



RNA 3D Structure Prediction Using Coarse-Grained Models

Jun Li and Shi-Jie Chen*

Departments of Physics and Biochemistry, and Institute of Data Science and Informatics, University of Missouri, Columbia, MO, United States

The three-dimensional (3D) structures of Ribonucleic acid (RNA) molecules are essential to understanding their various and important biological functions. However, experimental determination of the atomic structures is laborious and technically difficult. The large gap between the number of sequences and the experimentally determined structures enables the thriving development of computational approaches to modeling RNAs. However, computational methods based on all-atom simulations are intractable for large RNA systems, which demand long time simulations. Facing such a challenge, many coarse-grained (CG) models have been developed. Here, we provide a review of CG models for modeling RNA 3D structures, compare the performance of the different models, and offer insights into potential future developments.

Keywords: RNA, structure prediction, coarse-grained, molecular dynamics, monte carlo, statistical potential, all-atom force field

OPEN ACCESS

Edited by:

Yi He,
University of New Mexico,
United States

Reviewed by:

Adam Liwo,
University of Gdansk, Poland
Philippe Derreumaux,
UPR9080 Laboratoire de Biochimie
Théorique (LBT), France

*Correspondence:

Shi-Jie Chen
chenshi@missouri.edu

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 05 June 2021

Accepted: 17 June 2021

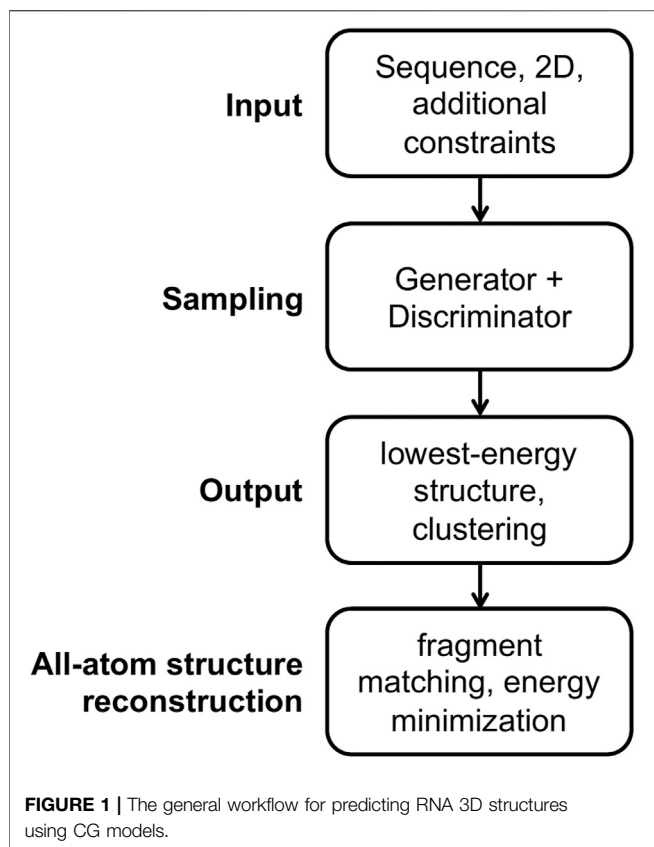
Published: 02 July 2021

Citation:

Li J and Chen S-J (2021) RNA 3D
Structure Prediction Using Coarse-
Grained Models.
Front. Mol. Biosci. 8:720937.
doi: 10.3389/fmolb.2021.720937

1 INTRODUCTION

Ribonucleic acid (RNA) is an unbranched polymeric macromolecule undertaking crucial and various biological functions, such as carrying genetic information, directing protein synthesis, regulating gene expression, catalyzing biological reactions (Gesteland et al., 2005; Cech and Steitz, 2014; Morris and Mattick, 2014; Mortimer et al., 2014). Exogenous RNAs can also exert deleterious effects on human health. For example, SARS-CoV-2, an RNA virus, caused the serious respiratory disease COVID-19 in the current pandemic (Tu et al., 2020; Oran and Topol, 2020). Moreover, recently developed mRNA vaccines suggest that RNAs can be used as an antiviral therapeutic agents against COVID-19 by making antibodies through human immune response (Amanat and Krammer, 2020; Krammer, 2020; Moore and Offit, 2021). RNA functions are tied to their 3D structures. For example, the formation of pseudoknot structure in SARS-CoV-2 frameshift stimulating element is required for ribosomal frameshifting which can lead to the production of alternative proteins from the same genomic region (Kelly et al., 2020). Therefore, detailed RNA 3D structures are important for understanding their functions and for drug design. To determine high-resolution 3D structures, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy experiments are conventionally performed, but it is laborious and tough to solve structures by these methods. Meanwhile, numerous RNA sequences have been obtained but their 3D structures remain unknown. Consequently, the huge gap between the number of sequences and the known structures stimulates the development of computational 3D structure prediction methods over the past decades (Miao and Westhof, 2017). In terms of the resolution of the RNA structural model, these methods can be classified into two categories, namely all-atom (AA) and coarse-grained (CG) approaches. For AA approaches, 20–23 heavy atoms and 10–11 hydrogen atoms are considered for each nucleotide, while for some CG approaches, several atoms are grouped together into a bead and only several beads are used to model a nucleotide, and for some other CG approaches, even more coarse-grained



representations are used by treating helix and loop structures as vertices and edges. For *de novo* modeling methods without using solved homologous structures, sampling in the energy landscape is often required and the lowest free energy structure is searched for (Anfinsen, 1973). Compared with AA methods, the energy landscape in CG models is smoother, enabling faster arrival at the global energy minima, and the computation should be more rapid due to fewer atoms. Next we will introduce a dozen of CG approaches to modeling RNA 3D structures.

2 COARSE-GRAINED MODELS FOR RIBONUCLEIC ACID 3D STRUCTURE MODELING

The general workflow for predicting RNA 3D structures using CG models is shown in **Figure 1**. There are four parts in the workflow, namely input, sampling, output and all-atom structure reconstruction. We will go through them next.

Input For the input part, an RNA sequence is usually required, and a secondary structure is essential for some models and is optional for others. Additional constraints are applicable for some models, such as experimental information, distance between specified atom pairs, initial 3D structures, and templates for motifs.

