HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2023 November 20.

Published in final edited form as:

J Chem Theory Comput. 2022 June 14; 18(6): 3930–3947. doi:10.1021/acs.jctc.2c00065.

Accelerating the Ensemble Convergence of RNA Hairpin Simulations with a Replica Exchange Structure Reservoir

Kenneth Lam[#,%], **Koushik Kasavajhala**[$,%], **Sarah Gunasekera**[&], **Carlos Simmerling**[#,%,$]

[#]Molecular and Cellular Biology, Stony Brook University, Stony Brook, NY 11794, United States

[$]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United State

[%]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States

[&]Program in Biology, Stony Brook University, Stony Brook, New York 11794, United States

Abstract

RNA is a key participant in many biological processes, but studies of RNA using computer simulations lag behind those of proteins, largely due to less-developed force fields and the slow dynamics of RNA. Generating converged RNA ensembles for force field development and other studies remains a challenge. In this study, we explore the ability of replica exchange molecular dynamics to obtained well-converged conformational ensembles for two RNA hairpin systems in implicit solvent. Even for these small model systems, standard REMD remains computationally costly, but coupling to a pre-generated structure library using the reservoir REMD approach provides a dramatic acceleration of ensemble convergence for both model systems. Such precise ensembles could facilitate RNA force field development and validation, and applications of simulation to more complex RNA systems. The advantages and remaining challenges of applying R-REMD to RNA are investigated in detail.

Graphical Abstract

Footnotes

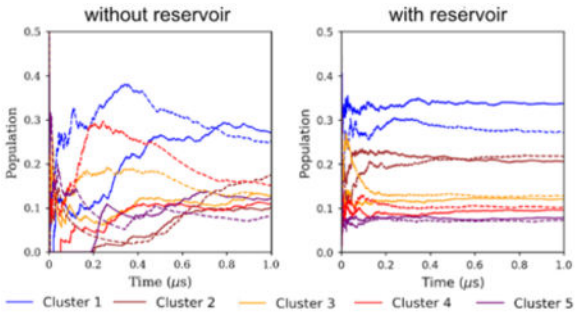
The authors declare no competing financial interest

Supporting Information

Coordinates for the restrained-extended structure are included for the CGUG and AUA hairpins, along with additional details, figures and tables.

CGUG_restrainedExtended.pdb

AUA_restrainedExtended.pdb



Introduction

RNA plays many roles in biology beyond transcription and translation through messenger, transfer, and ribosomal RNA. Ribozymes catalyze reactions, while riboswitches function as gene regulators. Some viruses, such as HIV and SARS-CoV-2, utilize RNA as genetic material. RNA plays roles in various diseases, from viral infections to cancer[1–3]. Due to its function in gene expression and regulation, RNA presents a promising avenue for drug targeting, and is also used as a medical tool. Most notably, several recent SARS-CoV-2 vaccines used mRNA as a vector to produce the spike protein within the body and elicit an immune response, with development of mRNA vaccines for other diseases under way[4]. To better understand the role of RNA in biology, improve the design of RNA-targeting drugs, or use RNA as a drug itself, a clear understanding is needed for RNA structure, dynamics and how RNA interacts with its environment and binding partners.

Experimental studies have been used to study RNA structure and interactions with other molecules, with well over 5000 RNA entries in the PDB as of 2021[5–6]. However, atomic detail structures obtained from experiment are typically static models that do not reveal the dynamics and energetics that can drive interactions between RNA and other molecules. Additionally, RNA structure is more complex than DNA structure due to the presence of the 2’ hydroxyl group, leading to interactions not present in DNA[7–9].

Computational studies have been used to complement experimental studies of biopolymers, providing insight into a range of behavior, including folding, binding, aggregation and catalytic mechanisms.[10] In this study, we focus on MD simulations of RNA. Hairpins and tetranucleotides are often simulated, along with simulations of larger duplexes[11–14]. Simulations can provide a direct connection between the ensemble-averaged experimental observables and the underlying individual conformations and energetics, thereby helping interpret experimental results or suggesting new experiments. For example, Bottaro et al used simulations and NMR data to generate MD ensembles of the UUCG RNA tetraloop that identified a second conformational state which lacked the stacking and base pairing present in the major conformation[15]. For example, studies have been performed on A-form RNA duplexes, examining the effects of substitutions on thermodynamic properties[11]. Other studies[16] have explored base pair stability and fraying, and the interactions between RNA and ions. Tetraloop simulations have been performed to identify native states and the underlying folding mechanisms[14, 17–23]. MD simulations have been used to study various non-coding RNA functions, including binding to target sites as well as interactions with

protein complexes[24]. Larger RNA systems have also been simulated, such as tRNA and ribosomes; the effects of ions on tRNA behavior were explored in 2019[25], and AspRS-tRNA complexes were simulated in 2020[26]. However, both studies used understandably short (50–100 ns) simulations as significant computational power is required to simulate systems of this size. Simulations of large RNA systems that require sampling of global changes remain largely inaccessible through conventional MD due to limitations in computational resources. Various reviews cover the timescales of RNA dynamics with respect to experiments and the techniques used to study them as well as their accessibility to molecular dynamics[7, 27–29].

While system size does dictate RNA simulation feasibility, the accuracy of RNA force fields also presents a key challenge.[7] Protein force fields are relatively mature and can accurately and precisely reproduce experimental structures, successfully folding multiple proteins of various sequences and lengths[30–32], but a similar ability has not yet been demonstrated for RNA. RNA force fields are still immature; many have been developed, yet no consensus as to the best force field exists.

The first modern RNA force field was ff94[33]. The ff99[34] force field iterated upon ff94's sugar puckers. The addition of the bsc0 parameters for α/γ torsions resulted in the ff99bsc0 force field[35]. This force field struggled to reproduce an A-form stem, instead preferring a staggered ladder conformation with a preference for the high-anti χ state[7, 17, 19, 36–37]. Subsequent force fields built on the existing ff99, with the f99χYIL[38] force field rebalancing syn and anti χ dihedrals. The f99bsc0χOL3[36] force field included the bsc0 correction, reparameterizing the χ dihedral to penalize the ladder conformation. These two force fields were trained using QM data on nucleosides and tested on small systems such as the r(GACC) tetranucleotide. Chen and Garcia[19] reparameterized vdW parameters in the ff99 force field in response to overly stable stacking interactions, using tetraloops as their test system. The f99bsc0χOL3_LJBB force field was developed in response to excess base-phosphate interactions and modified the phosphate oxygen vdW radius.[39] For all force fields listed above, testing was primarily done in few systems due to a lack of computational power, and few test systems were shared between studies to allow direct comparisons.

Improvement in computational power has allowed recent force field development to introduce more, larger systems in the testing process. The ff99ROC[40] force field parameterized backbone dihedrals primarily using dinucleotides, and was tested on a larger set of systems than previously used, ranging from tetranucleotides to duplexes. Tan et al[41] parameterized not only dihedrals, but vdW and electrostatic parameters. This force field also was tested on a more diverse set of structures than prior studies, including tetranucleotides, hairpins, and duplexes. Differentiating the Tan et al force field from other force fields was the length of simulations used to test their parameters, with simulations upwards of 30 μs in length.

Well converged simulations of multiple systems facilitate force field evaluation

Few direct performance comparisons between different RNA force fields have been reported. In 2010, a benchmark using sub-μs MD simulations for three tetraloops was reported.[37] In 2015, Bergonzo et al benchmarked modern force fields using a costly M-REMD method[42] applied to tetranucleotides and tetraloops. In 2016, Kuhrova et al

benchmarked force fields using the GAGA tetraloop and multiple enhanced sampling techniques[17]. No conclusions to a “best” force field were drawn in any of these studies, although weaknesses in each force field were identified.

In part, the lack of a consensus on force field accuracy comes as a result of slow RNA conformational sampling. Comparing a simulation ensemble to experiment, or comparing ensembles from different force fields, requires highly precise simulation data such that the uncertainties are smaller than the differences between the data sets. Ideally, these comparisons should be repeated using multiple RNA sequences with diverse structural motifs. To generate precise data across multiple RNA sequences and force fields, considerable sampling must be performed, and within a reasonable amount of time. Several methods for enhanced sampling in MD have been developed, with a few applications to RNA.

Enhanced sampling methods have been applied to RNA simulations. Accelerated MD was used to enhance sampling or torsions in combination with the enhanced sampling method replica exchange molecular dynamics (REMD), showing that convergence was accelerated compared to RNA simulations in the absence of accelerated MD[43]. Other sampling methods, like metadynamics[17–18] and umbrella sampling[44], require more extensive knowledge of the system being studied due to the need to define collective variables (CVs) along which sampling is to be enhanced. CV-based applications to RNA include force field development[17, 40], and studies of energies during conformational shifts of tetraloops[18, 45]. Restricting the amount of conformational space needing to be sampled also can be achieved by incorporating experimental data into simulations, as has been demonstrated for systems such as RNA hairpins[15, 46]. As with accelerated MD, umbrella sampling, and metadynamics, this approach requires additional information (here, NMR data) which may not be available in the general case.

Another approach to improving sampling is through use of implicit solvent. Implicit solvent models[47–49] (such as generalized Born, GB) have been widely used to accelerate conformational sampling in protein systems[50]. Implicit solvent replaces discrete water molecules by directly calculating solvation free energies using only solute coordinates. While potentially less accurate, implicit solvents have advantages. Since the water degrees of freedom are not simulated explicitly, the system has fewer degrees of freedom and the computational cost per time interval is reduced. In addition, solvent-related viscosity is eliminated, thus accelerating the timescale of global motions.[50] Implicit solvent has been used in studies of RNA dynamics, but the approach has not been as widely adopted as the (presumably) more accurate explicit solvents. In 2005, Sorin et al. identified structural water mediated interactions during the folding process in simulations of a GCAA tetraloop using explicit solvent.[51] Their 2002 study of the same system using a GB/SA implicit solvent was missing these interactions, resulting in two-state instead of downhill folding as observed with explicit solvent.[14]

Our GB-Neck2 implicit solvent model[47] has been shown to correctly sample the native structure for proteins of various sizes[30] and reasonably reproduce experimental melting curves for mini-proteins[52]. We also developed GB-Neck2 parameters for nucleic acids[53],

but this and other implicit solvent models for RNA have not been tested as rigorously as protein-specific models. In 2019, Bergonzo et al. refined PDB structures of an A-form helix using GB-Neck2 and various solvent models.[54] They obtained mixed results, where refinement using some types of NMR data was more successful with implicit solvent, while refinement with different NMR data was more successful with explicit solvent. However, GB-Neck2 performed worse when completely unrestrained.

Another solution to increase conformational transition rates is to use an elevated simulation temperature. Such simulations may show improved convergence, but the temperature dependence of thermodynamics alters the ensembles. The results may not be representative of populations at experimentally or biologically relevant temperatures, and may not even contain the native structure. Temperature replica exchange molecular dynamics (T-REMD, here denoted REMD)[55] overcomes this problem by coupling simulations at multiple temperatures, where simulations at high temperatures increases transition rates while the low temperatures provides ensembles at temperatures of interest, which have improved convergence through rigorous Monte Carlo exchanges with the higher temperatures. REMD has often been used to accelerate sampling for both proteins and RNA.[56–59] Another variant of REMD, Hamiltonian REMD (H-REMD)[60], couples replicas with different Hamiltonians instead of temperatures, where the modified Hamiltonians are typically designed to reduce specific energy barriers. REMD has been used in RNA force field development and folding studies.[11, 16, 56]

However, REMD simulations can be computationally costly. Simulations are typically performed using explicit solvent, necessitating large solvent boxes for systems if starting from an extended conformation. The large atom count can lead to a need for many replicas to ensure the adequate potential energy overlap that is needed for successful exchanges[61]. For example, convergence of REMD simulations of small tetranucleotides required over 2μs using 24 replicas.[62–63]

To improve efficiency, REMD variants have been combined. One RNA study[62] used multi-dimensional REMD (M-REMD), combining temperature and Hamiltonian REMD using dihedral parameter scaling in a 2D grid with 192 replicas. RNA sampling using M-REMD was faster than simply using T-REMD alone (300ns to nearly 4μs per replica) with more precise ensembles, however significantly more replicas were used (192 to 24). In terms of aggregate sampling time, M-REMD involved 58.6μs vs 96μs for REMD, but spread over significantly more simultaneous computer instances. The number of computers required makes this method largely inaccessible without a substantial computing facility.

Implicit solvent improves REMD efficiency[61], as fewer replicas are needed to span the same temperatures as an explicit solvent RNA REMD simulation. Thus, fewer computational resources are required. However, as we show in this work, good convergence remains highly challenging even for small RNA hairpin systems with implicit solvent REMD.

There remains a need to improve sampling for RNA, but ideally without significantly increasing the cost, nor requiring extensive prior knowledge of the system dynamics. Here, we test the potential of the reservoir variant of REMD (R-REMD)[64–66] to improve RNA

folding convergence over conventional REMD simulations. Using a reservoir has been shown to significantly accelerate convergence in peptide and protein folding simulations, while maintaining accurate ensembles[66]. R-REMD functions similarly to conventional REMD and is rigorously accurate[64], but adds a pre-generated reservoir of structures that is accessible to the REMD simulation through Monte Carlo exchanges. These exchanges allow instantaneous jumps to an alternate basin, eliminating the wait to cross intervening energy barriers. A reservoir effectively separates the search and scoring components of REMD, where resources can be focused on conformational sampling at high temperature without needing concurrent simulations at lower, less efficient temperatures as in conventional REMD. The costly multi-replica temperature ladder of REMD is used only for refinement and reweighting of the structures, but not while sampling global motions. Unlike other REMD variants such as M-REMD, or methods like metadynamics and umbrella sampling, R-REMD does not require CVs or significant prior knowledge of the system being simulated.

