Published in final edited form as:

Annu Rev Biophys. 2017 May 22; 46: 227–246. doi:10.1146/annurev-biophys-070816-033920.

Theory and Modeling of RNA Structure and Interactions with Metal Ions and Small Molecules

Li-Zhen Sun*, Dong Zhang*, and Shi-Jie Chen

Department of Physics, Department of Biochemistry, and MU Informatics Institute, University of Missouri, Columbia, MO 65211

Abstract

In addition to continuous rapid progress in RNA structure determination, probing, and biophysical studies, the past decade has seen remarkable advances in the development of a new generation of RNA folding theories and models. Here, we review RNA structure prediction models and models for ion-RNA and ligand-RNA interactions. These new models are becoming increasingly important for mechanistic understanding of RNA function and quantitative design of RNA nanotechnology. We focus on new methods for physics-based, knowledge-based, and experimental data-directed modeling for RNA structures, and explore the new theories for the predictions of metal ion and ligand binding sites and metal ion-dependent RNA stabilities. The integration of these new methods with theories for the cellular environment effects, such as molecular crowding and cotranscriptional folding, may ultimately lead to an all-encompassing RNA folding model.

Keywords

Three-dimensional structure prediction; knowledge-based approach; physics-based modeling; specific ion-binding; nonspecific ion-binding

INTRODUCTION

RNA functions, such as the replication of RNA viruses and RNA-mediated regulation of gene expression, depend on RNA structures and the ability to adopt a variety of conformational states. It is therefore essential to be able to quantitatively model RNA structures from the sequence and to predict folding stability in proper solution conditions. Moreover, because RNA is highly charged, metal ions are critical for RNA folding and ions in the solution are probably the most important solution condition for RNA folding. Therefore, a reliable RNA model must consider the metal ion effects. In this review we mainly focus on RNA structure prediction and the modeling of ion effects in RNA folding.

The past decade has seen remarkable progress in RNA structure prediction, especially for small size RNAs (33, 34, 41, 79, 130). However, the prediction accuracy for medium to large

*Co-first authors

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

(longer than 100 nucleotides) structures has been inconsistent for different RNA structures and sequences, even with the input from experimental data (101). The major problem comes primarily from structures with multibranched loops, base-base and base-backbone noncanonical interactions, and other long-range tertiary contacts. Furthermore, the problem of accurate prediction for ion and ligand binding sites and their binding affinities still remains unsolved, and the accuracy of theoretical predictions is usually acceptable only at low resolution level. The challenges mentioned above stem partly from the limited amount of experimental structural and biophysical data and partly from lack of physics-based free energy models for RNA tertiary folds. This review places emphasis on both the knowledge-based and physics-based approaches for the prediction of structures and modeling of ion and ligand binding.

TWO-DIMENSIONAL (2D) STRUCTURE PREDICTION

The fact that base pairing and stacking are dominant stabilizing forces in RNA structure supports the validity of a two-step structure prediction algorithm: first predict the two-dimensional (2D) structure of an RNA, then predict the three-dimensional (3D) structure based on the 2D structure. A 2D structure, like a contact map, describes the structure in terms of base pairs. A 2D structure can describe secondary as well as tertiary structural motifs, such as kissing loops and pseudoknots, at the base pairing level. Such a thermodynamics-based 2D to 3D approach, to some extent, bears resemblance the kinetic hierarchical pathway in RNA folding (15, 59).

Multiple sequence alignment (60, 108) can provide reliable predictions for the base pairs. However, the method is restricted by the availability of the homologous sequences, and the method can only provide conserved base pairs. For a given sequence, we often need a separate model to predict the remaining (non-conserved) base pairs. Helix assembly-based methods (9, 119) allow us to sample large, complex pseudoknotted 2D structures. However, without considering loop free energies, the prediction accuracies are not consistent, especially for large RNAs. Experimental data-directed free energy minimization approaches, such as chemical mapping, are highly promising (32, 42, 62, 73, 120). However, to interpret and understand the ambiguity of the experimental data for structured loops and populated distribution of multiple conformers requires a rigorous physical model.

A central issue in the physics-based 2D structure prediction is the free energy model, including the entropy model, for tertiary folds. Although folding entropy often receives less appreciation than energy in the field, it plays a significant role in the overall folding stability. For example, a rough estimate indicates that at room temperature, the loop entropy term $T S_{\text{loop}}$ for the T4-35 pseudoknot contributes a significant portion (~ 7.5 kcal/mol) to the total folding free energy G (~ 9.0 kcal/mol) (19).

One of the most significant distinctions between the free energy calculations for secondary and tertiary structures is the additivity issue. For secondary structures, one typically adds the free energies of the helices and loops, which can be calculated from the experimentally determined parameters (Turner Rules) (99), to estimate the total free energy. However, because the long-range cross-linkage causes coupling between the different loops and

helices, the free energy, especially the entropy, is nonadditive for tertiary folds. The nonadditivity underscores the importance to treat a tertiary fold as a whole in the entropy calculation. Due to the holistic nature of the tertiary structure entropy and the large number of tertiary motif types, it is unrealistic to rely on experiments to measure all the parameters. Instead, we need a theory to compute the parameters. By considering the nonadditivity effect, researchers have developed and applied a variety of free energy models for tertiary motifs, such as the following.

- Dirks and Pierce developed an empirical energy model (48) for multiloops/pseudoknots, where the parameters are optimized through training with both negative (non-pseudoknotted structures) and positive (pseudoknotted structures) controls.
- From the knowledge of end-to-end distance distribution for the two-length-scale (one for unpaired segments, one for paired segments) freely jointed chains, Aalbert *et al.* developed polymer theory-based entropy models for pseudoknot loops (1) and kissing hairpin loops (100), respectively.
- Chen *et al.* developed a statistical mechanics-based virtual bond lattice model (“Vfold”) to compute the entropy parameters through conformation enumeration for pseudoknotted structures, kissing loops, and RNA-RNA kissing complexes (19, 20, 23, 90).
- Based on a topological genus-based structure classification, Bon & Orland developed a free energy model for pseudoknots according to their genus (11).
- To predict pseudoknotted 2D structures using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) data, Mathews and Weeks *et al.* (62) proposed a heuristic free energy scoring function, where a jackknife approach is used to fit parameters from a training set with non-pseudoknotted and pseudoknot-containing RNAs.
- For unimolecular kissing hairpins (Fig. 1), Sperschneider *et al.* (131) developed a heuristic approximation: G (the total free energy) = sum of the helix free energies and a weighted sum of the number of the unpaired nucleotides in the different loop segments, where the weight coefficients are trained to fit the known structures.

These and other entropy/free energy parameters for tertiary folds have led to notable improvements in 2D structure prediction. However, the applicability of the parameters is still quite limited due to the following two reasons.

- *Lack of sequence-dependence.* For loops involving strong sequence-specific interactions, such as loop-helix triple base interactions, the loop free energy parameters require sequence-specific modifications (21).
- *Structure-specificity.* The above parameters may not be useful for other tertiary motifs. The non-physics-based parameters may not be transferable between the different structures and can thus cause inaccuracy in structure predictions (131).

We need an entropy/free energy model that is based on rigorous statistical mechanics. Developing such a model and parameter set demands a large-scale systematic computations for the various types of tertiary motifs/folds. The Vfold model above has the advantage of accounting for the chain connectivity and excluded volume effects while keeping the conformational realism. Computations with the Vfold model have demonstrated the ability to successfully predict the entropy parameters for a variety of tertiary motifs. Integrated with a proper sampling method, Vfold may serve as a promising framework for systematic calculations for various types of tertiary motifs. Furthermore, entropy is intrinsically tied to the conformational statistics and is thus a problem at the 3D conformational level (19, 20, 144). With advances in 3D structure modeling, we expect continuous improvements in the entropy model and hence the 2D structure prediction accuracy (93, 95, 99, 128).