Sampling For the sampling part, a conformational generator for producing candidate structures and a conformational

discriminator for assessing structures are essential. A generator is commonly implemented in a Monte Carlo (MC) or molecular dynamics (MD) simulation framework, and many variants of MC and MD engines are utilized to accelerate conformational sampling. Generally, a discriminator is called scoring function, (potential) energy function, or force field in structure prediction field. There are three main forms of energy functions used in the CG models. The first type of CG energy function mimics the all-atom force field including the bonded terms of bond lengths, angles and torsions, and the non-bonded terms of Van der Waals and electrostatic energies, as shown in **Eq. 1** for the AMBER force field (Wang et al., 2004) where harmonic potentials are used for bonds and angles, a cosine form potential is used for torsions, the Lennard-Jones potential is used for Van der Waals interactions, and the Coulomb interaction energy is used for electrostatic interactions.

$$\begin{aligned}
 E_{\text{AMBER}} = & \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\
 & + \sum_{i < j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right) + \sum_{i < j} \left(\frac{q_i q_j}{\epsilon R_{ij}} \right)
 \end{aligned} \quad (1)$$

Van der Waals
electrostatic

The second type of CG energy function is the potential of mean force (PMF) (Sippl, 1990) using the inverse Boltzmann formula as shown in **Eq. 2**, where $E(x)$ is the PMF for a structural variable x , k_B is the Boltzmann constant, T is the temperature, $p_{\text{obs}}(x)$ is the probability distribution for x observed in the experimental structures, and $p_{\text{ref}}(x)$ is the probability distribution for x in the reference state. The common structural variable x is the nucleotide- or atom-dependent interatomic distance, angle, torsion or the combination of them. We also call this potential statistical potential or knowledge-based potential (Capriotti et al., 2011; Wang et al., 2015; Tan et al., 2019).

$$E(x) = -k_B T \ln \left[\frac{p_{\text{obs}}(x)}{p_{\text{ref}}(x)} \right] \quad (2)$$

The third type of CG energy function is the hybrid of the former two energies. For example, for the bonded energies, the harmonic, Gaussian or other energy functions are used mimicking the all-atom force field, while for the non-bonded energies, the PMF energy is used to account for base pairing and base stacking interactions. The parameters in the bonded energy functions are usually fitted to the statistical potentials derived from the experimental structures (Tan et al., 2006; Jonikas et al., 2009b; Xia et al., 2013). In addition to the above three main forms of energy functions, some models use the sum and product of several harmonic potentials, Gaussian energy functions or other functions as a combined complex energy function (Šulc et al., 2014; Zhang et al., 2021). Some CG models can integrate additional experimental information into the energy functions by transforming the information into distance, angle, torsion and other constraints (Tan et al., 2006; Jonikas et al., 2009b; Xia et al.,

TABLE 1 | Summary of CG models for RNA 3D structure modeling.

Model	Representation	Energy	Applicable constraints ^a	Sampling	Output	Availability
Graph	Helix-based	KB ²	No	MC ⁷ + SA ⁸	AA	No
Ernwin	Helix-based	KB	No	MCMC ⁹	CG ¹⁵	Standalone + server
YUP	1 bead/nt ¹	KB	Footprinting, cross-link, etc.	MC + SA	CG	No
NAST	1 bead/nt	KB	Tertiary contact	MD ¹⁰	CG/AA ¹⁶	Standalone
iFoldRNA	3 beads/nt	KB + TD ³	HRP ⁴ , NMR ⁵	REDMD ¹¹	AA	Server
TOPRNA	3 beads/nt	KB	No	REMD ¹²	CG	Standalone
IsRNA1	4–5 beads/nt	KB	No	REMD	AA	Standalone + sever
Ren's	5 beads/nt	KB	NMR, SAXS ⁶	MD/SA	AA	No
oxRNA	5 beads/nt	TD	No	MC + US ¹³ + MD	CG	Standalone + server
SimRNA	5 beads/nt	KB	Distance	REMC ¹⁴	AA	Standalone + server
SPQR	5 beads/nt	KB	No	MC + SA	AA	Standalone
HIRE-RNA	6–7 beads/nt	KB	No	REMD	AA	No

^aApplicable constraints include 2D structure constraints for all models.

¹nt: nucleotide.

²KB: knowledge-based.

³TD: thermodynamics.

⁴HRP: hydroxyl radical probing.

⁵NMR: nuclear magnetic resonance.

⁶SAXS: small-angle X-ray scattering.

⁷MC: Monte Carlo.

⁸SA: simulated annealing.

⁹MCMC: Markov chain Monte Carlo.

¹⁰MD: molecular dynamics.

¹¹REDMD: replica exchange discrete molecular dynamics.

¹²REMD: replica exchange molecular dynamics.

¹³US: umbrella sampling.

¹⁴REMC: replica exchange Monte Carlo.

¹⁵CG: coarse grained.

¹⁶AA: all atom.

2013; Krokhotin et al., 2015). Generally the conformational generator and discriminator are not independent. Ideally the discriminator would lead the generator towards the low-energy conformational space. However, the generator may be trapped in the local minima in the energy landscape, and advanced sampling algorithms may help the generator traverse these local minima efficiently.

Output For the output part, some models select the structure with the lowest energy as the predicted native structure, while other models first group the low-energy structures into several clusters, and choose the centroid structures in the clusters as the predicted structures. The predicted best centroid structure is from the largest cluster, which accounts implicitly for the entropy effect.

All-atom structure reconstruction All-atom structures are essential for studying ligand binding and drug design and for structure-based understanding of RNA functions. For the all-atom structure reconstruction part, the fragment matching algorithm (Jonikas et al., 2009a) is often used, followed by structure refinement to remove steric clashes and chain breaks.

2.1 Models

Twelve CG models are described next and summarized in **Table 1**. The models are sorted first by the CG level and then by the published time.

Graph-based model (Kim et al., 2014) The model maps a 2D structure to a graph, where a helix is represented by two connected vertices bridging the helix ends, a hairpin loop by a

vertex attached to the stem, internal loops by a vertex connecting two helices, and N-way junction loops by a vertex connecting N helices. Like the all-atom force field, the connected vertices form bonds, angles, and torsions. For internal loops, a knowledge-based statistical potential is constructed in terms of the bond angles and torsions. For 3-way and 4-way junctions, RNAJAG program (Laing et al., 2013) predicts topologies and determines the coordinates of the junction vertices. In addition, a radius of gyration-related term is added to determine the global compactness. Monte Carlo/simulated annealing (MC/SA) (Van Laarhoven and Aarts, 1987; Robert and Casella, 2013) methods are used for the 3D graph sampling. The model can predict global helical arrangements compatible with a given secondary structure and give all-atom structures.

Ernwin (Kerpedjiev et al., 2015) Ernwin is also a helix-based coarse-grained model which maps helices to reduced cylinders and maps loops to edges attached to a helix or connecting two helices. The relative orientations and positions of two helices are determined by the loops connecting them. The model uses Markov chain Monte Carlo (MCMC) simulation to sample the 3D arrangements of the helices in a force field with five energy terms, including steric clash energy, junction closure energy, radius of gyration-related energy, A-minor energy favoring the long-range interaction in A-minor motif, and loop-loop interaction energy. The last three energy terms are knowledge-based potential of mean force. The model can efficiently predict plausible global shapes of large RNAs. The program is available at <http://github.com/pkerpedjiev/ernwin>. A server of this program

can be found at <http://rna.tbi.univie.ac.at/ernwin>. RNA sequence and secondary structure are essential inputs to the server and only coarse-grained structures can be retrieved at the present time.

