Chapter 9
Multiscale Modeling of RNA Structure and Dynamics

Feng Ding and Nikolay V. Dokholyan

Abstract We have developed a multiscale approach for RNA folding using discrete molecular dynamics (DMD), a rapid conformational sampling algorithm. We use a coarse-grained representation to effectively model RNA structures. Benchmark studies suggest that the DMD-based RNA model is able to accurately fold small RNA molecules (<50 nucleotides). However, the large conformational space and force field inaccuracies make it difficult to computationally identify the native states of large RNA molecules. We devised an automated modeling approach for prediction of large and complex RNA structures using experimentally derived structural constraints and tested it on several RNA molecules with known experimental structures. In all cases, we were able to bias the DMD simulations to the native states of these RNA molecules. Therefore, a combination of experimental and computational approaches has the potential to yield native-like models for the diverse universe of functionally important RNAs, whose structures cannot be characterized by conventional structural methods.

9.1 Introduction

RNA molecules play a wide range of functional roles in gene expression, from regulating transcription and translation [e.g., riboswitch regulator motifs (Edwards et al. 2007)] to decoding genetic messages (tRNA), catalyzing mRNA splicing [spliceosome RNA or self-splicing introns (Vicens and Cech 2006)] and protein synthesis (rRNA). Knowledge of the underlying RNA structure in these and many other molecules is a fundamental prerequisite to a complete understanding of RNA function. Methods such as X-ray crystallography and NMR spectroscopy offer critical

F. Ding • N.V. Dokholyan (✉)
Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA
e-mail: dokh@med.unc.edu

N. Leontis and E. Westhof (eds.), *RNA 3D Structure Analysis and Prediction*,
Nucleic Acids and Molecular Biology 27, DOI 10.1007/978-3-642-25740-7_9,

insight into the details of RNA structure–function relationships. However, many RNAs contain both structured and functionally important but flexible elements. These RNAs are not amenable to structure determination in their intact forms by crystallography or NMR. Hence, molecular modeling of RNA to predict three-dimensional structure and dynamics is crucial for our understanding of RNA functions.

Currently, RNA folding tools focus mainly on predicting RNA secondary structure (Hofacker 2003; Mathews 2006; Zuker 2003). Using a dynamic programming approach (Eddy 2004), secondary structures are inferred by scoring nearest-neighbor stacking interactions with adjacent base pairs (Mathews 2006). These RNA secondary structure prediction methods play an important role in the current study of RNA. However, in order to model the tertiary structure of RNA molecules, it is necessary to explicitly model RNA in 3D. Cao and Chen designed a simplified diamond-lattice model for predicting folded structure and thermodynamics of RNA pseudoknots (Cao and Chen 2005, 2006). This approach quantitatively predicts the free energy landscape for sequence-dependent folding of RNA pseudoknots, in agreement with experimental observations (Cao and Chen 2005, 2006). However, due to lattice constraints and the dynamic issues associated with predefined Monte Carlo moves (Baumgartner 1987), off-lattice models are necessary to accurately model RNA 3D structure.

Computational tools for manually constructing RNA models have been developed for RNA 3D structure prediction (Shapiro et al. 2007). These methods use comparative sequence analysis to manually construct 3D models, with or without reference to a known, homologous 3D structure. Their accuracy is enhanced by use of experimental probes of secondary or tertiary structure and libraries of modular 3D motifs (Jossinet and Westhof 2005; Major et al. 1991, 1993; Massire et al. 1998; Massire and Westhof 1998; Shapiro et al. 2007; Tsai et al. 2003). Recently, significant progress has been made toward ab initio modeling of RNA 3D structures (Das and Baker 2007; Ding et al. 2008; Parisien and Major 2008). These studies show that starting only with sequence, it is possible to predict the structures of some small RNA motifs with atomic-level accuracy. However, as RNA length increases, the conformational space increases exponentially and the inherent inaccuracies of the force field accumulate, limiting the ability of current methods to predict the structures of large RNAs automatically. De novo prediction of large RNA structures with nontrivial tertiary folds from sequence alone remains beyond the realm of current ab initio algorithms.

We have developed a multiscale approach (Ding and Dokholyan 2005) for RNA modeling based on a coarse-grained RNA model for discrete molecular dynamics (DMD) simulations (Ding et al. 2008). DMD is a special type of molecular dynamics simulation in which pairwise interactions are approximated by stepwise functions. This approximation enables DMD to sample conformational space more efficiently than traditional molecular dynamics simulations (Dokholyan et al. 1998). Using the coarse-grained RNA model with DMD simulations, we were able to accurately fold a set of 150 small RNA molecules (<50 nt) within 6 Å (a majority within 4 Å) to their native states (Ding et al. 2008). To solve the folding problem of large RNA molecules with complex tertiary 3D structures, we proposed to incorporate experimentally derived structural information into our structure determination protocol. Long-range constraints for RNA modeling can be inferred from a variety of biochemical and bioinformatic techniques, ranging from chemical footprinting and cross linking to sequence covariation (Gutell et al. 1992; Juzumiene et al. 2001; Michel and Westhof 1990; Ziehler and Engelke 2001). Experimental constraints derived from these biochemical and bioinformatics techniques are generally of lower than atomic resolution, but can be readily incorporated into the coarse-grained RNA model for structure determination. The all-atom RNA model can then be reconstructed from the coarse-grained structural model.

First, we will describe our coarse-grained representation of RNA models for DMD simulations. Then, we will describe and evaluate the applications of the DMD–RNA procedure to ab initio folding of a set of small RNA models and structure determination using experimental constraints.

9.2 Coarse-Grained RNA Modeling Using Discrete Molecule Dynamics

We use DMD as the conformational sampling engine. A detailed description of the DMD algorithm can be found elsewhere (Dokholyan et al. 1998; Rapaport 2004; Zhou and Karplus 1997). The difference between discrete molecular dynamics and traditional molecular dynamics is in the interaction potential functions. Interatomic interactions in DMD are governed by stepwise potential functions (Fig. 9.1a). Neighboring interactions (such as bonds, bond angles, and dihedrals) are modeled by infinitely high square well potentials (Fig. 9.1b). By approximating the continuous potential functions with step functions of pairwise distances, DMD simulations are reduced to event-driven (collision) molecular dynamics simulation. In a DMD simulation, atoms move with constant velocity until they collide with another atom. As soon as the potential of interaction between the two atoms changes (i.e., the pairwise distance is at the step of the stepwise potential function), the velocities of the two interacting atoms change instantaneously (Fig. 9.1a). These velocity changes are required to conform to the conservation laws of energy, momentum, and angular momentum. Each such collision is termed an "event." The sampling efficiency of DMD over traditional MD is mainly due to rapid processing of collision events and localized updates of collisions (only colliding atoms are updated at each collision). In the limit of infinitesimally small steps, the discrete step function approaches the continuous potential function, and DMD simulations become equivalent to traditional molecular dynamics.

