doi:10.1016/j.jmb.2012.05.027 J. Mol. Biol. (2012) xx, xxx–xxx

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

COMMENTARY

A New Toolkit for Modeling RNA from a Pseudo-Torsional Space Commentary on "Discrete RNA libraries from pseudo-torsional space" by Humphris-Narayanan and Pyle (*J. Mol. Biol.* March 2012)

RNA's diverse cellular roles from catalysis and molecular recognition to genetic regulation are highly dependent on its tertiary structure (see recent reviews in Refs. 1–3 and breakthrough papers in Refs. 4–8). Fortunately, the number of known RNA tertiary structures has greatly increased over the past 20 years through experimental studies. The emerging field of RNA computational biology has capitalized well upon this growing information to develop tools that model RNA tertiary structures (see recent reviews in Refs. 9 and 10). In turn, computational approaches have also contributed to annotating and analyzing RNA's modular and hierarchical features.[11–14] For example, complex RNA structures are abstracted through coarse-grained representations, networks,[15,16] beads,[17,18] or two-dimensional graphs,[19] and these have allowed classification and cataloguing of the resulting RNA shapes in the form of databases or libraries. Such cataloguing of RNA topologies has also led to RNA design.[19–23] In parallel, RNA structures have been predicted by assembling fragments from the coarse-grained libraries using programs such as MC-Sym[15,16] and NAST[17,18] or by borrowing from the fragment assembly/minimization protocol for proteins as in the RNA program FARNA.[24] In this issue of the *Journal of Molecular Biology*, Humphris-Narayanan and Pyle ("HP" for short) develop a discrete RNA rotameric library to analyze, catalogue, and assemble structural features of RNA without minimization or sampling.[25]

Specifically, HP represent RNA structures using a virtual bond system that reduces the backbone's seven torsion degrees to two pseudo-torsional angles. Earlier work by Richardson, Pyle, and coworkers provided a solid foundation for this modeling approach: in a collaboration with the RNA Ontology consortium,[26] the Richardson group developed the 8466 RNA rotameric data set for seven backbone torsion angles[27] and identified the 46 consensus clusters of RNA backbone conformations[26]; Pyle's group focused on coarse-grained modeling[28] and analysis of each backbone structure.[29] In particular, Pyle *et al.* earlier introduced the two pseudo-torsion angles (η–θ) defined by pseudo-bonds formed by the carbon (C4′) and phosphate (P) atoms to facilitate the classification of the RNA backbone of each nucleotide[28] (η: $C4'_{i-1}$-P_i-$C4'_i$-P_{i+1}, θ: P_i-$C4'_i$-P_{i+1}-$C4'_{i+1}$; see Fig.1 in the HP work[25] and Fig. 1 in this commentary). They then analyzed the η–θ angle distributions in Richardson's RNA backbone data set, an approach mimicking Ramachandran plots for protein backbone description.[29] Now, HP make the critical connection between these torsion angle analyses and the modeling of RNA structures: they identify basis sets of nucleotide conformers or "fragments" spanning the RNA pseudo-torsional space, and they assemble these fragments to construct RNAs, ranging from 2 to 174 nucleotides.

On its own, such a fragment assembly method would require additional tools for model building such as energy formulation and minimization. However, the discrete RNA nucleotide library based on a simplified pseudo-torsional representation of RNA backbone allows exhaustive sampling of RNA backbone structures.[25] As in a previous study,[29] nucleotides are classified into two groups (C3′-*endo* or C2′-*endo* sugar pucker), and two η–θ plots are generated for each group (see Fig. 2 in the HP work[25]). The further binning of these plots into six groups over regular angular intervals from 5° to

Please cite this article as: Kim, N. & Schlick, T. A New Toolkit for Modeling RNA from a Pseudo-Torsional Space, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.05.027