THREE-DIMENSIONAL (3D) STRUCTURE PREDICTION

Two main ingredients for a 3D structure prediction method are conformational sampling and scoring. For sampling, one may generate conformations through molecular dynamics (MD) or Monte Carlo simulations, or through assembly of the fragments/templates extracted from the known structures. For scoring, one may use sequence similarity, knowledge-based statistical potential, or molecular mechanics-based force field to rank candidate structures. The combinations of the different sampling and scoring methods lead to different 3D structure prediction methods. In fact, sampling and scoring can be interwoven with each other and the boundary between the different methods is blurry. For example, template-based sampling can dramatically reduce the search space for physics-based methods and physics-based energy functions can provide effective guidance for structure assembly in template-based approaches. In Table 1, Fig. 2, and Table S1, we summarize a set of selected methods.

Template-based methods

Template database: fragments vs. motif-based templates—A knowledge-based structure prediction method relies on a reference template database (22, 53, 54, 69, 98, 117, 121, 147). The database ranges from the whole structure (54, 121) to various structure motifs (22, 53, 117), including helices, hairpins, internal/bulge loops, pseudoknot loops, and junctions, down to small fragments with different chain lengths such as nucleotide cyclic motifs, trinucleotide fragments, and single base pairs (38, 40, 111). Recent studies on RNA-RNA interaction networks uncovered specific sequence patterns in many tertiary structural modules, including A-minors, ribose zippers, tetraloop receptors, and K-turns (18, 86). These structural modules can serve as excellent building blocks (templates) for 3D structure prediction. We refer to Westhof's article in this volume for more details about structural modules and related computational tools.

The performance of the different fragment/template methods varies from structure to structure (36, 79) and a single universal ranking of the different methods may not be available. In general, when there are reliable templates in the database, many different methods can provide reasonably good predictions. For instance, for the tRNA structures, the Root Mean Square Deviation (RMSD) given by various approaches ranges from 3.6 Å to 6.1

Å(22, 117, 121, 147). On the other hand, as indicated by Problem 3 in the blind test RNA-Puzzles (36) (see Westhof's article in this volume), the models that use relatively large size templates, such as Vfold (22), which searches the templates based on the whole loop/junction, have the advantage of accurately capturing the nonlocal intraloop interactions. In contrast, methods that use small size templates (fragments), such as SWA developed by Das *et al.* (36), have the advantage to accurately capture the local features, especially the local non-canonical interactions. Therefore, at the cost of a higher failure rate to find a template with an ideal match, the motif template-based methods, such as Vfold (22) and RNAComposer (117), may predict a reliable structure for complex structures such as multibranched loops and pseudoknot-ted loops (if the appropriate template is available). However, such motif-based templates are not always available. In that case, fragment-based methods, such as FARNA/FARFAR (38, 40) are particularly useful for generating 3D structures.

Sampling techniques: beyond the boundary of the template database—For template-based methods, the bottleneck is the lack of proper templates. Several novel methods have been developed to partially circumvent this problem. In RNAComposer, Adamiak *et al.* (117) utilized the machine translation system, where the translation search engine operates on the dictionary that maps the RNA 2D structure to 3D structure elements. In FARNA/FARFAR, Das *et al.* (132) developed an enumerative and recursive stepwise ansatz which enables the systematic sampling of RNA loop conformations at the atomic level. The fragment-based methods can be limited by the discrete nature of the fragments and their non-probabilistic nature. To remove the sampling limitations in the fragment assembly methods, Hamelryck and co-workers (55) developed a probabilistic model (BARNACLE) that allows efficient sampling of RNA conformations in continuous space with the associated probabilities. The Vfold (22) method exploits the fact that a 2D structure can be dynamic, especially at the loop-helix interface. The model allows a variable loop/junction size by closing/opening loop-closing base pairs. The approach can significantly augment the template database and enhance the possibility of finding the appropriate template.

In addition to the above sampling techniques for the candidate structures, global structural constraints encoded in the 2D structure can also be used to enhance the sampling efficiency, because the global constraints can influence the local structural preferences (7, 105). For example, the helix orientations (coaxial stacking, parallel, antiparallel, or perpendicular arrangement) in two-way (6), three-way (87), or higher-order junctions may subject to the restriction imposed by the long-range kissing contacts. A recent hierarchical graph theoretic algorithm guided by a knowledge-based scoring function (70, 72) shows great promise for efficiently sampling of large structures by incorporating the constraints from the 2D structure.

Scoring functions: guiding sampling/assembly process and evaluating structure candidates—A reliable scoring function is crucial for guiding the conformational sampling and ranking candidate structures. Ideally, such a function should be

based on rigorous physics. The essential elements for a physically reliable force field and scoring function will be discussed in “Physics-based methods” section below.

Physics-based methods

Energy functions—In Table S1, we summarize the energy/scoring functions, including CG energy functions, for RNA structure prediction. Atomistic models, such as FARFAR (40) and RASP (24), use an AMBER/CHARMM-like force field for local interactions and various forms of potential energies for nonlocal interactions. For CG models, based on fewer degrees of freedoms, the force field can be both distance- and orientation-dependent and is often model-dependent. CG energy functions describe base stacking and base pairing interactions by incorporating simple geometric constraint (68), pairwise (140) and many-body distance-dependent (46, 125) potentials, and specific distance- and orientation-dependent functions (34, 44, 65), or more sophisticated multi-atom potentials in the form of multidimensional grids (12). CG models based on highly simplified low-resolution representation (e.g., less than three interaction sites per nucleotide) cannot provide sufficient details for noncanonical base pairing (85) and base-phosphate interactions (148). Furthermore, most CG models tend to use the Debye-Hückel potential with the (partially) neutralized backbone charges (44, 125) to model the salt effect. The Debye-Hückel model has the advantage of being computationally efficient. However, the validity of this model for multivalent ions, such as Mg^{2+} ions, requires further rigorous tests with experimental data or detailed all-atom simulations.

Parameterization of the energy/scoring functions—There are two types of methods to derive the energy functions (109): physics-based and knowledge-based approaches. Physics-based methods are based on statistical mechanics models and compute energy parameters from thermodynamic data or atomistic force fields (17, 65). Knowledge-based methods use the probability distribution $p_{sta}(x)$ for the different structure elements x (e.g. torsional angles, base stacks, loops) from the known structures, and extract the statistical potential energies from the inverse of Boltzmann: $U(x) = -k_B T \ln[p_{sta}(x)/p_{ref}(x)]$. Here $p_{ref}(x)$ is the probability distribution in the reference state. Compared to the physics-based method, the knowledge-based approach has the advantage of being mathematically simple. However, unlike the physics-based energy functions, the knowledge-based energies suffer from two disadvantages: (a) bias toward the native structure and (b) the challenge to select a physically valid reference state.

A physical reference state should correctly reflect the inherent chain connectivity and excluded volume effects, and depending on the potential to be extracted, may also need to account for the correlation between the different intrachain contacts such as canonical and noncanonical base pairs. These features, however, are not fully captured in the current RNA models that use simplified reference states, such as the non-interacting ideal gas state (140), state treated with quasi-chemical approximation (8, 12) or uniform distribution in the reaction coordinate space (24, 34, 40, 125, 140). In a CG RNA loop structure prediction model, Liu and Chen (90) employed an iterative method, which was originally developed for protein folding (137), to circumvent the reference state problem. The model leads to much improved loop structure predictions at the CG level (90). Essentially, the iterative method

exploits statistical potentials such that the structure database-derived parameters can best reproduce the thermal equilibrium distribution for the complete conformational ensemble, including nonnative conformations, for the sequences in the training database. The method has the advantage of accounting for the full energy landscape, including the effects from nonnative structures.

Success and challenges—Simulations based on force fields such as AMBER or CHARMM have enabled us to extract detailed structural information about the native and alternative structures and folding and unfolding dynamics for small RNA structures (14,27,30,145) such as duplexes, hairpins, and pseudoknots. Moreover, other atomic force fields have been used to predict noncanonical motifs (40), to assess 3D structures (24), and to identify native-like conformations (8). Similarly, CG energy (scoring) functions have also led to success in structure prediction for small RNA (12, 34, 46, 125, 140), folding thermodynamics predictions for secondary structures and pseudoknots (34, 44, 46), and 3D structure refinement with structural constraints from experimental data (12, 68, 140). Furthermore, by taking noncanonical base pairs into account, CG models, such as HiRE-RNA (34) and SimRNA(12), can fold pseudoknots directly from the sequence (12, 34) and provide insights into the folding pathways of pseudoknots (12, 46), including triple helix pseudoknots (34).

