

## Structural bioinformatics

# Integration of accessibility data from structure probing into RNA–RNA interaction prediction

Milad Miladi <sup>1</sup>, Soheila Montaseri<sup>1</sup>, Rolf Backofen<sup>1,2,\*</sup> and Martin Raden <sup>1,\*</sup>

<sup>1</sup>Department of Computer Science, Bioinformatics Group, University of Freiburg, Freiburg D-79110, Germany and

<sup>2</sup>Center for Biological Signaling Studies (BIOSS), University of Freiburg, Freiburg D-79104, Germany

\*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on July 12, 2018; revised on November 20, 2018; editorial decision on December 8, 2018; accepted on December 18, 2018

## Abstract

**Summary:** Experimental structure probing data has been shown to improve thermodynamics-based RNA secondary structure prediction. To this end, chemical reactivity information (as provided e.g. by SHAPE) is incorporated, which encodes whether or not individual nucleotides are involved in intra-molecular structure. Since inter-molecular RNA–RNA interactions are often confined to unpaired RNA regions, SHAPE data is even more promising to improve interaction prediction. Here, we show how such experimental data can be incorporated seamlessly into accessibility-based RNA–RNA interaction prediction approaches, as implemented in IntaRNA. This is possible via the computation and use of unpaired probabilities that incorporate the structure probing information. We show that experimental SHAPE data can significantly improve RNA–RNA interaction prediction. We evaluate our approach by investigating interactions of a spliceosomal U1 snRNA transcript with its target splice sites. When SHAPE data is incorporated, known target sites are predicted with increased precision and specificity.

**Availability and implementation:** <https://github.com/BackofenLab/IntaRNA>

**Contact:** [backofen@informatik.uni-freiburg.de](mailto:backofen@informatik.uni-freiburg.de) or [mmann@informatik.uni-freiburg.de](mailto:mmann@informatik.uni-freiburg.de)

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

The function of many if not most non-coding (nc)RNA molecules is to act as platforms for inter-molecular interaction, which depends on their structure and sequence. A large number of ncRNAs regulate their target RNA molecules via base-pairing. For instance, small (s)RNAs regulate the translation of their target genes by direct RNA–RNA interactions with the respective messenger (m)RNAs (Backofen and Hess, 2010). To predict such interactions, regions not involved in intra-molecular base pairing have to be identified. This ‘free-to-interact’ potential of a region, i.e. its unpaired probability, is computed by assessing the fraction of structures where the region is free (unpaired) within the overall structure ensemble of an RNA [(Raden *et al.*, 2018) for a detailed introduction]. The probabilities are used by state-of-the-art prediction tools like IntaRNA (Mann *et al.*, 2017) to account for the regions’ accessibility. While correct within their thermodynamic models, such probabilities do

not incorporate all cellular constraints and dynamics that define accessible regions and thus the likelihood for interaction.

The accuracy of RNA structure prediction improves when experimental structure probing data, such as SHAPE (Wilkinson *et al.*, 2006), is incorporated (Hajdin *et al.*, 2013). This is done by converting the chemically sensed reactivity values to pseudo-energy terms. Pseudo-energies are combined with structure formation energies from thermodynamic models, that are used for RNA structure prediction (Lorenz *et al.*, 2016; Montaseri *et al.*, 2017; Spasic *et al.*, 2018). As SHAPE [For simplicity, SHAPE refers to any structure probing experiment (e.g. SHAPE, DMS)] reactivity is associated with the accessibility of nucleotides, the use of such experimental data is even more promising to improve the accuracy of RNA–RNA interaction prediction methods. For that reason, we introduce a seamless incorporation of SHAPE data into accessibility-based prediction approaches such as IntaRNA.