YUP (Malhotra et al., 1994; Tan et al., 2006) The model uses one bead to represent a nucleotide located at atom P, and uses harmonic potential functions to describe bond stretching, bond angle bending, and torsion angle twisting. The bond equilibrium values are determined from experiment structures and the force constants are fitted to reflect the uncertainties of the bond lengths, angles, and torsions. The model is used to refine low-resolution structures with additional experimental constraints.

NAST (Jonikas et al., 2009b) The NAST model simplifies a nucleotide by using only the C3' atom. Harmonic potentials are applied to the bond lengths and angles, and a three-term cosine function is used for torsions, and the energy parameters are fitted to the statistical potentials derived from three large ribosomal RNAs. Moreover, the repulsive term of the Lennard-Jones potential is applied to avoid steric overlap. Secondary structure and tertiary interaction constraints are included in the energy function. MD simulations are used for conformational sampling. It is computationally efficient to sample the conformational space for large RNAs with the given 2D and 3D constraints. This model, however, cannot *de novo* fold RNAs solely based on sequence information since the base information is missing in the model. All-atom structures can be reconstructed by the supporting software C2A (Jonikas et al., 2009a). The software can be downloaded from <https://simtk.org/projects/nast>.

iFoldRNA (Ding et al., 2008; Sharma et al., 2008; Krokhotin et al., 2015) The iFoldRNA model uses three beads to represent a nucleotide. Two beads are positioned at the center of mass of the phosphate group and the five-atom ring sugar. And the third bead is located at the center of the six-atom ring in the nucleobase. The energy functions include the bonded terms and the non-bonded terms. The bonded terms impose constraints on the bond lengths, angles, and torsions derived from experimental structures. For the non-bonded terms, base pairing interactions, phosphate-phosphate repulsion interactions, hydrophobic interactions (a general attraction between all bases) and nucleotide type-based base stacking interactions are considered in the energy function. The energy parameters for base pairing, stacking and hydrophobic interactions are computed by decomposing the sequence-dependent free energy parameters for individual nearest-neighbor hydrogen-bond model (INN-HB) (Mathews et al., 1999) and thus it can be used to study RNA thermodynamic properties. Moreover, loop entropy is explicitly considered in the model. Replica exchange discrete molecular dynamics (REDMD) simulations starting from a linear structure are utilized for conformational sampling. The model can *de novo* fold small RNAs without secondary structure information including pseudoknots and can fold large RNAs given secondary structure constraints. All-atom 3D structures can be generated. The iFoldRNA v2 server is available at <https://dokhlab.med.psu.edu/ifoldrna>.

TOPRNA (Mustoe et al., 2014) The TOPRNA model uses three beads to describe the phosphate, sugar, and base moieties of a nucleotide. The energy functions follow the form of the

standard CHARMM potential (Brooks et al., 2009), including three harmonic potentials for the bond lengths, angles and improper torsions, a cosine form potential for proper torsions, and a Lennard-Jones potential for the Van der Waals interactions. However, electrostatic interactions are ignored in the model. The parameters in the energy functions are fitted to the statistical potentials calculated according to a database of native structures. Replica exchange molecular dynamics (REMD) (Sugita and Okamoto, 1999) simulations are performed for conformational sampling. The link to the force field and the related software is as follows: <https://brooks.chem.lsa.umich.edu/index.php?page=toprna&subdir=articles/resources/software>.

IsRNA1 (Zhang and Chen, 2018; Zhang et al., 2021) For the IsRNA1 model, four or five beads are used to represent a nucleotide, including atoms P and C4' for the backbone, two beads for the pyrimidine bases, and three beads for the purine bases. A sum of a harmonic potential and a Gaussian potential is used for the bond length, and angle energy. A cosine form potential is used for the torsional angle energy. A combined function of two bond lengths, two angles, and two torsions is used to characterize the canonical base pairing interactions. A sum of two Morse potentials and two Gaussian potentials is used to account for non-canonical base pairing, base-stacking, base-backbone, and backbone-backbone interactions. Moreover, a repulsive Lennard-Jones potential is used to describe the excluded volume effects.

The parameters in the CG force field are optimized by using the statistical potential in Eq. 2. It is noteworthy that an iterative simulated reference state approach is implemented to parameterize the energy functions, which can account for the correlations between different structural variables in the energy functions. Specifically, energy terms are added to the force field sequentially and the probability distribution in the reference state is updated accordingly. In each step, an energy term $E(x)$ is computed from Eq. 2 and added to the existing CG force field U . A key point in the theory is that $p_{\text{ref}}(x)$ for x in Eq. 2 is computed from MD simulations with force field U . Such a reference state accounts for the chain connectivity, the excluded volume effects, and all the other interactions considered in U . Physically, $E(x)$ represents the interaction potential for x in excess of U . With the updated energy function $U + E(x)$, the simulated distribution $p_{\text{sim}}(x)$ of x would be equal to the observed result $p_{\text{obs}}(x)$ (Hurst et al., 2021).

In the IsRNA1 approach, REMD simulations are utilized to accelerate conformational sampling. Moreover, the model uses template-based Vfold3D (Zhao et al., 2017) and VfoldLA (Xu et al., 2019) programs to generate the initial structures. Users can also provide the initial structures generated by other models such as RNAComposer (Popenda et al., 2012) and RNA-MoIP (Yao et al., 2017; Reinhartz et al., 2021). Secondary structure is an input for IsRNA1 and energy-minimized all-atom 3D structures can be obtained. The standalone software and the server of the IsRNA1 model for RNA 3D structure prediction can be found at <http://rna.physics.missouri.edu/IsRNA/index.html>. The input to the server includes RNA sequence, secondary structure and optional initial structures, and the output is the energy-minimized all-atom 3D structures.

