RNA structure prediction by knowledge-based statistical potentials and Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE)

David Dufour and Marc A. Marti-Renom

Structural Genomics Laboratory, Centro de Investigación Principe Felipe, Valencia, Spain.

Summary

New RNA structure prediction tools are needed to fast obtaining detailed structural information of new non-coding RNA sequences. Here we propose to use knowledge-based statistical potentials and low-resolution experimental evidences as input to predict RNA structure from sequence. On the one hand, we have derived a series of knowledge-based statistical potentials extracted from the X-ray RNA structures available in the PDB. Such potentials describe the same-atom distances between neighboring residues and between opposed residues of a base pair, as well as the torsion angles defined between them. On the other hand, SHAPE[1] provides information about the secondary and tertiary structure of RNAs that can be translated into spatial structural restraints. Both structural data will be used as inputs in the Integrative Modeling Platform (IMP)[2] to perform simple modeling of RNA with canonical base-pairs. Moreover, we are working towards a classification of SHAPE reactivities based on the different base-pairs present in the RNA so that SHAPE can be used for refining the final RNA structure predicted by IMP.

Methods

Knowledge-based statistical potentials were derived from the RSCB PDB RNAs crystal structures of resolution ≤ 3 Å that contained only one RNA chain. They include each backbone same-atom consecutive nucleotide distance in the same strand as well as in nucleotides of opposite strands of a double helix (**Fig. 1**). The potential of torsion angles between the same-atom of two consecutive base pairs were calculated in order to address the double helix rotation (**Fig. 2**). The atoms included in this study were P, 05', C5', C4', C3' and O3'.

Using the MC-Annotate[3] program we compared the SHAPE data in W. Kladwang et al[4] with the experimentally determined crystal structure of six RNA molecules. The number of stacking interactions and the side of base-pairing interactions were correlated with the reactivity of each residue. These data allowed us to create a predictive model for input SHAPE reactivities.

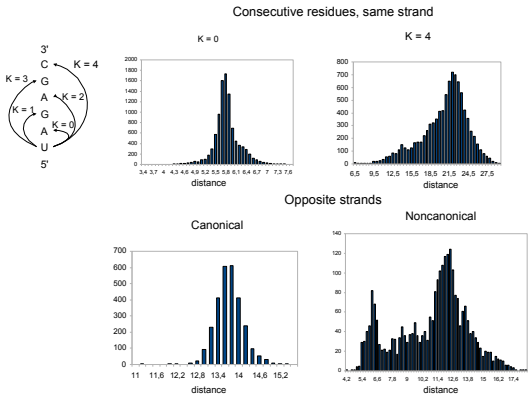

Figure 1. C3'-C3' distance distribution

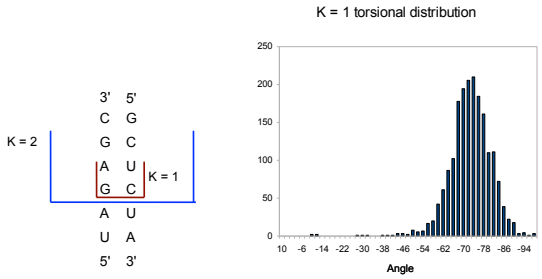

Figure 2. C3'-C3' torsion angle distribution

Results

The distance distributions as well as the torsion angle dihedral distributions showed a sharp profile in general, although the noncanonical interstrand distance distributions had a multimodal profile suggesting different populations that could be separated for further refinement. Nucleotides at each double helix end accounted for the major part of the standard deviation, suggesting a wider conformational space for them. SHAPE data from this experiment had huge standard deviations although separate distributions could be observed.

Conclusions

Statistical potentials obtained so far from canonical interactions can be used for modelling RNA double helices successful. The non-canonical interaction will be the ones that will provide in future the possibility of predicting the RNA structure successfully.

The SHAPE predictive model obtained assigns stacking and base-pairing side probabilities that help IMP to predict more precise RNA interactions, although further refinement is needed in order to be used an everyday tool.

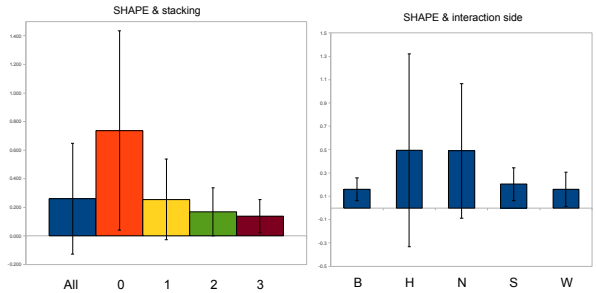

Figure 3. Average and standard deviation of SHAPE rectivities VS stacking and interaction side

References

1. EJ. Merino *et al. Journal of American Chemical Society* **vol 127** (2005) 4223-4231.
2. F. Alber, F. *et al. Nature* ***vol 450*** (2007) 683–694.
3. *P. Gendron et al. J Mol Biol.* ***vol 308(5)*** *(2001) 919-36.*
4. *W. Kladwang et al. Biochemistry* ***vol 50*** *(2011) 8049-56.*