Quantitative Biology 2020, 8(2): 109–118
https://doi.org/10.1007/s40484-020-0205-6

REVIEW

Prediction and differential analysis of RNA secondary structure

Bo Yu[1,2], Yao Lu[3], Qiangfeng Cliff Zhang[3,4,5], Lin Hou[1,2,4,*]

[1] Center for Statistical Science, Tsinghua University, Beijing 100084, China
[2] Department of Industrial Engineering, Tsinghua University, Beijing 100084, China
[3] School of Life Sciences, Tsinghua University, Beijing 100084, China
[4] MOE Key Laboratory of Bioinformatics, Tsinghua University, Beijing 100084, China
[5] Beijing Advanced Innovation Center for Structural Biology, Center for Synthetic and Systems Biology, Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China
* Correspondence: houl@tsinghua.edu.cn

Received December 6, 2019; Revised March 10, 2020; Accepted March 16, 2020

Background: **RNA structure is the crucial basis for RNA function in various cellular processes. Over the last decade, high throughput structure profiling (SP) experiments have brought enormous insight into RNA secondary structure.**
Results: **In this review, we first provide an overview of approaches for RNA secondary structure prediction, including free energy-based algorithms and comparative sequence analysis. Then we introduce SP technologies, databases to document SP data, and pipelines/algorithms to normalize and interpret SP data. Computational frameworks that incorporate SP data in RNA secondary structure prediction are also presented.**
Conclusions: **We finally discuss potential directions for improvement in the prediction and differential analysis of RNA secondary structure.**

Keywords: RNA secondary structure; prediction; differential analysis; structure profiling

Author summary: High throughput structure profiling (SP) experiments help the analysis of RNA secondary structure. In this review, we discuss existing frameworks for the prediction and differential analysis of RNA secondary structure, including computational methods and especially approaches incorporating SP data.

INTRODUCTION

Although primarily known for transmitting information from DNA to proteins, RNA has been discovered to play multiple important roles in various cellular processes in recent years [1–4]. To deepen understanding of the underlying mechanisms of multiple roles of RNA in cellular processes, it is crucial to outline the relationship between RNA structure and function, which is generally provided based on prediction and differential analysis of RNA secondary structure [5–7].

RNA secondary structure is formed by complementary base pairing, which exhibits dynamics in many aspects [3,8]. First, RNA secondary structure is flexible and dynamic even under the same environment. Second, many RNAs have been reported to adopt and convert between several secondary structures, thus different conformations of secondary structure may coexist in cellular context. Third, the structural landscape of RNA can change in response to experimental stimuli. The change may take place in the composition of different conformations, or in the adoption of new structures. These dynamics collectively contribute to the complexity of the depiction of the structural landscapes.

The dynamics of RNA secondary structure is extensive. Consequently, the combinatory landscape cannot be fully resolved by traditional experimental techniques, such as X-ray and cryogenic electron microscopy. Despite their high accuracy, traditional techniques are generally labor intensive. More importantly, the ability to conduct *in vivo* measurements is lacking, further limiting the application of traditional techniques in resolving RNA secondary

structure [7,9]. To overcome these difficulties, computational methods have been developed to predict RNA folding based on free energy or comparative sequence analysis. More recently, *in vivo* structure profiling (SP) experiments and computational tools to analyze such data have brought new insight into the problem.

While current computational approaches mainly focus on the prediction of RNA secondary structure, there is a growing demand for differential analysis of RNA secondary structure, *i.e.*, identifying regions where the structural landscape of an RNA differs between different conditions [10–13]. Applications of differential analysis of RNA secondary structure have shed new light on many related problems, such as RNA-protein interactions and RNA structural patterns [11,14,15].

The rest of the article is organized as follows. In Section of "Computational methods for RNA secondary structure prediction", we overview state of art computational methods for RNA secondary structure prediction. In Section of "Data-driven RNA secondary structure prediction", we overview the new generation of SP experiments as well as the corresponding databases and briefly introduce the analysis of SP data. Approaches combining computational methods and SP data for RNA secondary structure prediction will also be reviewed in this section. Lastly, we will review the differential analysis of RNA secondary structure in Section of "Discussion" . Potential directions for improvement in the prediction and differential analysis of RNA secondary structure will also be discussed in this section.

COMPUTATIONAL METHODS FOR RNA SECONDARY STRUCTURE PREDICTION

Due to the limitations of traditional techniques for measuring RNA secondary structure, for many years, computational methods for RNA secondary structure prediction are the primary origin of our knowledge on RNA structure and the corresponding potential functions for most RNAs. These computational methods are commonly based on thermodynamic models that calculate free energy of RNA secondary structure. An alternative approach, comparative sequence analysis, predicts RNA secondary structure by borrowing information from homologous RNA sequences. The computational approaches have been comprehensively reviewed elsewhere [16,17]. Here we briefly introduce several representative algorithms and benchmarking of computational methods. Combination of computational methods and recent SP data will be discussed in later sections.

