Structure
Article

Developing Three-Dimensional Models of Putative-Folding Intermediates of the HDV Ribozyme

Cédric Reymond,[1] Dominique Lévesque,[1] Martin Bisaillon,[1] and Jean-Pierre Perreault[1,*]
[1]RNA Group/Groupe ARN, Département de Biochimie, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Québec J1H 5N4, Canada
*Correspondence: jean-pierre.perreault@usherbrooke.ca
DOI 10.1016/j.str.2010.09.024

SUMMARY

Both the role and the interacting partners of an RNA molecule can change depending on its tertiary structure. Consequently, it is important to be able to accurately predict the complete folding pathway of an RNA molecule. The hepatitis delta virus (HDV) ribozyme is a small catalytic RNA with the greatest number of folding intermediates making it the model of choice with which to address this problem. The tertiary structures of the known putative intermediates along the folding pathway of the HDV ribozyme were predicted using the Macromolecular Conformations Symbolic programming (MC-Sym) software. The structures obtained by this method received physical support from Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE). The analysis of these structures elucidated several features of the HDV ribozyme. In addition, this report represents an application for MC-Sym that permits progression one step further toward the computer prediction of an RNA molecule-folding pathway.

INTRODUCTION

Understanding the RNA structure-function relationship is crucial in molecular biology, and the ability to accurately predict both the folding pathway and the active tertiary structure of RNA molecules from their primary sequences could have a major impact on human health. In the absence of specific chaperones, an RNA molecule folds sequentially from 5′ to 3′ and possesses a hierarchical structure in which the primary sequence determines the secondary structure that, in turn, determines the tertiary folding in a process that only minimally alters the secondary structure (Onoa and Tinoco, 2004; Zhuang et al., 2007; Greenleaf et al., 2008). The RNA-folding pathway is typically enthalpy driven, and the formation of numerous tertiary interactions appears to be crucial in obtaining increasingly stable intermediates. However, our understanding of how these interactions take place remains scanty (Moody and Bevilacqua, 2003; Reymond et al., 2009; Fiore et al., 2009). Due to the energies involved in the formation of RNA helices, the secondary structure is formed significantly faster than the tertiary structure (Sosnick and Pan, 2004). During the formation of the tertiary structure, the partially folded RNA molecules explore many structures with approximately similar energies until they find a transition structure that will allow them to proceed along the folding pathway (Cruz and Westhof, 2009). There are some exceptions to this rule where the tertiary interactions can induce a secondary structure rearrangement, but even so, the resulting structure possesses a similar energy. In general the overall shape of the molecule is conserved, and this kind of rearrangement usually leads to further stabilization through the formation of new tertiary interactions (Baird et al., 2005; Chauhan and Woodson, 2008).

Small catalytic RNAs are the models of choice with which to study RNA folding because modifications of their tertiary structure are reflected in their catalytic properties (Lilley, 2005; Cochrane and Strobel, 2008). This phenomenon is well illustrated by the shortened antigenomic *cis*-acting hepatitis delta virus (HDV) RNA motif and by the corresponding *trans*-acting ribozyme, both of which possess the most complex folding pathway elucidated to date among the catalytic RNAs (Reymond et al., 2009). This folding pathway is homogeneous and avoids most of the problems associated with folding into alternative, nonfunctional structures (Reymond et al., 2009). Structure-function assays and structural studies have resulted in the elucidation of the double-pseudoknot secondary structure of this catalytic RNA (Figure 1) (Been, 2006). This structure is composed of two stems (stem I, which permits binding of the substrate to the ribozyme in the *trans*-acting version; and stem II, which forms a pseudoknot in the *cis*-acting version), two stem-loops (III and IV), and three single-stranded junctions (I/II, I/IV, and IV/II). Both the junction I/IV and the loop III are single stranded in the initial stages of folding but are subsequently involved in the formation of the pseudoknot I.I. Both X-ray diffraction and NMR spectroscopy data have provided good pictures of the architecture of the catalytic core in both the pre- and the post-cleavage complexes. The stem I, the pseudoknot I.I, the stem-loop III, and the junction IV/II form the catalytic center, whereas the stems II and IV stabilize the structure. Previous experiments demonstrated that, after the formation of the secondary structure (Rz^1), five conformational steps have been reported to occur along the folding pathway (Figure 2A) (Reymond et al., 2009). The first conformational step is the docking of stem I within the catalytic core, which results in the production of the second stable intermediate (Rz^2). The second conformational step is

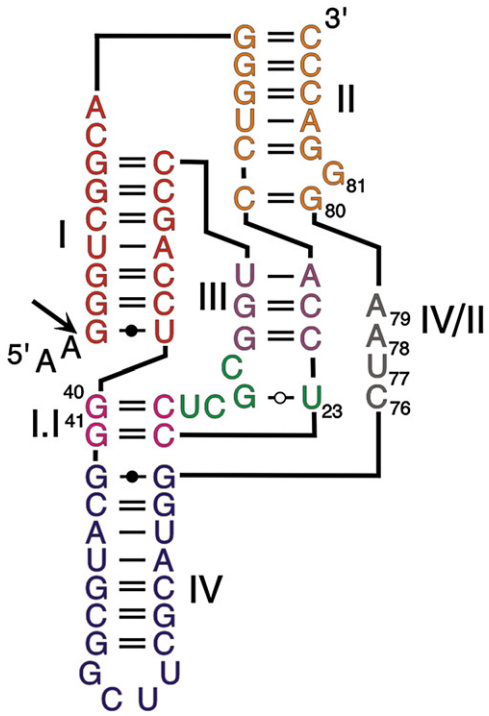

Figure 1. Secondary Structure of a Shortened *cis*-Acting Version of the Wild-Type Antigenomic HDV Ribozyme

Roman numerals indicate the stems (I, III, IV), the junction (IV/II), and the pseudoknots (II, I.I). Arabic numerals indicate the positions of the important nucleotides. The arrow indicates the cleavage site, and the 5′ leaving product is represented in black letters.

the formation of the pseudoknot I.I that includes two GC base pairs, which results in the production of the third stable intermediate (Rz^3). The third conformational step is the formation of the A-minor motif between the two consecutive adenosines of the junction IV/II and the minor groove of the stem III, more specifically with the two GC base pairs of the latter, which results in the production of the fourth stable intermediate (Rz^4). The fourth conformational step is a base pair switch at the bottom of stem II, which results in the production of the fifth intermediate (Rz^5). Finally, the fifth conformational step is the formation of the trefoil turn motif involving the catalytically active cytosine and its two flanking nucleotides, which results in the production of the catalytically active structure (Rz^6) (Figure 2B) (Reymond et al., 2009). All of these conformational steps lead to the in-line geometry required for the chemical step involving the catalytically active cytosine (C76). When the transesterification reaction occurs, the cleavage of the phosphodiester backbone is simultaneously accompanied by the product release.

