Recent advances in RNA folding

Jörg Fallmann[a], Sebastian Will[b], Jan Engelhardt[a], Björn Grüning[c], Rolf Backofen[c], Peter F. Stadler[a,b,d,e,f]

[a] *Bioinformatics Group, Department of Computer Science; and Interdisciplinary Center for Bioinformatics, University of Leipzig*
Härtelstraße 16-18, D-04107 Leipzig, Germany
[b] *Institute for Theoretical Chemistry, University of Vienna*
Währingerstraße 17, A-1090 Wien, Austria
[c] *Bioinformatics, University of Freiburg*
Georges-Köhler-Allee 106, 79110 Freiburg, Germany
[d] *RNomics Group, Fraunhofer Institute for Cell Therapy and Immunology*
Perlickstraße 1, 04103, Leipzig, Germany
[e] *Santa Fe Institute*
, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA
[f] Corresponding author: e-mail studla@bioinf.uni-leipzig.de

Abstract

Secondary structure is the natural level of coarse in the realm of nucleic acid structures. It forms a conceptually important intermediate level of description and explains the dominating part of the free energy of structure formation. Secondary structures are well conserved over evolutionary time-scales and for many classes of RNAs evolve slower than the underlying primary sequence. Given the close link between structure and function, secondary structure is routinely used as a basis to explain experimental findings. Recent technological advances, finally, have made it possible to assay secondary structure directly using high throughput methods. From a computational biology point of view, secondary structures have a special role because they can be computed efficiently using exact dynamic programming algorithms.

Keywords:

1. Introduction

Structure, in particular evolutionarily conserved structure is an excellent predictor of biological function. This is true for all classes of biopolymers, including

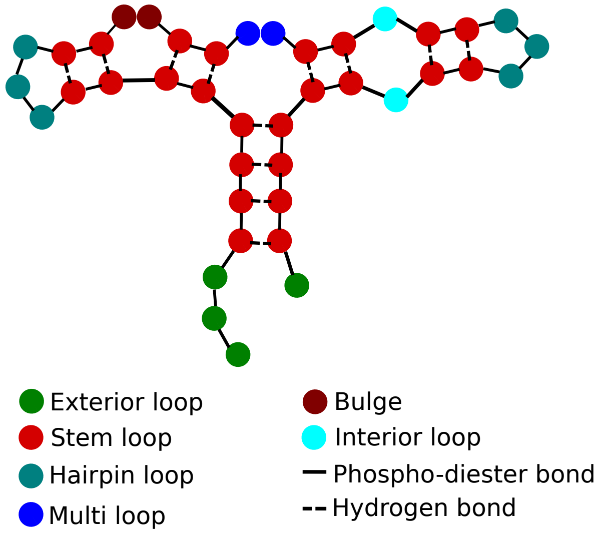

Figure 1: Decomposition of an RNA secondary structure into structural elements. Energy contributions are assigned to the plane faces depending on their type and the nucleotides delimiting them. “Stem loops”, i.e., stacked pairs of base pairs, from the unit of helices.

proteins and nucleic acids. The physics of structure formations, however, differs substantially between proteins and nucleic acids. The dominating process in protein folding is global, driven by hydrophobic forces. RNAs, on the other hand, exhibit a hierarchical folding process, where base pairs and thus helices, are rapidly formed, while the spatial arrangement of complex tertiary structures usually is a slow process.

RNA secondary structure elements (see Fig. 1 for an overview) are formed via intramolecular interactions of nucleotides. Such interactions form base-pairs via hydrogen bonds between corresponding nucleotides, enforcing restrictive local geometries. The standard set of RNA base-pairs (AU,GC) is known as Watson-Crick-base-pairs , named after the famous discoverers of DNAs double-helical structure [1]. GC-base-pairs can form three hydrogen bonds between their Watson-Crick edges, while AU-base-pairs can only form two. This is important considering their energy contributions, which is higher for GC- than for AU-base-pairs . The main part of the interaction energy, however, is contributed by the stacking interaction of the π-electron systems of the aromatic rings of the nucleobases. These energy contributions are large compared to the effects of hydrogen bonding. As a consequence, almost all RNAs form highly stable, well-defined secondary structures, while protein structures often remain flexible or are only marginally stable at room temperature [2].

At a more detailed level, other interactions between nucleotides beyond canonical base-pairs contribute to structure formation. Most prominently, GU wobble-base pairs regularly appear in native RNA structures. RNA bases not only interact via the "standard" Watson-Crick-edge. Instead, they can also form bonds between their Hoogsteen- or CH-edge and their Sugar-edge. These edges even allow the formation of base-pairs between three bases at once, known as base triplets, influencing the stability of helices and tertiary as well as quaternary structures. Long range interactions like pseudo-knots or kissing hairpins also contribute to RNA secondary structure formation. This form of intramolecular base-pairing happens when a stem or loop region interacts with another non-adjacent stem or loop region.

In this contribution we provide a short overview of the RNA folding algorithms and recent additions and variations. We briefly introduce current extensions beyond the basic secondary structure model and address methods to align, compare, and cluster RNA structures. The contribution ends with a tabular summary of the most important software suites in the fields, many of which are already integrated in the Galaxy-RNA-workbench [3].

2. Basic Secondary Structure Prediction Algorithms

The dominance of base stacking and loop entropies as energetic contribution and the restriction to a single interaction partner enables a purely combinatorial description of RNA (and DNA) secondary structures, and thus to completely ignore both, the atom-scale details and the actual spatial embedding of the molecule. Formally, an RNA secondary structure is simply a (labeled) graph whose nodes represent entire nucleotides and whose edges denote base pairs, so that

1. edges are formed only between nucleotides that form Watson-Crick or GU base pairs;
2. no two edges emanate from the same vertex, i.e., from the mathematical point of view, a secondary structure is a matching;
3. edges span at least three unpaired bases;
4. if the vertices are placed in $5'$ to $3'$ order on the circumference of a circle and edges are drawn as straight lines, no two edges cross.

The last condition ensures that the graph is outerplanar and therefore excludes so-called pseudo-knots, to which we will briefly return below.

Over the last two decades an additive energy model known as the "Turner parameters" has become the well-tested standard model for the energy of an RNA secondary structure. It stipulates that relevant energetic contribution are the stacking of base pairs, the entropic strain of loops, as well as partial stacking of unpaired bases at the ends of helical regions (usually referred to as dangling ends). These have been tabulated as function of the sequence compositions of stacked pairs and loops respectively, based on a wealth of detailed experimental evidence.

The dynamic programming approach to RNA secondary structure prediction relies on the fact that structures can be recursively decomposed into smaller components with independent energy contributions. In each of the decomposition steps only a single loop (or stacking of two consecutive base pairs) needs to be evaluated. Fig. 2 outlines this scheme in a graphical manner. This decomposition scheme has the form of a context free grammar. In the simplest model, Nussinov's maximum circular matching [5], the paired contribution C is interpreted as a single base pair around an arbitrary structure F. The more realistic Turner model requires a somewhat more complex grammar, distinguishing hairpin loops, interior loops (including stacking base pairs as a special case), and multi-branch loops. Again we refer to the literature for the details.