Free energy based algorithms

Under thermodynamical assumptions, the problem of RNA folding can be transformed into optimization problems based on free energy. To elaborate, the probability of observing an RNA secondary structure s at the Boltzmann equilibrium, $p(s)$, with the free energy of s, $\Delta G(s)$ is specified as follows:

$$p(s) \propto \exp\{\Delta G(s)/RT\},$$

where R,T are known physical constants [18]. Through this law, we can access the probability of observing a specific structure by estimating its free energy, which is the sum of free energies of its substructures. The calculation depends on experimentally measured parameters [19,20]. An in-depth description of free energy calculation can be found in Zuker and Stiegler [20].

Free energy minimization (MFE)

MFE is a pioneering computational algorithm that predicts RNA secondary structure through free energy minimization [20]. According to the above thermodynamic statistical law, by minimizing free energy, MFE identifies the most likely structure. Specifically, the number of possible secondary structures grows exponentially with RNA sequence length (denoted by L), posing a severe computational challenge. Zuker *et al.* devised a dynamic programming algorithm for the optimization problem, reducing the time and space complexity to $O(L^3)$ and $O(L^2)$, respectively. This algorithm provides a broadly applicable computational tool for RNA secondary structure prediction for the first time, through which RNA secondary structures are explored in batches.

However, cases of unsatisfactory prediction accuracy of MFE have been reported [21], probably due to simplifications in the calculation of free energy, inaccuracies of the experimentally measured parameters, and neglected effects of ions and other molecules. To account for such complexity, several algorithms adapted MFE to predict a list of suboptimal structures [19,22,23]. The length of such list of suboptimal structures, however, often grows exponentially with the length of RNA sequence, which is computationally extensive and brings greater difficulty to experimental validation.

Sfold

Instead of searching for structures with low free energy, the Sfold algorithm employs a sampling scheme in secondary structure prediction [24–26] that samples most representative structures from the Boltzmann ensemble. Specifically, the algorithm takes two steps. The forward step calculates the Boltzmann equilibrium probability for a secondary structure, while the key technique is the recursive calculation of equilibrium partition functions. The backward step samples structures from the Boltz-

mann distribution in a recursive fashion based on the sampling probabilities derived in the forward step. The algorithm has time and space complexity of $O(L^3)$ and $O(L^2)$, respectively. Next Sfold groups the sampled structures into clusters and identifies the centroid of each cluster. These centroids are thus representative of the ensemble and are supposed to approximate the true structures.

Prediction with the centroids has been reported to achieve higher accuracy on some RNAs compared to MFE [24]. Moreover, the number of clusters of sampled structures is usually small, which is more practical for experimental validation.

Expected accuracy maximization (MEA)

As another popular computational method, MEA predicts RNA secondary structure by maximizing expected accuracy of base pairing prediction [27]. Here, the expected accuracy of a structure s measures the consistency between s and the marginal pairing probabilities of the Boltzmann ensemble. Specifically,

$$\text{Expected Accuracy of } s = \sum_{(i,j)\in s} 2\gamma p_{ij} + \sum_{i\in s} q_i,$$

where

$$p_{ij} = \Pr(\text{nucleotide } i \text{ is paired with mucleotide } j),$$

$$q_i = 1 - \sum\nolimits_j p_{ij} = \Pr(\text{nucleotide } i \text{ is unpaired}),$$

and γ is a scaling factor that balances sensitivity and specificity in evaluating prediction accuracy. A parsing algorithm is developed to solve the optimization problem with time and space complexity of $O(L^3)$ and $O(L^2)$, respectively, which are the same as those of MFE and Sfold.

To address the computational difficulty to calculate the expected accuracy, pseudo-expected accuracy is developed as an approximation, which can be optimized via stochastic sampling. The pseudo-expected accuracy serves as a good alternative under various measures of prediction accuracy [28].

Comparative sequence analysis

Comparative sequence analysis often achieves higher prediction accuracy compared to free-energy based algorithms [29]. Comparative sequence analysis leverages the fact that RNA sequence and structure are closely related to RNA function and thus are conserved during evolution [30]. By integrating evolutionary information, a consensus structure from alignment of homologous sequences across different species is predicted. Some variants further incorporate other information including free energy of RNA structure to improve prediction accuracy [31–33]. Extensive reviews for tools for comparative sequence analysis can be found in [16,17,30]. Yet, several complications greatly limit the scope of application of these tools. In general, enough homologous sequences are prerequisites to guarantee prediction accuracy, which seldom happens. In addition, laborious collection of homologous sequences and computational challenges in multiple sequence alignment also put limits on the application of this approach. Moreover, similar to free-energy based algorithms, it remains unclear how to extend the current framework of comparative sequence analysis to construct structural landscapes.

Benchmarking of computational methods for RNA secondary structure prediction

Systematic benchmarking is necessary to evaluate the accuracy of RNA secondary structure prediction algorithms. Benchmark datasets haven been constructed by collecting experimentally solved RNA secondary structures from databases Protein Data Bank [34] and RNAstrand [35]. Hajiaghayi *et al.* [36] have evaluated the accuracy of energy-based prediction algorithms, and concluded that improvement in thermodynamic parameters can greatly increase prediction accuracy, and relative performance of methods can be affected by the parameter set used for evaluation. In general, average F measure of energy-based methods lies between 0.6 and 0.7, depending on the RNA classes being considered. Puton *et al.* have implemented a web server, CompaRNA, for continuous benchmarking of computational methods [29]. Rankings of computational methods for RNA secondary structure prediction under various scenarios are systematically reported, including free energy based algorithms and approaches to comparative sequence analysis.

