RNA-RNA interaction prediction based on multiple sequence alignments

Andrew X. Li, Manja Marz, Jing Qin, Christian M. Reidys

1 Center for Combinatorics, LPMC-TJKLC, Nankai University Tianjin 300071, P.R. China
2 RNA Bioinformatics Group, Philipps-University Marburg, Marbacher Weg 6, 34037 Marburg, Germany
3 Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany

Received on ****; revised on ****; accepted on ****

Associate Editor: *****

ABSTRACT

Motivation Many computerized methods for RNA-RNA interaction structure prediction have been developed. Recently, \( O(N^6) \) time and \( O(N^3) \) space dynamic programming algorithms have become available that compute the partition function of RNA-RNA interaction complexes. However, few of these methods incorporate the knowledge concerning related sequences, thus relevant evolutionary information is often neglected from the structure determination. Therefore, it is of considerable practical interest to introduce a method taking into consideration both thermodynamic stability and sequence covariation.

Results We present the a priori folding algorithm ripalign, whose input consists of two (given) multiple sequence alignments (MSA). ripalign outputs (1) the partition function, (2) base-pairing probabilities, (3) hybrid probabilities and (4) a set of Boltzmann-sampled suboptimal structures consisting of canonical joint structures that are compatible to the alignments. Compared to the single sequence-pair folding algorithm rip, ripalign requires negligible additional memory resource. Furthermore, we incorporate possible structure constraints as input parameters into our algorithm.

Availability The algorithm described here is implemented in C as part of the rip package. The supplemental material, source code and input/output files can freely be downloaded from [http://www.combinatorics.cn/cbpc/ripalign.html](http://www.combinatorics.cn/cbpc/ripalign.html).

Contact Christian Reidys duck@santafe.edu

Keywords multiple sequence alignment, RNA-RNA interaction, joint structure, dynamic programming, partition function, base pairing probability, hybrid, loop, RNA secondary structure.

1 INTRODUCTION

RNA-RNA interactions play a major role at many different levels of the cellular metabolism such as plasmid replication control, viral encapsidation, or transcriptional and translational regulation. With the discovery that a large number of transcripts in higher eukaryotes are noncoding RNAs, RNA-RNA interactions in cellular metabolism are gaining in prominence. Typical examples of interactions involving two RNA molecules are snRNAs [Forne et al. 1996]; snRNAs with their targets [Bachellerie et al. 2002]; micro-RNAs from the RNAi pathway with their mRNA target [Ambros 2004; Murchison and Hannon 2004]; snoRNAs from [Escherichia coli] (Hershberg et al. 2003; Repoila et al. 2003); and sRNA loop-loop interactions [Brunel et al. 2003]. The common feature in many noncoding RNA classes, especially prokaryotic small RNAs, is the formation of RNA-RNA interaction structures that are much more complex than the simple sense-antisense interactions.

As it is the case for the general RNA folding problem with unrestricted pseudoknots [Akutsu 2000], the RNA-RNA interaction problem (RIP) is NP-complete in its most general form [Alkan et al. 2006; Mneimneh 2009]. However, polynomial-time algorithms can be derived by restricting the space of allowed configurations in ways that are similar to pseudoknot folding algorithms [Rivas and Eddy 1999]. The simplest approach concatenates the two interacting sequences and subsequently employs a slightly modified standard secondary structure folding algorithm. The algorithms RNAcofold [Hofacker et al. 1994; Bernhart et al. 2006], pairfold [Andronescu et al. 2005], and NUFACK [Ren et al. 2009] subscribe to this strategy. A major shortcoming of this approach is that it cannot predict important motifs such as kissing-hairpin loops. The paradigm of concatenation has also been generalized to the pseudoknot folding algorithm of [Rivas and Eddy 1999]. The resulting model, however, still does not generate all relevant interaction structures [Chitsaz et al. 2009]. An alternative line of thought is to neglect all internal base-pairings in either strand and to compute the minimum free energy (MFE) secondary structure for their hybridization under this constraint. For instance, RNAduplex and RNAhybrid [Rehmsmeier et al. 2004] follows this line of thought. RNAup [Muckstein et al. 2008, 2008] and interRNA [Busch et al. 2008] restrict interactions to a single interval that remains unpaired in the secondary structure for each partner. These models have proved particularly useful for bacterial sRNA/mRNA interactions [Geissmann and Tootal 2004, Pervouchine 2004] and [Alkan et al. 2006] independently proposed MFE folding algorithms for predicting the joint structure of two interacting RNA molecules with polynomial time
complexity. In their model, a “joint structure” means that the intramolecular structures of each molecule are pseudoknot-free, the intermolecular binding pairs are noncrossing and there exist no so-called “zig-zags”, see supplement material (SM) for detailed definition. The optimal joint structure is computed in \( O(N^6) \) time and \( O(N^4) \) space via a dynamic programming (DP) routine.