Ren's model (Xia et al., 2010, 2013) In this model, a nucleotide is reduced to five beads. Specifically, atoms P and C4' are used to represent the backbone, and three bonded pseudo atoms are used to represent the nucleobase. Three harmonic potentials are implemented to characterize the bond lengths, angles, and torsions. And a Buckingham potential is used to describe the non-bonded interactions. The Debye-Huckel potential is utilized for electrostatic interactions. And a distance- and angle-based energy function is used to describe the hydrogen bonding interaction. The parameters in the potential energies are fitted to the statistical potentials derived from a database of native structures. SA simulations are performed to fold RNAs. Small RNAs of different topologies can be folded into near native structures and large RNAs can be folded with experimental constraints. Atomic 3D structures are obtained in the output. The program of this model is not released.

oxRNA (Šulc et al., 2014) In the oxRNA model, a nucleotide is treated as a rigid body with five interaction sites including backbone, hydrogen-bonding, cross-stacking and 3'- and 5'-stacking interactions. The energy functions adopt relatively complex forms and the energy parameters are fitted to reproduce the thermodynamics properties of the nearest-neighbor model (Tinoco et al., 1973) for short duplex and hairpin structures. The virtual move Monte Carlo (VCMC) algorithm combined with umbrella sampling (US) method or MD simulations are used for conformational sampling. It not only can fold small RNAs including pseudoknot and kissing hairpin structures, but also can characterize the thermodynamic and mechanical properties of RNAs. The model can be downloaded from https://dna.physics.ox.ac.uk/index.php/Main_Page. The web server (Poppleton et al., 2021) was recently released at <https://oxdna.org/>.

SimRNA (Boniecki et al., 2016; Magnus et al., 2016) The SimRNA model reduces 20–23 heavy atoms down to five, namely atoms P, C4', C2, N1 and C4 for pyrimidines, N9 and C6 for purines. The energy function in SimRNA model consists of sequence-independent local bonded terms and sequence-dependent non-bonded terms. The non-bonded terms include base-base, base-backbone, and backbone-backbone interactions, and discrete statistical potentials are constructed for these interactions based on a database of native structures. Replica exchange Monte Carlo (REMC) simulations are used to sample the conformational space. The model can fold small RNAs without secondary structure constraint and can fold large RNAs with secondary structure and other interatomic distance constraints. The standalone software can be found at <http://genesilico.pl/software/stand-alone/simrna>. The server of the model can be accessed from <https://genesilico.pl/SimRNAweb>. Users feed RNA sequence and 2D structure (optional) to the server, and can fetch the predicted all-atom 3D structures after the jobs are done.

SPQR (Poblete et al., 2018) The SPQR model uses a bead as the phosphate group, three beads for the nucleobase forming a rigid triplet, and a virtual bead for the sugar attached to the base. The non-bonded energy functions account for excluded volume effects, base pairing, base stacking, base-phosphate, and backbone-backbone interactions. The potentials of mean force

are applied to the non-bonded interactions and the relative strengths of different interactions are trained by distinguishing native structures from decoys for dozens of internal and hairpin loop structures. SA/MC protocol is adopted for sampling. All-atom structures are reconstructed finally. This model can provide accurate predictions for the secondary and tertiary structures of small RNA motifs. The program can be downloaded at <http://github.com/srnas/spqr>.

HiRE-RNA (Pasquali and Derreumaux, 2010; Cragnolini et al., 2013; Sterpone et al., 2014; Cragnolini et al., 2015) In the HiRE-RNA model, the coarse-grained representation of a nucleotide consists of six or seven beads: five beads for the backbone (atoms P, O5', C5', C4', C1'), one bead (for pyrimidine) or two beads (for purine) placed at the centers of mass of heavy atoms in the all-atom rings in the base. The energy functions include the local bonded terms and the long-range non-bonded terms. For the local energy terms, harmonic potentials are employed to constrain bond lengths and angles, and the cosine form potential is applied to torsions. For the non-local energy terms, excluded volume potential, electrostatic potential, base stacking potential, and base-pairing potential including non-canonical base pairs are considered. The key parameters in the non-bonded energies are the relative strengths between different energy terms. The parameters are trained using the genetic algorithm to distinguish native structures from decoys. REMD simulations are performed for conformational sampling. The model can *de novo* fold small RNAs of different topologies from fully extended structures without using secondary structure constraints, and can fold large RNAs with secondary structure and additional experimental constraints.

Besides 2D structure constraints, some of the models above can integrate additional constraints into energy functions to make conformational sampling more efficient and effective for structure prediction. The YUP model (Malhotra et al., 1994; Tan et al., 2006) can use footprinting data (Tijerina et al., 2007; Sclavi et al., 1998) and cross-linking (Harris and Christian, 2009) experimental information to refine structures. The iFoldRNA model (Ding et al., 2008; Sharma et al., 2008; Krokhotin et al., 2015) can utilize the hydroxyl radical probing (HRP) (Ding et al., 2012) and NMR experimental data. Ren's model (Xia et al., 2013) can transform NMR and small-angle X-ray scattering (SAXS) (Lipfert and Doniach, 2007) experimental data to constrain the sampling space. The NAST (Jonikas et al., 2009b) and SimRNA (Boniecki et al., 2016; Magnus et al., 2016) models can implement interatomic distance constraints when performing simulations.

The CG models reviewed above can be classified into two categories, namely the helix-based and the nucleotide-based models. The helix-based models can predict the global structures efficiently especially for large RNAs, but sequence information and structure details are missing in the sampling process and secondary structures are essential as input. The Graph-based model performs junction prediction first with RNAJAG program (Laing et al., 2013), which can reduce the conformational sampling space for RNAs with junction structures but may incur the risk of using a nonnative junction topology. The helix-based model Erwin implements new potentials favoring A-minor motif and loop-loop

interactions but the atomic 3D structures are not available at the present time.

The nucleotide-based models can capture structural details at the expense of computational cost. The low-resolution models YUP and NAST using one bead per nucleotide have the advantage over high resolution models in terms of computational speed but they cannot accurately model the different (non-)canonical base pairing interactions which are important in stabilizing RNA tertiary structures. Like the helix-based models, YUP and NAST models cannot fold an RNA solely based on sequence information and additional 2D and 3D restraints are required. The high resolution models using 3-7 beads per nucleotide introduce a force field consisting of the bonded and non-bonded energy terms and the differences between models lie in the energy components and energy function forms.

For the bonded energy terms, harmonic potentials or Gaussian potentials or combination of both are often used to restrain the bonds and angles, and a cosine form function is used for the torsions in the MD-based models TOPRNA, IsRNA1, Ren's model, and HiRE-RNA, while discrete statistical potentials are used for the bonds, angles, and torsions in the MC-based models SimRNA and SPQR.

