



OPEN ACCESS

EDITED BY

Alberto Paccanaro,
FGV EMAp—School of Applied Mathematics,
Brazil

REVIEWED BY

Joao Carlos Setubal,
University of São Paulo, Brazil

*CORRESPONDENCE

Peter F. Stadler,
✉ studla@bioinf.uni-leipzig.de

RECEIVED 29 November 2023

ACCEPTED 09 January 2024

PUBLISHED 22 February 2024

CITATION

von Löhneysen S, Mörl M and Stadler PF (2024),
Limits of experimental evidence in RNA
secondary structure prediction.
Front. Bioinform. 4:1346779.
doi: 10.3389/fbinf.2024.1346779

COPYRIGHT

© 2024 von Löhneysen, Mörl and Stadler. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Limits of experimental evidence in RNA secondary structure prediction

Sarah von Löhneysen¹, Mario Mörl² and Peter F. Stadler^{1,3,4,5,6,7*}

¹Bioinformatics Group, Department of Computer Science, Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig, Germany, ²Institute for Biochemistry, Leipzig University, Leipzig, Germany, ³Competence Center for Scalable Data Analytics and Artificial Intelligence, School of Embedded and Compositive Artificial Intelligence (SECAI), Leipzig University, Leipzig, Germany, ⁴Department of Theoretical Chemistry, University of Vienna, Wien, Austria, ⁵Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia, ⁶Center for Non-Coding RNA in Technology and Health, University of Copenhagen, Frederiksberg, Denmark, ⁷Santa Fe Institute, Santa Fe, NM, United States

KEYWORDS

RNA secondary structure, chemical probing, sequence covariation, pseudo-energies, loop-based energy model

1 Introduction

Chemical and enzymatic probing has a long history as an experimental source of information on RNA secondary structures. In recent years such protocols were interfaced with high-throughput sequencing methods to provide access to transcriptome-wide structural information (Kubote et al., 2015; Carlson et al., 2018). Despite the indisputable usefulness of structure probing, it is important to remember that any probing method provides a signal that *encodes* information on RNA structures, but remains far from directly measuring or unambiguously determining a structure.

An extensive body of empirical evidence of RNA structures has been integrated into the “standard model” for RNA secondary structure prediction. It defines an RNA secondary structure as a collection of Watson-Crick and GU base pairs such that i) each base has at most one pairing partner, ii) base pairs do not cross, i.e., if (i, j) is a pair, then there is no pair (k, l) with $i < k < j$ and $l < i$ or $l > j$, and iii) every base pair spans at least three unpaired positions (Lorenz et al., 2011). Every structure of this type is associated with an energy that can be computed as the sum of its loops (facets of its unique planar embedding), which correspond to stacked base pairs, hairpin loops, interior loops, and multi-branched loops. The energy contribution of each loop depends on its sequence, but is independent of its external context. Comprehensive tables of sequence-dependent loop energy contributions have been inferred (mostly) from melting experiments on small, specifically designed RNA molecules (Andronescu et al., 2014). Collected in the *standard energy model* (Turner and Mathews, 2010), they are used in exact dynamic programming algorithms that predict the ground state structure or the base pairing probabilities in the Boltzmann ensemble of secondary structures for arbitrary RNA sequences. We note in passing that Stochastic Context Free Grammars (SCFGs) use in essence the same model(s) for the structures (Rivas et al., 2012) and may serve as an alternative to the thermodynamic approach. Usually, SCFGs are parametrized using learning approaches from known structures, see, e.g., (Do et al., 2006). For the purpose of the present contribution, it is only important that there is a “universal” model that predicts (a reasonable approximation of) the secondary structure taking an arbitrary RNA sequence as input.

Empirical evidence, e.g., from probing experiments can be included in the universal structure prediction methods as a hard constraint forbidding structures that contradict the empirical evidence or as an additional energy term (soft constraint), favoring those structures that conform better to the empirical data over others, see, e.g., (Lorenz et al.,

2016c; Lorenz et al., 2016a). Since *unambiguous* experimental evidence that a base is paired or unpaired is difficult to obtain, we consider here only soft constraints that better reflect the probabilistic nature of the available evidence. This amounts to the inclusions of pseudo-energy terms that award a “bonus” to all secondary structures that exhibit a specific feature whose presence is supported by the external evidence.

