

Accepted Manuscript

A two-dimensional Replica-Exchange Molecular Dynamics method for simulating RNA folding using sparse experimental restraints

Parisa Ebrahimi, Simi Kaur, Lorenzo Baronti, Katja Petzold, Alan A. Chen

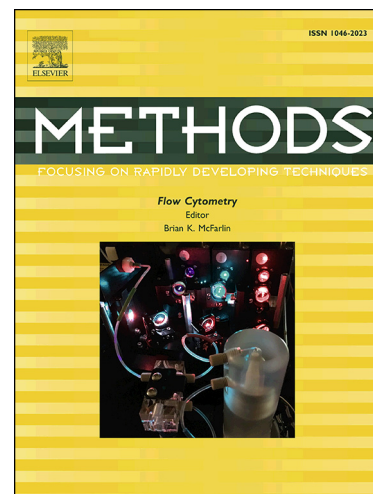
PII: S1046-2023(18)30338-4
DOI: <https://doi.org/10.1016/j.ymeth.2019.05.001>
Reference: YMETH 4698

To appear in: *Methods*

Received Date: 30 January 2019
Revised Date: 26 April 2019
Accepted Date: 1 May 2019

Please cite this article as: P. Ebrahimi, S. Kaur, L. Baronti, K. Petzold, A.A. Chen, A two-dimensional Replica-Exchange Molecular Dynamics method for simulating RNA folding using sparse experimental restraints, *Methods* (2019), doi: <https://doi.org/10.1016/j.ymeth.2019.05.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Title:

“A two-dimensional Replica-Exchange Molecular Dynamics method for simulating RNA folding using sparse experimental restraints”

Parisa Ebrahimi†, Simi Kaur†, Lorenzo Baronti, Katja Petzold and Alan A. Chen*

(University at Albany, SUNY)

(University at Albany, SUNY)

(Karolinska Institutet)

(Karolinska Institutet)

University at Albany, State University of New York, 1400 Washington Ave, Chemistry 123,
United States

† equal contributions

* corresponding author

ABSTRACT (250 Words)

We present a 2D replica exchange protocol incorporating secondary structure information to dramatically improve 3D RNA folding using molecular dynamics simulations. We show that incorporating base-pairing restraints into all-atom, explicit solvent simulations enables the accurate recapitulation of the global tertiary fold for 4 representative RNAs ranging in length from 24 to 68 nt. This method can potentially utilize base-pairing information from a wide variety of experimental inputs to predict complex RNA tertiary folds including pseudoknots, multi-loop junctions, and non-canonical interactions.

KEYWORDS (6)

RNA folding, Molecular Dynamics Simulations, Replica-Exchange, RNA Structure Prediction

INTRODUCTION

Molecular dynamics simulations have become an increasingly useful and reliable tool to study RNA. Simulations can complement experimental measurements by contextualizing how the 3D architecture and atomic motions of RNAs give rise to their molecular function. Explicit solvent, all-atom simulations are particularly attractive, as they can potentially capture the delicate competition between opposing physical forces (i.e. hydrogen bonding, stacking, steric repulsion, etc.) that ultimately dictate how an RNA will behave at the atomic level.

These simulations, however, require the 3D coordinates of the RNA's folded structure as a starting point - and unfortunately atomic resolution structures are typically unavailable for most RNA sequences of interest. Less than 1% of currently deposited structures in the RCSB [1] are of RNA (as of 1/2019), and this scarcity is unlikely to change due to persistent bottlenecks in experimental determination of RNA structures. RNA crystals suitable for X-Ray crystallography are notoriously difficult to obtain, compounded by the inherent flexibility and dynamic nature of RNA that is also often essential for its biological function. Alternatively, while NMR is able to probe the structure and dynamics of macromolecules in aqueous solution, in practice the severe spectral overlap often prevents atom-specific assignments of RNA spectra for all but the smallest of RNAs. As a result, there are only 9 NMR structures in total of RNAs greater than 65 nts in length, and they each represent many years of intensive labor [2].

In principle, molecular dynamics simulations can be used to fold RNA sequences from random unstructured starting conformations (i.e. "*de-novo*") given a perfect potential energy function (i.e. "force-field") and fully converged conformational sampling. In practice, while the force-fields used to model RNA have steadily improved in recent years[3,4], the required conformational sampling is too computationally demanding for all but the shortest of RNAs. Recent studies of 4-nucleotide RNA tetramers[5] were able to be completely converged utilizing petaflop-scale supercomputers; however achieving complete convergence in simulations of 8-10 nucleotide "tetraloop" hairpins [6] remains a challenge. Even the special-purpose ANTON molecular dynamics supercomputer, which has successfully folded 80 amino acid proteins[7], is only able to reversibly melt 5-8 basepair RNA duplexes[8].

RNA folding using all-atom molecular dynamics simulations is difficult for multiple, interrelated reasons. Compared to typical proteins, RNA is understood to have an inherently more rugged "folding landscape"[9]. Fast-folding proteins of 40-150 amino acids have been shown to fold in a cooperative, two-state manner - typically on the microsecond timescale and stabilized by just 1-6 kcal/mol for the entire folding process[10]. In contrast, RNA folding is a hierarchical, multi-step process: rapid initial formation of stable secondary structure (i.e. base-pair formation) is followed by slower conformational search for stable tertiary structure motifs (i.e. pseudoknots, loop-loop interactions, multi-loop junctions, etc)[11,12]. The highly favorable free-energy associated with base-pair formation (~ -1 to -3 kcal/mol *per base-pair*) and stacking of adjacent basepairs means that even a *suboptimally* basepaired configuration could be an order of magnitude *more stable* than an entire folded protein. Since the number of suboptimal folds increases exponentially with sequence length[13], the resulting folding landscape will be littered with deep kinetic traps that obstruct folding to the minimum free-energy state. Lastly, experimental estimates indicate that RNA tertiary folding occurs on a millisecond timescale or slower[11], so RNA folding is unlikely to be spontaneously observed in typical ns-??s molecular simulations. Therefore, *ab initio* folding of biologically relevant sized RNAs via "brute force" molecular dynamics remains infeasible at the current time. However, such approaches completely ignore the wealth of secondary-structure information that can be readily obtained during routine structure-function studies of RNA, even when no 3D structure is available.

Merely entering the RNA's primary sequence into any online secondary-structure predictor will immediately identify probable helix-forming regions with >80% accuracy depending on length[14]. This information can be improved with chemical probing experiments such as SHAPE, which report whether individual bases are single or double stranded[15]. Furthermore, even simple ¹H NMR spectra, recorded within 1-3 days, while insufficient for solving a full 3-D structure via traditional distance-geometry methods[16], can nonetheless reveal the sequence position of all base-pairing partners[17]. This requires substantially less effort than collecting the exhaustive amount of NMR data used in a traditional NMR structure determination.

In this work, we present a method to incorporate sparse base-pairing information to dramatically improve the accuracy and efficiency of 3D RNA folding using a novel 2D replica exchange protocol. We will show that just by incorporating a handful of secondary structure restraints into an otherwise unbiased all-atom simulation, accurate RNA tertiary folds can be modelled at the fraction of the effort of *de-novo* simulation of RNA folding. This method can potentially incorporate base-pairing information from a wide variety of experimental inputs to predict highly complex RNA tertiary folds including pseudoknots, multi-loop junctions, and non-canonical interactions.

THEORETICAL BACKGROUND

The 2-D replica exchange simulation algorithm described in this work is an extension of traditional temperature replica-exchange molecular dynamics (T-REMD)[18] combined with Hamiltonian replica-exchange[19], in which base-pairing connectivity is enforced by weak long-range restraints in order to narrow the conformational space that need be explored to find the free-energy minimum.

T-REMD - originally called "parallel tempering" in the study of spin glass physics [20], has proved to be an effective method for enhancing the equilibrium sampling of kinetically trapped systems. T-REMD utilizes a chain of nearly identical simulation replicas simulated in tandem but maintained at different temperatures (**Fig. 1b**). All replicas periodically attempt to swap configurations according to the Metropolis criterion, and the advantageous situation emerges where misfolded proteins are repeatedly heated and cooled until they arrive at the minimum free energy conformation. T-REMD has been shown to efficiently sample unfolded/folded transitions for proteins[21,22], but does not work nearly as well for folding RNA. The hierarchical, multi-state nature of RNA folding[12], where tertiary structure forms only after secondary structure is established, results in broad, rather than sharp melting transitions. Moreover, even modestly sized RNAs can have melting temperatures that exceed the boiling point of water, meaning in real life they are never completely unfolded and always possess some residual structure.

Additional disadvantages arise out of purely computational considerations. For T-REMD to reversibly fold a macromolecule, the hottest replica temperature must exceed the simulated melting point of the system, while the coldest temperature replica should be at the temperature of interest (i.e. 298 °K). Many intermediate temperature replicas must also be simulated in order

to maintain an average accepted swap rate of 15-20% between neighboring replicas[23]. Failure to maintain an adequate swap rate between replicas results in an uncoupled simulation in which individual trajectories cannot access both hot temperatures to unfold and cooler temperatures to reanneal, eliminating all advantages of T-REMD. Additionally, as the volume of an explicit-solvent simulation increases, so does the heat capacity (C_v) of the overall system, leading to intrinsically larger energy differences between replicas at different temperatures (i.e. proportional to $C_v \cdot dt$). The energy fluctuations from the extra solvent molecules, however, only grow proportional to $\sqrt{C_v}$, necessitating smaller temperature intervals between replicas to maintain a constant swap ratio. Since RNAs require higher simulation temperatures to melt than proteins, many more replicas are needed to span the necessary temperature range. For example, prior explicit solvent studies of 8-10 basepair tetraloop hairpins required simulation temperatures of 500-600 °K to fully melt, requiring 50-64 replicas to span the entire temperature range[24,25]. The number of replicas will be significantly greater for longer RNAs, as not only will the melting temperatures be higher, but also the temperature interval spacing between replicas will simultaneously decrease as well. For all these reasons, unmodified T-REMD will be inherently inefficient for folding RNAs of biologically relevant lengths (i.e. 20-100 nucleotides).

One approach to increasing the efficiency of T-REMD is Replica Exchange with Solute Tempering (REST) [26,27], in which only the temperature of the macromolecular solute (and not the solvent) is altered, allowing for greater effective temperatures to be sampled using a given number of replicas. While this approach has proven fruitful for certain protein and membrane systems where coupling to the water can be assumed to be weak, it has been shown to be problematic when applied RNA systems [28], where it caused rapid deterioration of the native state fold. This is likely due to the much stronger coupling between RNA and its solvation layer on account of extensive hydrogen bonding and ionic hydration.

