the puckering of G_{49} can briefly adopt a C2′-endo conformation and on the other hand, the local flexibility suggests the presence of two possible local conformations. A similar trend has been observed for the nucleotide G_{182} that has also a low reactivity with NMIA and 1M7 and a medium reactivity with BzCN.

Third, we focused on P10 and the nucleotides upstream and downstream from this helix (see Figure 4D). The change in reactivity for the region A_{41}-G_{44} may reflect the ability of these nucleotides to form triplets along the MD simulation. Although most of this region is not very dynamic, while A_{41} can form several non-canonical hydrogen bonds and at the same time the ribose adopts a C2′-endo conformation, the ribose of U_{42} often adopts intermediate conformations and the nucleobase stacks with A_{146} (downward stacking between U_{42} and A_{146} which also forms an outward stacking with A_{175}).

Finally, we analyzed the P15 region, G_{46}-A_{53}/U_{137}-U_{144} (see Figure 4E,F). Both strands of the helix are relatively flexible, however, in the region G_{46}-A_{53} the ribose moieties are on average more flexible than in a standard helix and often adopt the C2′-endo conformation. On the contrary, within the U_{137}-U_{144} strand, only the ribose of U_{144} is in the C2′-endo conformation. Moreover, on the 5′ strand nucleotides G_{46} and G_{49} stack with non-adjacent nucleotides C170 and A142, respectively. G_{46} also forms a non-canonical hydrogen bond with C_{118} (O2′B sides) along the MD simulation. These observations could explain why its reactivity increases up to 53 °C, although it is seen as a secondary structure pairing.
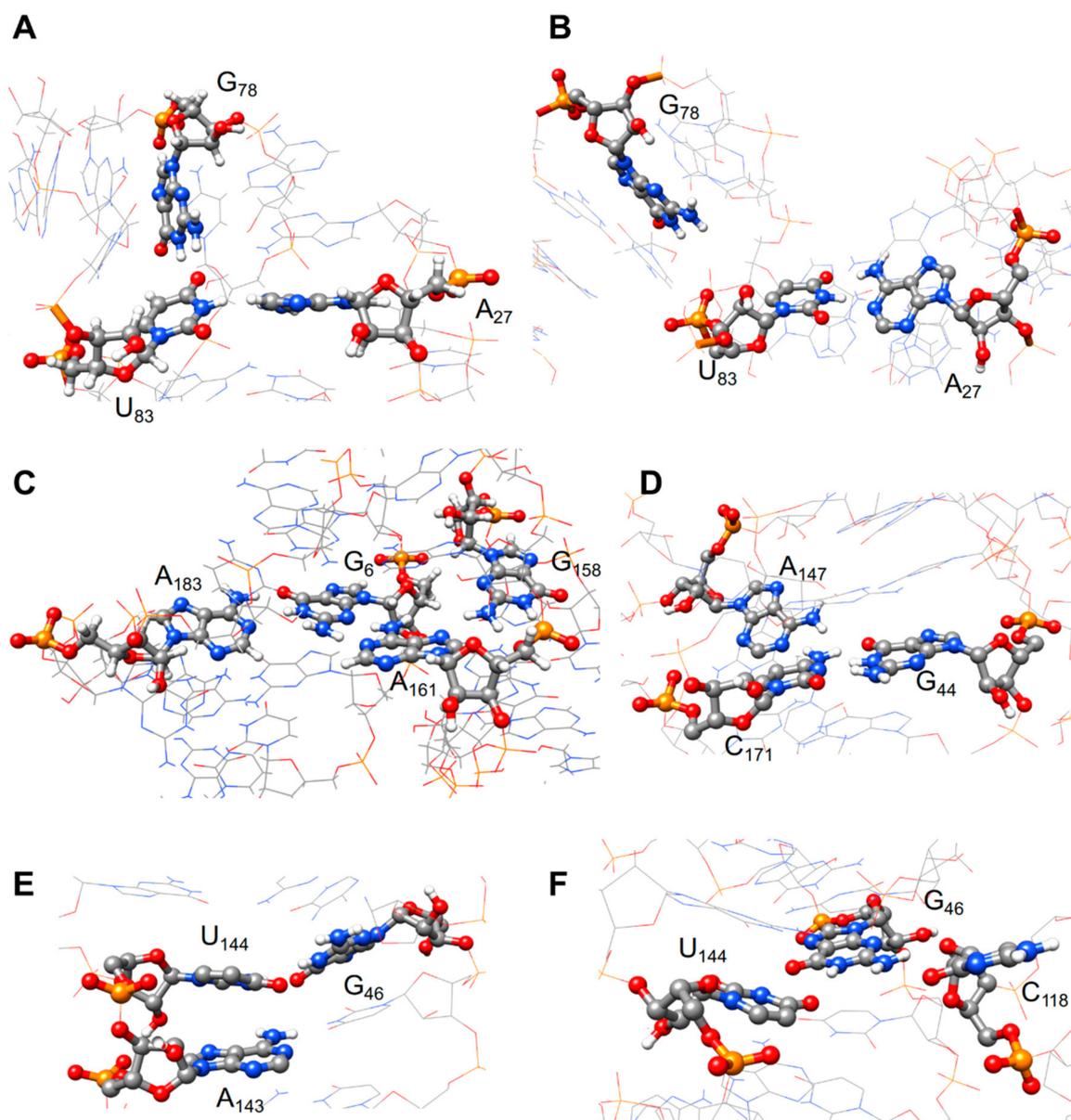
Figure 4. Different base pairing for some nucleotides of DiLCrz along MD simulation. Examples of different base pairing for the nucleotides U_{83}, A_{161} and X obtained along the MD simulations. (**A**) Triplet (U_{83}, A_{27} and G_{78}). U_{83}-A_{27}: cWW. U_{83}-G_{78}: tWH. (**B**) Canonical base pairing between U_{83} and A_{27}, the HB between U_{83}, and G_{78} is lost. (**C**) Quadruplet (A_{183}, G_6, A_{161} and G_{158}). G_{158}-A_{161}: tSH. G_6-A_{161}: tSW. G_6-A_{183}: cWW. (**D**) Triplet (C_{171}, A_{147} and G_{44}). G_{44}-C_{171}: cWW. C_{171}-A_{147}: cHW. (**E**) Triplet (U_{144}, A_{143} and G_{46}). G_{46}-U_{144}: cWW. U_{144}-A_{143}: H/O2′ (inward stacking). (**F**) Triplet (U_{144}, C_{118} and G_{46}). G_{46}-U_{144}: cWW. G_{46}-C_{118}: B/O2′.

