

The Multiscale Future of RNA Modeling

Petr Šulc^{1,*}

¹School of Molecular Sciences and Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University, Tempe, Arizona

RNA is a versatile molecule that plays essential roles in living organisms, ranging from information transfer to sensing, enzymatic, structural functions, and regulation (1). High-throughput sequencing methods continue to identify new noncoding RNA transcripts in cells, and their function (or lack of) remains an active topic of research. Designed RNAs also play an increasingly important role in the fields of nucleic acid nanotechnology and synthetic biology, in which they are engineered for diagnostic as well as therapeutic applications (2,3). Potential new therapies for COVID-19 involve targeting structural motifs in the folded RNA genome of SARS-CoV-2 to impede its function (4), highlighting the need for modeling the structure of large RNA sequences.

Recently, the RNA-Puzzles computational resource evaluation has been introduced (5) to assess different algorithms for blind de novo RNA structure prediction, an RNA equivalent of the long-running CASP-blind prediction experiment for protein folding software. It is an exciting time for RNA structure modeling, with new experimental data sets and new methods for RNA modeling being introduced. The increase in data has inspired novel modeling approaches and combinations of existing techniques. In partic-

ular, multiscale modeling combines multiple approaches by studying the same system at different levels of resolution (e.g., atomistic and residue-level) and integrating the results. Such approaches have been successfully used in modeling variety of systems such as actin and chromatin (6,7). The new work from Shi-Jie Chen's group, "Modeling loop composition and ion concentration effects in RNA hairpin folding stability" (8), published in the current issue of the *Biophysical Journal*, shows a promising method to use multiscale models for studying the conformational folding landscape of RNA molecules.

RNA function is closely related to its structure. RNA sequence adopts a secondary structure, defined by a list of canonical basepairs (Fig. 1). To predict secondary structure, dynamic programming-based algorithms (9,10) compute the structure with the lowest free energy based on the nearest-neighbor model developed by Turner and collaborators (11). RNA structures also contain tertiary structure interactions in which additional contacts give rise to a three-dimensional (3D) structure. Modeling these contacts has proven difficult because reliable thermodynamic data to estimate their stability are scarce, and missing these interactions often results in incorrect predictions.

At the same time, 3D molecular simulations have played an increasingly important role for molecular structure research, with fully atomistic models

being the most popular. However, because of limitations of timescales and system sizes amenable to study at such resolution, as well as the challenges involved in parametrizing the atomistic force fields to capture all relevant properties, alternatives to fully atomistic simulations for folding simulations or RNA structure predictions are also an active area of research.

One promising approach is to model RNA using coarse-grained models, in which groups of atoms are replaced by a single bead. Effective interactions between these beads are then parameterized to reproduce particular features of the molecule in question, such as known structural motifs or stem formation thermodynamics. Numerous coarse-grained RNA models have been proposed over the last few years (12–16), using various degrees of coarse graining. Another class of coarse-grained models uses graph-theory abstractions of motifs encountered in RNA structures (17–19). One of the most difficult problems in RNA modeling is the handling of ion effects. Although the thermodynamic data used to parametrize the Turner model were obtained at 1M sodium concentration, which screens long-range electrostatic effects, natural RNAs fold in physiological salt conditions that allow such electrostatic effects and include multivalent ions, particularly magnesium, that stabilize specific tertiary contacts. Ultimately every model is a trade-off between accurately capturing a certain set of RNA properties and

Submitted July 13, 2020, and accepted for publication August 25, 2020.

*Correspondence: psulc@asu.edu

Editor: Christine Heitsch.

<https://doi.org/10.1016/j.bpj.2020.08.026>

© 2020 Biophysical Society.

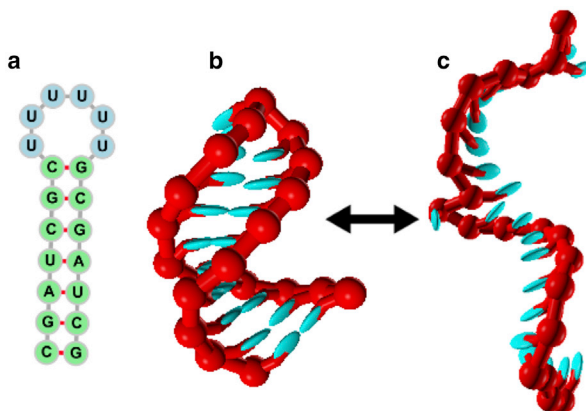


FIGURE 1 An RNA hairpin represented as a secondary structure with canonical base pairs shown in green (a) and a schematic 3D representation (b). In simulations and experiments, it is possible to measure fraction of the closed (b) and open (c) states. To see this figure in color, go online.

efficiency. The most likely path forward requires multiscale integration of various approaches, as showcased by Chen et al. (8).

For many years, Shi-Jie Chen's lab has been making progress on multiple aspects of RNA modeling, including secondary structure models (Vfold2D) as well as 3D coarse-grained models (16,20) and applying them, among other tools, successfully in the RNA-Puzzles challenge. In their latest work, published in the current issue, they introduce a new multiscale approach to study the folding of RNA loops in different ionic concentrations. They provide a general method that combines possible secondary structures, predicted by Vfold2D, with 3D molecular dynamics simulations using their coarse-grained model, IsRNA (16). They show that they can accurately reproduce yields observed in experimental studies of hairpin folding with long loop sequences and varying ionic conditions.