It is noteworthy that accuracies and rankings of computational methods can be strongly influenced by several factors, including the length of RNA sequence, RNA classes, and the existence of pseudoknots [29,37]. Moreover, there are several issues for the use of the benchmarks. First, it is hard to guarantee the separation of training data for existing computational methods and reference structures for benchmarking [29,38]. Second, in general the benchmarks are designed to evaluate single structure prediction [29,36,37]. For computational methods predicting multiple structures, one of the structures is properly selected for evaluation [29]. Third, the length of reference RNA sequences in benchmark datasets is limited by the computational burden for comparing computational methods [29]. Thus it is suggested the

rankings of methods should only be considered in specific context.

DATA-DRIVEN RNA SECONDARY STRUCTURE PREDICTION

Despite the successful prediction on a portion of RNAs, computational methods are hampered by their limitations, as described in Section of "Computational methods for RNA secondary structure prediction", from a wider application range and higher prediction accuracy. Moreover, as a common defect of computational methods, the prediction relies on the sequence information only and hence cannot be adapted to make predictions in varied conditions.

Recent advancements in SP experiments can help to overcome these difficulties. SP experiments utilize chemicals or enzymes to probe RNA structure. The enzymes act differently to RNA nucleotides depending on their pairing status, through which local structural characteristics are decoded into profiling data by tracking the footprints of the additives. The history of SP experiments to be used for RNA secondary structure profiling can be dated back to the 1970s [39,40]. In recent years, with the development of experimental technology, SP experiments have entered a new generation, from low resolution, low-throughput, *in vitro* experiments, to high resolution (nucleotide level), high-throughput experiments that are applicable under different conditions, such as various *in vivo* environments. Meanwhile, it is worth noting that profiling data from SP experiments (SP data) reflects local structural characteristics, and thus behaves stable when the length of RNA increases. This is valuable considering the instability of computational methods for long RNAs. SP experiments have become powerful and cost-effective tools for transcriptome-wide structure profiling under different conditions, thus are also promising for the study of relationship between RNA structure and function [5,7,14,41].

High-throughput SP experiments and databases

In this section, we overview SP experiments and the databases of SP data.

SP experiments

Once RNAs fold *in vivo* or *in vitro*, specific RNase can selectively cleave either unpaired regions or paired regions, generating corresponding cleavage ends. Combining such RNases with high-throughput sequencing technology, parallel analysis of RNA structures (PARS) and several alternative approaches have the capacity to interrogate RNA secondary structure at the transcriptome level [42–45].

However, approaches utilizing RNases are limited to *in vitro* structure profiling, since in general nucleases cannot penetrate cell membrane. An alternative approach is chemical modification, which enables RNA modification *in vivo*. Given the smaller molecular size of chemical probes [14,46–67], chemical modification approaches also have higher resolution than RNases-based approaches. DMS-seq, for example, utilizes dimethyl sulphate (DMS), which specifically reacts with unpaired nucleotides, for genome-wide structure profiling *in vivo* [60]. However, such reaction is limited to adenine and cytosine residues, providing only a partial view of RNA secondary structure. Alternatively, selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) can be used as the reagent. *In vivo* click selective 2′-hydroxyl acylation and profiling experiment (icSHAPE) [14,52,53] can probe unpaired nucleotides *in vivo* for all bases.

It's noteworthy that different chemical probes have distinct chemical reactivities, and probe combinations are suggested under certain conditions [41]. Thorough discussions on experimental details of SP experiments can be found in [7,41,68–71].

Databases

With the rapid development of SP experiments, numerous SP data of different types have been generated. There are comprehensive databases that systematically archive and manage SP data [72–77]. The databases differ by the type of SP experiments, the types of RNAs, and the source species. Moreover, some databases provide visualization tools and facilitate application of computational methods. Details of the databases are summarized in Table 1.

SP reactivity estimation

While a large amount of high resolution SP data have been generated at transcriptome level, estimation of nucleotide reactivities from SP data has proven challenging [15,78]. The reactivity of a nucleotide measures the extent of the nucleotide reacting to the probing enzymes and thus infers the pairing status of the nucleotide. The inference procedure is complicated by several factors. First, multiple secondary structures of an RNA can coexist, thus the observed reactivity of a nucleotide reflects the combination of pairing status in a mixture of secondary structures. Deconvolution of the reactivity measures is computationally challenging. Second, due to technical limitations of SP experiments, SP data are usually very noisy [15,79]. These collectively make the inference of RNA secondary structure very difficult. To account for the confounding factors, various approaches have been developed in the last decade. In particular,

