Supporting Information

Sim et al. 10.1073/pnas.1119918109

SI Methods

Modeling Mutations in the tRNA Monomer and tRNA-Square. Structures and sequences of the tRNA monomer and tRNA-square as studied experimentally (1) were obtained from the Jaeger lab. Mutated sequences were threaded into the RNA backbone using an in-house design software and all structures were subsequently minimized [using AMBER 99-bs0 (2) and dielectric damping model (3) in MOSAICS (4)] with only side-chain flexibility to preserve backbone topology.

To prevent steric clashes, the structures were modeled using a hard-sphere potential and base-pairs were preserved by the appropriate choice of sampling degrees of freedom. This was done by manipulating the sampling degrees of freedom such that base-pairs can only move in concert (except during chain closure) and there were no degrees of freedom that moves individual bases alone within the base-pair.

Three different sets of mutations were done on the tRNA monomer: U15G, C16G, and both U15G and C16G simultaneously. Besides moving each base-pair as one, helices were allowed to move independently, akin to the approach taken for the simple four-way junction. To determine the angle (θ) between the stacked helices of tRNA (see Fig. 3*A*), perfect A-form helices were superimposed (using backbone P only) to the stacked helices and the helical axes were determined using X3DNA (5). θ was then determined from the dot product of both axes.

Each monomer within the mutated tRNA-square carried the double mutation, and θ for each corner was determined as previously. For statistics presented in Fig. 3*B* (12,000 structures each for wild type and mutant), maximum $\Delta\theta$ is the maximum difference in θ for all four corners of each modeled structure.

Conformations of an RNA Four-Way Junction Derived from the RNA Hairpin Ribozyme. To simulate this junction, the starting structure was generated for the sequence in reference (6). Perfect A-form helices were used for the base-paired and stacked regions (i.e., a perfect A-form helix for each D on A and C on B stacked helices). The full RNA was assembled with a starting angle $\theta = 117.5°$ (see Fig. 4) and distances as determined visually between stacked helices; ensemble results were comparable, regardless of starting θ (Fig. S4). In order to stitch the RNA, minimization was done using Nucleic Acid Builder (7). In most cases, only base-pairs close to the junction were allowed to move during minimization, but additional nucleotides were also allowed to move as needed. Sampling was constrained such that all base-pairs were kept rigid and each pair of stacked helices was regarded as a continuous helix. A hard-sphere potential was used to isolate the effects of chain connectivity and sterics on junction handedness.

From our simulations, the right-handed antiparallel conformation of this four-way junction is most consistent with experimental FRET data (6) because the antiparallel to parallel structure transition can take place for the left-handed form but not the right-handed one (with stacking intact). However, there might be other biophysical forces at play, preventing the left-handed antiparallel conformation from rotating to the left-handed parallel conformation. Alternatively, the true experimental transition might be from the preferred antiparallel conformation to a differently stacked structure that is experimentally indistinguishable from the parallel conformation (6). If so, this set of experimental data and our simulations are insufficient to distinguish the handedness of the preferred antiparallel conformation. Nonetheless, the sterics and connectivity argument is simple, and the model interpretation straightforward. The experimental data together with our simulations give a simple plausible expectation that the junction prefers the right-handed antiparallel conformation. This conclusion is later supported by other experiments (8).

General Applicability of Hierarchical Natural Move Monte Carlo and Stochastic Chain Closure. While our stochastic chain closure algorithm is robust and able to restore chain connectivity in most instances, there are some cases when stochastic chain closure fails, in which case the proposed move is rejected. Because rejection of a proposed step due to unsuccessful chain closure affects the overall Monte Carlo acceptance ratio, the user-defined move step sizes can be adjusted such that the acceptance ratio is reasonable (approximately 0.3–0.4); chain closure is likely to succeed for smaller move steps. Alternatively, it is possible to increase the chain segment dedicated to solving the closure problem, as was discussed previously (9).

In our implementation, the total number of degrees of freedom (L) is partitioned into independent and dependent degrees of freedom ($L = L_i \cup L_d$). Hence our method can be considered as a Monte Carlo minimization technique in which minimization is performed on a set of dependent degrees of freedom, or equivalently we explore the energy surface:

$$\tilde{E}(L) = \tilde{E}(L_i \cup L_d) = \min_{L_d}\{E(L_i \cup L_d)\}$$

where E is the user-defined energy function. Our hierarchical natural move Monte Carlo (HNM-MC), similar to Monte Carlo minimization, is designed to sample a transformed energy surface in which some of the transition energy barriers are removed. Thus HNM-MC is not designed to preserve microscopic reversibility or to generate canonically distributed conformational ensembles along all degrees of freedom (L) over the original energy surface (E). Nevertheless, HNM-MC may be used as a canonical sampling method along the independent degrees of freedom, L_i (that are not affected by the minimization) over $\tilde{E}$. We are currently running more rigorous tests comparing simulations by HNM-MC to reference canonical distributions obtained by N dimensional numerical integrals based on quadrature rules (an approach successfully applied in ref. 10). Extensive discussion on microscopic reversibility is beyond the scope of the current paper.