We approximate the single-stranded RNA molecule as a coarse-grained "beads-on-a-string" polymer with three beads representing each nucleotide, one for sugar (S), one for phosphate (P), and one for nucleotide base (B) (Fig. 9.2a). The P and S beads are positioned at the centers of mass of the corresponding phosphate group and the 5-atom ring sugar, respectively. For both purines (adenine and guanine) and

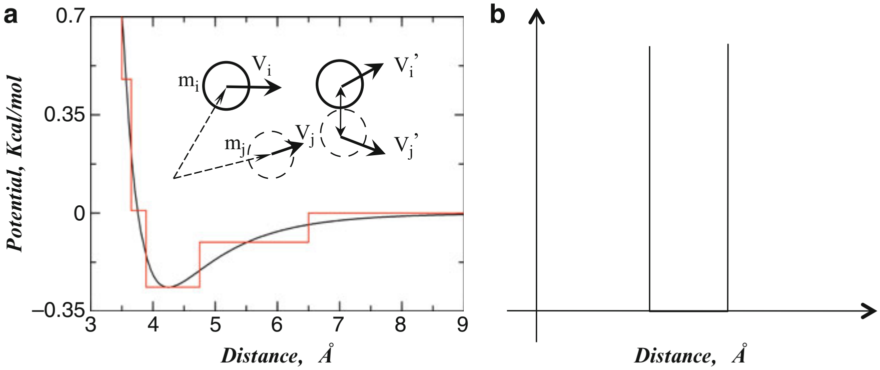


Fig. 9.1 Discrete molecular dynamics simulations. (**a**) Schematic of the DMD potential. The stepwise function used in DMD is the approximation of the continuous function in traditional molecular dynamics. The insert depicts the collision of two atoms with masses of m_i and m_j at the initial position of r_i and r_j, respectively. The two atoms move with constant velocities (v) until they meet at distance of R_{ij}. (**b**) Schematic of the potential energy of bonds in DMD. The atom pairs remain within the distance range during the simulation

pyrimidines (uracil and cytosine), we represent the base bead (B) as the center of the 6-atom ring. The neighboring beads along the sequence, which may represent moieties that belong to the same or a neighboring nucleotide, are constrained to mimic the chain connectivity and local chain geometry (Fig. 9.2a). Types of constraints include covalent bonds (solid lines), bond angles (dashed lines), and dihedral angles (dotted–dashed lines). The parameters for bonded interactions mimic the folded RNA structure and are derived from a high-resolution RNA structure database (Murray et al. 2003) (Table 9.1). Nonbonded interactions are crucial to model the folding dynamics of RNA molecules. In our model, we include base-pairing (Watson–Crick pairs of A–U and G–C and Wobble pair of U–G), base-stacking, short-range phosphate–phosphate repulsion, and hydrophobic interactions, which are described in the following section with the parameterization procedure.

Base Pairing. Two base-paired nucleotides have bases facing each other with the corresponding sugar and base beads aligned linearly. We use the "reaction" algorithm to model the orientation dependence of base-pairing interactions. The details of the algorithm can be found in (Ding et al. 2003). Briefly, to model the orientation dependence, we introduce auxiliary interactions in addition to the distance-dependent interactions between hydrogen bond donor and acceptor atoms (Fig. 9.2b). For example, when the two nucleotides (e.g., A–U, G–C, or U–G, represented as B_i and B_j in Fig. 9.2b) approach the interaction range, we evaluate the distances between S_iB_j and S_jB_i, which define the relative orientations of these two nucleotides. A hydrogen bond is allowed to form only when the distances fall within predetermined ranges. A schematic of the auxiliary interaction potential is shown in Fig. 9.2c, and the corresponding interaction parameters are listed in Table 9.2.

Hydrophobic Interactions and Overpacking. Buried inside the double-helix, the planar surface of bases are hydrophobic in nature. We include a weak attraction

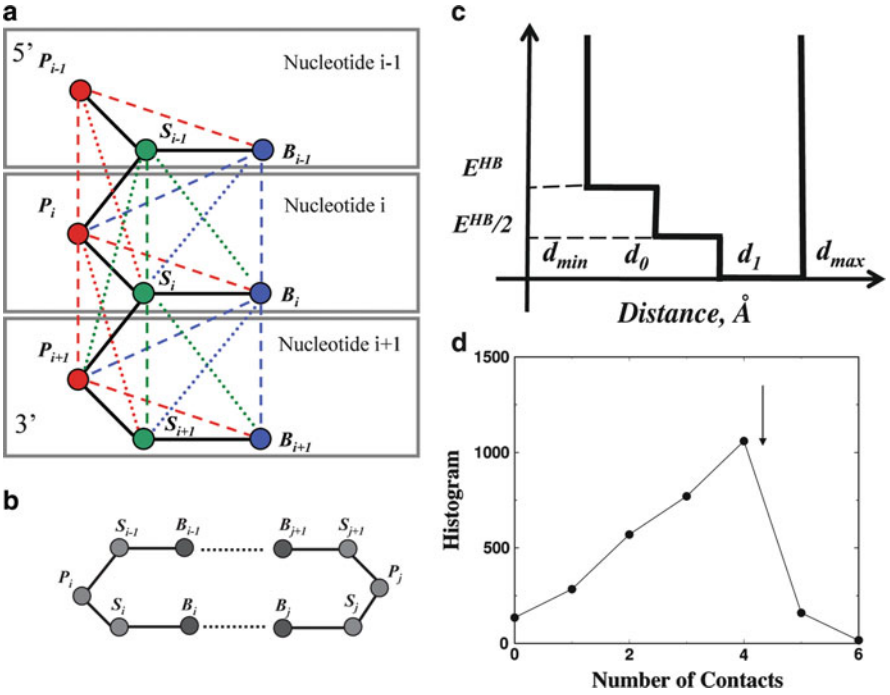


Fig. 9.2 Coarse-grained structural model of RNA employed in DMD simulations. (**a**) Three consecutive nucleotides, indexed $i-1$, i, $i+1$ are shown. The S, P, and B symbols correspond to loci of sugar, phosphate, and base beads in the RNA, respectively. Covalent interactions are shown as *thick lines*, angular constraints as *dashed lines*, and dihedral constraints as *dashed–dotted lines*. Additional steric constraints are used to model base stacking. (**b**) Hydrogen bonding in RNA base pairing. The base-pairing contacts between bases B_{i-1}:B_{j+1} and B_i:B_j are shown in *dashed lines*. A reaction algorithm is used (see Methods) for modeling the hydrogen-bonding interaction between specific nucleotide base pairs. (**c**) Schematic of the potential function for the auxiliary base-pairing interactions. (**d**) Histogram of the number of neighboring bases within a cutoff of 6.5 Å