A more reliable approach is to consider the partition function, which by construction integrates over the Boltzmann-weighted probability space, allowing for the derivation of thermodynamic quantities, like e.g. equilibrium concentration, melting temperature and base-pairing probabilities. The partition function of joint structures was independently derived by Chitsaz et al. (2009b) and Huang et al. (2009), while the base-pairing probabilities are due to Huang et al. (2009).

A key quantity here is the probability of hybrids, which cannot be recovered from base pairing probabilities since the latter can be highly correlated [Huang et al. 2010] presented a new hybrid-based decomposition grammar, facilitating the computation of the nontrivial hybrid-probabilities as well as the Boltzmann sampling of RNA-RNA interaction structures. The partition function of joint structures can be computed in \( O(N^6) \) and \( O(N^4) \) space and current implementations require very large computational resources. Salari et al. (2007) recently achieved a substantial speed-up making use of the observation that the external interactions mostly occur between pairs of unpaired regions of single structures. Chitsaz et al. (2009a) introduced tree-structured Markov Random Fields to approximate the joint probability distribution of multiple (≥3) contact regions.

Unfortunately, incompleteness of the underlying energy model, in particular for hybrid- and kissing-loops, may result in prediction inaccuracy. One way of improving this situation is to involve phylogenetic information of multiple sequence alignments (MSA).

In an MSA homologous nucleotides are grouped in columns, where homologous is interpreted in both: structural as well as evolutionary sense. I.e. a column of nucleotides occupies similar structural positions and all diverge from a common ancestral nucleotide. Also, many ncRNAs show clear signs of undergoing compensatory mutations along evolutionary trajectories. In conclusion, it seems reasonable to stipulate that a non-negligible component 3’–5’ orientation.

In the following we shall assume that a pair of RNA sequences can only interact if they belong to the same species. A pair \((R, S)\), can interact if for any row \(R^i\), there exist at least one row in \(S\) that can interact with \(R^i\).

Given a pair of interacting MSAs \((R, S)\), let \(m\) be the total number of potentially interacting pairs. ripalign exhibits a pre-processing step which generates a \(m \times N\)-matrix and a \(m \times M\)-matrix \(S\) such that \((R^i, S^\ell)\) range over all \(m\) potentially interacting RNA-pairs, see Tab. 1 and the SM, Section 2.1.

We shall refer in the following to \(R\) and \(S\) as MSAs ignoring the fact that they have multiple sequences.

We proceed by defining joint structures that are compatible to a fixed \((R, S)\). To this end, let us briefly review some concepts introduced in Huang et al. (2004).

A joint structure \(J(R, S, I)\) is a graph consisting of

(j1) Two secondary structures \(R\) and \(S\), whose backbones are drawn as horizontal lines on top of each other and whose arcs are drawn in the upper and lower halfplane, respectively. We consider \(R\) over a 3’ to 3’ oriented backbone \((R_1, \ldots, R_N)\) and \(S\) over a 3’ to 3’ oriented backbone \((S_1, \ldots, S_M)\) and refer to any \(R\)- and \(S\)-arc as interior arcs.

(j2) An additional set \(I\) of noncrossing arcs of the form \(R_i S_j, \text{exterior arc}\), where \(R_i\) and \(S_j\) are unpaired in \(R\) and \(S\).

(j3) \(J(R, S, I)\) contains no “zig-zags” (see SM).

The subgraph of a joint structure \(J(R, S, I)\) induced by a pair of subsequences \((R_i, R_{i+1}, \ldots, R_t)\) and \((S_h, S_{h+1}, \ldots, S_t)\) is denoted by \(J_{i,j,h,t}\). In particular, \(J(R, S, I) = J_{1,N,1,M}\) and \(J_{j,j,h,t} \subset J_{a,b,c,d}\) and if only if \(J_{j,j,h,t}\) is a subgraph of \(J_{a,b,c,d}\) induced by \((R_1, \ldots, R_t)\) and \((S_h, \ldots, S_t)\). In particular, we use \(S[i,j]\) to denote the subgraph of \(J_{1,3,1,3}\) induced by \((S_1, S_{i+1}, \ldots, S_t)\), where \(S[i,j] = S_1\) and \(S[i,j-1] = S_2\).