Due to the versatility of our algorithm and the generality of our implementation in MOSAICS (4), choices of moves for any particular system is completely user-defined and would clearly therefore depend on the purpose of simulation/sampling. If there are any known structural constraints, one can use correlated movesets to preserve them, as was done for applications 1 and 2 discussed in the main text. Based on our modeling experience, the effects of using different levels of moves on sampling efficiency depend heavily on the system studied. For example, sampling efficiency for an RNA with secondary structure constraints depends significantly on the length of single-stranded regions connecting helices as this changes RNA conformational flexibility.

1. Severcan I, Geary C, Verzemnieks E, Chworos A, Jaeger L (2009) Square-shaped RNA particles from different RNA folds. *Nano Lett* 9:1270–1277.
2. Perez A, et al. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* 92:3817–3829.
3. Rohs R, Etchebest C, Lavery R (1999) Unraveling proteins: A molecular mechanics study. *Biophys J* 76:2760–2768.
4. Minary P (2007) Methodologies for Optimization and SAmpling In Computational Studies (MOSAICS), version 3.8. http://csb.stanford.edu/~minary/MOSAICS.html.

5. Lu XJ, Olson WK (2008) 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* 3:1213–1227.
6. Hohng S, et al. (2004) Conformational flexibility of four-way junctions in RNA. *J Mol Biol* 336:69–79.
7. Macke TJ, Case DA (1998) Modeling unusual nucleic acid structures. *ACS Sym Ser* 682:379–393.
8. Goody TA, Lilley DM, Norman DG (2004) The chirality of a four-way helical junction in RNA. *J Am Chem Soc* 126:4126–4127.
9. Minary P, Levitt M (2010) Conformational optimization with natural degrees of freedom: A novel stochastic chain closure algorithm. *J Comput Biol* 17:993–1010.
10. Minary P, Tuckerman M, Martyna G (2008) Dynamical spatial warping: A novel method for the conformational sampling of biophysical structure. *SIAM J Sci Comput* 30:2055–2083.

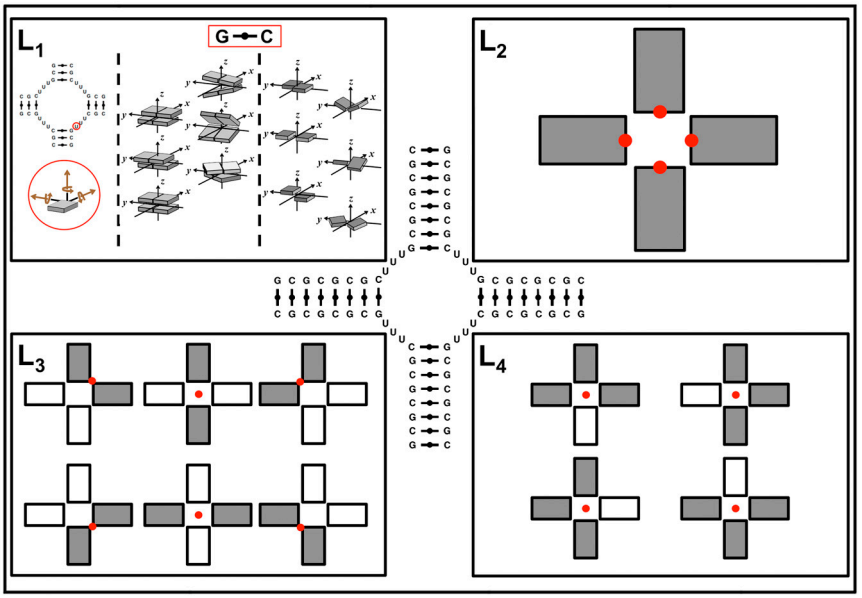

Fig. S1. Hierarchical moves for a simple four-way junction. Each nucleotide can move independently or as part of a base-pair (L_1). Regions of continuous base-pairing form helices (represented as rectangles), and these helices can be regarded as rigid bodies either independently (shaded in gray, L_2), as pairs of helices (L_3), or as groups of three (L_4). Rigid body motion requires the definition of rotation centers (shown as red dots). In the present case, the centers were defined to preserve symmetry of the system.