Two variants of R-REMD have been reported, with different weighting of the conformational ensemble in the reservoir. The original method uses a Boltzmann-weighted reservoir (B-R-REMD)[64], and the other a non-Boltzmann weighted reservoir (NB-R-REMD)[65]. In this study, we focus on B-R-REMD. With B-R-REMD, the reservoir is a converged, Boltzmann-weighted set of structures, typically generated via MD at a single high temperature.

R-REMD has previously been applied to RNA. In 2013, Henriksen et al used R-REMD to generate a converged set of structures for the small tetranucleotide r(GACC) in explicit water[63]. However, this implementation of the R-REMD method employed a Boltzmann-weighted reservoir but using the exchange probability appropriate for a non-Boltzmann-weighted reservoir. Nevertheless, the group successfully generated converged ensembles.

The present work seeks to develop best practices for R-REMD simulations on larger systems such as RNA hairpins. One significant challenge is testing the accuracy of the ensembles generated by R-REMD, isolated from inaccuracies in the force field that would convolute such evaluations if experimental data were used as the benchmark. This requires generation of well-converged ensembles using conventional simulation methods, but on non-trivial systems. In the proof-of-concept study presented here, we generate well-converged reference ensembles for two RNA hairpin systems using the GB-Neck2 implicit solvent model; these remain highly challenging in implicit solvent, and largely unfeasible in explicit solvent. We evaluate convergence of the entire ensemble via precision estimates on each conformational cluster rather than the single most-populated structure. These data sets provide the opportunity to rigorously quantify both the accuracy of R-REMD and potential advantages in terms of reduced computational cost. We test sensitivity of the results to several protocol choices that influence REMD, as well as additional factors specific to R-REMD. We demonstrate that the R-REMD ensembles do indeed reproduce the structures and populations of those sampled using conventional REMD. While not fully optimized, we show that R-REMD can significantly improve convergence of RNA ensembles compared to conventional REMD. Although RNA force field evaluation is not the focus of the present study, we compare our RNA hairpin ensembles to experimental structures as an example

of how the R-REMD protocol can facilitate identification of possible weaknesses in a force field / water model combination.

Materials and methods

System preparation and force field

Two small RNA model hairpin systems were selected, with sequences GGGCGUGCCC (CGUG tetraloop, PDB: 1IDV[67]) and GGUGCAUAGCACC (AUA triloop, PDB: 1ESH[68]). Two initial conformations were generated for each single-stranded sequence: the native structure from the PDB and an artificial, extended conformation (a single-stranded RNA adopting an A-form conformation). Per the PDB file, model 10 of the 10 submitted conformers of the 1IDV PDB ensemble was the best representative conformer and was designated as the native structure. Only one experimental model was available for the AUA hairpin. For each system, the extended conformation was generated by loading the sequence into the Amber *nab*[69] program and creating an A-form duplex. The complementary strand to the native sequence was then deleted.

Common simulation protocols

Amber simulation systems were generated for these initial structures using the *tleap* module. The ff99bsc0χOL3[36] parameters were used, along with GB-Neck2[53] implicit solvent combined with the mbondi2 intrinsic radii set[49]. A salt concentration of 0.1 M was used. Temperature was controlled with a Langevin thermostat with a collision frequency of 0.1/sec. SHAKE was employed to restrain bonds involving hydrogen. Hydrogen mass repartitioning[70] was performed to allow a 4 fs timestep. Minimization was performed for 1000 steps using the steepest descent algorithm. Because systems were expected to unfold and maintaining a stable initial structure was not the goal, further equilibration was not performed. Unless specified otherwise, these minimized structures were used as the initial coordinates in subsequent simulations.

REMD simulations

Standard REMD simulations included a temperature ladder with 6 replicas (Table S1). Temperatures were generated using an in-house script, estimating potential energy overlap based on number of atoms and desired exchange acceptance ratio (0.3). The 356.0 K temperature for the AUA hairpin was empirically adjusted from the suggested 353.2 K to increase number of transitions from low to high RMSD. Exchanges were attempted each 1 ps, with 10000000 exchange attempts for REMD (10 μsec) and 1000000 attempts for R-REMD (1 μsec).

Initial structures for REMD simulation

Two sets of REMD simulations (where a “set” is defined as one simulation run from the extended conformation, and one from the native conformation) were run: one set unrestrained, and one set with an added end-to-end distance restraint. The initially-extended structure for restrained simulations is described below.

End-to-end distance restraint

For some simulations, a flat-well harmonic distance restraint was imposed between the O5' and O3' on the first and last residues, respectively. For the CGUG hairpin this included O5' of residue 1 to O3' of residue 10, and for the AUA hairpin involved O5' of residue 1 to O3' of residue 13. A force constant of 50 kcal/mol was used between 0 and 1 Å, and between 20 and 25 Å. No penalty was present between 1 and 20 Å.

Since the extended initial conformation did not satisfy the restraint, an alternate starting conformation ("restrained-extended") for restrained runs was selected randomly from trajectories obtained from unrestrained REMD, with the criteria that no base pairing was present and the end-to-end distance was near, but not over, 20 Å. These coordinates are provided as Supporting Information.

MD simulations to evaluate potential energy distributions and select reservoir temperature

To assess the dependence of potential energy on structure for the CGUG hairpin, 2 μs MD simulations were generated at each of the top four temperatures in the REMD ladder, aiming to identify a temperature where there was a mix of folded and unfolded structures. End-to-end distance was selected as a measure of structure, as it roughly tracked the formation of a folded stem. The same simulations were used to select a reservoir temperature.

MD simulations to generate reservoirs

Ten sets of 10 μs simulations were run for each hairpin system; 5 each were run from the native and the restrained-extended conformations. Thermostat temperatures corresponded to a temperature from the REMD ladder; 361.3K for CGUG and 356.0 K for AUA (see R-REMD section in Results for rationale). End-to-end distance restraints were applied as described above. Structures were saved every 10ps (1,000,000 from each of the 10 runs).

R-REMD reservoir file generation

MD simulations were split into two groups, one including 3 simulations starting from the extended conformation and 2 starting from the native, and the second with the remaining simulations (2 extended and 3 native), denoted reservoirs 1 and 2 respectively. For each set, a reservoir was generated by extracting every 10000^{th} structure, evenly spaced, for an aggregate of 50000 structures per reservoir. Previous reservoir REMD simulations with a small protein showed that 10000 structures were sufficient[66]; here we increased this number to account for the possibility of greater diversity in RNA ensembles as compared to short proteins. Potential energies for each structure in the reservoir were calculated using the Amber *sander* program using the same force field options as described above. The *createreservoir* command of *cpptraj*[71] was used to merge the coordinates and energies into reservoir files.

R-REMD simulations

For each reservoir, two sets of R-REMD simulations were run, one starting from the restrained-extended conformation, and one from the native conformation. The temperature ladder for R-REMD included the lower four temperatures from the conventional REMD

ladder. R-REMD simulations were run for 1 μs, attempting exchanges every 1 ps. As with any pair of temperatures in the Amber REMD implementation, exchange attempts between the highest-temperature replica and the reservoir were performed every other exchange attempt. REMD inputs were otherwise the same as those for non-reservoir REMD. End-to-end distance restraints were imposed as described above.

Extracting trajectories

During REMD simulations in Amber, neighboring replicas exchange thermostat temperatures rather than structures, resulting in simulation trajectories that sample multiple temperatures. This continuous time-dependent evolution of the structure is referred to as a replica trajectory. Because each structure in an REMD replica trajectory also contains information about the current thermostat temperature, it is possible to extract temperature "trajectory" data sets for that include the structures sampled at each time point from the replica that adopts a specific thermostat temperature. Temperature trajectories were extracted using the *cpptraj* program with the *remdtraj* command. Once converged, these spatially discontinuous "temperature trajectories" correspond to the thermodynamic ensembles of the system at each temperature. R-REMD replica trajectories also can be converted to temperature trajectories; R-REMD temperature and replica trajectories have additional discontinuities when a replica exchanges with the reservoir.

Analysis of structure properties

RMSD values were calculated using *cpptraj*, using the native PDB coordinates as the reference structure for its respective system. All heavy atoms were selected (residues 1 – 10 for the CGUG hairpin, and 1 – 13 for the AUA hairpin). Histograms were normalized and used 50 bins equally spaced over the range of data.

End-to-end distances were calculated using *cpptraj* using the same atoms involved in the distance restraint described above. Time series plots and histograms were generated using python with matplotlib libraries.

Refolding rates

Transitions between low (folded) and high (unfolded) RMSD were counted, referred to as a refolding event. Low RMSD was defined as below 3 Å, and high RMSD was defined as above 9 Å, based on the distributions observed in CGUG and AUA hairpin RMSD plots (Figures S1 and S2 respectively). A refolding event was defined as a time where a replica trajectory that had sampled RMSD values above 9 Å switched to a value below 3 Å, or vice versa. These events are referred to in the text as refolding events. Total refolding event counts were converted to a refolding rate to obtain a value independent of simulation length.

Base pair counting

The number of base pairs in a given structure was calculated using the *nastruct* command in *cpptraj*, using the native PDB structure for each hairpin as a reference structure and otherwise using default options.

Cluster analysis

Cluster analyses were performed on temperature trajectories using the *kmeans* algorithm implemented in *cpptraj*, with a random seed of 1, targeting 100 clusters for 300 K ensembles. Increasing the number of clusters from 100 to 200 showed little qualitative difference (Figure S3). The number of clusters was increased to 1000 at higher temperature to account for increased structure diversity for the unfolded RNA. Clustering was performed on all heavy atoms. To facilitate comparison of populations between different simulations, common cluster definitions were created for each RNA sequence. All 300 K ensembles of the same sequence were clustered together, and populations of each cluster in individual ensembles were calculated separately. The *splitframe* option in cpptraj was used to separate ensembles by their number of frames, and the resulting output from the *summaryhalf* command was used to parse cluster populations for individual ensembles. Cluster populations were compared across data sets, with the line of best fit and R^2 generated using standard *python*. Time-dependent populations of the top 5 clusters for each ensemble were extracted from the cluster analysis population outputs, and plotted on a time series with *python*.

Simulations in explicit water

Additional standard MD simulations were carried out for the AUA hairpin in explicit water, using the OPC[72] water model. Three systems were built, using the ff99bsc0χOL3 force field with no counterions, ff99bsc0χOL3 with 12 neutralizing Na^+ ions[73], and ff99bsc0χOL3_LJBB with 12 Na^+ ions. The 1ESH PDB file was loaded into tleap and solvated using a minimum distance of 8 Å between solute atoms and the box boundary, resulting in addition of ~ 1600 water molecules. Na^+ ions were added using the addionsrand command in tleap. The system was minimized for 10000 steps with constant NTV and positional restraints on RNA heavy atoms and a force constant of 100 kcal/mol $Å^2$, with a direct space cutoff of 8 Å and particle mesh Ewald for long-range electrostatic interactions. Next, each system was heated from 100 K to 298 K over 1ns, maintaining the same restraints and other options while using a 1 fs time step. Next, the system volume was equilibrated at constant NTP over 1 ns with the same restraints. Next, the same conditions were continued for 1 ns with restraint force constant reduced to 10 kcal/mol $Å^2$. Restraints were released on all atoms except phosphorus, with 1000 steps of minimization followed by three consecutive steps of 1 ns MD at NTP with restraint force constant of 10, then 1, then 0.1 kcal/mol $Å^2$. A final 1 ns MD run was carried out with no restraints. This was followed by fully unrestrained production runs of 500 ns at 310 K, with a 4 fs time step enabled by hydrogen mass repartitioning[70].

General analysis and visualization

VMD[74] was used to generate all structure figures. Histograms, line plots, and scatter plots and corresponding best fit lines and correlations were generated using python with matplotlib libraries.

Results and Discussion

Inclusion of a structure reservoir in REMD simulations adds additional complexity compared to conventional REMD, with several aspects that could influence the precision and accuracy of the resulting ensemble. Possible challenges with R-REMD include:

1. Does including a reservoir accelerate REMD convergence for RNA?
2. Is the R-REMD low-temperature ensemble sensitive to the initial structure (as with conventional REMD)?
3. Does the precision of the reservoir have an impact on the resulting low-temperature ensembles?
4. Does R-REMD provide accurate ensembles with respect to conventional REMD?
5. How often should the replicas exchange with the reservoir?
6. How can a good reservoir be generated at the lowest possible computational cost? (T, length, multiple MD…)

In this study, we address points 1 through 4 to assess the applicability of R-REMD to RNA folding. For each point, detailed comparisons of structure ensembles at 300 K between R-REMD and conventional REMD simulations were performed. Points 5 and 6 pertain to important refinements of the method to reduce computational cost, but are beyond the scope of this article.

Model systems

In order to reduce the possibility that measured R-REMD benefits are not anecdotal or system-specific, two RNA hairpin model systems were chosen, one with a CGUG tetraloop (10 bases total) and the other with an AUA triloop (13 bases) (full sequences and PDB IDs as well as preparation can be found in Methods). The AUA system has a longer stem, as well as an AU base pair, while the CGUG system stem is shorter and comprised of only GC pairs.