2 Pseudo-energies from probing data

Most current methods for large-scale chemical probing use deep sequencing methods as read-out. The raw signal thus is a number of reads associated with each sequence position i . The probing methods most frequently employed at present detect unpaired positions. In SHAPE, RNA forms an adduct at conformationally flexible 2'-hydroxyl positions (Deigan et al., 2009), where flexibility serves as a proxy for unpairedness. Similarly, DMS treatment leads to a methylation of N¹ of adenine and N³ of cytosine in unpaired bases. In inline probing and related protocols using heavy metal ions as catalysts, the RNA is cleaved preferentially at unpaired positions. Both cleavage and bulky adducts (which lead to termination of reverse transcription) translate to read-ends in subsequent high-throughput sequencing.

Other chemically introduced modifications, in particular at the 2'-hydroxyl position of the ribose of structurally unconstrained positions, lead to misincorporations during cDNA synthesis, because reverse transcriptases incorporate non-templated nucleotides (Smola and Weeks, 2018). In the SHAPE-MaP approach, the nature of the misincorporated base can be identified as a base replacement in the resulting sequence alignments (mutational profiling). However, the efficiency of the ribose modification depends on the reactivity of the 2'-hydroxyl group, which itself is affected by the nature of the individual base (Wilkinson et al., 2009; Busan et al., 2019). Hence, the number of reads that is used as a proxy for signal reliability can vary to a certain extent. PORE-cupine, the combination of such structure-dependent nucleotide modifications with the Nanopore sequencing technology allows for an elegant direct read-out of these signals in individual RNA molecules, enabling single molecule structure analysis (Aw et al., 2021).

High throughput probing experiments require two distinct normalization steps since the observed, position-wise signal depends i) on the abundance of the probed RNA, i.e., the expression level, and ii) on the secondary structure. Normalization for expression levels requires annotated transcripts. While it would be desirable in principle to have RNA-seq data for the same sample to estimate expression levels, and possibly to refine the annotation, it is possible to use the probing signal itself. A reasonable normalized signal S can be obtained, e.g., by dividing the counts by the mean (adjusted to drop outliers) or the median of the read counts over a given annotation item. Much more elaborate statistical models have been developed to estimate *reactivities* from high-throughput sequencing data, taking into account both RT stops and misincorporations (Strobel et al., 2018; Yu et al., 2018).

The normalized signal then needs to be related to a probability or pseudo-energy contribution for the feature under consideration.

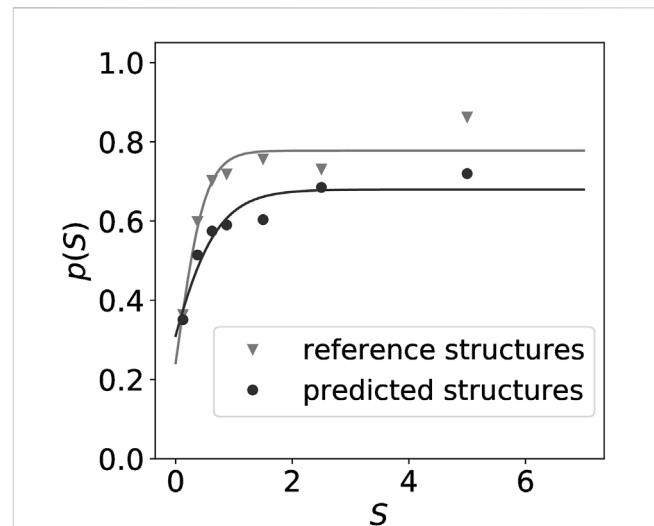


FIGURE 1
Normalized intensities S obtained from probing experiments [in this example from a Led-Seq cP library (Kolberg et al., 2023)] can be converted to probabilities of a structural feature (here the probability of a sequence position to be unpaired) by relating the empirical signal in a bin $[S, S + \Delta S]$ to the frequency of the feature of interest (here the frequency of observing an unpaired position) in a reference set. Here we compare a manually curated set of 32 reference secondary structures from *Escherichia coli* (∇) to secondary structures predicted from the thermodynamic model using the ViennaRNA software (\bullet) of all sequences with valid probing information from the same study (Kolberg et al., 2023). A smooth function $p(S)$ is then obtained by fitting a sigmoidal curve to the empirical data.