3.4. Using Probing Data for Secondary Structure Modeling

The probing results performed with the different SHAPE probes at 37 °C in the presence, or absence of Mg^{2+} ions were used as constraints for secondary structure prediction software. Most of the secondary structure was correctly predicted by RNAstructure and RNAfold, which both do not predict pseudoknots. However, in most of the proposed models the P7 part of the pseudoknot nucleotides are incorrectly paired (see Figure S2). In many of the models, A_{112}-G_{115} are paired to C_{70}-U_{73} compromising not only P7 but also the prediction of P4 and P6. Such misprediction can be expected as the P7 pseudoknot nucleotides are unreactive to the SHAPE reagents in native conditions, and only the 3′ part of the pseudoknot becomes reactive in absence of Mg^{2+}. In order to better inform the

prediction software about the tertiary structure and pseudoknots, the reactivity profile obtained at 53 °C was integrated in two different ways. In a first approach, 1M7 reactivity data collected at 53 °C were directly used as constraints. In a second approach, the nucleotides with significantly augmented reactivity when raising the temperature from 37 to 53 °C, were assigned a reactivity value of 10 while the reactivity of other nucleotides was kept to their value obtained at 37 °C with 1M7, NMIA or BzCN. This strategy prevents the formation of secondary structure pairings with the nucleotides involved in pseudoknots but does not improve the overall prediction. For instance, in some of the resulting models, P3 is destabilized to the benefit of spurious pairings.

We then evaluated the performance of our recently developed workflow IPANEMAP, which extends RNAfold-based probing-informed modeling by sampling and clustering steps that allow us to take multiple sets of probing data into account [87,88]. When using 1M7 reactivities as constraints IPANEMAP yields two models almost equiprobable. The first is the model described above with scrambled P7 and P4 for the prediction obtained with RNAstructure or RNAfold. In the second model, most pairings are correctly predicted including P4 and three out of five of the nucleotides involved in the P7 pseudoknot (G_{111}-C_{113}) are left unpaired in the loop of an alternative P6 hairpin (Figure S7A). Running IPANEMAP with multiple sets of constraints representing any combination of the data obtained with other reagents, and/or in any of the other conditions described does not improve the models. Interestingly, when informed with the reactivity data obtained using NMIA in the presence or absence of Mg^{2+} IPANEMAP yields one model which captures most of the secondary structure, but the two base-pair P6 (Figure S7B).

In another approach, we sought to predict the P7 pseudoknot pairings from the experimental results. To this end, we generated artificial constraints files aiming at indicating nucleotides potentially involved in pseudoknots to the prediction software. Reactivity of positions for which the SHAPE values significantly vary between two conditions was set to -0.2 while the value for other nucleotides was set to 5. When using such artificial constraints obtained from the reactivity with or without magnesium no base pair was predicted independently of the prediction software utilized. In contrast, when following this strategy with the probing results obtained with 1M7 at 37 and 53 °C, the only pairings predicted form P7. We conclude that the temperature differential SHAPE strategy allows for the prediction of DiLCrz pseudoknot.