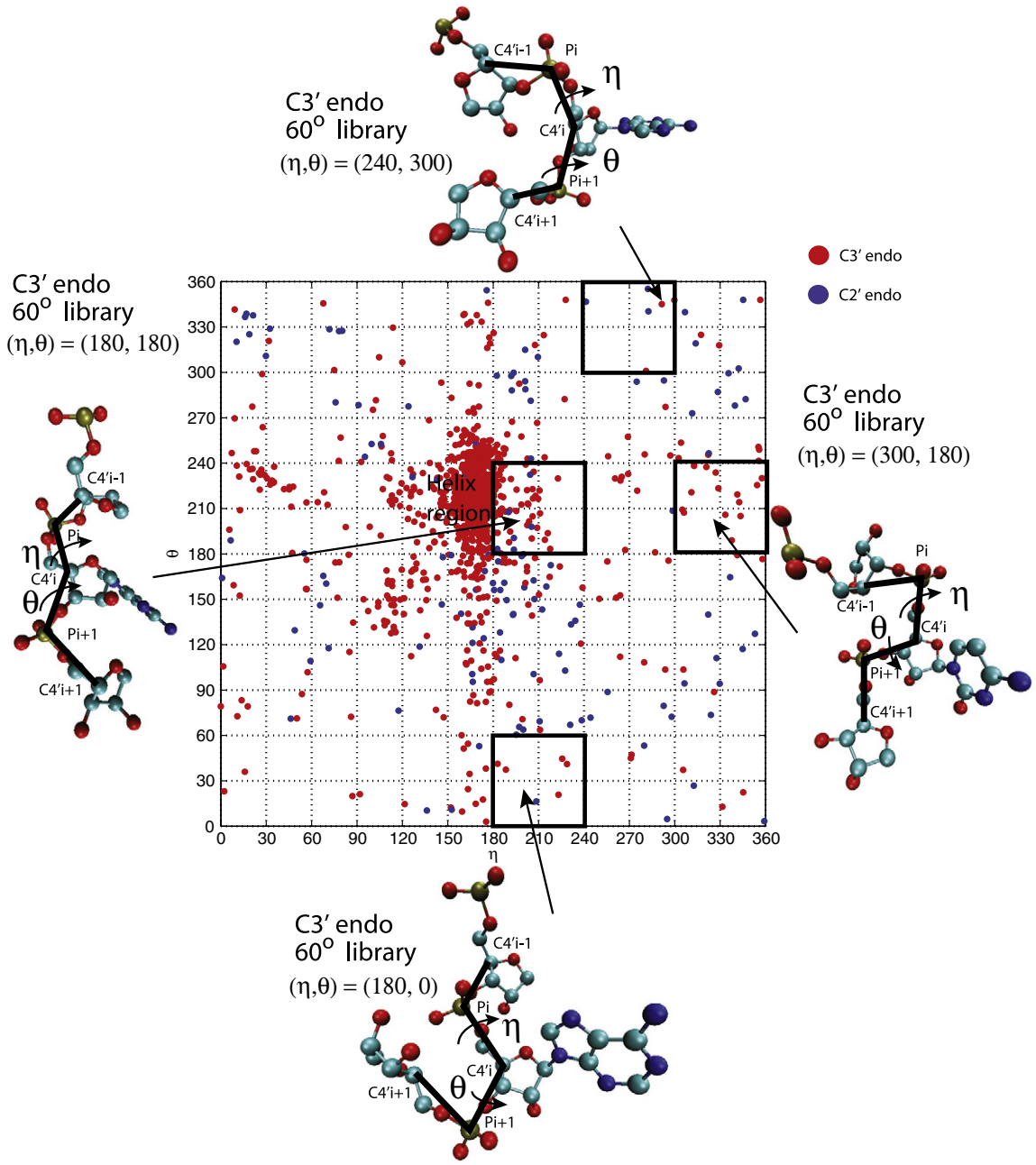

Fig. 1. Pseudo-torsional space of 1434 nucleotides found in the 25 RNA solved structures analyzed in our recent review (PDB entries: 1MFQ, 1LNG, 1MMS, 1MJI, 1DK1, 1F1T, 1KXK, 1I6U, 1MZP, 1OOA, 1RLG, 1S03, 1SJ4, 1XJR, 1U63, 1ZHO, 2GDI, 2GIS, 2HW8, 2IPY, 2OIU, 2OZB, 2PXB, 3D2G, and 3E5C). In the η–θ plot shown at the center, η and θ are the torsion angles for each nucleotide backbone, formed by $C4_{i-1}$-P_i-$C4_i$-P_{i+1} and P_i-$C4_i$-P_{i+1}-$C4_{i+1}$, respectively (see molecular models). The red and blue colors in the (η, θ) plot indicate C3′-*endo* and C2′-*endo* sugar puckers, respectively. Helical regions are located near the center of the plot. For non-helical regions, these angles are distributed diversely (see Fig. 2).

