BIOINFORMATICS ORIGINAL PAPER

Vol. 26 no. 2 2010, pages 175–181
doi:10.1093/bioinformatics/btp635

Sequence analysis

# Target prediction and a statistical sampling algorithm for RNA–RNA interaction

Fenix W. D. Huang[1], Jing Qin[1], Christian M. Reidys[1,2,*] and Peter F. Stadler[3–8]

[1]Center for Combinatorics, LPMC-TJKLC, [2]College of Life Science, Nankai University, Tianjin 300071, P. R. China, [3]Bioinformatics Group, Department of Computer Science, [4]Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, [5]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, [6]RNomics Group, Fraunhofer Institut for Cell Therapy and Immunology, Perlickstraße 1, D-04103 Leipzig, Germany, [7]Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria and [8]The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, USA

Received on September 9, 2009; revised on November 4, 2009; accepted on November 5, 2009

Advance Access publication November 13, 2009

Associate Editor: Ivo Hofacker

**ABSTRACT**

**Motivation:** It has been proven that the accessibility of the target sites has a critical influence on RNA–RNA binding, in general and the specificity and efficiency of miRNAs and siRNAs, in particular. Recently, $O(N^6)$ time and $O(N^4)$ space dynamic programming (DP) algorithms have become available that compute the partition function of RNA–RNA interaction complexes, thereby providing detailed insights into their thermodynamic properties.

**Results:** Modifications to the grammars underlying earlier approaches enables the calculation of interaction probabilities for any given interval on the target RNA. The computation of the 'hybrid probabilities' is complemented by a stochastic sampling algorithm that produces a Boltzmann weighted ensemble of RNA–RNA interaction structures. The sampling of $k$ structures requires only negligible additional memory resources and runs in $O(k{\cdot}N^3)$.

**Availability:** The algorithms described here are implemented in C as part of the `rip` package. The source code of `rip2` can be downloaded from http://www.combinatorics.cn/cbpc/rip.html and http://www.bioinf.uni-leipzig.de/Software/rip.html.

**Contact:** duck@santafe.edu

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 INTRODUCTION

RNA–RNA binding is a major mode of action of various classes of non-coding RNAs and plays a crucial role in many regulatory processes in all living organisms. Examples include the regulation of translation in both prokaryotes (Narberhaus and Vogel, 2007) and eukaryotes (Banerjee and Slack, 2002; McManus and Sharp, 2002), the targeting of chemical modifications (Bachellerie *et al.*, 2002), insertion editing (Benne, 1992) and transcriptional control (Kugel and Goodrich, 2007). Emerging evidence suggests, furthermore, that RNA–RNA interactions also play a role for the functionality of long mRNA-like ncRNAs (Hekimoglu and Ringrose, 2009). A common theme in many RNA classes, including miRNAs, snRNAs, gRNAs, snoRNAs and in particular many of the procaryotic small RNAs, is the formation of RNA–RNA interaction structures that are much more complex than simple complementary sense–antisense interactions. Thermodynamically, the binding of two RNA molecules $A$ and $B$ can be described by the binding energy $\Delta G^{\text{bind}} = G_{AB} - G_A - G_B$, i.e. by the difference of the energy of structure formation $G_{AB}$ of the $AB$ complex and the folding energies $G_A$ and $G_B$ of the two individual RNAs $A$ and $B$. Thus, the binding or hybridization energy has been widely used as a criterion to predict RNA–RNA interactions (Busch *et al.*, 2008; Rehmsmeier *et al.*, 2004; Tjaden *et al.*, 2006).

The interaction between two RNAs is governed by the same physical principles that determine RNA folding: the formation of specific base pairing patterns whose energy is largely determined by base pair stacking and loop strains. Secondary structures, therefore, are an appropriate level of description to quantitatively understand the thermodynamics of RNA–RNA binding. Just as the general RNA folding problem with unrestricted pseudoknots (Akutsu, 2000), the RNA–RNA interaction problem (RIP) is Non-Polynomial (NP)-complete in its most general form (Alkan *et al.*, 2006; Mneimneh, 2009). Polynomial-time algorithms can be derived, however, by restricting the space of allowed configurations in ways that are similar to pseudoknot folding algorithms (Rivas and Eddy, 1999). The simplest approach concatenates two (or more) interacting sequences and then employs the standard secondary structure folding algorithm with a slightly modified energy model that treats loops containing cut-points as external elements. The software tools `RNAcofold` (Bernhart *et al.*, 2006; Hofacker *et al.*, 1994), `pairfold` (Andronescu *et al.*, 2005) and `NUPACK` (Dirks *et al.*, 2007) subscribe to this strategy. The main problem 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.*, 2009b; Qin and Reidys, 2007). An alternative line of thought,

*To whom correspondence should be addressed.

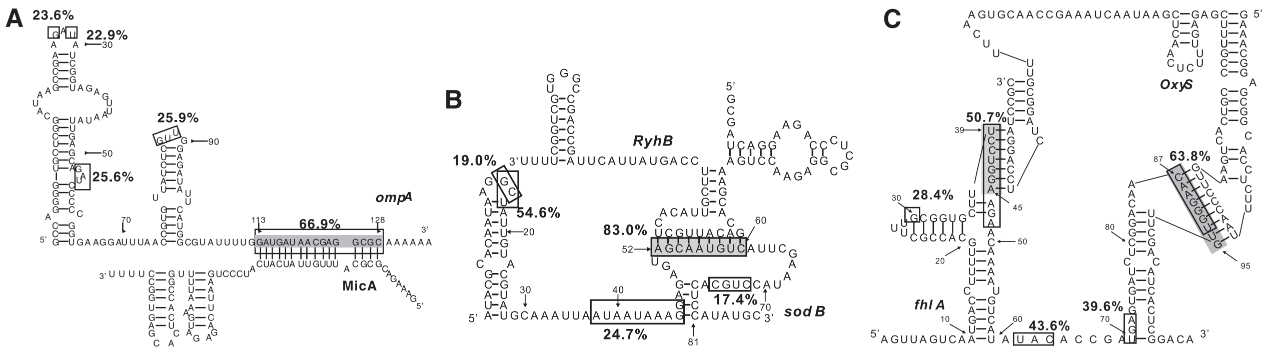

**Fig. 1.** Examples of RNA-RNA interactions structures. The primary interaction region(s) are highlighted in grey in the experimentally supported structural models from the literature: (**A**) *ompA-MicA*: (Udekwu *et al.*, 2005); (**B**) *sodB-RyhB*: (Geissmann and Touati, 2004); (**C**) *fhlA-OxyS*: (Argaman and Altuvia, 2000). Hybridization probabilities computed by `rip2` are annotated by black boxes for regions with a probability larger than 10%. In many cases, the computational predictions identify additional hybridization regions that may further stabilize the interaction.

implemented in `RNAduplex` and `RNAhybrid` (Rehmsmeier *et al.*, 2004), is to neglect all internal base pairings in either strand, i.e. to compute the minimum free energy (MFE) secondary structure of hybridization of otherwise unstructured RNAs. `RNAup` (Mückstein *et al.*, 2006, 2008) and `intaRNA` (Busch *et al.*, 2008) restrict interactions to a single interval that remains unpaired in the secondary structure for each partner. As a special case, snoRNA/target complexes are treated more efficiently using a specialized tool (Tafer *et al.*, 2009) due to the highly conserved interaction motif. Algorithmically, the approaches mentioned so far are close relatives of the RNA folding recursions given by Zuker and Sankoff (1984).