Table 1. Preprocessing in ripalign: Given a pair of MSAs \((R, S)\), where \(R\) consists of three aligned RNA sequences of species \(sp. \theta_1\) or \(\theta_2\), \(S\) in turn consists of four aligned sequences of species \(\theta_1\) and \(\theta_2\). Then we obtain the matrix-pair \((R, S)\), where \((R^1, S^\ell), 1 \leq i \leq 6\), ranges over all the six potentially interacting RNA-pairs.

| sp. | \(\theta_1\) | AGAACGGA | \(\theta_1\) | GGGCCG | \(\theta_1\) | AGAACGGA | \(\theta_1\) | GGGCCG |
|-----|-------------|-----------|-------------|--------|-------------|-----------|-------------|--------|
| \(\theta_1\) | AGAACGGA | \(\theta_1\) | AGUUAG | \(\theta_1\) | AGAACGGA | \(\theta_1\) | AGUUAG |
| \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA |
| \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA |
| \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA | \(\theta_2\) | AGAACGGA |

The symbol \(\theta\) represents a generalization of \(\theta\) to pairs of interacting MSAs and a new grammar of canonical interaction structures. The latter is of relevance since there are no isolated base pairs in molecular complexes.
introduce Watson-Crick, or

\[
\{ \delta_{i,j} \} \cup \{ \phi_{i,j} \} \in \{ \circ \cup \bigcirc \bigtriangleup \downarrow \} \cup \{ \triangle \downarrow \bigcirc \circ \bigtriangledown \}
\]

Given a joint structure, \( J_{a,b,c,d} \), a tight structure (TS), \( J_{i,j,h,\ell} \), \( \text{Huango et al. [2009]} \) is a specific subgraph of \( J_{a,b,c,d} \) indexed by its type \( \in \{ \circ \cup \bigcirc \bigtriangleup \downarrow \} \), see Fig. 1. For instance, we use \( J_{i,j,h,\ell} \) to denote a TS of type \( \square \).

A hybrid is a joint structure \( J_{i,j,h,\ell}^{\text{Hy}} \), i.e. a maximal sequence of intermolecular interior loops consisting of a set of exterior arcs \( (R_{i}, S_{i}, \ldots, R_{h}, S_{h}, \ell) \) where \( R_{i}, S_{i}, \ldots, R_{h}, S_{h} \) is nested within \( S_{i+1}, \ldots, S_{h+1} \) and where the internal segments \( R[j_{i} + 1, j_{h} - 1] \) and \( S[j_{h} + 1, j_{h} + 1 - 1] \) consist of single-stranded nucleotides only. That is, a hybrid is the maximal unbranched stem-loop formed by external arcs.

A joint structure \( J(R, S, I) \) is called canonical if and only if:

1. Each stack in the secondary structures \( R \) and \( S \) is of size at least two, i.e. there exist no isolated interior arcs,
2. Each hybrid contains at least two exterior arcs.

In the following, we always assume a joint structure to be canonical.

Next, we come to \( (R, S) \)-compatible joint structures. In difference to single sequence compatibility, this notion involves statistical information of the MSAs.

The key point consists in specifying under which conditions for interior arcs \( i,j \neq h,\ell \) the aligned sequences in \( R_{i}, R_{j}, S_{i}, S_{j} \) can pair. This is obtained by a generalization of the RNAfold approach \( \text{Hofacker et al. [2002]} \). We specify these conditions for interior \( (c_{i,j}^{R}), (s_{i,j}^{S}) \) and exterior pairs \( (c_{i,j}^{R}, S) \) in eq. (2.6).

For interior arcs \( (R_{i}, R_{j}) \), let \( X, Y \in \{ A, U, G, C \} \). Let \( f_{R}^{R}(X,Y) \) be the frequency of \( X, Y \) which exists in the 2-column sub-matrix \( (R_{i}, R_{j}) \) as a row-vector and

\[
C_{i,j}^{R} = \sum_{X,Y} f_{R}^{R}(X,Y)D_{X,Y}^{\text{R}}(X,Y)', \quad (2.1)
\]

Here \( XY \) and \( X'Y' \) independently range over all 16 elements of \( \{ A, U, G, C \} \times \{ A, U, G, C \} \) and \( D_{X,Y}^{\text{R}} = d_{H}(X,Y)' \), i.e. the Hamming distance between \( XY \) and \( X'Y' \) in case of \( XY \) and \( X'Y' \) being Watson-Crick, or GU wobble base pair and 0, otherwise.