However, all-atom and CG structure prediction remains a challenge even for medium size (e.g., 50–80 nts) RNAs. It suffers from three main problems. First, the force field needs further improvement. Even for the long-time tested AMBER and CHARMM force fields, folding studies for small systems such as an 8-nt tetraloop-forming hairpin (30) suggest the need for reparametrization of the force field. We expect the release of Amber16 force field, especially with the more accurate force field for metal ions (26), will lead to further improvements in prediction accuracy. For CG models, the energy function often lacks sufficiently detailed description for noncanonical base pairs and base-phosphate interactions. A correlated, many-body CG potential function may provide a viable way to alleviate the limitation. Second, efficient sampling of RNA conformations, especially for large RNAs, demands new approaches. Low-resolution representation, such as NAST (68), or modified force fields, such as DMD (46), can enhance the sampling efficiency, albeit at the cost of losing details about essential features for noncanonical interactions. Third, accurate modeling of RNA structures should consider the effect of the cellular environment. The problem requires incorporation of more sophisticated solvent models that can account for the effects from metal ions and other cofactors, including the molecular crowding effects, into the energy function.

DATA-DIRECTED STRUCTURE PREDICTION

Structural probing experiments, such as SHAPE (42, 62, 120), DMS (32) (dimethyl sulfate) structure detection, MOHCA (39) (multiplexed hydroxyl radical cleavage analysis), RING-MaP (66) (RNA interaction groups by mutational profiling), can provide a variety of low-resolution structure information. Other experiments, such as SAXS (small angle X-ray scattering) and Cryo-EM (cryogenic electron microscopy), can provide global shape of the structure. Integrating the experiment-generated low-resolution structure data into

computational models can significantly reduce the conformational search space and remarkable improvements in the scope and accuracy of structure prediction. Indeed, some of the most actively pursued structure prediction methods in recent years are based on data-directed algorithms, as shown in the following for several examples.

- Chemical mapping data (such as SHAPE)-directed 2D structure prediction based on data-derived pseudo-energy terms (32, 42, 62, 120);
- HRP (hydroxyl radical probing)-aided 3D structure refinement (47), and ^1H NMR chemical shifts-guided structure determination for noncanonical interactions (133);
- Multidimensional chemical mapping-guided structure prediction with Rosetta (39, 73);
- Chemical probing data-assisted discrete molecular dynamics simulation for 3D structure refinement (66);
- SAXS data-directed selection of the most native-like structure from decoys generated by MC-Fold/MC-Sym (142), NAST (68) or other CG models (140).

Although the data-assisted algorithms have led to remarkable success in RNA structure prediction, such algorithms cannot always lead to accurate structure predictions (74, 94), unless the data is properly interpreted and implemented. As indicated by Weeks *et al.* (66), the long-range tertiary interactions can be cooperatively coupled, thus, the structure determination should contain all the essential data for the tertiary interaction network. However, over-constraint by experimental data may cause trapping in local minimal for conformational sampling. How to use the data in an optimal way for computational modeling remains a nontrivial problem. Furthermore, as shown by Das and co-workers (39), the correlation between experimental data (e.g., strength of hits) and certain reaction coordinate (e.g. the pairwise distance between residues), can sometimes be poor. The situation highlights the demand for a physical model to correctly interpret the data and to eliminate the background noise, data fluctuation, and redundant/incorrect experimental information (139).

In the aforementioned RNA structure prediction models, ion-mediated electrostatic interactions are often treated by simplified models such as the counterion Condensation (CC) model or the Debye-Hückel model. However, the role of ions, especially the Mg^{2+} ions, goes beyond the description of these mean-field ion electrostatic models. A more accurate description for the ion effects is needed. Ion-RNA interactions can be roughly classified into two types: site-specific and nonspecific binding. Because the two types of interactions show quite different features and thus require different methods to model, we discuss them separately in the following sections.

ION EFFECTS: SPECIFIC BINDING

A fully hydrated Mg^{2+} ion has six water molecules in the first hydration shell and twelve in the second shell (97, 112). Specific binding of an ion is often associated with ion dehydration at various levels. For example, in the crystal structures of adenine riboswitch

(PDB ID: 1Y26 (124) and 4TZX (146), shown in Figs. 3A and B), Mg^{2+} ions keep only one water molecule at certain binding sites and retain several water molecules in other binding sites. Theoretical calculations show that the dehydration of the first and second shells cost 303.9 Kcal/mol and 156.9 Kcal/mol, respectively (112). Using MD simulations, Chen and coworkers investigated the dehydration of monovalent ions (Na^+ , K^+ , and Cs^+) bound to specific sites in an RNA kissing loop (28). The simulation led to the conclusion that (a) water molecules around smaller ions are bound more tightly thus smaller ions are less likely to dehydrate, and (b) ions at sites with strong ion-RNA attraction are more likely to dehydrate.

The specific ion binding sites are sensitive to ion species. Moreover, a recent study for ion binding to adenine riboswitch (146) (see Fig. 3C) showed that common binding sites for the different ion species can induce a competition between the different ions in RNA binding. Although Mg^{2+} has smaller radius (0.72 Å), Sr^{2+} (1.18 Å) appears to be a stronger competitor in this case as it replaces Mg^{2+} in several binding sites. The result depends on the interplay between Coulomb interaction, volume exclusion, and dehydration state of the ions. Because dehydration can reduce the effective size of an ion (97), smaller ions, which are less likely to dehydrate, may have a larger hydrated radius than an dehydrated ion (75). The large hydrated radius can cause stronger volume exclusion and weaker ion-RNA attraction. Similar competition has also been observed for Cu^{2+} and Ca^{2+} ions in the A-form RNA helix (123), Ca^{2+} and Mg^{2+} in the *Thi*-Box Riboswitch (50), and Mn^{2+} and Mg^{2+} in tetraloop-receptor complex (67).

Some binding sites can recognize only specific types of ions. An example is the Mg^{2+} ion trapped in the kissing loop of the adenine riboswitch. The specific binding may result from the direct binding between the ion and specific atoms of RNA such as the phosphate oxygen (123) instead of water-mediated indirect interactions. To quantitatively estimate the competition between the different types of ions, an atomistic model is needed. Recently, Panteva and coworkers (110) improved the force fields for Mg^{2+} , Mn^{2+} , Zn^{2+} , and Cd^{2+} ions using a “12–6–4” potential (88) to account for the electrostatic interaction and volume exclusion. The theory gives promising theory-experiment agreement (deviation within 0.1 Kcal/mol) for the binding free energies for four types of ions tested with different nucleic acid systems.

The specific binding is also dependent on the background salt conditions. In Fig. 3B (PDB ID: 4TZX (146)) and Fig. 3D (PDB ID: 4NXX (91)), we show a comparison of the site-specific binding of Mg^{2+} ions at high ($[\text{Mg}^{2+}]=200\text{mM}$) and low ($[\text{Mg}^{2+}]=100\text{mM}$) salt concentrations. The figures show that the low and high $[\text{Mg}^{2+}]$ lead to four binding sites in common. Five binding sites at high $[\text{Mg}^{2+}]$ disappear at dilute $[\text{Mg}^{2+}]$ and, surprisingly, two other binding sites observed at dilute $[\text{Mg}^{2+}]$ are missing at high $[\text{Mg}^{2+}]$. This phenomenon may result from two competitive effects induced by the change of the bulk ion concentration (84). An increase in the bulk concentration would cause more ions accumulated around the RNA. As a result, each binding site has a higher probability to capture an ion. However, accumulated ions tend to displace other like charge ions from the binding sites through Coulomb repulsion and volume exclusion, resulting in a negative impact on the specific binding.

Currently, both MD simulation-based and knowledge-based algorithms have been developed for the prediction of ion binding sites. For example, Philips and coworkers recently developed a knowledge-based algorithm to predict ion binding sites (114). The model uses more than one hundred ion-bound RNA crystal structures to deduce a statistical potential for metal ions such as Mg^{2+} , Na^+ , and K^+ . In a different approach, using MD simulation, Denesyuk and Thirumalai (45) investigated the effects of specific binding of ions for ribozyme folding. The theory predicted the binding sites for Mg^{2+} ions, including the two locations of co-binding Mg^{2+} ions at the active site. Based on a similar simulational approach, Lee and coworkers (82) found that the two co-binding Mg^{2+} ions can assist catalysis in different steps of the reaction. MD simulations have also been applied to the prediction of specific ion binding in other RNAs, including the 5S rRNA Loop E Motif (3), Hepatitis Delta Virus (HDV) Ribozyme (76), and HIV-1 TAR RNA (49).