between all the base beads. Due to the coarse-graining feature of our model, the assignment of attraction between bases results in overpacking (e.g., the symmetrically attractive interactions tend to form close packing). In order to avoid the artifact of overpacking, we first evaluate the packing observed in experimental 3D structures (http://ndbserver.rutgers.edu). We compute for each base the number of neighboring bases within a cutoff distance of 6.5 Å. The histogram of the number of neighbors is shown in Fig. 9.2d. Indeed, we find that the average number of neighbors is much smaller than that of close packing, 12. In order to avoid unrealistic close-packing due to the coarse-graining process, we introduce an effective energy term to penalize overpacking of bases:

$$E_{\text{overpack}} = \mathrm{d}E\Theta(n_{\text{c}} - n_{\max}), \qquad (9.1)$$

where $\Theta(x)$ is a step function,

$$\Theta(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}, \tag{9.2}$$

n_{c} is number of contacts, and $n_{\max}$ is the maximum number of contacts; dE is the repulsion coefficient. Based on the histogram of the number of base neighbors (Fig. 9.2d), we assign the value 4.2 for $n_{\max}$ and 0.6 kcal/mol for dE.

Base Stacking. To model stacking interactions, we assume that each base bead makes no more than two base–base stacking interactions and that three consecutively stacked base beads align approximately linearly. To determine the stacking interaction range between base beads, we compute center-to-center distances between base beads from known RNA structures. We find that distribution depends on base type (purine or pyrimidine) and identify stacking cutoff distances as 4.65 Å between purines, 4.60 between pyrimidines, and 3.80 Å between purine and pyrimidine. To approximately model the linearity of stacking interactions, two bases that form a stacking interaction to the same base are penalized for approaching closer than 6.5 Å. As a result, these three bases effectively define an obtuse angle. Next, we discuss the energy parameterization of base-stacking, base-pairing, and hydrophobic interactions.

Parameterization of Base-Pairing, Base-Stacking, and Hydrophobic Interactions. In order to determine the pairwise interaction parameters for stacking and hydrophobic interactions for all pairs of a base, we decomposed the sequence-dependent free energy parameters of the individual nearest-neighbor hydrogen bond model (INN-HB) (Mathews et al. 1999). We assume that the interaction of neighboring base pairs in INN-HB is the sum of all hydrogen-bond, base-stacking, and hydrophobic interactions. In a nearest neighboring base-pair configuration (Fig. 9.1), B_{i+1} and B_i (B_{j-1} and B_j) on one strand usually stack on top of each other. However, if both bases B_{i+1} and B_j are purines, we found that they tend to stack instead. The distance between bases B_i and B_{j-1} is usually greater than the cutoff distance of 6.5 Å for hydrophobic interactions. Therefore, we used the following equations to estimate the strength of pairwise interactions, where the first equation applies when B_{i+1}, B_j are both purines and the second equation applies otherwise:

$$E\begin{pmatrix} 5'\mathrm{B}_i\mathrm{B}_{i+1}3' \\ 3'\mathrm{B}_j\mathrm{B}_{j-1}5' \end{pmatrix} = \left(E^{\mathrm{HB}}_{\mathrm{B}_i\mathrm{B}_j} + E^{\mathrm{HB}}_{\mathrm{B}_{i+1}\mathrm{B}_{j-1}}\right) + E^{\mathrm{Stack}}_{\mathrm{B}_j\mathrm{B}_{i+1}} + E^{\mathrm{hydrophobic}}_{\mathrm{B}_i\mathrm{B}_{i+1}} + E^{\mathrm{hydrophobic}}_{\mathrm{B}_j\mathrm{B}_{j-1}}, \tag{9.3}$$

$$E\begin{pmatrix} 5'\mathrm{B}_i\mathrm{B}_{i+1}3' \\ 3'\mathrm{B}_j\mathrm{B}_{j-1}5' \end{pmatrix} = \left(E^{\mathrm{HB}}_{\mathrm{B}_i\mathrm{B}_j} + E^{\mathrm{HB}}_{B_{i+1}\mathrm{B}_{j-1}}\right) + E^{\mathrm{Stack}}_{\mathrm{B}_i\mathrm{B}_{i+1}} + E^{\mathrm{stack}}_{\mathrm{B}_j\mathrm{B}_{j-1}} + E^{\mathrm{hydrophobic}}_{\mathrm{B}_{i+1}\mathrm{B}_j}. \tag{9.4}$$

Here, E^{stack}, E^{HB}, and $E^{\mathrm{hydrophobic}}$ are the interaction strengths of base-stacking, base-pairing, and hydrophobic interactions, respectively. Given the experimentally tabulated energies between all possible neighboring base pairs (Mathews et al. 1999), we were able to determine values of E^{stack}, E^{HB}, and $E^{\mathrm{hydrophobic}}$ that are consistent with experimental measurements using singular value decomposition (Khatun et al. 2004; Press et al. 2002). The interaction parameters are listed in Tables 9.2 and 9.3.

Table 9.1 The averages and standard deviations of the bonded atom pairs

Bonded atom pair	Distance range (Å)
$P_i\ S_i$	4.55 ± 0.09
$S_i\ P_{i+1}$	4.10 ± 0.07
$S_i\ A_i$	4.85 ± 0.15
$S_i\ U_i$	3.74 ± 0.08
$S_i\ G_i$	4.81 ± 0.14
$S_i\ C_i$	3.70 ± 0.13
$P_i\ P_{i+1}$	6.25 ± 0.95
$S_i\ S_{i+1}$	5.72 ± 0.45
$P_i\ A_i$	7.45 ± 0.45
$P_i\ U_i$	5.57 ± 0.37
$P_i\ Gi$	7.43 ± 0.43
$P_i\ C_i$	5.57 ± 0.37
$A_i\ P_{i+1}$	7.25 ± 0.42
$U_i\ P_{i+1}$	6.40 ± 0.20
$G_i\ P_{i+1}$	7.20 ± 0.43
$C_i\ P_{i+1}$	6.40 ± 0.20
$P_{i-1}\ S_i$	9.25 ± 0.95
$S_{i-1}\ P_{i+1}$	8.96 ± 0.44
$A_{i-1}\ S_i$	5.68 ± 0.68
$U_{i-1}\ S_i$	6.38 + 0.73
$G_{i-1}\ S_i$	5.68 ± 0.68
$C_{i-1}\ S_i$	6.38 ± 0.73
$S_{i-1}\ A_i$	7.25 ± 0.60
$S_{i-1}\ U_i$	5.66 ± 0.54
$S_{i-1}\ G_i$	7.25 ± 0.60
$S_{i-1}\ C_i$	5.66 ± 0.54