In order to fully understand an RNA-folding pathway, it is preferable to visualize it in three dimensions. However, using the current computational tools, molecular dynamic simulations are not possible due to the calculation time involved for simulations covering the RNA-folding timescale, unless drastic approximations are made in the structure (Ding et al., 2008). It is for this reason that bioinformatic RNA tertiary structure prediction tools have been developed (e.g., Das and Baker, 2007; Parisien and Major, 2008; Shapiro et al., 2007). These prediction tools use many different approaches to build tertiary structures; however, for a highly structured RNA molecule of the length of the HDV motif, the only approach to date that has yielded interesting results was Macromolecular Conformation Symbolic programming (MC-Sym) (Parisien and Major, 2008). This program uses cyclic building blocks extracted from crystallographic data to solve a constraint satisfaction problem based on the secondary structure. The input is the nucleotide sequence and the secondary structure, whereas the output is a pool of the different tertiary structures that fulfill both the secondary structure constraints and take into account the steric repulsions. This pool of structures contains, at least theoretically, all of the potential tertiary structures that can be adopted by an RNA molecule. With the goal of visualizing the folding pathway of the HDV *cis*-acting RNA motif, we have undertaken the prediction of the tertiary structure of each isolated intermediate along this folding pathway. It is important to mention that these steps were isolated in conditions of equilibrium, and they may not be formed when RNA folding occurs cotranscriptionally. The predicted structures for each intermediate received physical support from chemical probing using the Selective 2′-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) method (Mortimer and Weeks, 2008; Watts et al., 2009). Detailed analysis of the predicted structures revealed several interesting details of HDV structure and permitted the generation of a new working hypothesis of the folding pathway.

RESULTS AND DISCUSSION

Building a Three-Dimensional Representation of the Different Intermediates

In order to model the different known intermediates of the *cis*-acting self-cleaving HDV RNA motif using MC-Sym, a standard building order of the catalytic RNA motif and harmonized variables were used. The building order represents the sequence in which the blocks are assembled in order to build the tertiary structures. The harmonized variables are defining important parameters such as the method used to build these structures, how the blocks are chosen, or how to deal with steric clashes. Preliminary scripts were run in order to test the building order and to optimize all of the variables so as to be finally able to obtain structures for each intermediate along the folding pathway (see Supplemental Experimental Procedures available online). A starting pool of structures for the HDV motif was generated by using only the nucleotide sequence and the secondary structure. This pool contains structures from the entire folding landscape of the *cis*-acting HDV motif. The subsequent addition of tertiary interaction constraints, known from previous biochemical studies, permitted the selection of populations of structures for the intermediates from this starting pool (for a list of constraints used for each tertiary interaction, see Supplemental Experimental Procedures). Specifically, the tertiary constraints sufficient to define an intermediate in the MC-Sym software are those that correspond to the interactions that are disrupted in the mutants previously used in the analysis of the folding pathway (Reymond et al., 2009) (Figure 2B). In addition the distances between two nucleotides shown to be in close proximity by UV-induced covalent bonds in previous crosslinking

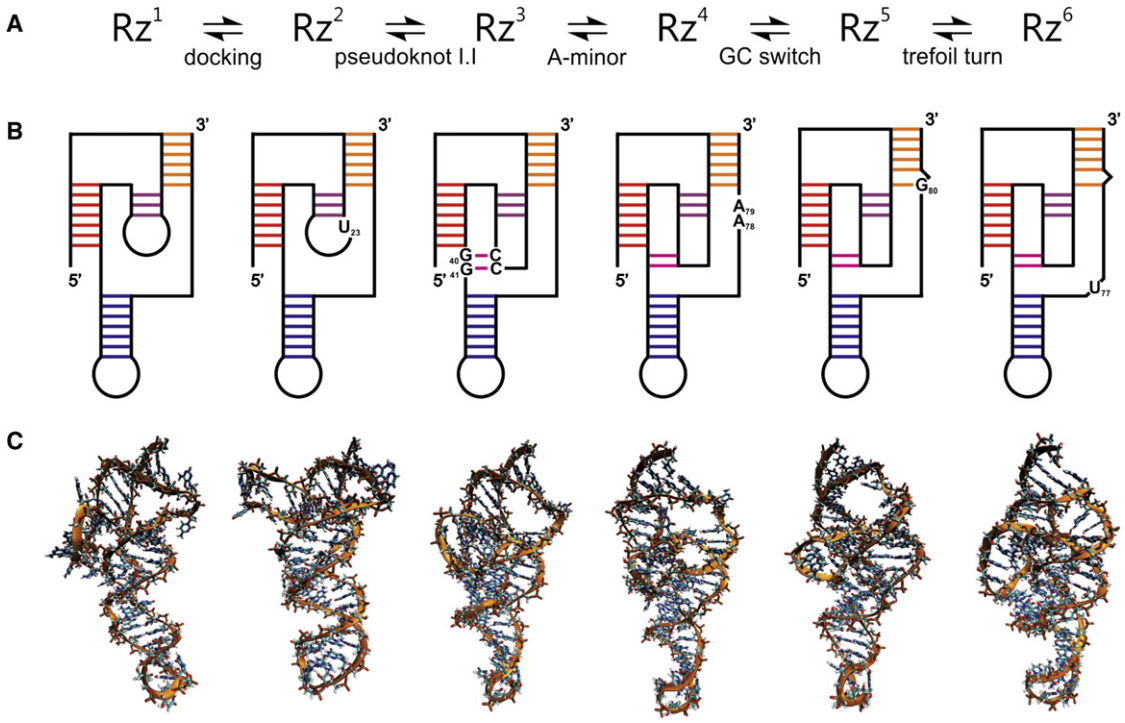

Figure 2. Folding Pathway of HDV Ribozyme

(A) Proposed folding pathway of HDV ribozyme, with the name of the conformational change between each intermediate indicated, yielding the fully folded structure (according to Reymond et al., 2009).