The grammar, whose exact form depends on the structural building blocks

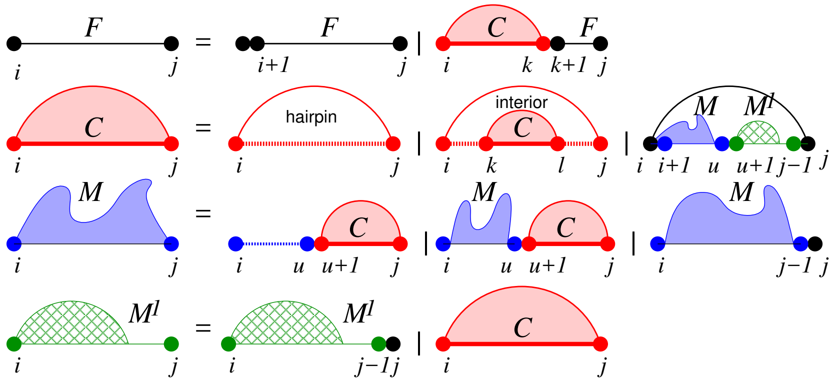

Figure 2: The classical recursions of the standard model of RNA folding (Drawing from [4]). The hieroglyphic symbols denote different types of RNA secondary structures: F is an arbitrary secondary structures, C a structure enclosed by a base pair, and M and M^1 denote components of multibranch loops. We refer to [4] and the references therein for a detailed description of the algorithms.

that are associated with energy contributions, pertains an identical form to the computation of the minimum free energy structure [6, 7], the partition function [8] or the density of states [9]. These algorithmic variants differ only in the way how the individual steps of the recursion are evaluated, i.e., whether energies are minimized, partition functions are summed, or histograms are convoluted over alternative decompositions. Instead of experimentally measured parameters, one can also employ machine learning techniques to infer parameters from training sets of known structures [10]. The machine learning approaches, usually phrased as stochastic context free grammars (SCFGs) [11], can afford more freedom in the choice of the details of the decomposition model [12].

Generic variations on the algorithms have been designed to retrieve a large collection of sub-optimal structures [13] instead of only a single representative minimum free energy structure. The exact computation of partition functions not only provides access to equilibrium base pairing probabilities but also to melting temperatures and specific heat profiles [14].

3. Variations on the Theme

3.1. Secondary Structures

Local Secondary Structures. RNAs much beyond the length of ribosomal RNAs presumably do not fold into their global minimum but form locally stable structural domains. This effect can be modeled by restricting the maximal span L of base pairs. This approach not only yields more plausible structure predictions, it also drastically increases the computational efficiency. The "scanning versions" [15, 16] of the standard folding recursions require only $O(nL^2)$ time and $O(n + L^2)$ space, where n is the sequence length. This makes them fast enough for genome and transcriptome-wide approaches. In [17], optimized parameter for local folding of mRNA were introduced. On a large set of benchmark, this work could also show that local folding is preferable to global folding for mRNAs.

Centroids and their Relatives. Centroids are structures with a minimum distance to all other structures in the ensemble of possible structures. Together with Maximum Expected Accuracy structures, which contain a maximal number of base pairs with high probability, they provide a measure for the confidence for a predicted structure, more details can be found in section 5.

Consensus Structures. Given a good alignment of a collection of related RNA structures, their consensus structure, i.e., a set of base pairs at corresponding alignment positions can be computed using the same dynamic programming approach. To this end, RNAalifold [18, 19] simply adds the sequence-dependent energy contribution over alignment columns in each evaluation of the energy model. The use of alignments as input considerably improves the accuracy of the predicted secondary structures. Consensus structure predictions are not only of interest in their own right but also form the basis for statistical measures of RNA secondary structure conservation [20, 21].

3.2. RNA Folding with Constraints

Although the Turner energy model provides a surprisingly accurate approximation of the RNA folding energies, it is not perfect. On the one hand, the energy parameters, which have been estimated by regression from large numbers of melting experiments, are afflicted by residual measurement errors. On the other hand, the secondary structure model is not perfect and neglects many weak interactions. As a consequence, secondary structure predictions are far from perfect. It is of great interest therefore to guide the prediction procedure with external information. This can be done in two ways: either by constraining the set of allowed structures using hard constraints or by encouraging or discouraging certain structural features with the help of bonus energies. Recently a generic framework to handle both types of constraints has been incorporated into the ViennaRNA Package [22].

Hard constraints either enforce or prevent pairing of a certain base or base-pair , usually implemented as high energy penalties. A less harsh way to implement constraints is to reward or penalize structures that match or contradict available information via moderate pseudo-energy terms, so called *soft constraints*. The latter can be set in proportion to some measure of confidence or signal strength.

In general, constraints become of interest in scenarios where RNAs interact, either with other RNAs, proteins, or ligands. Hard constraints can be used to model the exposition of binding sites, rendering them either accessible, or inaccessible for interaction partners. Soft constraints can be used to fine-tune RNA secondary structure predictions by incorporating chemical or enzymatic "reactivities" either directly, as energy contributions/penalties, or by minimizing the deviations between predicted and measured signal. In particular, the inclusion of SHAPE reactivities has been studied in much detail by several groups. A recent addition to the ViennaRNA package implements the most commonly used options [22]. These methods have become applicable to genome-wide surveys of condition-dependent secondary structure changes. An example is a recent study of temperature dependence of structures in bacterial pathogens [23].

RNA molecules *in vivo* usually interact with multiple partners simultaneously. These interactions can influence each other even if there is no direct competition of the same or overlapping binding sites since competitive and cooperative effects can be mediated by structural changes that can unblock or block previously paired or accessible regions. The magnitude of such effects can be computed when free energy of a RNA molecule bound by two interaction partners is derived from the difference $\Delta\Delta G$ between the sum of the energy of both partners interacting separately and the end state and ground state [24]. A negative value of $\Delta\Delta G$ indicates antagonistic binding effects, a positive $\Delta\Delta G$ indicates cooperative effects. Such effects can efficiently be modeled using the constraint folding option in the ViennaRNA package 2.0 [4], where a pair of binding sites constraints the structure ensemble by forcing these sites inaccessible.

Regional Accessibility. A parameter that is crucial for the analysis of interactions of RNAs with proteins or other nucleic acids is energy necessary to expose a local binding site region to the partner. It is of crucial importance for example in the context of microRNA/mRNA binding, siRNA efficiency, or bacterial sRNA function. This opening energy is conceptually the difference between the free energy of the equilibrium ensemble and the free energy of an ensemble constrained to leave a known binding site unpaired. Instead of using constrained folding framework to compute accessibilites for each individual region, it is possible to compute accessibilities for all intervals simultaneously using a much more efficient dedicated variation of the folding algorithms [25, 26].

3.3. RNA-RNA and RNA-Protein Interactions

The multi-faceted regulatory machinery of gene expression is based on the interplay between RNA and regulatory factors like other RNAs or proteins. It is crucial for the balance between synthesis (transcription), translation, transport and decay of mRNAs, ncRNAs and proteins to modulate the spatial-temporal expression of RNA molecules. Hundreds of RNA binding proteins and even more miRNAs are encoded in the human genome [27, 28, 29], emphasizing their role in gene regulation and thus the vitality of organisms. The extreme versatility of RNA molecules in terms of sequence and structure features and the complexity of RNA binding domains and binding preferences of proteins raise the need for advanced and efficient algorithms for interaction analysis.