60° (Fig. 2 in the HP work[25]) defines nucleotide conformer sets (or RNA filtered fragments). As a result, six libraries of filtered RNA fragments (for 5°, 10°, 15°, 20°, 30°, and 60° resolution) are created. For each library, structural analysis shows that only a fraction of the possible regions are occupied. For example, for 60° libraries, there are 72 bins, or 6×6=36 fragments for each set of C3′-*endo* or C2′-*endo* conformers; of these, 67 are occupied. For other refinements (30°, 20°, 15°, 10°, and 5°), the corresponding occupied bins to total possible bins (S_B/T_B) are 160/228, 242/648, 296/1152, 402/2592, and 577/10368, respectively. Thus, only a limited number of backbone motifs have been thus far observed,

Please cite this article as: Kim, N. & Schlick, T. A New Toolkit for Modeling RNA from a Pseudo-Torsional Space, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.05.027

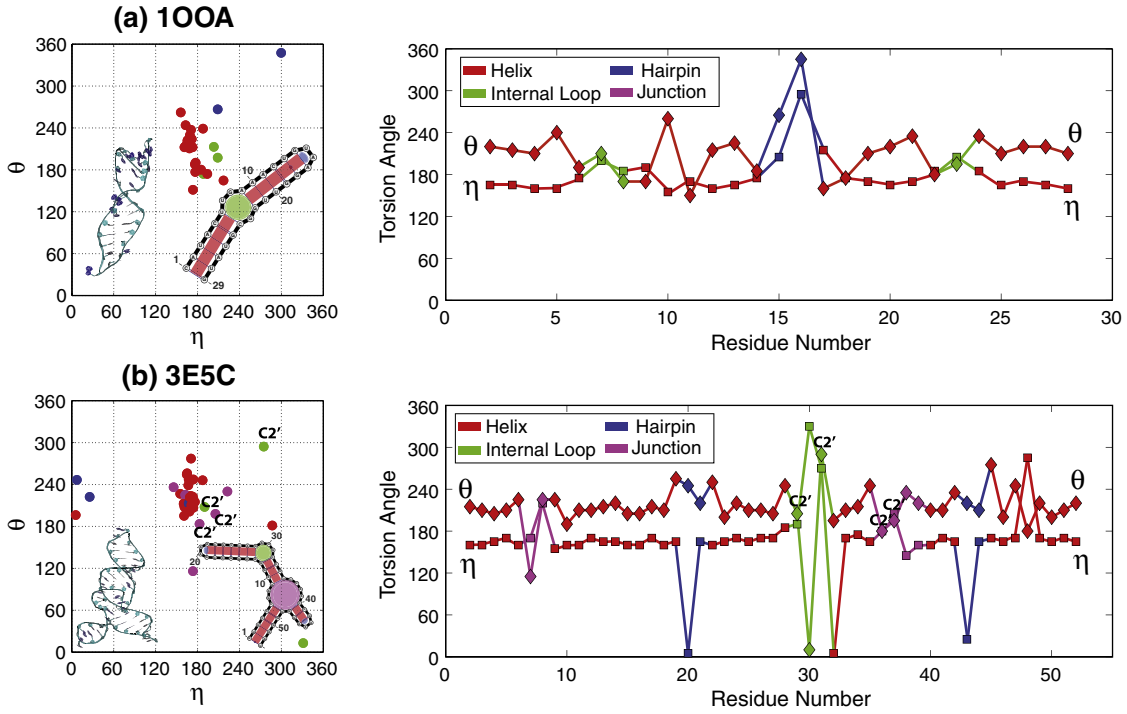

Fig. 2. Pseudo-torsional angles for two RNAs: (a) PDB: 1OOA (NF-κB aptamer) and (B) PDB: 3E5C (SAM riboswitch) using the 5°-binned HP library. On the left, η–θ plots with three- and two-dimensional sketches of each RNA are shown. The red, blue, green, and magenta colors represent helices, hairpins, internal loops, and junctions, respectively. Most nucleotides have C3′-*endo* sugar puckers. Only four nucleotides (residues 29, 31, 36, and 37) in 3E5C have C2′-*endo* sugar puckers, and these are indicated as "C2′". On the right, η (square) and θ (diamond) values for each nucleotide in the 5° library are shown along the residue numbers.

and this binning allows efficient exhaustive enumeration and sampling of RNA backbone structures.