Another approach is to gradually impose a biasing potential to smoothen the free energy landscape. This can be accomplished using a chain of replicas at identical temperatures where each replica employs a slightly different potential energy function; if these replicas periodically attempt to swap conformations according to the Metropolis criterion, then this is known as Hamiltonian Replica-Exchange Molecular Dynamics (H-REMD)[19,29–31]. Successful H-REMD methods generally require prior knowledge of the rate-limiting energy barrier such that the imposed biasing potential can be specifically tuned to ameliorate its impact. Cheatham and coworkers have shown that combining T-REMD with torsional-bias H-REMD resulted in more efficient sampling than spending equivalent CPU time using either T-REMD or H-REMD alone[5]. Inspired by these results, we investigated whether other choices of biasing Hamiltonians could lead to even more efficient sampling of RNA folded states.

Our 2D REMD RNA folding method improves upon T-REMD by incorporating an orthogonal perturbation in the form of soft restraints that favor formation of long-range native contacts. These native contacts effectively “sharpen” the free energy landscape by disfavoring exploration of conformations that violate the native contacts, an approach that has previously been shown to significantly increase efficiency of protein folding using approximate simulation models[32,33]. It should be noted, however, that it is tricky to accurately identify long-range native contacts for

proteins unless the 3D structure has already been solved. In contrast, we can often infer many long-range native contacts of an RNA sequence from secondary structure considerations alone, due to the hierarchical nature of RNA folding.

The central philosophy of our method is to conservatively restrain only unambiguously assigned regions of RNA secondary structure, to increase the efficiency of finding the most stable tertiary structure configuration. Incorporating even a small number of long-range native contacts dramatically reduces the conformational space to be searched by the simulation model, while still allowing the underlying all-atom interaction potential to decide how unrestrained bases will tend to behave (i.e. pairing, stacking, tertiary interactions, etc.)

Native contacts are incorporated as variable-strength, long-range restraints in an orthogonal replica dimension to temperature according to the scheme in **Figure 1a**. This results in a 2D simulation grid where each individual replica contains the same RNA molecule under identical simulation conditions except for the temperature and restraint strength. As with T-REMD, replicas periodically attempt to swap configurations with one another according to the Metropolis criterion, effectively coupling all simulations together while still maintaining equilibrium conditions within each replica. The restraint strengths are comparable to a long-ranged Hydrogen-bond, and crucially drop to zero strength in the low temperature, bottom row of the simulation grid.

This variation of temperatures and restraints across a single set of coupled simulations results in several features that greatly enhance reversible RNA folding. Restrained simulations are normally problematic in that if the restraints are too weak, they fail to bias the simulation enough to effect noticeable change; alternatively, if they are too strong to be overcome by thermal fluctuations, severe kinetic trapping will occur as the first misfolded conformation can never be undone to allow further sampling. The 2D replica-scheme avoids these problems in the following manner:

1. The replica with low-temperature and fully-on restraints (**Fig. 1a, upper left**) encourages targeted exploration of conformations that form as many base-pairs from the established secondary structure as possible.
2. Sub-optimally folded conformations from (1) can swap to the high temperature, zero restraint replica (**Fig 1a, upper right**) enabling it to unfold and escape kinetically trapped states
3. Unfolded structures from (2) can then anneal to lower temperature to attempt to find a better, more stable fold.
4. Well-folded RNA structures will be stable even in the absence of restraints. Only these conformations can effectively swap into the low temperature, zero restraint replica (**Fig. 1, lower left**).

This annealing process occurs continuously and simultaneously for all replicas of the RNA in a completely unsupervised fashion, since more stable configurations will naturally swap to lower temperatures and, in doing-so, “evict” less stable configurations to higher temperatures, where

they will unfold and eventually reanneal. In essence, the replica “swaps” are essentially an elaborate Monte-Carlo moveset that preferentially proposes conformations that satisfy base-pairing constraints, eliminating the long wait times needed to spontaneously find these conformations via a random walk at low temperature. Crucially, the presence of a replica with zero restraint strength filters out spurious conformations stabilized only by the restraints, and it is this replica that directly outputs the most physically realistic folded RNA conformations.

TEST SYSTEMS:

Four test cases with a range of tertiary structures are presented to illustrate the utility of the 2D REMD method for folding RNA. In each case, the RNA sequence was initiated in an extended, unstructured state; a minimal set of restraints were added corresponding to known base-pairing patterns (**Table S2**), and then the results of 2D REMD simulations using a minimal grid of 3 temperatures and 3 restraint strengths were compared to the experimental 3D structure to ascertain if the tertiary structure was properly recapitulated. For illustrative purposes, this work focuses only on RNAs with solved 3D structures such that the agreement with the true structure can be unambiguously assessed. However, these methods could in principle be applied to RNA systems lacking a solved structure, by combining predicted secondary structures and proposed tertiary contacts to efficiently predict the RNAs 3D fold.

The first test system was the 68 nt flavivirus xrRNA “knot”[34]. The 2.46 Å resolution crystal structure of this RNA (PDB ID: 4pqv) reveals a ring-like topology with the 5’ end of the molecule traversing the ring producing a knot-like structure (**Fig. 2a,d**), a feature that confers exonuclease resistance. Major features include two possible pseudoknots and a three-way junction produced by stacking of P1, P2 and P3. Correct formation of the three-helix junction is heavily dependent on interactions between several nucleotides, including a base triple, with the 5’ end of the molecule. Because of the unique “knotted” architecture of the RNA’s fold, it can easily get stuck in kinetic traps if base-pairs form in the incorrect order. Therefore, we wanted to test whether our method could achieve the correct tertiary fold despite the topological propensity of this RNA for forming kinetic traps.

The second test system was the 47 nt *Bacillus cereus* fluoride riboswitch [35]. The NMR structure (PDB ID: 5kh8) reveals a complex tertiary fold consisting of a central ligand-binding loop stabilized by two stem loops, a 4 base pair pseudoknot, and several additional long-range base-pairs that stabilize the tertiary structure (**Fig 3a, d**). The 20-structure NMR ensemble had an RMSD of 0.68 Å from the average structure. All stem loop base pairs observed in the secondary structure (**Fig. 3a**) and the two long-range base pairs were included as restraints. Pseudoknots are notoriously difficult to predict via secondary structure models, presumably because their stability is highly dependent on the local 3D architecture of the RNA which cannot be accounted for by a “nearest-neighbor” type model. We therefore included the fluoride riboswitch to demonstrate how all-atom models can provide the crucial structural context required to accurately model pseudoknot formation.

The third system, the 27 nt HBV tri-loop, is derived from the apical stem-loop of the human hepatitis B virus encapsidation signal (PDB: 2ixy). The NMR structure [36,37] reveals this RNA forms a pseudo-trinucleotide loop with 2 internal bulges (**Fig. 4a, e**). The reported 20-structure NMR ensemble of the HBV tri-loop had an RMSD of 1.92 to 2.71 Å from the average structure. This system has been thoroughly characterized by traditional NMR structural techniques, and therefore served as an ideal benchmark system to compare the relative effort required to obtain similar results via our improved simulation methods combined with a streamlined set of NMR experiments. For this study, just the ¹H NMR 2D NOESY spectra of the imino protons were acquired to ascertain the base-pairing connectivity of the system, which was then incorporated into our 2D-REMD simulation method to predict an equilibrium ensemble of 3D structures.

The final test system is the 24 nt SRP Internal Loop from Domain IV of E. coli SRP RNA (PDB ID: 1Int), whose structure was solved to 1.6 Å resolution via X-ray crystallography [38]. This system was chosen because of the high non-canonical base-pair content of the otherwise purely helical structure (**Fig 4b, e**). Most secondary structure models cannot account for non-Watson-Crick base pairs, and therefore would only predict a symmetric internal loop as indicated in **Figure 5a**. However, the all-atom simulation model should intrinsically allow non-Watson-Crick hydrogen bonds to form if the local packing correctly aligns the right hydrogen-bond donors and acceptors. We wanted to test whether just enforcing the three canonical base pairs on either side of the non-canonical region (which would have been straightforward to predict) would properly recapitulate the formation of a helix with non-Watson-Crick base-pairs. The original bi-molecular duplex was converted into a unimolecular stem-loop with the addition of a GAAA tetraloop sequence to fuse the two strands, which is both computationally expedient and more representative of the biological SRP RNA. It should be emphasized that no restraints were placed on the internal loop/non-Watson-Crick base pairs.

METHODS

All-atom molecular dynamics simulations were performed using the GROMACS 2016.4 molecular dynamics package[39] and the Amber-99-Chen-Garcia RNA force-field [24] with backbone phosphate modifications [40] for all test systems. Trajectories were propagated by integrating Newton's equations of motion using a velocity-Verlet algorithm with a 2 fs time-step. The TIP4P-Ewald model [41] was used to represent water molecules. Secondary structure information was incorporated into the test systems using distance restraints described by a flat-bottomed, harmonic potential with a piecewise linear extension between specific hydrogen bond donor and acceptor atoms. The flat bottom potential extends from 0 to 0.16 nm, the harmonic regime from 0.16 nm to 0.4 nm, and beyond 0.4 nm was the linear regime. In general, slightly higher force factors were used for tertiary vs secondary contacts (for details see **Table S1/S2**).

Two-dimensional replica exchange simulations, with a nine replica 3x3 grid model, were performed for all test systems at 3 different temperatures and with restraint strengths of 0%, 50%, and 100%. Base pairing interactions from the secondary structure were the only information provided as distance restraints in all test systems. Native base-pairs were restrained via a flat bottomed, harmonic potential connecting the central hydrogen bond donor and

acceptor of each pair. The restraint strength was defined using a force factor of 4000 kJ mol⁻¹ nm⁻² for most short range (secondary structure) base pairs and a slightly stronger force constant of 5000 kJ mol⁻¹ nm⁻² for long range (tertiary contact) base pairs.

Pymol, a molecular visualization program, was used to calculate the all-atom RMSD for all test systems. The align functionality superimposes structures following a sequence alignment, and then several refinement cycles are conducted to reject structural outliers. The all-atom RMSD was obtained by setting the number of refinement cycles to 0 in order to prevent outlier rejection. This functionality considers all provided states of a molecule, an NMR ensemble for example, when providing the final RMSD value after sequence alignment.

For the larger xrRNA and riboswitch structures, the ??RMSD metric of RNA structural similarity proposed by Bussi and coworkers[42] was calculated using the BARNABA software package[43]. ??RMSD calculates the structural similarity of two RNAs by comparing the relative spatial orientation of corresponding nucleobases which is highly sensitive to differences in base-pairing and stacking arrangements. This is a superior metric for larger RNAs with multiple helices, as even small differences in inter-helical angles result in large apparent RMSD differences even if base-pairing and stacking patterns are identical. **Fig. S3** and **Fig. S4** show the lack of correlation between RMSD and ??RMSD for these larger RNA structures.

Detailed simulation settings and preparative details for each individual system can be found in the **Appendix**.

RESULTS:

Flavivirus xrRNA “Knot”:

The global fold of the RNA appeared quickly, around 50 ns per replica, all basepairs were correctly formed by ~125 ns. Optimal temperatures for folding this molecule required several iterations to be correctly identified. Initially simulations at ~300 K were far too low to allow for adequate melting of suboptimal folds. Efficient folding was observed once the ~400K melting point was identified.