There are basically two different approaches for determining the interaction between two RNAs that takes into account both the sequence and structure of the participating RNAs. The first type of approaches defines the search for an RNA-target as the problem of predicting a common stable structure for the two interacting RNAs. This is in general an NP-complete problem [30]. Thus, existing approaches implement a partial structure model that can predict a certain class of interactions. The simplest model is implemented in RNAcofold [31], where only the class of nested interactions are considered, resulting in a complexity of $O(n^3)$ due to its similarity with normal RNA structure prediction.

However, many functional interactions such as kissing hairpins are not covered in this model. This led to the development of several extended structural models that provided a compromise between complexity and the structural class covered. As shown in several publications, excluding so-called zig-zag interactions does make the problem solvable in polynomial time. Roughly speaking, zig-zag interactions are structures where at least two inter-molecular base-pairs are covered by one intra-molecular base-pairs in one sequence, and two non-nested base-pairs in the other sequence in a way that disallows the split into two separate interaction sites. Once these interactions are excluded, the minimum free energy interaction structure can be predicted in several energy models [32, 30] in $O(n^6)$ time. Even the partition function and associated quantities such as melting temperature and base-pairing probabilities for inter-molecular base-pairs can be predicted with the same complexity [33, 34].

Albeit these approaches solve the problem of RNA-RNA interactions with kissing hairpins in polynomial time, the complexity of $O(n^6)$ time is too high for genome-wide screens. Here, accessibility-based approaches improve the situation while still being able to predict complex interactions like kissing hairpins. A region in an RNA structure is called *accessible* if it is free from internal structure. The energy required to make the interaction site accessible can be determined in a modified partition function approach for the individual sequences in cubic time [35]. RNAup [36] then combines this accessibility term for two interaction sites with the best energy for the duplex-formation for these sites, yielding an $O(n^2w^2)$ approach for target prediction, where w is the maximal length of the interaction sites. The resulting score corresponds to the partition function of all interacting structures that have the same duplex. IntaRNA [37, 38] reduces this runtime to $O(n^2)$ for the final duplex calculation using a heuristics for the right end of the interaction site. By combining this with a seed-based approach, the prediction quality is nearly the same as for RNAup . RNAplex is an even faster approach that uses a heuristic version for the calculation of accessibility. The energy required to make a region accessible is directly related to the probability that this region is free in the ensemble of all structures. This probability is now approximated in RNAplex using a Markov chain with limited memory.

One additional problem is that RNAup , IntaRNA and RNAplex predict only one continuous interaction site. However, there have been interactions experimentally validated that consist of several such interactions. There have been several approaches of different complexity to extend the accessibility concept to this extended class of interactions [39, 40].

The aforementioned approaches do not rely on conservation, which could drastically reduce the inherently high false positive rate for target prediction. One possibility for taking conservation into account is to use an alignment-folding approach as in RNAalifold (see above). Here, one predicts interactions between two different alignments [41, 42]. However, as shown in [43], the interaction sites is not necessarily conserved, especially on mRNAs. CopraRNA [44] does not attempt to predict conserved interaction sites, but conserved interactions by combining evidence for the interaction between two RNAs in different species. A recent benchmark on sRNA target prediction shows that CopraRNA clearly outperforms other target prediction tools. However, CopraRNA is limited to RNAs where conservation information is available.

While RNA-RNA interactions are directly related to RNA folding, this is not the case for the prediction of targets of RNA-binding proteins (RBPs). Instead, the approaches for finding binding sites of RBPs are more related to finding motifs in a set of bound sequences, which is the data provided by SELEX and CLIP experiments. Due to the similarity between this problem and the task of finding binding sites of transcription factors, motif discovery tools like MEME [45] have been used frequently. However, as already shown, one cannot ignore the contribution of the RNA secondary structure. Many RBPs for example prefer single-stranded regions as binding sites. Memeris [46] is an extension of MEME that uses accessibility as prior for motif discovery. RNAcontext [47] uses a physical energy model of motif binding that integrates structural information. Graphprot [48] extends the idea of k-mers with gaps to graphs, which are used to represent the folding of the binding sites and its context, using an efficient graph-kernel. It is currently one of the most reliable tools for predicting binding sites from CLIP data, as shown by several experimentally verified binding predictions [47, 49]. RNA secondary structure influences on RNA binding behavior of proteins has also been successfully used to discriminate actively bound sites from a list of potential binding sites [50]. This concept was one of the key motivations for the curation of AREsite2 [51], a database that combines genome wide motif annotation in human and several model organisms with RNA secondary structure and CLIP-derived binding site information. This serves as a basis for the analysis and prediction of RNA-protein interactions and their influence on RNA halflife.

3.4. RNA Gene Finding

Homology-based RNA gene finding. RNAs with conserved secondary structure are typically either short non-coding RNAs or relatively small structured domains that are part of larger transcripts. The short length, the small size of the nucleotide alphabet, and the usually relatively low level of sequence conservation conspire to make RNA homology search a difficult problem [52]. Still, the most commonly used tool is blastn and it works well in many circumstances. The conserved secondary structure of many RNA families, however, provides additional information that is harnessed by infernal to improve both sensitivity and specificity of the search [53]. Instead of single sequence, it starts from a structure-annotated multiple alignment, as available for many RNA families from the Rfam database [54]. The alignment is converted into a covariance model, a tree-like generalization of HMMs, which allows efficient search in genomic sequences. At present, infernal serves as *the* tool for RNA homology search.

De novo *detection of conserved RNA structure.* Our current knowledge of ncRNA genes is far from complete, however. Even in the age of efficient RNA-seq methods, it is still of interest to find evidence for evolutionarily conserved, and thus likely functional, RNA structure (see [55, 56] for recent reviews). Over the years, several types of tools have been devised for this purpose. QRNA [57] uses a fully probabilistic model and computes for a pairwise sequence alignment the posterior probabilities that it derived from a coding region, a conserved secondary structure, or neither. RNAdecoder [58] is an extension of this idea that considers the superposition of RNA structure and coding region. Tools such as AlifoldZ [59], SissiZ [60], and RNAz [61, 62, 63] start from multiple sequence alignments and evaluate descriptors such as folding energies and sequence diversity to decide whether the alignment harbors a conserved structure or not. To make this decision, RNAz , for example, uses a support vector machine trained from large sets of structured RNAs and shuffled decoys. cmfinder [64] considers a set of related, but unaligned sequences and their predicted secondary structures. Together with anchors of sequence similarity these are used to build CMs with the help of infernal , which in turn are used to search for further matches, which are used in an interactive expectation maximization step to refine the CM, whose significance is then evaluated. A common issue with all *de novo* RNA gene finders is a relatively high false discovery rate that needs to be estimated by comparing the foreground data with a control, which is usually constructed by column-wise shuffling of the input alignments.