All the bonds, angles, and dihedrals are effectively modeled using a bonded interaction in the DMD simulations (Fig. 9.1b). A, U, G, and C corresponds to four types of bases (B)

Table 9.2 The parameters for base pairing, modeled by hydrogen bonds between A–U, G–C, and U–G

Atom pair	$d_{\min}$ (Å)	d_0, (Å)	d_1, (Å)	$d_{\max}$ (Å)
C*i*–G*j* base pair				
Si Gj	7.70	8.08	8.63	9.00
Ci Sj	9.74	10.10	10.53	10.82
A*i*–U*j* base pair				
Si Uj	9.76	9.94	10.50	10.76
Ai Sj	7.72	7.92	8.82	9.00
U*i*–G*j* base pair				
Si Gj	7.00	7.44	8.24	8.70
Ui Sj	9.50	10.25	10.80	11.35

The details of the DMD algorithm for the hydrogen bond can be found in Ding et al. (2003). The schematic interaction potential is shown in Fig. 9.2c. The hydrogen bond strengths, E^{HB}, for A–U, G–C, and U–G are 0.5, 1.2, and 0.5 Kcal/mol, respectively. The interaction potential between the donor and acceptor is $-E^{\mathrm{HB}}$

Table 9.3 The stacking and hydrophobic interaction strengths, expressed in kcal/mol units

E^{Stack}	A_U	U_A	G_C	C_G	G_U	U_G
A_U	−0.45	−0.50	−0.75	−0.95	−0.42	−0.70
U_A	−0.50	−0.40	−0.55	−0.60	−0.35	−0.35
G_C	−0.75	−0.55	−0.81	−0.95	−0.48	−0.92
C_G	−0.95	−0.60	−0.95	−1.10	−0.47	−0.51
G_U	−0.42	−0.35	−0.48	−0.47	−0.52	0.62
U_G	−0.70	−0.35	−0.51	−0.51	0.62	−0.44
$E^{\mathrm{Hydrophobic}}$	A_U	U_A	G_C	C_G	G_U	U_G
A_U	−0.25	−0.40	−0.40	−0.50	−0.25	−0.35
U_A	−0.40	−0.30	−0.25	−0.25	−0.25	−0.25
G_C	−0.40	−0.25	−0.25	−0.45	−0.25	−0.41
C_G	−0.50	−0.25	−0.45	−0.50	−0.25	−0.41
G_U	−0.25	−0.25	−0.25	−0.25	−0.30	0.25
U_G	−0.35	−0.25	−0.41	−0.41	0.25	−0.25

The subscript indicates that the base bead is paired. For example, A_U is a base bead A that has been paired with a U bead. The cutoff distance for stacking interactions is 6.0 Å. The cutoff distance for hydrophobic interactions is 6.5 Å. The hardcore distance between all beads is set as 3.0 Å

Loop Entropy. Loop entropy plays a pivotal role in RNA folding kinetics and thermodynamics (Tinoco and Bustamante 1999). Hence, RNA folding prediction methods should take this entropic effect into account, either implicitly as in all-atom MD simulations (Sorin et al. 2004) or explicitly as in Monte Carlo or dynamic programming methods (Mathews 2006; Rivas and Eddy 1999). However, the reduction of degrees of freedom in our simplified RNA model causes entropy to be underestimated in DMD simulations. For example, we often observe formation of large loops that traps RNA molecules in nonnative conformations for significant simulation times. To overcome such artifacts arising from the coarse-graining process, we developed a simple modification of DMD simulation to model loop entropy explicitly. We use the free energy estimations for different types of loops, including hairpin, bulge, and internal loops (Mathews et al. 1999). Loop free energies were obtained from experimental fitting for small loops and extended to arbitrary lengths according to polymer theory. We compute the effective loop free energy in DMD simulations based on the set of base pairs formed in simulations. Upon the formation or breaking of each base pair, the total loop free energy changes according to the changes in either the number or size of loops. We estimate the changes in loop free energy, ΔG^{loop}, for each base pair formed during the simulation and determine the probability of forming such a base pair by coupling to a Monte Carlo procedure using a Metropolis algorithm with probability $p = \exp(-\beta\Delta G^{\mathrm{loop}})$. If the base pair is allowed to form stochastically, the particular base pair will form only if the kinetic energy is sufficient to overcome the possible potential energy difference before and after the base-pair formation. Upon breaking of a base pair, the stochastic procedure is not invoked since base-pair breakage is always entropically favorable. The breaking of a base pair is only governed by the conservation of momentum, energy, and angular momentum before and after the base-pair breakage.

The total potential energy, E, is obtained by adding all interaction terms, as given in (9.5):

$$E = E_{\mathrm{Bonded}} + E_{\mathrm{Hbond}} + E_{\mathrm{Stack}} + E_{\mathrm{Hydrophobic}} + E_{\mathrm{overpacking}} + G_{\mathrm{loop}}, \tag{9.5}$$

and is used to perform DMD simulations of RNA molecules. The energy landscape of RNA molecules is very rugged with a vast number of local minima due to the high degeneracy of nucleotide types (only 4 compared to the 20 different amino acids found in proteins). In order to efficiently sample the conformational space of RNAs, we utilize the replica-exchange sampling scheme (Okamoto 2004; Zhou et al. 2001).

Replica Exchange DMD. In replica exchange computing, multiple simulations or replicas of the same system are performed in parallel at different temperatures. Individual simulations are coupled through Monte Carlo-based exchanges of simulation temperatures between replicas at periodic time intervals. For two replicas, i and j, maintained at temperatures T_i and T_j and with energies E_i and E_j, temperatures are exchanged according to the canonical Metropolis criterion with exchange probability p, where $p = 1$ if $\Delta = \left(1/k_{\mathrm{B}}T_i - 1 - k_{\mathrm{B}}T_j\right)\left(E_j - E_i\right) \leq 0$, and $p = \exp(-\Delta)$, if $\Delta{>}0$. For simplicity, we use the same set of eight temperatures in all replica exchange simulations: 0.200, 0.208, 0.214, 0.220, 0.225, 0.230, 0.235, and 0.240. The temperature is in the abstract unit of kcal/(mol k_{B}). Note that we approximate the pairwise potential energy between coarse-grained beads with the experimentally determined free energy of nearest neighboring base pairs, instead of the actual enthalpy. As a result, the temperature does not directly correspond to physical temperatures. In DMD simulations, we maintain constant temperature using an Anderson thermostat (Andersen 1980).