(B) Secondary structure representation of each intermediate with the important nucleotides indicated by their abbreviations and positions.

(C) Representative example of the main family of tertiary structures obtained with MC-Sym for each intermediate.

studies were found to be present in these predicted structures (Ouellet and Perreault, 2004; Reymond et al., 2007). After minimization, all of the putative structures obtained for a given intermediate were visualized using the Visual Molecular Dynamics (VMD) software (Humphrey et al., 1996), and the most representative structure was selected based on both its stability and the average root-mean-square deviation (Figure 2C). It is noteworthy that all of the HDV ribozyme regions are required in order to obtain relevant structures and that the removal of the 5′ end nucleotides corresponding to the leaving product yielded structures with a slightly different global shape (data not shown). This observation is in agreement with the fact that the crystal structures of the pre- and post-cleaved ribozymes have been shown to be different (Ke et al., 2004).

The next important observation was the fact that the global shape of the *cis*-acting HDV motif is primarily defined by the secondary structure constraints. This is in agreement with a previously reported isothermal titration study of the HDV ribozyme showing that the Gibbs free-energy stabilization was relatively small for the tertiary structure, as compared to that of the secondary one, in both the *trans*- and the *cis*-acting versions of HDV ribozyme (Reymond et al., 2009). Prior to the formation of the pseudoknot I.I, the junction I/II and the loop III were already located in relatively close proximity to one another (Figure 3A). After the pseudoknot I.I formation, the geometry of the coaxial helices of the HDV ribozyme was fixed, and the 5′ extremity was docked inside the catalytic core in over 95% of the structures obtained (Figure 3B). This position was in agreement with the data from a previous crosslinking analysis (Reymond et al., 2007). In the remaining structures, the 5′ region was generally positioned outside of the catalytic core and, in a few exceptional

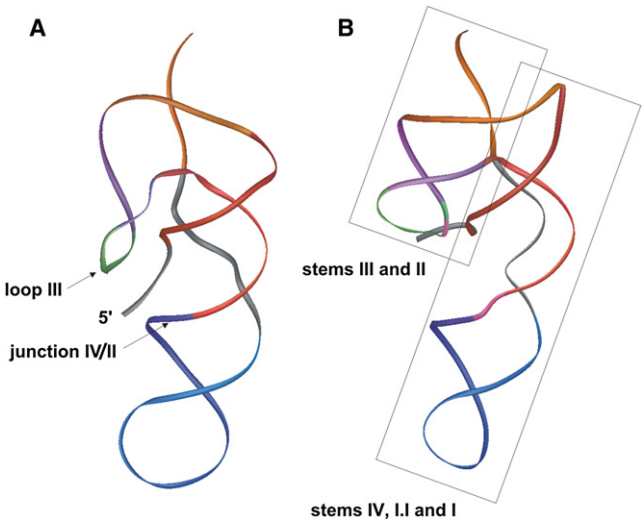

Figure 3. Global Features of the MC-Sym Generated HDV Structures

(A) Example of a prefolded pseudoknot I.I (the two regions involved in the pseudoknot I.I formation are indicated by the arrows).

(B) Example of the double-coaxial helical geometry found with the constraints used for pseudoknot I.I (the coaxial helices are boxed with thin straight lines).

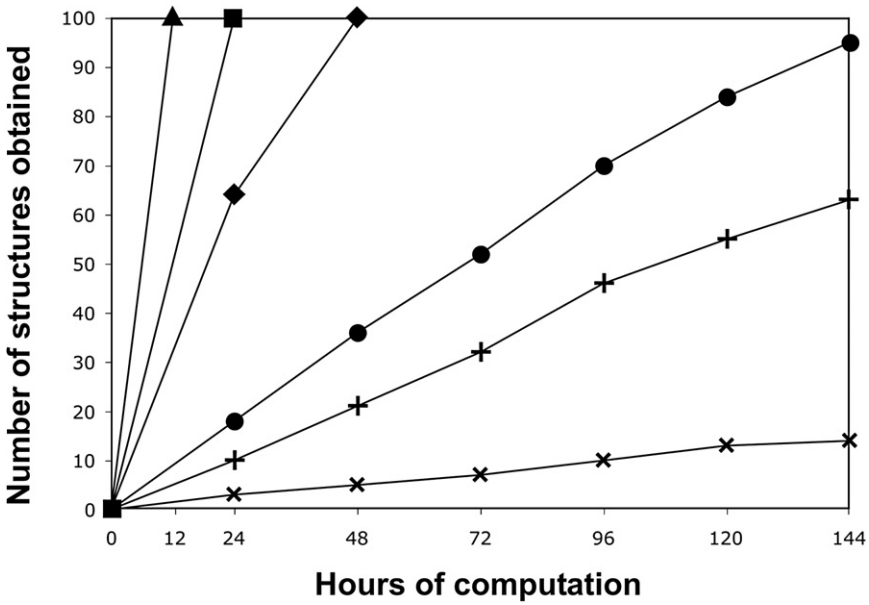

Figure 4. Graphical Representation of the Number of Structures Obtained as a Function of the Computation Time for the Different Scripts

Each script contains all of the constraints required to form an intermediate along the folding pathway (▲ = Rz^1, ■ = Rz^2, ◆ = Rz^3, + = Rz^4, ● = Rz^5, × = Rz^6).

cases, positioned through the formation of a kink turn located at the scissile phosphate bond. These exceptional cases are in agreement with the initial crystallographic data obtained for a genomic version of the HDV ribozyme (Ferré-D'Amaré et al., 1998).