The overall global fold of the simulated flavivirus xrRNA “Knot” structure (**Fig. 2c**) was highly similar to the experimental crystal structure, with an ??RMSD of 1.32 (see Fig. S3 which compares ??RMSD vs RMSD for this RNA). Key architectural features of the xrRNA Knot structure were recapitulated (**Fig 6**), including correct formation of the ring, pseudoknot and three-way junction, and correct placement of the 5' end. A ring-like topology is observed in the experimental and crystal structures, where nucleotides 33 to 49, from P1 and P3, form a concave continuous loop on one side of the molecule. The 5' end of the xrRNA is placed through this ring, passing from one side of the structure to the other side, and creating a knot-like structure. As in the crystal structure, the first proposed pseudoknot (nucleotides 29-31 and 50-52) is not observed in the simulated structure, even though the bases are positioned close

enough to potentially pair. The second pseudoknot, G3-C40, between S1 and S3 is observed in the simulated structure (**Fig. 6e,f**). Correct formation of the three-way junction is dependent on stacking of C5-G46 on U47-A49 (**Fig. 6b**), and a base triple formed by U4-A24-U38 (**Fig. 6d**). Although C5-G46 is observed to have similar placement as in the crystal structure, a stacking platform is not provided by U47-A49 (**Fig. 6a**). The placement of U4 is angled slightly higher, but the base triple is formed within the major groove of A24-U38 (**Fig. 6c**), as in the crystal structure (**Fig. 6d**).

The overall fold of the simulated structure is correct, and the major features defining the structure of the Dengue Knot were observed in the simulated structure. This combined with the fact that the restrained molecule was not kinetically trapped by forming base pairs in the incorrect order demonstrates the overall robustness of our 2D-REMD simulation method.

B. Cereus Fluoride Riboswitch:

The B. Cereus riboswitch was simulated for a total of 124 ns per replica, however the overall fold of the structure appeared fairly quickly, from 10 to 40 ns. Even though there are 16 total base-pairs in this system, the longest contiguous helix is just 6 base-pairs, which is why an effective melting temperature of ~400K was used, identical to that used for the Flavivirus xrRNA Knot system.

The simulated structure of the riboswitch (**Fig. 4c**), with an ??RMSD of 1.32 to the NMR ensemble (**Fig. 4d**), exhibits well-folded P1, P2 and P3 stem loops (see Fig. A4 which compares ??RMSD vs RMSD for this RNA). Tertiary junctions are well formed as long-range base pairs between A5/U35, and A37/U45 (**Fig. 7c,d**) are identical to the NMR structure. Long-range hydrogen bonding contacts between A37-2' OH and G7-N7 (**Fig. 7a,b**), U38-2'OH and A40-N7 and U38-H3 to C41-O2P (**Fig. 7e,f**) were correctly formed as seen in the experimental structure. The overall global fold is very similar to the NMR ensemble, and several of the long-range interactions, although not completely formed, were in close proximity to positions observed in the NMR ensemble. Therefore, this test demonstrated the ability of the 2D-REMD method to correctly recapitulate pseudoknot formation.

HBV pseudo-triloop:

The HBV ensemble was depicted by the centroid of the main cluster (**Fig. 4c**) along with 20 random structures from the main cluster (**Fig. 4d**), depicted alongside the previously published HBV NMR ensemble (**Fig. 4e**). The formation of the pseudo-triloop is well recapitulated overall, given the NOE imino walk constraints. The RMSD of the simulated ensemble range from 3.56 to 4.30 Å with respect to the NMR ensemble. Two internal bulges corresponding to nucleotides C16 and U23 and a flexible three nucleotide UGU loop were observed in both the NMR and simulation ensembles (**Fig. 8a-d**). In the original NMR studies [36,37], a substantial amount of data was collected to conclude that C16 and U23 were bulged out, including the sugar pucker, missing NOE's, and RDCs. It is quite promising that the simulation was also able to correctly

predict both of these residues as being bulged out without any restraints applied to these positions one way or another, it was merely the lowest energy conformation given the imposed base-pairing pattern of the surrounding nucleotides.

The most significant disagreement is for U12 which is flipped inwards in the NMR ensemble (**Fig. 8c**) but tends to be flipped out in the simulation (**Fig. 8d**). In the original NMR report, NOEs indicative of stacking between C11/U12/G13 as well as helical 31P chemical shifts for U12 provide direct evidence that U12 stacks inwards at least some of the time, while the simulation prefers a different stacking arrangement between U12 and its surrounding bases. However, the NMR ensemble exhibits significant heterogeneity in the conformations of the triloop, and Flodell et al note that the data is unable to specify a single well-defined loop structure [36]. However, they also note high force-field energies and a lack of experimental restraints for loop bases, making it unclear whether the reported NMR structures reflect actual conformational dynamics or merely artifacts of the NMR method itself.

SRP non-canonical Internal Loop:

The SRP non-canonical internal loop is distinctive in that it contains 4 sequential non-Watson-Crick base-pairs which ultimately fold into a helical segment (**Fig 5e**). Since non-canonical base-pairing is difficult to predict from sequence alone, it would be shown as a 4x4 internal loop in most secondary structure predictors (**Fig. 5a**). However, the 3 canonical base-pairs adjacent to either end of the internal loop would be readily identifiable. Therefore, we restrained just the adjacent canonical base-pairs (6 in all) to see whether the simulation could recapitulate the intervening non-Watson Crick base pairs. In the simulation, all the non-canonical base-pairs correctly formed such that the helical overall fold was recapitulated (**Fig. 5d**). Interestingly, the order in which basepairs were formed was not random but rather proceeded in a sequential fashion in which each new non-canonical basepair was seeded by stacking upon an adjacent, fully-formed basepair - thereby elongating an existing helix rather than nucleating a new one. This observation corroborates previous findings by Major et al that the local structural context of nearest-neighbor bases dictates when non-canonical base-pairs are most like to form [44]. The detailed geometric classification of each non-canonical base [45,46] was correctly recapitulated (**Fig 9a-f**) and the overall RMSD was 2.62 Å with respect to the crystal structure. The global fold, however, had a slightly different helical pitch as a consequence of the GAAA tetraloop which is not present in the crystal structure. However, the ability of the all-atom force-field to recapitulate non-Watson-Crick base pairs solely from the structural context of the surrounding sequences illustrates the power of our restrained simulation method for simulating RNA tertiary folding.

DISCUSSION

We have detailed in this report the 3D folding of four RNAs with diverse tertiary structures using sparse experimental restraints, exploiting the inherently hierarchical nature of RNA folding to deduce long-range native contacts from secondary structure profiles. Our 2D-REMD simulation protocol employs a strategy inspired by structure-based-models[47] in which native-contact

potentials are employed to smoothen RNAs inherently rugged folding landscape[48], while still retaining the predictive power of explicit-solvent, all-atom simulations for the unrestrained RNA nucleotides. Since RNAs are experimentally known to fold in a hierarchical, multi-state fashion with frequent kinetic trapping, one should expect that the same to hold true in a sufficiently accurate simulation of RNA folding. Therefore, even as RNA force-fields continue to mature, developing improved sampling strategies to overcome kinetically trapped RNA, such as the 2D-REMD method presented in this work, will be increasingly important.

Of course, in a real-world scenario, it would be difficult to know the precise tertiary contacts to restrain without prior knowledge of the solved 3D structure, which we have used purely for illustrative and benchmarking purposes. However our method is computationally efficient enough such that a range of potential tertiary contacts (perhaps from mutagenesis studies or biophysical experiments) could be comparatively simulated in order to deduce the most likely 3D fold given a range of constrained tertiary structure arrangements.

We anticipate this method will be particularly useful for reconstructing tertiary folds as well as folding pathways from SHAPE experiments[49], where the base-pairing patterns are well-defined in some regions while ambiguous in others. In such a scenario, only the most robust secondary structure annotations from a SHAPE experiment would be converted into simulation restraints, such that the all-atom potential itself would have to decide how the ambiguously structured regions prefer to behave. The solvent accessibility of individual nucleotides in the folded ensemble could then be quantified to back-calculate a simulated SHAPE profile as a means of experimental validation[50].

While our RNA folding method is superficially reminiscent of the restrained annealing calculations traditionally used to solve NMR structures, it is in fact a completely distinct approach both from a practical and theoretical standpoint. NMR structure calculations are typically carried out in vacuum or implicit solvent using approximate atomic potentials derived from decades-old precursors to modern force-fields. These models are completely incapable of predicting lowest energy conformations of RNA (i.e. folding) on their own. Therefore, an NMR structure calculation additionally requires hundreds of experimental restraints in the form of NOE or RDC measurements. This can easily require months or even years of work to collect, analyze, and assign all the requisite NMR spectra. The goal is then to over-restrain an inaccurate molecular model with an overwhelming number of experimental restraints such that the correctly folded structure emerges as the **only** possible solution. Repeated application of this approach yields an NMR “ensemble” of solutions, but it is always unclear whether variance amongst the annealed structures reflects actual molecular flexibility or simply a paucity of experimental constraints. This structure-solving protocol was expedient in an era where both computational power and accurate atomistic potentials were difficult to come by. However, in the ensuing decades, the quality of the all-atom potentials has improved significantly, and high-performance computing has become nearly commonplace.

The original NMR study of the HBV triloop [37], which exemplifies established best-practices in RNA NMR, involved over a dozen distinct multidimensional NMR experiments representing at

least several weeks of NMR time plus months of additional time for manual assignments, validation of all the acquired spectra and structure calculations. As NMR structures of RNA still requires large amount of manual labor, a faster structure determination method is urgently needed. Our HBV triloop structural ensemble is nearly comparable in quality to the published NMR ensemble but required only basic ¹H NMR experiments taking less than 3 days total time to collect and annotate the 11 base-pairing restraints used in our structure calculation. The 2D-REMD calculation itself took ~5 days for a 9-replica simulation on our local, modest-sized computing cluster. Alternatively, this calculation could have been outsourced to a cloud computing service for a few hundred dollars.

NMR data has the distinct advantage that exact base-pairing connectivity can be rapidly established without a full structure calculation (i.e. an imino-walk), as opposed to coarse-grained data such as SHAPE which only reports single vs double-stranded and not the identity of the pairing partners. Most significantly, since the resulting structural ensemble comes from a molecular dynamics simulation at thermal equilibrium with zero-strength restraints, the variance in the ensemble unambiguously arises from the internal dynamics of the molecule itself. This dynamics information can be valuable for studying not just structure, but also folding pathways and dynamics as well.

It should be noted that while the force-constants used in traditional NMR structure calculations may superficially appear comparable to those used in our method, the net perturbation resulting from dozens of harmonic restraints per nucleotide is considerably stronger than the single, flat-bottomed, piecewise-linear restraint per base-pair used in our method. The presence of explicit solvent also significantly dampens the effect of the bias forces compared to vacuum (or implicit solvent), and the flat-bottom close-range and linear long-range shape of the restraints results in much smaller forces overall than the pure harmonic restraints used to model NOEs.