A different approach was taken independently by Pervouchine (2004) and Alkan *et al.* (2006), who proposed MFE folding algorithms for predicting the *joint structure* of two interacting RNA molecules. In this model, “joint structure” means that the intramolecular structures of each partner is pseudoknot free, the intermolecular binding pairs are non-crossing and there is no so-called “zig-zag” configuration (see below for details). The optimal joint structure can be computed in $O(N^6)$ time and $O(N^4)$ space by means of dynamic programming (DP). More recently, extensions to the partition function were proposed by Chitsaz *et al.* (2009b) (`piRNA`) and Huang *et al.* (2009) (`rip1`). In contrast with the RNA folding problem, where minimum energy folding and partition functions can be obtained by very similar algorithms, this is much more complicated for joint structures. The reason is that simple unambiguous grammars are known for RNA secondary structures (Dowell and Eddy, 2004), while the disambiguation of grammar underlying the Alkan–Pervouchine algorithm requires the introduction of a large number of additional non-terminals (which algorithmically translate into additional DP tables). Although the partition function of joint structures can be computed in $O(N^6)$ time and $O(N^4)$ space, the current implementations require very large computational resources. Salari *et al.* (2009) 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), on the other hand, use tree-structured Markov random fields to approximate the joint probability distribution of multiple ($\geq$3) contact regions.

The binding energies provides a useful overall characterization of an RNA–RNA interaction. In many cases, however, the locations of the intermolecular base pairs and the detailed structure of the interaction complex is of crucial importance. Bacterial sRNAs, for example, may either up- or down-regulate mRNA translation depending on the structural changes induced by the interaction (Urban and Vogel, 2007). In particular, in RNA–RNA complexes with multiple interaction sites, i.e. in the class of structures for which the expensive computation of joint structures is necessary, one is interested in the probabilities of hybridization in individual regions and in the interdependencies of alternative conformations, see Fig. 1. The probabilities of the individual building blocks of the DP recursions of Huang *et al.* (2009), furthermore, do not lend themselves to direct biophysical interpretations (see Supplementary Material).

We therefore extend our previous framework in two directions: (i) A modification of the underlying grammar explicitly treats hybrids, i.e. maximal regions with exclusively intermolecular interactions. This allows us to investigate local aspects in much more detail. (ii) A stochastic bracktracing algorithm, in analogy to similar approaches for RNA secondary structure prediction (Ding and Lawrence, 2003; Tacker *et al.*, 1996), which can be used to produce representative structure and to generate samples from the thermodynamic properties. These samples can be useful to assess complex structural features for which it would be too tedious or expensive to design and implement dedicated exact backtracing algorithms.

## 2 THE HYBRID-PARTITION FUNCTION

### 2.1 Some basic facts

We briefly review some basic concepts and outline the notation introduced in Huang *et al.* (2009). Full details are given in the Supplementary Material.

Given two RNA sequences $R=(R_i)_1^N$ and $S=(S_j)_1^M$ (e.g. an antisense RNA and its target or an mRNA and its sRNA regulator) with $N$ and $M$ vertices, we label the vertices such that $R_1$ is the $5'$ end of $R$ and $S_1$ denotes the $3'$ end of $S$. The arcs of $R$ and $S$ then represent the respective, intramolecular base pairs. An arc is called *exterior* if it is of the form $R_iS_j$ and *interior*, otherwise.

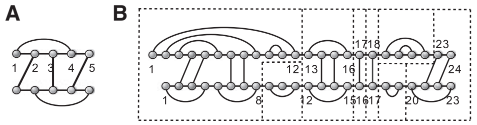

**Fig. 2.** (**A**) A zigzag, generated by $R_2S_1$, $R_3S_3$ and $R_5S_4$. (**B**) We partition the joint structure $J_{1,24;1,23}$ in segments and tight structures.

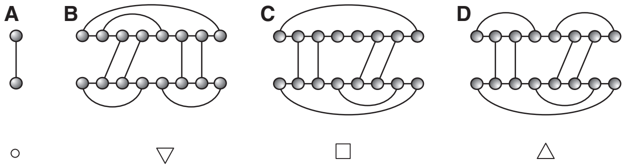

**Fig. 3.** The four basic types of TS. (**A**) $\circ$: $\{R_iS_h\}=J_{i,j;h,\ell}$ and $i=j$, $h=\ell$; (**B**) $\triangledown$: $R_iR_j\in J_{i,j;h,\ell}$ and $S_hS_\ell\notin J_{i,j;h,\ell}$; (**C**) $\square$: $\{R_iR_j,S_hS_\ell\}\in J_{i,j;h,\ell}$; (**D**) $\triangle$: $S_hS_\ell\in J_{i,j;h,\ell}$ and $R_iR_j\notin J_{i,j;h,\ell}$.

Next, we formally define joint structures (Alkan *et al.*, 2006; Chitsaz *et al.*, 2009b; Huang *et al.*, 2009; Pervouchine, 2004). A *joint structure*, $J(R,S,I)$, see Fig. 2B, is a graph such that

(1) $R$, $S$ are secondary structures (each nucleotide being paired with at most one other nucleotide via hydrogen bonds, without internal pseudoknots);

(2) $I$ is a set of exterior arcs without external pseudoknots, i.e. if $R_{i_1}S_{j_1}$, $R_{i_2}S_{j_2}\in I$ then $i_1<i_2$ implies $j_1<j_2$;

(3) $J(R,S,I)$ contains no 'zig-zags', see Fig. 2A;

where a zig-zag is defined as follows: suppose there is an exterior arc $R_aS_b$ with $R_iR_j$ and $S_{i'}S_{j'}$, where $i<a<j$ and $i'<b<j'$. Then $R_iR_j$ is *subsumed* in $S_{i'}S_{j'}$, if for any $R_kS_{k'}\in I$, $i<k<j$ implies $i'<k'<j'$. A *zigzag*, is a subgraph containing two dependent interior arcs $R_{i_1}R_{j_1}$ and $S_{i_2}S_{j_2}$ neither one subsuming the other (Fig. 2). Dependence here means that there exists at least one exterior arc $R_hS_\ell$ such that $i_1<h<j_1$ and $i_2<\ell<j_2$.