Furthermore, we introduce \( \Phi_{i,j}^{R} \) to deal with the inconsistent sequences

\[
d_{i,j}^{R} = 1 - \frac{1}{m} \sum_{h} \left( \Phi_{i,j}^{R} + \delta(\Phi_{i,j}^{R}, \text{gap}) \delta(\Phi_{i,j}^{R}, \text{gap}) \right), \quad (2.2)
\]

where \( \delta(x,y) \) is the Kronecker delta and \( \Phi_{i,j}^{R} \) is equal to 1 if \( R_{i}, R_{j} \) are Watson-Crick or GU wobble base pair and 0, otherwise. Now we obtain \( B_{i,j}^{R} = C_{i,j}^{R} - \Phi_{i,j}^{R} \). Based on sequence data, the threshold for pairing \( B_{i,j}^{R} \) as well as the weight of inconsistent sequences \( \Phi_{i,j}^{R} \) are computed we have

\[
(c_{i,j}^{R}) B_{i,j}^{R} \geq B_{i,j}^{R} \quad (2.3)
\]

The case of two positions \( S_{i} \) and \( S_{j} \) is completely analogous

\[
(s_{i,j}^{S}) B_{i,j}^{S} \geq B_{i,j}^{S} \quad (2.4)
\]

where \( B_{i,j}^{S} \) and \( B_{i,j}^{R} \) are analogously defined.

As for \( (c_{i,j}^{R}, S) \) a further observation factors in: since many ncRNA show clear signs of undergoing compensatory mutations in the course of evolution...
In order to decompose canonical joint structures via the unambiguous grammar introduced in Section 4, we distinguish the two types (Type cc and Type c) of TS’s of type \(\varnothing\), \(\Delta\) or \(\Box\). Given a TS of type \(\varnothing\), denoted by \(J_{i,j,h,t}^{\varnothing}\), we write depending on whether \(R_{i+1}R_{j-1} \in J_{i,j,h,t}^{\varnothing}\) and \(J_{i,j,h,t}^{\varnothing}\), respectively. Analogously, we define \(J_{i,j,h,t}^{\Box,cc}\) and \(J_{i,j,h,t}^{\Box,c}\), see Fig. 5.

Fig. 3. Examples of two TS-types. We display \(\triangledown\), \(\square\), or \(\triangle\)-tight structures: Type cc (top) and Type c (bottom).

2.4 Probabilities and the Boltzmann Sampling

A dynamic programming scheme for the computation of a partition function implies a corresponding computation of probabilities of specific substructures is obtained “from the outside to the inside” and a stochastic programming algorithm \(\text{ripalign}\), whose input consists of a pair of interacting MSAs. \(\text{ripalign}\) requires only marginally more computational resources but is, without doubt, still computationally costly. Approximation algorithms are much faster, for instance PETcofold \(\text{Seemann et al., 2010}\), having a time complexity of \(O(m(N + M)^3)\), where \(m\) is the number of sequences in MSAs, \(N\) and \(M\) being the sequence lengths of the longer and shorter alignment, respectively, and \(n < N/2\) is the number of iterations for the adaption of the threshold value to find likely partial secondary structures. Their basic assumption is that the two secondary structures fold independently and that intra-loop evaluation differences are negligible. The flip-side of reducing the complexity of a folding problem by

3 RESULTS AND DISCUSSION

In this paper we present an \textit{a priori} \(O(N^6)\) time and \(O(N^4)\) space dynamic programming algorithm \(\text{ripalign}\), whose input consists of a pair of interacting MSAs. \(\text{ripalign}\) requires only marginally more computational resources but is, without doubt, still computationally costly. Approximation algorithms are much faster, for instance PETcofold \(\text{Seemann et al., 2010}\), having a time complexity of \(O(m(N + M)^3)\), where \(m\) is the number of sequences in MSAs, \(N\) and \(M\) being the sequence lengths of the longer and shorter alignment, respectively, and \(n < N/2\) is the number of iterations for the adaption of the threshold value to find likely partial secondary structures. Their basic assumption is that the two secondary structures fold independently and that intra-loop evaluation differences are negligible. The flip-side of reducing the complexity of a folding problem by
On the flip side, due to the gaps in seven out of eight subsequences, the hybrid increases by nearly 40%.