Using these six pseudo-torsional libraries for each nucleotide, the total number of theoretically possible conformations for an *N*-polynucleotide RNA is S_B^N, which can be enumerated (e.g., 67^N or 402^N for 60° or 10° bin libraries, respectively). However, some of these theoretical structures for polynucleotides are not physically possible due to steric clashes and can be further eliminated. HP show that the resulting nucleotide sets cover most of the observed RNA backbone structures: the finer libraries ($B = 20°$, 15°, 10°, and 5°) represent 75–80% of the 8466 backbone conformers developed by the Richardson group[26] within 0.5 Å backbone RMSD for a *single* nucleotide (see Fig. 3 in the HP work[25]).

A second aspect of the HP work is application of this analysis to RNAs of interest, namely, assembling library conformers to build target RNA structures. These reconstructions based on RMSD comparisons are reported to succeed up to a level of 1.5 Å backbone RMSD for RNA molecules. That several hundred fragments may serve as a basis for building models of RNA is significant.

Figure 1 shows the η–θ distribution of 25 representative solved RNAs with diverse structural features and sequence lengths used in our recent review.[9] The figure indicates that each RNA motif can be distinguished by specific pseudo-torsional angles: in helices, these angles correspond to ~180°/180° of the η/θ grid, while in non-helical regions, the η/θ values are distributed diversely. Typical values for different secondary-structure elements can be gleaned from Fig. 2. Figure 2 shows η/θ values for two RNAs [Protein Data Bank (PDB) entries: 1OOA and 3E5C] using the 5° library (the other libraries have similar values). Residues in non-helical regions, such as hairpins (residues 15 and 16 in 1OOA and residues 20 and 43 in 3E5C) and internal loops (residues 29, 30, and 31 in 3E5C), display extreme (high or low) values, far from ~180°, which corresponds to double-stranded regions.

These correspondences between secondary structures and angular values suggest how fragment assembly may be accomplished. For example, PDB entry 1OOA, which has 29 nucleotides and an internal loop, requires at least 7 fragments in the 60° library or 24 fragments in the 5° library to be assembled. PDB entry 3E5C with 53 nucleotides corresponds to 11 fragments in the 60° library or 36 fragments in the 5° library. However, RNA models directly assembled using these fragments are likely to differ from the native structures due to both the discrete nature of the fragment library and the

Please cite this article as: Kim, N. & Schlick, T. A New Toolkit for Modeling RNA from a Pseudo-Torsional Space, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.05.027

flexibility possible between secondary-structure regions. Thus, for general applications to build target RNA structures using backbone fragments in the discrete libraries, a careful fragment selection procedure and/or optimization components may be required.

The present RNA filtered backbone fragment library presents a valuable tool for use in RNA modeling. It also offers a good starting point for RNA structure prediction and design protocols that employ conformational minimization since an initial structure can be assembled using backbone fragments with targeted torsion angles. The RNA structures assembled using the filtered backbone fragment library also have the additional advantage that they need not to be translated back into all-atom structures, unlike those generated via other more abstract coarse-grained models.

As demonstrated by advances in the field of protein modeling that occurred after the introduction of the expanding rotamer libraries,[30,31] such RNA pseudo-torsional fragment libraries are expected to offer a promising avenue for exploring the diverse structural repertoire of the RNA backbone space. The new tool from the Pyle lab is a welcome addition to the field of RNA computational biology. Together with other innovative approaches to modeling, predicting, and designing RNAs, RNA enthusiasts are quickly catching up with analogous protein efforts.

Acknowledgements

This work is supported by the National Science Foundation (DMS-0201160 and CCF-0727001) and the National Institutes of Health (GM100469 and GM081410) awards to T.S.