The current model of the HDV-folding pathway is based on the generation of an increasingly stable and complex structure through the formation of tertiary interactions. This model was verified by plotting the number of structures generated by MC-Sym as a function of the calculation time for each intermediate. Preliminary results showed that there is no direct correlation between the number of constraints in the different scripts and the number of structures obtained per amount of time. Moreover, these preliminary results revealed that a single constraint can have a dramatic effect on the structure output. Because the constraints used in the scripts have been proven to be present in the structure of the intermediates in solution, the output has to contain structures representative of the intermediates. This approach takes advantage of the fact that the algorithm used by MC-Sym to build the 3D structures does not mimic RNA folding but rather gives an approximation of the complexity of an RNA structure without having to go through the folding pathway, which would effectively be a measure of the complexity of the transition states. The general trend observed is that the further advanced an intermediate is along the folding pathway, the more its structure is rigid and the lesser the number of structures that are retrieved per amount of calculation time (Figure 4). One noticeable exception is the inversion that occurs between the intermediates Rz^4 and Rz^5. In both of these intermediates the A-minor motif is formed, whereas the only difference is either the absence or the presence of the GC base pair switch. The conclusion drawn is that the A-minor motif is easier to form if the GC base pair switch is present, thus yielding a greater number of structures per amount of calculation time. This observation supports the previously published results that suggest the A-minor motif introduces backbone tension in the structure and that the GC base pair switch is involved in backbone relaxation (Nehdi et al., 2007). In addition this observation is very interesting because it shows that synergy can be found between constraints. It also supports the idea that tertiary constraints can locally change the secondary structure and that the alteration of the latter can be predicted by computational approaches.

Finally, we noted that over defining regions through additional tertiary constraints often leads to an absence of results. A good example of this problem was the interactions between stem III and junction IV/II. In the crystal structure of the genomic HDV ribozyme, the two adenosines of the latter junction interact with the minor groove of stem III to form two A-minor motifs of different geometries (Ke et al., 2004). Application of the script that tried to reproduce the geometry found in the crystal structure to the antigenomic HDV ribozyme through distance constraints did not produce any results. Removal of some of these constraints, and adjusting the distances, led to structures containing two A-minor motifs with the correct geometries. This approach raises the question of the objectivity behind the constraint choices, not only for MC-Sym, but also more generally for any program solving constraint satisfaction problems. In order to answer this question, the specific distance constraints were replaced by pairing constraints that permitted the exploration of all of the possible geometries and distances that support a given interaction. The geometries of the A-minor motifs in the resulting structures were very similar to those of both the crystal and the scripts, based on specific distance constraints. Thus, using pairing constraints is the best approach with which to introduce tertiary constraints in a structure, and it avoids, as much as possible, the subjectivity inherent in human scripting.

Validation of the 3D Representation

Although the MC-Sym building blocks are extracted from crystal structures, and the constraints required are known from previous biochemical studies, the tertiary structures obtained by this program remain predictions. As such, they have to be validated, or at least receive indirect physical support. The correlations between the MC-Sym predicted tertiary structures and the tertiary structure of the putative intermediates in solution at equilibrium were assessed by the Selective 2′-Hydroxyl Acylation reactions analyzed by Primer Extension (SHAPE) method (Mortimer and Weeks, 2008). The SHAPE reagents acylate 2′-hydroxyl functional groups based on the local flexibility of