Table 1 Databases of structure probing data

Database	Probed RNAs	SP experiments	Species	Prediction methods	Website	Reference
RMDB	Single RNAs (riboswitches, tRNAs, ribozyme, ribosomal domains and human-designed sequences)	Diverse experiments	Multiple species and artificial RNAs	RNAstructure package	https://rmdb.stanford.edu/	[72–74]
Structure surfer	Transcriptome-wide	*In vivo* and *in vitro*, icSHAPE, DMS-Seq, PARS and ds/ssRNA-Seq	*Homo sapiens*, *Mus musculus*	NA	http://tesla.pcbi.upenn.edu/strucuturesurfer/	[75]
FoldAtlas	Genome-wide	*In vivo*, DMS and structure-seq	*Arabidopsis thaliana*	RNAstructure package	http://www.foldatlas.com/	[76]
RNAex	Genome-wide	*In vivo* and *in vitro*, diverse experiments	*Homo sapiens*, *Mus musculus*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*	RME, SeqFold, RNAstructure package and RNAfold	http://rnaex.ncrnalab.org/	[77]

computational effort roughly falls into two classes, heuristic approaches and statistical model-based approaches.

Heuristic approaches

Previous to interpreting data from SP experiments, substantial preprocessing, or normalization, is needed, and the approach is usually tailored to specific experiment protocols. For example, PARS first normalizes SP counts to account for different sequencing depth and then derives reactivities as the log2 ratio between the normalized counts of RNase V1 and of RNase S1 [45]. Here, RNase V1 and RNase S1 are enzymes that preferentially cleave paired and unpaired regions, respectively. Another SP experiment, Structure-seq [65], first normalizes natural logs of SP counts by transcript length and abundances in case experiment and control experiment separately, and then subtracts the normalized counts in control experiment from those in case experiment as raw reactivity measurements. Finally, 2%/8% normalization technique [80] and a reactivity-capping procedure are sequentially applied to the raw reactivities to obtain final reactivity estimations [65].

The normalization steps are customized to specific experimental procedures in SP experiments, hence they seldom directly apply to data generated by other SP experiments. Nevertheless, SP experiments share a set of core experimental procedures [7,41,71], therefore general pipelines of reactivity estimation for various SP data have been proposed, which provided integrated computational framework for reads mapping, background correction, and reactivity derivation [78,81].

Statistical model-based approaches

The output of normalization approaches is the estimation of reactivity, and it is straightforward to determine RNA secondary structures by thresholding. However, model-based approaches often yield improved estimations and can be adapted to more complicated situations.

Aviran *et al.* have devised a rigorous probabilistic model to describe the fragment distribution in SHAPE-Seq (*selective 2'-hydroxyl acylation analyzed by primer extension sequencing*) experiment, which accounts for the effect of chemical modification and natural polymerase drop-off [82]. In contrast to heuristic approaches that rely on expert knowledge, the structure inference is automatic through maximum likelihood estimation. Extending the strategy of generative models, another approach, PROBer, has been developed to simultaneously model reactivities of multiple RNAs in a Bayesian framework [83]. Specifically, reactivities, as well as noise parameters, of multiple RNA isoforms are set to follow a common prior distribution, and maximum posterior estimates are further employed to incorporate the generative model of SP data. The Bayesian framework of PROBer not only greatly reduces the number of parameters, but also enables borrowing information across RNA transcripts, which is important in the estimation of reactivities considering the high noise level and limited number of replicates in SP data.

Unlike generative modeling approaches, BUM-HMM (Beta-Uniform Mixture Hidden Markov Model) first calculated empirical P values for treatment-control LDR (log-ratio of drop-off rates) and then made structure inference by modeling the distribution of the P values

[79]. In detail, an empirical null distribution of the LDR was constructed to account for the variability of replicates in SP data. By comparing the observed LDRs, log-ratios of drop-off rates of case experiments with those of control experiments, to the empirical null distribution, quantitative differences between case and control experiments are transferred into *P* values. Next, BUM-HMM utilizes a hidden Markov model to leverage the dependency of RNA secondary structure between neighboring nucleotides. Hidden states in the model correspond to the status of modification. Several confounding factors, including coverage biases and sequence biases, are explicitly accounted for in the model. It is worth noticing that these factors, along with the way BUM-HMM deals with them, have general implications for the analysis of SP data beyond the estimation of reactivities. The idea to model structural dependencies in adjacent nucleotides has also been exploited in JPGM, a statistical model-based approach to interpret RNase footprint sequencing data [84].

RNA secondary structure prediction integrating SP data

Free-energy based algorithms and comparative sequence analysis are based on nucleotide sequences, while SP experiment directly interrogates the RNA secondary structure by chemical probing regardless of the sequence context. Thus, these approaches provide complementary information, and it is reasonable to hypothesize that combining SP data with computational methods could potentially improve RNA secondary structure prediction.