4. Beyond the Standard Model

Folding in "$2\frac{1}{2}D$". Several structural motifs go well beyond the secondary structure model but can still be accommodated in the same computational framework. This pertains in particular to local motifs. Well-studies examples are G-quadruplexes [65] and local 3D-motifs such as kink-turns [66]. Computationally these are treated like special loop types, which is made easy by the constraint handling framework in the ViennaRNA package. In addition, however, motif specific energy models are necessary to handle these cases consistently. So far, these are only available for some motif classes.

Folding in the Leontis-Westhof Representation. The Leontis-Westhof representation of RNA structures goes beyond secondary structure in that it also accommodates all types of non-standard base pairs and classifies them by isostericity classes [67, 68]. This leads to a natural extension of the standard energy model in which interior and hairpin loops are decomposed further into small components delimited by non-standard base pairs; in addition, the energy model now takes into account that adjacent loop components strongly influence each other. This type of extended model serves as starting point for *de novo* 3D structure prediction tools such as mc-sym [69]. It can be dealt with by dynamic programming, albeit the recursions are substantially more involved than those of Fig. 2, see [70].

Pseudoknots. The topic of RNA pseudoknots has received much attention in the past, albeit to a large extent from a more theoretical and algorithmic point of view. There are several competing models describing the different classes of pseudoknotted structures, most of which fall into the realm of multi-context-free grammars (MCFGs) and can be handled by dynamic programming [11, 71], albeit at computational complexities that are prohibitive for molecules larger than a few hundred nucleotides. Enumerative approaches for non-MCFG classes of structures are discussed in [72]. At present, the practical applicability of pseudoknots is largely limited by accuracy of energy models, which have to be estimated from small sets of examples.

5. Comparison of RNAs Based on Secondary Structures

Tree Editing and Tree Alignment. Secondary structures are naturally represented in the "dot-parenthesis" notation, which consists of a pair of matching parentheses for each base pair and and dot for each unpaired position. The example of Fig. 1, for example reads

...((((((((...))..))..((.((...)).)))))).

Such expressions of nested parenthesis have a natural interpretation as rooted, ordered trees in computer science. In consequence, tree alignment and tree editing algorithms, which generalize familiar sequence alignment methods, can be adapted for comparing RNAs based on their sequence *and* structure [73, 74]. Both approaches were extended to multiple RNAs following the progressive alignment scheme [73, 75]. Furthermore, the tree-based approach can even be extended to pseudoknotted structures for a large variety of pseudoknot types [76, 77].

Such methods however, especially if based on tree-alignment, are very sensitive to the compared secondary structures. This limits their practical use for analyzing RNAs of a priori unknown structure, since secondary structures have to be predicted from the sequence of each single RNA.

Simultaneous Folding and Alignment. The quality of secondary structure prediction increases substantially, when the structure is computed from an alignment of related sequences. While sequences of high similarity can be aligned sufficiently well by traditional sequence alignment methods, such alignments tend to become inaccurate, when pairwise identities drop below about 60%; then compromising comparative structure prediction.

In such cases, the simultaneous computation of alignment and secondary structure folding, originally proposed by Sankoff [78], remedies this RNA structure analysis dilemma. In practice, the original Sankoff algorithm suffers from considerable computational cost, due to its extreme time complexity of $O(n^6)$ and overhead due to the computation of minimum energy in the Turner model.

The tool LocARNA [79] substantially improves computation time over the Sankoff algorithm by utilizing information from the single RNAs structure ensembles. Building on pairwise Sankoff-like simultaneous alignment and folding (SA&F), it aligns multiple structures following the progressive alignment scheme (realized in the tool mlocarna). LocARNA-P [80] takes LocARNA 's idea of fast SA&F to a new level by computing partition functions over simultaneous alignments and foldings. This allows the efficient computation of alignment reliability profiles as well as probabilistic consistency-transformation to improve the quality of multiple RNA alignments [80].

Due to their RNA ensemble-based optimizations, LocARNA and LocARNA-P reduce the time and space complexity over Sankoff's algorithm each by a quadratic factor (in sequence length n). Nevertheless, its $O(n^4)$ time complexity is still limiting. In practice, LocARNA tackles this by a series of further heuristics as well as alignment constraints.

Taking a different route, SPARSE [81] performs SA&F in a similar model as LocARNA—in fact, it improves structure prediction flexibility—without relying on prior knowledge or sequence-based heuristics, but reduces the time complexity of the alignment algorithm by another quadratic factor over LocARNA , resulting in $O(n^2)$ time. This is achieved by exploiting even more features of the RNAs' secondary structure ensembles.

In a variation of SA&F, which compares RNAs based on (simultaneously) predicted non-crossing structures of the RNAs, CARNA [82] computes alignments that optimize similarity across the entire secondary structure ensembles. This strategy allows pseudoknots and is potentially advantageous for alignments of multi-structure RNAs with complex dependencies. Another interesting SA&F-related problem is the prediction of local secondary structures with exactly matching sequences. As in SA&F, such structures are not known a priori but are predicted simultaneously to the comparison. This *simultaneous matching and folding* problem is efficiently solved by ExpaRNA-P [83] in $O(n^2)$ time and space. Due to the efficiency of this method, the exactly matching substructures enable fast analysis of very large RNAs and serve as anchors to speed up SA&F in LocARNA [83]. SPARS [81] is an extension of ExpaRNA-P that solves the complete SA&F problem in $O(n^2)$ time and space.

Measures of reliability. Prediction always includes some amount of uncertainty. For the user it is important to get some information on how reliable a prediction is, for a detailed review refer to [84]. In case of RNA secondary structure prediction, base pair probabilities and the partition function can be used to derive some measures for reliability. This includes *Ensemble Diversity*, which is the average distance of two structures drawn from the Boltzmann ensemble, in the simplest case the base pair distance. *Positional Entropy* captures whether a nucleotide is found mainly paired or unpaired. *Ensemble Centroids* are structures that minimize the weighted average (base-pair) distance to all other structures in the ensemble. *Maximum Expected Accuracy* structures are predicted by maximizing the number of correct base pairs.

Clustering of structured RNAs. As discussed before, finding new RNA genes is a hard problem. One important approach is to use computational screens for conserved structured RNAs. However, such screens result in large sets (typically several thousands if not hundered's of thousand) of putative ncRNAs. The main problem is to annotate these newly detected putative RNA genes. The detection of individual domains, as very successfully used in the annotation of protein coding genes, can currently not be applied for RNA genes due to the flexibility of the RNA structure.

The successful classification of known RNA-genes in families (i.e., RNAs related by evolution like tRNAs) and classes (i.e., RNAs related by same functional structure like miRNA and snoRNAs) has opened up a possibility for a structure-based annotation approach by clustering putative ncRNAs according their sequence *and* structure to detect new RNA classes. One possibility is to directly use the score produced by sequence-structure alignment as for the hierarchical clustering of RNAs [85, 79]. RNAclust [79] is a dedicated pipeline facilitating complete clustering of RNAs. SoupViewer allows to semi-automatically analyze such a complete RNA cluster tree, easing the otherwise manual process of inspection for potential misclassifications. However, the problem of this clustering approach is twofold. First, determining the similarity between two different RNAs in the clustering procedure is complex (at least $O(n^4)$ time). Second, this score has to be calculated for all pairs of RNAs, which restricts its application to a small sets of RNAs, typically in range of few thousands. This is circumvented by, so-called alignment-free approaches [86, 87] that avoid the calculation of a quadratic number of alignments and thus are able to cluster hundreds of thousand of RNAs. GraphClust [86] even avoids any quadratic step by using an inverse index based on a structure-aware hashing approach to determine dense RNA neighborhoods.