In summary, while all-atom, explicit solvent models for RNA are potentially accurate enough to capture RNA tertiary interactions, in practice much computational effort is wasted exploring dead-end kinetic traps using conventional sampling methods such as molecular dynamics and T-REMD. Conversely, experimental determination of a new RNA structure is a challenging process measured in postdoc-years and with a high chance of failure (i.e. partial NMR assignments, poorly diffracting crystals, etc.). The 2D REMD method presented in this work combines simulations and sparse experimental data to enjoy the best of both worlds, by eliminating the wasted computational effort spent exploring conformations that disagree with experimentally measured basepairing patterns. We anticipate that these methods could be used to revisit previously unsuccessful RNA structure determination projects as even a partial basepairing annotation could provide long-range constraints crucial for successfully determining and RNAs 3D fold.

HINTS FOR TROUBLESHOOTING

In all the systems studied, the minimal 3x3 2D-REMD grid was sufficient to sample productive folding events without excessive kinetic trapping. An ideal temperature range allows for

configurationally stable states with no bias forces to exist at lower temperatures, but also allows for kinetically trapped states to melt by swapping with a state with higher temperature and no restraints. The minimal 3x3 grid includes just 3 temperatures, the middle of which must be optimized to reflect the melting temperature of common misfolded intermediates.

Single MD simulations at varying temperatures were initially used to identify the simulation melting temperature of each specific system. Slightly higher temperature and lower temperature are introduced to encourage further melting, and continuance of stable conformations, while maintaining ~15% swap success rates in the REMD simulation. It should be noted that re-annealing of misfolded states does not require melting of the entire RNA, but rather just the misfolded portion which will be at a lower temperature.

The minimal 3x3 replica grid can be extended to include additional temperature and restraint strengths. Specifying additional restraint lambdas increases the number of replicas and did not measurably enhance sampling for the cases presented in this work - on account of the small number of weak restraining potentials that sum up to a small fraction of the overall energy of the system. If a large number of very strong restraints were utilized, then this would potentially benefit from additional replicas in the restraint strength dimension to increase swap acceptance between adjacent replicas. Additional temperatures may be useful for larger systems, as reaching higher temperatures will enable melting of deeper kinetic traps, while lower temperatures will more accurately capture the stability and dynamics of the system under more physiological conditions.

Simulation clock time is heavily dependent on the size and complexity of the molecule. In general, we observed that the temperature range was not suitable if hydrogen bond breaking and reforming is not observed within the first 10 ns. As with canonical T-REMD, the temperature schedule is too broad if there were no successful swap attempts in the replica-exchange step of the simulation.

It should be noted that the simulation box must be appropriately sized to contain the unstructured RNA without interacting with its periodic image. Melting the initial single stranded conformation into a more compact structure greatly reduces simulation time by reducing the total number of solvent molecules in the box compared to one large enough to contain the entire elongated single-strand.

It may be helpful to lower restraint strength for RNAs with lower melting temperatures and increase restraint strength for RNAs with higher melting temperatures. A further optimization would be to apply stronger bias forces for A-U base pairs, non-canonical interactions and base pairs closer to bulges, while weakening restraint strength for G-C base pairs, on account of the current force-fields being somewhat overly GC-friendly. This of course could also be remedied simply by using more thermodynamically accurate all-atom force-fields for RNA as they become available.

Encouragingly, we found exact specification of all long-range hydrogen bonding contacts was not always necessary as often similar results could be achieved including only partial tertiary structure restraints. However, a single mis-assigned base-pairing restraint could easily prevent the correct structure from forming at all, so it is always better to use less restraints which will simply require more sampling to compensate rather than incorrectly over-restrain the system. For 2D structure and SHAPE in particular, it would be expedient to only use only the subset of completely unambiguous secondary structure restraints in the 2D-REMD calculation.

APPENDIX

Flavivirus xrRNA “Knot” Details:

The flavivirus xrRNA “knot” crystal structure (PDB ID: 4pqv) was melted at 500 K, and then solvated with 18,331 mol of TIP4P-ew waters with 428 sodium ions and 361 chloride ions in a cubic box of dimensions 8.43 nm per side. A 9 replica 2D simulation at temperatures of 398 K, 400 K and 402 K was propagated for 262 ns per replica. Several tertiary contacts, G3-C40, U4-A24 and U47-U49 (**Fig 2a**), were identified from the crystal structure and included as restraints in the 2D-REMD simulation. The folded states were determined by clustering the lowest temperature and restraint trajectory using the “gromos” method with a 0.9 Å cutoff. The average swap rate was 25% for corresponding replicas, 20% for replicas with the same restraint strength at different temperatures and 7% for all other exchanges.

B. cerius Fluoride Riboswitch Details:

An initial, single stranded conformation (**Fig. 3b**) was solvated in 40,442 TIP4P-ew waters with 783 chloride ions and 829 sodium ions in a 10 nm x 10 nm x 13 nm orthorhombic box and energy minimized. The riboswitch was simulated in two-dimensional parameter and temperature space for 124 ns per replica, at 398 K, 400 K, and 402 K. The average swap rate was 15% for all replicas. Folded states were determined from the lowest temperature trajectory with no restraints.

HBV Triloop Details:

The melted HBV triloop sequence (**Fig 4b**) was solvated in 18,600 TIP4P-ew waters with 26 potassium ions in a cubic box 8.3 nm on a side, energy minimized and equilibrated. The melted structure was simulated with a temperature range of 505.10 K, 507.67 K and 510.25 K, with stronger force factors on base-pairs next to internal bulges and AUs (6) and weaker bias forces on CGs in the bottom stem (4), for 108.6 ns per replica. The folded states were determined from clustering the lowest temperature trajectory with zero restraints within a 5.0 Å cutoff to the lowest NMR energy ensemble.

The base pairing pattern, including the non-canonical pseudo-triloop was inferred using 1H-1H imino-imino NOE correlations. Unpaired nucleotides likewise can be identified by the same

method and by the characteristic ^1H - ^{13}C 6/8 chemical shift values that typically deviates from those of Watson-Crick base-paired nucleotides.

SRP Non-Canonical Helix Details:

The initial coordinates (**Fig. 5c**) reflect a partially melted structure where the middle symmetric internal loop is unstructured but the canonical base pairs on either end are intact. The system was solvated with 11,451 TIP4P-ew waters and 23 sodium ions in a 7.07 nm^3 cubic box, and energy minimized. Restraints were only applied to canonical base pairs at the top and the bottom of the stem, leaving the middle unrestrained. The molecule was simulated at temperatures of 358 K, 361 K and 364 K for 28 ns per replica. In addition, temperature REMD with hydrogen bond restraints on canonical base-pairs with a force of 500 kJ/mol was conducted over 9 temperatures: 350.0, 353.5, 357.0, 360.6, 364.2, 367.2, 371.5, 375.2, 378.9 K. Each replica was sampled for 30 ns. The folded states were determined from the lowest temperature and restraint trajectory. Average exchange rates were 20% for corresponding replicas and 10% for all other exchanges.

Acknowledgements:

AAC was supported by the National Science Foundation grant MCB1651877. This work used the computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE) [allocation TG-MCB140273 to AAC], which is supported by National Science Foundation grant number ACI-1548562

FIGURES

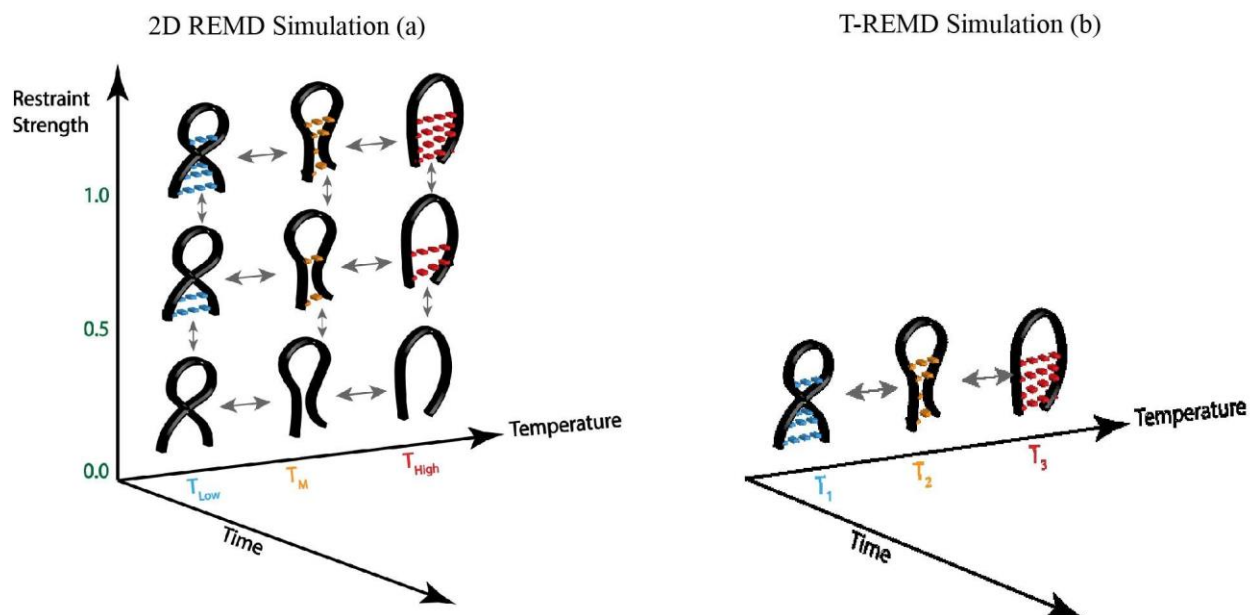


Figure 1: (a) 2D grid of replicas spanning both temperatures and restraint strengths. (b) regular 1D array of replicas for an unmodified T-REMD simulation

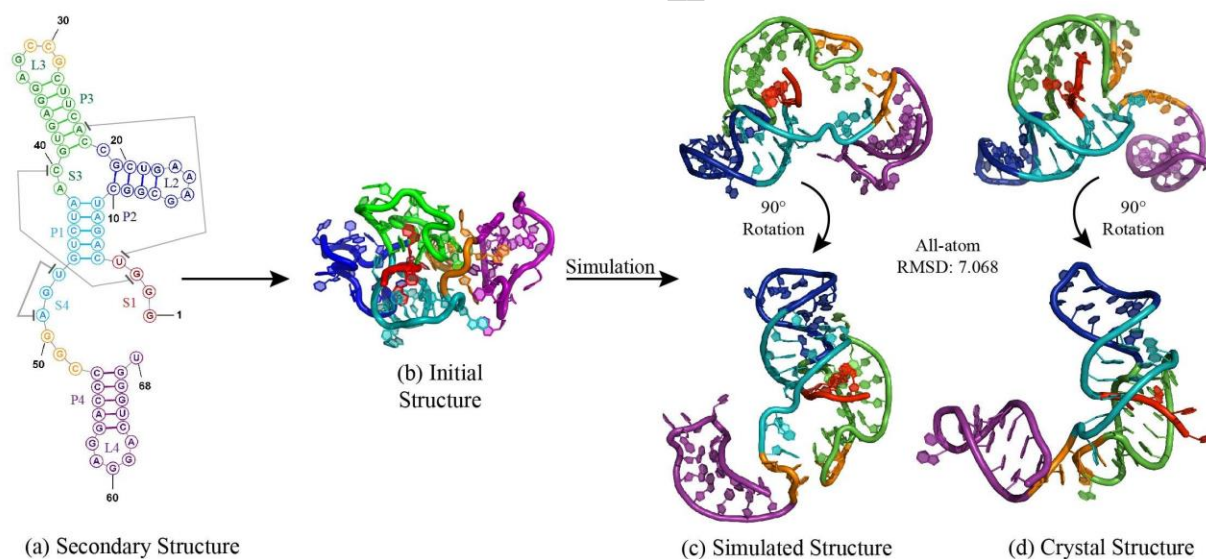


Figure 2: Flavivirus xrRNA "Knot" (a) secondary structure (b) initial collapsed conformation (c) 2D-REMD simulated structure (d) crystal structure