ION EFFECTS: NONSPECIFIC BINDING

Nonspecific Coulombic attraction between RNA and ions results in significant accumulation of ions around the RNA. The pronounced ion accumulation could extend more than 20 Å away from the surface of RNA (63). Because of the large number of bound ions, the nonspecific ion binding contributes significantly to the overall RNA folding stability (102). In recent years, several novel experimental methods, such as “ion counting” (IC) (4, 56), SAXS (5), X-ray scattering interferometry (XSI) (126), and single-molecule optical trapping (SMOT) (2, 10) methods have been developed to probe ion binding properties. These experiments provided highly needed information about ion distribution and the number of bound ions around an RNA. In parallel with the experimental development, several new theories have been proposed to predict ion distribution and electrostatic free energies of the system.

Traditional theories such as counterion condensation theory (CC) and nonlinear Poisson-Boltzmann equation (NLPB) can give accurate treatment for monovalent ion solutions. However, for multivalent ions, such as Mg^{2+} , the traditional theories become less reliable. For example, for a tethered 12-basepair DNA duplex in a Mg^{2+} solution (5), NLPB significantly underestimates the fraction of the relaxed conformations and the number of excess Mg^{2+} ions around the DNA (4). More recently, Gebala *et al.* found that NLPB also underestimates the excess number of monovalent cations in a high concentration salt solution due to neglecting the cation-anion correlation effect (56).

One of the problems in the traditional theories is neglecting ion correlation and fluctuation effects. When ions accumulate around the RNA surface, ions are close to each other. The high local concentration of the bound ions leads to two effects. First, the long-range strong Coulomb force can cause a strong coupling between the ions, such that ion binding at one location is correlated to the binding of ions at other locations. Second, the electrostatic interactions become highly sensitive to the discrete distributions as well as the discrete properties (such as the size) of the ions.

MD simulations with proper force field with explicit ions and solvents can intrinsically account for ion-ion correlations. Indeed, simulations have provided significant insights into

the detailed information about ion-RNA interactions (28, 29, 45). A meaningful simulation requires a large number of ions in the systems, rendering a large simulation system and long simulation time. The problem is more pronounced for low ion concentration solutions. The difficulty in MD simulation has inspired the development of alternative approaches, such as the modified CC-like and NLPB-like theories (64,96), the three dimensional interaction site model (3D-RISM) (57), and the Tightly Bound Ions model (134, 136).

In the generalized Manning condensation (GMC) model (64), an explicit Mg^{2+} – Mg^{2+} ion correlation term is introduced to the classical Manning's ion condensation layer, and interaction potentials for the three different types of ions (explicit, condensed, and screening ions) are used in the Langevin dynamics. Moreover, to account for the volume exclusion effect of ions, the size-modified Poisson–Boltzmann equation (SMPBE) was developed by incorporating the finite ion sizes into NLPB through a lattice gas formulation (13). Applications of the SMPBE to ion-RNA interactions have led to improved predictions for the ion distributions (31).

Furthermore, in a recent study, Sushko *et al.* treated ion correlation and solvation effects in a density functional framework (135). Comparisons with SAXS data for the ion distribution support the importance of ion-ion correlation and hydration effects. 3D-RISM is another promising model for solvation and correlation effects based on the density distribution of ions (57). The model solves the Ornstein and Zernike (OZ) integral equation by averaging the solvent degrees of freedom. Applications of the 3D-RISM model to ion distribution around nucleic acids have led to reliable results. For example, the model gives two peaks in the ion density distribution around a nucleic acid helix, a result observed in explicit MD simulation. Because the 3D-RISM model takes into account both the ion-ion and the ion-solvent correlations, it can also be used to investigate specific binding of ions (118).

To treat ion correlation and fluctuation effects, Chen *et al.* developed the Tightly Bound Ion (TBI) model (136) in the statistical mechanical framework. The basic idea in the TBI model is to enumerate discrete ion distributions for strongly correlated ions (“tightly bound ions”), which are distributed in a thin layer around the RNA, and to use NLPB to treat the rest (weakly correlated) ions. For the original TBI model (136), the computational efficiency is severely limited by the time-consuming enumeration of the discrete ion distributions. As a result, the applicability of the model is restricted to RNAs less than 80 nucleotides long. Recently, using Monte Carlo simulation to sample the ion distributions (134), Chen *et al.* developed a new generation of the TBI model (“MCTBI”). The MCTBI model allows users to treat large RNAs of several hundreds of nucleotides long. The key idea in the new TBI model is a re-sampling algorithm for ion distribution: using Metropolis algorithm, to first randomly insert the tightly bound ions with bias toward low electric potential sites, followed by random removal of the ions with bias toward high electric potential sites.

LIGAND BINDING

Inspired by the tremendous potential of RNA-targeted therapeutics (127), researchers have developed a numbers of docking/scoring models for RNA-ligand binding (116). Different from ion binding, where an ion can be approximately modeled as a sphere, ligand binding

models involve sampling of different orientations of the (anisotropic) ligand. In Table 2 we show several computational tools developed in recent years. As shown in the table, these docking/scoring tools involve a variety of sampling and physics-based or knowledge-based scoring methods, and have been successful at different levels. In contrast to ligand-protein docking models, which are supported by extensive database of the available compounds and the affinity data, such a database for ligand-RNA docking is highly limited. Therefore, physics-based modeling may be an important approach of choice. In the future model development, a successful physics-based for ligand-RNA binding should account for several unique, important factors for ligand-RNA binding, including the ones listed below.

First, ligand-RNA interaction is highly sensitive to the charge distribution on the ligand. A ligand may contain positively charged (protonated) subunits such as amidinium, guanidinium and imidazole, and negatively charged (deprotonated) subunits such as ethanesulfonic acid and acetic acid. The positively charged subunits are more likely to be attracted to the (negatively charged) RNA backbone. For example, amino-glycosides that contain several positive charges are potent inhibitors for RNA binding (52). By comparing the binding affinities between the different aminoglycosides, including paromomycin, kanamycin-B, gen-tamycin, and tobramycin, Smith *et al.* demonstrated that the charge distribution is critical for specific binding (129). To model the effect, Moitessier and coworkers developed a specific scoring function for aminoglycosides-RNA binding (103).

Second, the nucleobase elements in some ligands can induce specific hydrogen bonding and stacking interaction. Experimental evidence shows that 2,6-diaminopurine could bind to the C74U mutation of the guanine riboswitch because of the formation of eight hydrogen bonds, while benzimidazole couldn't bind to the same riboswitch because it could only form two hydrogen bonds (58). In general, stacking interaction is much stronger than hydrogen bonding and hence stacking interaction is quite common in the stabilization of a ligand-riboswitch complex. For example, SAM riboswitch involves a full stacking between two nucleotides (92) at the ligand binding site.

Third, the interactions between RNA and the different cofactors, such as ions, proteins, and ligands, are often cooperatively or competitively coupled. For example, aminoglycoside binding to the bacterial ribosomal A-site RNA is in competition with metal ion binding, and binding of a synthetic ligand for an HCV RNA target is also found to be in competition with Mg^{2+} binding (25). For many riboswitches, ions facilitate ligand-induced conformational switch. For example, without Mg^{2+} , the ligand binding may not be sufficient to cause the structure change in the expression platform (83). Similarly, for a large class of riboswitches, such as metal ion sensing riboswitches (35,37), the thiamine pyrophosphate (TPP) riboswitches (141), and glycine riboswitches (89), specific ion-binding is often required for ligand binding.

Furthermore, ligand structure is often sensitive to RNA structural changes in the ligand-RNA binding process, such as the binding of RBT550 to HIV-1 TAR RNA (PDB ID: 1UTS) (106) and Argininamide to HIV-2 TAR RNA (PDB ID: 1AJU) (16). Therefore, conformational flexibility is another critical challenge that must be taken in account in ligand-RNA interaction models.