References

1. Lilley, D. M. (2011). Catalysis by the nucleolytic ribozymes. *Biochem. Soc. Trans.* **39**, 641–646.
2. Breaker, R. R. (2011). Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.
3. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. (2012). Functional complexity and regulation through RNA dynamics. *Nature*, **482**, 322–330.
4. Winkler, W., Nahvi, A. & Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
5. Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A. *et al.* (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747–756.
6. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I. *et al.* (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679–688.
7. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D. *et al.* (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615–623.
8. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
9. Laing, C. & Schlick, T. (2010). Computational approaches to 3D modeling of RNA. *J. Phys.: Condens. Matter*, **22**, 283101–283118.
10. Laing, C. & Schlick, T. (2011). Computational approaches to RNA structure prediction, analysis, and design. *Curr. Opin. Struct. Biol.* **21**, 306–318.
11. Laing, C., Wen, D., Wang, J. T. & Schlick, T. (2012). Predicting coaxial helical stacking in RNA junctions. *Nucleic Acids Res.* **40**, 487–498.
12. Xin, Y., Laing, C., Leontis, N. B. & Schlick, T. (2008). Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA*, **14**, 2465–2477.
13. Leontis, N. B., Lescoute, A. & Westhof, E. (2006). The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16**, 279–287.
14. Leontis, N. B., Altman, R. B., Berman, H. M., Brenner, S. E., Brown, J. W., Engelke, D. R. *et al.* (2006). The RNA Ontology Consortium: an open invitation to the RNA community. *RNA*, **12**, 533–541.
15. Parisien, M. & Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
16. St-Onge, K., Thibault, P., Hamel, S. & Major, F. (2007). Modeling RNA tertiary structure motifs by graph-grammars. *Nucleic Acids Res.* **35**, 1726–1736.
17. Jonikas, M. A., Radmer, R. J. & Altman, R. B. (2009). Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models. *Bioinformatics*, **25**, 3259–3266.
18. Jonikas, M. A., Radmer, R. J., Laederach, A., Das, R., Pearlman, S., Herschlag, D. & Altman, R. B. (2009). Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, **15**, 189–199.
19. Izzo, J. A., Kim, N., Elmetwaly, S. & Schlick, T. (2011). RAG: an update to the RNA-As-Graphs resource. *BMC Bioinformatics*, **12**, 219.
20. Kim, N., Gan, H. H. & Schlick, T. (2007). A computational proposal for designing structured RNA pools for in vitro selection of RNAs. *RNA*, **13**, 478–492.
21. Kim, N., Izzo, J. A., Elmetwaly, S., Gan, H. H. & Schlick, T. (2010). Computational generation and screening of RNA motifs in large nucleotide sequence pools. *Nucleic Acids Res.* **38**, e139.
22. Kim, N., Shiffeldrim, N., Gan, H. H. & Schlick, T. (2004). Candidates for novel RNA topologies. *J. Mol. Biol.* **341**, 1129–1144.
23. Quarta, G., Kim, N., Izzo, J. A. & Schlick, T. (2009). Analysis of riboswitch structure and function by an energy landscape framework. *J. Mol. Biol.* **393**, 993–1003.
24. Das, R. & Baker, D. (2007). Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl Acad. Sci. USA*, **104**, 14664–14669.

Please cite this article as: Kim, N. & Schlick, T. A New Toolkit for Modeling RNA from a Pseudo-Torsional Space, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.05.027

25. Humphris-Narayanan, E. & Pyle, A. M. (2012). Discrete RNA libraries from pseudo-torsional space. *J. Mol. Biol.* http://dx.doi.org/10.1016/j.jmb.2012.03.002.
26. Richardson, J. S., Schneider, B., Murray, L. W., Kapral, G. J., Immormino, R. M., Headd, J. J. *et al.* (2008). RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, **14**, 465–481.
27. Murray, L. J. W., Arendall, W. B., Richardson, D. C. & Richardson, J. S. (2003). RNA backbone is rotameric. *Proc. Natl Acad. Sci. USA*, **100**, 13904–13909.
28. Wadley, L. M., Keating, K. S., Duarte, C. M. & Pyle, A. M. (2007). Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. *J. Mol. Biol.* **372**, 942–957.
29. Keating, K. S. & Pyle, A. M. (2010). Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc. Natl Acad. Sci. USA*, **107**, 8177–8182.
30. Xiang, Z. & Honig, B. (2001). Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.* **311**, 421–430.
31. Peterson, R. W., Dutton, P. L. & Wand, A. J. (2004). Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library. *Protein Sci.* **13**, 735–751.

Namhee Kim
Tamar Schlick
Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA
E-mail address: schlick@nyu.edu.

Please cite this article as: Kim, N. & Schlick, T. A New Toolkit for Modeling RNA from a Pseudo-Torsional Space, *J. Mol. Biol.* (2012), doi:10.1016/j.jmb.2012.05.027