For the non-bonded energy terms, excluded volume potential, electrostatic repulsive energy between phosphate-phosphate groups, and backbone-backbone, base-backbone, and base-base interactions are the main energy components. Although different models may consider different interactions, the base-base energies are included in all models because base pairing and base stacking interactions are the most important forces stabilizing RNA structures. The different models adopt different strategies to account for base-base energies. For the canonical base pairing interactions, all the models use explicit terms of various energy functions to favor the canonical base pairs, and in these energy terms, distances, angles and torsions are introduced to characterize the co-planarity and the relative positions and orientations of two paired bases. For the non-canonical base pairing and base stacking interactions, several models use explicit terms to describe the interaction potential energy. For example, the iFoldRNA model uses a general base attraction energy to mimic the non-canonical base pairing interactions and uses distance-based energy functions to favor base stacking interactions. The oxRNA model does not consider non-canonical base pairing interactions, but uses complex distance- and angle-dependent energy functions to describe the stacking interactions, and the energy parameters are fitted by reproducing the melting temperatures. The SPQR model uses the distance-, angle-, and torsion-based statistical potentials to describe the different non-canonical base pairing and base stacking interactions. The SimRNA model implements statistical potentials to extract the energy functions using six discrete distances as the variables. The HiRE-RNA model uses functions of distances and angles to achieve base pairing and stacking interactions and the energy parameters are determined by using the genetic algorithm to distinguish native structures from decoys. In contrast, other models use implicit energy terms to describe non-canonical base pairing and base stacking

interactions. For example, the IsRNA1 model and the Ren's model integrate the non-canonical base pairing, base stacking, base-backbone and backbone-backbone interactions into an effective potential as a function of the distances between non-bonded atom pairs such as the Van der Waals energy in the all-atom force field. The IsRNA1 model uses a sum of a Morse potential and two Gaussian potentials while Ren's model uses a Buckingham potential, and the corresponding energy parameters are fitted to the interatomic distance statistical potentials. The non-canonical base pairing interactions are of great importance but the current models either ignore them or describe them with low accuracy. The statistical potential-based energy for the non-canonical base pairs could underestimate the interaction strength of relatively rare base pairs in the native structures. The implicit and CG potentials for non-canonical base pairs have the difficulty in accurate description of the geometries and the interaction strength.

It is important to note that other excellent CG RNA modeling methods have been reported elsewhere but cannot be covered with great detail in this short mini-review. For example, NARES-2P (He et al., 2013), a physics-based 2-bead model, can fold DNA and RNA double helical structures, featuring the mean-field dipole-dipole interactions between bases. Combined with global optimization conformational space annealing algorithm and limited distance restraints, NARES-2P model can also deal with complex RNA folds (Sieradzan et al., 2018). RECER, a 5-bead CG model, can accurately predict the native structures and capture the folding free energy of RNA hairpins and duplexes (Bell et al., 2017). Fyta et al. has developed a four-bead CG model for double-stranded RNA which uses density functional theory to determine the parameters in energy functions and can reproduce the structural and mechanical properties (Cruz-León et al., 2018). Tan et al. managed to predict the 3D structures and stability of RNA kissing complexes in monovalent/divalent ion solutions by using a 3-bead CG model (Jin et al., 2019). Even for the large ribosome system, a 1-bead CG model for RNAs and proteins has been applied to study its dynamics (Trylska et al., 2005). Recently a three-bead CG model has been used to study the folding of the ribosomal subunit where the Mg^{2+} effect can be explicitly considered (Hori et al., 2021).

2.2 Performance Comparison Between the iFoldRNA, SimRNA and IsRNA1 Models

Three CG methods, namely iFoldRNA, SimRNA, and IsRNA1 models, are compared in terms of performance for RNA 3D structure prediction in the reference paper (Zhang et al., 2021). The other methods are not chosen here because the programs are not open to the public, or generate only CG structures, or are designed mainly for studying the folding thermodynamics of small RNAs instead of structure prediction of large RNAs. A large dataset of 130 RNAs are used for benchmark test. The dataset includes 44 stem-loops, 43 multi-way junctions, and 43 structures of long-range tertiary interactions, and the shortest and longest RNAs contain 40 and 161 nucleotides, respectively. RNA sequences and native secondary structures are used as inputs to the three models. The IsRNA1 model uses template-based

Vfold3D (Cao and Chen, 2011; Xu and Chen, 2015; Zhao et al., 2017) and loop-based VfoldLA (Xu and Chen, 2017; Xu et al., 2019) algorithms to generate initial 3D structures. As an objective test, the native 3D structures in the benchmark dataset and their homologous entries are excluded from the template/loop database when generating initial structures. For the 130 RNAs, the average/median RMSDs for the predicted top-ranked structure by IsRNA1, SimRNA, and iFoldRNA models are 9.51/8.12, 11.26/10.95, and 11.87/11.37 Å, respectively. Considering the top-three predicted structures, the average RMSDs for the best model predicted by IsRNA1, SimRNA, and iFoldRNA models are 8.34, 9.73, and 10.88 Å, respectively. For all the three models, however, the average RMSDs for multi-way junction structures are always larger than those for structures of other topologies, which indicates that the relative positions of the helices in the junction structures are more difficult to predict. For RNAs larger than 100 nucleotides, the RMSDs for the three models are almost larger than 20 Å and the average clash score (the number of serious all-atom steric overlaps per 1,000 atoms) (Word et al., 1999) for the IsRNA1 model (4.0) is lower than that for the SimRNA model (139.7) and the iFoldRNA model (170.4) as energy minimization is performed for the predicted structures in IsRNA1 model. Detailed results can be found in the reference paper (Zhang et al., 2021).

The runtime for 3D structure prediction is dependent on RNA size. For an RNA of about 100 nucleotides, the runtimes for IsRNA1 (50 million steps), SimRNA (16 million steps), and iFoldRNA (2 million steps) models are about 1 day, 1 day, and 15 h, respectively. For an RNA of about 150 nucleotides, the runtimes increase to about 2 days, 3.5 days, and 20 h. All the three models provide server services to users, and users can adjust and test the simulation time or step based on the previous runtime estimation.

The performance improvement for the IsRNA1 model may result from the following aspects: 1) a set of more detailed energy functions are used for the bonds, angles, and torsions instead of the simple harmonic potentials in the previous models, enabling broader and more accurate conformational coverage for RNA backbone; 2) the iterative simulated reference state approach takes into consideration the correlation between different energy terms to avoid over-constraint; 3) the iteratively constructed energy functions learn knowledge not only from the native structures but also from the non-native ones in the simulated reference state; 4) using the structures generated by the template-based methods Vfold3D and VfoldLA as the initial structures may help folding from an initial state not too far away from the native structures. Meanwhile, the relatively poor performance for the junction structures of the CG models including the IsRNA1 model necessitates a more detailed energy function, especially for non-canonical base pairing and stacking interactions, for this structural topology.

3 FUTURE DEVELOPMENTS IN MODELING RIBONUCLEIC ACID 3D STRUCTURES

Compared with all-atom models, CG models have a smoother energy landscape and far fewer atoms (pseudo-atoms), and thus

the conformational space can be explored more efficiently. To further speed up the computation, GPU-driven CG models could be a promising option justified by the GPU version of all-atom MD simulation software AMBER. General MD simulation engines LAMMPS (Plimpton, 1995) and OpenMM (Eastman et al., 2017) have already supported GPU computation, enabling new CG models on these platforms.