Since the DMD code is highly optimized, we have found that the computational timescales linearly with respect to the system size. The folding simulation of a 50-nucleotide-long RNA sequence (median size of RNA chains in the sample) for 2×10^6 DMD simulation time units takes approximately 7 h of total wall-clock time, utilizing eight 2.33-GHz Intel Xeon compute nodes.

9.3 Ab Initio Folding of Small RNA Molecules

For each RNA molecule, we initially generated a linear conformation using the nucleotide sequence alone. Starting from this extended conformation, we performed replica exchange simulations at different temperatures as described above. From the simulation trajectories, we extracted sampled RNA conformational states, including the lowest energy state, the folding intermediate state, and the corresponding thermodynamic data. In Fig. 9.3, we illustrate the folding trajectory of one of the replicas for a turnip yellow mosaic virus (TYMV) pseudoknot (PDB ID: 1A60). An RNA pseudoknot structure has nonnested base pairing and minimally comprises base pairing between a loop region and a downstream RNA segment. Pseudoknots serve diverse biological functions, including

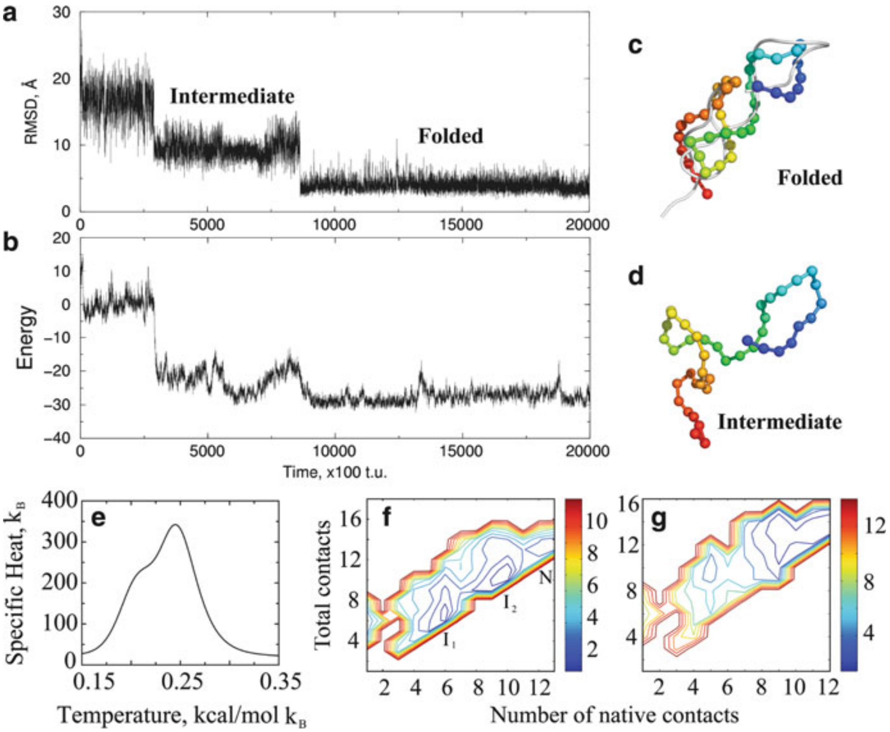


Fig. 9.3 Folding of a pseudoknot. For one replica, we present the RMSD (**a**) and energy (**b**) as the function of simulation time. Before folding into its native state (**c**), the molecule samples a folding intermediate state (**d**). (**e**) Specific heat is computed from the replica exchange trajectories using WHAM. (**f**) Two-dimensional potential of mean force 2D-PMF (potential mean force) for pseudoknot folding at $T^* = 0.245$ (corresponds to the major peak in the specific heat). The two intermediate states and the native state are indicated by I_1, I_2, and N, respectively. (**g**) The 2D-PMF plot at $T^* = 0.21$

formation of protein recognition sites that mediate replication and translational initiation, participation in self-cleaving ribozyme catalysis, and induction of frameshifts in translation of mRNA by ribosomes (Staple and Butcher 2005). For example, 1A60 is composed of a 5′-stem and a 3′-pseudoknot (Fig. 9.3c). From the simulation trajectory (Fig. 9.3), we observe folding of the RNA model within 5 Å root-mean-square deviation (RMSD) to the native state, and the lowest RMSD from the simulations is 2.03 Å. The lowest potential energy conformation, computed across all replicas using the effective free energy function in (9.5), has all native base pairs formed and an RMSD of 4.58 Å to the native state. Interestingly, we find that during the folding process the RNA molecule samples a stable folding intermediate state (Fig. 9.3a, b). The intermediate state forms a 5′-stem and a partially folded 3′-pseudoknot with one of the stems. Our identified folding intermediate state is consistent with the NMR studies of the solution structures of the TYMV pseudoknot and its 3′-stem (Kolk et al. 1998). Therefore, our DMD simulation not only allows the prediction of the native state but also enables us to identify folding intermediate states that might be important for the function of the RNA. The availability of multiple folding trajectories at different temperatures allows quantitative characterization of the folding thermodynamics.

We used the weighted histogram analysis method (WHAM) to calculate folding thermodynamics. The WHAM method utilizes multiple simulation trajectories with overlapping sampling along the reaction coordinates. The density of states $\rho(E)$ is self-consistently computed by combining histograms from different simulation trajectories (Kumar et al. 1992). Given the density of states, the folding specific heat (C_v) can be computed at different temperatures according to the partition function, $Z = \int \rho(E)\exp(-E/K_\mathrm{B}T)\mathrm{d}E$. To compute the potential of mean force (PMF) as a function of reaction coordinate A, we compute the conditional probability $P(A|E)$ of observing A at given energy E, which is evaluated from all simulation trajectories. Here, the reaction coordinate A can be any physical parameter describing the folding transitions, such as the number of native base pairs, the radius of gyration, or RMSD. The conditional probability $P(A|E)$ can be estimated from the histogram of parameter A for conformation states whose potential energies are within the range of $[E, E + dE]$. The PMF is computed as

$$\mathrm{PMF}(A) = -\ln\left(\int P(A|E)\rho(E)\exp(-E/K_\mathrm{B}T)\mathrm{d}E\right) + C. \tag{9.6}$$

Here, C is the reference constant, and we assign the lowest PMF a value of zero. Since our simulations start from fully extended conformations, we exclude the trajectories from the first 5×10^5 time units and use those of the last 1.5×10^6 time units for WHAM analysis. We used the trajectories from all replicas to compute histograms. In Fig. 9.3e–g, we illustrate the folding thermodynamics of 1A60 using WHAM analysis, including the specific heat and potential mean field. The specific heat (Fig. 9.3e) has one peak centered at temperature $T^* = 0.245$ and a shoulder near $T^* = 0.21$, suggesting the presence of intermediate states in the folding pathway. The thermodynamic folding intermediate species is characterized by computing the two-dimensional potential of mean force (2D-PMF) as a function of the total number of base pairs (N) and the number of native base pairs (NN). The 2D-PMF plots at temperatures corresponding to the two peaks in the specific heat (Fig. 9.3f, g) show two intermediate states with distinct free energy basins: the first intermediate state corresponds to the folded 5′-hairpin, while the second intermediate corresponds to the formation of one of the helix stems for the 3′-pseudoknot. For example, the 2D-PMF plot at $T^* = 0.21$ (Fig. 9.3g) shows that the shoulder in the specific heat plot corresponds to the formation of the second intermediate state. The basins corresponding to the two intermediate states have a weak barrier, resulting in a lower peak height in the specific heat plot. Therefore, the coarse-grained RNA model combined with the DMD sampling algorithm allows the modeling of RNA structure as well as folding thermodynamics.