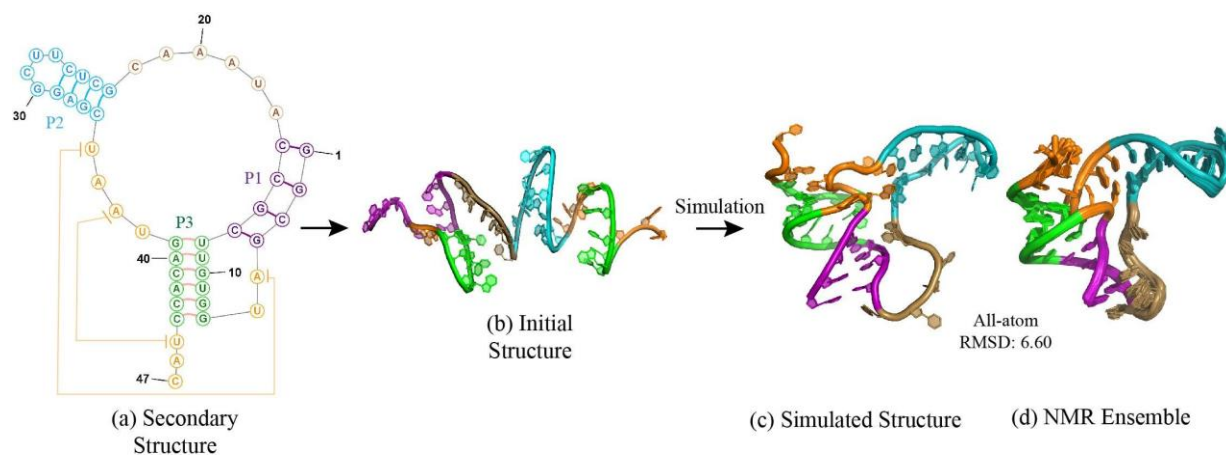


Figure 3: Fluoride Riboswitch: a) secondary structure (b) initial conformation (c) 2D-REMD simulated structure (d) NMR structure

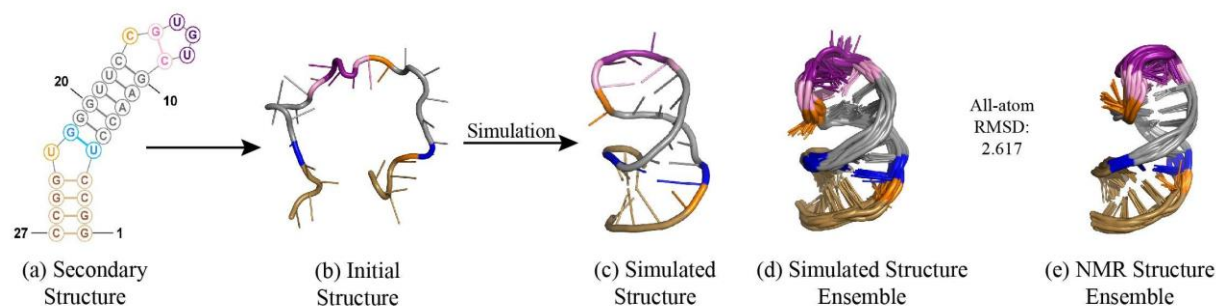


Figure 4: HBV Triloop: (a) secondary structure (b) initial conformation (c) 2D-REMD simulated structure (d) simulation ensemble of 20 random structures (e) NMR ensemble

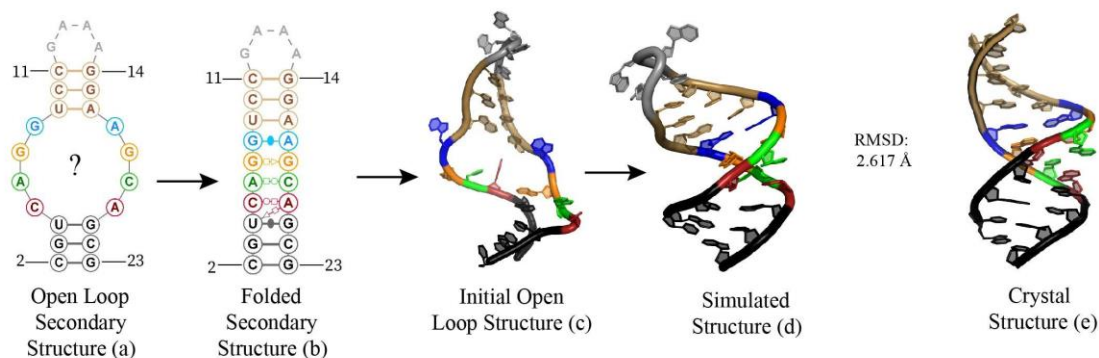




Figure 5: SRP non-canonical internal loop: (a) symmetric internal loop (b) secondary structure including non-canonicals (c) initial conformation with featureless internal loop (d) simulated structure (e) crystal structure

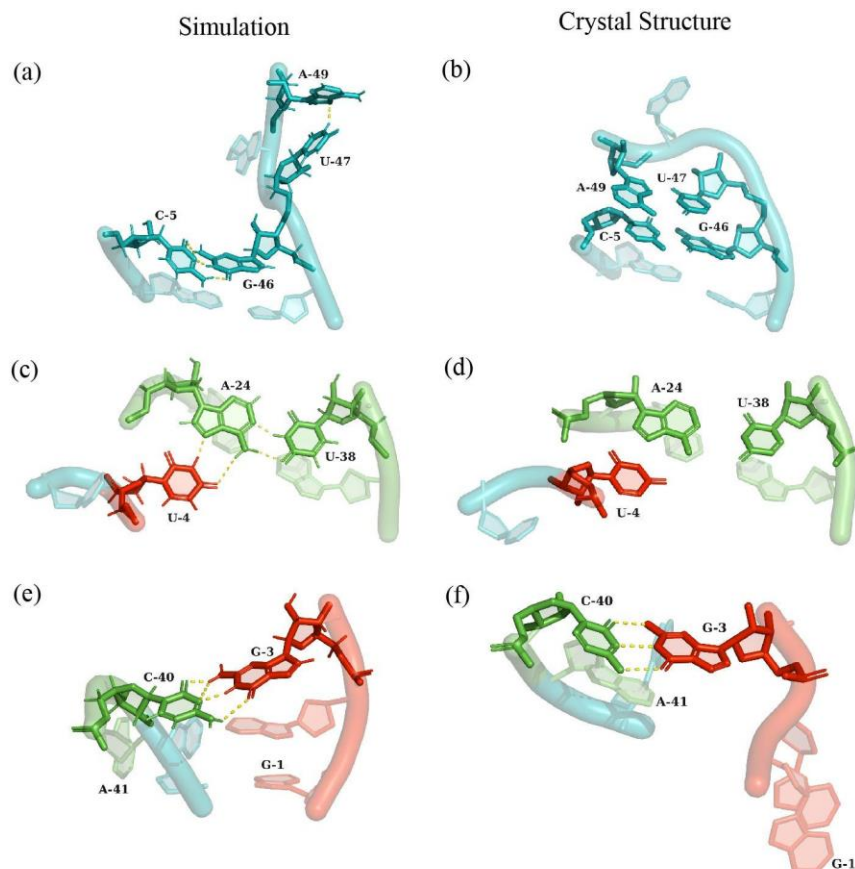


Figure 6: Flavivirus xrRNA “Knot” Key Interactions

- (a) C5-G46 base pair is well formed in simulated structure, but does not stack with A49-U47
- (b) C5-G46 base pair and stacking with A49 - U47 observed in crystal structure.
- (c) U4-A24-U38 base triple in simulated structure. U4-A24-U38 base triple in the crystal structure
- (d) Second Pseudoknot (G3-C40) in the simulated structure. Second Pseudoknot (G3-C40) in crystal structure

