

# RNA structure prediction: From 2D to 3D

Bernhard C. Thiel<sup>1</sup>      Christoph Flamm<sup>1</sup>  
Ivo L. Hofacker<sup>1,2\*</sup>

July 19, 2017

<sup>1</sup>Department of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Austria

<sup>2</sup>Research Group BCB, Faculty of Computer Science, University of Vienna, Austria

\*To whom correspondence should be addressed.

## Abstract

We summarize different levels of RNA structure prediction, from classical 2D structure to extended secondary structure and motif based research towards 3D structure prediction of RNA. We outline the importance of classical secondary structure during all those levels of structure prediction.

## 1 Introduction

The secondary structure model of RNA is at the center of most computational work related to RNA. In this review we will sketch how recent research has taken advantage of the increasing amount of experimentally solved RNA 3D structures to move beyond the secondary structure view, but how it cannot move away from it.

RNA secondary structure is defined by canonical AU, GC base-pairs and GU wobble-pairs, which form between nucleotides typically within one RNA strand to create anti-parallel A-type helices. Pseudoknot-free secondary structures can be encoded as strings, e.g. in the Vienna dot-bracket notation, where nested pairs of left and right parentheses signify nested base pairs, while dots signify unpaired nucleotides (See Figure 1, box “Secondary Structure”).

RNA secondary structure can be efficiently predicted from the primary sequence using dynamic programming approaches. The easiest such algorithm maximizes the number of base-pairs for a given sequence [1], while state-of-the-art prediction tools [2] use an energy model with different energy contributions for different loop types. This model takes the stacking of neighboring base pairs into account and is often referred to as nearest neighbor model. The energy parameters of the nearest neighbor model were derived from RNA melting experiments collected by Turner et al [3]. These secondary structure prediction algorithms can be used not only to predict the structure with the lowest free energy, but also to generate the Boltzmann ensemble of suboptimal structures [2, 4, 5]. By including data from chemical probing experiments such as SHAPE, the reliability of the generated secondary structures can be further increased. For an overview over state of the art techniques in secondary structure prediction and their integration with chemical probing data, see the recent review by

Lorenz et al.[6]. Pseudoknots increase the computational complexity, but simple types can be predicted with dedicated software [7]

The great success of the secondary structure description of RNA is grounded in the hierarchical nature of RNA folding [8]. Intra-helical interactions have a significantly greater energy contribution than tertiary interactions. Thus, energy barriers and in most cases energy differences between different secondary structures are significantly higher than energy differences and barriers between different 3D structures that correspond to the same secondary structure.

Secondary structure is useful in many approaches beyond simple structure prediction: Sequence design [9, 10] aims at finding sequences that fold into one or more predefined (meta-)stable structures. Folding kinetics [11] and co-transcriptional folding [12, 13] look at the dynamic nature of RNA secondary structure formation. Functional RNA secondary structures are expected to be evolutionary conserved. Hence they can be detected in genomic data via compensatory mutations and structure stability [14] or co-variance models (RNA families) [15, 16]).

While some biological mechanisms can be understood on the basis of secondary structure alone, others require more detailed structure knowledge. In this contribution we will therefore focus on approaches that go beyond classical secondary structure.

On the other end of the spectrum lie methods that model RNA structures in full atomistic resolution. In particular, Molecular Dynamics simulations can be readily applied to RNA, with the AMBER force-field being most popular in the RNA community. However, despite many RNA related corrections to the AMBER force field, the current version still sometimes fails to identify the correct native state among all possible states [17, 18]. Despite these challenges, reduced model [19] and explicit solvent MD simulations[20] of the full ribosome have been successfully performed, using experimental crystal structures as a starting point and, in the case of the reduced model, for the construction of a structure-based (Gō-like) potential. The reduced model simulation covered biologically relevant time-scales, while the explicit solvent MD simulation covered over a microsecond, enough to separate rapid local fluctuations from large-scale collective movements. The latter simulation took several months on over 1000 computer cores.

In contrast to molecular dynamics from a given starting conformation, true de-novo simulations are possible only for very small molecules, due to the huge space of possible conformations and rugged energy landscapes, even when combined with enhanced sampling techniques [21] Recent applications of classical force fields to RNA folding range from the folding of G-quadruplexes [22] and RNA-ion interactions [23] to the folding of tetra-loop hairpins [17].

While all-atom approaches are limited to small molecules or fragments, somewhat larger RNAs can be treated by methods working with a coarse grained RNA representation, which we will review in sections 3 and 4. In both cases, prior knowledge of the secondary structure is not required. Rather, correct secondary structure should emerge as a by-product of tertiary structure prediction. Occasionally, known secondary structures are used as constraints to reduce the search space. However, we will also discuss tertiary structure methods that explicitly build on secondary structure knowledge and use it to enable a more aggressive coarse-graining.