4. Discussion

Evaluation of the reactivity profiles regarding the X-ray crystal structure confirms that SHAPE probing very accurately reflects RNA folding. Nucleotides highly reactive are single-stranded, while positions not susceptible to modification are involved in canonical pairings. Finally, nucleotides showing a mild reactivity are most of the time engaged in dynamic tertiary pairings. Molecular dynamics simulation further shows that paradoxical reactivity observed for nucleotides within stable helices most probably reflects the fast equilibrium of these nucleotides between two conformations. We then followed two strategies to pinpoint nucleotides involved in tertiary structure. The differential Mg^{2+} SHAPE analysis selectively reveals many of the nucleotides involved in non-canonical interactions. As expected, nucleotides involved in pseudoknots are also more reactive in absence of Mg^{2+}, although they are WC cis base pairs. However, if the 3′ strand of P7 is well highlighted in the absence of Mg^{2+}, the 5′ strand remains poorly reactive toward all probes. As suggested by some of the models obtained, it may reflect the involvement of these strands in alternative structure in absence of Mg^{2+} since this region mostly relies on tertiary interactions. The temperature-dependent SHAPE probing seems to be promising to spot residues involved in long-range pairings forming pseudoknots.

Several studies have reported the effect of magnesium and/or temperature on RNA chemical reactivity [89–95]. We observed an asymmetric change in P7 reactivity when magnesium was omitted. This observation is reminiscent of what has been observed with the env25 pistol pseudoknot. In this case, the 5′ strand reactivity is rather insensitive

to the presence of magnesium while the 3′ strand reactivity increases in the absence of magnesium [95]. It also resembles temperature-dependent probing of tRNA^{Asp} where one strand of the D-stem becomes reactive at a temperature higher than the second one [92]. The authors mentioned that purines may maintain stacking interactions in the absence of base-pairing better than pyrimidines, which may be due to their bicyclic chemical structure. In line with this hypothesis, the purine-rich strand is the most stable in each of the three examples of asymmetric change in reactivity discussed here (DiLCrz P7, pistol pseudoknot and tRNA^{Asp} D-stem). Although P7 pseudoknot reactivity is asymmetrically affected by magnesium, it is symmetrically affected by the increase of temperature. In fact, probing upon temperature ramping and probing in presence and absence of magnesium reflect well distinct processes. While in the magnesium-dependency experiment, the RNA is folded either in the presence or absence of magnesium, in the temperature-dependent experiment it is first folded at 37 °C then progressively denatured. Thus, in the first case, we observe two independent conformations while in the second case the successive conformations are dependent on the initial conformation.

4.1. DiLCrz Hierarchical Thermal Unfolding

The four different temperature clusters suggest somehow an unfolding model of the DiLCrz. Below 37 °C, the reactivity of P8 and P5 loops (L8 and L5) indicate intrinsic molecular motions that are observed in the crystal structure. While there is no electronic density for residues from L8 in the X-ray DiLCrz crystal structure, L5 is observed since the kissing interaction with L2.1 somehow maintains the L5 structure, in spite of greater than average temperature factors (B factors), which express the uncertainty of the atomic coordinates. Highly reactive residues from J6/7 (C_{109}-A_{110}) have also higher temperature factors. The same observation applies to residues from J15/7 (G_{145}, A_{146}) which interact with the three-way junction (P2/P2.1/P10) and are also characterized by higher B factors. At the same time, L9 is also reactive, indicating a strong cooperativity between the tertiary interactions L9-P2, L2.1-L5, and the J15/7-3wj interaction. The 3wj acts as a receptor for A_{146}. This interaction is instrumental to stabilize the trans WC pair G_{46}-U_{144}. Heating beyond 37°C provokes melting of this G-U pair and other residues from cluster 2 (red). Opening up the P15 G-U pair is accompanied by melting of P10 and P7. Simultaneously, the tertiary interactions L9-P2 and L2.1-L5 are lost. The A-minor motif from J5/4 is also lost, breaking J3/4, which cannot stabilize P6 anymore. Then, when 65 °C is reached, residues from cluster 3 (green) start to melt. It is worth noting that, at this stage, the secondary structure elements that are intimately linked to the tertiary structure have already been partially lost, like P10 and P4P6. The fourth cluster regroups residues that remain folded in spite of a high temperature (blue). Surprisingly, the most stable element is P2.1. Since P2.1 is on the 5′ side of the ribozyme, it is transcribed early during transcription and is presumably the first element of the ribozyme to be folded. Moreover, this element, together with P2 and P10, constitutes a three-way junction, which binds A_{146} in J15/7, a critical region for catalysis, thus potentially regulating the activity of the ribozyme [86]. The present data thus suggest that P2.1 may be a folding nucleation element toward the active structure of the ribozyme.

The present scheme is supported by previous studies on the tRNA^{Asp} thermal unfolding using NMIA [92]. The authors identified five melting transitions that can be clustered in three main unfolding events. The initial event is described by the first two transitions (51 and 53 °C, respectively), first concerning the melting of the tertiary interactions mediated between the D- and T-loops, secondly followed by D-stem strands separation (53 °C). The second event at 58 °C mainly consists in melting of the acceptor and anticodon regions, which reorganize by forming alternative pairings between the 3′ residues from the acceptor stem, the 5′ strand form the anticodon stem, and the 3′ end from the anticodon loop. The formation of this alternative secondary structure presumably compensates for the tertiary structure loss. The third event occurs when the alternative stems melt individually when the temperature rises. The T-stem loop remains incredibly stable along the whole process

and may be compared to P2.1 from DiLCrz. Although we cannot exclude that alternative secondary structures could be formed during thermal unfolding of DiLCrz, the present SHAPE data do not present differential reactivities between cognate helical strands at temperatures <65 °C. The DiLCrz unfolding scheme thus appears as a more hierarchical process than the one from tRNA^{Asp}.