The *(induced) subgraph* of $G$ induced by $V$ has vertex set $V$ and contains all $G$-edges having both incident vertices in $V$. The subgraph of a joint structure $J(R,S,I)$ induced by a pair of subsequences $(R_i,R_{i+1},\dots,R_j)$ and $(S_h,S_{h+1},\dots,S_\ell)$ is denoted by $J_{i,j;h,\ell}$. In particular, $J(R,S,I)=J_{1,N;1,M}$ and $J_{i,j;h,\ell}\subset J_{a,b;c,d}$ if and only if $J_{i,j;h,\ell}$ is a subgraph of $J_{a,b;c,d}$ induced by $(R_i,\dots,R_j)$ and $(S_h,\dots,S_\ell)$. In particular, we use $S[i,j]$ to denote the subgraph of the pre-structure $G(R,S,I)$ induced by $(S_i,S_{i+1},\dots,S_j)$, where $S[i,i]=S_i$ and $S[i,i-1]=\varnothing$.

Given a joint structure, $J_{a,b;c,d}$, a tight structure (TS), $J_{i,j;h,\ell}$, (Huang *et al.*, 2009) is a specific subgraph of $J_{a,b;c,d}$. A TS contains a rightmost exterior arc whose $J_{a,b;c,d}$-ancestors (see Supplementary Material for more details) with maximal length give rise to one of the four types of joint structures illustrated in Fig. 3. Intuitively, a TS is obtained as follows: given an exterior arc, $\alpha$, consider its ancestors of maximal length. If there is none, then TS equals $\alpha$. If there is (at least) one, $\beta$, then the TS is determined by the maximal ancestor of the leftmost exterior arc descending from $\beta$ or its endpoint if there is none.

In the following, a TS is denoted by $J^{T}_{i,j;h,\ell}$. If its type is known, then $T$ can be replaced by its type$\in\{\circ,\triangledown,\square,\triangle\}$, see Fig. 3. For instance, we use $J^{\square}_{i,j;h,\ell}$ to denote a TS of type $\square$.

## 2.2 The hybrid grammar

A *hybrid* structure, $J^{\mathsf{Hy}}_{i_1,i_\ell;j_1,j_\ell}$, is a maximal sequence of intermolecular interior loops consisting of exterior arcs $(R_{i_1}S_{j_1},\dots,R_{i_\ell}S_{j_\ell})$ where $R_{i_h}S_{j_h}$ is nested within $R_{i_{h+1}}S_{j_{h+1}}$ and where the internal segments $R[i_h+1,i_{h+1}-1]$ 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. Each hybrid thus forms a distinctive region of interaction between the two RNAs. Note that we can interpret interactions admitted by `intaRNA`/`RNAup` (Busch *et al.*, 2008; Mückstein *et al.*, 2008) as joint structures with at most one hybrid.

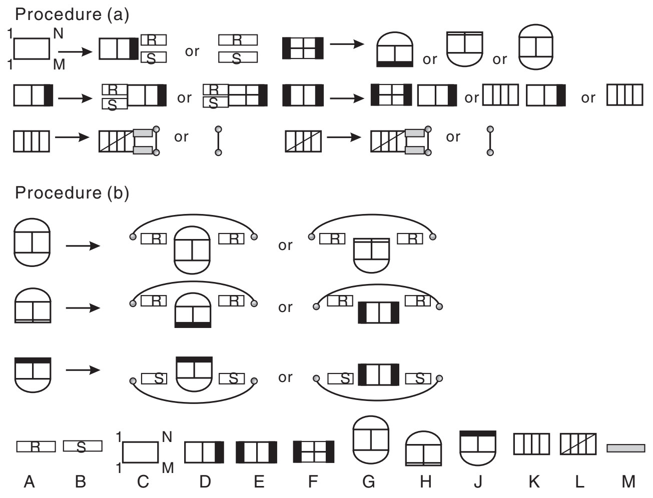

**Fig. 4.** Illustration of the reduction of arbitrary joint structures and of right-tight structures, Procedure (a), and of tight structures, Procedure (b). In the bottom row the symbols for the 10 distinct types of structural components are listed: **A**, **B** maximal secondary structure segments $R[i,j]$, $S[r,s]$; **C** arbitrary joint structure $J_{1,N;1,M}$; **D** right-tight structures $J^{RT}_{i,j;r,s}$; **E** double-tight structure $J^{DT}_{i,j;r,s}$; **F** tight structure of type $\triangledown$, $\triangle$ or $\square$; **G** type $\square$ tight structure $J^{\square}_{i,j;r,s}$; **H** type $\triangledown$ tight structure $J^{\triangledown}_{i,j;r,s}$; **J** type $\triangle$ tight structure $J^{\triangle}_{i,j;r,s}$; **K** hybrid structure $J^{\mathsf{Hy}}_{i,j;h,\ell}$; **L** substructure of a hybrid $J^{\mathsf{h}}_{i,j;h,\ell}$ such that $R_iS_j$ and $R_hS_\ell$ are exterior arcs and $J^{\mathsf{h}}_{i,j;h,\ell}$ itself is not a hybrid since it is not maximal; **M** isolated segment $R[i,j]$ or $S[h,\ell]$.

In the following, we redesign the grammar outlined by Huang *et al.* (2009) so that it explicitly makes use of hybrids. An efficient solution of the partition function problem for RIP requires an unambiguous context-free grammar with the constraint that the number of break points, i.e. the number of non-terminals in each individual production, is as small as possible. This is achieved by introducing several specific types of joint structures that are described in detail in the following. We call a joint *right-tight structure* (RTS), $J^{RT}_{i,j;r,s}$ in $J_{i_1,j_1;r_1,s_1}$, if its rightmost block is a $J_{i_1,j_1;r_1,s_1}$-TS and *double-tight* structure (DTS), $J^{DT}_{i,j;r,s}$ in $J_{i_1,j_1;r_1,s_1}$, if both of its leftmost and rightmost blocks are $J_{i_1,j_1;r_1,s_1}$-TS's. We remark that this definition is a bit different from the notion of the DTS defined in Huang *et al.* (2009). In particular, we consider single interaction arcs as particular DTS. Adopting the point of view of Algebraic Dynamic Programming (Giegerich and Meyer, 2002), we regard each decomposition rule as a production in a suitable grammar. Fig. 4 summarizes the three basic steps of the hybrid grammar: (I) "interior arc-removal" to reduce TS. The scheme is complemented by the usual loop decomposition of secondary structures, and (II) "block-decomposition" to split a joint structure into two smaller blocks.