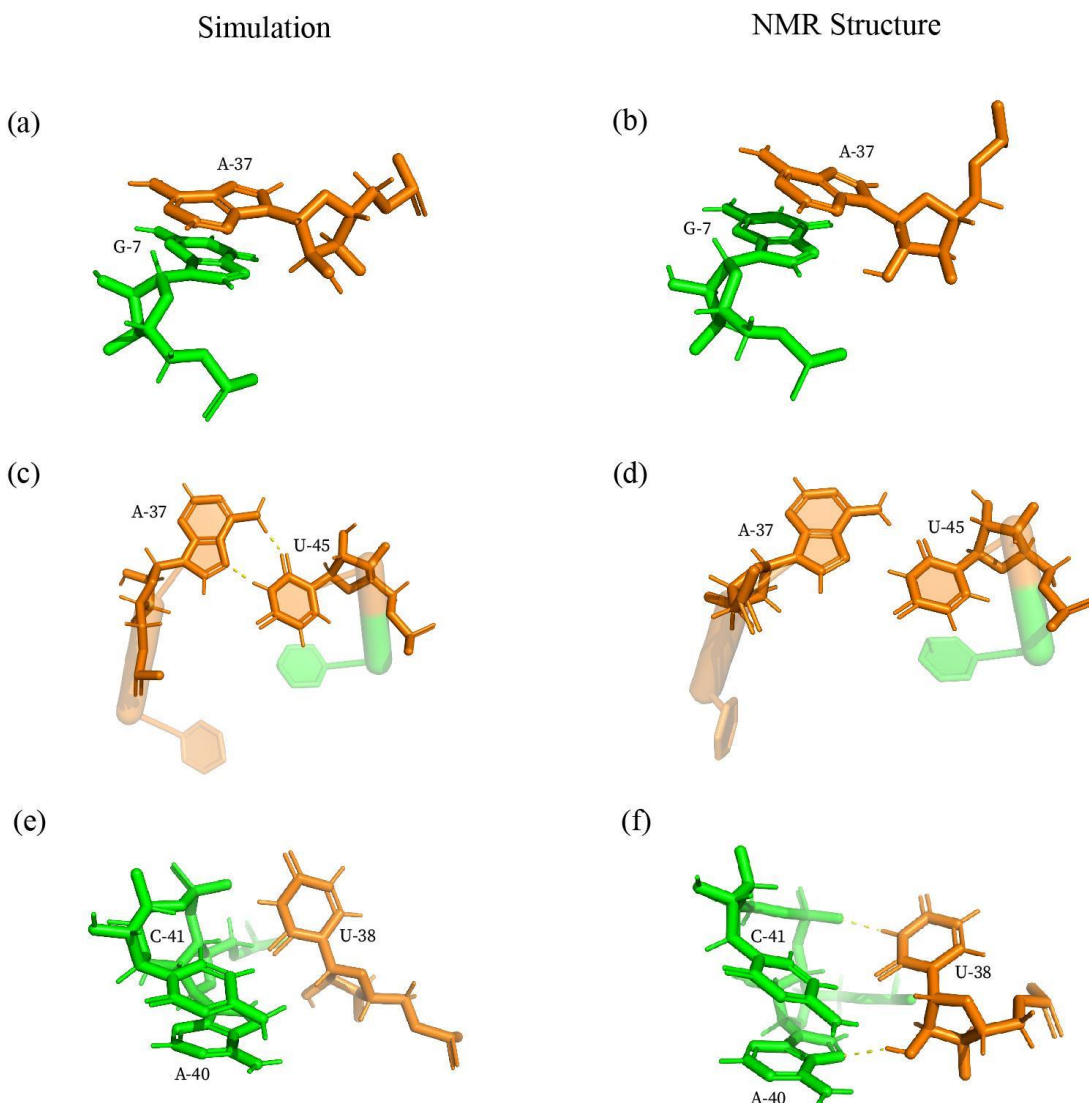


Figure 7: Fluoride Riboswitch Key Interactions

- Rare long-range hydrogen bond between highly conserved residues G7 (N7) and A37 (H2'C2') in simulated structure (a) and in NMR structure (b).
- A37 - U45 trans Watson-Crick/Hoogsteen base pair in simulated structure (c) compared with the NMR structure (d).
- Comparison of hydrogen bonding between U38 (2'OH) - A40 (N7) and U38 (H3) - C41 (O2P) in simulated structure (e) and NMR structure (f).

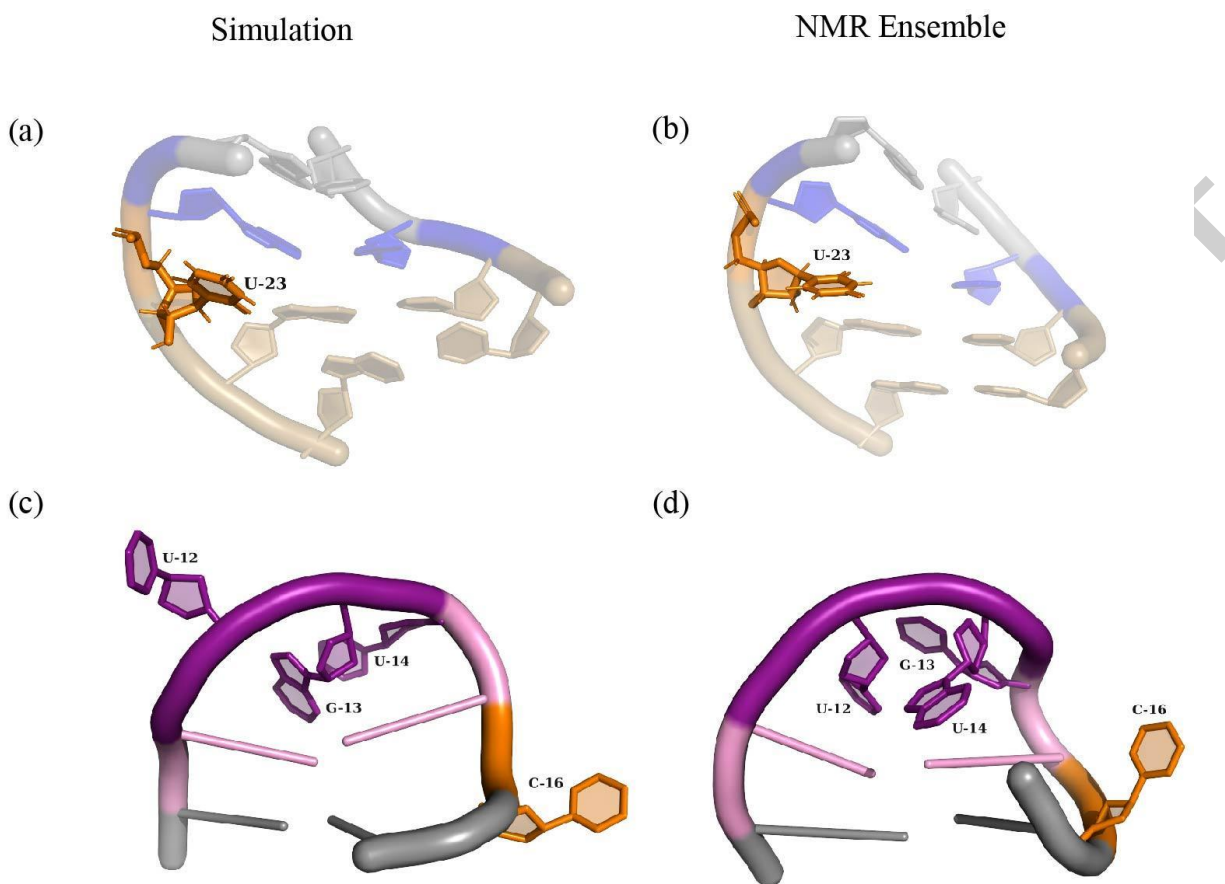


Figure 8: HBV Triloop Key Interactions

- (a) U23 bulge in simulation ensemble
- (b) U23 bulge in NMR ensemble
- (c) Bulge/bending due to C-16, tri-nucleotide loop
- (d) Pseudo-trinucleotide loop in NMR ensemble

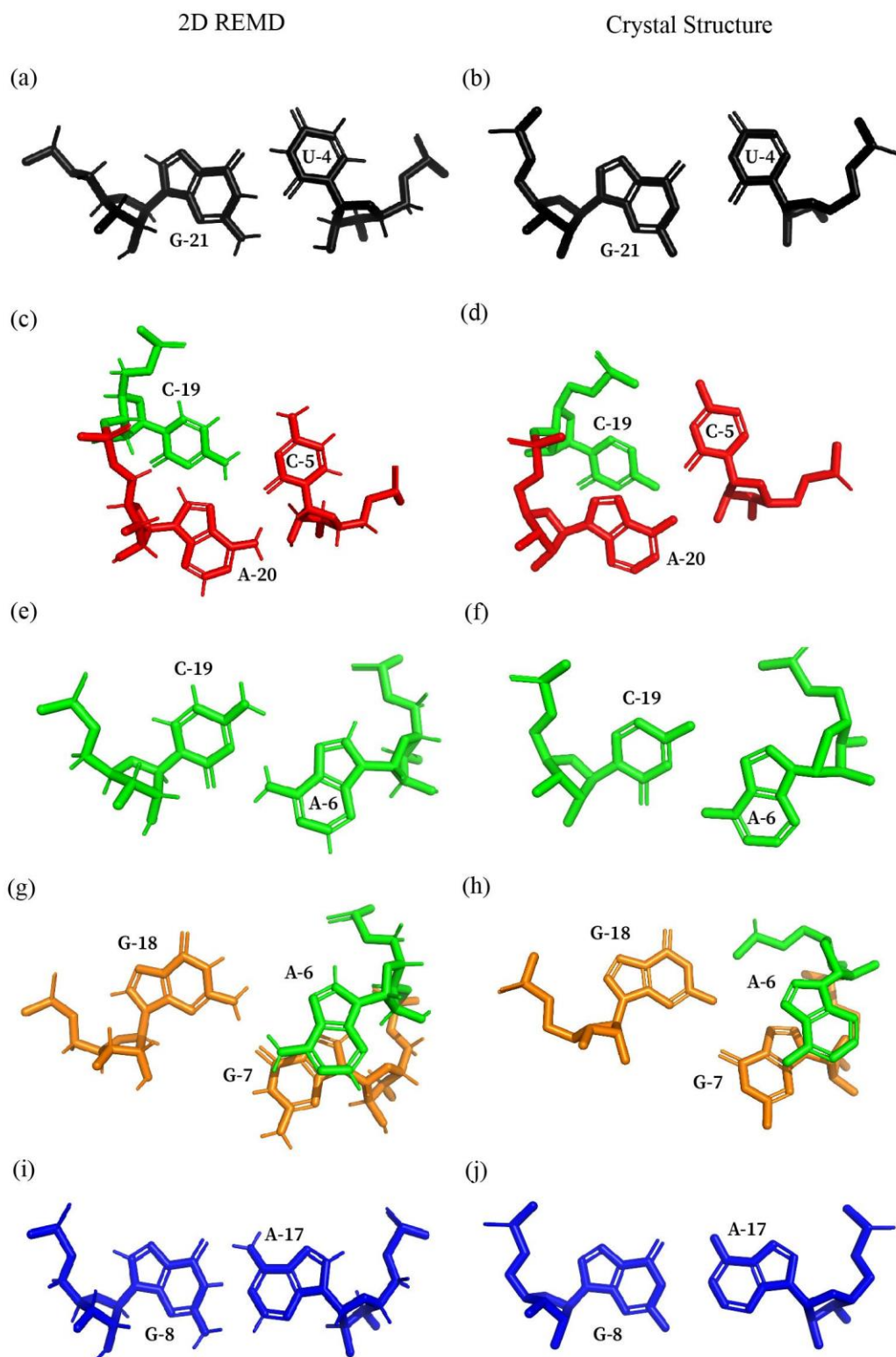


Figure 9: SRP RNA Non-Canonical Internal Loop Key Interactions

(a) Wobble GU base pair between U4 and G21 in simulation, (b) crystal structure

(c) C5/U4/A20 base triple in the simulation, (d) crystal structure - N4(C5) and O2P(C19) are closer together

- (e) A6 - C19 trans WC/Hoogsteen base pair in simulated structure, (f) crystal structure
 (g) N1(G18) - O2P(A6), O2P(A6) - N2(G18) and N2(G18) - O6(G7) in simulation, (h) crystal structure
 (i) A17 - G8 “symmetric” cis-WC/WC base-pair in simulation, (j) crystal structure

System	λ_1	λ_2	λ_3
Fluoride Riboswitch	0	5	10
Dengue Knot	0	5	10
HBV	0	4	8
Internal Loop	0	5	10

Table S1. ?? schedule used in the H-REMD dimension of the 2D REMD simulations for each RNA system.