4.2. Molecular Dynamics Shed Light on Unexpected Reactivity

As previously reported, the different SHAPE probes show very similar but not identical reactivity patterns [77,96]. The analysis of the structure dynamics of DiLCrz shows that when significant differences are observed between the probes, the reactivities are medium to high, and the nucleotides are most of the time involved in multiplets. For example, the nucleotide A_{161} is mostly involved in a quadruplet and there are significant differences between each pair of probes. Another example is given by the nucleotide C_{171} that is involved in either a triplet or a multiplet with four other nucleotides.

In the case of G_{182} where low reactivity towards both NMIA and 1M7 was observed, while a high/medium reactivity was observed with BzCN, the nucleotide adopts a transient conformation before reaching a final and stable state (see Video S1). Indeed, along the MD simulation the puckering of the nucleotide briefly adopts a C2′-endo conformation, changing from C3′-endo and some intermediate states close to either the C2′-endo or the C3′-endo conformation. In addition, the structural local flexibility suggests the presence of two possible conformations.

Finally, our study confirms that the flexibility of the sugar is a key element for the chemical reactivity but that further investigations are still required to better characterize the SHAPE reagent reactivity and clarify what appears to be a complex scenario. For instance, the high to medium reactivity observed with both NMIA and BzCN within P15 were quite puzzling and could lead one to question a model displaying P15. Molecular dynamics suggests that albeit stable this helix is flexible, showing transient alternative conformation and unusual stacking.

4.3. Informing Prediction Software to Integrate Pseudoknots and Other Tertiary Motifs

Although pseudoknots are present in many RNA, predicting them is still a challenge in terms of modeling and computation. Thus, as exemplified in this study, failing to take them into account may lead to incorrect secondary structure models. While more general and exact methods seem precluded, predicting pseudoknots in two rounds has been applied successfully; most prominently in ShapeKnots [43]: a first round predicts several models for secondary structure, and a second round searches for complementarity between the remaining single stranded nucleotides; identified pseudoknots are rewarded with a thermodynamic bonus. Using the 1M7 reactivity pattern obtained in this study, ShapeKnots delivers ten possible models among which the second most stable is a model including the P7 pseudoknot and otherwise correct but for the absence of P6 and the kissing-loop between P5 and P2.1. Despite this unquestionable success, in absence of additional data, it is impossible for the experimenter to decide which of the 10 models is the most likely. Data such as those presented here (Mg^{2+} or temperature differential SHAPE) undoubtedly indicate which model to choose. As shown above, the data obtained with the temperature differential SHAPE allow to identify unambiguously the P7 pseudoknot, and when constrained with the 1M7 or the NMIA reactivity map IPANEMAP yields models in which respectively, part or all of the P7 sequences are left available. The combination of both steps would lead to a model identical to the one obtained with ShapeKnots. However, such a post modeling evaluation for ShapeKnots or semi-automated modeling for IPANEMAP is feasible in the modeling of a short RNA such as DiLCrz, but does not appear realistic in the context of longer RNA. In the near future we will extend the IPANEMAP workflow to infer tertiary interactions based on multi-condition probing data. The system will be informed of nucleotides with a high probability to form tertiary structure. The software will at first exclude them from secondary structure modeling, which allows it to identify

pseudoknot helices and even recurrent tertiary structure motifs in a second modeling phase. For instance, comparing the profiles of the three probes on DiLCrz highlights A_{161}, a nucleotide part of a GAAA tetraloop that interacts with a receptor. Such motifs are frequent within RNA structures, the combination of multiple SHAPE probing and the signature sequence (GNRA) could help to identify such structures ab initio. One of the challenges in modeling DiLcRz structure is the prediction of the L2.1/L5 kissing loop. Both temperature- and Mg^{2+}- differential probing strategies detect most of the nucleotides involved in the kissing loop (4 and 5 out of 6 nt, respectively). However the informed second round of modeling does not predict 2 bp long interactions to avoid the prediction of spurious pairings. In contrast, predicting three base pairs long range or cross interactions should be possible—it would require the prior detection of all the nucleotides involved. Comparing SHAPE data performed in different conditions or with different probes reveals not only the secondary but also some elements of the tertiary contacts. In this context, it is crucial to further understand the mechanisms behind SHAPE reactivities and the interplay between SHAPE probes, their reactivities, and RNA structures using methods at different length scales and by comparing diverse experimental data on a RNA benchmark dataset. This will allow deriving more general rules to inform 2D and 3D modeling software. In particular, the integration of this new information in an innovative theoretical framework for molecular dynamics simulations will open the route to bias three dimensional RNA structures and to better predict their structural motifs.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ncrna7040071/s1, Figure S1: Comparison of the reactivity obtained with the three probes, Figure S2: Reactivity profiles in absence of Mg^{2+}, Figure S3: Model obtained with RNAfold and RNAstructure using 1M7 reactivities as soft constraints, Figure S4: Melting profiles of the wild-type DiLCrz with error bars, Figure S5: 1M7 probing profiles at different temperature, Figure S6: Clustering nucleotides according to 1M7 probing profiles at different temperature, Figure S7: Secondary structure models obtained with IPANEMAP, Video S1: Trajectory of the region around the nucleotide G_{182}, Table S1: in this file is summarized the pairing status of the nucleotide and some differential probing (exposition (red) means the difference is positive; protection (green) means the difference is negative). We reported the type of base pairing computed in the x-ray structure (columns B to E) and along the MD simulation (columns F to J) (presence >5%) using MCAnnotate [69–71]. The HBs in italic are non-canonical. c: cys. t: trans. W: Watson-Crick side. H: Hoogsten side. B: B side, the position is between the W and H side. S: sugar side. We also detail if the HBs are formed between the O2′ and an atom in the Watson–Crick (W), Hoogsten (H) or the B side (the position is between the W and H side) or between OP and an atom in these sides [68]. PK: Pseudo-knot. KL: kissing loop.