6. Tools and suites for RNA analysis

This section presents a collection of the most relevant RNA-centric software available. Table 6 lists a selection of tools or suites of tools which are concerned with RNA secondary structure prediction, design, homology and more. This collection of software enables researchers to investigate virtually all aspects of RNA biology. Although some tasks are covered by more than one program, they each have their specifics and features making them a valid contribution to this collection. The RNAshape [88] algorithm of the Bielefeld RNA tools for example, allows the abstraction of RNA secondary structure to a tree-like domain of shapes which integrates well with dynamic programming algorithms and avoids exponential explosion while providing a non-heuristic and complete account of properties of the molecule's folding space. Many of the tools presented here were curated in a collection, the Galaxy-RNA-Workbench , which provides users with a virtual box containing pre-installed versions of the tools. This workbench enables researches to investigate RNA in silico , even without detailed knowledge of the command line or the overhead of installation and dependency resolution, all in a dockerized galaxy instance.

Tool/Suite	Description	URL	Citation*
ViennaRNA	The ViennaRNA Package consists of a C-library and a set of RNA secondary structure related standalone programs covering topics like (sub)optimal structure prediction, RNA-RNA interaction analysis, energy evaluation, folding path identification, consensus structure prediction and structural alignments. It includes interactive command-line tools like RNAfold , RNAsubopt , RNAplfold , RNAalifold , RNAinverse and many more.	`http://rna.tbi.univie.ac.at/`	[4]
Freiburg RNA tools	The Freiburg RNA tools provides access to a series of RNA research tools for sequence-structure alignments (LocARNA , CARNA , MARNA), clustering (ExpaRNA), interaction prediction (IntaRNA , CopraRNA , metaMIR), identification of homologs (GLASSgo), sequence design (AntaRNA , INFORNA , SECISDesign) and many more tasks.	`http://rna.informatik.uni-freiburg.de/`	[89]
RNAsoft	RNAsoft is a collection of online services for the computational prediction and design of RNA/DNA structures. It provides access to PairFold , CombFold and RNA Designer .	`http://www.rnasoft.ca/`	[90]
mfold/UNAfold	UNAFold is a comprehensive software package for nucleic acid folding and hybridization prediction. It provides methods for folding of single-stranded RNA or DNA, or hybridization between two single-strands. It can compute partition functions, minimum free energy foldings or hybridizations and also suboptimal foldings.	`http://unafold.rna.albany.edu/?q=mfold`	[91]
RNAstructure	The RNAstructure suite provides many tools for tasks like the prediction of secondary structures, secondary structures common to two or more sequences as well as the prediction of bimolecular secondary structures.	`http://rna.urmc.rochester.edu/RNAstructureWeb/`	[92]
rtools	The rtools web server provides access to tools for RNA secondary structure prediction with and without homology information, pseudoknot and accessibility, as well as mutation change predictions.	`http://rtools.cbrc.jp/`	[93]
Bielefeld RNA tools	The RNA processing tools of Bielefeld work on RNA data and provide, among others, shape prediction/abstraction and hybridization solutions.	`https://bibiserv.cebitec.uni-bielefeld.de/rna`	[94]

Table 1: RNA-centric tools and suites * If no suite is available, the web server interface is considered

7. Concluding Remarks

RNA structure prediction is a fast evolving topic, both, in regard to computational as well as experimental methods. The recent emerge of experimental techniques for (high-throughput) capture of *in vivo* RNA secondary structures and RNA interactions further speeds up this process. In this review, we present an overview from general concepts of RNA secondary structure prediction to recent advances in computational RNA folding, which deal with existing challenges in the field and address new challenges introduced by experimentally derived structure constraints.

The major concept for the inclusion of experimental data into prediction of RNA secondary structure and influences of interactions is the definition of constraints. No matter if hard or soft constraints are used, the integration of experimental data has to be handled with care, as there is no guarantee that predictions become indeed more accurate.

However, the here presented tools and suites allow to investigate virtually all aspects of RNA secondary structure and thereby affected features. Most of them are either available as suites, or have web server interfaces that allow the non-commandline affine user to benefit from their features. In a collaborative effort, many of these tools have additionally been collected in the Galaxy-RNA-workbench , which makes them available in a virtualized box, featuring a Galaxy brand easy to use interface.

8. Acknowledgement

This work was supported by the BMBF-funded projects (FKZ 031A538A+B) within The German Network for Bioinformatics Infrastructure (de.NBI).

References

[1] J. Watson, F. Crick, Molecular structure of nucleic acids, Nature 171 (4356) (1953) 737–738.
URL http://www.nature.com/physics/looking-back/crick/

[2] M. J. Sippl, M. Ortner, M. Jaritz, P. Lackner, H. Flöckner, Helmholtz free energies of atom pair interactions in proteins, Folding & Design 1 (1996) 289–298.

[3] R. Backofen, J. Engelhardt, A. Erxleben, J. Fallmann, B. Grüning, U. Ohler, N. Rajewsky, P. F. Stadler, The german center for RNA-bioinformatics (RBC), J. Biotech.This issue.

[4] R. Lorenz, S. H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, I. L. Hofacker, ViennaRNA Package 2.0., Algorithms for molecular biology : AMB 6 (2011) 26. doi:10.1186/1748-7188-6-26.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3319429&tool=pmcentrez&rendertype=abstract

[5] R. Nussinov, a. B. Jacobson, Fast algorithm for predicting the secondary structure of single-stranded RNA., Proceedings of the National Academy of Sciences of the United States of America 77 (11) (1980) 6309–13.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=350273&tool=pmcentrez&rendertype=abstract

[6] M. Zuker, P. Stiegler, Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information., Nucleic Acids Research 9 (1) (1981) 133–148.
URL http://nar.oxfordjournals.org/cgi/content/abstract/9/1/133

[7] M. Zuker, D. Sankoff, RNA secondary structures and their prediction, Bulletin of Mathematical Biology 46 (4) (1984) 591–621.
URL http://www.springerlink.com/index/4122K5036L1Q7429.pdf

[8] J. S. McCaskill, The equilibrium partition function and base pair binding probabilities for RNA secondary structure., Biopolymers 29 (6-7) (1990) 1105–19. doi:10.1002/bip.360290621.
URL http://www.ncbi.nlm.nih.gov/pubmed/1695107http://www.hubmed.org/display.cgi?uids=1695107

[9] J. Cupal, C. Flamm, A. Renner, P. F. Stadler, Density of states, metastable states, and saddle points. Exploring the energy landscape of an RNA molecule, in: T. Gaasterland, P. Karp, K. Karplus, C. Ouzounis, C. Sander, A. Valencia (Eds.), Proceedings of the ISMB-97, AAAI Press, Menlo Park, CA, 1997, pp. 88–91.

[10] C. B. Do, D. A. Woods, S. Batzoglou, CONTRAfold: RNA secondary structure prediction without physics-based models, Bioinformatics 22 (2006) e90–98.