For a discussion on different ways of coarse graining in RNA 3D structure

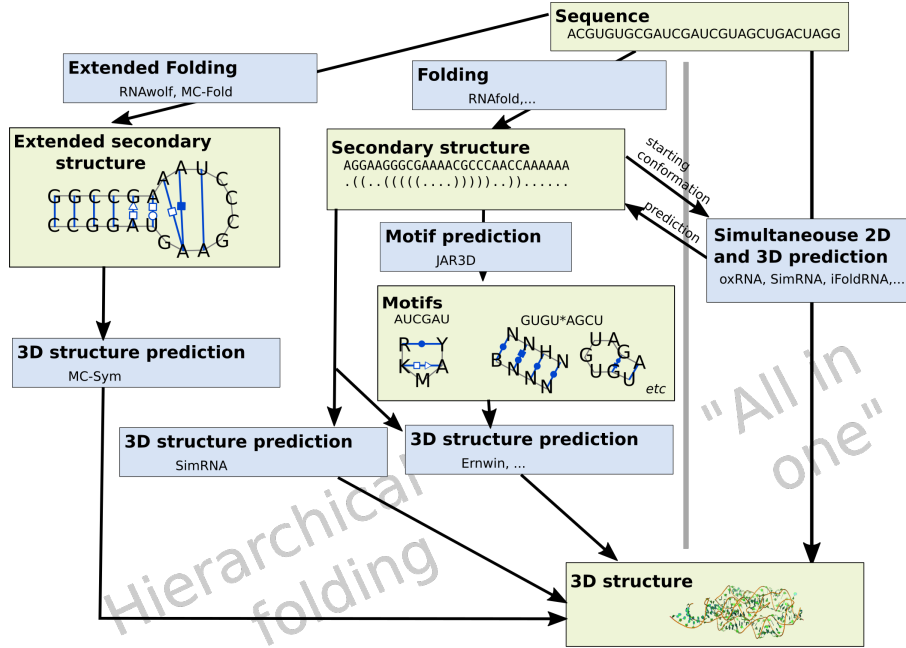


Figure 1: RNA structure prediction programs can be classified into hierarchical folding approaches and all-in-one programs.

prediction and the distinction between theory based and knowledge based potentials, see the recent review by Dawson et al. [24].

## 2 Extended secondary structure

While prediction of full tertiary RNA structures remains an extremely difficult task, several promising avenues have emerged to go beyond classical secondary structure without tackling actual 3D structure, e.g. by extending the notion of secondary structure to include non-canonical interactions. In addition to the Watson-Crick and GU wobble pairs, RNA nucleotides can form a wide variety of interactions, both pairwise and between more than two nucleotides.

A highly useful classification of these non-canonical interactions was introduced by Leontis and Westhof [25], by noting that almost all interactions happen on one of 3 “edges” (the Watson-Crick, Hoogsteen, and sugar edge). Together with the relative orientation of the glycosidic bond (cis or trans) this results in 12 basepairing types, each of which has at least two hydrogen bonds and was observed for at least some combinations of bases [26]. Additionally, weaker interactions involving a single hydrogen bond, base triplets or multiplets and G-quadruplexes have been observed in RNA molecules.

An RNA structure including some or all of the above-mentioned non-canonical interactions, but not the full 3D information, will be called an *extended secondary structure*. *Decomposition* and *motif search* are the two main approaches for the prediction of extended secondary structures.

Extended secondary structures play an important role in many natural RNAs, including ribosomal RNAs[27], where non-canonical basepairs contribute to the movement of the ribosome[28].

## 2.1 Loop decomposition with non-canonical basepairs

Classical secondary structure prediction is based upon the unique decomposition of the RNA structure into loops delimited only by canonical and G-U basepairs. The program **MC-Fold** [29] generalizes the classical decomposition by considering all types of non-canonical basepairs as loop delimiters. The resulting loops, termed NCMs (Nucleotide cyclic motifs) in **MC-Fold**, are defined via a minimal cycle basis of the RNA structure graph [30].

The programs **MC-Fold** [29] and the faster dynamic programming implementation **MC-Fold-dp** [31] find optimal and sub-optimal combinations of NCMs for a given sequence according to a statistical energy function. This energy function is composed of probability terms for finding individual and combinations of two NCMs.

**RNAwolf** [31] further generalizes the **MC-Fold** approach by allowing each nucleotide to form two interactions, thus supporting base-triplets which are observed quite frequently in nature. **RNAwolf**, however, uses a simplified Nussinov-like energy model, limiting its accuracy. Parametrization of a full featured energy model remains the big challenge for extended secondary structure prediction, and it is unclear if enough data is available to estimate the large number of parameters required. Moreover, while the nearest neighbor approximation has been shown to be quite accurate for canonical secondary structures, the distortion of the double helix through non-canonical pairs could well exert an influence beyond its direct neighbors.

## 2.2 Motif based

An alternative to the loop decomposition above is to treat substructures containing several non-canonical pairs as a single unit, called a tertiary structure motif. Such an RNA motif is characterized by well-defined (base-pairing) interactions [32, 33], a well-defined geometry [34, 35, 36],[37], or both. The prime example of such a recurrent motif would be the kink turn motif, a sharp turn between two adjacent helices introduced by a number of bases forming a dense network of non-canonical interactions. This and three other common motifs can be detected in single sequences or multiple sequence alignments using **RMDetect** [38].

A well-maintained collection of such motifs can be found in the Motif Atlas [39]. An accompanying search tool, **JAR3D** [40], can be used to predict the presence of any motif from the Atlas in interior or hairpin loops of a secondary structure.

Motif based approaches circumvent the problem of decomposing complex interaction networks, such as the kink turn, and even allow for crossing interaction within these motifs. On the downside, they cannot predict novel motifs and are limited to the set of motifs that we observe in known tertiary structures. One application of motifs is homology modeling, where non-homologous regions can be filled in using motifs.[41]

Prediction of G-quadruplexes can also be seen as a motif based extension of secondary structure. G-quadruplex prediction has recently been integrated in the folding algorithms of the ViennaRNA package [42]. This integration presents a significant advantage over pure sequence searches, as it properly treats the competition between formation of G-quads and normal secondary structure.

Extended secondary structures are of great interest because functionally important interaction with proteins and other factors, typically take place in regions of irregular structure, rather than perfect helices. Moreover, extended secondary structures provide a much better starting place for modeling tertiary structures. The kink turn again serves as a perfect example: The sharp bend introduced by this motif will strongly affect the overall shape of the molecule, but would be very hard to predict by any approach that is not aware of the motif.

### 3 Hierarchical folding enables aggressive coarse-graining

While secondary structure can be used in most 3D structure prediction programs to constrain the search space, for some programs the secondary structure is at the core of the abstraction, on which the coarse grained structure representation is based.