Author Contributions: Conceptualization, B.S., G.D.B.; Data curation, G.D.B.; Formal analysis, G.D.B., E.F., B.M., S.W. and B.S.; Funding acquisition, Y.P. and B.S; Investigation, G.D.B., D.A. and E.F.; Methodology, B.S.; Supervision, B.S.; Visualization, E.F. and B.M.; Writing—original draft, E.F., B.M., S.W. and B.S.; Writing—review & editing, G.D.B., D.A. and Y.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not applicable.

References

1. Kehr, J.; Kragler, F. Long distance RNA movement. *New Phytol.* **2018**, *218*, 29–40. [CrossRef] [PubMed]
2. Flynn, R.A.; Pedram, K.; Malaker, S.A.; Batista, P.J.; Smith, B.A.H.; Johnson, A.G.; George, B.M.; Majzoub, K.; Villalta, P.W.; Carette, J.E.; et al. Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* **2021**, *184*, 3109–3124.e22. [CrossRef] [PubMed]
3. Langdon, E.M.; Qiu, Y.; Niaki, A.G.; McLaughlin, G.A.; Weidmann, C.A.; Gerbich, T.M.; Smith, J.A.; Crutchley, J.M.; Termini, C.M.; Weeks, K.M.; et al. mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **2018**, *360*, 922–927. [CrossRef] [PubMed]
4. Iserman, C.; Roden, C.A.; Boerneke, M.A.; Sealfon, R.S.G.; McLaughlin, G.A.; Jungreis, I.; Fritch, E.J.; Hou, Y.J.; Ekena, J.; Weidmann, C.A.; et al. Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* **2020**, *80*, 1078–1091.e6. [CrossRef]
5. Beaudoin, J.-D.; Novoa, E.M.; Vejnar, C.E.; Yartseva, V.; Takacs, C.M.; Kellis, M.; Giraldez, A.J. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* **2018**, *25*, 677–686. [CrossRef]
6. Mustoe, A.M.; Busan, S.; Rice, G.M.; Hajdin, C.E.; Peterson, B.K.; Ruda, V.M.; Kubica, N.; Nutiu, R.; Baryza, J.L.; Weeks, K.M. Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **2018**, *173*, 181–195.e18. [CrossRef]
7. Watts, J.M.; Dang, K.K.; Gorelick, R.J.; Leonard, C.W.; Bess, J.W., Jr.; Swanstrom, R.; Burch, C.L.; Weeks, K.M. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **2009**, *460*, 711. [CrossRef]
8. Dethoff, E.A.; Boerneke, M.A.; Gokhale, N.S.; Muhire, B.M.; Martin, D.P.; Sacco, M.T.; McFadden, M.J.; Weinstein, J.B.; Messer, W.B.; Horner, S.M.; et al. Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11513–11518. [CrossRef]
9. Koculi, E.; Cho, S.S.; Desai, R.; Thirumalai, D.; Woodson, S.A. Folding path of P5abc RNA involves direct coupling of secondary and tertiary structures. *Nucleic Acids Res.* **2012**, *40*, 8011–8020. [CrossRef]
10. Deigan, K.E.; Li, T.W.; Mathews, D.H.; Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 97–102. [CrossRef]
11. Washietl, S.; Hofacker, I.L.; Stadler, P.F.; Kellis, M. RNA folding with soft constraints: Reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Res.* **2012**, *40*, 4261–4272. [CrossRef]
12. Brunel, C.; Romby, P. Probing RNA structure and RNA-ligand complexes with chemical probes. In *Methods in Enzymology*; Elsevier BV: Amsterdam, The Netherlands, 2000; Volume 318, pp. 3–21.
13. McGinnis, J.L.; Dunkle, J.A.; Cate, J.H.D.; Weeks, K.M. The Mechanisms of RNA SHAPE Chemistry. *J. Am. Chem. Soc.* **2012**, *134*, 6617–6624. [CrossRef]
14. Frezza, E.; Courban, A.; Allouche, D.; Sargueil, B.; Pasquali, S. 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*, 108–127. [CrossRef]
15. Pinamonti, G.; Bottaro, S.; Micheletti, C.; Bussi, G. Elastic network models for RNA: A comparative assessment with molecular dynamics and SHAPE experiments. *Nucleic Acids Res.* **2015**, *43*, 7260–7269. [CrossRef]
16. Mlýnský, V.; Bussi, G. Molecular dynamics simulations reveal an interplay between SHAPE reagent binding and RNA flexibility. *J. Phys. Chem. Lett.* **2018**, *9*, 313–318. [CrossRef]
17. Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A.M.; Weissman, J.S.; Rouskin, S. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat. *Methods* **2017**, *14*, 75–82.
18. Lucks, J.B.; Mortimer, S.A.; Trapnell, C.; Luo, S.; Aviran, S.; Schroth, G.P.; Pachter, L.; Doudna, J.A.; Arkin, A.P. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11063–11068. [CrossRef]
19. Watters, K.E.; Yu, A.M.; Strobel, E.J.; Settle, A.H.; Lucks, J.B. Characterizing RNA structures in vitro and in vivo with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* **2016**, *103*, 34–48. [CrossRef]
20. Flynn, R.A.; Zhang, Q.C.; Spitale, R.C.; Lee, B.; Mumbach, M.; Chang, H.Y. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.* **2016**, *11*, 273–290. [CrossRef]