[11] E. Rivas, R. Lang, S. R. Eddy, A range of complex probabilistic models for RNA secondary structure prediction that include the nearest neighbor model and more, RNA 18 (2012) 193–212.

[12] R. D. Dowell, S. R. Eddy, Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction, BMC Bioinformatics 5 (2004) 71.

[13] S. Wuchty, W. Fontana, I. L. Hofacker, P. Schuster, Complete suboptimal folding of RNA and the stability of secondary structures., Biopolymers 49 (2) (1999) 145–65. doi:10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G.
URL http://www.ncbi.nlm.nih.gov/pubmed/10070264

[14] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, P. Schuster, Fast folding and comparison of RNA secondary structures, Monatsh. Chem. 125 (1994) 167–188.

[15] I. L. Hofacker, B. Priwitzer, P. F. Stadler, Prediction of locally stable RNA secondary structures for genome-wide surveys, Bioinformatics 20 (2004) 191–198.

[16] S. H. Bernhart, I. L. Hofacker, P. F. Stadler, Local RNA base pairing probabilities in large sequences, Bioinformatics (Oxford, England) 22 (5) (2006) 614–5. `doi:10.1093/bioinformatics/btk014`.
URL `http://www.ncbi.nlm.nih.gov/pubmed/16368769`

[17] S. J. Lange, D. Maticzka, M. Möhl, J. N. Gagnon, C. M. Brown, R. Backofen, Global or local? Predicting secondary structure and accessibility in mRNAs, Nucleic Acids Res 40 (12) (2012) 5215–26, sJL and DM contributed equally to this work. `doi:10.1093/nar/gks181`.

[18] I. L. Hofacker, M. Fekete, P. F. Stadler, Secondary structure prediction for aligned RNA sequences, J. Mol. Biol. 319 (2002) 1059–1066.

[19] S. H. Bernhart, I. L. Hofacker, S. Will, A. R. Gruber, P. F. Stadler, `RNAalifold`: improved consensus structure prediction for RNA alignments, BMC Bioinformatics 9 (2008) 474.

[20] A. R. Gruber, S. H. Bernhart, I. L. Hofacker, S. Washietl, Strategies for measuring evolutionary conservation of RNA secondary structures, BMC Bioinformatics 9 (2008) 122.

[21] J. Gorodkin, I. L. Hofacker, From structure prediction to genomic screens for novel non-coding RNAs, PLoS Comput Biol 7 (2011) e1002100. `doi:10.1371/journal.pcbi.1002100`.

[22] R. Lorenz, D. Luntzer, I. L. Hofacker, P. F. Stadler, M. T. Wolfinger, SHAPE directed RNA folding, Bioinformatics 32 (2016) 145–147. `doi:10.1093/bioinformatics/btv523`.

[23] F. Righetti, A. M. Nuss, C. Twittenhoff, S. Beele, K. Urban, S. Will, S. H. Bernhart, P. F. Stadler, P. Dersch, F. Narberhaus, The temperature-responsive RNA structurome of *Yersinia pseudotuberculosis*, Proc. Natl.

Acad. Sci. USA 113 (2016) 7237–7242. doi:doi:10.1073/pnas.1523004113.

[24] Y.-H. Lin, R. Bundschuh, RNA structure generates natural cooperativity between single-stranded rna binding proteins targeting 5 and 3UTRs, Nucleic Acids Res 43 (2015) 1160–1169.

[25] A. Busch, A. S. Richter, R. Backofen, IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, Bioinformatics 24 (2008) 2849–2856.

[26] U. Mückstein, H. Tafer, S. H. Bernhart, M. Hernandez-Rosales, J. Vogel, P. F. Stadler, I. L. Hofacker, Translational control by RNA-RNA interaction: Improved computation of RNA-RNA binding thermodynamics, in: M. Elloumi, J. Küng, M. Linial, R. Murphy, K. Schneider, C. Toma (Eds.), Bioinformatics Research and Development, Vol. 13 of Communications in Computer and Information Science, Springer-Verlag Berlin Heidelberg, 2008, pp. 114–127.

[27] A. G. Baltz, M. Munschauer, B. Schwanhusser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, M. Landthaler, The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts., Molecular cell 46 (5) (2012) 674–90. doi:10.1016/j.molcel.2012.05.021.
URL http://www.ncbi.nlm.nih.gov/pubmed/22681889

[28] A. Castello, B. Fischer, K. Eichelbaum, R. Horos, B. Beckmann, C. Strein, N. Davey, D. Humphreys, T. Preiss, L. Steinmetz, J. Krijgsveld, M. Hentze, Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins, Cell 149 (6) (2012) 1393–1406. doi:10.1016/j.cell.2012.04.031.
URL http://linkinghub.elsevier.com/retrieve/pii/S0092867412005764

[29] D. Ray, H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. A. Smith, C. A. Yarosh, S. M. Kelly, B. Nabet, D. Mecenas, W. Li, R. S. Laishram, M. Qiao, H. D. Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. A. Anderson, K. W. Lynch, L. O. F. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q. D. Morris, T. R. Hughes, A compendium of RNA-binding motifs for decoding gene regulation, Nature 499 (7457) (2013) 172–177. doi:10.1038/nature12311. URL http://www.nature.com/doifinder/10.1038/nature12311

[30] C. Alkan, E. Karakoç, J. H. Nadeau, S. C. Sahinalp, K. Zhang, RNA-RNA interaction prediction and antisense RNA target search, J Comput Biol 13 (2) (2006) 267–282.

[31] S. H. Bernhart, H. Tafer, U. Mückstein, C. Flamm, P. F. Stadler, I. L. Hofacker, Partition function and base pairing probabilities of RNA heterodimers, Algorithms Mol Biol 1 (1) (2006) 3.

[32] D. D. Pervouchine, IRIS: intermolecular RNA interaction search, Genome Inform 15 (2) (2004) 92–101.

[33] F. W. D. Huang, J. Qin, C. M. Reidys, P. F. Stadler, Partition function and base pairing probabilities for RNA-RNA interaction prediction, Bioinformatics 25 (2009) 2646–2654.

[34] H. Chitsaz, R. Salari, S. C. Sahinalp, R. Backofen, A partition function algorithm for interacting nucleic acid strands, Bioinformatics 25 (2009) i365–i373.

[35] S. H. Bernhart, U. Mückstein, I. L. Hofacker, RNA Accessibility in cubic time, Algorithms Mol Biol 6 (1) (2011) 3. doi:10.1186/1748-7188-6-3.

[36] U. Muckstein, H. Tafer, J. Hackermuller, S. H. Bernhart, P. F. Stadler, I. L. Hofacker, Thermodynamics of RNA-RNA binding, Bioinformatics 22 (10) (2006) 1177–82. doi:10.1093/bioinformatics/btl024.

[37] A. Busch, A. S. Richter, R. Backofen, IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, Bioinformatics 24 (24) (2008) 2849–56. doi:10.1093/bioinformatics/btn544.

[38] P. R. Wright, J. Georg, M. Mann, D. A. Sorescu, A. S. Richter, S. Lott, R. Kleinkauf, W. R. Hess, R. Backofen, CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains, Nucleic Acids Res 42 (Web Server issue) (2014) W119–23, pRW, JG and MM contributed equally to this work. doi:10.1093/nar/gku359.