CONCLUSIONS

In the past decade we witnessed substantial achievements in physics-based and knowledge-based predictions for RNA folding. With the remarkable progress in structural and biophysical experiments, such as chemical mapping of RNA structures, SAXS and ion counting experiment for the probing of ion distributions and the number of bound ions, we begin to see successful integration of physics-based theoretical models for RNA folding and experimental data for the predictions of large RNA structures and the detailed mechanism of ion-RNA interactions. Furthermore, with the expansion of our knowledge about the known structures, we can realistically expect continuous improvements in the accuracy of knowledge-based structure predictions. Nevertheless, accurate predictions for large structures and for folding under cellular conditions are not solved problems. Many of the limitations in the current models stem from the inaccuracy in physics-based free energy models for RNA tertiary folds and ion binding and dehydration models of multivalent metal ions.

Furthermore, to understand RNA function, we need models for RNA folding *in vivo*. In the cellular environment, an RNA is surrounded by other macromolecules and small molecules, causing a crowded environment with up to 30% volume fraction(51). Crowders can cause significant differences between *in vivo* and *in vitro* RNA structures (78, 138). Recent experimental and theoretical analyses suggested that the excluded volume of the crowders can reduce the accessible space of RNA conformations, especially for the unfolded state, resulting in stabilization of the folded state (43, 71, 78, 107). Furthermore, recent theoretical studies demonstrated that organic crowders can weaken or strengthen ion-RNA binding as a result of the interplay between three crowding effects: steric competition between ions and crowders in RNA binding, crowder-induced change in the dielectric environment, which can change ion-RNA Coulomb interaction strength, and changes in the hydration state due to crowder-RNA binding(143).

In addition, RNA folding *in vivo* is cotranscriptional and kinetically controlled. Cotranscriptional folding of RNA often leads to very different structures (81) from *in vitro* refolding. The above complications combined highlight the demand for an all-encompassing physical model to incorporate the *in vivo* solution effect and cotranscriptional kinetics into structure prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Xiaojun Xu and Travis Hurst for critical reading of the manuscript and providing many insightful comments. This research was supported by National Institutes of Health grant R01-GM063732.

LITERATURE CITED

1. Aalberts DP, Hodas NO. Asymmetry in RNA pseudoknots: observation and theory. *Nucleic Acids Res.* 2005; 33:2210–4. [PubMed: 15831794]

2. Anthony PC, Sim AYL, Chu VB, Doniach S, Block SM, Herschlag D. Electrostatics of nucleic acid folding under conformational constraint. *J Am Chem Soc.* 2012; 134:4607–4614. [PubMed: 22369617]
3. Auffinger P, Bielecki L, Westhof E. The Mg²⁺ Binding Sites of the 5S rRNA Loop E Motif as Investigated by Molecular Dynamics Simulations. *Chem Biol.* 2003; 10:551–561. [PubMed: 12837388]
4. Bai Y, Greenfeld M, Herschlag D. Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J Am Chem Soc.* 2007; 129:14981–14988. [PubMed: 17990882]
5. Bai Y, Chu VB, Lipfert J, Pande VS, Herschlag D, Doniach S. Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble. *J Am Chem Soc.* 2008; 130:12334–12341. [PubMed: 18722445]
6. Bailor MH, Sun X, Al-Hashimi HM. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science.* 2010; 327:202–6. [PubMed: 20056889]
7. Bailor MH, Mustoe AM, Brooks CL III, Al-Hashimi HM. Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation. *Curr Opin Struct Biol.* 2011; 21:296–305. [PubMed: 21497083]
8. Bernauer J, Huang X, Sim AYL, Levitt M. Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA.* 2011; 17:1066–75. [PubMed: 21521828]
9. Bindewald E, Kluth T, Shapiro BA. CyloFold: secondary structure prediction including pseudoknots. *Nucleic Acids Res.* 2010; 38:W368W372.doi: 10.1093/nar/gkq432 [PubMed: 20501603]
10. Bizarro CV, Alemany A, Ritort F. Non-specific binding of Na⁺ and Mg²⁺ to RNA determined by force spectroscopy methods. *Nucleic Acids Res.* 2012; 40:6922–6935. [PubMed: 22492710]
11. Bon M, Orland H. Prediction of RNA secondary structures with pseudoknots. *Physica A.* 2010; 389:298792.
12. Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, Rother KM, Bujnicki JM. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* 2016; 44:e63. [PubMed: 26687716]
13. Borukhov I, Andelman D, Orland H. Steric Effects in Electrolytes: A Modified Poisson-Boltzmann Equation. *Phys Rev Lett.* 1997; 79:435–438.
14. Bowman GR, Huang X, Yao Y, Sun J, Carlsson G, Guibas LJ, Pande VS. Structural insight into RNA hairpin folding intermediates. *J Am Chem Soc.* 2008; 130:9676–8. [PubMed: 18593120]
15. Brion P, Westhof E. Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct.* 1997; 26:113–37. [PubMed: 9241415]
16. Brodsky AS, Williamson JR. Solution structure of the HIV-2 TAR-argininamide complex. *J Mol Biol.* 1997; 267:624–639. [PubMed: 9126842]
17. Brooks BR, Brucceeri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem.* 1983; 4:187–217.
18. Butcher SE, Pyle AM. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res.* 2011; 44:1302–11. [PubMed: 21899297]
19. Cao S, Chen SJ. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.* 2006; 34:2634–52. [PubMed: 16709732]
20. Cao S, Chen SJ. Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA.* 2009; 15:696–706. [PubMed: 19237463]
21. Cao S, Giedroc DP, Chen SJ. Predicting loop-helix tertiary structural contacts in RNA pseudoknots. *RNA.* 2010; 16:538–52. [PubMed: 20100813]
22. Cao S, Chen SJ. Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B.* 2011; 115:4216–26. [PubMed: 21413701]
23. Cao S, Xu X, Chen SJ. Predicting structure and stability for RNA complexes with intermolecular loop-loop base-pairing. *RNA.* 2014; 20:835–45. [PubMed: 24751648]

24. Capriotti E, Norambuena T, Marti-Renom MA, Melo F. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics*. 2011; 27:1086–93. [PubMed: 21349865]
25. Carnevali M, Parsons J, Wyles DL, Hermann T. A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *ChemBiochem*. 2010; 11:1364–7. DOI: 10.1002/cbic.201000177 [PubMed: 20564282]
26. Case, DA., Betz, RM., Botello-Smith, W., Cerutti, DS., Cheatham, TE., III, Darden, TA., Duke, RE., Giese, TJ., Gohlke, H., Goetz, AW., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, TS., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, KM., Monard, G., Nguyen, H., Nguyen, HT., Omelyan, I., Onufriev, A., Roe, DR., Roitberg, A., Sagui, C., Simmerling, CL., Swails, J., Walker, RC., Wang, J., Wolf, RM., Wu, X., Xiao, L., York, DM., Kollman, PA. AMBER 2016. University of California; San Francisco: 2016.
27. Chakraborty D, Collepardo-Guevara R, Wales DJ. Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins. *J Am Chem Soc*. 2014; 136:18052–61. [PubMed: 25453221]
28. Chen AA, Draper DE, Pappu RV. Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. *J Mol Biol*. 2009; 390:805819.
29. Chen H, Meisburger SP, Pabit SA, Sutton JL, Webb WW, Pollack L. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc Natl Acad Sci USA*. 2012; 17:799–804.
30. Chen AA, Garca AE. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci USA*. 2013; 110:16820–5. [PubMed: 24043821]
31. Chu V, Bai Y, Lipfert J, Herschlag D, Doniach S. Evaluation of Ion Binding to DNA Duplexes Using a Size-Modified Poisson–Boltzmann Theory. *Biophys J*. 2007; 93:3202–3209. [PubMed: 17604318]
32. Cordero P, Kladwang W, VanLang CC, Das R. Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry*. 2012; 51:7037–9. [PubMed: 22913637]
33. Cragolini T, Derreumaux P, Pasquali S. Ab initio RNA folding. *J Phys: Condens Matter*. 2015; 27:233102. [PubMed: 25993396]
34. Cragolini T, Laurin Y, Derreumaux P, Pasquali S. Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base pairings. *J Chem Theory Comput*. 2015; 11:3510–22. [PubMed: 26575783]
35. Cromie MJ, Shi Y, Latifi T, Groisman EA. An RNA sensor for intracellular Mg²⁺. *Cell*. 2006; 125:7184.
36. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang L, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Philips A, Puton T, Santalucia J, Sijen Y, Hermann T, Rother K, Rother M, Serganov A, Skorski M, Soltysinski T, Sripakdeevong P, Tuszynska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012; 18:610–25. [PubMed: 22361291]
37. Dann CE III, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. Structure and mechanism of a metal-sensing regulatory RNA. *Cell*. 2007; 130:878892.
38. Das R, Baker D. Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci USA*. 2007; 104:14664–69. [PubMed: 17726102]
39. Das R, Kudaravalli M, Jonikas M, Laederach A, Fong R, Schwans JP, Baker D, Piccirilli JA, Altman RB, Herschlag D. Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc Natl Acad Sci USA*. 2008; 105:4144–9. [PubMed: 18322008]
40. Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods*. 2010; 7:291–4. [PubMed: 20190761]
41. Dawson WK, Bujnicki JM. Computational modeling of RNA 3D structures and interactions. *Curr Opin Struct Biol*. 2016; 37:22–8. [PubMed: 26689764]
42. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA*. 2009; 106:97–102. [PubMed: 19109441]