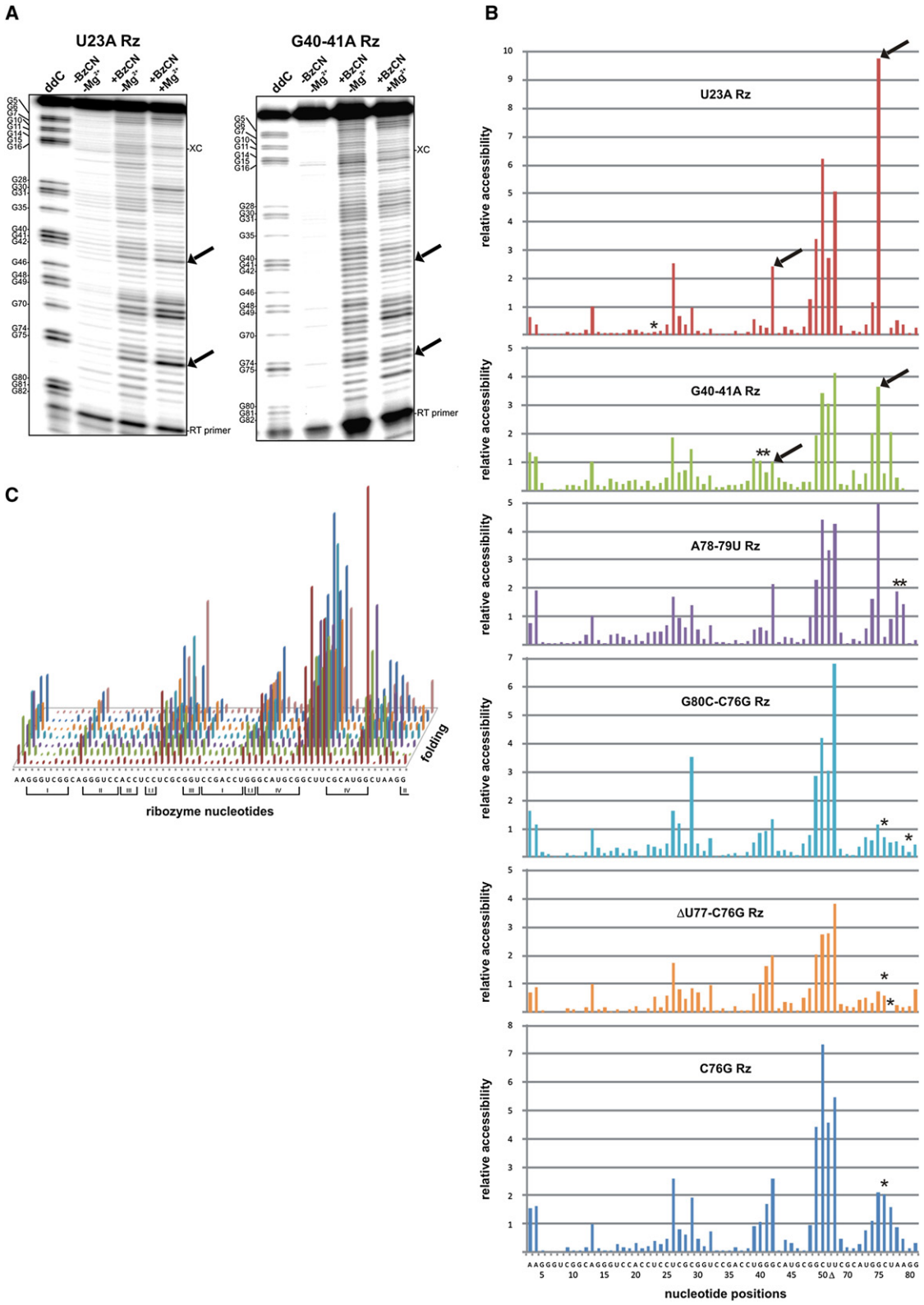

Figure 5. HDV Ribozyme Structure Probing Experiments

(A) SHAPE gels for the mutants U23A and G40A/G41A, respectively, of the Rz^1 and the Rz^2 intermediates of the folding pathway. In both cases the first lane is a guanosine ladder, the second is a negative control, the third is the SHAPE reaction performed in the absence of magnesium, and the last is the SHAPE reaction performed in the presence of magnesium. The bands corresponding to the homopurine G42-G75 base pair located at the top of stem IV are indicated by the

the backbone. This flexibility differs depending upon whether or not the nucleotide in question is accessible. In other words this method permits the elucidation of the secondary structure of an RNA molecule. The reaction is almost instantaneous; therefore, it avoids conformational changes that could lead to misfolded HDV ribozyme structures. The acylation reactions were performed at 37°C, which is the temperature of the HDV ribozyme natural environment. Because RNA polymerization stops upon reaching a nucleotide with an acylated 2′-hydroxyl group, the modified positions were determined by reverse transcription using a radioactive primer.

As an example, the results for the mutant U23A (intermediate Rz^1) and G40A/G41A (intermediate Rz^2) of the *cis*-acting HDV ribozyme are presented in Figure 5A. These two consecutive steps, which occur at the beginning of the folding pathway, are only separated by the docking of stem I. Semi-automated footprinting analysis (SAFA) software was used to extract the information contained in the gels and yielded the corresponding band intensities (Figure 5B; Figure S1) (Laederach et al., 2008). Considering the decreases in the band intensities between those two mutants, the docking of stem I appears to contribute to the stabilization of the homopurine base pair. SHAPE probings were then performed (in triplicate) on all of the previously identified mutants along the HDV-folding pathway (Figure 5C). These experiments confirmed the overall secondary structure of the various mutants and proved that the mutants are properly folded in solution. They also showed that both the targeted interactions and the local structures around the mutations are disrupted with the exception of the mutant U23A. This explains how the folding pathway is interrupted in the different mutants. Globally, the intensities of the bands decrease while the ribozyme progresses along the folding pathway and forms the catalytic core tertiary network, supporting the idea that the secondary structure becomes increasingly stable and less accessible. This implies an enthalpically driven folding, in agreement with previous experiments performed on the HDV ribozyme (Reymond et al., 2009). At the tertiary interaction level, all of the known features of HDV can be seen by SHAPE probing; however, any new interaction was revealed by this approach (Figure 5B).

Interesting correlations between the SHAPE data and the MC-Sym predictions were observed. Most of the structures generated by MC-Sym using only the secondary structure information are prefolded, with the substrate located inside the catalytic core and the pseudoknot I.I almost formed. However, when performing the probing with a *trans*-acting ribozyme in the absence of its cognate substrate, the strand I involved in the substrate recognition appears highly accessible to acylation (Supplemental Experimental Procedures). Interestingly, the C9 and the A36 involved in base pairs located in the middle of the stem I (i.e., C9-G35 and U8-A36) tend to become more accessible along the folding pathway. This observation is in agreement with enzymatic probing showing the same behavior of the stem I (Ouellet and Perreault, 2004). In all of the SHAPE experiments, the bands corresponding to both the pseudoknot I.I and the homopurine located at the top of stem IV have strong intensities, suggesting that these base pairs are not always formed and that this region is probably highly dynamic. These data support the idea that the main role of both the pseudoknot I.I and the homopurine base pair is likely to stabilize the stacking of stem I on stem IV and would explain why mutants possessing only one base pair in the pseudoknot I.I retain some level of catalytic activity (Nishikawa and Nishikawa, 2000; Deschênes et al., 2003). This weak stabilization brought about by the pseudoknot I.I and the entropic loss involved by fixing the loop III position have been proposed to explain why this conformational change is the rate-limiting step of the folding pathway (Reymond et al., 2009). The two last steps of the folding pathway are the base pair switch at the bottom of stem II and the formation of the trefoil turn. Both of which involve extruding a nucleotide outside of the catalytic center, thereby relaxing the backbone of junction IV/II (Nehdi et al., 2007). Both of these extruded nucleotides (G81 and U77, respectively) are easily detected by both SHAPE experiments and in-line probing assays (Figure 5C) (Nehdi et al., 2007). It is interesting to note that in the structures obtained by MC-Sym, both of these nucleotides are bulged out without additional constraints if the A-minor motif is formed (Figures 6A and 6B) and are located inside if it is not. This provides a good indication that both the trefoil turn and the base pair switch are the consequences of the formation of the A-minor motif. However, bulging out U77 is not sufficient to both fully form the trefoil turn and correctly position the catalytic cytosine. Some structure predictions generated by MC-Sym pointed out that this positioning seems to happen through stacking interactions with A78 inside of the catalytic core (Figure 6A). This hypothesis still needs to receive physical support, but if it is indeed true, it may explain why C76 is less accessible during probing toward the end of the folding pathway. Finally, upon inspecting the A-minor motif region in the SHAPE experiments, it is evident that A79 is less accessible than is A78. This observation correlates with the MC-Sym structures showing that A79 prefers a type I A-minor interaction, whereas A78 prefers a type II A-minor interaction (Figures 6C and 6D), as is found in the crystal structure of the genomic HDV ribozyme (Ke et al., 2004). Overall, the predicted structures generated by MC-Sym received strong support from the SHAPE data, suggesting that these predicted structures most likely occur in solution.