The grammar in Fig. 4 corresponds to the decomposition (parsing) of a joint structure into interior arcs and hybrids. Fig. 5A shows the corresponding parse tree. The full details of the decomposition procedures are described in Section 2 of the Supplementary Material, where we show that for each joint structure $J_{1,N;1,M}$, we indeed obtain a unique decomposition tree (parse tree), denoted by $T_{J_{1,N;1,M}}$. More precisely, $T_{J_{1,N;1,M}}$ has root $J_{1,N;1,M}$ and all other vertices correspond to a specific substructure of $J_{1,N;1,M}$ obtained by the successive application of the decomposition steps of Fig. 4 and the loop decomposition of the secondary structures. Thus, the hybrid grammar

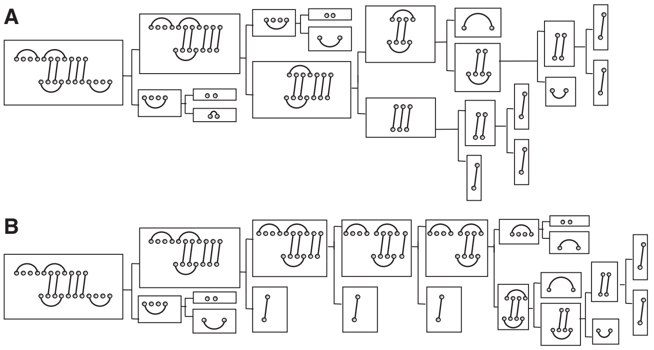

**Fig. 5.** Different grammars lead to different (parse) trees. We show the parse tree $T_{J_{1,11;1,11}}$ for the same joint structure $J_{1,11;1,11}$ according to the grammars of `rip2` (**A**) and `rip1` (**B**), respectively.

is unambiguous. The two panels of Fig. 5 contrast the grammars of `rip1` (Huang *et al.*, 2009) and the hybrid grammar of `rip2` introduced here. In `rip1`, hybrids were immediately decomposed into individual external base pairs and their associated interior loops, so that individual hybrids were not tractable in a straightforward manner.

Let us now have a closer look at the energy evaluation of $J_{i,j;h,\ell}$. Each decomposition step in Fig. 4 results in substructures whose energies are assumed to contribute additively and generalized loops that can be evaluated directly. There are the following two basic scenarios:

(I) Interior Arc removal: the first type of decomposition is derived from the decomposition of TS of Huang *et al.* (2009). Most of the decomposition operations in Procedure (b) displayed in Fig. 4 can be viewed as the "removal" of an arc (corresponding to the closing pair of a loop in secondary structure folding) followed by decomposition. Both, loop type as well as the subsequent decomposition steps depend on the newly exposed structural elements. Following the approach of Zuker and Sankoff (1984) for secondary structures, we treat the loop-decomposition problem by introducing additional matrices. Without loss of generality, we can assume that we open an interior base pair $R_iR_j$.

The set of base pairs on $R[i,j]$ consists of all interior pairs $R_pR_q$ with $i \leq p < q \leq j$ and all exterior pairs $R_pS_h$ with $i \leq p \leq j$. An interior arc is *exposed* on $R[i+1,j-1]$ if and only if it is not enclosed by any interior arc in $R[i,j]$. An exterior arc is *exposed* on $R[i+1, j-1]$ if and only if it is not a descendant of any interior arc in $R[i+1,j-1]$. Given $R_iR_j$, the arcs exposed on $R[i+1,j-1]$ correspond to the base pairs *immediately interior* of $R_iR_j$. Let us write $E_{R[i,j]} = E^i_{R[i,j]} \dot\cup E^e_{R[i,j]}$ for this set of 'exposed base pairs' and its subsets of interior and exterior arcs. As in secondary structure folding, the loop type is determined by $E_{R[i,j]} := E_R$ as follows: $E_R = \varnothing$, hairpin loop; $E_R = E^i_R$ and $|E_R| = 1$, interior loop (including bulge and stacks); $E_R = E^i_R$, $|E_R| \geq 2$, multi-branch loop; $E_R = E^e_R$, kissing-hairpin loop; $|E^i_R|, |E^e_R| \geq 1$, general kissing loop.

This picture needs to be refined even further since the arc removal is coupled with further decomposition of the interval $R[i+1,j-1]$. This prompts us to distinguish TS and DTS with different classes of exposed base pairs on one or both strands. It will be convenient, furthermore, to include information on the type of loop in which it was found.

A TS $J^{\triangledown}_{i,j;h,\ell}$ is of type $\mathsf{E}$, if $S[h,\ell]$ is not enclosed in any base pair ($J^{\triangledown,\mathsf{E}}_{i,j;h,\ell}$). Suppose $J^{\triangledown}_{i,j;h,\ell}$ is located immediately interior to the closing pair $S_pS_q$ ($p < h < \ell < q$). If the loop closed by $S_pS_q$ is a multi-loop, then $J^{\triangledown}_{i,j;h,\ell}$ is of type $\mathsf{M}$ ($J^{\triangledown,\mathsf{M}}_{i,j;h,\ell}$). If $S_pS_q$ is contained in a kissing loop, we distinguish the types $\mathsf{F}$ and $\mathsf{K}$, depending on whether or not $E^e_{S[h,\ell]} = \varnothing$.

Analogously, there are in total four types of a hybrid $J^{\mathsf{Hy}}_{i,j;h,\ell}$, i.e. $\{J^{\mathsf{Hy,EE}}_{i,j;h,\ell}, J^{\mathsf{Hy,EK}}_{i,j;h,\ell}, J^{\mathsf{Hy,KE}}_{i,j;h,\ell}, J^{\mathsf{Hy,KK}}_{i,j;h,\ell}\}$.

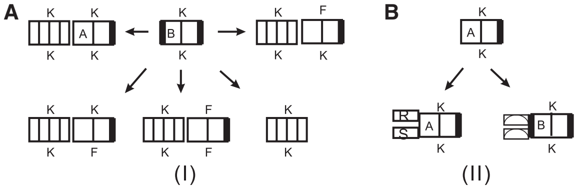

**Fig. 6.** Decomposition of $J^{DT,\mathsf{KKB}}_{i,j;h,\ell}$ (l.h.s.) and $J^{RT,\mathsf{KKA}}_{i,j;h,\ell}$ (r.hs.).

(II) Block decomposition: the second type of decomposition is the splitting of joint structures into 'blocks'. Here, the hybrid grammar differs from the grammar of Huang *et al.* (2009) in two ways. First, we use the hybrid as a new block of the grammar, decomposing a hybrid by removing its exterior arcs in parallel simultaneously starting from the right. Second, we split a joint structure into blocks via alternating decompositions of RTS and DTS as shown in the Procedure (a) of Fig. 4.

In order to guarantee the maximality hybrids, we observe that the RTS's $J^{RT,KK}_{i,j;h,\ell}$, $J^{RT,KE}_{i,j;h,\ell}$, $J^{RT,EK}_{i,j;h,\ell}$ and $J^{RT,EE}_{i,j;h,\ell}$ can appear in two scenarios, depending on whether or not there exists an exterior arc $R_{i_1}S_{h_1}$ such that $R[i,i_1-1]$ and $S[h,h_1-1]$ are isolated segments. In case such an exterior arc exists, we say the RTS is of type ($\mathsf{B}$) or ($\mathsf{A}$), otherwise. Similarly, a DTS, $J^{DT,KK}_{i,j;h,\ell}$, $J^{DT,KE}_{i,j;h,\ell}$, $J^{DT,EK}_{i,j;h,\ell}$ or $J^{DT,EE}_{i,j;h,\ell}$ is of type ($\mathsf{B}$) or ($\mathsf{A}$) depending on whether $R_iS_h$ is an exterior arc. In Fig. 6A, we display the decomposition of $J^{DT,\mathsf{KKB}}_{i,j;h,\ell}$ into hybrids and RTS of type ($\mathsf{A}$) and in Fig. 6B, we display the decomposition of $J^{RT,\mathsf{KKA}}_{i,j;h,\ell}$ into secondary structure segments and DTS accordingly.

