The influence of the local sequence environment on RNA loop structures

RNA (2011) 17: 1247-1257 doi: 10.1261/rna.2550211

CHRISTIAN SCHUDOMA,[1,3] ABDELHALIM LARHLIMI,[2] and DIRK WALTHER[1]
[1]Bioinformatics Group, Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany
[2]Bioinformatics Group, Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam-Golm, Germany

ABSTRACT

RNA folding is assumed to be a hierarchical process. The secondary structure of an RNA molecule, signified by base-pairing and stacking interactions between the paired bases, is formed first. Subsequently, the RNA molecule adopts an energetically favorable three-dimensional conformation in the structural space determined mainly by the rotational degrees of freedom associated with the backbone of regions of unpaired nucleotides (loops). To what extent the backbone conformation of RNA loops also results from interactions within the local sequence context or rather follows global optimization constraints alone has not been addressed yet. Because the majority of base stacking interactions are exerted locally, a critical influence of local sequence on local structure appears plausible. Thus, local loop structure ought to be predictable, at least in part, from the local sequence context alone. To test this hypothesis, we used Random Forests on a nonredundant data set of unpaired nucleotides extracted from 97 X-ray structures from the Protein Data Bank (PDB) to predict discrete backbone angle conformations given by the discretized h/u-pseudo-torsional space. Predictions on balanced sets with four to six conformational classes using local sequence information yielded average accuracies of up to 55%, thus significantly better than expected by chance (17%–25%). Bases close to the central nucleotide appear to be most tightly linked to its conformation. Our results suggest that RNA loop structure does not only depend on long-range base-pairing interactions; instead, it appears that local sequence context exerts a significant influence on the formation of the local loop structure.

Keywords: RNA; 3D structure; structure prediction; Random Forests; machine learning; backbone conformation