Vfold3D [43, 44] creates a coarse grained 3D scaffold from the RNA secondary structure and sequence using a template based approach. For each loop region, the best matching template is selected from a template library constructed from PDB structures, where the match quality is based on sequence similarity. Helices are modeled as ideal A-type helices. From this scaffold, a full atom model is constructed which is then relaxed using an AMBER all-atom force field.

More than one fragment can be tried for each loop in an exhaustive way (5 fragments were tried in the RNA puzzles entries - see below). However, Vfold3D does not contain any energy function to score these coarse-grained scaffolds, nor does it contain a sampling protocol to sample combinations of such fragments. Since it is unlikely that AMBER energy minimization will overcome larger energy barriers, the quality of the prediction depends on the correct choice of fragments. Thus Vfold works best when the structure of loops strongly depends on the sequence (see section 2.2 about motifs) or Vfold contains close homologs of the target RNA in its knowledge-base, while it has limitations whenever similar loop conformations.

RNAComposer [45] uses a machine translation system that translates RNA

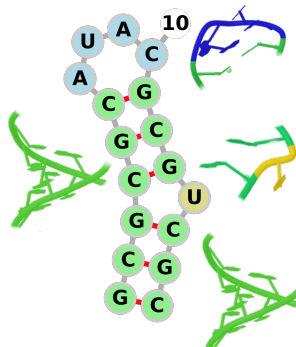


Figure 2: The secondary structure of RNA defines fragments for fragment assembly or aggressive coarse-graining.

sequence and secondary structure into 3D structure. Similar to **VFold** it selects the best matching fragment for every loop and helix from a database of fragments. **RNAComposer** provides fallback mechanisms for cases where no fragment is found as well as two final refinement steps using **CYANA** (for refinement in torsion angle space) and the **CHARMM** force field (for refinement in cartesian space). Like **Vfold3D**, it allows for the creation of random suboptimal structures but does not provide sampling methods or a fast to evaluate energy function for a systematic exploration of the conformational space.

**ERNWIN** [46] and **RAGTOP** [47] use the user-supplied secondary structure to guide their aggressive coarse-graining of the RNA into helices (stems) and connecting loops. In contrast to **RNAComposer** and **Vfold3D**, these tools explore the conformational space on the level of loops and helices.

**RAGTOP** [47], first employs machine learning to predict the topology of multi-loops (junctions) [48]. Next interior loop angles are sampled with a Monte Carlo/ Simulated Annealing algorithm using a knowledge-based potential. Finally an all-atom representation is recovered from the sampled conformations [49]. With RMSD values from 2.38 to 14.56 (for structures of 25 to 158 nucleotides) the reported prediction accuracy of **RAGTOP** is slightly better than previous tools which use less aggressive coarse-graining.

**ERNWIN** [46] uses a similar coarse-graining based on secondary structure. Helices are described by 10 parameters: The position of the helix’s start and end (or equivalently the start and the direction) and 4 parameter for the position of the minor groove along the length of the helix. This model assumes a regular helix where the position of the minor groove with respect to the helices axis changes by a fixed angle with each subsequent nucleotide. The helix axis and the vector pointing to the minor groove form a local coordinate system for residues. Using average atom position data in this local coordinate system gathered from a non-redundant list of PDB files, so-called “virtual atom positions” of all the helix atoms can be quickly calculated just from the 10 parameters defining the helix. Another major difference between **RAGTOP** and **ERNWIN** is the fact that the later can sample different multiloop-conformations. This means that the prediction quality of **ERNWIN** is independent of the correctness of the junction topology prediction, but comes at the cost of a higher number of rejections during Monte Carlo sampling. Finally, sampling and energy evaluation are significantly different between **RAGTOP** and **ERNWIN**. In **ERNWIN**, local loop and helix conformations are sampled directly from a fragment library, whereas **RAGTOP** samples the angle from continuous space and therefore needs local energy terms to evaluate the likeliness of a local angle. As for the contribution of global features to the energy function, both programs have a term for the radius of gyration. **ERNWIN** in addition contains a term for loop-loop interactions and a sophisticated term for interactions of single stranded Adenines with the minor groove of a helix (A-Minor motif [50]). **ERNWIN** also has direct support for motif search via **JAR3D** (see above).

While the idea of using helices without internal degrees of freedom is common to **RAGTOP** and **ERNWIN**, it is the details that matter: The sampling strategy that explores the conformational space and the energy function that detects native-like conformations require a lot of fine-tuning and still have room for improvement.

**MC-Sym** [29] uses a different approach towards tertiary structure prediction. It is built on top of **MC-Fold** which can predict extended secondary structures.

It uses a library of 3D fragments for each nucleotide cyclic motif (NCM) predicted by MC-Fold to assemble 3D structures. In a Las Vegas algorithm, 3D structures are sampled for 12 hours by aligning adjacent NCM-3D fragments to form structures. The final result is an ensemble of 3D structures.

## 4 Predicting secondary and tertiary structure together (“all in one”)

While the tools in the previous section built their model on top of a given secondary structure, the following programs add secondary structure constraints into their model via a force field.

The program NAST[51] represents each nucleotide by a single point, which means that it does not hold any information about the orientation of the base with respect to the backbone. The program requires an input secondary structure, which is used to create energy potentials on the lengths, angles and dihedrals between residues. These potentials direct the sampling of the RNA towards the desired secondary structure. We argue that the requirement for a secondary structure is inherent to the coarse grained model used, because formation of new basepairs would depend on the orientation of the base, which is not part of the model. In NAST, sampling via a Molecular Dynamics approach starting from the extended chain is followed by filtering according to tertiary structure constraints (e.g. from SAXS or SHAPE experiments or from phylogenetic observations [52]) and clustering using an k-means algorithm based on a simplified pairwise GDT-TS [53] distance. For one RNA molecule the effect of errors in the secondary structure input was investigated and the authors concluded that up to 35% wrong base-pairs did not significantly reduce the RMSD of their prediction.