Reactivities are often directly converted to pseudo-energies using simple empirical formulas (Strobel et al., 2018). A more principled approach proposed by Zarringhalam et al. (2012) is to first estimate the probability $p(S)$ of the features as a function of the observed signal strength and then to convert $p(S)$ to a pseudo-energy via

$$E(S) = -RT \ln \frac{p(S)}{1 - p(S)} \quad (1)$$

Here $p(S)$ will in general be a monotonic function of the normalized signal. Typically, $p(S)$ will be sigmoidal to limit the impact of outliers. We note in passing that the conventional conversion of SHAPE reactivities to pseudo-energies (Low and Weeks, 2010), $E = m \ln(S + 1) + b$, is a particular case of Eq. 1 using the sigmoidal function $p(S) = 1/[1 + (b/RT)(S + 1)^{m/RT}]$ with empirical fitting parameters b and m . The pseudo-energy terms are included as additional position-dependent contributions in the thermodynamic RNA folding algorithms, see e.g., (Lorenz et al., 2016c; Lorenz et al., 2016b).

The advantage of Eq. 1 is not only a more direct interpretation as a log-odds ratio. It is also readily extended to aggregating evidence from different sources, e.g., from different probing experiments. This is used in practice, e.g., in the Led-Seq approach (Kolberg et al., 2023). There, each lead-induced cleavage at single-stranded positions is assayed both via the 2',3'-cyclophosphate end and the 5'-OH end and modeled via a two-dimensional sigmoidal fit $p(S_1, S_2)$. Of course, it is also possible to stratify the signal for instance by the identity of the cleaved di-nucleotide. The function $p(S)$ can be estimated by comparing the observed signal S with the

frequency of the assayed feature in a set of reference structures. Typically, a collection of well-known structures is used for this purpose. It should be kept in mind, however, that few RNA structures are perfectly known at present. Recent advance in Cryo-EM methods, however, may alleviate this bottleneck (Ma et al., 2022). Moreover, RNA structures can change with temperature (Narberhaus et al., 2006; Marz et al., 2010), salt concentration (Yao et al., 2023) and with the presence of binding partners such as proteins or other RNAs (Sutandy et al., 2018). In practice, therefore, the calibration of the function $p(S)$ will have to be performed from imperfect data.

Figure 1 shows that the thermodynamic model for RNA folding is surprisingly accurate in predicting the distinction of paired versus unpaired nucleotides. Predicted structures can therefore serve as an alternative to reference structures when estimating $p(S)$. The difference in accuracy between a manually curated reference set and predictions from the standard energy model translates to a lower saturation value in terms of $p(S)$ for the thermodynamic predictions. Importantly, the curve for the curated reference data also saturates at a value well below the theoretical upper bound of $p(S) = 1$ for $S \rightarrow \infty$. While possible errors in the reference data may contribute to this effect, it is most likely dominated by the fact that probing methods assay chemical properties that are only correlated with structural features such as the unpairedness of a nucleotide instead of directly measuring them.

Moreover, several RNA molecules are known to contain high affinity binding sites for metal ions (Pyle, 2002) that can lead to disproportionately high cleavage results (Ciesiolka et al., 1994). Nevertheless, such prominent cleavage sites are highly informative in structural analysis based on metal cleavage (Behlen et al., 1990; Kolberg et al., 2023). Another point to consider is the fact that temperature can have a dramatic impact on the probing result. Here, lead-dependent probing can be applied to a wide range of temperatures, allowing for structural investigation in psychro- as well as thermophilic organisms (Kolberg et al., 2023). Other approaches depending on more temperature-sensitive reagents might be limited in this aspect, and the corresponding half-lives at increased temperatures must be considered (Smola and Weeks, 2018; Busan et al., 2019).