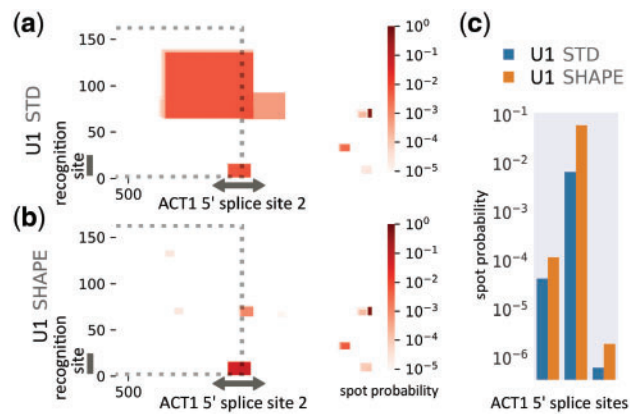
Recently, structure probing has been complemented by next-generation sequencing techniques to efficiently obtain transcriptome-wide reactivity information (Choudhary *et al.*, 2017; Kutchko and Laederach, 2017). This produces large datasets that demand for fast methods incorporating the probing data, which is met by our extension of IntaRNA introduced in the following.

For a given RNA–RNA interaction  $I$  (see [Supplementary Material](#) for detailed formalisms), its accessibility-based free energy is defined by  $E(I) = E^{\text{hyb}}(I) + ED^1(I) + ED^2(I)$ . Therein,  $E^{\text{hyb}}(I)$  provides the hybridization energy from intermolecular base pairing while the  $ED^{1,2}$  terms represent the energy (penalty) needed to make the respective interacting subsequences unpaired/accessible (Mückstein *et al.*, 2006). ED terms are defined by unpaired probabilities  $Pr^{\text{ss}}$  of the subsequences via  $ED(I) = -RT \log(Pr^{\text{ss}})$ , where  $R$  is the gas constant and  $T$  the temperature. Detailed introductions on ED computation are provided e.g. in (Raden *et al.*, 2018; Wright *et al.*, 2018). Computation of unpaired probabilities can be guided by SHAPE data (Lorenz *et al.*, 2016). While SHAPE-guided energy evaluations cannot be compared to unconstrained energy values (due to the introduced pseudo-energy terms), unpaired probabilities are compatible, since they are reflecting the accessible structure space rather than individual structures. Thus, SHAPE-constrained  $Pr^{\text{ss}}_{\text{SHAPE}}$  values can be directly used for ED and thus  $E$  computation while preserving comparability of the resulting energies.

Now, we show that SHAPE-guided accessibility prediction improves RNA–RNA interaction prediction. To this end, we study the probabilities of U1 small nuclear (sn)RNA interacting with pre-mRNA target sites, which is an established example of intermolecular RNA interaction essential for RNA-splicing in eukaryotes. U1 is involved in pre-mRNA splicing by recognizing the 5' site of introns via inter-molecular base pairing (Hertel and Graveley, 2005). Due to the dynamics and constraints imposed by the spliceosome, it is generally challenging to avoid false positive interaction predictions, which are either predictions of U1's recognition site with random regions of the mRNA or predicted interactions of other U1 accessible regions with the mRNA. For that reason, we used U1 as an example to show that *in vivo* probing data effectively reduces false positive predictions in RNA–RNA interaction prediction.

SHAPE data for U1 was obtained from *in vivo* DMS-seq RNA structure probing of *Arabidopsis thaliana* (Ding *et al.*, 2014). We selected the U1 homolog transcript bearing the largest secondary structure distance between the unconstrained and SHAPE-constrained structure prediction. The pre-mRNA sequences for five genes were extracted, which have been previously validated to perform U1-dependent splicing (Yeh *et al.*, 2017). [Figure 1a and b](#) exemplify the effect of SHAPE-constrained predictions for ACT1 mRNA. Without SHAPE constraints, the splice site is predicted to interact with various regions of U1 with high probability; see [Supplementary Material](#) for formalisms. In contrast, the splice site's interaction with U1's recognition site is dominant with high specificity in SHAPE-guided IntaRNA predictions. Furthermore, SHAPE-guided prediction has a higher precision. Among all predicted interactions of U1 with the ACT1 mRNA, the ranking of the known interaction is improved from 9 to 3 in SHAPE-guided mode. When investigating the accessibility profile of U1 ([Supplementary Material](#)), this mainly results from an increased SHAPE-guided unpaired probability  $Pr^{\text{ss}}_{\text{SHAPE}}$  of U1's recognition site and thus reduced ED penalties.