The most aggressive coarse-graining possible that allows for de-novo prediction of base-pairs uses one rigid body per nucleotide (in contrast to one point), as implemented in oxRNA [54, 55] (see below). In a similar fashion, FARNA uses rigid fragments of 1 to 3 nucleotides to assemble the final RNA structure. Next to energy terms for clashes and the radius of gyration, FARNA uses a base-pairing statistical potential that can be seen as a heatmap around the center of an ideal base. If two bases are placed in a relative orientation that is frequently found in base-pairs in solved RNA structures, these bases receive a favorable energy contribution. Structures generated by FARNA can be refined using an all-atom force field, giving rise to the combined method FARFAR [56, 57]. In theory such an all-atom refinement would be useful for all coarse grained approaches (but see introduction for the limits of current force-fields), however its integration in one framework (namely the Rosetta framework) is a great technical advantage of FARNA/FARFAR.

In contrast to FARNA, oxRNA [54, 55] samples structures from a continuous space. It can be used for Molecular Dynamics Simulations and for Monte Carlo calculations. The energy function of oxRNA is parametrized to favor RNA A-type helices by using potentials for sequence dependent hydrogen bonding, stacking and cross-stacking in helices in addition to the backbone potential. Since all nucleotides are equally sized, mismatches in helices do not disturb the overall helix geometry in oxRNA and thus the stability of such helices is over-estimated

despite the lack of the energy contribution from the hydrogen bonding term. The quality of the secondary structure predicted by **oxRNA** is comparable with secondary structure prediction tools using the nearest neighbor model for small RNAs, as shown by melting temperature calculations done with **oxRNA**. Since **oxRNA** does not include any non-canonical or long-range interactions (except excluded volume effects), it is no surprise that the use of a 3-dimensional model does not bring any significant advantage over a 2 dimensional model with respect to the prediction of secondary structures. The **oxRNA** paper studied mechanical properties like force-extension and overstretching properties, persistence length and modeled hairpin-unzipping. These applications, while leading to results that are only in the same order of magnitude as the experimental values, could not be done with most of the other RNA structure prediction models.

In contrast to the continuous energy model used in **oxRNA**, **simRNA** [58, 59] and **iFoldRNA** [60] use discrete, grid-based statistical potentials. While **iFoldRNA** uses Discrete Molecular Dynamics [61, 62], **simRNA** uses Monte Carlo simulations.

**iFoldRNA** [60] uses 3 beads per nucleotide: one for the sugar, one for the phosphate and only one for the base (see Figure 3). The position of the base relative to the sugar is used to determine the direction of hydrogen bonding, but tilting of the base’s plane cannot be modeled with this coarse-graining. Non-canonical interactions are implicitly modeled by the use of a general hydrophobic attraction between all kinds of bases. **iFoldRNA**’s energy function only contains local terms (for bond length, angle and dihedral angle, base pairing, phosphate-phosphate repulsion, hydrophobic interactions and base stacking). Furthermore an additional energy term for the loop entropy is used to compensate for the bias in loop entropy introduced by the coarse-graining. According to the data reported in the paper’s supplement [60], **iFoldRNA** predicts pseudoknot-free secondary structures slightly better than **MFold** (average Q-value of 0.953 vs 0.948) and can additionally predict pseudoknots. However, we note that the reported benchmark of secondary structure prediction only includes the Q-value (true positives) and does not take false positives (additional basepairs) into account.

**SimRNA** [58, 59] models the base with 3 points and is thus able to capture the full 3-dimensional orientation of the base’s plane. The statistical energy terms for base-base, base-backbone and backbone-backbone interactions can be visualized as heatmaps similar to the ones used in the original **FARNA** publication [63]. The authors report that their model performs better than other 3D structure prediction programs for short RNA sequences, but needs explicit secondary structure (and potentially long-range) restraints to model longer RNA molecules correctly.

Among the last 3 discussed models, **simRNA** is the only one which fully incorporates non-canonical base-pairs in its energy model. Since secondary structure constraints are modeled as distance constraints, (extended) secondary structures could in principle be used as input, but the type of the base-pair in the prediction might be different from the type in the input structure.

## 5 RNA puzzles

As of now, 3 rounds of RNA puzzles, a CASP-like blind experiment in RNA 3D structure prediction, have been completed [64, 65, 66]. In these competitions,



participating experimental groups provide solved but unpublished tertiary structures. Computational groups then have a few weeks time to perform tertiary structure prediction given the sequence and - in some cases - chemical probing information.

The submitted models are scored using different measures: The root mean square deviation (RMSD) is a common measure for comparison of macromolecules which can be quickly calculated [67], but has received some critiques [68]. The Interaction Network Fidelity (INF) can be calculated for stacking and (non-)canonical hydrogen bonding interactions. While the RMSD is sensitive to errors in flexible loop regions, the INF is sensitive to errors in structured regions. Additionally, RMSD and INF are combined to the Deformation Index (DI). From round II of RNA puzzles onwards, the mean of circular quantities (MCQ)[69] which is based on angular coordinates, was used as additional measure. In contrast to the RMSD, errors in the relative orientation of two relatively large RNA domains do not have a higher impact on the MCQ score than errors in the orientation of small stems or hairpins.

In the first round of RNA puzzles, one riboswitch domain, a stem-loop structure with two interior loops and a square formed by 4 helices and interior loops in the corners were modeled. Of these challenges, only the riboswitch domain contained true higher order 3D folds. Interestingly, a fragment based approach (Vfold) combined with relaxation in an all-atom AMBER force-field yielded the result with the lowest RMSD for this puzzle. This probably means that fragments from homologous RNA molecules were selected based on the sequence similarity score. The second best result of this puzzle used a coarse grained approach with secondary structure constraints, followed by full-atom refinement in a Discrete Molecular Dynamics framework (iFoldRNA). One lesson learned from this RNA puzzle is certainly that the huge size of the sampling space requires some sort of coarse-grained initial step which uses the secondary structure, followed by an all-atom refinement.