Building an HDV-Folding Landscape

Based on these results, and the fact that the starting pool of structures from the output of MC-Sym theoretically contains all of the intermediates' structures, transition structures, and potential tertiary interactions, the next logical step was to use MC-Sym

arrows. The offset of one nucleotide between the guanosine ladder and the other lanes is due to the presence of the acylated nucleotides that block the reverse-transcription reaction.

(B) Quantification of the SHAPE gels for the different mutants using the SAFA software (Laederach et al., 2008). The HDV ribozyme sequence and nucleotide positions corresponding to Figure 1 are displayed on the x axis, and the relative accessibility for each position is depicted. The arrows are pointing to the quantification of the bands corresponding to the homopurine G42-G75 base pair from (A), and the asterisks indicate the position of the mutations.

(C) Average of SHAPE-probing triplicates for all of the different intermediates. The sequence is displayed on the x axis, the relative accessibility on the y axis, and the folding pathway on the z axis. Roman numbers indicate the stems (I, III, IV) and the pseudoknots (II, I.I). See also Figure S1.

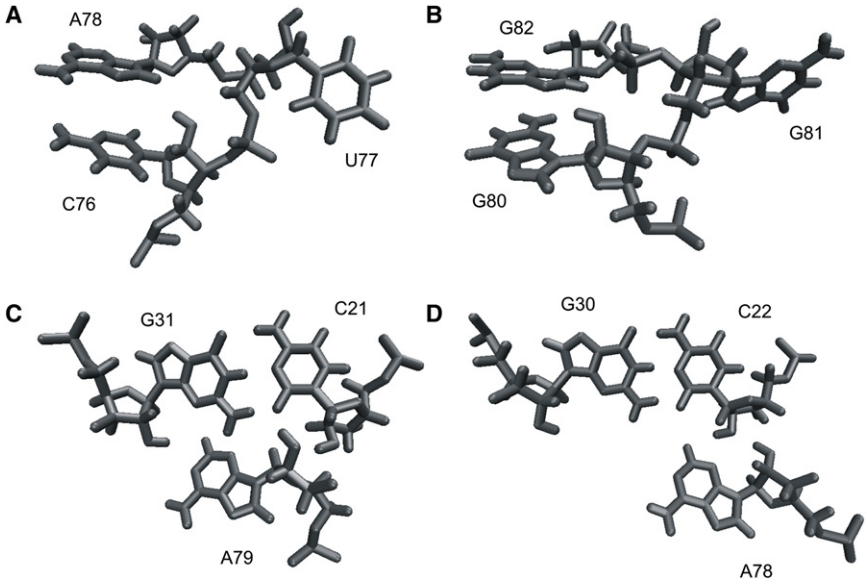

Figure 6. Specific Features of the MC-Sym Generated HDV Structures

(A) Example of a trefoil turn showing U77 bulged out and C76 stacked under the A78 that is involved in an A-minor motif.

(B) Example of a GC switch showing G81 bulged out and G80 stacked under G82 in order to shorten the junction IV/II.

(C) Example of A79 involved in a type I A-minor motif.

(D) Example of A78 involved in a type II A-minor motif.

to build a complete RNA-folding landscape. The purpose of this folding landscape is the prediction of the folding pathway based on the progression of the tertiary structure. In order to build this folding landscape, all of the tertiary structures have to be compared and classified. The first method that comes to mind in order to achieve this is to calculate the root-mean-square deviation between all of the tertiary structures in order to obtain all of the possible folding pathways. The Gibbs free energy is then calculated for each structure in order to measure the evolution of the energy along each folding pathway. The folding pathway with the lowest energy intermediates and transition states will naturally occur most of the time. The problem with this method is that the tertiary structures produced by MC-Sym possess very high Gibbs free energies due to the manner in which the structures are built and how the hydrogen atoms are added afterward. When a few minimization steps are performed in order to relax these structures, the high-energy intermediates and transition states are lost, even with the use of carefully chosen force fields and specific constraints to conserve the secondary structure. This implies that the structures deduced from the MC-Sym output cannot be used directly and that a significant portion of the folding landscape is lost when the minimization is performed. This is not an issue for the other applications of MC-Sym, but in this specific case all of the structures are equally important in order to obtain the complete folding landscape. In order to overcome this problem, the MC-Annotate program has been used to generate a symbolic representation of each tertiary structure. These symbolic representations are files listing all of the secondary and tertiary interactions present in the tertiary structures and have been introduced in a distance matrix used to calculate the Matthews correlation coefficient between each interaction network. Dendogram representations of hierarchical clustering based on the value of this coefficient have been used to classify the interaction networks. Because RNA folding is enthalpically driven (Moody and Bevilacqua, 2003; Reymond et al., 2009; Fiore et al., 2009), the hypothesis used to build the in silico-folding pathway of the HDV ribozyme is that its interaction network is evolving toward complexity and stability. Sadly, the differences between the clusters consensus interaction networks were insufficient to permit the drawing of a clear folding pathway. The clusters are built based on the Matthews correlation coefficient, which is only one number; therefore, two unrelated groups of structures can have the same coefficient. This explains why the clusters consensus interaction networks are rather small. In order to really classify these structures into a coherent landscape, more complex methods, such as using graph theory to calculate the distances between interaction maps, will have to be used (which goes beyond the scope of this work). Nevertheless, the use of the symbolic representations instead of the full tertiary structure presents a very powerful approach. The Matthews correlation itself does not permit the classification of the structures derived from the output of MC-Sym, but it is not far from achieving this goal.