To leverage on SP data, several approaches use reactivities as constraints to guide RNA folding. For example, Deigan *et al*. have developed an approach to combine the MFE algorithm with reactivities measurements from SHAPE experiments [85]. Instead of using hard thresholding, *i.e.*, forcing nucleotides with high (low) reactivities to be unpaired (paired) [86], reactivities are incorporated as pseudo free energy terms for paired nucleotides. Specifically, the pseudo free energy term of nucleotide i is defined as

$$\Delta G_i = m\ln(1 + \text{reactivity of nucleotide } i+1) + b,$$

where m and b are parameters that are pre-trained on datasets with reference structures. By adding pseudo free energy terms to the original energy function of MFE, structures that assign nucleotides with high reactivities as paired are penalized. In consequence, RNA folding is guided to match the structural information reflected by reactivities. This approach has been shown to improve prediction accuracy compared to MFE [85]. However, the choice of parameters (m,b) for the whole transcriptome reflects the contribution of SP data, and it is not straightforward to specify their values due to their nonphysical nature [87]. Moreover, it remains an open problem how to decide the format of pseudo free energy terms [88–90].

Similarly, SeqFold extended the Sfold algorithm to combine diverse types of SP data as a structure preference profile [91]. After grouping structures sampled by Sfold into clusters, SeqFold assigns the structure preference profile into the most likely cluster and makes prediction with centroid of the cluster. This approach is robust and is flexible to accommodate a variety of SP data. In addition to integrate SP data, Rsample predicts multiple RNA structures, and provides estimates of the mixing proportions [89]. Particularly, it allows more flexible options for RNA secondary structure prediction, either to predict a single structure with the MEA algorithm or to predict multiple representative structures with the Sfold algorithm.

Methods that combine SP experiment with free energy based prediction of RNA secondary structure have achieved comparable and even better performance compared to existing methods [88,90]. While integrating SP data as reactivities is usually convenient to implement, compressing SP data into reactivities can lead to loss of information. Alternatively, SLEQ (Structure Landscape Explorer and Quantifier) incorporates SP data at the reads level [92]. Moreover, structural landscapes, that is, different conformations of secondary structure along with their relative abundances, are considered in this approach. In SLEQ, candidate structures are linked with patterns of sequencing reads through a generative model, where candidate structures can be flexibly selected based on prior knowledge, such as results from various computational methods. Then the approach maximizes the agreement between candidate structures and patterns of sequencing reads to reconstruct the structural landscape. Remarkably, SLEQ alleviates the dependence on thermodynamic models, allowing for versatile RNA secondary structure modeling. Nevertheless, the method enforces stringent constraints on completeness of candidate structures, *i.e.*, all structures in the landscape are required to be included in the list of candidate structures.

Some of the above approaches that combine computational methods with SP data have been integrated into databases listed in Table 1.

DISCUSSION

RNA secondary structure prediction has been a classical problem in computational biology. However, with the rapid development of SP experiments, the problem has attracted new attention in the last decade. Emerging approaches that combine SP data with conventional algorithms have been shown to improve prediction

accuracy. However, existing approaches mainly focus on identification of single optimized structure, leaving deciphering the comprehensive structural landscapes, *i.e.*, estimating the content and abundance of multiple co-existing structures, open problems.

In addition to RNA secondary structure prediction, SP data can also facilitate comparison of RNA secondary structure in different conditions, which brings insight regarding the conversion of RNA structure and their cellular roles [10–15]. Recently, dStruct proposed a statistical framework to measure dissimilarity of SP data across conditions and to screen for regions with significant difference [13]. Undoubtfully, the identification of regions with structure change under different conditions brings important functional implications. More effort is needed in this line of work.

Last, we highlight a few emerging themes in analysis of SP data, which should be properly addressed in future development of models and algorithms. First, statistical models that accommodate the systematic biases and high noise nature of SP data are often required to make accurate and robust inference [15,79,84,93]. However, it is a challenging task as the number of replicates is limited and variability between replicates is usually substantial. Second, considering the complications in preprocessing SP data, it would be beneficial to incorporate SP data at raw reads level when combined with free energy based predictions. The reason is that summarizing SP data into reactivity profiles might lose information, and the procedure is highly dependent on the normalization method of choice. More importantly, for the problem of construction of structural landscapes, utilizing SP data at a less compressed level rather than reactivities might retain quantitative information. Similarly, it is important to balance between complexity and information retention of SP data in other related problems. Third, recent advances in high throughput mapping of RNA-RNA interactions, such as PARIS [94], SPLASH [95] and LIGR-seq [96], shed new light on the prediction of RNA secondary structure. Rather than detecting pairing states of single nucleotides, these experiments directly capture the pairing of two nucleotides. Such two-dimensional information is valuable as it provides direct evidences for alternative conformations (if exist) and lays crucial foundations for constructing long-range structures.