Even with native secondary structures, the accuracy of structure prediction for large RNAs (>100nt) is low. Large RNAs often contain multi-way junctions, and an incorrect arrangement of the helices in the junction structures could result in large deviation from native structures. The problem can be caused by inaccurate energy functions or insufficient sampling or both. Insufficient sampling problem could be alleviated by extending simulation time or adopting advanced accelerating algorithms. However, inaccurate energy problem is more serious because energy functions are used not only to assess structures but also to bias the sampling direction in the conformational space. Inaccuracy in energy functions may come from the inaccurate characterization of both the intra-RNA interactions and RNA-environment interactions. The RNA-environment interactions, including RNA-solvent, RNA-ion, RNA-ligand and RNA-protein interactions, are missing or not accurate in the CG models. Instead of *de novo* folding, the combination of CG models and homology modeling of whole structures or motifs has the potential to improve the performance as in the IsRNA1 model.

Solving structures by traditional X-ray crystallography and NMR spectroscopy experiments is laborious and technically challenging. However, low-resolution information (overall shape, tertiary contact, local properties, etc.) may be readily available from experiments such as SAXS (Small-angle X-ray scattering) (Lipfert and Doniach, 2007), cryo-EM (Electron microscopy) (Bai et al., 2015), FRET (Förster resonance energy transfer) (Jares-Erijman and Jovin, 2003; Stephenson et al., 2016), chemical cross-linking (Harris and Christian, 2009), footprinting (Tijerina et al., 2007; Sclavi et al., 1998). CG models combined with the low-resolution information can significantly improve the structure prediction results (Ponce-Salvatierra et al., 2019; Li et al., 2020). Efficient and accurate integration of experimental data and CG models is a highly promising approach for the prediction of large RNA structures.

AUTHOR CONTRIBUTIONS

JL performed literature search. All authors wrote the manuscript, read, and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health under Grant R35-GM134919 to S-JC.