We benchmarked the DMD–RNA model on a set of 153 RNAs with length up to 100 nucleotides (Ding et al. 2008). For a majority of the simulated RNA sequences, the lowest energy structures from simulations have a percentage of native base pairs, or *Q*-value, close to unity, suggesting the correct formation of native base pairs in simulations. Here, we only considered the base pairs of A–U, G–C, and U–G. The other commonly observed Wobble pairing, A–G, was not included in the benchmark study but will be included in future studies. The average *Q*-value for all 153 RNA molecules studied is 94%. For comparison with available secondary structure prediction methods, we also computed the *Q*-values using Mfold, which yielded an average *Q*-value of 91%. Given the high percentage of correctly predicted base pairs (94%) and the relatively simple topology of the studied RNA molecules, the average number of incorrectly predicted base pairs is less than one.

The RMSD between predicted and experimental structures is often computed to evaluate the accuracy of predicted tertiary structures. Although the RMSD calculation does not provide detailed information on local structural features such as base pairing and base stacking, it gives a straightforward measure of the overall structure prediction. Recently, we have developed an approach to evaluate the statistical significance of RNA 3D structure prediction with a given RMSD for different lengths (Hajdin et al. 2010). Alternatively, Parisien et al. (2009) have proposed new metrics to account for both local and global structural information during structural comparison. However, their calculation requires the atomic structure of the prediction. To evaluate the overall 3D fold of our coarse-grained models, we computed the RMSD to compare our predictions with experimental structures. We found that for RNA molecules with nucleotide length $<$ 50 nt, the RMSD of predicted structures are less than 6 Å. Predictions of longer RNAs exhibit larger RMSD due to the highly flexible nature of RNA molecules. Among the 153 sequences simulated, 84% of the predicted tertiary structures have an RMSD of <4 Å with respect to the experimentally derived native RNA structure. The benchmark results highlight the predictive power of the DMD–RNA methodology, at least for small RNA molecules.

Three out of 153 RNA molecules studied are longer than 65 nucleotides, where the DMD–RNA method cannot be applied to predict the native secondary and tertiary structure from sequence alone. The challenges to predict large RNA folding ab initio arise from the exponentially increasing size of the conformational space and inaccuracies in the force field. Therefore, it is important to develop new approaches to predict the 3D fold of large RNA molecules.

9.4 Automated RNA Structure Determination Using Experimental Constraints

RNA structural information including secondary structure and some tertiary interactions can often be derived experimentally and computationally prior to the determination of high-resolution 3D structure. Accurate RNA secondary structures can be obtained from comparative sequence analysis (Gutell et al. 2002; Michel and Westhof 1990) and experimentally constrained prediction (Deigan et al. 2009).

SHAPE chemistry (selective 2′-hydroxyl acylation analyzed by primer extension) was recently shown to be a powerful approach for analyzing secondary structure at single nucleotide resolution for RNAs of any length (Merino et al. 2005; Wilkinson et al. 2006). SHAPE exploits the discovery that the 2′-OH group in unconstrained or flexible nucleotides reacts preferentially with hydroxyl-selective electrophilic reagents. In contrast, nucleotides constrained by base-pairing or tertiary interactions are unreactive. The resulting reactivity information can be used, in concert with a secondary structure prediction algorithm, to obtain accurate secondary structures (Deigan et al. 2009; Mathews et al. 2004; Mortimer and Weeks 2007; Wang et al. 2008; Wilkinson et al. 2008). Long-range interactions of RNA molecules can also be inferred by biochemical and bioinformatic methods, such as dimethyl sulfate (DMS) modification (Jan and Sarnow 2002; Flor et al. 1989), hydroxyl radical protection (Murphy and Cech 1994), mutational analysis (Kanamori and Nakashima 2001; De la Pena et al. 2003; Khvorova et al. 2003; Murphy and Cech 1994; Wang et al. 1995), and sequence covariation (Cannone et al. 2002). Therefore, we propose to incorporate experimentally determined secondary and tertiary structure information into DMD simulations to reconstruct a conformational ensemble that is consistent with experimental measurements.

In general, existing programs for modeling complex RNAs use either computationally intensive all-atom reconstruction, which limits their applications to small RNAs, or overly simplified models that omit key structural details. Other challenges in many current approaches are requirements for high levels of expert user intervention or comparative sequence information and the reliance on chemical intuition derived from preexisting information on tertiary interactions [reviewed in (Shapiro et al. 2007)]. Here, we developed an approach for accurate de novo determination of RNA tertiary fold that does not require expert user intervention nor impose heavy computational requirements, and that is efficient for large RNAs (Fig. 9.4). The approach takes an input list of base pairs and distance constraints between specific pairs of nucleotides and outputs a structural ensemble that is consistent with the input constraints. Starting from the extended conformation, we performed DMD simulations with biased potential for base-pairing constraints. Iterative DMD optimization was performed until all base pairs formed. After base-pair formation was confirmed, long-range interaction constraints were added for DMD simulated annealing simulations. At the end of each simulated annealing simulation, we devised filters to evaluate the simulation results, including radius of gyration and/or number of satisfied long-range constraints. We performed iterative annealing simulations until all filters were satisfied and, after constructing the structural ensemble from simulation trajectories, performed cluster analysis to identify representative structures. In all DMD simulations, only serial computation (instead of replica exchange) was used, which also reduced the computational requirement.