Suppose $J^{DT}_{i,j;r,\ell}$ is a DTS contained in a kissing loop, that is, we have either $E^e_{R[i,j]} \neq \varnothing$ or $E^e_{S[h,\ell]} \neq \varnothing$. Without loss of generality, we may assume $E^e_{R[i,j]} \neq \varnothing$. Then, at least one of the two 'blocks' contains at least an exterior arc belonging to $E^e_{R[i,j]}$ labeled by $\mathsf{K}$ or $\mathsf{F}$, otherwise, see Fig. 6A.

## 2.3 Forward recursions

The computation of the partition function proceeds 'from the inside to the outside', see Equation (3). The recursions are initialized with the energies of individual external base pairs and empty secondary structures on subsequences of length up to 4. In order to differentiate multi- and kissing-loop contributions, we introduce the partition functions $Q^{\mathsf{m}}_{i,j}$ and $Q^{\mathsf{k}}_{i,j}$. Here, $Q^{\mathsf{m}}_{i,j}$ denotes the partition function of secondary structures on $R[i,j]$ or $S[i,j]$ having at least one arc contained in a multi-loop. Similarly, $Q^{\mathsf{k}}_{i,j}$ denotes the partition function of secondary structures on $R[i,j]$ or $S[i,j]$ in which at least one arc is contained in a kissing loop. Let $\mathbb{J}^{\xi,Y_1Y_2Y_3}_{i,j;h,\ell}$ be the set of substructures $J_{i,j;h,\ell} \subset J_{1,N;1,M}$, induced from some joint structure $J_{1,N;1,M}$, such that $J_{i,j;h,\ell}$ appears in $T_{J_{1,N;1,M}}$ as an interaction structure of type $\xi \in \{DT, RT, \triangledown, \vartriangle, \square, \circ\}$ with loop-subtypes $Y_1, Y_2 \in \{\mathsf{M},\mathsf{K},\mathsf{F}\}$ on the subintervals $R[i,j]$ and $S[h,\ell]$, $Y_3 \in \{\mathsf{A},\mathsf{B}\}$. Let $Q^{\xi,Y_1Y_2Y_3}_{i,j;h,\ell}$ denote the partition function of the set $\mathbb{J}^{\xi,Y_1Y_2Y_3}_{i,j;h,\ell}$. All recursions for $Q^{\xi,Y_1Y_2Y_3}_{i,j;h,\ell}$ represent a reformulation of the hybrid grammar specified in Fig. 4.

For instance, the recursion for $Q^{DT,\mathsf{KKB}}_{i,j;h,\ell}$ displayed in Fig. 6A is given by:

$$Q^{DT,\mathsf{KKB}}_{i,j;h,\ell} = \sum_{i_1,h_1} Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{KKA}}_{i_1+1,j;h_1+1,\ell} + Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{KF}}_{i_1+1,j;h_1+1,\ell} + Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{FF}}_{i_1+1,j;h_1+1,\ell} + Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{FK}}_{i_1+1,j;h_1+1,\ell} + Q^{\mathsf{Hy,KK}}_{i,j;h,\ell}, \tag{1}$$

where the corresponding recursion for $Q^{\mathsf{Hy,KK}}_{i,j;h,\ell}$ is

$$Q^{\mathsf{Hy,KK}}_{i,j;h,\ell} = \sum_{i_1,h_1} Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} e^{-(\sigma_0 + \sigma G^{\mathsf{Int}}_{i_1,h_1,j,\ell} + (j+\ell-i_1-h_1-2)\beta_3)}. \tag{2}$$

Analogously, the recursions for $Q^{\mathsf{Hy,EE}}_{i,j;h,\ell}$, $Q^{\mathsf{Hy,EK}}_{i,j;h,\ell}$ and $Q^{\mathsf{Hy,KE}}_{i,j;h,\ell}$ read:

$$\begin{aligned} Q^{\mathsf{Hy,EE}}_{i,j;h,\ell} &= \sum_{i_1,h_1} Q^{\mathsf{Hy,EE}}_{i,i_1;h,h_1}\, e^{-(\sigma_0+\sigma G^{\mathsf{Int}}_{i_1,h_1,j,\ell})}; \\ Q^{\mathsf{Hy,EK}}_{i,j;h,\ell} &= \sum_{i_1,h_1} Q^{\mathsf{Hy,EK}}_{i,i_1;h,h_1}\, e^{-(\sigma_0+\sigma G^{\mathsf{Int}}_{i_1,h_1,j,\ell}+(\ell-h_1-1)\beta_3)}; \\ Q^{\mathsf{Hy,KE}}_{i,j;h,\ell} &= \sum_{i_1,h_1} Q^{\mathsf{Hy,KE}}_{i,i_1;h,h_1}\, e^{-(\sigma_0+\sigma G^{\mathsf{Int}}_{i_1,h_1,j,\ell}+(j-i_1-1)\beta_3)}. \end{aligned} \tag{3}$$

## 2.4 Hybrid probabilities

Since the probabilities of individual base pairs are not independent, it is not possible to compute the probabilities for particular hybrids directly from them. Hybrid probabilities thus cannot be obtained in a simple way from the backward recursions described by Huang *et al.* (2009).

Given two RNA sequences, our notion of probability is based on the ensemble of all possible joint interaction structures. Let $Q^I$ denote the partition function of all these joint structures that can formed by two input RNA sequences. The probability of a fixed joint structure $J_{1,N;1,M}$ is given by

$$\mathbb{P}_{J_{1,N;1,M}} = \frac{Q_{J_{1,N;1,M}}}{Q^I}. \tag{4}$$

In difference to the computation of the hybrid-partition function 'from the inside to the outside' (IO), the computation of probabilities of specific substructures is obtained 'from the outside to the inside'. The same principle applies to the computation of base pairing computation of base pairing probabilities of secondary structures (McCaskill, 1990) and joint structures (Huang *et al.*, 2009).