Comparing the prediction based on the MSAs (Fig. 6, middle) with the one induced by the secondary structure of fhlA based on the consensus sequence (Fig. 6, bottom), we observe:

(a) we incorporate evolutionary factors into the RNA-RNA interaction structure prediction via alignments as input,
(b) we introduce the grammar of canonical joint structures of interacting-alignments,
(c) we a priori factor in structural-constraints, like for instance, knowledge on Sm-binding sites.

Below we shall discuss (a), (b) and (e) in more detail in the context of concrete examples. All the MSAs involving in (a), (b) and (e) are listed in SM, Section 2.

(a): The fhlA/OxyS interaction

The OxyS RNA represses fhlA mRNA translation initiation through base-pairing with two short sequence [Agrawal and Altuvia 2000], one of which overlaps the ribosome binding sequence and the other resides further downstream, within the coding region of fhlA. Our algorithm predicts correctly both interaction sites based on MSAs, see Fig 6. In addition, most predicted stacks in the secondary structures of fhlA and OxyS agree well with the most frequent Boltzmann sampled structure. Two more hybrids, \(J^\text{Hy}_{i,j|h,l} \) and \(J^\text{Hy}_{i,j|I} \) are predicted in our output. The two additional contact regions, identified in the partition function, exhibit a significantly lower probability. An additional hairpin over two additional contact regions, identified in the partition function, exhibit slightly.

(b): The SmY-10/SL-1 interaction of C. elegans

MacMorris et al. (2007) stipulated that SmY-10 RNA, possible involved in trans-splicing, interacts with the splice leader RNA (SL1 RNA). In Fig 8 we show that the Sm-binding sites (colored in red) of the RNA molecules SmY-10 and SL-1 are \(R[56, 62] \) and \(S[25, 31] \), respectively. In Fig 8 we show that the Sm-binding sites are predicted by ripalign without any structure constraint. The hybrids listed in Column III are predicted by ripalign under the structural constraints that \(S'\text{-AUUUUUUG-3'}(R[56, 62]) \) and \(3'\text{-GUUUAAP-5'}(S[25, 31]) \) are Sm-binding sites (colored in red) in SmY-10 and SL-1, respectively. Here, we use \(J^\text{Hy}_{i,j|h,l} \) to denote the hybrid induced by \(R[i, j] \) and \(S[h, l] \).

(c): The U4/U6 interaction

Two of the snRNAs involved in pre-mRNA splicing, U4 and U6, are known to interact by base pairing [Zucker-Aprison et al. 1988]. We divided all known metazoan U4 and U6 snRNAs into three distinct groups and alignments: protostomia without insects, insects and deuterostomia [Martz et al. 2008; Martz et al. 2008] observed that insects behave in their secondary structure different from other protostomes, see Fig 8. Comparing all the predicted U4/U6 interactions, displayed in Fig 8 we can conclude:

(1) the secondary partial structures of the U4/U6 complex for all three groups predicted by ripalign agree predominantly with the described secondary structures in metazoans [Thomas et al. 1990; Tatch et al. 2002; Shamburek et al. 1994; López et al. 2008; Shukla et al. 2002], e.g. as depicted in Fig 8(top) for C. elegans [Zucker-Aprison et al. 1988].

(2) for all three groups, Stem I and II (Fig 8 top) are highly conserved. External ascendancies, such as protein interactions may stabilize stem II additionally.

(3) for all three groups, the 5' hairpin of U4 snRNA seems highly conserved to interact with the U6 snRNA. This RNA feature is not fully understood, since this element is also believed to contain intraloop interactions and may bind to a 15.6kDa protein Vidovic et al. 2003.

(4) for all metazoans, the U6 snRNA shows conserved intramolecular interactions between the 5' hairpin.