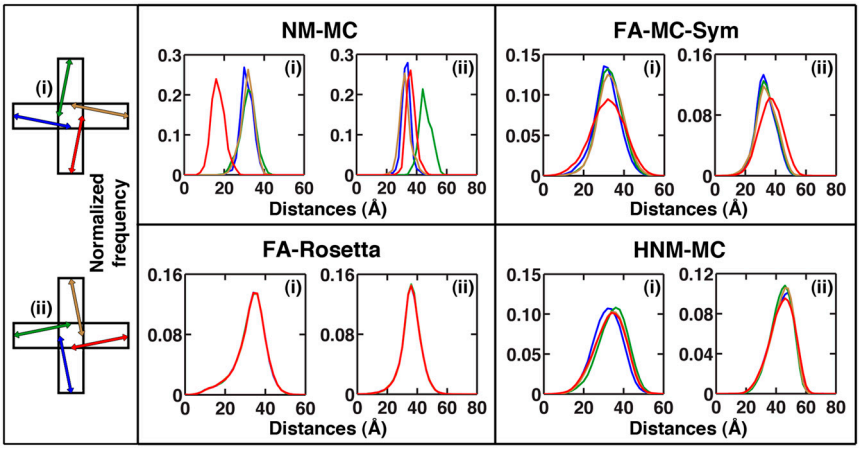

Fig. S2. Distance distributions of a simple four-way junction obtained by sampling using four different sampling protocols: natural move Monte Carlo (NM-MC), fragment assembly implemented by MC-Sym (FA-MC-Sym), fragment assembly implemented by Rosetta (FA-Rosetta) and hierarchical natural move Monte Carlo (HNM-MC). See Fig. 1 in main text for the distance distributions between ends of neighboring helices.

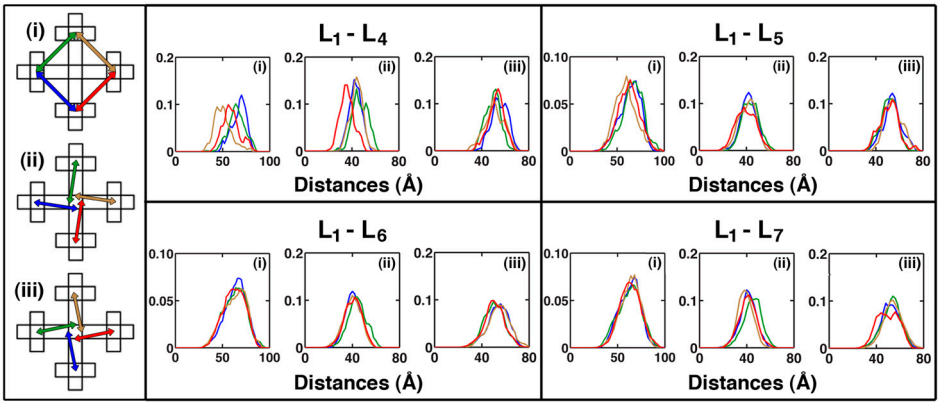

Fig. S3. Distance distributions within a large and complex symmetric RNA for four different sets of hierarchical moves (L_1-L_4, L_1-L_5, L_1-L_6, and L_1-L_7; see Fig. 2*A* for illustration of each level). Addition of higher order collective rigid body motions (i.e., adding L_5-L_7 to L_1-L_4) improved sampling efficiency significantly, and the distance distributions converged, as expected for this symmetrical system. Distance distributions were obtained from the statistics based on 2×10^4 iterations indicated by the vertical dashed line labeled * in Fig. 2*B*.

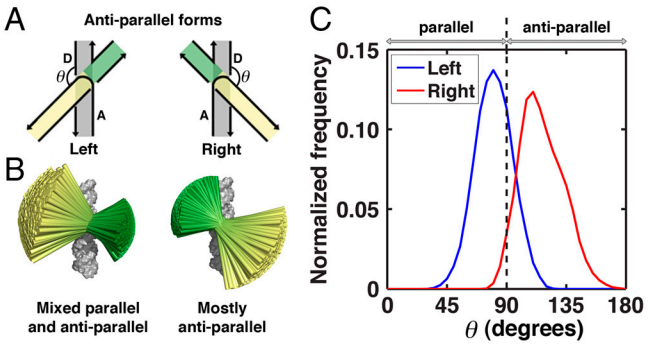

Fig. S4. Constrained sampling of a four-way RNA junction derived from the RNA hairpin ribozyme. The actual conformational distributions illustrated in Fig. 4 of the main text are dependent on the initial starting structure, due to the highly constrained nature of the junction and move-sets used. However, the conclusion that the left-handed form is more flexible (switches readily between parallel and antiparallel forms) than the right-handed one is independent of starting choice of θ. Here, a starting $\theta = 100°$ was used, compared to a starting $\theta = 117.5°$ in Fig. 4 of the main text.