Their work tackles some of the challenges that one faces when modeling RNA experiments, such as inaccurate description of possible states when considering secondary structure alone and difficulty with the correct treatment of ion effects. Chen et al. use Vfold2D to generate possible secondary structures that are then converted into 3D representations using their Vfold3D model (20). These distinct

structures are then the starting points of IsRNA simulations that sample (within few hours) various configurations that the RNA can adopt. The authors then use a genetic algorithm to parametrize a weighted combination of effective coarse-grained interaction terms to obtain a free-energy difference between the open and closed states of the hairpin. Thus, the model can accurately extrapolate yields in experimental conditions with different salt concentrations. The authors use their previously developed method, Monte Carlo Tightly Bound Ion Model, which allows for representation of ion electrostatic effects at physiological salt concentrations or lower. In such conditions, the Debye-Hückel approximation, a popular choice among coarse-graining community, is no longer valid. Their approach can be straightforwardly extended to divalent ions too, opening a pathway for more accurate modeling of RNA behavior in different ionic conditions.

Even though their article currently only deals with hairpins, the researchers illustrate a promising way forward for the coarse-grained modeling field of RNA. We will likely see more multiscale approaches in the future, combining two-dimensional and 3D structure prediction tools and potential functions fitted to experiments to inform the coarse-grained simulations of RNA folding. Additionally, the article shows

how experimental data can be incorporated into model parametrization: the field of RNA modeling currently lacks reliable free-energy estimates outside the data sets that were used to parametrize Turner model. Even if we resolve a structure of an RNA molecule with tertiary contacts, there can be other transient folds or populations of alternative structures, which make parametrization of a tertiary contact's free-energy stabilization difficult. Combinations of structure prediction and modeling, such as the one demonstrated in the new study, are required to identify these structures.

Given the overall efficiency of the methods presented in this new study (8), the approach can likely be extended to larger RNAs, paving new ways of integrative modeling across more scales to study large RNA structures in complex monovalent and divalent ionic environments. Given the growing interest in artificial intelligence methods, one might be tempted to think that with more RNA structures available, the folding-problem solution will be based on data-driven machine-learning approaches such as neural networks. However, a true advance toward accurate long-running simulations will almost certainly require careful combination of experiments and multiscale models of two-dimensional and 3D structures that integrate knowledge-based potential with physics-based interaction terms. The work by Chen and collaborators highlights how such an approach might be developed, warranting their selection for a New and Notable article.

REFERENCES

1. Elliott, D., and M. Ladomery. 2017. *Molecular Biology of RNA*. Oxford University Press, Oxford, UK.
2. Green, A. A., P. A. Silver, ..., P. Yin. 2014. Toehold switches: de-novo-designed regulators of gene expression. *Cell*. 159:925–939.
3. Qi, X., X. Liu, ..., Y. Chang. 2020. RNA origami nanostructures for potent and safe anticancer immunotherapy. *ACS Nano*. 14:4727–4740.
4. Rangan, R., I. N. Zheludev, ..., R. Das. 2020. RNA genome conservation and

- secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA*. 26:937–959.
5. Miao, Z., R. W. Adamiak, ..., E. Westhof. 2017. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*. 23:655–672.
 6. Zhou, J., I. F. Thorpe, ..., G. A. Voth. 2007. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophys. J.* 92:4289–4303.
 7. Portillo-Ledesma, S., and T. Schlick. 2020. Bridging chromatin structure and function over a range of experimental spatial and temporal scales by molecular modeling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 10:e1434.
 8. Zhao, C., D. Zhang, ..., S. -J. Chen. 2020. Modeling loop composition and ion concentration effects in RNA hairpin folding stability. *Biophys. J.* 119:1439–1455.
 9. Zadeh, J. N., C. D. Steenberg, ..., N. A. Pierce. 2011. NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* 32:170–173.
 10. Lorenz, R., S. H. Bernhart, ..., I. L. Hofacker. 2011. ViennaRNA package 2.0. *Algorithms Mol. Biol.* 6:26.
 11. Mathews, D. H., J. Sabina, ..., D. H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288:911–940.
 12. Šulc, P., T. E. Ouldridge, ..., A. A. Louis. 2015. Modelling toehold-mediated RNA strand displacement. *Biophys. J.* 108:1238–1247.
 13. Denesyuk, N. A., and D. Thirumalai. 2013. Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B.* 117:4901–4911.
 14. Pasquali, S., and P. Derreumaux. 2010. HiRE-RNA: a high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B.* 114:11957–11966.
 15. Xia, Z., D. R. Bell, ..., P. Ren. 2013. RNA 3D structure prediction by using a coarse-grained model and experimental data. *J. Phys. Chem. B.* 117:3135–3144.
 16. Zhang, D., and S. -J. Chen. 2018. IsRNA: an iterative simulated reference state approach to modeling correlated interactions in RNA folding. *J. Chem. Theory Comput.* 14:2230–2239.
 17. Jain, S., S. Saju, ..., T. Schlick. 2019. An extended dual graph library and partitioning algorithm applicable to pseudoknotted RNA structures. *Methods*. 162-163:74–84.
 18. Jain, S., Q. Zhu, ..., T. Schlick. 2020. Identification of novel RNA design candidates by clustering the extended RNA-As-Graphs library. *Biochim. Biophys. Acta Gen. Subj.* 1864:129534.
 19. Schlick, T. 2018. Adventures with RNA graphs. *Methods*. 143:16–33.
 20. Cao, S., and S. -J. Chen. 2011. Physics-based de novo prediction of RNA 3D structures. *J. Phys. Chem. B.* 115:4216–4226.