The distribution of errors in thermodynamic secondary structure predictions could in principle be determined empirically by comparison with curated reference data. This opens the possibility to devise estimates for $p(S)$ that compensate for the imperfections of thermodynamic predictions. We are not aware, however, that such a method has become available.

3 Pseudo-energies from phylogenetic conservation

Evolutionary conservation of structures leads to mutual constraints at spatial contacts. Given a multiple sequence alignment, this effect can be quantified as covariation or mutual information between alignment columns. In MIFOLD (Freyhult et al., 2005), consensus secondary structures are predicted with reasonable accuracy directly from the mutual information of alignment columns. The thermodynamic model can also be readily extended to multiple sequence alignments by averaging

energy contributions for stacking and loops over the rows (Hofacker et al., 2002; Bernhart et al., 2008). In practice, the RNAalifold program also uses covariance-based pseudo-energies. Both MIFOLD and RNAalifold are based on the assumption that there is a global consensus structure that is present in each of the aligned sequences in essence without variations. This is not always the case, however.

In many cases evolutionary conserved secondary structures are only local elements in often much larger RNA molecules. This is most obvious for features such as Selenocystein Insertion (SECIS) elements or structured Internal Ribosomal Entry Sites (IRES) on protein-coding mRNAs. Conserved structures in long non-coding RNAs also seem to be local in general. In such cases, consensus methods typically predict large unstructured regions. For any particular RNA molecule, these regions will typically form secondary structures, which, however, are not conserved across the aligned sequences. von Löhneysen et al. (2023) therefore proposed to convert predicted consensus structures into pseudo-energies. In the setting of the RNAalifold approach this is most directly achieved by using $\Gamma_{ij} = \min\{0, -RT \ln[p_{ij}^{\circ}/(1 - p_{ij}^{\circ})]\}$ as pseudo-energy, where p_{ij}° is the probability of i and j being paired in the alignment according to the averaged energy model. The values of Γ_{ij} are truncated at 0 since the absence of conserved base pairs does not imply that the two positions are excluded from base pairing in particular sequences. The statistical significance of covarying alignment columns can be tested using RSCAPE (Rivas et al., 2017).

Not surprisingly, the beneficial effect of conservation-derived pseudo-energies on the accuracy of the structure prediction depends strongly on the quality of the multiple sequence alignment (von Löhneysen et al., 2023). In addition, the phylogenetic distribution of the input sequences may play a role. Some consensus structure prediction methods such as pfold (Knudsen and Hein, 2003) explicitly require a phylogenetic tree. Alternatively, the sequences (rows) in an alignment may be given weights depending on the similarity to the other aligned sequences (Vingron and Sibbald, 1993); this is used, e.g., in RNAalifold (Bernhart et al., 2008). Both explicit and implicit phylogenetic information incurs the danger, however, that the effect of alignment errors are aggravated. Misaligned sequences can be expected to have larger distances from other members of the alignments. As a consequence they appear less redundant and thus contribute with higher weight.

4 External evidence versus energy-based prediction

In Figure 2 we compare the accuracy of RNA secondary structures generated by using different combinations of evidence. For the sake of this exposition, we only analyzed a limited data set. While this cannot replace a thorough benchmarking, it shows the salient trends, and illustrates some basic facts.