Let $J = J_{1,N;1,M}$, with associated decomposition tree $T(J)$ and let $\Lambda_{J_{i,j;h,\ell}} = \{J \mid J_{i,j;h,\ell} \in T(J)\}$ denote the set of all joint structures $J$ such that $J_{i,j;h,\ell}$ is contained in the decomposition tree $T(J)$. Then we have, by construction,

$$\mathbb{P}_{J_{i,j;h,\ell}} = \sum_{J \in \Lambda_{i,j;h,\ell}} \mathbb{P}_J. \tag{5}$$

Following the (OI)-paradigm, the probability of a parent structure, $\mathbb{P}_{\theta_s}$, is computed prior to the calculation of $\mathbb{P}_{J_{i,j;h,\ell}}$. The conditional probability $\mathbb{P}_{J_{i,j;h,\ell}|\theta_s}$ equals $Q_{\theta_s}(J_{i,j;h,\ell})/Q(\theta_s)$, where $Q(\theta_s)$ is the partition function of $\theta_s$, and $Q_{\theta_s}(J_{i,j;h,\ell})$ the partition function of all those $\theta_s$, that have in addition $J_{i,j;h,\ell}$ as a child in their parse trees. Consequently, $\mathbb{P}_{J_{i,j;h,\ell}}$ can inductively be computed by summing over all probabilities $\mathbb{P}_{\theta_s}$, i.e.

$$\mathbb{P}_{J_{i,j;h,\ell}} = \sum_{\theta_s} \mathbb{P}_{J_{i,j;h,\ell}|\theta_s} \mathbb{P}_{\theta_s} = \sum_{\theta_s} \big[Q_{\theta_s}(J_{i,j;h,\ell})/Q(\theta_s)\big] \mathbb{P}_{\theta_s}. \tag{6}$$

Let $\mathbb{P}^{\mathsf{Hy}}_{i,j;h,\ell}$ denote the probability of the set of substructures $J$ such that the specific hybrid substructure, $J^{\mathsf{Hy}}_{i,j;h,\ell}$, appears in the decomposition tree $T(J)$, i.e. $J^{\mathsf{Hy}}_{i,j;h,\ell} \in T(J)$. Since each joint structure $J^{\mathsf{Hy}}_{i,j;h,\ell}$ is either one of the four types $J^{\mathsf{Hy,EE}}_{i,j;h,\ell}$, $J^{\mathsf{Hy,EK}}_{i,j;h,\ell}$, $J^{\mathsf{Hy,KE}}_{i,j;h,\ell}$ or $J^{\mathsf{Hy,KK}}_{i,j;h,\ell}$, we arrive at

$$\mathbb{P}^{\mathsf{Hy}}_{i,j;h,\ell} = \mathbb{P}^{\mathsf{Hy,EE}}_{i,j;h,\ell} + \mathbb{P}^{\mathsf{Hy,EK}}_{i,j;h,\ell} + \mathbb{P}^{\mathsf{Hy,KE}}_{i,j;h,\ell} + \mathbb{P}^{\mathsf{Hy,KK}}_{i,j;h,\ell}. \tag{7}$$

We remark that, by construction, for $[h_1,\ell_1] \neq [h_2,\ell_2]$, the hybrid probabilities $\mathbb{P}^{\mathsf{Hy}}_{i,j;h_1,\ell_1}$ and $\mathbb{P}^{\mathsf{Hy}}_{i,j;h_2,\ell_2}$ quantify disjoint classes of joint structures. This is a consequence of the maximality of hybrids, which implies that, for fixed interval $[i,j]$, each $[h_1,\ell_1]$ corresponds to a unique hybrid $J^{\mathsf{Hy}}_{i,j;h_1,\ell_1}$. Based on the notion of hybrid probability, we can introduce

$$\mathbb{P}^{\mathsf{target}}_{[i,j]} = \sum_{h,\ell} \mathbb{P}^{\mathsf{Hy}}_{i,j;h,\ell}, \tag{8}$$

which is, according to the above, the probability of the target site $[i,j]$ and furthermore

$$\pi_R(i) = \sum_{p,q:\ p\le i\le q} \sum_{h,\ell} \mathbb{P}^{\mathsf{Hy}}_{p,q;h,\ell}, \tag{9}$$

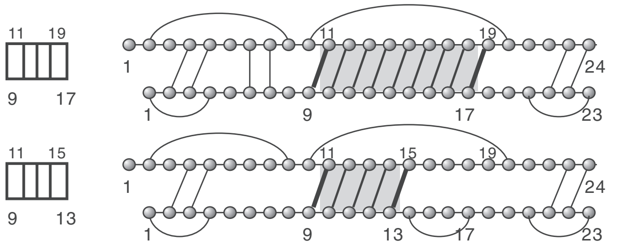

**Fig. 7.** Hybrid probability: the maximality of hybrids implies that—although the intervals $[h_1,\ell_1]$ and $[h_2,\ell_2]$ overlap—they belong to two distinct hybrids (gray).

measuring, for each base $i$ in $R$ the probability that $i$ is contained in a hybrid. A particulary instructive observable is the interaction base pairing matrix, given by

$$\pi_{i,k} = \sum_{p,q:\ p\le i\le q} \sum_{r,s: r\le k\le s} \mathbb{P}^{\mathsf{Hy}}_{p,q;r,s}. \tag{10}$$

Clearly, $\pi_{i,k}$ measures the probability that a pair of nucleotides $(i,k)$, located on different strands, is contained in an interaction region. In contrast with the base pairing probabilities, large values of $\pi_{i,k}$ do not imply that $i$ and $k$ actually form an exterior base pair. Instead, it highlights regions of intermolecular interactions.

## 2.5 Boltzmann sampling

A dynamic programming scheme for the computation of a partition function implies a corresponding stochastic backtracing procedure that can be used to sample from the associated distribution (Tacker *et al.*, 1996). The usefulness of this approach for RNA secondary structures is discussed by Ding and Lawrence (2003). The same ideas can of course also produce representative samples from the Boltzmann equilibrium distribution of RNA interaction structures (Fig. 8).

The basic data structure of the algorithm is a stack $\mathcal{A}$ that stores tuples of the form $\{(i,j;h,\ell;\xi)\}$ describing a pair of intervals $[i,j]$ in $R$ and $[h,\ell]$ in $S$ and the type $\xi$ of the—not further specified—joint structure formed by the two intervals. The stack $\mathcal{A}$, initialized with $(1,N;1,M,?)$ where '?' denotes the unspecified type, guides the backtracing which is complete as soon as $\mathcal{A}$ is empty. A list $\mathcal{L}$ is used to collect the interior and exterior arcs and unpaired bases generated by the decompositions and eventually define the sampled interaction structure. In the first step, $(1,N;1,M,?)$ is decomposed according to the grammar in Fig. 4 into either (i) a pair of secondary structures, or (ii) a RTS $(i,N;j,M;RT\mathsf{EE})$ with probabilities derived as explained above. Depending on the stochastic choice, we push either (i) $(1,N;0,0;\mathsf{sec})$ and $(0,0;1,M;\mathsf{sec})$ or (ii) $(1,i-1;0,0,\mathsf{sec})$, $(0,0;1,j-1;\mathsf{sec})$ and $(i,N;j,M;RT\mathsf{EE})$ into the stack $\mathcal{A}$.