ACKNOWLEDGEMENTS

H. L. acknowledge the following fundings: the National Natural Science Foundation of China (No. 11601259) and Shanghai Municipal Science and Technology Major Project (No. 2017SHZDZX01).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Bo Yu, Yao Lu, Qiangfeng Cliff Zhang and Lin Hou declare that they have no conflict of interests.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Luco, R. F. and Misteli, T. (2011) More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Curr. Opin. Genet. Dev., 21, 366–372
2. Licatalosi, D. D. and Darnell, R. B. (2010) RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet., 11, 75–87
3. Cech, T. R. and Steitz, J. A. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell, 157, 77–94
4. Cech, T. R. (2012) The RNA worlds in context. Cold Spring Harb. Perspect. Biol., 4, a006742
5. Strobel, E. J., Watters, K. E., Loughrey, D. and Lucks, J. B. (2016) RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. Curr. Opin. Biotechnol., 39, 182–191
6. Halvorsen, M., Martin, J. S., Broadaway, S. and Laederach, A. (2010) Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet., 6, e1001074
7. Piao, M., Sun, L. and Zhang, Q. C. (2017) RNA regulations and functions decoded by transcriptome-wide RNA structure probing. Genom. Proteom. Bioinf., 15, 267–278
8. Schroeder, R., Barta, A. and Semrad, K. (2004) Strategies for RNA folding and assembly. Nat. Rev. Mol. Cell Biol., 5, 908–919
9. Keel, A. Y., Rambo, R. P., Batey, R. T. and Kieft, J. S. (2007) A general strategy to solve the phase problem in RNA crystallography. Structure, 15, 761–772
10. Sun, L., Fazal, F. M., Li, P., Broughton, J. P., Lee, B., Tang, L., Huang, W., Kool, E. T., Chang, H. Y. and Zhang, Q. C. (2019) RNA structure maps across mammalian cellular compartments. Nat. Struct. Mol. Biol., 26, 322–330
11. Smola, M. J., Calabrese, J. M. and Weeks, K. M. (2015) Detection of RNA-protein interactions in living cells with SHAPE. Biochemistry, 54, 6867–6875
12. Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R. C., Snyder, M. P., Segal, E., *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. Nature, 505, 706–709
13. Choudhary, K., Lai, Y. H., Tran, E. J. and Aviran, S. (2019) dStruct: identifying differentially reactive regions from RNA structurome profiling data. Genome Biol., 20, 40
14. Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J. W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., *et al.* (2015) Structural imprints *in vivo* decode RNA regulatory mechanisms. Nature, 519, 486–490
15. Choudhary, K., Deng, F. and Aviran, S. (2017) Comparative and integrative analysis of RNA structural profiling data: current practices and emerging questions. Quant. Biol., 5, 3–24
16. Yan, K., Arfat, Y., Li, D., Zhao, F., Chen, Z., Yin, C., Sun, Y., Hu, L., Yang, T. and Qian, A. (2016) Structure prediction: new insights

into decrypting long noncoding RNAs. Int. J. Mol. Sci., 17, 132
17. Lorenz, R., Wolfinger, M. T., Tanzer, A. and Hofacker, I. L. (2016) Predicting RNA secondary structures from sequence and probing data. Methods, 103, 86–98
18. McCaskill, J. S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers, 29, 1105–1119
19. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res., 31, 3406–3415
20. Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res., 9, 133–148
21. Carvalho, L. E. and Lawrence, C. E. (2008) Centroid estimation in discrete high-dimensional spaces with applications in biology. Proc. Natl. Acad. Sci. USA, 105, 3209–3214
22. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. Science, 244, 48–52
23. Wuchty, S., Fontana, W., Hofacker, I. L. and Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers, 49, 145–165
24. Ding, Y., Chan, C. Y. and Lawrence, C. E. (2005) RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. RNA, 11, 1157–1166
25. Ding, Y. and Lawrence, C. E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. Nucleic Acids Res., 31, 7280–7301
26. Ding, Y., Chan, C. Y. and Lawrence, C. E. (2006) Clustering of RNA secondary structures with application to messenger RNAs. J. Mol. Biol., 359, 554–571
27. Do, C. B., Woods, D. A. and Batzoglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics, 22, e90–e98
28. Hamada, M., Sato, K. and Asai, K. (2010) Prediction of RNA secondary structure by maximizing pseudo-expected accuracy. BMC Bioinformatics, 11, 586
29. Puton, T., Kozlowski, L. P., Rother, K. M. and Bujnicki, J. M. (2013) CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. Nucleic Acids Res., 41, 4307–4323
30. Hamada, M. (2015) RNA Secondary Structure Prediction from Multi-Aligned Sequences. In: RNA Bioinformatics, Picardi, E., (ed.), pp. 17–38. Totowa: Humana Press Inc.
31. Mathews, D. H. and Turner, D. H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. Tinoco. J. Mol. Biol., 317, 191–203
32. Knudsen, B. and Hein, J. (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. Nucleic Acids Res., 31, 3423–3428
33. Hofacker, I. L. (2003) Vienna RNA secondary structure server. Nucleic Acids Res., 31, 3429–3431
34. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res., 28, 235–242
35. Andronescu, M., Bereg, V., Hoos, H. H. and Condon, A. (2008) RNA STRAND: the RNA secondary structure and statistical analysis database. BMC Bioinformatics, 9, 340
36. Hajiaghayi, M., Condon, A. and Hoos, H. H. (2012) Analysis of energy-based algorithms for RNA secondary structure prediction. BMC Bioinformatics, 13, 22
37. Xu, Z., Almudevar, A. and Mathews, D. H. (2012) Statistical evaluation of improvement in RNA secondary structure prediction. Nucleic Acids Res., 40, e26
38. Mathews, D. H. (2019) How to benchmark RNA secondary structure prediction accuracy. Methods, 162–163, 60–67
39. Peattie, D. A. and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. Proc. Natl. Acad. Sci. USA, 77, 4679–4682
40. Noller, H. F. and Chaires, J. B. (1972) Functional modification of 16S ribosomal RNA by kethoxal. Proc. Natl. Acad. Sci. USA, 69, 3115–3118
41. Strobel, E. J., Yu, A. M. and Lucks, J. B. (2018) High-throughput determination of RNA structures. Nat. Rev. Genet., 19, 615–634
42. Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, D. L., Nutter, R. C., Segal, E. and Chang, H. Y. (2012) Genome-wide measurement of RNA folding energies. Mol. Cell, 48, 169–181
43. Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L. S. and Gregory, B. D. (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. PLoS Genet., 6, e1001141
44. Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., Lowe, T. M., Salama, S. R. and Haussler, D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat. Methods, 7, 995–1001.
45. Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y. and Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast. Nature, 467, 103–107
46. Smola, M. J. and Weeks, K. M. (2018) In-cell RNA structure probing with SHAPE-MaP. Nat. Protoc., 13, 1181–1195
47. Saus, E., Willis, J. R., Pryszcz, L. P., Hafez, A., Llorens, C., Himmelbauer, H. and Gabaldn, T. (2018) nextPARS: parallel probing of RNA structures in Illumina. RNA, 24, 609–619
48. Busan, S. and Weeks, K. M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. RNA, 24, 143–148
49. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S. and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. Nat. Methods, 14, 75–82
50. Ritchey, L. E., Su, Z., Tang, Y., Tack, D. C., Assmann, S. M. and Bevilacqua, P. C. (2017) Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure *in vivo*. Nucleic Acids Res., 45, e135
51. Incarnato, D., Anselmi, F., Morandi, E., Neri, F., Maldotti, M., Rapelli, S., Parlato, C., Basile, G. and Oliviero, S. (2017) High-throughput single-base resolution mapping of RNA 2-O-methylated residues. Nucleic Acids Res., 45, 1433–1441