REFERENCES

- Amanat, F., and Krammer, F. (2020). SARS-CoV-2 Vaccines: Status Report. *Immunity* 52, 583–589. doi:10.1016/j.immuni.2020.03.007
- Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science* 181, 223–230. doi:10.1126/science.181.4096.223
- Bai, X.-C., McMullan, G., and Scheres, S. H. W. (2015). How Cryo-EM Is Revolutionizing Structural Biology. *Trends Biochem. Sci.* 40, 49–57. doi:10.1016/j.tibs.2014.10.005
- Bell, D. R., Cheng, S. Y., Salazar, H., and Ren, P. (2017). Capturing RNA Folding Free Energy with Coarse-Grained Molecular Dynamics Simulations. *Sci. Rep.* 7, 1–14. doi:10.1038/srep45812
- Boniecki, M. J., Lach, G., Dawson, W. K., Tomala, K., Lukasz, P., Soltysinski, T., et al. (2016). SimRNA: a Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction. *Nucleic Acids Res.* 44, e63. doi:10.1093/nar/gkv1479
- Brooks, B. R., Brooks, C. L., III, Mackerell, A. D., Jr, Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: the Biomolecular Simulation Program. *J. Comput. Chem.* 30, 1545–1614. doi:10.1002/jcc.21287
- Cao, S., and Chen, S.-J. (2011). Physics-based De Novo Prediction of RNA 3D Structures. *J. Phys. Chem. B* 115, 4216–4226. doi:10.1021/jp112059y
- Capriotti, E., Norambuena, T., Marti-Renom, M. A., and Melo, F. (2011). All-atom Knowledge-Based Potential for RNA Structure Prediction and Assessment. *Bioinformatics* 27, 1086–1093. doi:10.1093/bioinformatics/btr093
- Cech, T. R., and Steitz, J. A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* 157, 77–94. doi:10.1016/j.cell.2014.03.008
- Cragolini, T., Derreumaux, P., and Pasquali, S. (2013). Coarse-grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* 117, 8047–8060. doi:10.1021/jp400786b
- Cragolini, T., Laurin, Y., Derreumaux, P., and Pasquali, S. (2015). Coarse-grained HiRE-RNA Model for Ab Initio RNA Folding beyond Simple Molecules, Including Noncanonical and Multiple Base Pairings. *J. Chem. Theor. Comput.* 11, 3510–3522. doi:10.1021/acs.jctc.5b00200
- Cruz-León, S., Vázquez-Mayagoitia, A., Melchionna, S., Schwierz, N., and Fyta, M. (2018). Coarse-grained Double-Stranded RNA Model from Quantum-Mechanical Calculations. *J. Phys. Chem. B* 122, 7915–7928. doi:10.1021/acs.jpcc.8b03566
- Ding, F., Sharma, S., Chalasani, P., Demidov, V. V., Broude, N. E., and Dokholyan, N. V. (2008). Ab Initio RNA Folding by Discrete Molecular Dynamics: from Structure Prediction to Folding Mechanisms. *RNA* 14, 1164–1173. doi:10.1261/rna.894608
- Ding, F., Lavender, C. A., Weeks, K. M., and Dokholyan, N. V. (2012). Three-dimensional RNA Structure Refinement by Hydroxyl Radical Probing. *Nat. Methods* 9, 603–608. doi:10.1038/nmeth.1976
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *Plos Comput. Biol.* 13, e1005659. doi:10.1371/journal.pcbi.1005659
- Gesteland, R. F., Cech, T. R., and Atkins, J. F. (2005). *The RNA World*. Third Edition. New York: Cold Spring Harbor Laboratory Press.
- Harris, M. E., and Christian, E. L. (2009). RNA Crosslinking Methods. *Methods Enzymol.* 468, 127–146. doi:10.1016/S0076-6879(09)68007-1
- He, Y., Maciejczyk, M., Oldziej, S., Scheraga, H. A., and Liwo, A. (2013). Mean-field Interactions between Nucleic-Acid-Base Dipoles Can Drive the Formation of a Double helix. *Phys. Rev. Lett.* 110, 098101. doi:10.1103/PhysRevLett.110.098101
- Hori, N., Denesyuk, N. A., and Thirumalai, D. (2021). Shape Changes and Cooperativity in the Folding of the central Domain of the 16S Ribosomal RNA. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2020837118. doi:10.1073/pnas.2020837118
- Hurst, T., Zhang, D., Zhou, Y., and Chen, S.-J. (2021). A Bayes-Inspired Theory for Optimally Building an Efficient Coarse-Grained Folding Force Field. *Commun. Assoc. Inf. Syst.* 21, 65–83. doi:10.4310/CIS.2021.v21.n1.a4
- Jares-Erijman, E. A., and Jovin, T. M. (2003). FRET Imaging. *Nat. Biotechnol.* 21, 1387–1395. doi:10.1038/nbt896
- Jin, L., Tan, Y.-L., Wu, Y., Wang, X., Shi, Y.-Z., and Tan, Z.-J. (2019). Structure Folding of RNA Kissing Complexes in Salt Solutions: Predicting 3D Structure, Stability, and Folding Pathway. *RNA* 25, 1532–1548. doi:10.1261/rna.071662.119
- Jonikas, M. A., Radmer, R. J., and Altman, R. B. (2009a). Knowledge-based Instantiation of Full Atomic Detail into Coarse-Grain RNA 3D Structural Models. *Bioinformatics* 25, 3259–3266. doi:10.1093/bioinformatics/btp576
- Jonikas, M. A., Radmer, R. J., Laederach, A., Das, R., Pearlman, S., Herschlag, D., et al. (2009b). Coarse-grained Modeling of Large RNA Molecules with Knowledge-Based Potentials and Structural Filters. *RNA* 15, 189–199. doi:10.1261/rna.1270809
- Kelly, J. A., Olson, A. N., Neupane, K., Munshi, S., San Emeterio, J., Pollack, L., et al. (2020). Structural and Functional Conservation of the Programmed-1 Ribosomal Frameshift Signal of SARS Coronavirus 2 (SARS-CoV-2). *J. Biol. Chem.* 295, 10741–10748. doi:10.1074/jbc.AC120.013449
- Kerpedjiev, P., Zu Siederdisen, C. H., and Hofacker, I. L. (2015). Predicting RNA 3D Structure Using a Coarse-Grain helix-centered Model. *RNA* 21, 1110–1121. doi:10.1261/rna.047522.114
- Kim, N., Laing, C., Elmetwaly, S., Jung, S., Curuku, J., and Schlick, T. (2014). Graph-based Sampling for Approximating Global Helical Topologies of RNA. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4079–4084. doi:10.1073/pnas.1318893111
- Krammer, F. (2020). SARS-CoV-2 Vaccines in Development. *Nature* 586, 516–527. doi:10.1038/s41586-020-2798-3
- Krokhotin, A., Houlihan, K., and Dokholyan, N. V. (2015). iFoldRNA V2: Folding RNA with Constraints. *Bioinformatics* 31, 2891–2893. doi:10.1093/bioinformatics/btv221
- Laing, C., Jung, S., Kim, N., Elmetwaly, S., Zahran, M., and Schlick, T. (2013). Predicting Helical Topologies in RNA Junctions as Tree Graphs. *PLoS One* 8, e71947. doi:10.1371/journal.pone.0071947
- Li, B., Cao, Y., Westhof, E., and Miao, Z. (2020). Advances in RNA 3D Structure Modeling Using Experimental Data. *Front. Genet.* 11, 1147. doi:10.3389/fgene.2020.574485
- Lipfert, J., and Doniach, S. (2007). Small-angle X-ray Scattering from RNA, Proteins, and Protein Complexes. *Annu. Rev. Biophys. Biomol. Struct.* 36, 307–327. doi:10.1146/annurev.biophys.36.040306.132655
- Magnus, M., Boniecki, M. J., Dawson, W., and Bujnicki, J. M. (2016). SimRNAweb: a Web Server for RNA 3D Structure Modeling with Optional Restraints. *Nucleic Acids Res.* 44, W315–W319. doi:10.1093/nar/gkw279
- Malhotra, A., Tan, R., and Harvey, S. C. (1994). Modeling Large RNAs and Ribonucleoprotein Particles Using Molecular Mechanics Techniques. *Biophys. J.* 66, 1777–1795. doi:10.1016/S0006-3495(94)80972-5
- Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* 288, 911–940. doi:10.1006/jmbi.1999.2700
- Miao, Z., and Westhof, E. (2017). RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annu. Rev. Biophys.* 46, 483–503. doi:10.1146/annurev-biophys-070816-034125
- Moore, J. P., and Offit, P. A. (2021). SARS-CoV-2 Vaccines and the Growing Threat of Viral Variants. *JAMA* 325, 821–822. doi:10.1001/jama.2021.1114
- Morris, K. V., and Mattick, J. S. (2014). The Rise of Regulatory RNA. *Nat. Rev. Genet.* 15, 423–437. doi:10.1038/nrg3722
- Mortimer, S. A., Kidwell, M. A., and Doudna, J. A. (2014). Insights into RNA Structure and Function from Genome-wide Studies. *Nat. Rev. Genet.* 15, 469–479. doi:10.1038/nrg3681
- Mustoe, A. M., Al-Hashimi, H. M., and Brooks, C. L., III (2014). Coarse Grained Models Reveal Essential Contributions of Topological Constraints to the Conformational Free Energy of RNA Bulges. *J. Phys. Chem. B* 118, 2615–2627. doi:10.1021/jp411478x
- Oran, D. P., and Topol, E. J. (2020). Prevalence of Asymptomatic SARS-CoV-2 Infection: a Narrative Review. *Ann. Intern. Med.* 173, 362–367. doi:10.7326/M20-3012
- Pasquali, S., and Derreumaux, P. (2010). HiRE-RNA: a High Resolution Coarse-Grained Energy Model for RNA. *J. Phys. Chem. B* 114, 11957–11966. doi:10.1021/jp102497y
- Plimpton, S. (1995). Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* 117, 1–19. doi:10.1006/jcph.1995.1039
- Poblete, S., Bottaro, S., and Bussi, G. (2018). A Nucleobase-Centered Coarse-Grained Representation for Structure Prediction of RNA Motifs. *Nucleic Acids Res.* 46, 1674–1683. doi:10.1093/nar/gkx1269