We tested the automated structure refinement method on tRNA^{asp} (Gherghe et al. 2009). Base pairing from the X-ray crystallography structure was consistent with the SHAPE-derived secondary structures. Long-range distance constraints were determined using a site-directed footprinting experiment. An Fe(II)-EDTA moiety was tethered specifically to RNA using the site-selective intercalation reagent

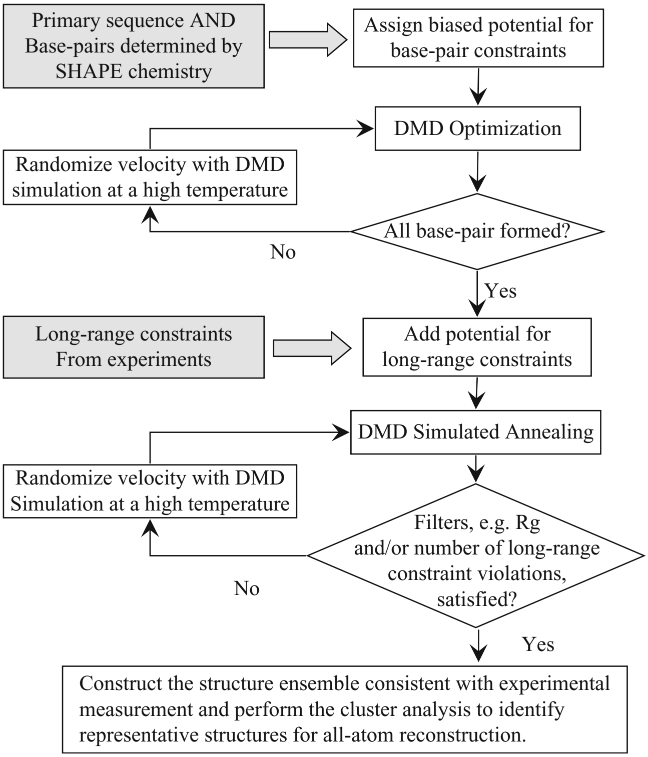
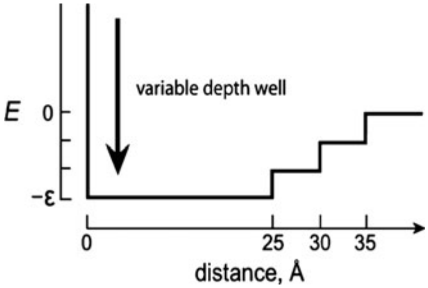

Primary sequence AND Base-pairs determined by SHAPE chemistry
Assign biased potential for base-pair constraints
DMD Optimization
Randomize velocity with DMD simulation at a high temperature
All base-pair formed?
No
Yes
Long-range constraints From experiments
Add potential for long-range constraints
DMD Simulated Annealing
Randomize velocity with DMD Simulation at a high temperature
Filters, e.g. Rg and/or number of long-range constraint violations, satisfied?
No
Yes
Construct the structure ensemble consistent with experimental measurement and perform the cluster analysis to identify representative structures for all-atom reconstruction.

Fig. 9.4 Flowchart of the DMD–RNA structure determination method using experimentally derived structural information

methidiumpropyl-EDTA (MPE) (Hertzberg and Dervan 1982). MPE preferentially intercalates at CpG steps in RNA at sites adjacent to a single-nucleotide bulge (White and Draper 1987; White and Draper 1989), which can be introduced by mutations in helical regions. To apply the cleavage information to bias DMD simulations, we developed a generic approach to interpret each cleavage event as a distance constraint (Fig. 9.5). The interaction potential features a "soft" energy wall at 25 Å, with smaller energy bonuses extending out to 35 Å (Fig. 9.5). The 25-Å barrier corresponds to the distance cutoff within which the nucleotides exhibit strong cleavage and beyond which the nucleotides have weak cleavage. The interaction strength is assigned according to the cleavage intensity [$E \propto \ln(I/<I>)$]. This approach has two advantages: (1) no user input is required to decide whether a given cleavage is significant or not and (2) structure refinement is highly tolerant of measurement errors inherent in any hydroxyl radical footprinting experiment. By using this structure determination approach (Fig. 9.5), we were able to refine the structure of tRNA^{asp} to 6.4 Å RMSD relative to the crystal structure (Gherghe et al. 2009).

Recently, we applied the structure refinement methodology on four RNAs: domain III of the cricket paralysis virus internal ribosome entry site (CrPV)

Fig. 9.5 Potential function used to convert experimental cleavage information into DMD potential energy constraints



(49 nts), a full-length hammerhead ribozyme from *S. mansoni* (HHR) (67 nts), *S. cerevisiae* tRNA^{Asp} (75 nts), and the P546 domain of the *T. thermophilia* group I intron (P546) (158 nts). Each of these RNAs has a complex three-dimensional fold, involving more than simple intrahelix interactions. Prior to publication of the high-resolution structures (Cate et al. 1996; Costantino et al. 2008; Martick and Scott 2006; Westhof et al. 1988), significant biochemical or bioinformatic data describing tertiary interactions were available for each RNA. The secondary structure was also known to high accuracy in each case. Only this prior information was used during DMD refinement. In all cases, we were able to generate a low-RMSD structure. The RMSD between the predicted structure and the native state for the CrPV, HHR, tRNA^{Asp}, and P546 RNAs are 3.6, 5.4, 6.4, and 11.3 Å, respectively (Lavender et al. 2010). Calculations were performed on a Linux workstation (Intel Pentium 4 processor, 3.2 GHz) and the CPU times ranged from 18 (CrPV, 49 nts) to 42 h (P546, 158 nts). Therefore, the combination of efficient DMD simulations and sufficient biochemical experiments can accurately determine RNA structure of arbitrary length.

9.5 Conclusions

We have developed a multiscale RNA modeling approach to model 3D structure and dynamics of RNAs having a wide range of lengths. We use a coarse-grained representation of the RNA to efficiently model the conformational space. For short RNA molecules (<50 nt), we are able to capture the folded state from the sequence alone. The availability of replica-exchange simulation trajectories at multiple temperatures allows for the characterization of folding thermodynamics as well as capture of the final folded state. To efficiently sample the exponentially increasing conformational space of large RNA molecules, we devised an automated modeling approach to determine large and complex RNA structures using experimentally derived structural information. A benchmark study (Lavender et al. 2010) highlights the application of combining DMD simulation and experimental structural information to yield native-like models for the diverse universe of functionally important RNAs whose structures cannot be characterized by conventional methods.