In round II [65] of the RNA puzzles contest, SHAPE data was provided to

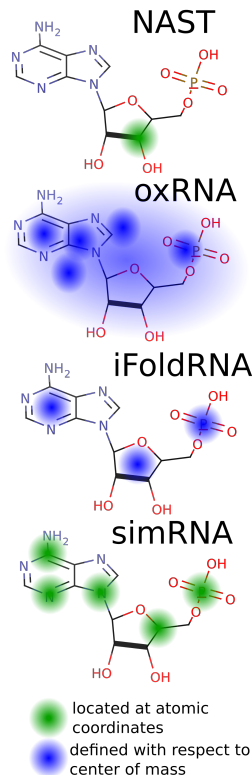


Figure 3: Different coarse grained representations of nucleotides. **NAST** uses a single point located at the C3' atom. **oxRNA** uses a single rigid body with multiple interaction sites defined with respect to the RNA's center of mass (the positions in the figure are rough estimates of this interaction sites) **iFoldRNA** uses 3 beads per residue, located at the center of the phosphate, the sugar and the base respectively. **simRNA** uses 5 beads per nucleotide, 3 of which define the plane of the base.

all participants by one group. Since this round’s target structures were longer RNA molecules with more complex 3D folds (a ribozyme, a riboswitch and a T-box-tRNA complex), these chemical probing data were crucial for 3D structure prediction. Despite the use of experimental secondary structures, no perfect predictions were obtained, which shows the open challenges in RNA 3D structure prediction.

In round III it became apparent how homology modeling can achieve good results when a homolog with solved structure exists, and how targets without homology to solved structures still pose huge challenges to modelers. It is also notable that many groups who submitted several models could not correctly rank the best prediction as their top choice.

## 6 Summary

While RNA secondary structure prediction is well established and widely used, tertiary structure prediction has long seemed out of reach. Recently, however, two directions have emerged that promise RNA structure models that go beyond secondary structure. Inclusion of structure motifs and non-canonical base pairs could yield extended secondary structures that are more detailed and will perhaps even improve prediction accuracy. Coarse grained models of tertiary structures may overcome the sampling problems that limit all-atom methods.

A successful RNA 3D structure prediction pipeline, as illustrated in figure 4, will need several ingredients: It should start from a reliable secondary structure, preferably including tertiary motifs and supported by experimental evidence, such as probing data. Exploration of the conformation space will be best done using coarse grained models that provide efficient sampling. Empirical scoring functions need to be able to identify coarse-grained conformation(s) close to the native state. Additional experimental restraints can be incorporated at this step, as reviewed by Magnus et al[70]. The final step will reconstruct and refine all-atom models from the best coarse grained conformations. While further improvements are needed in all of these stages, reliable RNA tertiary structure prediction is slowly getting within reach.

## 7 Acknowledgments

This work was funded, in part, by the Austrian FWF, project “SFB F43 RNA regulation of the transcriptome”. We would like to thank Roman Ochsenreiter for proof-reading the manuscript.

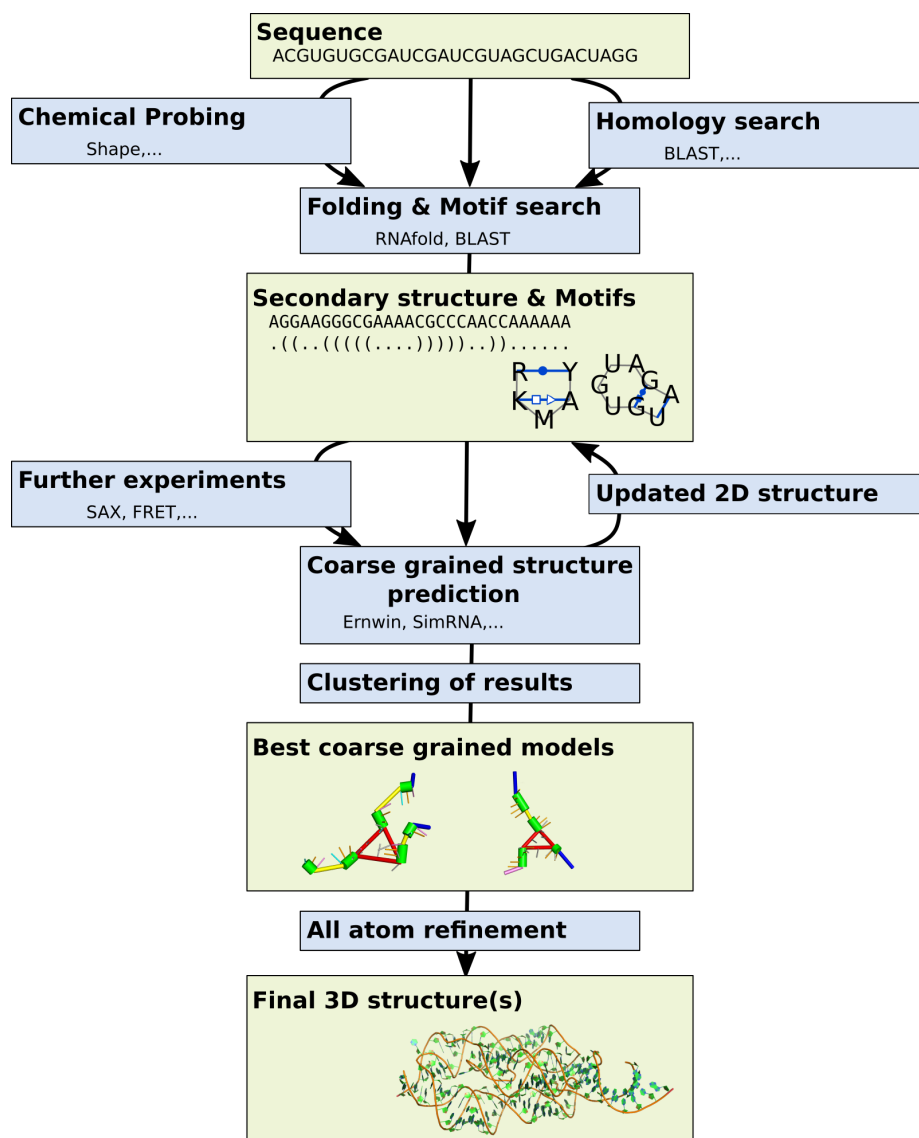


Figure 4: A proposed RNA 3D structure prediction pipeline, as described in Section 6. In addition to the structure prediction (middle), additional experimental (left) and computational steps (right) can improve the prediction accuracy.