43. Denesyuk NA, Thirumalai D. Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA. *J Am Chem Soc.* 2011; 133:11858–61. [PubMed: 21736319]
44. Denesyuk NA, Thirumalai D. Coarse-grained model for predicting RNA folding thermodynamics. *J Phys Chem B.* 2013; 117:4901–11. [PubMed: 23527587]
45. Denesyuk NA, Thirumalai D. How do metal ions direct ribozyme folding. *Nat Chem.* 2015; 7:793–801. [PubMed: 26391078]
46. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA.* 2008; 14:116473.
47. Ding F, Lavender CA, Weeks KM, Dokholyan NV. Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat Methods.* 2012; 9:603–8. [PubMed: 22504587]
48. Dirks RM, Pierce NA. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput Chem.* 2003; 24:1664–77. [PubMed: 12926009]
49. Do TN, Ippoliti E, Parrinello M. Counterion Redistribution upon Binding of a Tat-Protein Mimic to HIV-1 TAR RNA. *J Chem Theory Comput.* 2012; 8:688–694. [PubMed: 26596616]
50. Edwards TE, Ferré-D'Amaré AR. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure.* 2006; 14:1459–1468. [PubMed: 16962976]
51. Ellis RJ. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci.* 2001; 26:597–604. [PubMed: 11590012]
52. Ennifar E, Paillart JC, Bodlenner A, Walter P, Weibel JM, Aubertin AM, Pale P, Dumas P, Marquet R. Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: From crystal to cell. *Nucleic Acids Res.* 2006; 34:2328–2339. [PubMed: 16679451]
53. Eriksson ES, Joshi L, Billeter M, Eriksson LA. De novo tertiary structure prediction using RNA123—benchmarking and application to Macugen. *J Mol Model.* 2014; 20:2389. [PubMed: 25107358]
54. Flores SC, Wan Y, Russell R, Altman RB. Predicting RNA structure by multiple template homology modeling. *Pac Symp Biocomput.* 2010:216–27. [PubMed: 19908374]
55. Frellsen J, Moltke I, Thiim M, Mardia KV, Ferkinghoff-Borg J, Hamelryck T. A probabilistic model of RNA conformational space. *PLoS Comput Biol.* 2009; 5:e1000406. [PubMed: 19543381]
56. Gebala M, Giambasu GM, Lipfert J, Bisaria N, Bonilla S, Li G, York DM, Herschlag D. Cation-Anion Interactions within the Nucleic Acid Ion Atmosphere Revealed by Ion Counting. *J Am Chem Soc.* 2015; 137:14705–14715. [PubMed: 26517731]
57. Giambasu GM, Gebala MK, Panteva MT, Luchko T, Case DA, York DM. Competitive interaction of monovalent cations with DNA from 3D-RISM. *Nucleic Acids Res.* 2015; 43:8405–15. [PubMed: 26304542]
58. Gilbert SD, Mediatore SJ, Batey RT. Modified Pyrimidines Specifically Bind the Purine Riboswitch. *J Am Chem Soc.* 2006; 128:14214–5. [PubMed: 17076468]
59. Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM. Direct observation of hierarchical folding in single riboswitch aptamers. *Science.* 2008; 319:630–3. [PubMed: 18174398]
60. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res.* 2003; 31:439–41. [PubMed: 12520045]
61. Guilbert C, James TL. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J Chem Inf Model.* 2008; 48:1257–68. [PubMed: 18510306]
62. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci USA.* 2013; 110:5498–503. [PubMed: 23503844]
63. Hayes RL, Noel JK, Mohanty U, Whitford PC, Hennelly SP, Onuchic J, Sanbonmatsu KY. Magnesium fluctuations modulate RNA dynamics in the SAM-I riboswitch. *J Am Chem Soc.* 2012; 134:C12043–C12053.

64. Hayes RL, Noel JK, Mandic A, Whitford PC, Sanbonmatsu KY, Mohanty U, Onuchic JN. Generalized Manning Condensation Model Captures the RNA Ion Atmosphere. *Phys Rev Lett*. 2015; 114:258105. [PubMed: 26197147]
65. He Y, Maciejczyk M, Odziej S, Scheraga HA, Liwo A. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys Rev Lett*. 2013; 110:098101. [PubMed: 23496746]
66. Homan PJ, Favorov OV, Lavender CA, Kursun O, Ge X, Busan S, Dokholyan NV, Weeks KM. Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci USA*. 2014; 111:13858–63. [PubMed: 25205807]
67. Jared HD, Trenton RF, Marco T, Samuel EB. Role of metal ions in the tetraloopreceptor complex as analyzed by NMR. *RNA*. 2007; 13:79–86.
68. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*. 2009; 15:189–99. [PubMed: 19144906]
69. Jossinet F, Ludwig TE, Westhof E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*. 2010; 26:2057–9. [PubMed: 20562414]
70. Kerpedjiev P, Hner Zu Siederdisen C, Hofacker IL. Predicting RNA 3D structure using a coarse-grain helix-centered model. *RNA*. 2015; 21:1110–21. [PubMed: 25904133]
71. Kilburn D, Roh JH, Behrouzi R, Briber RM, Woodson SA. Crowders perturb the entropy of RNA energy landscapes to favor folding. *J Am Chem Soc*. 2013; 135:10055–63. [PubMed: 23773075]
72. Kim N, Zahran M, Schlick T. Computational prediction of riboswitch tertiary structures including pseudoknots by RAGTOP: a hierarchical graph sampling approach. *Methods Enzymol*. 2015; 553:115–35. [PubMed: 25726463]
73. Kladwang W, VanLang CC, Cordero P, Das R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat Chem*. 2011; 3:954–62. [PubMed: 22109276]
74. Kladwang W, VanLang CC, Cordero P, Das R. Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry*. 2011; 50:8049–56. [PubMed: 21842868]
75. Korolev N, Lyubartsev AP, Rupprecht A, Nordenskiöld L. Competitive binding of Mg²⁺, Ca²⁺, Na⁺, and K⁺ ions to DNA in oriented DNA fibers: experimental and Monte Carlo simulation results. *Biophys J*. 1999; 77:2736–2749. [PubMed: 10545373]
76. Krasovska MV, Sefcikova J, Sponer J. Cations and Hydration in Catalytic RNA: Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. *Biophys J*. 2006; 91:626–638. [PubMed: 16617077]
77. Krüger DM, Bergs J, Kazemi S, Gohlke H. Target Flexibility in RNA-Ligand Docking Modeled by Elastic Potential Grids. *ACS Med Chem Lett*. 2011; 2:489–493. [PubMed: 24900336]
78. Kwok CK, Ding Y, Tang Y, Assmann SM, Bevilacqua PC. Determination of in vivo RNA structure in low-abundance transcripts. *Nat Commun*. 2013; 4:2971. [PubMed: 24336128]
79. Laing C, Schlick T. Computational approaches to 3D modeling of RNA. *J Phys: Condens Matter*. 2010; 22:283101. [PubMed: 21399271]
80. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA*. 2009; 15:1219–1230. [PubMed: 19369428]
81. Leamy KA, Assmann SM, Mathews DH, Bevilacqua PC. Bridging the gap between in vitro and in vivo RNA folding. *Q Rev Biophys*. 2016; 49:e10. [PubMed: 27658939]
82. Lee TS, Radak BK, Harris ME, York DM. A Two-Metal-Ion-Mediated Conformational Switching Pathway for HDV Ribozyme Activation. *ACS Catal*. 2016; 6:1853–1869. [PubMed: 27774349]
83. Leipply D, Draper DE. Effects of Mg²⁺ on the Free Energy Landscape for Folding a Purine Riboswitch RNA. *Biochemistry*. 2011; 50:2790–2799. [PubMed: 21361309]
84. Leipply D, Draper DE. Evidence for a Thermodynamically Distinct Mg²⁺ Ion Associated with Formation of an RNA Tertiary Structure. *J Am Chem Soc*. 2011; 133:13397–405. [PubMed: 21776997]
85. Leontis NB, Stombaugh J, Westhof E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res*. 2002; 30:3497–531. [PubMed: 12177293]