Evaluation metrics for accuracy, precision and efficiency

In order to evaluate the accuracy of a simulation method, it is important that the final ensembles be precise (reproducible, with uncertainties that are smaller than differences between the methods being compared). Furthermore, evaluating the potential speedup from including a reservoir requires estimation of how long in simulation time was required for these ensembles to attain a defined precision.

Our method comparisons use four metrics: folding reversibility (refolding rate), ensemble precision, ensemble convergence time, and ensemble accuracy. We define convergence time here as when simulations reproducibly plateau near their equilibrium cluster populations.

The most qualitative measure, *refolding rate*, refers to how frequently each replica trajectory transitions between folded and unfolded structures (see Methods). These trajectories track individual replicas that have continuous structure sampling, while the thermostat temperature varies as REMD exchanges occur. These REMD exchanges help drive

refolding.[75] We quantified refolding efficiency by calculating the average per-replica "refolding rate" (in μs^{-1}) from below 3 to above 9 Å (or vice versa) RMSD to the native PDB structure (see Methods for details on calculation).

The RMSD to a single native reference structure is a poor metric for gauging the reproducible sampling of the broader energy landscape. We quantify *ensemble precision* here using the low-temperature (300 K) populations of a set of clusters that are communally defined among the simulations being compared.[61] Independent simulations producing the same cluster populations provides additional reassurance that the final ensemble is converged. Precise populations then allow us to quantify external impacts on the ensemble accuracy, such as whether the ensemble obtained using a reservoir (R-REMD) matches the reference REMD data, and whether using a different reservoir results in a different converged ensemble.

While the cluster population comparison indicates *if* simulations are converged, *ensemble convergence time* assesses how much simulation time was required to achieve the convergence. Time to converge is defined here as the estimated time required for cluster populations to plateau at a reproducible value, when visualized on a time-resolved plot. We expected that the timescale to converge an RNA system would be reduced when the reservoir was included. More importantly, for the method to have value we expect R-REMD to be able to converge in cases where conventional REMD remains too challenging.

The fourth metric, *ensemble accuracy*, involves validation of R-REMD ensembles by comparison to precise reference ensembles obtained using conventional REMD, using otherwise identical protocols and models, and thus eliminating any convoluting factors that are present when comparing to experiments. If R-REMD converges quickly but generates the wrong ensemble at low temperatures, it would not be a viable protocol. Clustering R-REMD ensembles concurrently with the reference ensembles enables direct comparison of populations[61].

Simulations of the CGUG tetraloop RNA hairpin

Generating reference data with conventional REMD—We begin with the CGUG tetraloop system, since it is smaller and thus had fewer degrees of freedom to sample than the AUA system. We generated data using unbiased conventional REMD of 10 μs for each of the six temperatures (see Methods for full details). Precision was estimated via two independent REMD simulations initiated from different conformations. The data were used to determine refolding rates, cluster populations, and convergence time.

Assessing folding reversibility

To calculate the refolding rate for each REMD replica, we calculated time-dependent heavy atom RMSDs to the PDB reference structure using replica trajectories (Figure 1a). Each replica shows multiple refolding events. Replica 6 shows periods of several μs when only high RMSD structures are sampled, while the other five replicas appear less trapped. Simulations starting from the native structure exhibited similar behavior (Figure S4). Histograms of RMSD values show varying peak heights and locations for different

replicas (Figure S1); two replicas in initially extended simulations sampled a 6 Å peak not present in other replicas. Despite inconsistent peak heights and presence, replicas generally sampled the same range of RMSDs. The average refolding rate across all replicas is 2.4 μs^{-1}, with the value for each replica indicated in Figure 1a. As expected, the average rate in simulations starting from the native conformation was very similar at 3.7 μs^{-1} (Figure S4).

Quantifying ensemble convergence with cluster analysis

We compared ensembles at 300 K obtained from the two independent conventional REMD simulations of the CGUG hairpin (Figure 2a). Ideally, rank ordering and fractional populations of clusters would match between these independent runs. Both prefer the same most-populated structure, although the absolute population of the top cluster shows a modest variation with 0.15 fractional population in simulations starting from the extended conformation vs. 0.22 in simulations from the native conformation.

One cluster, with fractions of 0.067 and 0.004 in initially-extended and initially-native simulations respectively, appears to be an outlier with reduced precision compared to the others. Poor precision for this cluster reinforces the observations of inconsistent peaks at ~ 6 Å in per-replica histograms of RMSD relative to the native structure (Figure S1). To explore if this peak corresponds to the outlier cluster, RMSDs were calculated for each replica using the outlier cluster as the reference (Figure S5). Two replicas in the initially-extended REMD continuously sample this cluster for periods of over 1 μs, while other replicas do not sample the structure at all. These data support that the disparity in population in the two ensembles is due to inconsistent sampling between replicas.

While there is modest spread in cluster populations along the diagonal, clusters other than the outlier demonstrate reasonable agreement and the remaining errors in cluster populations are relatively low (Figure 2a). We concluded that the two ensembles were reasonably converged with respect to the top clusters.

Quantifying convergence as a function of simulation time

We calculated the time-dependent populations of the top five largest clusters, including all structures sampled up to that point in time (Figure 3a). We compared these time-dependent cluster populations across independent runs, expecting the differences in populations between simulations to become smaller as the simulated time progressed, and reaching a plateau as the ensembles converged.

Populations approach their final equilibrium values after ~ 6 μs, but with the top cluster still showing an uncertainty range of ~ 10%. This is consistent with the time-dependent RMSD data in Figure 1a which shows that replicas can be trapped on the μs timescale. The rank ordering of the top clusters appears to take ~ 8 μs to become reproducible, and does not change thereafter. We concluded that simulations were reasonably converged beyond 8μs, with some remaining uncertainty in the absolute populations.

Employing a distance restraint to reduce sampling of less important basins

The time-dependent RMSDs and the refolding rates (Figure 1a) suggest that replicas can spend ~ 1 μs or more exploring the diverse unfolded landscape between folding events. We aimed to reduce this time interval to improve statistical sampling of basins that are more relevant at lower temperatures. Protein folding studies suggest that long-distance contacts can be slow to form; indeed, RNA hairpin folding dynamics are a slow process with some conformational changes on the order of μs and longer[20–23, 76].

To characterize the diversity of structures in unfolded basins, end-to-end distances (see Methods) were calculated for the 384.3 K REMD ensemble. A broad range of values up to 52 Å is observed (Figure 4a). Ensembles sampled in REMD at lower temperatures give a narrower distribution that is strongly shifted toward the ~16 Å value calculated for the NMR-based conformation (Figure 4c). The broadening of distance distributions with increasing temperature suggests that the conformational space to be explored may increase dramatically with temperature, forming an entropy bottleneck that can counteract the ability of REMD to cross enthalpy barriers more efficiently. Such anti-Arrhenius behavior has been shown to impede convergence in REMD simulations of protein folding[77]. This suggests that extensive sampling of the very extended RNA structures driven by the elevated temperature in REMD is counterproductive to improving convergence at lower temperatures where they are not expected to make a significant contribution.

Prior work has shown that introduction of knowledge-based restraints can improve sampling of relevant parts of the conformational landscape and speed convergence for simulations of protein folding and binding[32, 78–79]. For the RNA model systems here, it seems reasonable to assume that the terminal nucleotides would be close in space for a hairpin structure motif. We decided to introduce an end-to-end restraint between the two terminal residues that would allow the stem to fully unzip to facilitate rearrangements in the loop, but also reduce sampling of very extended structures that are less relevant at lower temperatures. This differs from prior work[42] that employed base pairing distance restraints for the first few base pairs in a stem to restrict conformational changes to the loop region only. Here, no assumptions are made about base pairing or stem geometry other than assuming general proximity of the terminal oxygens.

REMD ensembles at 300 K primarily contained structures with end-to-end distances near 16 Å, the value seen in the NMR-based structure (Figure 4c). In contrast, unrestrained REMD and MD at 384.3 K showed sampling of structures with distances shifted to much longer values (Figures 4a and 4b). An ideal restraint would reduce sampling of these long distances; a target maximum distance of 20 Å was selected, as it encompassed 94% of structures shown in Figure 4c. Indeed, sampling of structures near 16 Å increased significantly in restrained MD simulations at 384.3 K (Figure 4d) compared to the unrestrained REMD and MD. Any residual effects of including this restraint are tested in the *ensemble accuracy* section. While not explored here, future protocol optimization could assess the impact of using longer target distances, or weaker force constants.

This end-to-end distance restraint was employed in conventional REMD simulations (“restrained REMD”), keeping all other parts of the protocol identical to the unrestrained

REMD described above. Two independent simulations were carried out, again starting from extended and native initial conformations (see Methods for the description of the extended structure for restrained simulations). We expected the improved sampling of experimentally-relevant distances at high temperature to lead to improvements in folding reversibility, ensemble precision, and ensemble convergence time.

Time-dependent RMSD values relative to the native PDB structure for individual replicas are shown in Figure 1b for restrained REMD starting from the extended conformation (initially-native REMD data are shown in Figure S6). The highest RMSD sampled in all replicas is ~ 10 Å, compared to 15 Å in unrestrained REMD. This is expected since the restraints prevent sampling of the very extended structures that lead to high RMSD values. Refolding rates show a modest 2-fold increase compared to unrestrained REMD, with an average rate of 4.7 μs^{-1} (5.1 μs^{-1} in initially-native simulations). This increase suggests more efficient sampling per replica in comparison to unrestrained simulations. Overall, less kinetic trapping is present, although periods over 1 μs with no refolding events remain.

RMSD histograms also demonstrate moderate improvements in sampling compared to unrestrained REMD. While unrestrained REMD showed inconsistent peak presence and heights between replicas, replicas using the distance restraint exhibit more uniform profiles, though peak heights remain inconsistent (Figure S7).

To quantify improvement in ensemble convergence with the addition of the end-to-end restraint, cluster populations for the two restrained REMD 300 K ensembles were compared. Good agreement is observed, with populations close to the diagonal (Figure 2b). A single cluster is preferred in unrestrained REMD (Figure 2a), but with inclusion of the restraint, the two top clusters show similar fractional populations (within 0.03). Unlike unrestrained REMD, no outliers are observed and none of the clusters are highly populated in only one run.

To assess ensemble accuracy, and the potential influence of the end-to-end distance restraint, cluster populations at 300 K from restrained REMD were compared to those obtained from unrestrained REMD. It appears that the restraint did not introduce any bias into the ensemble; no clusters present in unrestrained REMD were missing in restrained REMD, and the average populations show good match, with the same rank ordering and all points near the diagonal (Figure 2d). It is apparent that unrestrained REMD simulations are less precise; the uncertainties observed in unrestrained REMD are significantly reduced in restrained REMD, with the average populations in restrained REMD falling within error bars of unrestrained REMD. We concluded that the unrestrained and restrained ensembles are essentially equivalent within their precision, and that application of the end-to end distance restraint on this RNA hairpin provided improved precision but had little discernable impact on accuracy.

Based on the improved precision at 10 μs, we expected the restrained REMD simulations to have converged more quickly than unrestrained REMD, resulting in a faster ensemble convergence time. Time-dependent cluster population analysis (Figure 3b) suggests that including the distance restraint somewhat reduces the time required to obtain precise

populations, from ~8 μs in unrestrained REMD down to ~ 6 μs with the end-to-end restraint. However, the uncertainty in rank ordering between the top two clusters persists at 10 μs, and the population for cluster 1 exhibits an upward drift in both runs of restrained REMD (Figure 3b).

Overall, inclusion of the end-to-end restraint appears to moderately improve the precision of the entire ensemble in restrained REMD as compared to unrestrained REMD for the same simulation timescale. A significant improvement is seen in the convergence of the outlier cluster that was poorly converged in unrestrained REMD. Therefore, the restrained REMD data were designated as the reference ensemble for evaluating the applicability of the reservoir approach for RNA.

REMD simulations using a reservoir (R-REMD)

We aimed to generate a well-converged set of structures at elevated temperature, and determine if it would accelerate subsequent R-REMD simulations while maintaining accuracy. First, two independent reservoirs were generated so that we could evaluate the impact of reservoir variability on the final R-REMD ensemble. Next, each reservoir was coupled to two independent R-REMD simulations initiated from different structures (4 R-REMD simulations total), as was done for the conventional REMD simulations described above. The end-to-end distance restraint was included in reservoir generation and during all R-REMD simulations, to be consistent with the protocol used to obtain the reference ensemble.

Generating reservoirs for R-REMD

The R-REMD variant tested in this work requires the reservoir populations to be consistent with a Boltzmann-weighted ensemble at a defined temperature.[66] Selecting the reservoir temperature is a non-trivial step. At an ideal temperature, sampling is diverse and rapid, but structures relevant to lower temperatures should still be present in the reservoir for seeding replicas during the subsequent REMD phase. The top four temperatures in our conventional REMD ladder were considered by running 2 μs MD simulation at each temperature using 20 Å end-to-end restraints. Extensive sampling of compact and extended structures (indicated by the end-to-end distance, Figure S8) at 361.3 K suggests that this is a reasonable temperature for reservoir generation.