Given $\mathcal{A}$ and $\mathcal{L}$, we can associate a probability by considering the decomposition of the particular type of joint structure. For instance, suppose we have extracted $(i,j;h,\ell,DT\mathsf{KKB})$ from stack $\mathcal{A}$, see Fig. 6. Then, the probabilities for continuing with one of the five decompositions displayed in Fig. 6, for each position of the break points $i_1 \in [i,j]$ and $h_1 \in [h,\ell]$, is given by

$$\begin{aligned} \mathbb{P}^0_{i_1,h_1} &= Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{KKA}}_{i_1+1,j;h_1+1,\ell} / Q^{DT,\mathsf{KKB}}_{i,j;h,\ell}, \\ \mathbb{P}^1_{i_1,h_1} &= Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{KF}}_{i_1+1,j;h_1+1,\ell} / Q^{DT,\mathsf{KKB}}_{i,j;h,\ell}, \\ \mathbb{P}^2_{i_1,h_1} &= Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{FF}}_{i_1+1,j;h_1+1,\ell} / Q^{DT,\mathsf{KKB}}_{i,j;h,\ell}, \\ \mathbb{P}^3_{i_1,h_1} &= Q^{\mathsf{Hy,KK}}_{i,i_1;h,h_1} Q^{RT,\mathsf{FK}}_{i_1+1,j;h_1+1,\ell} / Q^{DT,\mathsf{KKB}}_{i,j;h,\ell}, \\ \mathbb{P}^4_{i_1,h_1} &= Q^{\mathsf{Hy,KK}}_{i,j;h,\ell} / Q^{DT,\mathsf{KKB}}_{i,j;h,\ell}. \end{aligned}$$

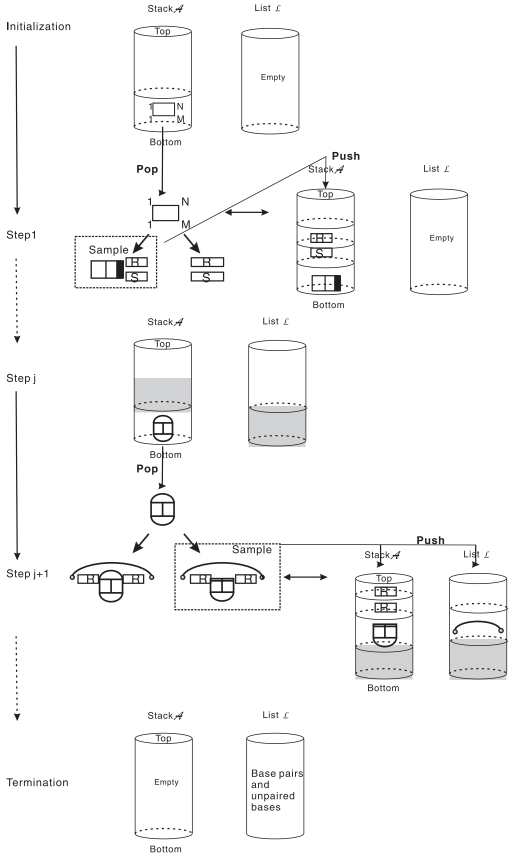

**Fig. 8.** Stochastic backtracing algorithm: elements of stack $\mathcal{A}$ are successively decomposed according to the hybrid-grammar. The resulting arcs and unpaired vertices are stored in the list $\mathcal{L}$ which, once $\mathcal{A}$ is empty, eventually contains the Boltzmann-sampled interaction structure.

One of these decompositions is accordingly sampled and the respective output is pushed back into stack $\mathcal{A}$. For instance, if $\mathbb{P}^1_{i_1,h_1}$ is selected, then we push $(i,i_1;h,h_1;\mathsf{HyKK})$ and $(i_1+1,j;h_1+1,\ell;RT\mathsf{KF})$ back into stack $\mathcal{A}$.

## 3 RESULTS AND CONCLUSIONS

We presented here a modified and improved unambiguous grammar for the RIP. Compared with `rip1` (Huang *et al.*, 2009), it reduces the computational efforts, in particular the memory consumption, by about a third. In the Supplementary Material, we contrast `rip2` with `rip1` and show that hybrids (as opposed to TS, RTS or DTS) are uniquely suited for identifying the interaction regions of two RNA molecules. The complete set of recursions is compiled in Section 3 of the Supplementary Material. It comprises 9 4D-arrays $Q^{\triangle,\triangledown,\Box}_{i,j;r,s}$ for TS of various types, 20 4D-arrays $Q^{RT}_{i,j;r,s}$ for RTS and 20 4D-arrays $Q^{DT}_{i,j;r,s}$ for DTS. The implementation has been complemented by a stochastic backtracing facility. Fig. 9 gives an example of the output produced by `rip2` (see also Supplementary Material, Fig. 4). Despite algorithmic improvements, `rip2` still requires quite substantial computational resources for practical applications. `rip2` is in practise limited to problem sizes of $N_1+N_2 \lessapprox 250$ on current hardware. While `rip2` is still not an efficient tool for large-scale routine applications, it is suitable for investigating the fine details of particular interactions. Future work will thus focus on controlled approximations with the aim of a drastic reduction of both: CPU and memory consumption.

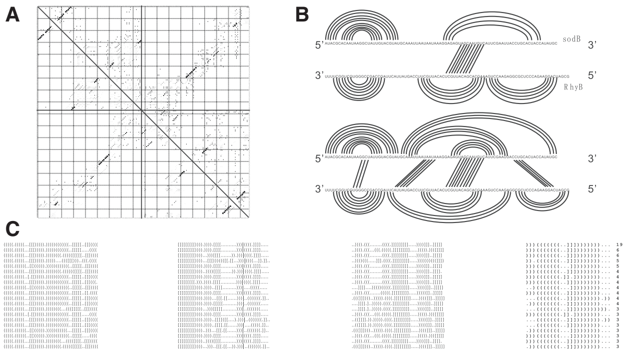

**Fig. 9.** Interaction of *sodB–RhyB*. (**A**) Base-pairing probability matrix. The upper right triangle shows the probabilities obtained from the exact backwards recursion, the lower left triangle is the estimate from a sample of 10 000 structures obtained by stochastic backtracing, showing that the estimates converge quickly. (**B**) Comparison of the structure proposed in Geissmann and Touati (2004) and the `rip2` prediction. While the major stable hairpins agree and `rip2` correctly predicts the primary interaction region, `rip2` also identifies additional interaction regions that may stabilize the interaction. (**C**) Sampled joint structures (here the 20 most frequent ones) are represented as dot-bracket strings: () and [] represent pairs of interior and exterior arc, respectively, while dots indicate unpaired bases. | separates the two RNA sequences which are both written in $5'\to3'$ direction.

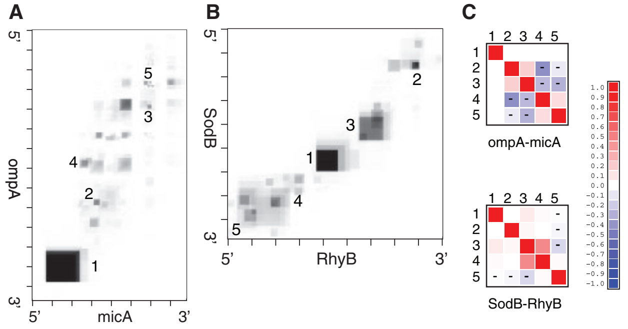

**Fig. 10.** Interaction maps. The *ompA–MicA* interaction (**A**) has a dominating interaction region that brings together the 3′ end of *ompA* and the 5′ terminus of *MicA*. The *sodB–RhyB* interactions (**B**) has two clear hybridization regions in the middle of the molecules and a diffuse contact area at the 3′ end of *sodB*. The grayscale show the probabilities $\pi_{ik}$. Tick marks indicate every 10th nucleotide. The correlations between the major binding regions can be computed easily from Boltzmann samples. The heatmaps show the correlation coefficients for the most probable interaction regions (indicated by numbers in the interaction maps). (**C**) For *sodB–RhyB*, we observe fairly weak correlations, except for the cooperative interaction between contacts 3 and 4. In case of *ompA–MicA*, we observe strong negative correlations between conflicting hybridization regions.