The interaction probability of U1's recognition site with all three 5' splice sites of ACT1's coding sequence are increased when SHAPE data are incorporated ([Fig. 1c](#)). This effect results from a decreased number of false positive predictions ([Fig. 1a and b](#)).



**Fig. 1.** RNA–RNA interaction prediction between spliceosomal RNA U1 with ACT1 mRNA of *A. thaliana*. Interaction probabilities predicted between U1 (y-axis) and the region around the second intron splice site of ACT1 coding sequence mRNA using (a) unconstrained (STD) and (b) SHAPE-constrained accessibility estimates for U1. The dotted lines enclose U1 interactions with exon 2. (c) Spot probabilities of U1 recognition site (spot index = 8) interacting with the 5' splice sites of ACT1 (spot = 1st intron index), with SHAPE constraints (orange) and without (blue) (Color version of this figure is available at [Bioinformatics online](#).)

Following this trend, the probabilities of splice site recognition are improved among all the mRNAs and are on average 3.08 times higher (SHAPE/STD). Further details about data, evaluation and analyses are provided in the [Supplementary Material](#).

## Funding

This work was supported by Bundesministerium für Bildung und Forschung [031A538A RBC, 031L0106B] and Deutsche Forschungsgemeinschaft [BA 2168/14-1, BA 2168/16-1].

*Conflict of Interest:* none declared.

## Acknowledgement

The authors thank Dr. Ronny Lorenz for discussions on SHAPE integration.

## References

- Backofen, R. and Hess, W.R. (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biol.*, **7**, 33–42.
- Choudhary, K. *et al.* (2017) Comparative and integrative analysis of RNA structural profiling data: current practices and emerging questions. *Quant. Biol.*, **5**, 3–24.
- Ding, Y. *et al.* (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696.
- Hajdin, C.E. *et al.* (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. USA*, **110**, 5498–5503.
- Hertel, K.J. and Graveley, B.R. (2005) RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem. Sci.*, **30**, 115–118.
- Kutchko, K.M. and Laederach, A. (2017) Transcending the prediction paradigm: novel applications of SHAPE to RNA function and evolution. *Wiley Interdiscip. Rev. RNA*, **8**, e1374.
- Lorenz, R. *et al.* (2016) SHAPE directed RNA folding. *Bioinformatics*, **32**, 145–147.
- Mann, M. *et al.* (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *NAR*, **45**, W435–W439.
- Montaseri, S. *et al.* (2017) Evaluating the quality of SHAPE data simulated by k-mers for RNA structure prediction. *J. Bioinform. Comput. Biol.*, **15**, 1750023.

- Mückstein,U. *et al.* (2006) Thermodynamics of RNA-RNA binding. *Bioinformatics*, **22**, 1177.
- Raden,M. *et al.* (2018) Interactive implementations of RNA structure and RNA-RNA interaction prediction approaches for example-driven teaching. *PLoS Comput. Biol.*, **14**, e1006341.
- Spasic,A. *et al.* (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.*, **46**, 314–323.
- Wilkinson,K.A. *et al.* (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (Shape): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610.
- Wright,P.R. *et al.* (2018) Structure and interaction prediction in prokaryotic RNA biology. *Microbiol. Spectr.*, **6**.
- Yeh,C.-S. *et al.* (2017) The conserved AU dinucleotide at the 5' end of nascent U1 snRNA is optimized for the interaction with nuclear cap-binding-complex. *Nucleic Acids Res.*, **45**, 9679–9693.