86. Leontis NB, Lescoute A, Westhof E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol.* 2006; 16:279–87. [PubMed: 16713707]
87. Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. *RNA.* 2006; 12:83–93. [PubMed: 16373494]
88. Li P, Merz KM Jr. Taking into account the ion-induced dipole interaction in the non-bonded model of ions. *J Chem Theory Comput.* 2014; 10:289–297. [PubMed: 24659926]
89. Lipfert J, Sim AY, Herschlag D, Doniach S. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA.* 2010; 16:708–719. [PubMed: 20194520]
90. Liu L, Chen SJ. Computing the conformational entropy for RNA folds. *J Chem Phys.* 2010; 132:235104. [PubMed: 20572741]
91. Liu Y, Holmstrom E, Zhang J, Yu P, Wang J, Dyba MA, Chen D, Ying J, Lockett S, Nesbitt DJ, Ferré-D'Amaré AR, Sousa R, Stagno JR, Wang YX. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature.* 2015; 522:368–372. [PubMed: 25938715]
92. Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, Ke A. Crystal structures of the SAM-III/SMK riboswitch reveal the SAM-dependent translation inhibition mechanism. 2008
93. Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL. Predicting RNA secondary structures from sequence and probing data. *Methods.* 2016; 103:86–98. [PubMed: 27064083]
94. Lorenz R, Luntzer D, Hofacker IL, Stadler PF, Wolfinger MT. SHAPE directed RNA folding. *Bioinformatics.* 2016; 32:145–7. [PubMed: 26353838]
95. Low JT, Weeks KM. SHAPE-directed RNA secondary structure prediction. *Methods.* 2010; 52:150–8. [PubMed: 20554050]
96. Mak CH, Henke PS. Ions and RNAs: Free Energies of Counterion-Mediated RNA Fold Stabilities. *J Chem Theory Comput.* 2013; 9:621–639. [PubMed: 26589060]
97. Markham GD, Glusker JP, Bock CW. The arrangement of first- and second- sphere water molecules in divalent magnesium complexes: results from molecular orbital and density functional theory and from structural crystallography. *J Phys Chem B.* 2002; 106:51185134.
98. Martinez HM, Maizel JV Jr, Shapiro BA. RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn.* 2008; 25:669–83. [PubMed: 18399701]
99. Mathews DH, Turner DH. Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol.* 2006; 16:270–8. [PubMed: 16713706]
100. Meng Y, Aalberts DP. Free energy cost of stretching mRNA hairpin loops inhibits small RNA binding. *Biophys J.* 2013; 104:482–7. [PubMed: 23442870]
101. Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cheng C, Chojnowski G, Chou FC, Cordero P, Cruz JA, Ferré-D'Amaré AR, Das R, Ding F, Dokholyan NV, Dunin-Horkawicz S, Kladwang W, Krokhotin A, Lach G, Magnus M, Major F, Mann TH, Masquida B, Matelska D, Meyer M, Peselis A, Popenda M, Purzycka KJ, Serganov A, Stasiewicz J, Szachniuk M, Tandon A, Tian S, Wang J, Xiao Y, Xu X, Zhang J, Zhao P, Zok T, Westhof E. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA.* 2015; 21:1–19. [PubMed: 25516996]
102. Misra VK, Draper DE. On the role of magnesium ions in RNA stability. *Biopolymers.* 1998; 48:113–135. [PubMed: 10333741]
103. Moitessier N, Westhof E, Hanessian S. Docking of aminoglycosides to hydrated and flexible RNA. *J Med Chem.* 2006; 49:10231033.
104. Morley SD, Afshar M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *J Comput Aided Mol Des.* 2004; 18:189–208. [PubMed: 15368919]
105. Mustoe AM, Al-Hashimi HM, Brooks CL III. Coarse Grained Models Reveal Essential Contributions of Topological Constraints to the Conformational Free Energy of RNA Bulges. *J Phys Chem B.* 2014; 118:2615–2627. [PubMed: 24547945]
106. Murchie AI, Davis B, Isel C, Afshar M, Drysdale MJ, Bower J, Potter AJ, Starkey ID, Swarbrick TM, Mirza S, Prescott CD, Vaglio P, Aboul-ela F, Karn J. Structure-based drug design targeting

- an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J Mol Biol.* 2004; 336:625–638. [PubMed: 15095977]
107. Nakano S, Karimata HT, Kitagawa Y, Sugimoto N. Facilitation of RNA enzyme activity in the molecular crowding media of cosolutes. *J Am Chem Soc.* 2009; 131:16881–8. [PubMed: 19874030]
 108. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 2015; 43:D130–7. [PubMed: 25392425]
 109. Noid WG. Perspective: Coarse-grained models for biomolecular systems. *J Chem Phys.* 2013; 139:090901. [PubMed: 24028092]
 110. Panteva MT, Giambasu GM, York DM. Force Field for Mg(2+), Mn(2+), Zn(2+), and Cd(2+) Ions That Have Balanced Interactions with Nucleic Acids. *J Phys Chem B.* 2015; 119:15460–70. [PubMed: 26583536]
 111. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* 2008; 452:51–5. [PubMed: 18322526]
 112. Pavlov M, Siegbahn PEM, Sandstrom M. Hydration of beryllium, magnesium, calcium, and zinc ions using density functional theory. *J Phys Chem A.* 1998; 102:219–228.
 113. Pfeffer P, Gohlke H. DrugScoreRNA—knowledge-based scoring function to predict RNA-ligand interactions. *J Chem Inf Model.* 2007; 47:1868–1876. [PubMed: 17705464]
 114. Philips A, Milanowska K, Lach G, Boniecki M, Rother K, Bujnicki JM. MetalionRNA: computational predictor of metal-binding sites in RNA structures. *Bioinformatics.* 2012; 28:198–205. [PubMed: 22110243]
 115. Philips A, Milanowska K, Lach G, Bujnicki JM. LigandRNA: computational predictor of RNA-ligand interactions. *RNA.* 2013; 19:1605–1616. [PubMed: 24145824]
 116. Philips A, Lach G, Bujnicki JM. Computational methods for prediction of RNA interactions with metal ions and small organic ligands. *Methods Enzymol.* 2015; 553:261–285. [PubMed: 25726469]
 117. Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 2012; 40:e112. [PubMed: 22539264]
 118. Radak BK, Lee TS, Harris ME, York DM. Assessment of metal-assisted nucleophile activation in the hepatitis delta virus ribozyme from molecular simulation and 3D-RISM. *RNA.* 2015; 21:1566–1577. [PubMed: 26170378]
 119. Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA.* 2005; 11:14941504.
 120. Rice GM, Leonard CW, Weeks KM. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA.* 2014; 20:846–54. [PubMed: 24742934]
 121. Rother M, Rother K, Puton T, Bujnicki JM. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* 2011; 39:4007–22. [PubMed: 21300639]
 122. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD. rDock: A fast, versatile and open source code for docking ligands to proteins and nucleic acids. *PLoS Comput Biol.* 2014; 10:e1003571. [PubMed: 24722481]
 123. Schaffer MF, Peng G, Spingler B, Schnabl J, Wang M, Olieric V, Sigel RKO. The X-ray Structures of Six Octameric RNA Duplexes in the Presence of Different Di- and Trivalent Cations. *Int J Mol Sci.* 2016; 17:988.
 124. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol.* 2004; 11:1729–1741. [PubMed: 15610857]
 125. Shi YZ, Wang FH, Wu YY, Tan ZJ. A coarse-grained model with implicit salt for RNAs: Predicting 3D structure, stability and salt effect. *J Chem Phys.* 2014; 141:105102. [PubMed: 25217954]

