Challenges in RNA Structural Modeling and Design

Over the past several decades, it has become increasingly clear that functionally active RNA molecules are not restricted to a few unusual examples that are involved in translation (such as transfer RNA and ribosomal RNA) and RNA processing (like spliceosomal RNAs, self-splicing ribozymes, self-cleaving ribozymes and ribonuclease P). Rather, large and small regulatory RNAs are transcribed in abundance and are involved in every aspect of metabolism. Even messenger RNA molecules contain highly structured components, and there is growing awareness that coding regions themselves contain functional structural motifs. While the expanding number of regulatory RNAs is astonishing, even more daunting is our lack of information about their structures. Most of our knowledge comes exclusively from primary sequence data, and our tools for inferring secondary and tertiary structural information are limited.

To address this challenge, Anna Marie Pyle and Tamar Schlick organized a small, focused meeting in Telluride, Colorado, during the summer of 2014, titled: "Challenges in RNA structural modeling and design", which brought together structural biologists, computational biologists and mathematicians to explore methods for leveraging the abundance of new information on RNA in order to gain structural and functional knowledge. To foster continuing exploration in this area, Pyle and Schlick set out to highlight the topic in the *Journal of Molecular Biology* by creating this special issue by the same name. The set of papers within this issue explore a diversity of new strategies for solving, modeling and designing RNA structures and many of them expand upon a set of working themes that were generated by the Telluride meeting group. These themes formulated a set of challenges for future work that will facilitate our understanding of RNA structural modeling and design. As a preface to the Special Issue, Pyle and Schlick summarize them below:

- **Improve the "input data" for prediction algorithms.** To maximize the power of new mathematical and computational modeling approaches, it is essential that our input data be constantly improved, and three areas were highlighted:
  A. *Continue to improve and optimize our utilization of the fundamental principles that determine RNA structure* (*also called Turner's rules*). Nearest-neighbor interactions are robust parameters that represent a central anchor for all of secondary structure prediction. However the context dependence of these terms and their conditional variation make it important to continually improve them in order to advance the power of prediction algorithms. See the contribution from Somarowthu [1] and from Das and the EteRNA laboratory [2].
  B. *Improve and diversify biochemical probes of RNA structure.* SHAPE and DMS probing have revolutionized our ability to visualize RNA secondary structural elements. See the contribution from the Bevilacqua group in this collection [3]. However, it remains important to continually explore methods for accurately assessing SHAPE data, to incorporate data from other types of probes (such as hydrolytic metal ions, cross-links and enzymes), and perhaps most importantly, for biochemists to design new probes, particularly those that are uniquely sensitive to double-stranded RNA structure.
  C. *Incorporate phylogenetic data into prediction and design*. RNA structures of functional RNAs that do not code for proteins are conserved during evolution. The functions of these molecules are mediated by their structures on which selective pressure acts. However, sequence motifs can sometimes be replaced without functional consequence. Indeed, variation in sequence may hide phylogenetic signals. The freely available functional data in phylogenetic covariation can be readily mined from alignments of RNA sequence,

and it is vital that we identify new ways to incorporate these data into our pipelines for predicting structures and macromolecular interfaces.

D. *Integrate tertiary structural features into secondary structural prediction*. Long-range tertiary interactions are not readily predicted and robust energetic terms for other types of long-range tertiary interactions (such as tetraloop–receptor interactions) are lacking. Accounting for RNA tertiary structural elements is essential for accurately predicting secondary structures, as RNA folding is not purely hierarchical, and also for visualizing the final, native tertiary structure of an RNA molecule. See, for example, the contribution from the Encarna-Salas group in this collection [4].