## References

- [1] Nussinov R, Pieczenik G, Griggs JR, Kleitman DJ. Algorithms for Loop Matchings. *SIAM Journal on Applied Mathematics*. 1978 jul;35(1):68–82. Available from: <https://doi.org/10.1137/2F0135006>.
- [2] Lorenz R, Bernhart SH, zu Siederdissen CH, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*. 2011;6(1):26. Available from: <https://doi.org/10.1186/2F1748-7188-6-26>.
- [3] Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*. 2009 oct;38(Database):D280–D282. Available from: <https://doi.org/10.1093/2Fnar/2Fgkp892>.
- [4] Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie - Chemical Monthly*. 1994 feb;125(2):167–188. Available from: <https://doi.org/10.1007/2Fbf00818163>.
- [5] Wuchty S, Fontana W, Hofacker IL, Schuster P. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*. 1999;49(2):145–165. Available from: [http://dx.doi.org/10.1002/\(SICI\)1097-0282\(199902\)49:2<145::AID-BIP4>3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G).
- [6] Lorenz R, Wolfinger MT, Tanzer A, Hofacker IL. Predicting RNA secondary structures from sequence and probing data. *Methods*. 2016 jul;103:86–98. Available from: <https://doi.org/10.1016/2Fj.ymeth.2016.04.004>.
- [7] Bellaousov S, Mathews DH. ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. *RNA*. 2010 aug;16(10):1870–1880. Available from: <https://doi.org/10.1261/2Frna.2125310>.
- [8] Mustoe AM, Brooks CL, Al-Hashimi HM. Hierarchy of RNA Functional Dynamics. *Annual Review of Biochemistry*. 2014 jun;83(1):441–466. Available from: <https://doi.org/10.1146/2Fannurev-biochem-060713-035524>.
- [9] Wolfe BR, Porubsky NJ, Zadeh JN, Dirks RM, Pierce NA. Constrained Multistate Sequence Design for Nucleic Acid Reaction Pathway Engineering. *Journal of the American Chemical Society*. 2017 feb;139(8):3134–3144. Available from: <https://doi.org/10.1021/2Fjacs.6b12693>.
- [10] Taneda A. MODENA: a multi-objective RNA inverse folding. *Advances and applications in bioinformatics and chemistry: AABC*. 2011;4:1–12. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169953/>.
- [11] Kucharik M, Hofacker IL, Stadler PF, Qin J. Pseudoknots in RNA folding landscapes. *Bioinformatics*. 2016;32(2):187. Available from: [+http://dx.doi.org/10.1093/bioinformatics/btv572](http://dx.doi.org/10.1093/bioinformatics/btv572).

- [12] Badelt S, Hammer S, Flamm C, Hofacker IL. Thermodynamic and Kinetic Folding of Riboswitches. In: *Methods in Enzymology*. Elsevier BV; 2015. p. 193–213. Available from: <https://doi.org/10.1016%2Fbs.mie.2014.10.060>.
- [13] Proctor JR, Meyer IM. COFOLD: an RNA secondary structure prediction method that takes co-transcriptional folding into account. *Nucleic Acids Research*. 2013 mar;41(9):e102–e102. Available from: <https://doi.org/10.1093%2Fnar%2Fgkt174>.
- [14] Gruber AR, Findeiß S, Washietl S, Hofacker IL, Stadler PF. RNAZ 2.0:. In: *Biocomputing 2010*. World Scientific Pub Co Pte Lt; 2009. p. 69–79. Available from: [https://doi.org/10.1142%2F9789814295291\\_0009](https://doi.org/10.1142%2F9789814295291_0009).
- [15] Eddy SR, Durbin R. RNA sequence analysis using covariance models. *Nucleic Acids Research*. 1994 jun;22(11):2079–2088. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308124/>.
- [16] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29(22):2933–2935. Available from: <http://dx.doi.org/10.1093/bioinformatics/btt509>.
- [17] Kührová P, Best RB, Bottaro S, Bussi G, Šponer J, Otyepka M, et al. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *Journal of Chemical Theory and Computation*. 2016 sep;12(9):4534–4548. Available from: <https://doi.org/10.1021%2Facs.jctc.6b00300>.
- [18] Gil-Ley A, Bottaro S, Bussi G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *Journal of Chemical Theory and Computation*. 2016 jun;12(6):2790–2798. Available from: <https://doi.org/10.1021%2Facs.jctc.6b00299>.
- [19] Whitford PC, Geggier P, Altman RB, Blanchard SC, Onuchic JN, Sanbonmatsu KY. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA*. 2010;16(6):1196–1204. Available from: <http://rnajournal.cshlp.org/content/16/6/1196.abstract>.
- [20] Whitford PC, Blanchard SC, Cate JHD, Sanbonmatsu KY. Connecting the Kinetics and Energy Landscape of tRNA Translocation on the Ribosome. *PLOS Computational Biology*. 2013 03;9(3):1–10. Available from: <https://doi.org/10.1371/journal.pcbi.1003003>.
- [21] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*. 2014 feb;185(2):604–613. Available from: <https://doi.org/10.1016%2Fj.cpc.2013.09.018>.
- [22] Šponer J, Bussi G, Stadlbauer P, Kührová P, Banáš P, Islam B, et al. Folding of guanine quadruplex molecules–funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2016 dec; Available from: <https://doi.org/10.1016%2Fj.bbagen.2016.12.008>.