Conclusion

This work demonstrates that it is possible to obtain a 3D structure for all of the putative intermediates of a complex folding pathway with MC-Sym using only the basic secondary structure and carefully selected additional tertiary constraints previously known from biochemical experiments. MC-Sym appears to be an interesting tool with which to illustrate a folding pathway. This represents a new application for this kind of software. The probing results confirmed the predicted structures of the putative intermediates along the folding pathway. This work brings new insights into the folding of the HDV ribozyme. More specifically, the structure is prefolded, the stacking interactions in the coaxial helices are crucial not only for the stability of the pseudoknot but probably for the positioning of the catalytically active cytosine as well, and the two final steps of the folding pathway occur as the result of the formation of the A-minor motif. This work also demonstrates that it is possible to propose secondary structure changes based on their ability to generate tertiary structures and that synergy can happen between constraints. Finally, the attempts to predict HDV-folding landscape showed some very interesting interactions, conformations, and distance distributions in the clusters consensus interaction networks. The HDV ribozyme, with its complex folding pathway, is an ideal model with which to study RNA folding, and hopefully the approach presented in this work will be applied to other, less

well-understood RNA molecules. At present, a view of RNA as a modular structure constructed from a combination of building blocks and tertiary linkers is beginning to emerge. Therefore, it is tempting to propose that MC-Sym may constitute an approach to provide a qualitative representation of the folding of nascent RNA during transcription if it is used in order to model building blocks that will be then assembled to give putative structures of longer molecules.

EXPERIMENTAL PROCEDURES

MC-Sym

All of the information required to generate, edit, and optimize the scripts used in MC-Sym can be found on the MC-Pipeline web page (http://www.major.iric.ca/MC-Pipeline). The three basic templates used to fold HDV ribozyme can be found in the Supplemental Experimental Procedures (one template for the secondary structure, one for the secondary structure including the pseudoknot I.I, and one for the secondary structure including both the pseudoknot I.I and the GC base pair switch). In all of these templates, the specific order used to build the tertiary structures started with stem I (from the bottom to the top), stem III (from the top to the bottom), loop III, and the pseudoknot I.I (if present), followed by the bottom of stem II, junction IV/ II, the top of stem IV, junction I/IV (if the pseudoknot I.I is not present), stem II (from the bottom to the top), junction I/II, stem IV (from the top to the loop), and finally the 5′ end of the ribozyme. At each position, 25% of the cyclic building blocks were tried, the backtrack was 25% of the structure, and the method was probabilistic. The maximum number of structures was fixed to 100, the computation time to 144 hr, and the minimal difference between two structures to 1 Å. The energies of the structures were directly minimized on the web server until the root-mean-square gradient (G_{RMS}) was <5 kcal/mol/Å. Upon visual inspection, a portion of the structures obtained for each intermediate was manually discarded because they were irrelevant in the context of RNA transcription.

Ribozyme DNA Templates

DNA template production for each ribozyme was achieved by using two overlapping DNA oligonucleotides: a forward oligonucleotide containing the T7 promoter sequence in order to allow for RNA transcription, and a reverse oligonucleotide that introduces extra sequence in the 3′ region of the ribozyme in order to be able to anneal a DNA oligonucleotide during reverse transcription (5′-GTTTGTTTGTTTGTTGAGG-3′). The filling reaction was performed in a final volume of 100 μl containing 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 2 μM of each dNTP, 1 μM of each DNA oligonucleotide, and 5 U of Pwo DNA polymerase (Roche Diagnostics). The reaction products were ethanol precipitated, washed, and the resulting DNA pellets dissolved in 56 μl of deionized water.

RNA Synthesis

RNA transcriptions were performed as described previously (Reymond et al., 2009). The dissolved DNA pellets (56 μl) were used in 100 μl transcription reactions containing 80 mM HEPES-KOH (pH 7.5), 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 5 mM of each rNTP, 0.01 U of pyrophosphatase (Roche Diagnostics), 20 U of RNaseOUT RNase inhibitor (Invitrogen), and 10 μg of purified T7 RNA polymerase, and the reactions were allowed to proceed for 3 hr at 37°C. The reactions were then treated with 4 U of RQ1 DNase (Promega), phenol-chloroform extracted, the RNA ethanol was precipitated, washed, and finally dissolved in 40 μl of deionized water. One volume of loading buffer (97.5% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol, 10 mM EDTA) was added, and the samples were then fractionated on 8% denaturing (8 M urea) polyacrylamide gels (PAGE, 19:1 ratio of acrylamide to bisacrylamide) using 45 mM Tris-borate (pH 7.5) and 1 mM EDTA solution as running buffer. The RNA was visualized by UV shadowing, and the gel slices were cut out and eluted overnight using 500 mM ammonium acetate, 1 mM EDTA, and 0.1% SDS solution. After ethanol precipitation, the RNA transcripts were dissolved in deionized water and quantified by spectrometry at 260 nm.

SHAPE Probing

The selective 2′-hydroxyl acylation reaction has been described previously (Mortimer and Weeks, 2008). Briefly, 5 pmol of RNA in 9 μl of water was heated at 65°C for 2 min, put on ice for 5 min, and then 1 μl of a solution containing 500 mM Tris-HCl (pH 7.5), either without or with 100 mM $MgCl_2$, was added, and the reactions were incubated at 37°C for 15 min. In order to start the SHAPE reactions, either 1 μl of benzoyl cyanide (Sigma) (600 mM in DMSO) or 1 μl of DMSO (negative control) was added to the RNA solution. Glycogen (20 μg; Roche Diagnostics) was then added and the RNA ethanol precipitated, washed twice with 70% ethanol, and dissolved in 9 μl of deionized water. For the primer extension reaction, 10 pmol of the DNA primer (5′-GTTTGTTTGTTTGTTGAGGGGG-3′) was 5′ end-labeled using 2 μl of [γ-^{32}P]-ATP (6000 Ci/mmol; New England Nuclear) and 3 U of T4 polynucleotide kinase (USB), as recommended by the manufacturer. After purification by phenol-chloroform extraction, gel purification, and ethanol precipitation, the radioactivity was adjusted to 5000 CPM/μl and the concentration to 0.3 μM by adding cold primer. The benzoyl cyanide-treated RNA and 3 μl of the radioactive primer mix were heated at 65°C for 5 min, then cooled at 37°C for 5 min, and finally incubated at 4°C for 1 min. Next, 4 μl of 5X First-Strand Buffer (Invitrogen), 1 μl of 100 mM DTT, 1 μl of 10 mM dNTP, and 2 μl of DMSO were added and the reactions preincubated at 61°C for 1 min prior to adding SuperScript III (100 U; Invitrogen) and incubating for another 10 min at 61°C. In the case of the ladder, untreated RNA and an additional 1 μl of 10 mM ddCTP were used in the reaction. The reactions were stopped by adding 1 μl of 4 N NaOH and heating at 95°C for 5 min in order to hydrolyze the RNA. The cDNA was ethanol precipitated, dissolved in 10 μl of loading buffer, and electrophoresed on 8% denaturing PAGE gels. The gels were exposed to phosphoscreens, scanned using a Storm apparatus (Molecular Dynamics), and the band intensities quantified using the SAFA software (Stanford University) (Laederach et al., 2008). The background from the reverse transcription was subtracted using the result obtained with DMSO-treated RNA. The gels were normalized with the band intensity of nucleotide A13.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and one figure and can be found with this article online at doi:10.1016/j.str.2010.09.024.