Reservoir generation was performed using 10 MD simulations at 361.3K, each 10 μs long using a 20 Å end-to-end restraint. 10 μs was chosen because the conventional REMD simulations were converged on this timescale, and we aimed to ensure a fully converged reservoir to permit identification of any possible accuracy errors arising from coupling to a reservoir. Five simulations were started from the native conformation, and five from the restrained-extended structure. Snapshots from each set of five independent MD simulations were used to build an independent reservoir. A detailed description of the reservoir generation procedure is provided in Methods. Cluster analysis and population comparison between the two independent long reservoirs indicates that the two reservoirs represent comparable ensembles and are well-converged (Figure S9).

Since each reservoir was generated via 50 μs aggregate MD simulation, compared to the 60 μs aggregate simulation to converge REMD without the reservoir, the overall computational cost for our R-REMD simulations is similar to that of our conventional REMD. While 10 μs per MD simulation was selected to ensure reservoir convergence, we expect that the reservoirs likely converged earlier. This was tested by generating reservoirs using only the first 5 μs of the same MD simulations, for an aggregate of 25 μs, referred to as the short reservoirs; those from the aggregate 50 μs are referred to as the long reservoirs. Cluster populations correlate well between the long and short reservoirs (Figure S10), indicating that the reservoir populations were likely converged prior to 25 μs aggregate sampling, potentially reducing reservoir generation time by half. Future work will explore how much these simulations can be shortened while still providing a converged reservoir.

A broad range of clusters is present at 361.3 K when comparing populations between the two long reservoirs (Figure S9), but a single cluster shows approximately double the population than the next most populated, comprising nearly 1% of each reservoir. If this cluster was the same as that preferred at low temperature, then high-temperature simulations could be a strategy to quickly predict the preferred conformation. However, the cluster preferred at 361.3 K is a poor match to any of the top five clusters at 300 K, with RMSD values of at least 5.0 Å (Table S2).

Workflow for validating R-REMD results

A Boltzmann-weighted reservoir replaces the corresponding simulation in the temperature ladder, and only the lower temperatures are retained.[64] Here, the temperature ladder includes the lower 4 temperatures in the original REMD ladder (282.0, 300.0, 319.2, and 339.6 K), coupled to the reservoir at 361.3 K. Each of the four R-REMD simulations was run for 1 μs, as compared to 10 μs for conventional REMD, since the reservoir was expected to accelerate convergence.

R-REMD simulations were evaluated using the same metrics applied to conventional REMD (refolding rate, ensemble precision, and ensemble accuracy vs. the reference ensemble, and ensemble convergence time). Refolding events in R-REMD are driven largely by MC exchanges with reservoir structures, and do not depend on individual replicas crossing energy barriers as in conventional REMD methods. Thus, we expected refolding rates to be significantly higher than in conventional REMD, improving statistical sampling. Based on similar work using small proteins[66], we also expected R-REMD to converge to the same ensemble as conventional REMD (maintain accuracy), with equal or better precision on a shorter timescale (more efficient). Once precision is quantified, the accuracy of the cluster populations from R-REMD can be validated against the reference data.

R-REMD has an additional factor not relevant in conventional REMD. Precision in R-REMD can be influenced by both the starting structure and the composition of the reservoir. Both aspects of R-REMD precision were tested here. For each system, four simulations were carried out: two independent reservoirs were generated as described above, and for each reservoir, two R-REMD simulations were run starting from different initial conformations. Since all 300 K ensembles were clustered together and thus share cluster definitions, cluster populations can be compared directly between R-REMD simulations that vary the starting

structure, the reservoir, or both. We expected that any population uncertainties would originate from reservoir differences or simulation length, since the initial structures should rapidly exchange with the reservoir and not provide a lasting impact on the simulation. Since the reservoirs are well-converged (Figure S9), R-REMD simulations using independent reservoirs should generate the same low-temperature ensembles.

Reservoir REMD simulations converge more quickly than conventional REMD simulations

Time-dependent RMSD values for each replica are shown in Figure 1e for R-REMD (from the extended conformation, using reservoir 1, with the other R-REMD simulations in Figure S11). The first 1 μs of corresponding unrestrained and restrained conventional REMD are presented in Figure 1c and 1d for comparison. On this timescale, the kinetic trapping in the conventional REMD data is even more apparent than in Figures 1a and 1b, which include all 10 μs. As expected, the replicas in R-REMD show markedly improved sampling across the entire RMSD range, and per-replica RMSD histograms (Figure S12) show uniform sampling of RMSDs across all replicas.

R-REMD shows substantially higher refolding rates as compared to either unrestrained or restrained REMD, since the MC jumps using the reservoir permit instantaneous switching between folding or unfolded basins. The average refolding rate is 61 μs^{-1}, which is 13 times greater than that of the corresponding restrained REMD. The R-REMD simulation starting from the native structure using the same reservoir also averages 61 μs^{-1}, and simulations using reservoir 2 averaged 63 and 61 μs^{-1} (Figure S11), showing little overall dependence on the reservoir or starting structure.

Precision via sensitivity to 1) starting structure and 2) reservoir is assessed in Figure 2c, which depicts cluster populations from all four R-REMD 300 K ensembles. Populations for the two independent runs using a single reservoir were averaged, with error bars reflecting the impact of the starting structure; the small error bars suggest that the initial structure has little impact on the final populations. These values are compared to the average values from the alternate reservoir, again using two independent starting structures.

There is also good agreement in most populations between R-REMD ensembles generated using different reservoirs. Average cluster populations for the majority of clusters are near the diagonal. However, reduced precision was observed for one cluster in R-REMD, suggesting that the populations for this cluster were modestly sensitive to reservoir composition. The relatively small error bar suggests that the difference in population is a result of the reservoir composition, not the starting structure or simulation length. These analyses demonstrate that the R-REMD ensembles are reproducible, with little impact from different initial structures and a modest impact for one cluster when switching to a different, independently generated reservoir.

R-REMD structure ensembles reproduce the reference ensembles

The reproducibility of cluster populations discussed above reinforces that R-REMD can generate precise ensembles. However, it is important that these ensembles also are accurate and reproduce the reference data (ensemble accuracy). Since the method is formally rigorous, we expected cluster populations in R-REMD to agree with those in REMD.

However, an incomplete reservoir could lead to rapid convergence to an ensemble that lacks structures that are significantly populated in the reference data. Errors in the reservoir could also lead to a systematic bias in the low-temperature ensembles sampled in the REMD step of R-REMD. We did not expect these to be issues here due to the good convergence of the simulations used to generate the reservoirs (Figure S9), and the low sensitivity of the results to switching the reservoir.

To evaluate R-REMD accuracy, the final R-REMD cluster populations at 300 K were compared to the reference data (from restrained conventional REMD), enabled by the common cluster definitions (Figure 2e). The deviation from the diagonal (perfect match) is reasonable, especially with consideration of the error bars. Notably, the error bars for the lower-precision cluster in R-REMD cross the diagonal, indicating that population of the cluster in the reference data falls within the range from R-REMD. Both REMD methods select the same two most populated clusters, with slightly higher population in R-REMD. However, the top cluster exhibited upwards drift in population vs time at the end of the restrained conventional REMD, suggesting that the small error bars for the reference data may underestimate the true uncertainty, and the apparent inaccuracy of the R-REMD population may result from imprecise reference data.

Analyzing R-REMD convergence speed vs. conventional REMD

After determining that R-REMD provides reasonably precise ensembles with good accuracy compared to conventional REMD, the ensemble convergence time for R-REMD was analyzed. Populations of the top five clusters were calculated as a function of time for two R-REMD simulations using different reservoirs (Figure 3d). Compared to conventional REMD, the cluster populations reach plateau values much earlier in the simulation, near 0.5 μs. At the same timescale, conventional restrained REMD does not even show correct rank order (Figure 3c), and uncertainties in cluster populations are larger than the differences between clusters. Permitting MC exchanges to a reservoir appears to dramatically improve REMD convergence for this RNA system.

Assessing generalizability: application to an RNA hairpin with AUA triloop

The REMD protocols described above were repeated for an RNA hairpin with an AUA triloop and a 5 base-pair stem containing an AU pair. Unrestrained conventional REMD simulations for AUA remained very poorly converged, with refolding rates less than 1 μs^{-1} (lower than CGUG), and dramatic dependence of cluster populations on the initial structure even after 10 μs per replica (Figures S13, S14 and S15). Given this poor convergence, unrestrained simulations were not explored further (refer to SI for additional discussion and data).

AUA triloop restrained conventional REMD simulations show poor per-replica sampling but reasonable ensemble precision

Following the protocol used for CGUG, a 20 Å end-to-end distance restraint was added to the conventional REMD simulations for the AUA triloop system (see Methods). We used the same four metrics (refolding rate, ensemble precision, ensemble accuracy, and ensemble

convergence time) as for CGUG. Two conventional REMD simulations were run, one starting from the restrained-extended conformation, and one from the native PDB structure.

Addition of the distance restraint resulted in a refolding rate of 0.5 μs^{-1} in initially-extended simulations (Figure 5a and Figure S16) and 0.8 in initially-native simulations, a 3-fold improvement over unrestrained REMD when comparing initially-extended simulations. Replicas still spent multiple μs trapped in low or high RMSDs, with inconsistent refolding rates (Figure 5a); RMSD histograms confirm inconsistent sampling across replicas, particularly near 5 – 10 Å (Figure S2).

Despite the relatively few refolding events and poor RMSD profile consistency, cluster populations at 300 K for restrained REMD (Figure 6a) show reasonable precision, and substantial improvement compared to unrestrained REMD (Figure S15). We conclude that the two ensembles are reasonably converged despite sampling relatively few refolding events. Unlike the CGUG tetraloop which exhibited multiple dominant clusters, REMD simulations for AUA consistently produce a single dominant cluster (30 – 40 %).

Cluster populations as a function of time were evaluated to determine the time needed to achieve a reliable ensemble (Figure 7a). We expected simulations to take longer to converge than comparable CGUG simulations, as the system size was larger and the RMSD vs. time sampling in AUA was inferior to that observed for CGUG. Overall rank ordering is inconsistent between the two conventional restrained REMD runs and continues to change even at 10 μs. Cluster 2 shows the largest uncertainty in population and rank order, and populations for clusters 1 and 2 do not plateau; the time to converge was estimated as 9 to 10 μs. Ensembles generated by restrained REMD were determined reliable enough to be used as reference data, with limitations to quantitative convergence that should be kept in mind.

Generating reservoirs for the AUA hairpin system

The reservoir generation protocol outlined above for the CGUG hairpin system was repeated for the AUA hairpin. For CGUG, reservoir temperature was selected based on short MD simulations at the top four temperatures of the REMD ladder. In lieu of generating additional MD simulations for AUA, end-to-end distances during the first 2 μs of restrained REMD were examined (Figure S17). 356.0 K appears sufficiently hot, as the broad distance profile was similar at 372.9 K, while lower temperatures showed a narrowed range. As with CGUG, two independent reservoirs of 10,000 structures each were generated from 10 MD independent MD simulations.

Reservoir convergence was tested by comparing cluster populations after 5 and 10 μs (Figure S18). As with CGUG, small populations for all clusters suggest a diverse unfolded ensemble at this elevated temperature, with no strongly preferred structures and all populations less than 2.5 % of the ensemble. Despite the small values, the populations are remarkably close to the diagonal, suggesting that the additional 5 μs had little impact on reservoir populations. The independent reservoirs also show good agreement (Figure S19). Unlike CGUG, the largest cluster at high T matches one of the larger clusters at low T; the RMSD to the second most populated cluster in the reference ensemble is 1.4 Å (Table S3).

Inclusion of the reservoir improves sampling and convergence time

As with CGUG, four R-REMD runs of 1 μs each were generated for the AUA hairpin (two independent R-REMD simulations for each of the two reservoirs). Time-dependent RMSDs compared to the native structure are shown for all replicas in Figures 5c and S20. As with CGUG, these demonstrate significantly improved sampling with R-REMD as compared to conventional REMD over the same 1 μs timescale (Figure 5b). Refolding rates are improved 142-fold over restrained REMD, with an average rate of 71 μs^{-1}. Similar refolding rates are observed for the remaining R-REMD simulations (Figure S20). RMSD histograms are highly uniform across replicas (Figure S21), an improvement over the inconsistent sampling in non-reservoir REMD (Figure S2). Overall, the RMSD analysis illustrates a dramatic improvement in sampling with R-REMD.

There is notably less sampling of high (above 5 Å) RMSD structures by replicas in R-REMD than in REMD, as observed when comparing per-replica RMSD histograms for the two methods in Figures S2 and S21. In order to validate that this behavior does not reflect a problem with the reservoir approach, we examined the temperature-dependent RMSD distributions for both methods (Figure S22). RMSD profiles from R-REMD match well to those from restrained REMD. The higher temperatures simulated in conventional REMD (356.0 K and 372.9 K) show a dramatic shift to higher RMSD values, as does the reservoir itself (generated at 356.0 K, with a profile that matches well to the conventional REMD data at the same temperature). This indicates that the melting temperature of the hairpin in this model is between that of the reservoir and the highest temperature in the R-REMD ladder, with a sharp transition that leads to much higher population of folded hairpin in the R-REMD simulation than in the reservoir itself. The ability of the reservoir to accurately seed ensembles across a melting transition supports the value of the reservoir, despite having a significantly different distribution of structures in the reservoir than are sampled in the R-REMD replicas.

Average cluster populations at 300 K from four R-REMD simulations are compared in Figure 6b. Precision is improved compared to restrained conventional REMD simulations. Error bars reflecting sensitivity to initial structure are small, within 1–2% of the average value (too small to be seen beyond the points in Figure 6b). Average populations from simulations using different reservoirs are in excellent agreement along the diagonal.