The major advantage of stochastic sampling is that it provides a generic and convenient means to estimate quantities that cannot be easily computed directly by backwards recursion (Ding and Lawrence, 2003). Both, the *ompA-MicA* and *sodB-RhyB* complexes show a primary, highly likely, hybrid region and several additional less stable points of contact, see Fig. 10. In these examples, it is of interest to investigate in detail how the putative interaction regions influence each other: is the binding cooperative so that the major hybrids in Fig. 10 are positively correlated, or do they constitute mutually exclusive contacts? Once a sufficiently large Boltzmann sample is obtained, we can easily compute, e.g. correlations $\rho_{PQ}$ between indicator variables $P$ and $Q$ that measure the existence of external base pairs in two different hybrids. Fig. 10C provide examples, showing that there are strong correlations between hybridization regions. These multiple contacts can contribute substantially to the total interaction energy.

## ACKNOWLEDGEMENTS

We want to thank Sven Findeiß for discussions. We are grateful to Sharon Selzo of the Modular and BICoC Benchmark Center, IBM and Kathy Tzeng of IBM Life Sciences Solutions Enablement. Their great support was vital for all computations presented here.

*Funding*: 973 Project of the Ministry of Science and Technology; PCSIRT Project of the Ministry of Education; National Science Foundation of China (to C.M.R. and his lab); Deutsche Forschungsgemeinschaft under the auspices of SPP-1258 ‘Small Regulatory RNAs in Prokaryotes’ (grant No. STA 850/7-1to P.F.S. and his lab); European Community FP-6 project SYNLET (Contract Number 043312 to P.F.S. and his lab).

*Conflict of Interest*: none declared.

## REFERENCES

Akutsu,T. (2000) Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Disc. Appl. Math.*, **104**, 45–62.

Alkan,C. *et al.* (2006) RNA-RNA interaction prediction and antisense RNA target search. *J. Comput. Biol.*, **13**, 267–282.

Andronescu,M. *et al.* (2005) Secondary structure prediction of interacting RNA molecules. *J. Mol. Biol.*, **345**, 1101–1112.

Argaman,L. and Altuvia,S. (2000) *fhlA* 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. *et al.* (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.

Banerjee,D. and Slack,F. (2002) Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays*, **24**, 119–129.

Benne,R. (1992) RNA editing in trypanosomes. the use of guide RNAs. *Mol. Biol. Rep.*, **16**, 217–227.

Bernhart,S. *et al.* (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.*, **1**, 3.

Busch,A. *et al.* (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, **24**, 2849–2856.

Chitsaz,H. *et al.* (2009a) biRNA: fast RNA-RNA binding sites prediction. In *Proceedings of the 9th Workshop on Algorithms in Bioinformatics (WABI)*, Vol. 5724 of *Lectures Notes in Computer Science*. Springer, pp. 25–36.

Chitsaz,H. *et al.* (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.

Dirks,R. *et al.* (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.*, **49**, 65–88.

Dowell,R.D. and Eddy,S.R. (2004) Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics*, **5**, 7.

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.

Giegerich,R. and Meyer,C. (2002) Algebraic Dynamic Programming. In Vol. 2422 of *Lecture Notes in Computer Science*. Springer, London, pp. 349–364 .

Hekimoglu,B. and Ringrose,L. (2009) Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol.*, **6**, 129–137.

Hofacker,I.L. *et al.* (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.

Huang,F.W.D. *et al.* (2009) Partition function and base pairing probabilities for RNA-RNA interaction prediction. *Bioinformatics*, **25**, 2646–2654.

Kugel,J. and Goodrich,J. (2007) An RNA transcriptional regulator templates its own regulatory RNA. *Nat. Struct. Mol. Biol.*, **3**, 89–90.

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

McManus,M.T. and Sharp,P.A. (2002) Gene silencing in mammals by small interfering RNAs. *Nat. Rev.*, **3**, 737–747.

Mneimneh,S. (2009) On the approximation of optimal structures for RNA-RNA interaction. *IEEE/ACM Trans. Comp. Biol. Bioinform.*, **6**, 682–688.

Mückstein,U. *et al.* (2006) Thermodynamics of RNA-RNA binding. *Bioinformatics*, **22**, 1177–1182.

Mückstein,U. *et al.* (2008) Translational control by RNA-RNA interaction: improved computation of RNA-RNA binding thermodynamics. In Elloumi,M. *et al.* (eds) *Bioinformatics Research and Development — BIRD 2008*, Vol. 13 of *Communication in Computer and Information Science*. Springer, Berlin, pp. 114–127.

Narberhaus,F. and Vogel,J. (2007) Sensory and regulatory RNAs in prokaryotes: A new german research focus. *RNA Biol.*, **4**, 160–164.

Pervouchine,D. (2004) IRIS: intermolecular RNA interaction search. *Proc. Genome Inform.*, **15**, 92–101.

Qin,J. and Reidys,C.M. (2007) A combinatorial framework for RNA tertiary interaction. *Technical Report 0710.3523*, arXiv. Available at http://arxiv.org/PS_cache/arxiv/pdf/0710/0710.3523v3.pdf.

Rehmsmeier,M. *et al.* (2004) Fast and effective prediction of microRNA/target duplexes. *Gene*, **10**, 1507–1517.

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. *et al.* (2009) Fast prediction of RNA-RNA interaction. In *Proceedings of the 9th Workshop on Algorithms in Bioinformatics (WABI)*, Vol. 5724 of *Lecture Notes in Computer Science*. Springer, pp. 261–272.

Tacker,M. *et al.* (1996) Algorithm independent properties of RNA structure prediction. *Eur. Biophy. J.*, **25**, 115–130.

Tafer,H. *et al.* (2009) RNAsnoop: efficient target prediction for box H/ACA snoRNAs. *Bioinformatics*, University of Leipzig. Available at http://www.bioinf.uni-leipzig.de/Publications/PREPRINTS/0 9-025.pdf

Tjaden,B. *et al.* (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res.*, **34**, 2791–2802.

Udekwu,K. *et al.* (2005) Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev.*, **19**, 2355–2366.

Urban,J.H. and Vogel,J. (2007) Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res.*, **35**, 1018–1037.

Zuker,M. and Sankoff,D. (1984) RNA secondary structures and their prediction. *Bull. Math. Biol.*, **46**, 591–621.