(5) for deuterostomes (Fig 8 bottom), with a contact-region probability of 45.5%, our algorithm identifies a third U4/U6 interaction, Stem III, to be conserved, which agrees with the findings in Iakub et al. 1997.

|   | I      | II    | III   |
|---|--------|-------|-------|
| 1 | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} |
| 2 | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} |
| 3 | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} |
| 4 | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} |
| 5 | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} |
| 6 | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} | J^\text{Hy}_{12|11|0} |

Table 2. Top 6 probable hybrids predicted by rip and ripalign: Interaction of two specific RNA molecules, SL1 and SmY-10 of Caenorhabditis elegans as illustrated in Fig. 8. The top 6 probable hybrids predicted by rip implemented by [Huang et al. 2010] is shown in column I. The hybrids listed in column II are predicted by ripalign without any structure constraint. The hybrids listed in Column III are predicted by ripalign under the structural constraints that 5'-AUUUUUUG-3'(R[56, 62]) and 3'-GUUUAAP-5'(S[25, 31]) are Sm-binding sites (colored in red) in SmY-10 and SL-1, respectively. Here, we use \(J^\text{Hy}_{i,j|h,l} \) to denote the hybrid induced by \(R[i, j] \) and \(S[h, l] \).
of the constrained stems based on the same MSAs showed in Fig. 6. Here, the prediction was performed (top) without and (bottom) with the extension of constrained stems. The assumption may cause, for instance, the existence of some interior hairpin. However, this interaction can also be assumed.

We want to thank Fenix W.D. Huang and Jan BICoC Benchmark Center, IBM and Kathy Tzeng of IBM Life Sciences Solutions Enablement. Their support was vital for all computations and BICoC Benchmark Center, IBM and Kathy Tzeng of IBM Life Sciences Solutions Enablement. Their support was vital for all computations.
Fig. 8. ripalign versus rip: Interaction of two specific RNA molecules, SLI and SmY-10 of Caenorhabditis elegans. The Sm-binding sites (colored in red) in the RNA molecules SmY-10 and SLI are 5’-AAUUUUUG-3’(R[56, 62]) and 3’-GUUUUUAA-5’(S[25, 31]), respectively. The joint structure contain a single interior arc \( R_{24} S_{67} \) (top) is predicted by rip implemented by [Huang et al. 2011]. The joint structure (middle) is predicted by ripalign without any structural constraint. The joint structure (bottom) is predicted by ripalign with a single structural constraint. The Sm-Y-10 or just by R. The Sm-binding sites prediction. In [56, 61.2%] and [55%], otherwise.

Argaman, L. and Altuvia, S. (2000) \( \beta \)I4A repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. J. Mol. Biol., 300, 1101–1112.

Bachellerie, J., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. Biochimie, 84, 775–779.

Bernhart, S., Hofacker, I., Will, S., Gruber, A. and Stadler, P. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics, 9, 474–487.

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

Brow, D. and Vidaver, R. (1995) An element in human U6 RNA destabilizes the U4/U6 splicesomeal RNA complex. RNA, 1, 122–131.

Brunel, C., Marquet, R., Romby, P. and Ehresmann, C. (2003) RNA loop-loop interactions as dynamic functional motifs. Biochimie, 84, 925–944.

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

Chitsaz, H., Backofen, R. and Sahinalp, S. C. (2009a) biRNA: Fast RNA-RNA binding sites prediction. In Proceedings of the 9th Workshop on Algorithms in Bioinformatics (WABI), volume 5724 of LNCS, pp. 25–36. Springer, Berlin / Heidelberg.

Zucker-Aprison et al. (1988)
Chitsaz, H., Saliari, R., Sahinalp, S. C. and Backofen, R. (2009b) A partition function algorithm for interacting nucleic acid strands. Bioinformatics, 25, i365–i373.

Ding, Y. and Lawrence, C. E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. Nucleic Acid Res., 31, 7280–7301.

Forne, T., Labouvier, E., Antoine, E., Rossi, F., Gallouzi, I., Cathala, G., Tazi, J. and Brunel, C. (1996) Structural features of U6 snRNA and dynamic interactions with other spliceosomal components leading to pre-mRNA splicing. Biochimie, 78, 434–442.

Gaspin, C. and Westhof, E. (1995) An interactive framework for RNA secondary structure prediction with a dynamical treatment of constraints. J. Mol. Biol., 254.

Geissmann, T. and Touati, D. (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. EMBO J., 23, 396–405.

Hershberg, R., Altuvia, S. and Margalit, H. (2003) A survey of small RNA-encoding genes in Escherichia coli. Nucleic Acids Res., 31, 1813–1820.

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

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

Huang, F., Qin, J., Reidys, C. and Stadler, P. (2010) Target prediction and a statistical sampling algorithm for RNA-RNA interaction. Bioinformatics, 26, 175–181.