As with restrained conventional REMD, a single most populated cluster is observed, at ~ 35% of the ensemble in initially-extended simulations, and 27% in initially-native simulations. The second most populated cluster comprises ~ 22% of the ensemble. These populations correspond to the bimodal RMSD histograms in Figure S22c and d, where two peaks at 300 K can be observed at 3.1 and 3.8 Å RMSD.

R-REMD simulations reproduce the reference ensemble for the AUA hairpin system

Based on the good accuracy obtained when applying R-REMD to the CGUG hairpin, we expected cluster populations from R-REMD to match the reference results from conventional REMD, indicating that coupling to the reservoir accelerated convergence but did not alter the low temperature ensembles. The R-REMD cluster populations at 300 K

were compared to the reference 300 K data from restrained conventional REMD (Figure 6c). The most populated cluster is reproduced, and all of the significant clusters in the reference data are present with R-REMD. Average populations generally fall along the diagonal, with one cluster (cluster 2) showing differences in population larger than the error bars. However, this is likely due to weaknesses in the reference data, since the reference population for this cluster was still changing at 10 μs of simulation (cluster 2 in Figure 7a). Thus, we concluded that R-REMD accurately reproduced the reference ensemble, within convergence limitations of the reference data.

Populations vs. time for the top five clusters demonstrate that coupling to a reservoir provides a dramatic improvement in convergence rate as compared to conventional REMD (Figure 7c). The populations plateau within 200 ns of R-REMD, and are reproducible between the independent simulations. The conventional REMD simulations were unable to achieve a comparable level of precision after 1 μs (Figure 7b) or even after 10 μs (Figure 7a), suggesting a gain in efficiency of at least 50x by coupling to a reservoir. However, this comparison does not include the cost of generating the reservoir, discussed in more detail below.

Precise ensembles permit analysis of the accuracy of this RNA force field + water model combination

One metric for force field accuracy is the ability to reproduce experimental structures. While force field development is outside of the scope of this study, we show an example of how precise RNA ensembles generated by R-REMD can provide feedback on force field accuracy. Since only two RNA systems were simulated, the observations are likely not general enough to suggest specific force field modifications. As we demonstrate, the comparisons performed in this section underscore a greater need for more sequence variety in force field parameterization and testing.

Structure representatives from the top five clusters in the R-REMD ensembles (aggregate > 50% of the ensemble for each hairpin) were compared to the respective native structure (Table 1). Ideally, the most highly populated cluster in simulations would show high similarity to that based on experiments. Alternatively, a high population of non-native structures could suggest biases in the force field. For brevity, the representative structures were analyzed for the most populated cluster and the cluster with the lowest RMSD among the top 5 clusters.

The most populated cluster for the CGUG hairpin does not adopt any native-like properties (Figure 8a), with an RMSD of 4.0 Å (Table 1). The stem is incorrectly formed, with G1 base pairing with C9 instead of C10, resulting in a staggered stem. This may occur due to the stem sequence, which is composed entirely of CG pairs and therefore lacks the specificity to strongly favor the correct alignment. Drawing conclusions about base slipping during stem formation, however, requires more data on systems where slipping could occur, i.e. other RNA systems where the stem contains repeats. Additionally, the G3 base forms hydrogen bonds with a phosphate oxygen on G7, and the G7 base interacts with a U6 phosphate oxygen. Indeed, interactions between base and phosphate were identified as too strong in the ff99bsc0χOL3 force field used here, and adjusted in the ff99bsc0χOL3_LJBB force field.[39]

The second most populated cluster is the closest to the native structure, with a representative RMSD of 2.4 Å, exhibiting a native stem and a loop that contains some native-like properties (Figure 8b). Residues G5, U6, and G7 stack similarly to the native structure. However, G7 is flipped along the χ dihedral, and forms a base pair with C4 that is not present in the native structure, burying the C4 inside the loop and showing distortion in the loop backbone. Unlike the most populated cluster, there is no excessive base pairing between bases and phosphates.

The cluster representatives for the AUA hairpin more closely resemble the corresponding native structure. Cluster 4 has the lowest heavy-atom RMSD (2.1 Å), and the RMSDs for the third, fourth, and fifth clusters are between 2 and 3 Å (Table 1). However, the RMSD of most-populated cluster is 3.1 Å, indicating that a structure somewhat more distant from the native is preferred by this combination of force field and solvent model. Notably, the OP1 phosphate oxygens of A6 and U7 are in close contact (3.5 Å) in the experimental native structure (indicated in Figure 9), suggesting possible charge-charge repulsion mitigated by local structured water or ions; these effects are not expected to be reproduced accurately in an implicit solvent model.

The stem in the most populated cluster reproduces the native base-pairing pattern (Figure 9a). The presence of an A-U base pair reduces the likelihood of base slipping as observed in the CGUG hairpin stem. Native-like stacking is present for U7, A8, and G9. However, in the simulation structure the A6 base is flipped into solvent relative to the native structure, indicating poorer match to experiment. There is also a clockwise twist in the loop backbone compared to the native structure. Additionally, the A6-U7 phosphate oxygen distance is higher in the most populated cluster than in the native structure (8.3 Å vs 3.5 Å).

The cluster representative with a 2.1 Å RMSD to the native structure shows geometry more similar to, but still distorted relative to the native structure (Figure 9b). Unlike the most populated cluster, A6 is not flipped into solvent. There is some overlap in base positions for the loop between the cluster representative and the native structure, but backbone overlap is less accurate. There is also a clockwise twist in U7 and A8 relative to the native structure. This is accompanied by a 6.4 Å A6-U7 phosphate oxygen distance, still significantly higher than in the native structure. This suggests that the distortions may result from inability to stabilize the close-packed phosphate groups.

We have shown that the combination of GB-Neck2 and ff99bsc0χOL3 struggles to reproduce native loop and stem conformations for the CGUG hairpin, and loop conformations for the AUA hairpin. Neither system is able to fully reproduce the native structure. Notably, a staggered stem is observed in the most populated cluster for the CGUG hairpin, and the A6 in the AUA hairpin is flipped out in the most populated cluster. Both AUA clusters are unable to reproduce the short phosphate oxygen distance present in the native structure. The closest structure to the native conformation for both hairpins contained native stems, but loops which contained distortions in the backbone relative to the reference.

It is possible that the lack of agreement with experiment arises from the implicit water model used here. We carried out preliminary MD simulations of the AUA hairpin using

the OPC[80] explicit solvent starting from the NMR-based structure. Simulations using the ff99bsc0χOL3 force field[36], both with and without 12 neutralizing Na^+ ions, as well as simulations with the ff99bsc0χOL3_LJBB force field[39] with neutralizing Na^+, were all unable to maintain the phosphate pair close contact. This distance plateaus at 6 – 8 Å (Figure S23), the same range as seen in the highly populated clusters from implicit solvent. Although a more detailed investigation of the conformational preferences in explicit water is beyond the scope of this study, these limited results suggest that the deviation from experiment in our simulations may be a result of divalent ions present in the experimental sample, or errors in these RNA force fields, rather than the implicit water model.

The overall anecdotal nature and lack of obvious trends between the two systems underscores a greater need for future RNA force field development to use a diverse set of test structures, with a range of sequence and structure motifs as demonstrated in our prior study with proteins[30].

Conclusions

Generating converged ensembles for RNA via MD can be a costly process, requiring multiple long simulations. As a result, RNA force field development and subsequent applications of MD simulations to RNA are computationally expensive. While enhanced sampling methods such as REMD accelerate sampling, converging simulations can still be difficult even for small hairpins. In this study, we explored coupling a reservoir to allow MC jumps between structures, while simultaneously reducing sampling space by introducing a general knowledge-based end-to-end restraint to eliminate sampling of structures that are not expected to be significantly populated for an RNA hairpin.

The introduction of an end-to-end restraint to REMD simulations resulted in a **modest 2-fold** increase in refolding rates for the CGUG hairpin, and a similarly **modest 3-fold** increase for the AUA hairpin, with accurate ensembles relative to unrestrained REMD. Coupling a reservoir to REMD simulations provided a much more dramatic acceleration of sampling and convergence time. Ensembles from R-REMD were precise, and reasonably accurate compared to restrained REMD. The CGUG hairpin showed a **13-fold increase** in refolding rates over restrained REMD simulations, while the AUA hairpin showed an impressive **142-fold increase**. R-REMD simulations converged within 1 μs, an improvement over the multiple μs required for restrained REMD for both systems. However, this neglects the additional cost of generating the reservoir.

Future work should focus on additional optimization of the protocol. Generating a reservoir as described in this study is resource intensive, using 50 μs of simulation per reservoir. However, we demonstrated that there is potential to significantly reduce the computational cost, and future directions should explore how much simulation time can be reduced while maintaining accuracy. Furthermore, identifying an optimal reservoir temperature must be streamlined. Future RNA R-REMD simulations may also consider using explicit solvent, which incurs an additional computational cost compared to implicit solvents. This is particularly important for applications like RNA force field development, where the approximations from implicit solvent may not be sufficient. However, generating well-

converged reservoirs using MD may remain intractable in explicit solvent, even at elevated temperatures.

A possible approach to overcoming these challenges is the use of non-Boltzmann (NB) R-REMD. Non-Boltzmann-weighted reservoirs can be generated by a wider variety of methods, such as metadynamics or accelerated MD, which do not rely on elevated temperature to accelerate sampling. Additionally, NB R-REMD reservoirs in protein systems require less simulation to generate than an equivalent Boltzmann-weighted reservoir[66], making them a promising option for generating RNA explicit solvent reservoirs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant GM107104. We gratefully acknowledge support from Henry and Marsha Laufer.

Abbreviations

RNA
Ribonucleic Acid

GB
Generalized Born

MD
Molecular Dynamics

MC
Monte Carlo

CV
Collective Variable

REMD
Replica Exchange MD

T-REMD
Temperature REMD

H-REMD
Hamiltonian REMD

aMD
Accelerated MD

US
Umbrella sampling

R-REMD
Reservoir REMD

RMSD
Root Mean Square Deviation

PDB
Protein Data Bank

Amber Assisted Model Building with Energy Refinement

NB
Non-Boltzmann

Works Cited

1. Maegdefessel L, The emerging role of microRNAs in cardiovascular disease. J. Intern. Med 2014, 276 (6), 633–644. [PubMed: 25160930]
2. Goodall GJ; Wickramasinghe VO, RNA in cancer. Nat. Rev. Cancer 2021, 21 (1), 22–36. [PubMed: 33082563]
3. Cross ST; Michalski D; Miller MR; Wilusz J, RNA regulatory processes in RNA virus biology. WIREs RNA 2019, 10 (5), e1536. [PubMed: 31034160]
4. Chaudhary N; Weissman D; Whitehead KA, mRNA vaccines for infectious diseases: principles, delivery and clinical translation. Nat. Rev. Drug Discovery 2021, 20 (11), 817–838. [PubMed: 34433919]
5. Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE, The Protein Data Bank. Nucleic Acids Res 2000, 28 (1), 235–242. [PubMed: 10592235]
6. RCSB PDB. rcsb.org (accessed Nov 20, 2021).
7. Šponer J; Bussi G; Krepl M; Banáš P; Bottaro S; Cunha RA; Gil-Ley A; Pinamonti G; Poblete S; Jurečka P; Walter NG; Otyepka M, RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. Chem. Rev 2018, 118 (8), 4177–4338. [PubMed: 29297679]
8. Leontis NB; Stombaugh J; Westhof E, The non-Watson–Crick base pairs and their associated isostericity matrices. Nucleic Acids Res 2002, 30 (16), 3497–3531. [PubMed: 12177293]
9. Stombaugh J; Zirbel CL; Westhof E; Leontis NB, Frequency and isostericity of RNA base pairs. Nucleic Acids Res 2009, 37 (7), 2294–2312. [PubMed: 19240142]
10. Brini E; Simmerling C; Dill K, Protein storytelling through physics. Science 2020, 370 (6520).
11. Sakuraba S; Asai K; Kameda T, Predicting RNA Duplex Dimerization Free-Energy Changes upon Mutations Using Molecular Dynamics Simulations. J. Phys. Chem. Lett 2015, 6 (21), 4348–4351. [PubMed: 26722970]
12. Garcia AE; Paschek D, Simulation of the Pressure and Temperature Folding/Unfolding Equilibrium of a Small RNA Hairpin. J. Am. Chem. Soc 2008, 130 (3), 815–817. [PubMed: 18154332]
13. Krepl M; Réblová K; Koča J; Šponer J, Bioinformatics and Molecular Dynamics Simulation Study of L1 Stalk Non-Canonical rRNA Elements: Kink-Turns, Loops, and Tetraloops. J. Phys. Chem. B 2013, 117 (18), 5540–5555. [PubMed: 23534440]
14. Sorin EJ; Engelhardt MA; Herschlag D; Pande VS, RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop11Edited by J. Dovdna. J. Mol. Biol 2002, 317 (4), 493–506. [PubMed: 11955005]
15. Bottaro S; Nichols PJ; Vögeli B; Parrinello M; Lindorff-Larsen K, Integrating NMR and simulations reveals motions in the UUCG tetraloop. Nucleic Acids Res 2020, 48 (11), 5839–5848. [PubMed: 32427326]