21. Ziv, O.; Gabryelska, M.M.; Lun, A.T.L.; Gebert, L.F.R.; Sheu-Gruttadauria, J.; Meredith, L.W.; Liu, Z.-Y.; Kwok, C.K.; Qin, C.-F.; MacRae, I.J.; et al. COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* **2018**, *15*, 785–788. [CrossRef]

22. Cheng, C.Y.; Chou, F.-C.; Das, R. Modeling Complex RNA Tertiary Folds with Rosetta. In *Methods in Enzymology*; Elsevier BV: Amsterdam, The Netherlands, 2015; Volume 553, pp. 35–64.

23. Smola, M.J.; Rice, G.M.; Busan, S.; Siegfried, N.A.; Weeks, K.M. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **2015**, *10*, 1643. [CrossRef]

24. Leontis, N.B.; Stombaugh, J.; Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002**, *30*, 3497–3531. [CrossRef]

25. Meyer, M.; Nielsen, H.; Olieric, V.; Roblin, P.; Johansen, S.D.; Westhof, E.; Masquida, B. Speciation of a group I intron into a lariat capping ribozyme. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7659–7664. [CrossRef]

26. Beckert, B.; Nielsen, H.; Einvik, C.; Johansen, S.D.; Westhof, E.; Masquida, B. Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes. *EMBO J.* **2008**, *27*, 667–678. [CrossRef]

27. Mathews, D.H. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* **2004**, *10*, 1178–1190. [CrossRef]

28. Mathews, D.H.; Burkard, M.E.; Freier, S.M.; Wyatt, J.R.; Turner, D.H. Predicting oligonucleotide affinity to nucleic acid targets. *RNA* **1999**, *5*, 1458–1469. [CrossRef]

29. Turner, D.H.; Mathews, D.H. NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* **2009**, *38*, D280–D282. [CrossRef]

30. Watkins, A.M.; Rangan, R.; Das, R. FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Struct.* **2020**, *28*, 963–976.e6. [CrossRef]

31. Watkins, A.M.; Geniesse, C.; Kladwang, W.; Zakrevsky, P.; Jaeger, L.; Das, R. Blind prediction of noncanonical RNA structure at atomic accuracy. *Sci. Adv.* **2018**, *4*, eaar5316. [CrossRef]

32. Boniecki, M.J.; Lach, G.; Dawson, W.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K.M.; Bujnicki, J.M. SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2016**, *44*, e63. [CrossRef]

33. Rybarczyk, A.; Szostak, N.; Antczak, M.; Zok, T.; Popenda, M.; Adamiak, R.; Blazewicz, J.; Szachniuk, M. New in silico approach to assessing RNA secondary structures with non-canonical base pairs. *BMC Bioinform.* **2015**, *16*, 276. [CrossRef] [PubMed]

34. Antczak, M.; Zablocki, M.; Żok, T.; Rybarczyk, A.; Blazewicz, J.; Szachniuk, M. RNAvista: A webserver to assess RNA secondary structures with non-canonical base pairs. *Bioinformatics* **2019**, *35*, 152–155. [CrossRef] [PubMed]

35. Parisien, M.; Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nat. Cell Biol.* **2008**, *452*, 51–55. [CrossRef] [PubMed]

36. Sloma, M.F.; Mathews, D.H. Base pair probability estimates improve the prediction accuracy of RNA non-canonical base pairs. *PLoS Comput. Biol.* **2017**, *13*, e1005827. [CrossRef]

37. Siederdissen, C.H.Z.; Bernhart, S.H.; Stadler, P.F.; Hofacker, I. A folding algorithm for extended RNA secondary structures. *Bioinformatics* **2011**, *27*, i129–i136. [CrossRef]

38. Lyngsø, R.B.; Pedersen, C.N.S. RNA Pseudoknot Prediction in Energy-Based Models. *J. Comput. Biol.* **2000**, *7*, 409–427. [CrossRef]

39. Sheikh, S.; Backofen, R.; Ponty, Y. Impact of the energy model on the complexity of RNA folding with pseudoknots. In *Human-Computer Interaction—INTERACT 2011*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2012; pp. 321–333.