[39] H. Chitsaz, R. Backofen, S. C. Sahinalp, biRNA: Fast RNA-RNA binding sites prediction, in: S. Salzberg, T. Warnow (Eds.), Proc. of the 9th Workshop on Algorithms in Bioinformatics (WABI), Vol. 5724 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2009, pp. 25–36. doi:10.1007/978-3-642-04241-6.

[40] R. Salari, R. Backofen, S. C. Sahinalp, Fast prediction of RNA-RNA interaction, Algorithms Mol Biol 5 (2010) 5. doi:10.1186/1748-7188-5-5.

[41] S. E. Seemann, A. S. Richter, T. Gesell, R. Backofen, J. Gorodkin, PETcofold: predicting conserved interactions and structures of two multiple alignments of RNA sequences, Bioinformatics 27 (2) (2011) 211–219. doi:10.1093/bioinformatics/btq634.

[42] A. X. Li, M. Marz, J. Qin, C. M. Reidys, RNA-RNA interaction prediction based on multiple sequence alignments, Bioinformatics 27 (4) (2011) 456–63. doi:10.1093/bioinformatics/btq659.

[43] A. S. Richter, R. Backofen, Accessibility and conservation: General features of bacterial small RNA-mRNA interactions?, RNA Biol 9 (7) (2012) 954–65. doi:10.4161/rna.20294.

[44] P. R. Wright, A. S. Richter, K. Papenfort, M. Mann, J. Vogel, W. R. Hess, R. Backofen, J. Georg, Comparative genomics boosts target prediction for

bacterial small RNAs, Proc Natl Acad Sci USA 110 (37) (2013) E3487–96. doi:10.1073/pnas.1303248110.

[45] T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME SUITE: tools for motif discovery and searching, Nucleic Acids Res 37 (Web Server issue) (2009) W202–8. doi:10.1093/nar/gkp335.

[46] M. Hiller, R. Pudimat, A. Busch, R. Backofen, Using RNA secondary structures to guide sequence motif finding towards single-stranded regions, Nucleic Acids Res 34 (17) (2006) e117. doi:10.1093/nar/gkl544.

[47] H. Kazan, D. Ray, E. T. Chan, T. R. Hughes, Q. Morris, RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins, PLoS Comput Biol 6 (2010) e1000832. doi:10.1371/journal.pcbi.1000832.

[48] D. Maticzka, S. J. Lange, F. Costa, R. Backofen, GraphProt: modeling binding preferences of RNA-binding proteins, Genome Biol 15 (1) (2014) R17.
URL http://www.biomedcentral.com/content/pdf/gb-2014-15-1-r17.pdf

[49] Y. S. Niknafs, S. Han, T. Ma, C. Speers, C. Zhang, K. Wilder-Romans, M. K. Iyer, S. Pitchiaya, R. Malik, Y. Hosono, J. R. Prensner, A. Poliakov, U. Singhal, L. Xiao, S. Kregel, R. F. Siebenaler, S. G. Zhao, M. Uhl, A. Gawronski, D. F. Hayes, L. J. Pierce, X. Cao, C. Collins, R. Backofen, C. S. Sahinalp, J. M. Rae, A. M. Chinnaiyan, F. Y. Feng, The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression, Nat Commun 7 (2016) 12791. doi:10.1038/ncomms12791.

[50] V. Sedlyarov, J. Fallmann, F. Ebner, J. Huemer, L. Sneezum, M. Ivin, K. Kreiner, A. Tanzer, C. Vogl, I. Hofacker, P. Kovarik, Tristetraprolin binding site atlas in the macrophage transcriptome reveals a switch for

inflammationresolution, Molecular Systems Biology 12 (5) (2016) n/a–n/a. doi:10.15252/msb.20156628.
URL http://dx.doi.org/10.15252/msb.20156628

[51] J. Fallmann, V. Sedlyarov, A. Tanzer, P. Kovarik, I. Hofacker, AREsite2: an enhanced database for the comprehensive investigation of AU/GU/U-rich elements, Nucleic Acids Research 44 (D1) (2016) D90–D95. doi:10.1093/nar/gkv1238.
URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv1238

[52] P. Menzel, J. Gorodkin, P. F. Stadler, The tedious task of finding homologous non-coding RNA genes, RNA 15 (2009) 2075–2082.

[53] E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches, Bioinformatics 29 (2013) 2933–2935.

[54] E. P. Nawrocki, S. W. Burge, A. Bateman, J. Daub, R. Y. Eberhardt, S. R. Eddy, E. W. Floden, P. P. Gardner, T. A. Jones, J. Tate, R. D. Finn, Rfam 12.0: updates to the RNA families database, Nucl. Acids Res. 43 (2015) D130–D137.

[55] J. Gorodkin, I. L. Hofacker, E. Torarinsson, Z. Yao, J. H. Havgaard, W. L. Ruzzo, De novo prediction of structured RNAs from genomic sequences, Trends Biotechnol. 28 (1) (2010) 9–19.

[56] R. Backofen, W. R. Hess, Computational prediction of sRNAs and their targets in bacteria, RNA Biol 7 (1) (2010) 33–42.

[57] E. Rivas, S. R. Eddy, Noncoding RNA gene detection using comparative sequence analysis, BMC Bioinformatics 2 (2001) 8.

[58] J. S. Pedersen, I. M. Meyer, R. Forsberg, P. Simmonds, J. Hein, A comparative method for finding and folding RNA secondary structures within protein-coding regions, Nucleic Acids Res. 32 (2004) 4925–4936.

[59] S. Washietl, I. L. Hofacker, Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics, J Mol Biol 342 (2004) 19–30.

[60] T. Gesell, S. Washietl, Dinucleotide controlled null models for comparative RNA gene prediction, BMC Bioinformatics 9 (2008) 248. `doi:10.1186/1471-2105-9-248`.

[61] S. Washietl, I. L. Hofacker, P. F. Stadler, Fast and reliable prediction of noncoding RNAs, Proc. Natl. Acad. Sci. USA 102 (2005) 2454–2459.

[62] S. Washietl, I. L. Hofacker, M. Lukasser, A. Hüttenhofer, P. F. Stadler, Mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome, Nature Biotech. 23 (2005) 1383–1390.

[63] A. R. Gruber, S. Findeiß, S. Washietl, I. L. Hofacker, P. F. Stadler, `RNAz 2.0`: improved noncoding RNA detection, Pac. Symp. Biocomput. 15 (2010) 69–79.

[64] W. L. Ruzzo, J. Gorodkin, De novo discovery of structured ncRNA motifs in genomic sequences, Methods Mol. Biol. 1097 (2014) 303–318.

[65] R. Lorenz, S. H. Bernhart, J. Qin, C. Höner zu Siederdissen, A. Tanzer, F. Amman, I. L. Hofacker, P. F. Stadler, 2D meets 4G: G-quadruplexes in RNA secondary structure prediction, IEEE Trans. Comp. Biol. Bioinf. 10 (2013) 832–844, doi: 10.1109/TCBB.2013.7.