References

Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 72:10

Baumgartner A (1987) Applications of the Monte-Carlo simulations in statistical physics. Springer, New York

Cannone J, Subramanian S, Schnare M, Collett J, D'Souza L, Du Y, Feng B, Lin N, Madabusi L, Müller K et al (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3:e2

Cao S, Chen SJ (2005) Predicting RNA folding thermodynamics with a reduced chain representation model. RNA 11:1884–1897

Cao S, Chen SJ (2006) Predicting RNA pseudoknot folding thermodynamics. Nucleic Acids Res 34:2634–2652

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. Science 273:1678–1685

Costantino DA, Pfingsten JS, Rambo RP, Kieft JS (2008) tRNA-mRNA mimicry drives translation initiation from a viral IRES. Nat Struct Mol Biol 15:57–64

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

De la Pena M, Gago S, Flores R (2003) Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. EMBO J 22:5561–5570

Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci USA 106:97–102

Ding F, Borreguero JM, Buldyrey SV, Stanley HE, Dokholyan NV (2003) Mechanism for the alpha-helix to beta-hairpin transition. Proteins 53:220–228

Ding F, Dokholyan NV (2005) Simple but predictive protein models. Trends Biotechnol 23:450–455

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

Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI (1998) Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des 3:577–587

Eddy SR (2004) How do RNA folding algorithms work? Nat Biotechnol 22:1457–1458

Edwards TE, Klein DJ, Ferre-D'Amare AR (2007) Riboswitches: small-molecule recognition by gene regulatory RNAs. Curr Opin Chem Biol 17:273–279

Flor PJ, Flanegan JB, Cech TR (1989) A conserved base pair within helix P4 of the Tetrahymena ribozyme helps to form the tertiary structure required for self-splicing. EMBO J 8:3391–3399

Gherghe CM, Leonard CW, Ding F, Dokholyan NV, Weeks KM (2009) Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. J Am Chem Soc 131:2541–2546

Gutell RR, Lee JC, Cannone JJ (2002) The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol 12:301–310

Gutell RR, Power A, Hertz GZ, Putz EJ, Stormo GD (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res 20:5785–5795

Hajdin CE, Ding F, Dokholyan NV, Weeks KM (2010) On the significance of an RNA tertiary structure prediction. RNA 16:1340–1349

Hertzberg RP, Dervan PB (1982) Cleavage of double helical DNA by (Methidiumpropyl-EDTA) iron(II). J Am Chem Soc 104:313–315

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431

Jan E, Sarnow P (2002) Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. J Mol Biol 324:889–902

Jossinet F, Westhof E (2005) Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. Bioinformatics 21:3320–3321

Juzumiene D, Shapkina T, Kirillov S, Wollenzien P (2001) Short-range RNA-RNA crosslinking methods to determine rRNA structure and interactions. Methods 25:333–343

Kanamori Y, Nakashima N (2001) A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. RNA 7:266–274

Khatun J, Khare SD, Dokholyan NV (2004) Can contact potentials reliably predict stability of proteins? J Mol Biol 336:1223–1238

Khvorova A, Lescoute A, Westhof E, Jayasena SD (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10:708–712

Kolk MH, van der Graaf M, Fransen CT, Wijmenga SS, Pleij CW, Heus HA, Hilbers CW (1998) Structure of the 3′-hairpin of the TYMV pseudoknot: preformation in RNA folding. EMBO J 17:7498–7504

Kumar S, Bouzida D, Swendswn RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method. J Computat Chem 13:11

Lavender CA, Ding F, Dokholyan NV, Weeks KM (2010) Robust and generic RNA modeling using inferred constraints: a structure for the hepatitis C virus IRES pseudoknot domain. Biochemistry 49:4931–4933

Major F, Gautheret D, Cedergren R (1993) Reproducing the three-dimensional structure of a tRNA molecule from structural constraints. Proc Natl Acad Sci U S A 90:9408–9412

Major F, Turcotte M, Gautheret D, Lapalme G, Fillion E, Cedergren R (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA. Science 253:1255–1260

Martick M, Scott WG (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 126:309–320

Massire C, Jaeger L, Westhof E (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. J Mol Biol 279:773–793

Massire C, Westhof E (1998) MANIP: an interactive tool for modelling RNA. J Mol Graph Model 16(197–205):255–197

Mathews DH (2006) Revolutions in RNA secondary structure prediction. J Mol Biol 359:526–532

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA 101:7287–7292

Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288:911–940

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127:4223–4231

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610

Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc 129:4144–4145

Murphy FL, Cech TR (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. J Mol Biol 236:49–63

Murray LJ, Arendall WB 3rd, Richardson DC, Richardson JS (2003) RNA backbone is rotameric. Proc Natl Acad Sci U S A 100:13904–13909

Okamoto Y (2004) Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. J Mol Graph Model 22:425–439

Parisien M, Cruz JA, Westhof E, Major F (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. RNA 15:1875–1885

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

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2002) Numerical Recipes in C, 2nd edn. Cambridge University Press, Cambridge

Rapaport DC (2004) The art of molecular dynamics simulation. Cambridge University Press, Cambridge

Rivas E, Eddy SR (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. J Mol Biol 285:2053–2068

Shapiro BA, Yingling YG, Kasprzak W, Bindewald E (2007) Bridging the gap in RNA structure prediction. Curr Opin Struct Biol 17:157–165

Sorin EJ, Nakatani BJ, Rhee YM, Jayachandran G, Vishal V, Pande VS (2004) Does native state topology determine the RNA folding mechanism? J Mol Biol 337:789–797

Staple DW, Butcher SE (2005) Pseudoknots: RNA structures with diverse functions. PLoS Biol 3: e213

Tinoco I Jr, Bustamante C (1999) How RNA folds. J Mol Biol 293:271–281

Tsai HY, Masquida B, Biswas R, Westhof E, Gopalan V (2003) Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. J Mol Biol 325:661–675

Vicens Q, Cech TR (2006) Atomic level architecture of group I introns revealed. Trends Biochem Sci 31:41–51

Wang B, Wilkinson KA, Weeks KM (2008) Complex ligand-induced conformational changes in tRNA^{Asp} revealed by single nucleotide resolution SHAPE chemistry. Biochemistry 47:3454–3461

Wang C, Le SY, Ali N, Siddiqui A (1995) An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5′ noncoding region. RNA 1:526–537

Westhof E, Dumas P, Moras D (1988) Restrained refinement of 2 crystalline forms of yeast aspartic-acid and phenylalanine transfer-Rna crystals. Acta Crystallographica Sect A 44:112–123

White SA, Draper DE (1987) Single base bulges in small RNA hairpins enhance ethidium binding and promote an allosteric transition. Nucleic Acids Res 15:4049–4064

White SA, Draper DE (1989) Effects of single-base bulges on intercalator binding to small RNA and DNA hairpins and a ribosomal RNA fragment. Biochemistry 28:1892–1897

Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol 6:e96

Wilkinson KA, Merino EJ, Weeks KM (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protocol 1:1610–1616

Zhou R, Berne BJ, Germain R (2001) The free energy landscape for beta hairpin folding in explicit water. Proc Natl Acad Sci U S A 98:14931–14936

Zhou Y, Karplus M (1997) Folding thermodynamics of a model three-helix-bundle protein. Proc Natl Acad Sci USA 94:14429–14432

Ziehler WA, and Engelke DR (2001). Probing RNA structure with chemical reagents and enzymes. Curr Protoc Nucleic Acid Chem *Chapter 6*, Unit 6 1

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415