126. Shi X, Huang L, Lilley DMJ, Harbury PB, Herschlag D. The Solution Structural Ensembles of RNA Kink-turn Motifs and Their Protein Complexes. *Nat Chem Biol.* 2016; 12:146–152. [PubMed: 26727239]
127. Shortridge MD, Varani G. Structure based approaches for targeting non-coding RNAs with small molecules. *Curr Opin Struct Biol.* 2015; 30:79–88. [PubMed: 25687935]
128. Sloma MF, Mathews DH. Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol.* 2015; 553:91–114. [PubMed: 25726462]
129. Smith AL, Kassman J, Srouf KJ, Soto AM. Effect of salt concentration on the conformation of TAR RNA and its association with aminoglycoside antibiotics. *Biochemistry.* 2011; 50:9434–9445. [PubMed: 21916409]
130. Somarowthu S. Progress and Current Challenges in Modeling Large RNAs. *J Mol Biol.* 2016; 428:736–47. [PubMed: 26585404]
131. Sperschneider J, Datta A, Wise MJ. Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. *RNA.* 2011; 17:27–38. [PubMed: 21098139]
132. Sripakdeevong P, Kladwang W, Das R. An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc Natl Acad Sci USA.* 2011; 108:20573–8. [PubMed: 22143768]
133. Sripakdeevong P, Cevic M, Chang AT, Erat MC, Ziegeler M, Zhao Q, Fox GE, Gao X, Kennedy SD, Kierzek R, Nikonowicz EP, Schwalbe H, Sigel RK, Turner DH, Das R. Structure determination of noncanonical RNA motifs guided by H NMR chemical shifts. *Nat Methods.* 2014; 11:413–6. [PubMed: 24584194]
134. Sun LZ, Chen SJ. Monte Carlo Tightly Bound Ion model: Predicting ion binding properties of RNA with ion correlations and fluctuations. *J Chem Theory Comput.* 2016; 12:33703381.
135. Sushko ML, Thomas DG, Pabit SA, Pollack L, Onufriev AV, Baker NA. The role of correlation and solvation in ion interactions with B-DNA. *Biophys J.* 2016; 110:315–326. [PubMed: 26789755]
136. Tan ZJ, Chen SJ. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J Chem Phys.* 2005; 122:44903. [PubMed: 15740294]
137. Thomas PD, Dill KA. An iterative method for extracting energy-like quantities from protein structures. *Proc Natl Acad Sci USA.* 1996; 93:11628–33. [PubMed: 8876187]
138. Tyrrell J, McGinnis JL, Weeks KM, Pielak GJ. The cellular environment stabilizes adenine riboswitch RNA structure. *Biochemistry.* 2013; 52:8777–85. [PubMed: 24215455]
139. Wu Y, Shi B, Ding X, Liu T, Hu X, Yip KY, Yang ZR, Mathews DH, Lu ZJ. Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.* 2015; 43:7247–59. [PubMed: 26170232]
140. Xia Z, Bell DR, Shi Y, Ren P. RNA 3D structure prediction by using a coarse-grained model and experimental data. *J Phys Chem B.* 2013; 117:313544.
141. Yamauchi T, Miyoshi D, Kubodera T, Nishimura A, Nakai S, Sugimoto N. Roles of Mg²⁺ in TPP-dependent riboswitch. *FEBS Lett.* 2005; 579:2583–2588. [PubMed: 15862294]
142. Yang S, Parisien M, Major F, Roux B. RNA structure determination using SAXS data. *J Phys Chem B.* 2010; 114:10039–48. [PubMed: 20684627]
143. Yu T, Zhu Y, He Z, Chen SJ. Predicting Molecular Crowding Effects in Ion-RNA Interactions. *J Phys Chem B.* 2016; doi: 10.1021/acs.jpcb.6b05625
144. Zhang J, Dundas J, Lin M, Chen R, Wang W, Liang J. Prediction of geometrically feasible three-dimensional structures of pseudoknotted RNA through free energy estimation. *RNA.* 2009; 15:2248–63. [PubMed: 19864433]
145. Zhang Y, Zhang J, Wang W. Atomistic Analysis of Pseudoknotted RNA Unfolding. *J Am Chem Soc.* 2011; 133:6882–5. [PubMed: 21500824]
146. Zhang J, Ferré-D'Amaré AR. Dramatic improvement of crystals of large RNAs by cation replacement and dehydration. *Structure.* 2014; 22:1363–1371. [PubMed: 25185828]
147. Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. Automated and fast building of three-dimensional RNA structures. *Sci Rep.* 2012; 2:734. [PubMed: 23071898]

148. Zirbel CL, Sponer JE, Sponer J, Stombaugh J, Leontis NB. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* 2009; 37:4898–918. [PubMed: 19528080]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Loop entropy and free energy parameters for the formation of a unimolecular long-range kissing structure.

Figure 2. The main elements for several computational approaches for RNA 3D structure modeling. Other models, such as MC-Fold/MC-Sym (111), may use a hybrid form of the elements listed in the figure.

Figure 3. Specific binding of ions in the adenine riboswitch. (A) PDB ID: 1Y26 (124), [Mg²⁺]=200mM; (B) PDB ID: 4TZX (146), [Mg²⁺]=200mM; (C) PDB ID: 4TZY (146), [Mg²⁺]=150mM and [Sr²⁺]=50mM; (D) PDB ID: 4NXR (91), [Mg²⁺]=100mM. The red and orange balls present Mg²⁺ ions and Sr²⁺ ions.

Table 1

An overview of a selected set of 3D structure prediction models

Name/Reference	Operation material	Sampling technique	Scoring function
FARNA/FARFAR(38,40)	Trinucleotide fragment	Monte Carlo fragment assembly	Knowledge-based/full atomic force field
MC-Fold/MC-Sym(111)	Nucleotide cyclic motif	Las Vegas algorithm	Knowledge-based scoring function
RNA2D3D(98), Assemble(69)	User-defined	Interactive graphical editing	User-defined
RNABuilder(54)	Whole structure	User-defined constraints satisfaction	User-defined
ModeRNA(121)	Whole structure	Sequence alignment	Sequence similarity
Vfold(22)	2D structure based templates	Template match	Loop size and sequence similarity
RNAComposer(117)	Stems, loops and single strands	Fragment match	2D structure topology and sequence similarity
3dRNA(147)	Smallest secondary elements (SSEs)	SSE topology match	Sequence identity
RNA123(53)	Fragment and template	Template match	Knowledge-based force field
Graph theory approach(70, 72)	Graph representation	Monte Carlo graph sampling	Knowledge-based energy function
DMD(46), HiRE-RNA(34), SimRNA(12)	Coarse-grained beads	MD/MC simulation	Knowledge/physics-based force field

Table 2

Summary of a selected set of ligand-RNA docking tools

Name/Reference	Binding target(s)	Descriptions
DrugScore ^{RNA} (77, 113)	RNA	The knowledge-based scoring function contains a set of distance-dependent potentials between two atoms in ligand and flexible target, respectively.
MORDOR (61)	RNA and Protein	The CHARMM-27 and general AMBER force fields are applied to RNA and ligand, respectively. The model considers flexible conformations of the ligand and the RNA target.
DOCK6 (80)	RNA and Protein	An extended version is developed for targeting RNA. The physics-based scoring function contains the effects of solvation and sodium ion effects on docking.
ligandRNA (115)	RNA	The model scores and ranks ligand poses in ligand-RNA complexes. The knowledge-based function contains a set of distance- and angle-dependent statistical potentials for atom pairs between ligand and RNA.
rDOCK(104, 122)	Nucleic acids and protein	The package is open source. The scoring function is specifically developed for RNA-ligand docking by taking into account the stacking interactions and the charge-distributions in both RNA and ligand.