In order to assess the structural information contained in the probing signals alone, we exclusively used the position-wise pseudo-energies as energy model. The contribution for potential base pair (i, j) is set to $E_{ij} = -(E_i + E_j)$, where E_i and E_j is the pseudo-energy for positions i and j to be unpaired. Since this “energy model” scores the base pairs and unpaired positions independently of each other, the optimal secondary structure can be computed using Nussinov’s

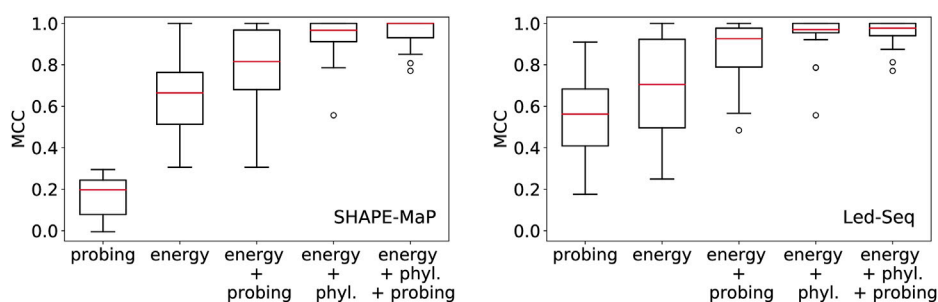


FIGURE 2

Structural information in probing data and comparative analysis of RNA structures. Position-wise probing information alone yield less accurate structures. Combining probing or conservation data (phyl.) with the energy model improves predictions considerably. Left: SHAPE-MaP data of 12 ncRNAs from *Escherichia coli* (Mustoe et al., 2018) downloaded from RASP (Li et al., 2020), Right: Led-Seq data of 24 ncRNAs from *Escherichia coli* (Kolberg et al., 2023). Probing data were converted to pseudo-energies as described by Kolberg et al. (2023) for probing data and by von Löhneysen et al. (2023) for phylogenetic information. MCC, Matthews correlation coefficient.

circular matching algorithm (Nussinov and Jacobson, 1980) in the same manner as for the calculation of maximum expected accuracy structures (Lu et al., 2009). For both SHAPE-MaP and Led-Seq data, the predicted structures are rather inaccurate. In particular, the probing data alone yield structures that are systematically worse than the un-aided energy-based predictions. We note that the SHAPE-MaP and Led-Seq data in Figure 2 are not directly comparable since they were retrieved from unrelated experiments. In SHAPE-MaP, Led-Seq, and DMS probing, structural signals are only obtained for unpaired nucleotides, while base-paired regions are inert. The protocol commonly employed for SHAPE data is to interpret low signals as a double-stranded segment in the RNA structure. Here, the misinterpretation of unreacted but single-stranded regions is possible, resulting in misleading structural models. Therefore Kolberg et al. (2023), instead interprets strong signals as unpaired regions, reducing the danger of misreading the lack of signal.

Including either probing data or conservation information substantially improves the structure prediction. This shows that the experimental evidence from probing can be meaningfully accessed only in conjunction with a universal folding model, in our example the thermodynamic model. We also observe that the inclusion of phylogenetic information yields substantially better structural models than the probing data. This is probably a consequence of the fact that probing data offer only position specific constraints, while phylogenetic methods introduce specific base pairs and thus restrict the search space quite drastically.

5 Discussion

The main purpose of this short opinion piece is to highlight the fundamental importance of the loop-based thermodynamic energy model (or one of its SCFG-based variants) for RNA secondary structure determination. Although chemical and enzymatic probing methods provide invaluable additional structural information, they cannot unambiguously determine RNA structures on their own. This begs the question to what extent probing data can identify

pseudoknotted structures and whether this can be achieved without a reasonably accurate pseudoknot-aware thermodynamic folding algorithm.

Tb-seq exploits the fact that Tb^{3+} causes backbone cleavage in regions where the compression of the phosphate backbone causes sharp, stable turns in the RNA structure. Since such regions are typically associated with stable tertiary interactions, Tb-seq provides key information beyond secondary structures (Patel et al., 2023), even though its signal yields a position-dependent profile just like other probing methods. The interpretation of Tb-seq data at present also requires a good structural model to start with.