- [23] Cunha RA, Bussi G. Unravelling  $Mg^{2+}$ -RNA binding with atomistic molecular dynamics. *RNA*. 2017 feb;p. rna.060079.116. Available from: <https://doi.org/10.1261%2Frna.060079.116>.
- [24] Dawson WK, Maciejczyk M, Jankowska EJ, Bujnicki JM. Coarse-grained modeling of RNA 3D structure. *Methods*. 2016 jul;103:138–156. Available from: <https://doi.org/10.1016%2Fj.ymeth.2016.04.026>.
- [25] Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA*. 2001;7(4):499–512. Available from: <http://rnajournal.cshlp.org/content/7/4/499.abstract>.
- [26] RNA Basepair Catalogue;. Accessed: 2017-04-03. <http://ndbserver.rutgers.edu/ndbmodule/services/BPCatalog/bpCatalog.html>.
- [27] Petrov AS, Bernier CR, Gulen B, Waterbury CC, HersHKovits E, Hsiao C, et al. Secondary Structures of rRNAs from All Three Domains of Life. *PLoS ONE*. 2014 feb;9(2):e88222. Available from: <https://doi.org/10.1371/journal.pone.0088222>.
- [28] Mohan S, Noller HF. Recurring RNA structural motifs underlie the mechanics of L1 stalk movement. *Nature Communications*. 2017 feb;8:14285. Available from: <https://doi.org/10.1038/ncomms14285>.
- [29] Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*. 2008 mar;452(7183):51–55. Available from: <https://doi.org/10.1038%2Fnature06684>.
- [30] Lemieux S. Automated extraction and classification of RNA tertiary structure cyclic motifs. *Nucleic Acids Research*. 2006 apr;34(8):2340–2346. Available from: <https://doi.org/10.1093%2Fnar%2Fgk1120>.
- [31] zu Siederdisen CH, Bernhart SH, Stadler PF, Hofacker IL. A folding algorithm for extended RNA secondary structures. *Bioinformatics*. 2011 jun;27(13):i129–i136. Available from: <https://doi.org/10.1093%2Fbioinformatics%2Fbtr220>.
- [32] Djelloul M, Denise A. Automated motif extraction and classification in RNA tertiary structures. *RNA*. 2008 oct;14(12):2489–2497. Available from: <https://doi.org/10.1261%2Frna.1061108>.
- [33] Zhong C, Zhang S. Clustering RNA structural motifs in ribosomal RNAs using secondary structural alignment. *Nucleic Acids Research*. 2011 oct;40(3):1307–1317. Available from: <https://doi.org/10.1093%2Fnar%2Fgkr804>.
- [34] Wadley LM. The identification of novel RNA structural motifs using COM-PADRES: an automated approach to structural discovery. *Nucleic Acids Research*. 2004 dec;32(22):6650–6659. Available from: <https://doi.org/10.1093%2Fnar%2Fgkh1002>.
- [35] HUANG HC. The application of cluster analysis in the intercomparison of loop structures in RNA. *RNA*. 2005 apr;11(4):412–423. Available from: <https://doi.org/10.1261%2Frna.7104605>.

- [36] Wang X, Huan J, Snoeyink JS, Wang W. Mining RNA Tertiary Motifs with Structure Graphs. In: 19th International Conference on Scientific and Statistical Database Management (SSDBM 2007). Institute of Electrical and Electronics Engineers (IEEE); 2007. p. 31–40. Available from: <https://doi.org/10.1109/2Fssdbm.2007.38>.
- [37] Chojnowski G, Waleń T, Bujnicki JM. RNA Bricks—a database of RNA 3D motifs and their interactions. *Nucleic Acids Research*. 2014;42(D1):D123–D131. Available from: [+http://dx.doi.org/10.1093/nar/gkt1084](http://dx.doi.org/10.1093/nar/gkt1084).
- [38] Cruz JA, Westhof E. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nature Methods*. 2011 may;8(6):513–519. Available from: <https://doi.org/10.1038/2Fnmeth.1603>.
- [39] Petrov AI, Zirbel CL, Leontis NB. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA*. 2013 aug;19(10):1327–1340. Available from: <https://doi.org/10.1261/2Frna.039438.113>.
- [40] Zirbel CL, Roll J, Sweeney BA, Petrov AI, Pirrung M, Leontis NB. Identifying novel sequence variants of RNA 3D motifs. *Nucleic Acids Research*. 2015 jun;43(15):7504–7520. Available from: <https://doi.org/10.1093/2Fnar/2Fgkv651>.
- [41] Tung CS, Joseph S, Sanbonmatsu KY. All-atom homology model of the Escherichia coli 30S ribosomal subunit. *Nature Structural Biology*. 2002 sep;9(10):750–755. Available from: <https://doi.org/10.1038/nsb841>.
- [42] Lorenz R, Bernhart SH, Qin J, Höner zu Siederdissen C, Tanzer A, Amman F, et al. 2D meets 4G: G-Quadruplexes in RNA Secondary Structure Prediction. *IEEE Trans Comp Biol Bioinf*. 2013;10:832–844.
- [43] Xu X, Zhao P, Chen SJ. Vfold: A Web Server for RNA Structure and Folding Thermodynamics Prediction. *PLoS ONE*. 2014 sep;9(9):e107504. Available from: <https://doi.org/10.1371/2Fjournal.pone.0107504>.
- [44] Cao S, Chen SJ. Physics-Based De Novo Prediction of RNA 3D Structures. *The Journal of Physical Chemistry B*. 2011 apr;115(14):4216–4226. Available from: <https://doi.org/10.1021/2Fjp112059y>.
- [45] Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, et al. Automated 3D structure composition for large RNAs. *Nucleic Acids Research*. 2012;40(14):e112. Available from: [+http://dx.doi.org/10.1093/nar/gks339](http://dx.doi.org/10.1093/nar/gks339).
- [46] Kerpedjiev P, zu Siederdissen CH, Hofacker IL. Predicting RNA 3D structure using a coarse-grain helix-centered model. *RNA*. 2015 apr;21(6):1110–1121. Available from: <https://doi.org/10.1261/2Frna.047522.114>.
- [47] Kim N, Laing C, Elmetwaly S, Jung S, Curuksu J, Schlick T. Graph-based sampling for approximating global helical topologies of RNA. *Proceedings of the National Academy of Sciences*. 2014 mar;111(11):4079–4084. Available from: <https://doi.org/10.1073/2Fpnas.1318893111>.