[66] J. A. Cruz, E. Westhof, Sequence-based identification of 3D structural modules in RNA with RMDetect, Nature Meth. 8 (2011) 513519. `doi:10.1038/nmeth.1603`.

[67] N. B. Leontis, J. Stombaugh, E. Westhof, The non-watson-crick base pairs and their associated isostericity matrices, Nucleic Acids Res 30 (16) (2002) 3497–3531.

[68] N. B. Leontis, A. Lescoute, E. Westhof, The building blocks and motifs of RNA architecture, Current Opinion Struct. Biol. 13 (2003) 300–308.

[69] M. Parisien, F. Major, The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data, Nature 452 (7183) (2008) 51–55.

[70] C. Höner zu Siederdissen, S. H. Berhart, P. F. Stadler, I. L. Hofacker, A folding algorithm for extended RNA secondary structures, Bioinformatics 27 (2011) i129–i137, iSMB.

[71] M. Riechert, C. Höner zu Siederdissen, P. F. Stadler, Algebraic dynamic programming for multiple context-free grammars, Theor. Comp. Sci. 639 (2016) 91–109. `doi:10.1016/j.tcs.2016.05.032`.

[72] C. Reidys, Combinatorial Computational Biology of RNA, Springer Verlag, Berlin, Heidelberg, 2011.

[73] M. Höchsmann, T. Töller, R. Giegerich, S. Kurtz, Local similarity in RNA secondary structures, in: Proc of the Computational Systems Bioinformatics Conference, Stanford, CA, August 2003 (CSB 2003), 2003, pp. 159–168.

[74] T. Jiang, J. Wang, K. Zhang, Alignment of trees — an alternative to tree edit, Theor. Comp. Sci. 143 (1995) 137–148.

[75] S. Siebert, R. Backofen, MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons, Bioinformatics 21 (16) (2005) 3352–9.

[76] M. Möhl, S. Will, R. Backofen, Lifting prediction to alignment of RNA pseudoknots, in: S. Batzoglou (Ed.), Proc.of the 13^{th} Annual International Conferences on Computational Molecular Biology (RECOMB'09), Vol. 5541 of LNBI, Springer, 2009, pp. 285–301.

[77] M. Möhl, S. Will, R. Backofen, Lifting prediction to alignment of RNA pseudoknots, J Comput Biol 17 (3) (2010) 429–42. `doi:10.1089/cmb.2009.0168`.

[78] D. Sankoff, Simultaneous solution of the rna folding, alignment and protosequence problems, SIAM Journal on Applied Mathematics 45 (5) (1985) 810–825.

[79] S. Will, K. Reiche, I. L. Hofacker, P. F. Stadler, R. Backofen, Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering, PLoS Computational Biology 3 (4) (2007) e65. doi:10.1371/journal.pcbi.0030065.
URL http://dx.plos.org/10.1371/journal.pcbi.0030065

[80] S. Will, T. Joshi, I. L. Hofacker, P. F. Stadler, R. Backofen, LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs, RNA 18 (5) (2012) 900–914. doi:10.1261/rna.029041.111.
URL http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.029041.111

[81] S. Will, C. Otto, M. Miladi, M. Möhl, R. Backofen, SPARSE: quadratic time simultaneous alignment and folding of RNAs without sequence-based heuristics, Bioinformaticsdoi:10.1093/bioinformatics/btv185.

[82] A. D. Palu, M. Möhl, S. Will, A propagator for maximum weight string alignment with arbitrary pairwise dependencies, in: Proceedings of the 16th International Conference on Principles and Practice of Constraint Programming (CP-2010), 2010, p. 8.

[83] C. Otto, M. Mohl, S. Heyne, M. Amit, G. M. Landau, R. Backofen, S. Will, ExpaRNA-P: simultaneous exact pattern matching and folding of RNAs, BMC Bioinformatics 15 (1) (2014) 6602. doi:10.1186/s12859-014-0404-0.

[84] R. Lorenz, M. T. Wolfinger, A. Tanzer, I. L. Hofacker, Predicting RNA secondary structures from sequence and probing data, Methods 103 (2016) 86–98. doi:10.1016/j.ymeth.2016.04.004.
URL http://linkinghub.elsevier.com/retrieve/pii/S1046202316300743

[85] J. H. Havgaard, E. Torarinsson, J. Gorodkin, Fast pairwise structural RNA alignments by pruning of the dynamical programming matrix, PLoS Comput. Biol. 3 (10) (2007) 1896–1908.

[86] S. Heyne, F. Costa, D. Rose, R. Backofen, GraphClust: alignment-free structural clustering of local RNA secondary structures., Bioinformatics (Oxford, England) 28 (12) (2012) i224–i232. doi:10.1093/bioinformatics/bts224.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3371856&tool=pmcentrez&rendertype=abstract

[87] S. A. Middleton, J. Kim, NoFold: RNA structure clustering without folding or alignment, RNA 20 (11) (2014) 1671–1683. arXiv:http://rnajournal.cshlp.org/content/early/2014/09/18/rna.041913.113.full.pdf+html, doi:10.1261/rna.041913.113.
URL http://rnajournal.cshlp.org/content/early/2014/09/18/rna.041913.113.abstract

[88] P. Steffen, B. Vo, M. Rehmsmeier, J. Reeder, R. Giegerich, Rnashapes: an integrated rna analysis package based on abstract shapes, Bioinformatics 22 (4) (2006) 500. doi:10.1093/bioinformatics/btk010.
URL +http://dx.doi.org/10.1093/bioinformatics/btk010

[89] C. Smith, S. Heyne, A. S. Richter, S. Will, R. Backofen, Freiburg RNA Tools: a web server integrating IntaRNA, ExpaRNA and LocARNA, Nucleic Acids Res 38 Suppl (2010) W373–7. doi:10.1093/nar/gkq316.

[90] M. Andronescu, R. Aguirre-Hernndez, A. Condon, H. H. Hoos, RNAsoft: a suite of RNA secondary structure prediction and design software tools, Nucleic Acids Research 31 (13) (2003) 3416–3422.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC169018/

[91] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Research 31 (13) (2003) 3406–3415.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC169194/

[92] J. S. Reuter, D. H. Mathews, RNAstructure: software for RNA secondary structure prediction and analysis, BMC Bioinformatics 11 (1) (2010) 129. doi:10.1186/1471-2105-11-129.
URL http://www.biomedcentral.com/1471-2105/11/129

[93] M. Hamada, Y. Ono, H. Kiryu, K. Sato, Y. Kato, T. Fukunaga, R. Mori, K. Asai, Rtools: a web server for various secondary structural analyses on single rna sequences, Nucleic Acids Research 44 (W1) (2016) W302. arXiv:/oup/backfile/Content_public/Journal/nar/44/W1/10.1093_nar_gkw337/3/gkw337.pdf, doi:10.1093/nar/gkw337.
URL +http://dx.doi.org/10.1093/nar/gkw337

[94] S. Janssen, R. Giegerich, The rna shapes studio, Bioinformatics 31 (3) (2014) 423. arXiv:/oup/backfile/Content_public/Journal/bioinformatics/31/3/10.1093_bioinformatics_btu649/2/btu649.pdf, doi:10.1093/bioinformatics/btu649.
URL +http://dx.doi.org/10.1093/bioinformatics/btu649