System	H _{BOND} Donor-Acceptor	Force Factor (kJ mol ⁻¹ nm ⁻²)	Region
Fluoride Riboswitch	1G (H1) – 16C (N3)	5	P1
	2G (H1) – 15C (N3)	5	P1
	3C (N3) – 14G (H1)	5	P1
	4G (H1) – 13C (N3)	5	P1
	7G (H1) – 44C (N3)	5	P3
	8G (H1) – 43C (N3)	5	P3
	9U (H3) – 42A (N1)	5	P3
	10G (H1) – 41C (N3)	5	P3

	11U (H3) – 40A (N1)	5	P3
	12U (O2) – 39G (H1)	5	P3
	23G (H1) – 34C (N3)	5	P2
	24C (N3) – 33G (H1)	5	P2
	25U (H3) – 32A (N1)	5	P2
	26C (N3) – 31G (H1)	5	P2
	5A (N1) – 35U (H3)	6	Tertiary Contact
	37A (N7) – 45U (H3)	6	Tertiary Contact
Dengue Knot	5C (N3) – 46G (H1)	4	P1
	6A (N1) – 45U (H3)	4	P1
	7G (H1) – 44C (N3)	4	P1
	8A (N1) – 43U (H3)	4	P1
	9U (H3) – 42A (N1)	4	P1
	10C (N3) – 21G (H1)	4	P2
	11G (H1) – 20C (N3)	4	P2
	12G (H1) – 19U (O2)	4	P2
	13C (N3) – 18G (H1)	4	P2
	23C (N3) – 39G (H1)	4	P3
	24A (N1) – 38U (H3)	4	P3

	25C (N3) – 37G (H1)	4	P3
	26U (H3) – 36A (N1)	4	P3
	27U (O2) – 35G (H1)	4	P3
	28C (N3) – 34G (H1)	4	P3
	53C (N3) – 67G (H1)	4	P4
	54C (N3) – 66G (H1)	4	P4
	55C (N3) – 65G (H1)	4	P4
	56A (N1) – 64U (H3)	4	P4
	57G (H1) – 63C (N3)	4	P4
	3G (H1) – 40C (N3)	6	Tertiary Contact
	47U (H3) – 49A (N7)	6	Tertiary Contact
	4U (H3) – 24A (N7)	6	Tertiary Contact
HBV	1G (H1) – 27C (N3)	4	Stem 1
	2G (H1) – 26C (N3)	4	Stem 1
	3C (N3) – 25G (H1)	4	Stem 1
	4C (N3) – 24G (H1)	4	Stem 1
	5U (O2) – 22G (H1)	6	Stem 2
	6C (N3) – 21G (H1)	5	Stem 2

	7C (N3) – 20G (H1)	4	Stem 2
	8A (N1) – 19U (H3)	6	Stem 2
	18U (H3) – 9A (N1)	6	Stem 2
	10G (H1) – 17C (N3)	5	Stem 2
	11C (N3) – 15G (H1)	6	Triloop Lone Pair
Internal Loop	1C (N3) – 24G (H1)	5	Stem 1
	2G (H1) – 23C (N3)	5	Stem 1
	3U (O2) – 22G (H1)	5	Stem 1
	8U (H3) – 17A (N1)	5	Stem 2
	9C (N3) – 16G (H1)	5	Stem 2
	10C (N3) -15G (H1)	5	Stem 2

Table S2: List of restrained H-bond atoms used for each RNA system, restraint force factors, and annotation as secondary structural element (colored to match figures) or tertiary contact.

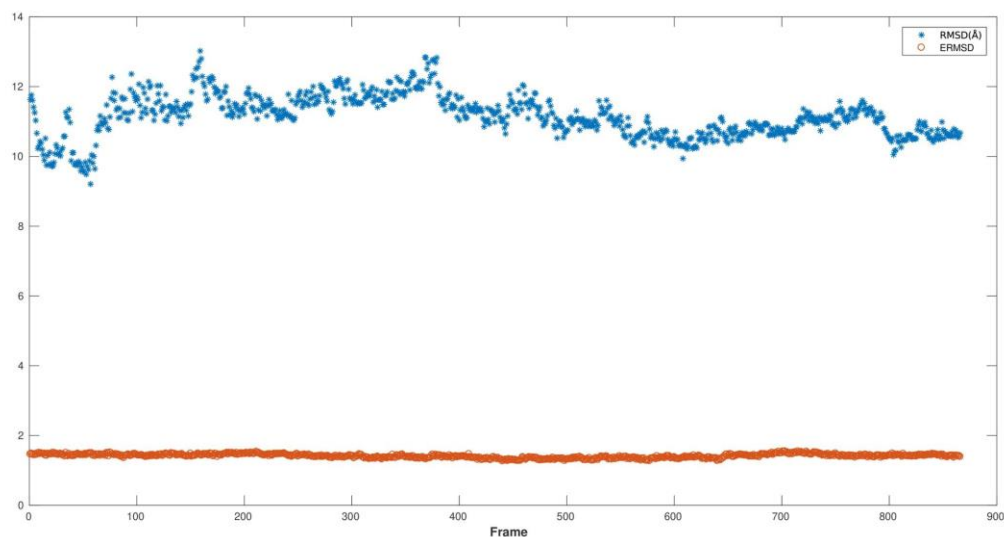


Figure S1. RMSD vs. ??RMSD time variation for xrRNA Knot folding trajectory

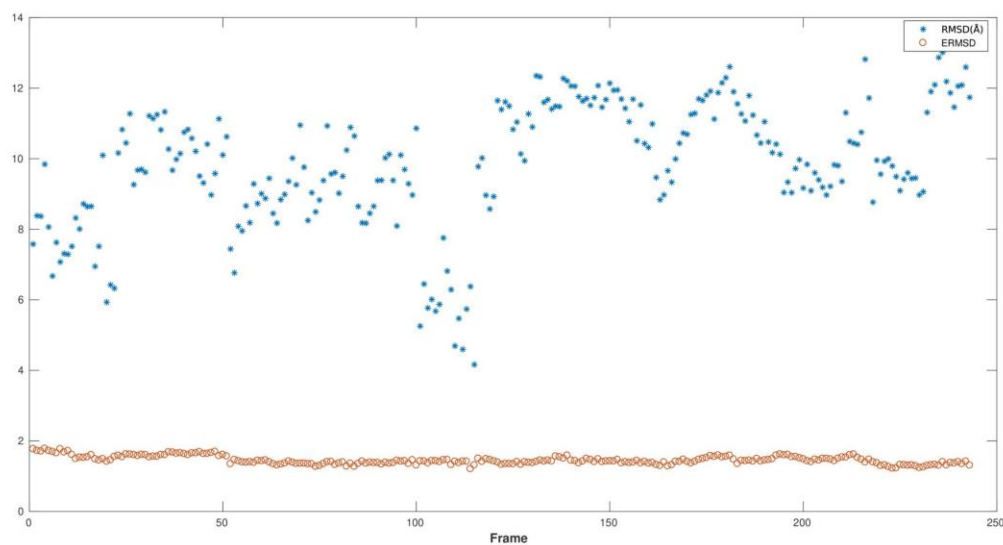


Figure S2. RMSD vs. ??RMSD time variation for Fluoride Riboswitch folding trajectory

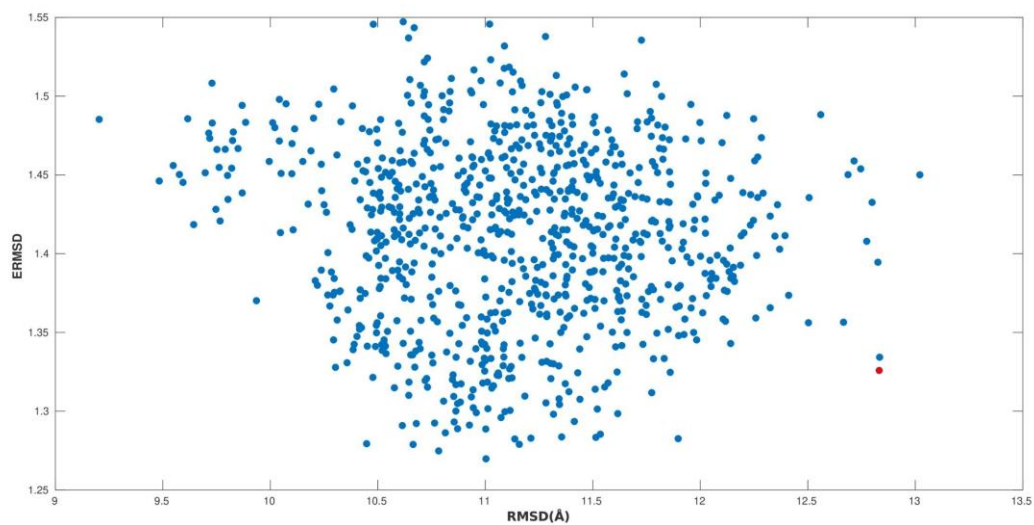
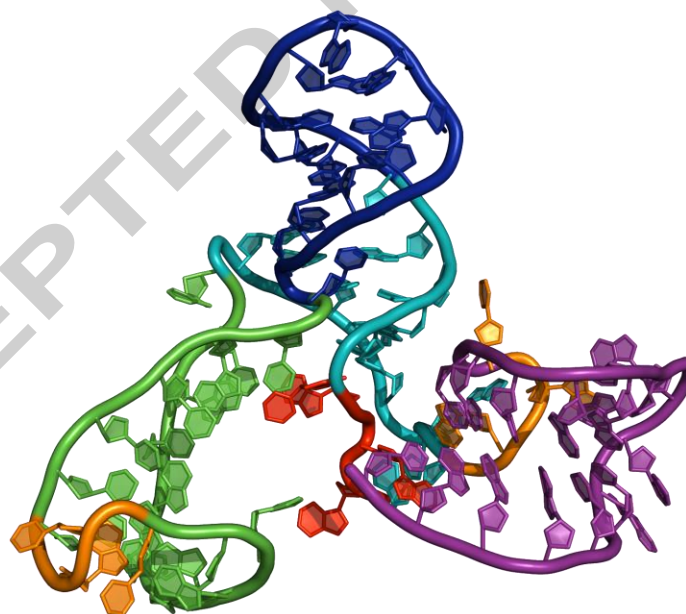


Figure S3. Heavy-atom RMSD from native versus ERMSD of 850 structures from folded cluster of xrRNA
Red circle represents an extracted conformation with high RMSD from native structure and low ERMSD (shown below)