40. Jabbari, H.; Wark, I.; Montemagno, C. RNA secondary structure prediction with pseudoknots: Contribution of algorithm versus energy model. *PLoS ONE* **2018**, *13*, e0194583. [CrossRef]

41. Rivas, E.; Eddy, S.R. A dynamic programming algorithm for RNA structure prediction including pseudoknots11Edited by I. Tinoco. *J. Mol. Biol.* **1999**, *285*, 2053–2068. [CrossRef]

42. Jabbari, H.; Condon, A. A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures. *BMC Bioinform.* **2014**, *15*, 147. [CrossRef]

43. Hajdin, C.E.; Bellousov, S.; Huggins, W.; Leonard, C.W.; Mathews, D.H.; Weeks, K.M. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5498–5503. [CrossRef]

44. Xayaphoummine, A.; Bucher, T.; Thalmann, F.; Isambert, H. Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15310–15315. [CrossRef]

45. Miao, Z.; Adamiak, R.W.; Blanchet, M.-F.; Boniecki, M.; Bujnicki, J.M.; Chen, S.-J.; Cheng, C.; Chojnowski, G.; Chou, F.-C.; Cordero, P.; et al. RNA-Puzzles Round II: Assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* **2015**, *21*, 1066–1084. [CrossRef]

46. Miao, Z.; Adamiak, R.W.; Antczak, M.; Batey, R.T.; Becka, A.J.; Biesiada, M.; Boniecki, M.J.; Bujnicki, J.M.; Chen, S.-J.; Cheng, C.Y.; et al. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **2017**, *23*, 655–672. [CrossRef]

47. Miao, Z.; Adamiak, R.W.; Antczak, M.; Boniecki, M.J.; Bujnicki, J.; Chen, S.-J.; Cheng, C.Y.; Cheng, Y.; Chou, F.-C.; Das, R.; et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* **2020**, *26*, 982–995. [CrossRef]

48. Deforges, J.; Chamond, N.; Sargueil, B. Structural investigation of HIV-1 genomic RNA dimerization process reveals a role for the Major Splice-site Donor stem loop. *Biochimie* **2012**, *94*, 1481–1489. [CrossRef]

49. De Bisschop, G.; Sargueil, B. RNA Footprinting Using Small Chemical Reagents. *Funct. Proteom.* **2021**, *2323*, 13–23. [CrossRef]

50. Karabiber, F.; McGinnis, J.L.; Favorov, O.V.; Weeks, K.M. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* **2012**, *19*, 63–73. [CrossRef]

51. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef]

52. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [CrossRef]

53. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [CrossRef]

54. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.R.; Smith, J.C.; Kasson, P.M.; van der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854. [CrossRef]

55. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T.E.; Laughton, C.A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **2017**, *92*, 3817–3829. [CrossRef]

56. Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R.A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; et al. RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chem. Rev.* **2018**, *118*, 4177–4338. [CrossRef]

57. Horn, H.W.; Swope, W.C.; Pitera, J.W. Characterization of the TIP4P-Ew water model: Vapor pressure and boiling point. *J. Chem. Phys.* **2005**, *123*, 194504. [CrossRef]

58. Joung, I.S.; Cheatham, T.E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041. [CrossRef]

59. Besseová, I.; Otyepka, M.; Réblová, K.; Sponer, J. Dependence of A-RNA simulations on the choice of the force field and salt strength. *Phys. Chem. Chem. Phys. PCCP* **2009**, *11*, 10701–10711. [CrossRef]

60. Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502. [CrossRef]

61. Cunha, R.A.; Bussi, G. Unraveling Mg^{2+}–RNA binding with atomistic molecular dynamics. *RNA* **2017**, *23*, 628–638. [CrossRef]

62. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

63. Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]

64. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [CrossRef]

65. Harvey, S.C.; Tan, R.K.-Z.; Cheatham, T.E. The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **1998**, *19*, 726–740. [CrossRef]

66. Berendsen, H.J.C.; Postma, J.P.M.; Van Gunsteren, W.F.; Di Nola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690. [CrossRef]

67. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [CrossRef]

68. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [CrossRef]

69. Lisi, V.; Major, F. A comparative analysis of the triloops in all high-resolution RNA structures reveals sequence structure relationships. *RNA* **2007**, *13*, 1537–1545. [CrossRef]

70. Lemieux, S.; Major, F. RNA canonical and non-canonical base pairing types: A recognition method and complete repertoire. *Nucleic Acids Res.* **2002**, *30*, 4250–4263. [CrossRef]

71. Gendron, P.; Lemieux, S.; Major, F. Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.* **2001**, *308*, 919–936. [CrossRef]

72. Leontis, N.B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499–512. [CrossRef]

73. Lu, X.-J.; Bussemaker, H.J.; Olson, W.K. DSSR: An integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **2015**, *43*, e142. [CrossRef]