The information obtainable from crosslinking approaches such as SPLASH, RIC-seq, PARIS, LIGR-seq and others [reviewed, e.g., by Zhang et al. (2022)] goes beyond position-wise base pairing propensities and provides direct evidence on interacting RNA regions, see, e.g., (Schäfer and Voß, 2021). It seems fair to say, however, that the problem of deriving detailed secondary structure models from such data has not yet been solved in a satisfactory manner. Comparative sequence analysis provides an attractive alternative since it yields direct evidence on specific base pairs. This information can then be included into folding algorithms in the same way as probing data, namely, by adding a pseudo-energy E_{ij} to the base pair (i, j) . Wherever comparative data on a conserved structural consensus is applicable and available, furthermore, these tend to have a larger beneficial impact on prediction accuracy than probing data. We believe that this is due to the fact that the specific consensus base pairs are much more informative than the position-wise status of being paired or unpaired. It appears, furthermore, that very little is gained by combining conservation and probing information in cases where the entire RNA structure is well-conserved over a long evolutionary time scale. On the other hand, probing data are invaluable in the much more frequent scenario that only certain functional elements of an RNA are well-conserved. In this setting we expect that the combination of probing and conservation data is particularly useful.

The energy-directed model also seems to be sufficient at least in principle to gauge the conversion of (normalized) probing signals to pseudo-energies. This alleviates the need to build large collections of manually curated reference structures, which would be hard to obtain in

many cases, in particular when analyzing transcriptomes of non-model organism. Less naïve methods than the simple fitting procedure of Figure 1, however, will need to be developed for this purpose.

Despite these limitations of probing data and the superiority of comparative information, where it is available, probing is indispensable in many situations. For example, structural changes caused by binding partners or chemical modifications are difficult, and usually impossible, to capture by conservation data, even though it is possible in some cases to identify conserved alternative folds, see, e.g., (Meyer, 2017). Similarly, probing data are key whenever phylogenetic evidence does not exist, as is the case for RNAs designed in synthetic biology applications (Domin et al., 2017) and in evolutionary novelties appearing in loci with accelerated evolution (Benjaminov et al., 2008).

Author contributions

SL: Conceptualization, Software, Writing—original draft, Writing—review and editing. MM: Conceptualization, Funding acquisition, Writing—original draft, Writing—review and editing. PS: Conceptualization, Funding acquisition, Writing—original draft, Writing—review and editing.