16. Zgarbová M; Otyepka M; Šponer J; Lankaš F; Jurečka P, Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. J. Chem. Theory Comput 2014, 10 (8), 3177–3189. [PubMed: 26588288]
17. Kührová P; Best RB; Bottaro S; Bussi G; Šponer J; Otyepka M; Banáš P, Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. J. Chem. Theory Comput 2016, 12 (9), 4534–4548. [PubMed: 27438572]
18. Haldar S; Kührová P; Banáš P; Spiwok V; Šponer J; Hobza P; Otyepka M, Insights into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics. J. Chem. Theory Comput 2015, 11 (8), 3866–3877. [PubMed: 26574468]
19. Chen AA; García AE, High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. Proc. Natl. Acad. Sci. USA 2013, 110 (42), 16820–16825. [PubMed: 24043821]
20. Bowman GR; Huang X; Yao Y; Sun J; Carlsson G; Guibas LJ; Pande VS, Structural Insight into RNA Hairpin Folding Intermediates. J. Am. Chem. Soc 2008, 130 (30), 9676–9678. [PubMed: 18593120]
21. Ma H; Proctor DJ; Kierzek E; Kierzek R; Bevilacqua PC; Gruebele M, Exploring the Energy Landscape of a Small RNA Hairpin. J. Am. Chem. Soc 2006, 128 (5), 1523–1530. [PubMed: 16448122]
22. Sarkar K; Meister K; Sethi A; Gruebele M, Fast Folding of an RNA Tetraloop on a Rugged Energy Landscape Detected by a Stacking-Sensitive Probe. Biophys. J 2009, 97 (5), 1418–1427. [PubMed: 19720030]
23. Sarkar K; Nguyen DA; Gruebele M, Loop and stem dynamics during RNA hairpin folding and unfolding. RNA 2010, 16 (12), 2427–2434. [PubMed: 20962040]
24. Palermo G; Casalino L; Magistrato A; Andrew McCammon J, Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. J. Struct. Biol 2019, 206 (3), 267–279. [PubMed: 30880083]
25. Fandilolu PM; Kamble AS; Dound AS; Sonawane KD, Role of Wybutosine and Mg(2+) Ions in Modulating the Structure and Function of tRNA(Phe): A Molecular Dynamics Study. ACS Omega 2019, 4 (25), 21327–21339. [PubMed: 31867527]
26. Ramakrishnan C; Nagarajan R; Sekijima M; Michael Gromiha M, Molecular dynamics simulations of cognate and non-cognate AspRS-tRNAAsp complexes. J. Biomol. Struct. Dyn 2020, 1–9.
27. Al-Hashimi HM; Walter NG, RNA dynamics: it is about time. Curr. Opin. Struct. Biol 2008, 18 (3), 321–329. [PubMed: 18547802]
28. Bothe JR; Nikolova EN; Eichhorn CD; Chugh J; Hansen AL; Al-Hashimi HM, Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. Nat. Methods 2011, 8 (11), 919–931. [PubMed: 22036746]
29. Chen S-J, RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. Annu. Rev. Biophys 2008, 37 (1), 197–214. [PubMed: 18573079]
30. Nguyen H; Maier J; Huang H; Perrone V; Simmerling C, Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. J. Am. Chem. Soc 2014, 136 (40), 13959–13962. [PubMed: 25255057]
31. Lindorff-Larsen K; Piana S; Dror RO; Shaw DE, How Fast-Folding Proteins Fold. Science 2011, 334 (6055), 517–520. [PubMed: 22034434]
32. Robertson JC; Nassar R; Liu C; Brini E; Dill KA; Perez A, NMR-assisted protein structure prediction with MELDxMD. Proteins: Struct., Funct., Bioinf 2019, 87 (12), 1333–1340.
33. Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc 1995, 117 (19), 5179–5197.
34. Wang J; Cieplak P; Kollman PA, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem 2000, 21 (12), 1049–1074.

35. Pérez A; Marchán I; Svozil D; Sponer J; Cheatham TE 3rd; Laughton CA; Orozco M, Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys. J 2007, 92 (11), 3817–3829. [PubMed: 17351000]
36. Zgarbová M; Otyepka M; Sponer J; Mládek A; Banáš P; Cheatham TE 3rd; Jurečka P, Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. J. Chem. Theory Comput 2011, 7 (9), 2886–2902. [PubMed: 21921995]
37. Banáš P; Hollas D; Zgarbová M; Jurečka P; Orozco M; Cheatham TE; Šponer J; Otyepka M, Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. J. Chem. Theory Comput 2010, 6 (12), 3836–3849. [PubMed: 35283696]
38. Yildirim I; Stern HA; Kennedy SD; Tubbs JD; Turner DH, Reparameterization of RNA chi Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. J. Chem. Theory Comput 2010, 6 (5), 1520–1531. [PubMed: 20463845]
39. Bergonzo C; Cheatham TE, Improved Force Field Parameters Lead to a Better Description of RNA Structure. J. Chem. Theory Comput 2015, 11 (9), 3969–3972. [PubMed: 26575892]
40. Aytenfisu AH; Spasic A; Grossfield A; Stern HA; Mathews DH, Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. J. Chem. Theory Comput 2017, 13 (2), 900–915. [PubMed: 28048939]
41. 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. USA 2018, 115 (7), E1346–E1355. [PubMed: 29378935]
42. 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 (9), 1578–1590. [PubMed: 26124199]
43. Roe DR; Bergonzo C; Cheatham TE, Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. J. Phys. Chem. B 2014, 118 (13), 3543–3552. [PubMed: 24625009]
44. Torrie GM; Valleau JP, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J. Comput. Phys 1977, 23 (2), 187–199.
45. Deng N-J; Cieplak P, Free Energy Profile of RNA Hairpins: A Molecular Dynamics Simulation Study. Biophys. J 2010, 98 (4), 627–636. [PubMed: 20159159]
46. Reißer S; Zucchelli S; Gustincich S; Bussi G, Conformational ensembles of an RNA hairpin using molecular dynamics and sparse NMR data. Nucleic Acids Res 2019.
47. Nguyen H; Roe DR; Simmerling C, Improved Generalized Born Solvent Model Parameters for Protein Simulations. J. Chem. Theory Comput 2013, 9 (4), 2020–2034. [PubMed: 25788871]
48. Hawkins GD; Cramer CJ; Truhlar DG, Pairwise solute descreening of solute charges from a dielectric medium. Chem. Phys. Lett 1995, 246 (1), 122–129.
49. Onufriev A; Bashford D; Case DA, Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins: Struct., Funct., Bioinf 2004, 55 (2), 383–394.
50. Onufriev AV; Case DA, Generalized Born Implicit Solvent Models for Biomolecules. Annu. Rev. Biophys 2019, 48, 275–296. [PubMed: 30857399]
51. Sorin EJ; Rhee YM; Pande VS, Does Water Play a Structural Role in the Folding of Small Nucleic Acids? Biophys. J 2005, 88 (4), 2516–2524. [PubMed: 15681648]
52. Huang H; Simmerling C, Fast Pairwise Approximation of Solvent Accessible Surface Area for Implicit Solvent Simulations of Proteins on CPUs and GPUs. J. Chem. Theory Comput 2018, 14 (11), 5797–5814. [PubMed: 30303377]
53. Nguyen H; Pérez A; Bermeo S; Simmerling C, Refinement of Generalized Born Implicit Solvation Parameters for Nucleic Acids and Their Complexes with Proteins. J. Chem. Theory Comput 2015, 11 (8), 3714–3728. [PubMed: 26574454]
54. Bergonzo C; Grishaev A, Accuracy of MD solvent models in RNA structure refinement assessed via liquid-crystal NMR and spin relaxation data. J. Struct. Biol 2019, 207 (3), 250–259. [PubMed: 31279068]

55. Sugita Y; Okamoto Y, Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett 1999, 314 (1), 141–151.
56. Mlýnský V; Bussi G, Exploring RNA structure and dynamics through enhanced sampling simulations. Curr. Opin. Struct. Biol 2018, 49, 63–71. [PubMed: 29414513]
57. Geng H; Chen F; Ye J; Jiang F, Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. Comput. Struct. Biotechnol. J 2019, 17, 1162–1170. [PubMed: 31462972]
58. Ostermeir K; Zacharias M, Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. Biochim. Biophys. Acta, Proteins Proteomics 2013, 1834 (5), 847–853.
59. Nymeyer H; Gnanakaran S; García AE, Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm. In Methods Enzymol, Academic Press: 2004; Vol. 383, pp 119–149. [PubMed: 15063649]
60. Fukunishi H; Watanabe O; Takada S, On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. J. Chem. Phys 2002, 116 (20), 9058–9067.
61. Okur A; Wickstrom L; Layten M; Geney R; Song K; Hornak V; Simmerling C, Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. J. Chem. Theory Comput 2006, 2 (2), 420–433. [PubMed: 26626529]
62. Bergonzo C; Henriksen NM; Roe DR; Swails JM; Roitberg AE; Cheatham TE 3rd, Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. J. Chem. Theory Comput 2014, 10 (1), 492–499. [PubMed: 24453949]
63. Henriksen NM; Roe DR; Cheatham TE 3rd, Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations. J. Phys. Chem. B 2013, 117 (15), 4014–4027. [PubMed: 23477537]
64. Okur A; Roe DR; Cui G; Hornak V; Simmerling C, Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir. J. Chem. Theory Comput 2007, 3 (2), 557–568. [PubMed: 26637035]
65. Roitberg AE; Okur A; Simmerling C, Coupling of replica exchange simulations to a non-Boltzmann structure reservoir. J. Phys. Chem. B 2007, 111 (10), 2415–2418. [PubMed: 17300191]
66. Kasavajhala K; Lam K; Simmerling C, Exploring Protocols to Build Reservoirs to Accelerate Temperature Replica Exchange MD Simulations. J. Chem. Theory Comput 2020, 16 (12), 7776–7799. [PubMed: 33142060]
67. Rijnbrand R; Thiviyanathan V; Kaluarachchi K; Lemon SM; Gorenstein DG, Mutational and Structural Analysis of Stem-loop IIIc of the Hepatitis C Virus and GB Virus B Internal Ribosome Entry Sites. J. Mol. Biol 2004, 343 (4), 805–817. [PubMed: 15476802]
68. Kim C-H; Kao CC; Tinoco I, RNA motifs that determine specificity between a viral replicase and its promoter. Nat. Struct. Biol 2000, 7 (5), 415–423. [PubMed: 10802741]
69. Macke TJ; Case DA, Modeling Unusual Nucleic Acid Structures. In Molecular Modeling of Nucleic Acids, American Chemical Society: 1997; Vol. 682, pp 379–393.
70. Hopkins CW; Le Grand S; Walker RC; Roitberg AE, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J. Chem. Theory Comput 2015, 11 (4), 1864–1874. [PubMed: 26574392]
71. Roe DR; Cheatham TE, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput 2013, 9 (7), 3084–3095. [PubMed: 26583988]
72. Izadi S; Anandakrishnan R; Onufriev AV, Building Water Models: A Different Approach. J. Phys. Chem. Lett 2014, 5 (21), 3863–3871. [PubMed: 25400877]
73. Joung IS; Cheatham TE, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. J. Phys. Chem. B 2008, 112 (30), 9020–9041. [PubMed: 18593145]
74. Humphrey W; Dalke A; Schulten K, VMD: Visual molecular dynamics. J. Mol. Graphics 1996, 14 (1), 33–38.

75. Seibert MM; Patriksson A; Hess B; van der Spoel D, Reproducible Polypeptide Folding and Structure Prediction using Molecular Dynamics Simulations. J. Mol. Biol 2005, 354 (1), 173–183. [PubMed: 16236315]

76. Chakraborty D; Collepardo-Guevara R; Wales DJ, Energy Landscapes, Folding Mechanisms, and Kinetics of RNA Tetraloop Hairpins. J. Am. Chem. Soc 2014, 136 (52), 18052–18061. [PubMed: 25453221]

77. Zheng W; Andrec M; Gallicchio E; Levy RM, Simulating replica exchange simulations of protein folding with a kinetic network model. Proc. Natl. Acad. Sci. USA 2007, 104 (39), 15340–15345. [PubMed: 17878309]

78. Brini E; Kozakov D; Dill KA, Predicting Protein Dimer Structures Using MELD × MD. J. Chem. Theory Comput 2019, 15 (5), 3381–3389. [PubMed: 30908034]

79. Liu C; Brini E; Perez A; Dill KA, Computing Ligands Bound to Proteins Using MELD-Accelerated MD. J. Chem. Theory Comput 2020, 16 (10), 6377–6382. [PubMed: 32910647]

80. Izadi S; Anandakrishnan R; Onufriev AV, Building Water Models: A Different Approach. The Journal of Physical Chemistry Letters 2014, 5 (21), 3863–3871. [PubMed: 25400877]