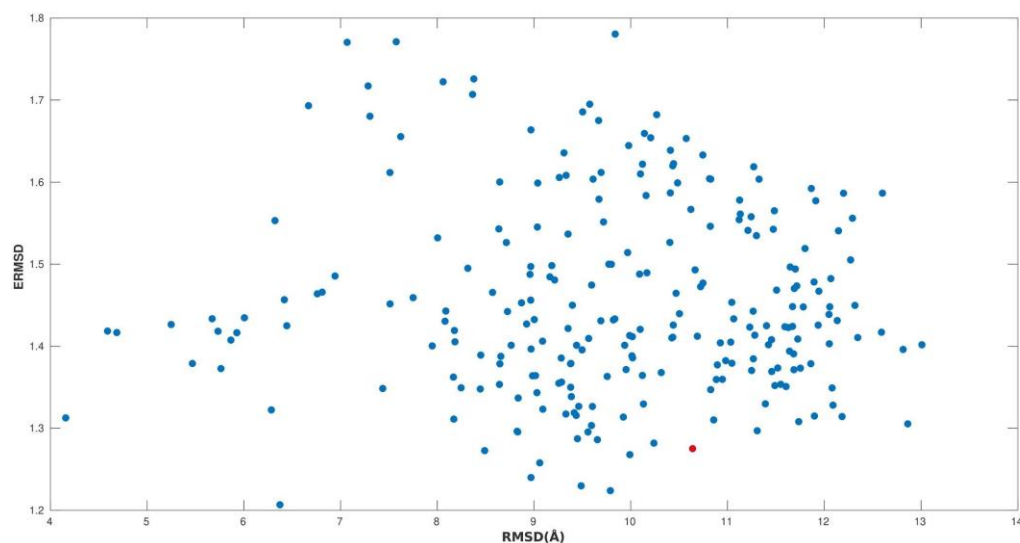


Figure S4. Heavy-atom RMSD from native versus ??RMSD of 250 ensembles from folded cluster of Fluoride Riboswitch
Red circle represents an extracted structure from clustered trajectory with high RMSD from native structure and low ??RMSD (shown below)



REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* Oxford University Press; 2000;28: 235–242. doi:10.1093/nar/28.1.235
2. Barnwal RP, Yang F, Varani G. Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys.* 2017;628: 42–56. doi:10.1016/j.abb.2017.06.003
3. Vangaveti S, Ranganathan SV, Chen AA. Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdiscip Rev RNA.* 2017;8. doi:10.1002/wrna.1396
4. Smith LG, Zhao J, Mathews DH, Turner DH. Physics-based all-atom modeling of RNA energetics and structure. *Wiley Interdiscip Rev RNA.* 2017;8. doi:10.1002/wrna.1422
5. Bergonzo C, Henriksen NM, Roe DR, Swails JM, Roitberg AE, Cheatham TE III. Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J Chem Theory Comput.* ACS Publications; 2013;10: 492–499. Available: <https://pubs.acs.org/doi/abs/10.1021/ct400862k>
6. Bergonzo C, Henriksen NM, Roe DR, Cheatham TE 3rd. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA.* 2015;21: 1578–1590. doi:10.1261/rna.051102.115
7. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science.* 2011;334: 517–520. doi:10.1126/science.1208351
8. Tan D, Piana S, Dirks RM, Shaw DE. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci U S A.* 2018;115: E1346–E1355. doi:10.1073/pnas.1713027115
9. Chen SJ, Dill KA. RNA folding energy landscapes. *Proc Natl Acad Sci U S A.* 2000;97: 646–651. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10639133>
10. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, Stability, Sequence, and Length: Defining the Determinants of Two-State Protein Folding Kinetics. *Biochemistry.* American Chemical Society; 2000;39: 11177–11183. doi:10.1021/bi000200n
11. Schlatterer JC, Martin JS, Laederach A, Brenowitz M. Mapping the kinetic barriers of a Large RNA molecule's folding landscape. *PLoS One.* 2014;9: e85041. doi:10.1371/journal.pone.0085041
12. Brion P, Westhof E. Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct.* 1997;26: 113–137. doi:10.1146/annurev.biophys.26.1.113
13. Clote P. Combinatorics of saturated secondary structures of RNA. *J Comput Biol.* 2006;13: 1640–1657. doi:10.1089/cmb.2006.13.1640
14. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A.* National Acad Sciences; 2004;101:

- 7287–7292. doi:10.1073/pnas.0401799101
15. Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc.* 2006;1: 1610–1616. doi:10.1038/nprot.2006.249
 16. Varani G, Aboul-ela F, Allain FH-T. NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc.* 1996;29: 51–127. doi:10.1016/0079-6565(96)01028-X
 17. Fürtig B, Richter C, Wöhnert J, Schwalbe H. NMR spectroscopy of RNA. *Chembiochem.* 2003;4: 936–962. doi:10.1002/cbic.200300700
 18. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett.* 1999;314: 141–151. doi:10.1016/S0009-2614(99)01123-9
 19. Jang S, Shin S, Pak Y. Replica-exchange method using the generalized effective potential. *Phys Rev Lett.* 2003;91: 058305. doi:10.1103/PhysRevLett.91.058305
 20. Swendsen RH, Wang JS. Replica Monte Carlo simulation of spin glasses. *Phys Rev Lett.* 1986;57: 2607–2609. doi:10.1103/PhysRevLett.57.2607
 21. Nymeyer H, Gnanakaran S, García AE. Atomic simulations of protein folding, using the replica exchange algorithm. *Methods Enzymol.* 2004;383: 119–149. doi:10.1016/S0076-6879(04)83006-4
 22. Garcia AE, Herce H, Paschek D. Chapter 5 Simulations of Temperature and Pressure Unfolding of Peptides and Proteins with Replica Exchange Molecular Dynamics. In: Spellmeyer DC, editor. *Annual Reports in Computational Chemistry.* Elsevier; 2006. pp. 83–95. doi:10.1016/S1574-1400(06)02005-6
 23. Rathore N, Chopra M, de Pablo JJ. Optimal allocation of replicas in parallel tempering simulations. *J Chem Phys.* 2005;122: 024111. doi:10.1063/1.1831273
 24. Chen AA, García AE. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci U S A.* 2013;110: 16820–16825. doi:10.1073/pnas.1309392110
 25. Kührová P, Best RB, Bottaro S, Bussi G, Šponer J, Otyepka M, et al. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies [Internet]. *arXiv [q-bio.BM]*. 2016. Available: <http://arxiv.org/abs/1609.08276>
 26. Liu P, Kim B, Friesner RA, Berne BJ. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc Natl Acad Sci U S A.* 2005;102: 13749–13754. doi:10.1073/pnas.0506346102
 27. Wang L, Friesner RA, Berne BJ. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J Phys Chem B.* 2011;115: 9431–9438. doi:10.1021/jp204407d
 28. Bussi G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol Phys.* Taylor & Francis; 2014;112: 379–384. doi:10.1080/00268976.2013.824126
 29. Okamoto Y. Generalized-ensemble algorithms: enhanced sampling techniques for Monte

- Carlo and molecular dynamics simulations. *J Mol Graph Model*. 2004;22: 425–439. doi:10.1016/j.jmglm.2003.12.009
30. Faraldo-Gómez JD, Roux B. Characterization of conformational equilibria through Hamiltonian and temperature replica-exchange simulations: assessing entropic and environmental effects. *J Comput Chem*. 2007;28: 1634–1647. doi:10.1002/jcc.20652
 31. Bitetti-Putzer R, Yang W, Karplus M. Generalized ensembles serve to improve the convergence of free energy simulations. *Chem Phys Lett*. 2003;377: 633–641. doi:10.1016/S0009-2614(03)01057-1
 32. Perez A, MacCallum JL, Dill KA. Accelerating molecular simulations of proteins using Bayesian inference on weak information. *Proc Natl Acad Sci U S A*. 2015;112: 11846–11851. doi:10.1073/pnas.1515561112
 33. MacCallum JL, Perez A, Dill KA. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. *Proc Natl Acad Sci U S A*. 2015;112: 6985–6990. doi:10.1073/pnas.1506788112
 34. Chapman EG, Costantino DA, Rabe JL, Moon SL, Wilusz J, Nix JC, et al. The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science*. 2014;344: 307–310. doi:10.1126/science.1250897
 35. Zhao B, Guffy SL, Williams B, Zhang Q. An excited state underlies gene regulation of a transcriptional riboswitch. *Nat Chem Biol*. 2017;13: 968–974. doi:10.1038/nchembio.2427
 36. Flodell S, Petersen M, Girard F, Zdunek J, Kidd-Ljunggren K, Schleucher J, et al. Solution structure of the apical stem-loop of the human hepatitis B virus encapsidation signal. *Nucleic Acids Res*. Oxford University Press; 2006;34: 4449–4457. doi:10.1093/nar/gkl582
 37. Flodell S, Schleucher J, Cronsigt J, Ippel H, Kidd-Ljunggren K, Wijmenga S. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res*. Oxford University Press; 2002;30: 4803–4811. doi:10.1093/nar/gkf603
 38. Deng J, Xiong Y, Pan B, Sundaralingam M. Structure of an RNA dodecamer containing a fragment from SRP domain IV of *lit* Escherichia coli. *Acta Crystallogr D Biol Crystallogr*. 2003;59: 1004–1011. doi:10.1107/S0907444903006747
 39. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2: 19–25. doi:10.1016/j.softx.2015.06.001
 40. Steinbrecher T, Latzer J, Case DA. Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput*. 2012;8: 4405–4412. doi:10.1021/ct300613v
 41. Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys*. 2004;120: 9665–9678. doi:10.1063/1.1683075
 42. Bottaro S, Di Palma F, Bussi G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res*. 2014;42: 13306–13314. doi:10.1093/nar/gku972

43. Bottaro S, Bussi G, Pinamonti G, Reißer S, Boomsma W, Lindorff-Larsen K. Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA*. 2019;25: 219–231. doi:10.1261/rna.067678.118
44. Lemieux S, Major F. Automated extraction and classification of RNA tertiary structure cyclic motifs. *Nucleic Acids Res*. 2006;34: 2340–2346. doi:10.1093/nar/gkl120
45. Leontis NB, Stombaugh J, Westhof E. The non- Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res*. Oxford University Press; 2002;30: 3497–3531. doi:10.1093/nar/gkf481
46. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA*. 2001;7: 499–512. Available: <https://www.ncbi.nlm.nih.gov/pubmed/11345429>
47. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins: Struct Funct Bioinf*. Wiley Online Library; 2009;75: 430–441. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.22253>
48. Whitford PC, Schug A, Saunders J, Hennelly SP, Onuchic JN, Sanbonmatsu KY. Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function. *Biophys J*. 2009;96: L7–9. doi:10.1016/j.bpj.2008.10.033
49. Yu AM, Gasper PM, Strobel EJ, Watters KE, Chen AA, Lucks JB. Computationally Reconstructing Cotranscriptional RNA Folding Pathways from Experimental Data Reveals Rearrangement of Non-Native Folding Intermediates [Internet]. *bioRxiv*. 2018. p. 379222. doi:10.1101/379222
50. Takahashi MK, Watters KE, Gasper PM, Abbott TR, Carlson PD, Chen AA, et al. Using in-cell SHAPE-Seq and simulations to probe structure–function design principles of RNA transcriptional regulators. *RNA*. 2016; doi:10.1261/rna.054916.115

Highlights:

- A new simulation method for simulating RNA 3D Folding is presented
- Secondary structure restraints are incorporated in a 2D grid of simulation replicas
- Creates accurate 3D structures of RNA using limited experimental data
- Correctly captures tertiary motifs such as pseudoknots and non-canonical basepairs
- Substantially more computationally efficient than de-novo RNA folding