References

- Andronescu, M., Condon, A., Turner, D. H., and Mathews, D. H. (2014). The determination of RNA folding nearest neighbor parameters. *Methods Mol. Biol.* 1097, 45–70. doi:10.1007/978-1-62703-709-9_3
- Aw, J. G. A., Lim, S. W., Wang, J. X., Lambert, F. R. P., Tan, W. T., Shen, Y., et al. (2021). Determination of isoform-specific RNA structure with nanopore long reads. *Nat. Biotechnol.* 39, 336–346. doi:10.1038/s41587-020-0712-z
- Behlen, L. S., Sampson, J. R., DiRenzo, A. B., and Uhlenbeck, O. C. (1990). Lead-catalyzed cleavage of yeast tRNA^{Phe} mutants. *Biochemistry* 23, 2515–2523. doi:10.1021/bi00462a013
- Benjaminov, A. D., Westhof, E., and Krol, A. (2008). Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* 14, 1270–1275. doi:10.1261/rna.1054608
- Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R., and Stadler, P. F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinforma.* 9, 474. doi:10.1186/1471-2105-9-474
- Busan, S., Weidmann, C. A., Sengupta, A., and Weeks, K. M. (2019). Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry* 58, 2655–2664. doi:10.1021/acs.biochem.8b01218
- Carlson, P. D., Evans, M. E., Yu, A. M., Strobel, E. J., and Lucks, J. B. (2018). Snapshot RNA structure probing technologies. *Cell* 175, 600–600.e1. doi:10.1016/j.cell.2018.09.024
- Ciesiolka, J., Hardt, W. D., Schlegl, J., Erdmann, V. A., and Hartmann, R. K. (1994). Lead-ion-induced cleavage of RNase P RNA. *Eur. J. Biochem.* 219, 49–56. doi:10.1111/j.1432-1033.1994.tb19913.x
- Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A.* 106, 97–102. doi:10.1073/pnas.0806929106
- Do, C. B., Woods, D. A., and Batzoglou, S. (2006). CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* 22, e90–e98. doi:10.1093/bioinformatics/btl246
- Domin, G., Findeiß, S., Wachsmuth, M., Will, S., Stadler, P. F., and Mörl, M. (2017). Applicability of a computational design approach for synthetic riboswitches. *Nucleic Acids Res.* 45, 4108–4119. doi:10.1093/nar/gkw1267
- Freyhult, E., Moulton, V., and Gardner, P. P. (2005). Predicting RNA structure using mutual information. *Appl. Bioinf.* 4, 53–59. doi:10.2165/00822942-200504010-00006
- Hofacker, I. L., Fekete, M., and Stadler, P. F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319, 1059–1066. doi:10.1016/s0022-2836(02)00308-x
- Knudsen, B., and Hein, J. (2003). Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* 31, 3423–3428. doi:10.1093/nar/gkg614
- Kolberg, T., von Löhneysen, S., Ozerova, I., Wellner, K. W., Hartmann, R. K., Stadler, P. F., et al. (2023). Led-seq – ligation-enhanced double-end sequence-based structure analysis of RNA. *Nucleic Acids Res.* 51, e63. doi:10.1093/nar/gkad312
- Kubote, 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. doi:10.1038/nchembio.1958
- Li, P., Zhou, X., Xu, K., and Zhang, Q. C. (2020). RASP: an atlas of transcriptome-wide RNA secondary structure probing data. *Nucleic Acids Res.* 49, D183–D191. doi:10.1093/nar/gkaa880
- Lorenz, R., Bernhart, S. H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA package 2.0. *Alg. Mol. Biol.* 6, 26. doi:10.1186/1748-7188-6-26
- Lorenz, R., Hofacker, I. L., and Stadler, P. F. (2016a). RNA folding with hard and soft constraints. *Alg. Mol. Biol.* 11, 8. doi:10.1186/s13015-016-0070-z
- Lorenz, R., Luntzer, D., Hofacker, I. L., Stadler, P. F., and Wolfinger, M. T. (2016b). SHAPE directed RNA folding. *Bioinformatics* 32, 145–147. doi:10.1093/bioinformatics/btv523
- Lorenz, R., Wolfinger, M. T., Tanzer, A., and Hofacker, I. L. (2016c). Predicting RNA secondary structures from sequence and probing data. *Methods* 103, 86–98. doi:10.1016/j.ymeth.2016.04.004
- Low, J. T., and Weeks, K. M. (2010). SHAPE-directed RNA secondary structure prediction. *Methods* 52, 150–158. doi:10.1016/j.ymeth.2010.06.007
- Lu, Z. J., Gloor, J. W., and Mathews, D. H. (2009). Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* 15, 1805–1813. doi:10.1261/rna.1643609
- Ma, H., Jia, X., Zhang, K., and Su, Z. (2022). Cryo-EM advances in RNA structure determination. *Signal Transduct. Target. Ther.* 7, 58. doi:10.1038/s41392-022-00916-0
- Marz, M., Vanzo, N., and Stadler, P. F. (2010). Temperature-dependent structural variability of RNAs: spliced leader RNAs and their evolutionary history. *J. Bioinf. Comp.* 8, 1–17. doi:10.1142/s0219720010004525
- Meyer, I. M. (2017). *In silico* methods for co-transcriptional RNA secondary structure prediction and for investigating alternative RNA structure expression. *Methods* 120, 3–16. doi:10.1016/j.ymeth.2017.04.009
- Mustoe, A. M., Busan, S., Rice, G. M., Hajdin, C. E., Peterson, B. K., Ruda, V. M., et al. (2018). Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173, 181–195.e18. doi:10.1016/j.cell.2018.02.034
- Narberhaus, F., Waldminghaus, T., and Chowdhury, S. (2006). RNA thermometers. *FEMS Microbiol. Rev.* 30, 3–16. doi:10.1111/j.1574-6976.2005.004.x