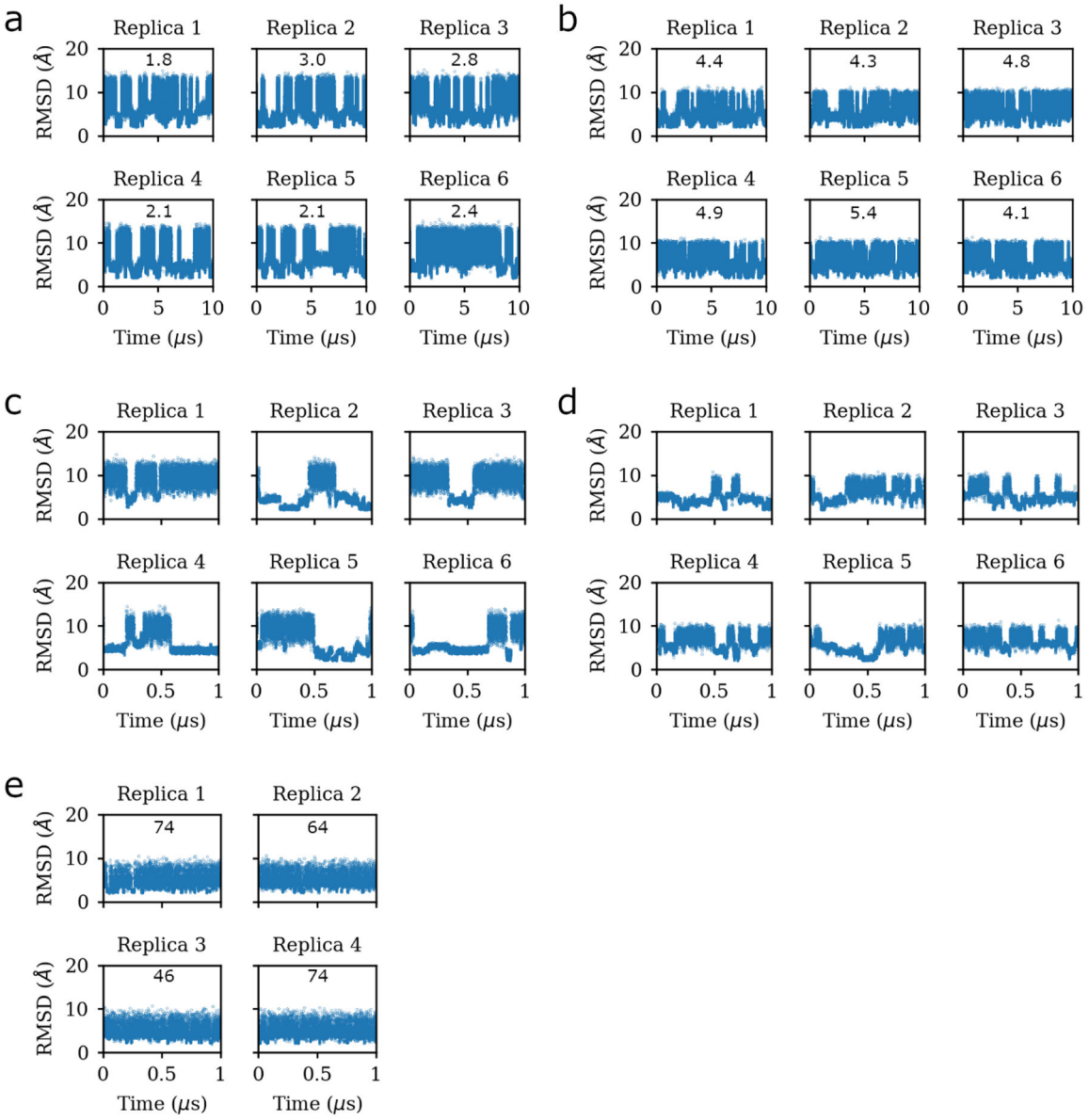


Figure 1.

(upper) Time-dependent heavy atom RMSD to the CGUG reference PDB structure for 10 μs of initially-extended (a) unrestrained and (b) restrained conventional REMD, first 1 μs of (c) unrestrained and (d) restrained conventional REMD, and (e) 1 μs R-REMD. The refolding rate (in μs^{-1}) for each replica is indicated in each subplot. The average refolding rates are 2.3, 4.7, and 61 μs^{-1} for (a), (b), and (e) respectively.

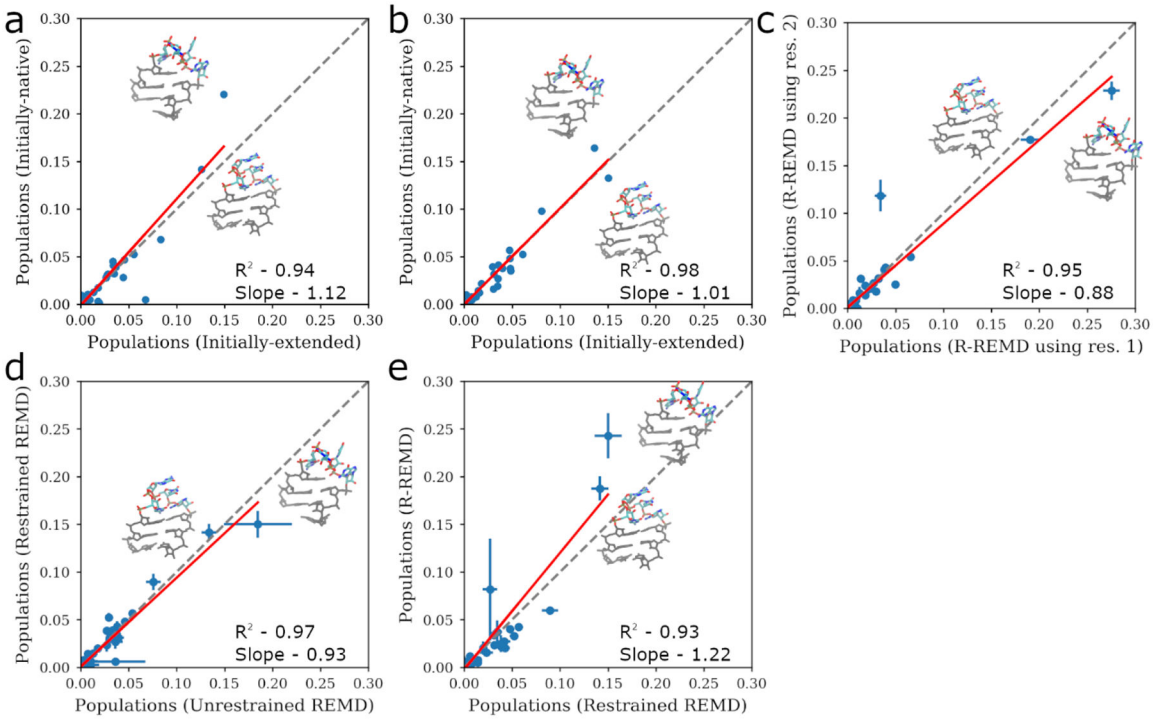


Figure 2.

Comparison of cluster populations in independent 300 K ensembles of the CGUG hairpin for various REMD methods. Populations are compared between initially extended and native runs from (a) unrestrained and (b) restrained conventional REMD simulations. Average populations are compared for initially-extended and native simulations in (c) R-REMD using different reservoirs, (d) unrestrained REMD and restrained REMD, and (e) restrained REMD and R-REMD. Error bars represent the difference from the average for two independent runs. Clustering was performed on all heavy atoms. The diagonal indicating hypothetical perfect match is in gray; populations for most clusters are near the diagonal. Best fit lines are in red, with R^2 and slope in plots. Representative structures for the top two clusters are pictured.

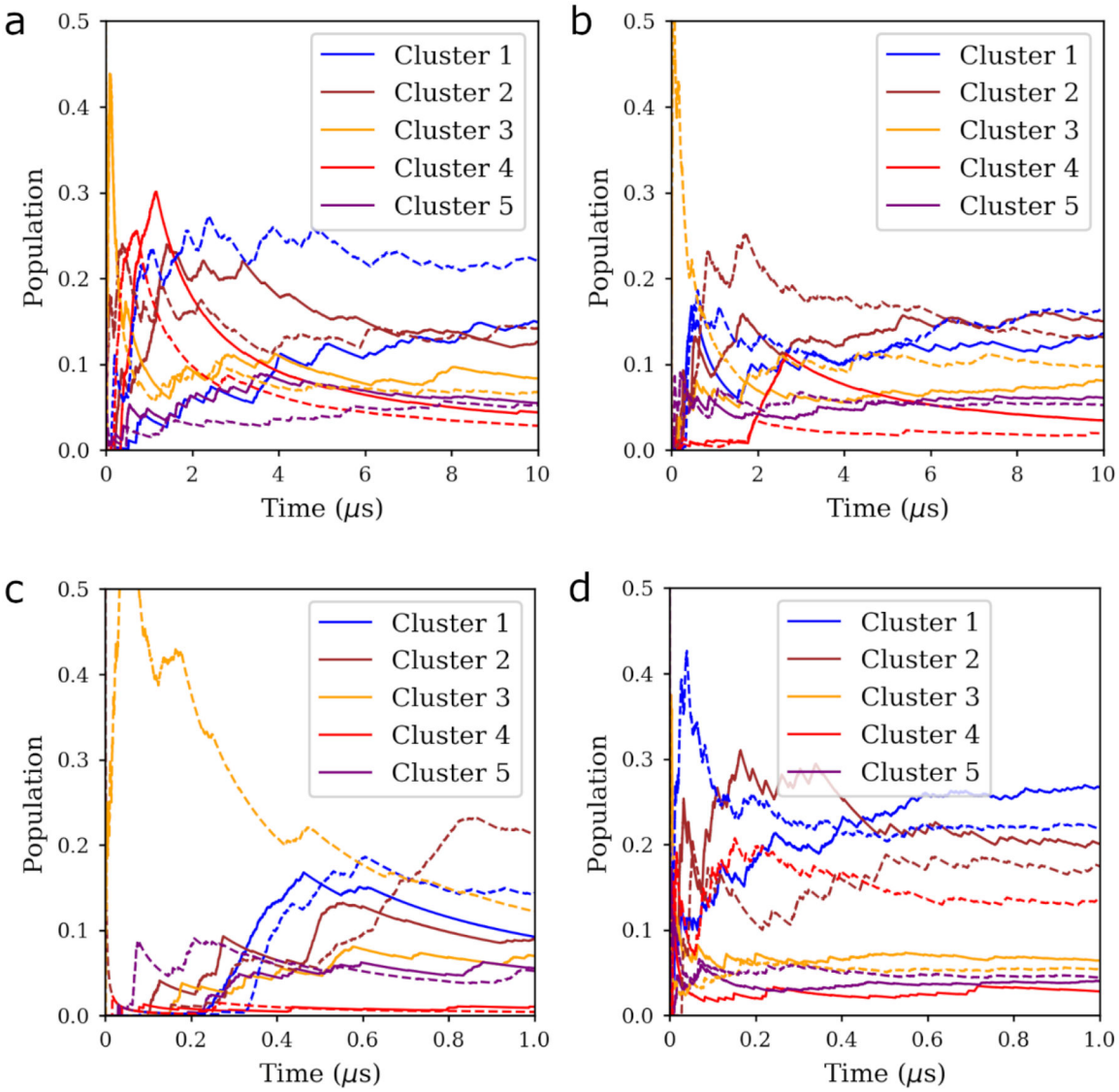


Figure 3.

CGUG hairpin cluster populations vs. time at 300 K for the top 5 most populated clusters in two runs of **10 μs** conventional (a) unrestrained REMD and (b) restrained REMD, and for **1 μs** of (c) restrained REMD and (d) R-REMD. Clustering was performed on all heavy atoms. Solid and dashed lines represent independent simulations with extended and native initial conformations, respectively. Rank order takes at least 6–8 μs to become reliable for conventional REMD, with some average populations still changing at 10 μs. R-REMD shows convergence around 0.5 μs.

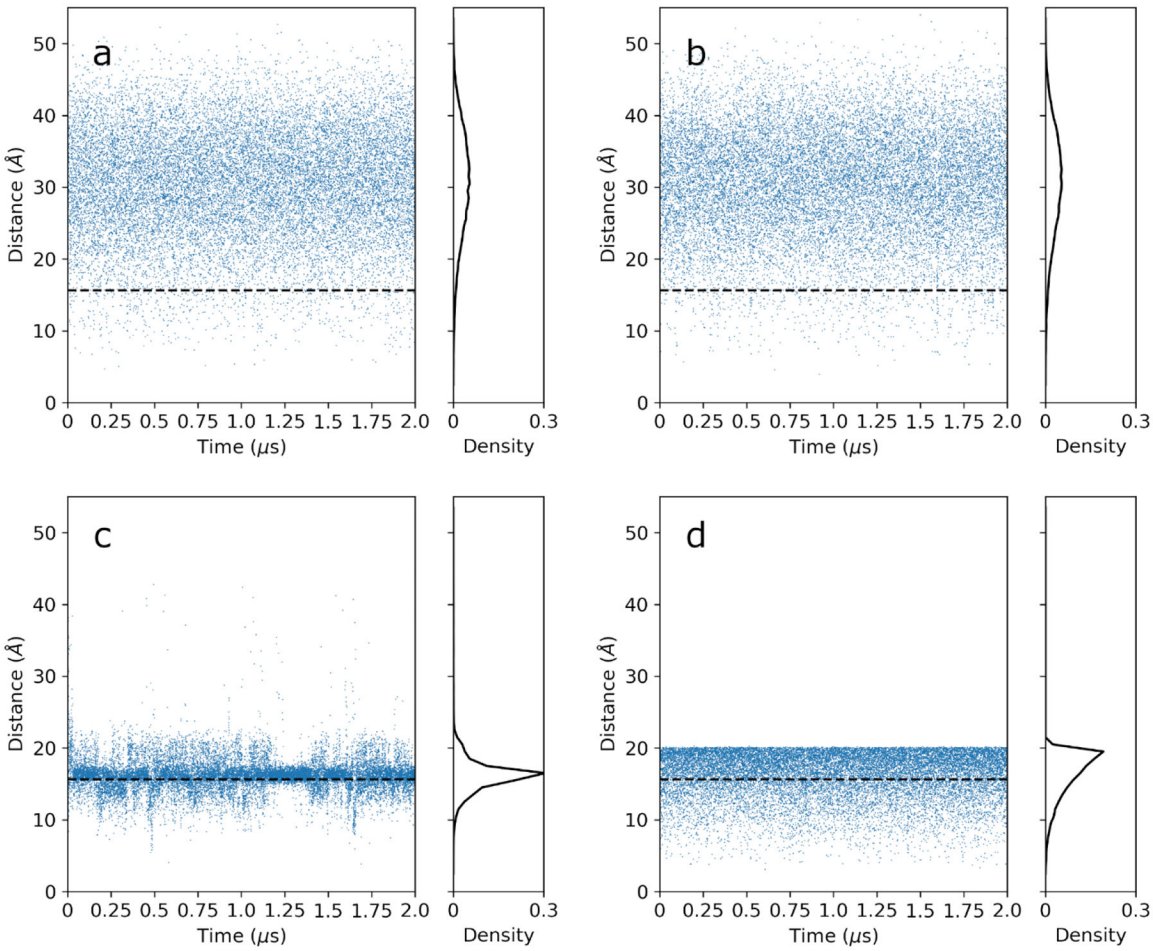


Figure 4.