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

Jabbari, H., Condon, A., Pop, A., Pop, C. and Zhao, Y. (2007) Hfold:RNA pseudoknotted secondary structure prediction using hierarchical folding. In, R., G. (ed.), in algorithms in Bioinformatics, 7th international workshop, WABI 2007. Philadelphia, PA, USA.

Jakab, G., Mougin, A., Kis, M., Pollák, T., Antal, M., Branlant, C. and Solymosy, F. (1997) Chlamydomonas U2, U4 and U6 snRNAs. An evolutionary conserved putative third interaction between U4 and U6 snRNAs which has a counterpart in the U4atac-U6atac snRNA duplex. Biochimie, 79, 387–395.

López, M., Rosenblad, M. and Samuelsson, T. (2008) Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. Nucleic Acids Res., 36, 3001–3010.

MacMorris, M., Kumar, M., Lisda, E., Larsen, A., Kraemer, B. and Blumenthal, T. (2007) A novel family of C. elegans snRNPs contains proteins associated with Trans-splicing. RNA, 13, 511–520.

Marz, M., Kirsten, T. and Stadler, P. F. (2008) Evolution of spliceosomal snRNA genes in metazoan animals. J. Mol. Evol., 67, 594–607.

Mathews, D., Sabina, J., Zuker, M. and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol., 288, 911–940.

McCaskill, J. S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers, 29, 1105–1119.

Mneimneh, S. (2009) On the approximation of optimal structures for RNA-RNA secondary structure prediction. and Brunel, C. (1996) Structural features of U6 snRNA and dynamic interactions with other spliceosomal components leading to pre-mRNA splicing. Biochimie, 78, 434–442.

Otake, L., Scamborova, P., Hashimoto, C. and Steitz, J. (2002) The divergent U12-type splicesome is sequired for pre-mRNA splicing and is essential for development in Drosophila. Mol. Cell, 9, 439–446.

Pervouchine, D. (2004) Intermolecular RNA interaction search. Proc. Genome Informatics, 15, 92–101.

Rehmsoeier, M., Steffen, P., Hochmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. Gene, 10, 1507–1517.

Ren, J., Rastegari, B., Condon, A. and Hoos, H. H. (2005) Hotknots: heuristic prediction of microRNA secondary structures including pseudoknots. RNA, 11, 1494–1504.

Repola, F., Majdalani, N. and Gottesman, S. (2003) Small non-coding RNAs, coordinators of adaptation processes in Escherichia coli: The RpoS paradigm. Mol. Microbiol., 48, 855–861.

Rivas, E. and Eddy, S. R. (1999) A dynamic programming algorithms for RNA structure prediction including pseudoknots. J. Mol. Biol., 285, 2053–2068.

Salari, R., Backofen, R. and Sahinalp, S. (2009) Fast prediction of RNA-RNA interaction. In Proceedings of the 9th Workshop on Algorithms in Bioinformatics (WABI), volume 5724 of LNCS, pp. 261–272. Springer, Berlin / Heidelberg.

Seemann, S., Gorodkin, J. and Backofen, R. (2008) Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. Nucleic Acids Res., 36.

Seemann, S., Richter, A., Gorodkin, J. and Backofen, R. (2010) Hierarchical folding of multiple sequence alignments for the prediction of structures and RNA-RNA interactions. Algorithms for Molecular Biology, 5. Doi:10.1186/1748-7188-5-22.

Shambbaugh, J., Hannon, G. and Nilsen, T. (1994) The spliceosomal U small nuclear RNAs of Ascaris lumbricoides. Mol. Biochem. Parasitol., 64, 349–352.

Shukla, G., Cole, A., Dietrich, R. and Peggert, R. (2002) Domains of human U4atac snRNA required for U12-dependent splicing in vivo. Nucleic Acids Res., 30, 4650–4657.

Thomas, J., Lea, K., Zucker-Aprison, E. and Blumenthal, T. (1990) The spliceosomal snRNAs of Caenorhabditis elegans. Nucleic Acids Res., 18, 2633–2642.

Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R. and Picher, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol. Cell, 6, 1331–1342.

Zucker-Aprison, E., Thomas, J. and Blumenthal, T. (1988) C. elegans snRNAs: a model for U4/U6 base pairing. Nucleic Acids Res., 16, 7188–7188.