- [48] Laing C, Schlick T. Analysis of Four-Way Junctions in RNA Structures. *Journal of Molecular Biology*. 2009 jul;390(3):547–559. Available from: <https://doi.org/10.1016%2Fj.jmb.2009.04.084>.
- [49] Zahran M, Bayrak CS, Elmetwaly S, Schlick T. RAG-3D: a search tool for RNA 3D substructures. *Nucleic Acids Research*. 2015 aug;43(19):9474–9488. Available from: <https://doi.org/10.1093%2Fnar%2Fgkv823>.
- [50] Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proceedings of the National Academy of Sciences*. 2001 apr;98(9):4899–4903. Available from: <https://doi.org/10.1073%2Fpnas.081082398>.
- [51] Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, et al. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*. 2009 feb;15(2):189–199. Available from: <https://doi.org/10.1261%2Frna.1270809>.
- [52] Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS. 3D RNA and Functional Interactions from Evolutionary Couplings. *Cell*. 2016 may;165(4):963–975. Available from: <https://doi.org/10.1016%2Fj.cell.2016.03.030>.
- [53] Zemla A, Venclovas Č, Moult J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. *Proteins: Structure, Function, and Bioinformatics*. 1999;37(S3):22–29.
- [54] Šulc P, Romano F, Ouldrige TE, Doye JPK, Louis AA. A nucleotide-level coarse-grained model of RNA. *The Journal of Chemical Physics*. 2014 jun;140(23):235102. Available from: <https://doi.org/10.1063%2F1.4881424>.
- [55] Matek C, Šulc P, Randisi F, Doye JPK, Louis AA. Coarse-grained modelling of supercoiled RNA. *The Journal of Chemical Physics*. 2015 dec;143(24):243122. Available from: <https://doi.org/10.1063%2F1.4933066>.
- [56] Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nature Methods*. 2010 feb;7(4):291–294. Available from: <https://doi.org/10.1038%2Fnmeth.1433>.
- [57] Cheng CY, Chou FC, Das R. Modeling Complex RNA Tertiary Folds with Rosetta. In: *Methods in Enzymology*. Elsevier BV; 2015. p. 35–64. Available from: <https://doi.org/10.1016%2Fbs.mie.2014.10.051>.
- [58] Boniecki MJ, Lach G, Dawson WK, Tomala K, Lukasz P, Soltysinski T, et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Research*. 2015 dec;44(7):e63–e63. Available from: <https://doi.org/10.1093%2Fnar%2Fgkv1479>.
- [59] Magnus M, Boniecki MJ, Dawson W, Bujnicki JM. SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Research*. 2016 apr;44(W1):W315–W319. Available from: <https://doi.org/10.1093%2Fnar%2Fgkw279>.



- [60] Sharma S, Ding F, Dokholyan NV. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics*. 2008 jun;24(17):1951–1952. Available from: <https://doi.org/10.1093%2Fbioinformatics%2Fbtn328>.
- [61] Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA*. 2008 apr;14(6):1164–1173. Available from: <https://doi.org/10.1261%2Frna.894608>.
- [62] Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. *Folding and Design*. 1998 nov;3(6):577–587. Available from: <https://doi.org/10.1016%2Fs1359-0278%2898%2900072-8>.
- [63] Das R, Baker D. Automated de novo prediction of native-like RNA tertiary structures. *Proceedings of the National Academy of Sciences*. 2007 aug;104(37):14664–14669. Available from: <https://doi.org/10.1073%2Fpnas.0703836104>.
- [64] Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, et al. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012 feb;18(4):610–625. Available from: <https://doi.org/10.1261%2Frna.031054.111>.
- [65] Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, et al. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA*. 2015 apr;21(6):1066–1084. Available from: <https://doi.org/10.1261%2Frna.049502.114>.
- [66] Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, et al. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*. 2017 jan;p. rna.060368.116. Available from: <https://doi.org/10.1261%2Frna.060368.116>.
- [67] Liu P, Agrafiotis DK, Theobald DL. Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*. 2009;p. n/a–n/a. Available from: <https://doi.org/10.1002%2Fjcc.21439>.
- [68] Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A. A study of quality measures for protein threading models. *BMC Bioinformatics*. 2001;2(1):5. Available from: <https://doi.org/10.1186%2F1471-2105-2-5>.
- [69] Zok T, Popenda M, Szachniuk M. MCQ4Structures to compute similarity of molecule structures. *Central European Journal of Operations Research*. 2013 apr;22(3):457–473. Available from: <https://doi.org/10.1007%2Fs10100-013-0296-5>.
- [70] Magnus M, Matelska D, Łach G, Chojnowski G, Boniecki MJ, Purta E, et al. Computational modeling of RNA 3D structures, with the aid of experimental restraints. *RNA Biology*. 2014 apr;11(5):522–536. Available from: <https://doi.org/10.4161/rna.28826>.