ACKNOWLEDGMENTS

We would like to thank Marc Parisien and François Major for sharing their expertise in the use of the MC-Sym platform as well as for the interesting discussions, and Luc Grenier for the clustering experiments. The work performed in J.-P.P.'s laboratory was supported by a grant from the Canadian Institute for Health Research (CIHR: MOP-44022). The RNA group is supported by grants from the CIHR and the Université de Sherbrooke. J.-P.P. holds the Canada Research Chair in Genomics and Catalytic RNA and is a member of the Infectious Diseases group of the Centre de Recherche Clinique Étienne-Lebel.

Received: June 25, 2010
Revised: September 13, 2010
Accepted: September 17, 2010
Published: December 7, 2010

REFERENCES

Baird, N.J., Westhof, E., Qin, H., Pan, T., and Sosnick, T.R. (2005). Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. J. Mol. Biol. *352*, 712–722.

Been, M.D. (2006). HDV ribozymes. Curr. Top. Microbiol. Immunol. *307*, 47–65.

Chauhan, S., and Woodson, S.A. (2008). Tertiary interactions determine the accuracy of RNA folding. J. Am. Chem. Soc. *130*, 1296–1303.

Cochrane, J.C., and Strobel, S.A. (2008). Catalytic strategies of self-cleaving ribozymes. Acc. Chem. Res. *41*, 1027–1035.

Cruz, J.A., and Westhof, E. (2009). The dynamic landscapes of RNA architecture. Cell *136*, 604–609.

Das, R., and Baker, D. (2007). Automated de novo prediction of native-like RNA tertiary structures. Proc. Natl. Acad. Sci. USA *104*, 14664–14669.

Deschênes, P., Ouellet, J., Perreault, J., and Perreault, J.P. (2003). Formation of the P1.1 pseudoknot is critical for both the cleavage activity and substrate specificity of an antigenomic *trans*-acting hepatitis delta ribozyme. Nucleic Acids Res. *31*, 2087–2096.

Ding, F., Sharma, S., Chalasani, P., Demidov, V.V., Broude, N.E., and Dokholyan, N.V. (2008). Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. RNA *14*, 1164–1173.

Ferré-D'Amaré, A.R., Zhou, K., and Doudna, J.A. (1998). Crystal structure of a hepatitis delta virus ribozyme. Nature *395*, 567–574.

Fiore, J.L., Kraemer, B., Koberling, F., Edmann, R., and Nesbitt, D.J. (2009). Enthalpy-driven RNA folding: single-molecule thermodynamics of tetraloop receptor tertiary interaction. Biochemistry *48*, 2550–2558.

Greenleaf, W.J., Frieda, K.L., Foster, D.A., Woodside, M.T., and Block, S.M. (2008). Direct observation of hierarchical folding in single riboswitch aptamers. Science *319*, 630–633.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

Ke, A., Zhou, K., Ding, F., Cate, J.H., and Doudna, J.A. (2004). A conformational switch controls hepatitis delta virus ribozyme catalysis. Nature *429*, 201–205.

Laederach, A., Das, R., Vicens, Q., Pearlman, S.M., Brenowitz, M., Herschlag, D., and Altman, R.B. (2008). Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. Nat. Protoc. *3*, 1395–1401.

Lilley, D.M. (2005). Structure, folding and mechanisms of ribozymes. Curr. Opin. Struct. Biol. *15*, 313–323.

Moody, E.M., and Bevilacqua, P.C. (2003). Folding of a stable DNA motif involves a highly cooperative network of interactions. J. Am. Chem. Soc. *125*, 16285–16293.

Mortimer, S.A., and Weeks, K.M. (2008). Time-resolved RNA SHAPE chemistry. J. Am. Chem. Soc. *130*, 16178–16180.

Nehdi, A., Perreault, J., Beaudoin, J.D., and Perreault, J.P. (2007). A novel structural rearrangement of hepatitis delta virus antigenomic ribozyme. Nucleic Acids Res. *35*, 6820–6831.

Nishikawa, F., and Nishikawa, S. (2000). Requirement for canonical base pairing in the short pseudoknot structure of genomic hepatitis delta virus ribozyme. Nucleic Acids Res. *28*, 925–931.

Onoa, B., and Tinoco, I., Jr. (2004). RNA folding and unfolding. Curr. Opin. Struct. Biol. *14*, 374–379.

Ouellet, J., and Perreault, J.P. (2004). Cross-linking experiments reveal the presence of novel structural features between a hepatitis delta virus ribozyme and its substrate. RNA *10*, 1059–1072.

Parisien, M., and Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature *452*, 51–55.

Reymond, C., Ouellet, J., Bisaillon, M., and Perreault, J.P. (2007). Examination of the folding pathway of the antigenomic hepatitis delta virus ribozyme reveals key interactions of the L3 loop. RNA *13*, 44–54.

Reymond, C., Bisaillon, M., and Perreault, J.P. (2009). Monitoring of an RNA multistep folding pathway by isothermal titration calorimetry. Biophys. J. *96*, 132–140.

Shapiro, B.A., Yingling, Y.G., Kasprzak, W., and Bindewald, E. (2007). Bridging the gap in RNA structure prediction. Curr. Opin. Struct. Biol. *17*, 157–165.

Sosnick, T.R., and Pan, T. (2004). Reduced contact order and RNA folding rates. J. Mol. Biol. *342*, 1359–1365.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature *460*, 711–716.

Zhuang, Z., Jaeger, L., and Shea, J.E. (2007). Probing the structural hierarchy and energy landscape of an RNA T-loop harpin. Nucleic Acids Res. *35*, 6995–7002.