End-to-end distances during simulations of the CGUG hairpin. From top left to bottom right: (a) unrestrained REMD 384.2 K, (b) unrestrained MD 384.3 K, (c) unrestrained REMD 300 K, (d) restrained MD 384.3 K. Normalized histograms of the distances are provided on the right, while the distance in the native structure is indicated with a dashed line. Unrestrained REMD and MD data at 384.3 K in (a) and (b) tend to sample predominantly structures with much longer distances than seen in the NMR-based structure or in REMD at 300 K (b). Adding a restraint (d) improves sampling of shorter distances at elevated temperatures.

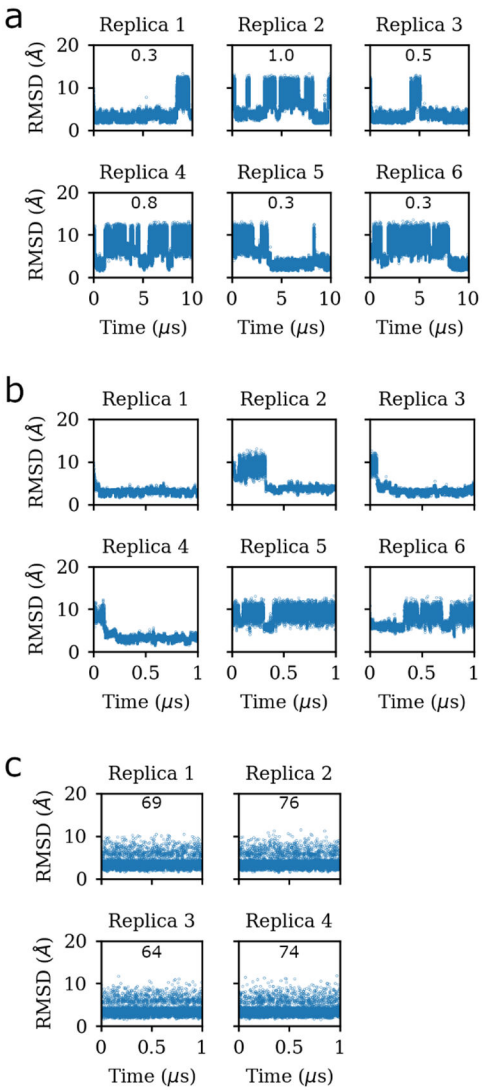


Figure 5.

AUA hairpin system per-replica heavy atom RMSDs to the reference PDB structure as a function of simulation time for initially-extended (a) all 10 μs restrained REMD, (b) first 1 μs restrained REMD, and (c) 1 μs R-REMD. Refolding rates for each replica (in μs^{-1}) are provided in the subplots. The average refolding rate is 0.5 and 71 μs^{-1} for (a) and (c) respectively.

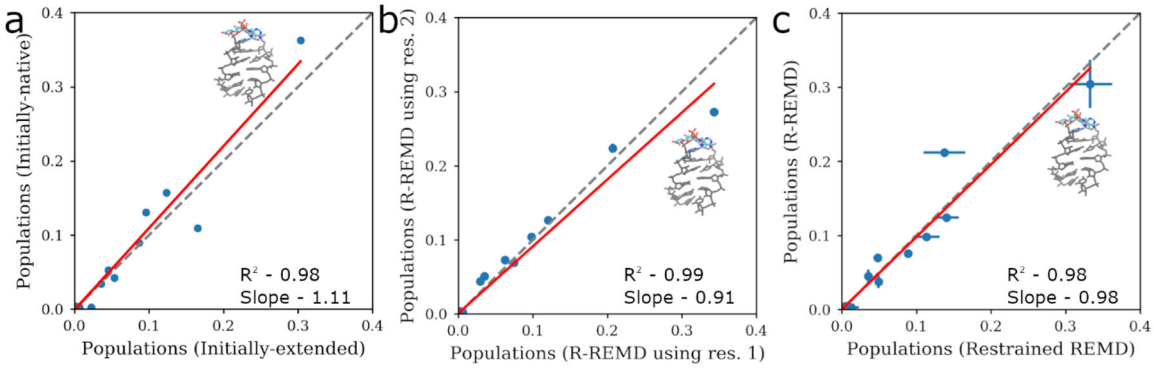


Figure 6.

Comparison of cluster populations in 300 K ensembles for the AUA hairpin system. Comparisons are made between (a) 10 μs restrained REMD simulations, (b) R-REMD runs using different reservoirs, and (c) restrained REMD and R-REMD. Clustering was performed on all heavy atoms. Error bars in (b) and (c) represent the difference from the average for independent runs. Error bars for (b) are too small to be clearly seen. The diagonal for perfect match is shown in gray. The line of best fit is in red, with slope and R^2 in each subplot. The representative structure is shown for the top cluster.

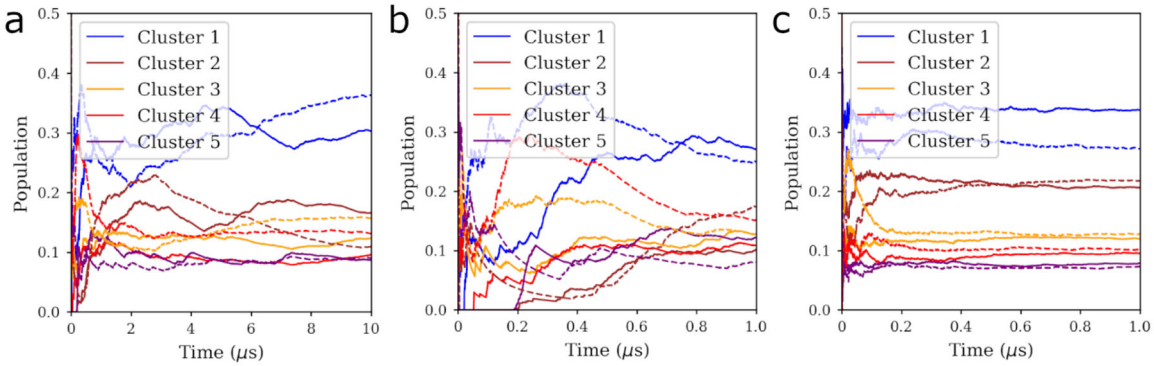


Figure 7.

AUA hairpin system cluster populations vs. time at 300 K for the top 5 most populated clusters in two runs of (a) all 10 μs conventional restrained REMD, (b) first 1 μs restrained REMD and (c) 1 μs R-REMD. Clustering was performed on all heavy atoms on all residues. Solid and dashed lines represent independent simulations with extended and native initial conformations for restrained REMD, and independent runs using different reservoirs for R-REMD, respectively.

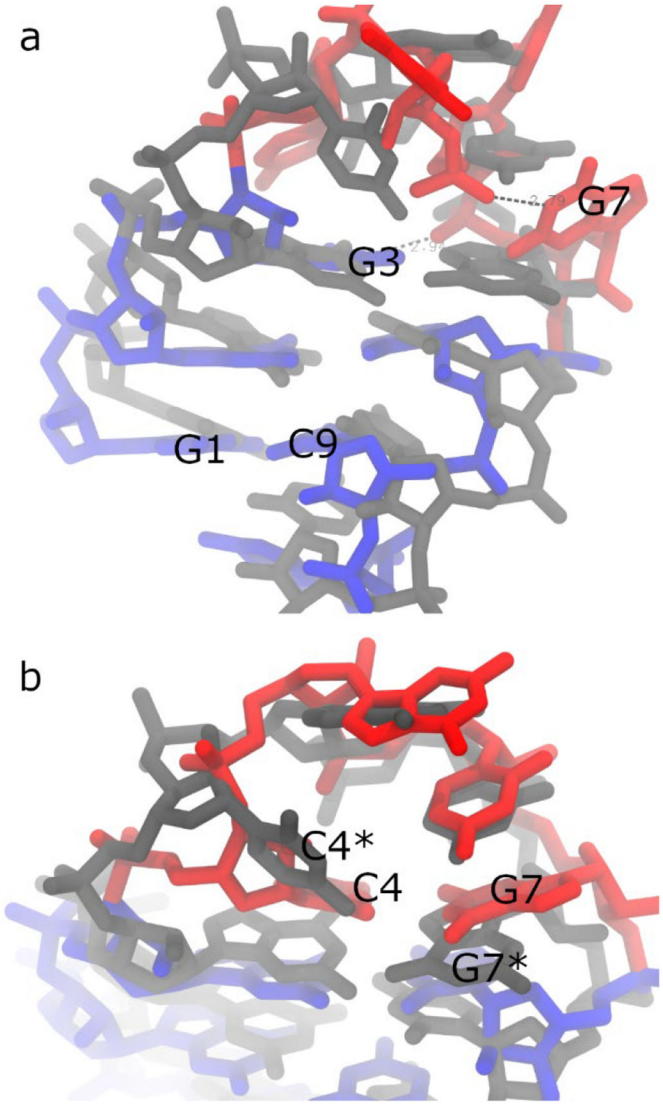


Figure 8.

CGUG hairpin representative structures for clusters with the (a) largest population and (b) lowest heavy atom RMSD to the native structure. The native structure is colored in gray, and the loop and stem for cluster representatives in red and blue, respectively. Hydrogen bonds between the G7 base and U6 phosphate, and the G3 base and G7 phosphate are indicated in dashed lines in (a). Select residues are labeled, and C4* and G7* in (b) indicate the native structure.

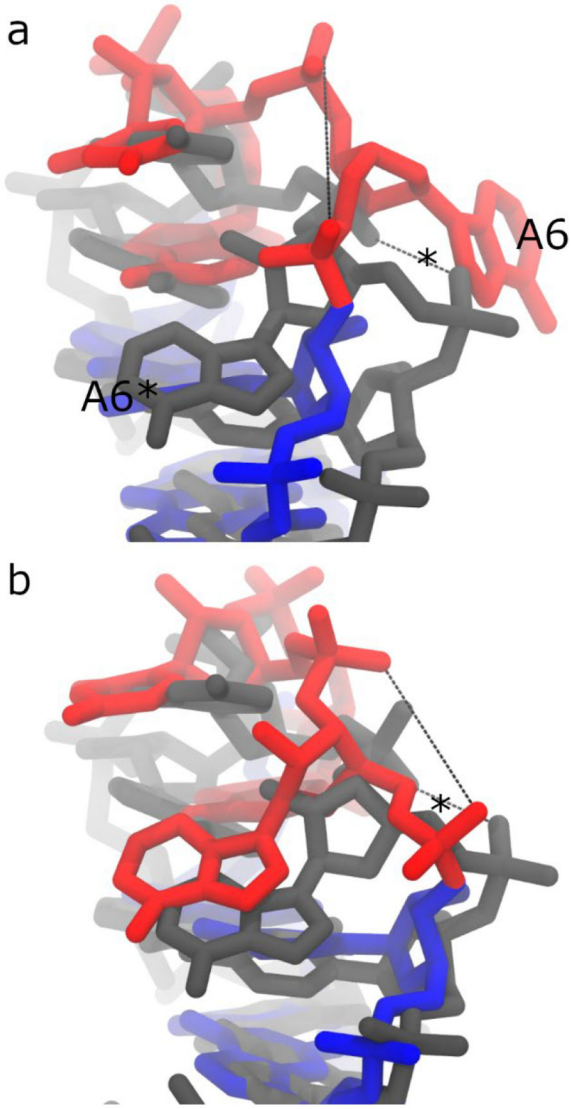


Figure 9.

AUA hairpin representative structures for clusters with the (a) largest population and (b) lowest heavy atom RMSD to the native structure. The native structure is colored in gray, and the loop and stem for cluster representatives in red and blue, respectively. The distances between phosphate oxygens on A6 and U7 are indicated with dashed lines, and the native distance is marked with an asterisk. A6 is labeled, with A6* indicating the residue in the native structure.

Table 1.

Heavy-atom RMSDs in Å for the top 5 clusters of the CGUG and AUA hairpins, relative to the corresponding experimental structure.

Cluster	CGUG Hairpin	AUA Hairpin
1	4.0	3.1
2	2.4	3.8
3	4.2	2.6
4	4.3	2.1
5	3.6	3.0