- Ponce-Salvatierra, A., Merdas, K., Nithin, C., Ghosh, P., Mukherjee, S., Bujnicki, J. M., et al. (2019). Computational Modeling of RNA 3D Structure Based on Experimental Data. *Biosci. Rep.* 39. doi:10.1042/BSR20180430
- Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., et al. (2012). Automated 3D Structure Composition for Large RNAs. *Nucleic Acids Res.* 40, e112. doi:10.1093/nar/gks339
- Poppleton, E., Romero, R., Mallya, A., Rovigatti, L., and Šulc, P. (2021). OxDNA.Org: a Public Webserver for Coarse-Grained Simulations of DNA and RNA Nanostructures. *Nucleic Acids Res.* gkab324 doi:10.1093/nar/gkab324
- Reinharz, V., Sarrazin-Gendron, R., and Waldspühl, J. (2021). Modeling and Predicting RNA Three-Dimensional Structures. *Methods Mol. Biol.* 2284, 17–42. doi:10.1007/978-1-0716-1307-8_2
- Robert, C., and Casella, G. (2013). *Monte Carlo Statistical Methods*. New York, NY: Springer Science & Business Media.
- Slavi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S. A. (1998). RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting. *Science* 279, 1940–1943. doi:10.1126/science.279.5358.1940
- Sharma, S., Ding, F., and Dokholyan, N. V. (2008). iFoldRNA: Three-Dimensional RNA Structure Prediction and Folding. *Bioinformatics* 24, 1951–1952. doi:10.1093/bioinformatics/btn328
- Sieradzan, A. K., Golon, L., and Liwo, A. (2018). Prediction of DNA and RNA Structure with the NARES-2P Force Field and Conformational Space Annealing. *Phys. Chem. Chem. Phys.* 20, 19656–19663. doi:10.1039/C8CP03018A
- Sippl, M. J. (1990). Calculation of Conformational Ensembles from Potentials of Mean Force: an Approach to the Knowledge-Based Prediction of Local Structures in Globular Proteins. *J. Mol. Biol.* 213, 859–883. doi:10.1016/S0022-2836(05)80269-4
- Stephenson, J. D., Kenyon, J. C., Symmons, M. F., and Lever, A. M. (2016). Characterizing 3D RNA Structure by Single Molecule FRET. *Methods* 103, 57–67. doi:10.1016/j.jymeth.2016.02.004
- Sterpone, F., Melchionna, S., Tuffery, P., Pasquali, S., Mousseau, N., Cragolini, T., et al. (2014). The OPEP Protein Model: from Single Molecules, Amyloid Formation, Crowding and Hydrodynamics to DNA/RNA Systems. *Chem. Soc. Rev.* 43, 4871–4893. doi:10.1039/C4CS00048J
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* 314, 141–151. doi:10.1016/S0009-2614(99)01123-9
- Šulc, P., Romano, F., Ouldridge, T. E., Doye, J. P., and Louis, A. A. (2014). A Nucleotide-Level Coarse-Grained Model of RNA. *J. Chem. Phys.* 140, 235102. doi:10.1063/1.4881424
- Tan, R. K., Petrov, A. S., and Harvey, S. C. (2006). YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models. *J. Chem. Theor. Comput.* 2, 529–540. doi:10.1021/ct050323r
- Tan, Y.-L., Feng, C.-J., Jin, L., Shi, Y.-Z., Zhang, W., and Tan, Z.-J. (2019). What Is the Best Reference State for Building Statistical Potentials in RNA 3D Structure Evaluation? *RNA* 25, 793–812. doi:10.1261/rna.069872.118
- Tijerina, P., Mohr, S., and Russell, R. (2007). DMS Footprinting of Structured RNAs and RNA–Protein Complexes. *Nat. Protoc.* 2, 2608–2623. doi:10.1038/nprot.2007.380
- Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., et al. (1973). Improved Estimation of Secondary Structure in Ribonucleic Acids. *Nat. New Biol.* 246, 40–41. doi:10.1038/newbio246040a0
- Trylska, J., Tozzini, V., and McCammon, J. A. (2005). Exploring Global Motions and Correlations in the Ribosome. *Biophys. J.* 89, 1455–1463. doi:10.1529/biophysj.104.058495
- Tu, Y.-F., Chien, C.-S., Yarmishyn, A. A., Lin, Y.-Y., Luo, Y.-H., Lin, Y.-T., et al. (2020). A Review of SARS-CoV-2 and the Ongoing Clinical Trials. *Int. J. Mol. Sci.* 21, 2657. doi:10.3390/ijms21072657
- Van Laarhoven, P. J., and Aarts, E. H. (1987). Simulated Annealing. In *Simulated Annealing: Theory and Applications* (Dordrecht: Springer), 7–15. doi:10.1007/978-94-015-7744-1_2
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and Testing of a General Amber Force Field. *J. Comput. Chem.* 25, 1157–1174. doi:10.1002/jcc.20035
- Wang, J., Zhao, Y., Zhu, C., and Xiao, Y. (2015). 3dRNAcore: a Distance and Torsion Angle Dependent Evaluation Function of 3D RNA Structures. *Nucleic Acids Res.* 43, e63. doi:10.1093/nar/gkv141
- Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., et al. (1999). Visualizing and Quantifying Molecular Goodness-Of-Fit: Small-Probe Contact Dots with Explicit Hydrogen Atoms. *J. Mol. Biol.* 285, 1711–1733. doi:10.1006/jmbi.1998.2400
- Xia, Z., Gardner, D. P., Gutell, R. R., and Ren, P. (2010). Coarse-grained Model for Simulation of RNA Three-Dimensional Structures. *J. Phys. Chem. B* 114, 13497–13506. doi:10.1021/jp104926t
- Xia, Z., Bell, D. R., Shi, Y., and Ren, P. (2013). RNA 3D Structure Prediction by Using a Coarse-Grained Model and Experimental Data. *J. Phys. Chem. B* 117, 3135–3144. doi:10.1021/jp400751w
- Xu, X., and Chen, S.-J. (2015). “A Method to Predict the 3D Structure of an RNA Scaffold,” In *RNA Scaffolds*. New York, NY: Springer. 1–11. doi:10.1007/978-1-4939-2730-2_1
- Xu, X., and Chen, S.-J. (2017). Hierarchical Assembly of RNA Three-Dimensional Structures Based on Loop Templates. *J. Phys. Chem. B* 122, 5327–5335. doi:10.1021/acs.jpcc.7b10102
- Xu, X., Zhao, C., and Chen, S.-J. (2019). VfoldLA: a Web Server for Loop Assembly-Based Prediction of Putative 3D RNA Structures. *J. Struct. Biol.* 207, 235–240. doi:10.1016/j.jsb.2019.06.002
- Yao, J., Reinharz, V., Major, F., and Waldspühl, J. (2017). RNA-MoIP: Prediction of RNA Secondary Structure and Local 3D Motifs from Sequence Data. *Nucleic Acids Res.* 45, W440–W444. doi:10.1093/nar/gkx429
- Zhang, D., and Chen, S.-J. (2018). ISRNA: an Iterative Simulated Reference State Approach to Modeling Correlated Interactions in RNA Folding. *J. Chem. Theor. Comput.* 14, 2230–2239. doi:10.1021/acs.jctc.7b01228
- Zhang, D., Li, J., and Chen, S.-J. (2021). ISRNA1: De Novo Prediction and Blind Screening of RNA 3D Structures. *J. Chem. Theor. Comput.* 17, 1842–1857. doi:10.1021/acs.jctc.0c01148
- Zhao, C., Xu, X., and Chen, S.-J. (2017). Predicting RNA Structure with Vfold. In *Functional Genomics* (New York, NY: Springer), 3–15. doi:10.1007/978-1-4939-7231-9_1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.