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the Deutsche Forschungsgemeinschaft (Grant Numbers MO 634/18-1 and STA 850/48-1).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Nussinov, R., and Jacobson, A. B. (1980). Fast algorithm for predicting the secondary structure of single stranded RNA. *Proc. Natl. Acad. Sci. U. S. A.* 77, 6309–6313. doi:10.1073/pnas.77.11.6309
- Patel, S., Sexton, A. N., Strine, M. S., Wilen, C. B., Simon, M. D., and Pyle, A. M. (2023). Systematic detection of tertiary structural modules in large RNAs and RNP interfaces by Tb-seq. *Nat. Commun.* 14, 3426. doi:10.1038/s41467-023-38623-1
- Pyle, A. (2002). Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* 7, 679–690. doi:10.1007/s00775-002-0387-6
- Rivas, E., Clements, J., and Eddy, S. R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* 14, 45–48. doi:10.1038/nmeth.4066
- Rivas, E., Lang, R., and Eddy, S. R. (2012). A range of complex probabilistic models for RNA secondary structure prediction that includes the nearest-neighbor model and more. *RNA* 18, 193–212. doi:10.1261/rna.030049.111
- Schäfer, R. A., and Voß, B. (2021). RNAnue: efficient data analysis for RNA-RNA interactomics. *Nucleic Acids Res.* 49, 5493–5501. doi:10.1093/nar/gkab340
- Smola, M. J., and Weeks, K. M. (2018). In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* 13, 1181–1195. doi:10.1038/nprot.2018.010
- Strobel, E. J., Yu, A. M., and Lucks, J. B. (2018). High-throughput determination of RNA structures. *Nat. Rev. Genet.* 19, 615–634. doi:10.1038/s41576-018-0034-x
- Sutandy, F. X. R., Ebersberger, S., Huang, L., Busch, A., Bach, M., Kang, H. S., et al. (2018). *In vitro* iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Res.* 28, 699–713. doi:10.1101/gr.229757.117
- Turner, D. H., and Mathews, D. H. (2010). NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucl. Acids Res.* 38, D280–D282. doi:10.1093/nar/gkp892
- Vingron, M., and Sibbald, P. R. (1993). Weighting in sequence space: a comparison of methods in terms of generalized sequences. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8777–8781. doi:10.1073/pnas.90.19.8777
- von Löhneysen, S., Spicher, T., Varenky, Y., Yao, H.-T., Lorenz, R., Hofacker, I., et al. (2023). Phylogenetic information as soft constraints in RNA secondary structure prediction. *Bioinforma. Res. Appl.* 14248, 267–279. doi:10.1007/978-981-99-7074-2_21
- Wilkinson, K. A., Vasa, S. M., Deigan, K. E., Mortimer, S. A., Giddings, M. C., and Weeks, K. M. (2009). Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* 15, 1314–1321. doi:10.1261/rna.1536209
- Yao, H.-T., Lorenz, R., Hofacker, I. L., and Stadler, P. F. (2023). Mono-valent salt corrections for RNA secondary structures in the ViennaRNA package. *Alg. Mol. Biol.* 18, 8. doi:10.1186/s13015-023-00236-0
- Yu, A. M., Evans, M. E., and Lucks, J. B. (2018). Estimating RNA structure chemical probing reactivities from reverse transcriptase stops and mutations. *Tech. Rep. Biorxiv.* doi:10.1101/292532
- Zarringhalam, K., Meyer, M. M., Dotu, I., Chuang, J. H., and Clote, P. (2012). Integrating chemical footprinting data into RNA secondary structure prediction. *PLOS ONE* 7, e45160. doi:10.1371/journal.pone.0045160
- Zhang, J., Fei, Y., Sun, L., and Zhang, Q. C. (2022). Advances and opportunities in RNA structure experimental determination and computational modeling. *Nat. Methods* 19, 1193–1207. doi:10.1038/s41592-022-01623-y