52. Chan, D., Feng, C. and Spitale, R. C. (2017) Measuring RNA structure transcriptome-wide with icSHAPE. Methods, 120, 85–90
53. Flynn, R. A., Zhang, Q. C., Spitale, R. C., Lee, B., Mumbach, M. R. and Chang, H. Y. (2016) Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc., 11, 273–290
54. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. and Weeks, K. M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. Nat. Protoc., 10, 1643–1669
55. Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A. and Vinther, J. (2015) SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. RNA, 21, 1042–1052
56. Ding, Y., Kwok, C. K., Tang, Y., Bevilacqua, P. C. and Assmann, S. M. (2015) Genome-wide profiling of *in vivo* RNA structure at single-nucleotide resolution using structure-seq. Nat. Protoc., 10, 1050–1066
57. Talkish, J., May, G., Lin, Y., Woolford, J. L. Jr and McManus, C. J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA, 20, 713–720
58. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. and Weeks, K. M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat. Methods, 11, 959–965
59. Seetin, M.G., Kladwang, W., Bida, J.P. and Das, R. (2014) Massively Parallel RNA Chemical Mapping with a Reduced Bias MAP-Seq Protocol. In: RNA Folding. Methods in Molecular Biology (Methods and Protocols), Waldsich, C. (ed.), vol 1086, pp. 95–117. Totowa: Humana Press
60. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. Nature, 505, 701–705
61. Loughrey, D., Watters, K. E., Settle, A. H. and Lucks, J. B. (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res., 42, e165
62. Incarnato, D., Neri, F., Anselmi, F. and Oliviero, S. (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. Genome Biol., 15, 491
63. Homan, P. J., Favorov, O. V., Lavender, C. A., Kursun, O., Ge, X., Busan, S., Dokholyan, N. V. and Weeks, K. M. (2014) Single-molecule correlated chemical probing of RNA. Proc. Natl. Acad. Sci. USA, 111, 13858–13863
64. Hector, R. D., Burlacu, E., Aitken, S., Le Bihan, T., Tuijtel, M., Zaplatina, A., Cook, A. G. and Granneman, S. (2014) Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. Nucleic Acids Res., 42, 12138–12154
65. Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C. and Assmann, S. M. (2014) *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature, 505, 696–700
66. Mortimer, S. A., Trapnell, C., Aviran, S., Pachter, L. and Lucks, J. B. (2012) SHAPE-Seq: high-throughput RNA structure analysis. Curr. Protoc. Chem. Biol., 4, 275–297
67. Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., Pachter, L., Doudna, J. A. and Arkin, A. P. (2011) Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. USA, 108, 11063–11068
68. Silverman, I. M., Berkowitz, N. D., Gosai, S. J. and Gregory, B. D. (2016) Genome-Wide Approaches for RNA Structure Probing. In: RNA Processing. Advances in Experimental Medicine and Biology, Yeo, G. (eds.), vol 907, pp. 29–59. Cham: Springer
69. Bevilacqua, P. C., Ritchey, L. E., Su, Z., and Assmann, S. M. (2016) Genome-wide analysis of RNA secondary structure. Annu. Rev. Genet., 50, 235–266
70. Kubota, M., Tran, C. and Spitale, R. C. (2015) Progress and challenges for chemical probing of RNA structure inside living cells. Nat. Chem. Biol., 11, 933–941
71. Kwok, C. K., Tang, Y., Assmann, S. M. and Bevilacqua, P. C. (2015) The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. Trends Biochem. Sci., 40, 221–232
72. Yesselman, J. D., Tian, S., Liu, X., Shi, L., Li, J. B. and Das, R. (2018) Updates to the RNA mapping database (RMDB), version 2. Nucleic Acids Res., 46, D375–D379
73. Cordero, P., Lucks, J. B. and Das, R. (2012) An RNA Mapping DataBase for curating RNA structure mapping experiments. Bioinformatics, 28, 3006–3008
74. Rocca-Serra, P., Bellaousov, S., Birmingham, A., Chen, C., Cordero, P., Das, R., Davis-Neulander, L., Duncan, C. D., Halvorsen, M., Knight, R., *et al.* (2011) Sharing and archiving nucleic acid structure mapping data. RNA, 17, 1204–1212
75. Berkowitz, N. D., Silverman, I. M., Childress, D. M., Kazan, H., Wang, L. S. and Gregory, B. D. (2016) A comprehensive database of high-throughput sequencing-based RNA secondary structure probing data (Structure Surfer). BMC Bioinformatics, 17, 215
76. Norris, M., Kwok, C. K., Cheema, J., Hartley, M., Morris, R. J., Aviran, S. and Ding, Y. (2017) FoldAtlas: a repository for genome-wide RNA structure probing data. Bioinformatics, 33, 306–308
77. Wu, Y., Qu, R., Huang, Y., Shi, B., Liu, M., Li, Y. and Lu, Z. J. (2016) RNAex: an RNA secondary structure prediction server enhanced by high-throughput structure-probing data. Nucleic Acids Res., 44, W294–W301
78. Kladwang, W., Mann, T. H., Becka, A., Tian, S., Kim, H., Yoon, S. and Das, R. (2014) Standardization of RNA chemical mapping experiments. Biochemistry, 53, 3063–3065
79. Selega, A., Sirocchi, C., Iosub, I., Granneman, S. and Sanguinetti, G. (2017) Robust statistical modeling improves sensitivity of high-throughput RNA structure probing experiments. Nat. Methods, 14, 83–89
80. Low, J. T. and Weeks, K. M. (2010) SHAPE-directed RNA secondary structure prediction. Methods, 52, 150–158
81. Tang, Y., Bouvier, E., Kwok, C. K., Ding, Y., Nekrutenko, A., Bevilacqua, P. C. and Assmann, S. M. (2015) StructureFold:

genome-wide RNA secondary structure mapping and reconstruction *in vivo*. Bioinformatics, 31, 2668–2675

82. Aviran, S., Trapnell, C., Lucks, J. B., Mortimer, S. A., Luo, S., Schroth, G. P., Doudna, J. A., Arkin, A. P. and Pachter, L. (2011) Modeling and automation of sequencing-based characterization of RNA structure. Proc. Natl. Acad. Sci. USA, 108, 11069–11074
83. Li, B., Tambe, A., Aviran, S., and Pachter, L. (2017) PROBer provides a general toolkit for analyzing sequencing-based toeprinting assays. Cell Syst., 4, 568–574 e7
84. Zou, C. and Ouyang, Z. (2015) Joint modeling of RNase footprint sequencing profiles for genome-wide inference of RNA structure. Nucleic Acids Res., 43, 9187–9197
85. Deigan, K. E., Li, T. W., Mathews, D. H. and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. USA, 106, 97–102
86. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA, 101, 7287–7292
87. Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K. and Arkin, A. P. (2012) Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. Nucleic Acids Res., 40, 5775–5786
88. Deng, F., Ledda, M., Vaziri, S. and Aviran, S. (2016) Data-directed RNA secondary structure prediction using probabilistic modeling. RNA, 22, 1109–1119
89. Spasic, A., Assmann, S. M., Bevilacqua, P. C. and Mathews, D. H. (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. Nucleic Acids Res., 46, 314–323
90. Wu, Y., Shi, B., Ding, X., Liu, T., Hu, X., Yip, K. Y., Yang, Z. R., Mathews, D. H. and Lu, Z. J. (2015) Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. Nucleic Acids Res., 43, 7247–7259
91. Ouyang, Z., Snyder, M. P. and Chang, H. Y. (2013) SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. Genome Res., 23, 377–387
92. Li, H. and Aviran, S. (2018) Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. Nat. Commun., 9, 606
93. Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M. and Simon, M. D. (2017) Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. Biochemistry, 56, 4713–4721
94. Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S., *et al.* (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. Cell, 165, 1267–1279
95. Aw, J. G. A., Shen, Y., Wilm, A., Sun, M., Lim, X. N., Boon, K. L., Tapsin, S., Chan, Y. S., Tan, C. P., Sim, A. Y., *et al.* (2016) *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. Mol. Cell, 62, 603–617
96. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. and Blencowe, B. J. (2016) Global mapping of human RNA-RNA interactions. Mol. Cell, 62, 618–626