74. Marzec, C.J.; Day, L.A. An exact description of five-membered ring configurations. I. Parameterization via an amplitude S, an angle gamma, the pseudorotation amplitude q and phase angle P, and the bond lengths. *J. Biomol. Struct. Dyn.* **1993**, *10*, 1091–1123. [CrossRef]

75. Westhof, E.; Sundaralingam, M. 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. *J. Am. Chem. Soc.* **1983**, *105*, 970–976. [CrossRef]

76. Lavery, R.; Moakher, M.; Maddocks, J.H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929. [CrossRef]

77. Busan, S.; Weidmann, C.A.; Sengupta, A.; Weeks, K.M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **2019**, *58*, 2655–2664. [CrossRef]

78. Rice, G.; Leonard, C.W.; Weeks, K.M. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* **2014**, *20*, 846–854. [CrossRef]

79. Butcher, S.E.; Burke, J.M. Structure-mapping of the Hairpin Ribozyme: Magnesium-dependent folding and evidence for tertiary interactions within the ribozyme-substrate complex. *J. Mol. Biol.* **1994**, *244*, 52–63. [CrossRef]

80. Horiya, S.; Li, X.; Kawai, G.; Saito, R.; Katoh, A.; Kobayashi, K.; Harada, K. RNA LEGO: Magnesium-dependent formation of specific RNA assemblies through kissing interactions. *Chem. Biol.* **2003**, *10*, 645–654. [CrossRef]

81. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic Acid–Ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813–841. [CrossRef]

82. Tan, Z.-J.; Chen, S.-J. Importance of diffuse metal ion binding to RNA. *Met. Ions Life Sci.* **2011**, *9*, 101–124.

83. Lorenz, R.; Bernhart, S.H.F.; Zu Siederdissen, C.H.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [CrossRef]

84. Mathews, D.H. RNA Secondary structure analysis using RNA structure. *Curr. Protoc. Bioinform.* **2006**, *13*, 12.6.1–12.6.14. [CrossRef]

85. Silvers, R.; Keller, H.; Schwalbe, H.; Hengesbach, M. Differential scanning fluorimetry for monitoring RNA stability. *Chem. BioChem.* **2015**, *16*, 1109–1114. [CrossRef]

86. Nielsen, H.; Westhof, E.; Johansen, S. An mRNA Is Capped by a 2′, 5′ Lariat Catalyzed by a Group I-Like Ribozyme. *Science* **2005**, *309*, 1584–1587. [CrossRef]

87. Saaidi, A.; Allouche, D.; Regnier, M.; Sargueil, B.; Ponty, Y. IPANEMAP: Integrative probing analysis of nucleic acids empowered by multiple accessibility profiles. *Nucleic Acids Res.* **2020**, *48*, 8276–8289. [CrossRef]

88. Deforges, J.; de Breyne, S.; Ameur, M.; Ulryck, N.; Chamond, N.; Saaidi, A.; Ponty, Y.; Ohlmann, T.; Sargueil, B. Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. *Nucleic Acids Res.* **2017**, *45*, 7382–7400. [CrossRef]

89. Banerjee, A.R.; Jaeger, J.A.; Turner, D.H. Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. *Biochemistry* **1993**, *32*, 153–163. [CrossRef]

90. Brion, P.; Michel, F.; Renée, S.; Westhof, E. Analysis of the cooperative thermal unfolding of the td intron of bacteriophage T4. *Nucleic Acids Res.* **1999**, *27*, 2494–2502. [CrossRef]

91. Jaeger, L.; Westhof, E.; Michel, F. Monitoring of the cooperative unfolding of the sunY Group I intron of bacteriophage t4: The active form of the suny ribozyme is stabilized by multiple interactions with 3′ terminal intron components. *J. Mol. Biol.* **1993**, *234*, 331–346. [CrossRef]

92. Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. RNA SHAPE Chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNAAsp transcripts. *J. Am. Chem. Soc.* **2005**, *127*, 4659–4667. [CrossRef]

93. Strulson, C.A.; Boyer, J.A.; Whitman, E.E.; Bevilacqua, P.C. Molecular crowders and cosolutes promote folding cooperativity of RNA under physiological ionic conditions. *RNA* **2014**, *20*, 331–347. [CrossRef]

94. Choi, S.; Park, C.; Kim, K.E.; Kim, K.K. An in vitro technique to identify the RNA binding-site sequences for RNA-binding proteins. *BioTechniques* **2017**, *63*, 28–33. [CrossRef] [PubMed]

95. Gasser, C.; Gebetsberger, J.; Gebetsberger, M.; Micura, R. SHAPE probing pictures Mg^{2+}-dependent folding of small self-cleaving ribozymes. *Nucleic Acids Res.* **2018**, *46*, 6983–6995. [CrossRef] [PubMed]

96. Steen, K.-A.; Rice, G.M.; Weeks, K.M. Fingerprinting Noncanonical and Tertiary RNA Structures by Differential SHAPE Reactivity. *J. Am. Chem. Soc.* **2012**, *134*, 13160–13163. [CrossRef] [PubMed]