- **Obtain consensus on standardization and normalization for transcriptome-mapping**. A wealth of data is being produced through genome-wide approaches. Our ability to leverage this for structure determination will benefit from creating standards for quantitation and interpretation of this data.
- **Obtain more high-resolution RNA structures and confront the technical issues with large RNA structural determination**. It is important to support new initiatives to solve diverse RNA tertiary structures, as the motifs in these structures are the building blocks for modeling and, ultimately, for designing novel RNA structures from sequence. See the contribution from the Butcher group [5].
- **Develop new approaches for annotating RNA tertiary structural motifs.** There should be integrated resources to provide multivariable, multifunctional data on each type of tertiary interaction so that they can be utilized in modeling and design, much as it currently exists for the K-turn motif. See the perspective from Huang and Lilley in this collection [6].
- **Appreciate the potential significance of co-transcriptional folding.** It is important to recognize the potential significance of polarity in RNA synthesis in folding and assembly of certain RNA molecules. The order in which domains of RNA are synthesized introduces unique computational challenges to structure prediction. See the original research article from Das and the EteRNA global laboratory in this collection [2].
- **Less is more: Appreciate the importance of diverse coarse-grained approaches for RNA modeling.** Often, more accurate RNA models are generated with data that are less precise, but which reflect attributes of the global and basic thermodynamic parameters of RNA. It is therefore important to "imagine" RNA in new ways that accurately represent its properties in a coarse-grained manner. Getting from these to all-atom representations is important, but it is a different challenge. See examples from the Jernigan and Schlick groups in this collection [7,8].
- **Acknowledge the presence of multiple RNA structural solutions.** RNA adopts multiple stable structural states and often needs to sample among them. It is therefore an important challenge to calculate all of these, or at least include their sampling, in our approaches. See the original research article from Das and the EteRNA global laboratory in this collection [2].

Other ideas like the role of modification on the structure and function of RNA molecules are woven into the articles of this special issue as in the original research article from the group of Zhou and Pan [9]. We hope that by bringing these thoughts to the community, it will be possible to stimulate innovation in the development of tools for visualizing and understanding RNA structure.

Acknowledgements

The TSRC meeting "Challenges in RNA modeling and design" was supported in part by the National Institutes of General Medical Sciences under Award Number R13 GM112216 to the scientific organizers Tamar Schlick and Anna Marie Pyle. The content of the research is solely the responsibility of the meeting attendees and does not represent views of the National Institutes of Health.

References

[1] Srinivas Somarowthu, Progress and current challenges in modeling large RNAs, J. Mol. Biol. (14 November 2015), http://dx.doi.org/10.1016/j.jmb.2015.11.011 (Available online).

[2] Janderson Lee (EteRNA player); Eli Fisker (EteRNA player); Vineet Kosaraju; Michelle J Wu; Justin Kong; Minjae Lee; mat747 (EteRNA player); Adrien Treuille, Ph.D.; Rhiju Das, Ph.D.; EteRNA players Principles for predicting RNA secondary structure design difficulty, J. Mol. Biol.
[3] Yin Tang, Sarah M. Assmann, Philip C. Bevilacqua, Protein structure is related to RNA structural reactivity *in vivo*, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2015.11.012.
[4] Gloria Lozano, Noemi Fernandez, Encarnacion Martinez-Salas, Modeling three-dimensional structural motifs of viral IRES, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2016.01.005.
[5] Gabriel Cornilescu, Allison L. Didychuk, Margaret L. Rodgers, Lauren A. Michael, Jordan E. Burke, Eric J. Montemayor, Aaron A. Hoskins, Samuel E. Butcher, Structural analysis of multi-helical rnas by NMR–SAXS/WAXS: application to the U4/U6 di-snRNA, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2015.11.026.
[6] Lin Huang, David M.J. Lilley, The kink turn, a key architectural element in RNA structure, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2015.09.026.
[7] Michael T. Zimmermann, Kejue Jia, Robert L. Jernigan, Ribosome mechanics informs about mechanism, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2015.12.003.
[8] Naoto Baba, Shereef Elmetwaly, Namhee Kim, Tamar Schlick, Predicting large RNA-like topologies by a knowledge-based clustering approach, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2015.10.009.
[9] Katherine I. Zhou, Marc Parisien, Qing Dai, Nian Liu, Luda Diatchenko, Joseph R. Sachleben, Tao Pan, N6-methyladenosine modification in a long noncoding RNA hairpin predisposes its conformation to protein binding, J. Mol. Biol., http://dx.doi.org/10.1016/j.jmb.2015.08.021.

Anna Marie Pyle*
Yale University and Howard Hughes Medical Institute
E-mail address: anna.pyle@yale.edu.

Tamar Schlick*
New York University
E-mail address: schlick@nyu.edu.

*Corresponding authors