Translational Control by RNA-RNA Interaction

Improved Computation of RNA-RNA Binding Thermodynamics

Ulrike Mückstein[1] **, Hakim Tafer[1], Stephan H. Bernhart[2], Maribel Hernandez-Rosales[2], Jörg Vogel[3], Peter F. Stadler[2,1,4,5], and Ivo L. Hofacker[1]

[1] Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria
{ulim,htafer,ivo}@tbi.uvivie.ac.at,
http://www.tbi.univie.ac.at/∼ivo/RNA/

[2] Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany
{bstephan,maribel,studla}@bioinf.uni-leipzig.de

[3] RNA Biology, Max Planck Institut für Infektionsbiologie, Charitéplatz 1, Campus Charité Mitte, D-10117 Berlin, Germany
vogel@mpiib-berlin.mpg.de

[4] RNomics Group, Fraunhofer Institut for Cell Therapy and Immunology (IZI), Deutscher Platz 5e, D-04103 Leipzig, Germany

[5] Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Abstract. The thermodynamics of RNA-RNA interaction consists of two components: the energy necessary to make a potential binding region accessible, i.e. unpaired, and the energy gained from the base pairing of the two interaction partners. We show here that both components can be efficiently computed using an improved variant of `RNAup`. The method is then applied to a set of bacterial small RNAs involved in translational control. In all cases of biologically active sRNA target interactions, the target sites predicted by `RNAup` are in perfect agreement with literature. In addition to prediction of target site location, `RNAup` can also be used to determine the mode of sRNA action. Using information about target site location and the accessibility change resulting from sRNA binding we can discriminate between positive and negative regulators of translation.

1 Introduction

A series of high-throughput transcriptomics projects, among them ENCODE [1] and FANTOM [2] have demonstrated that mammalian genomes are pervasively transcribed, and that a large fraction of the transcripts does not code for proteins. Concurrently, small RNAs, in particular microRNAs and siRNAs have been identified as crucial regulators of gene expression, reviewed e.g. in [3].

** the first two authors contributed equally to this work

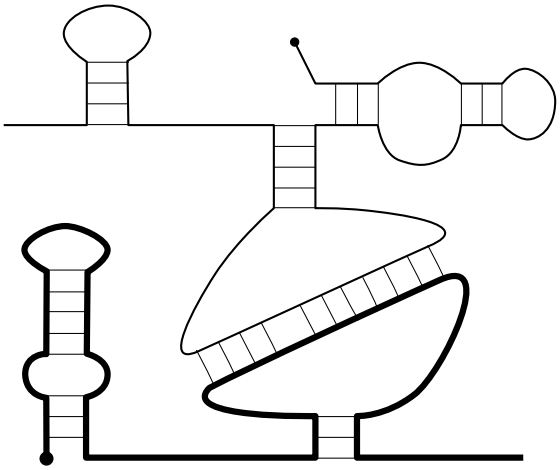

Fig. 1. Interaction between two RNAs of comparable length. Since each molecule forms intramolecular structures, the accessibility for an interaction differs along the molecule: Unstructured regions can easily take part in an interaction. Regions that are involved in an intramolecular structure, e.g. the left hand side of the molecule drawn as a bold line, are not easily accessible for intermolecular binding.

Genome-wide mapping of small ncRNAs [4] revealed novel classes of ncRNAs, implying that ncRNAs act by several, if not many, different mechanisms.

MicroRNAs, siRNAs and snoRNAs require the direct interaction of ncRNAs and their target by means of base-pairing [5]. The same is true for many of the bacterial small RNAs discovered during the last decade, see e.g. [6]. Computational evidence [7] suggests, furthermore, that a significant fraction of RNA candidates with evolutionary conserved RNAs [8] binds to mRNAs.

These observations have triggered increasing interest in methods to predict "targets" via the evaluation of RNA-RNA interactions. For microRNAs, the available tools are almost too numerous to list (see [9, 10] for recent reviews), `targetRNA` [11] is frequently used for bacteria, and a specific heuristic for orphan snoRNAs was presented recently [12]. In the most simple case, only the base pairing between the two interacting partners is taken into account [13–16, 11]. In most cases, however, RNA-RNA interaction does not cover the entire target. This is maybe most evident in the case of short siRNAs or miRNAs targeting long mRNAs. It becomes necessary in such cases, to explicitly consider the structure of the target. In [17], anti-sense targets are predicted as unpaired regions on the target molecules. For siRNA and microRNA it was shown that the accessibility of the target site correlates directly with the efficiency of cleavage [18, 19].

Instead of treating the target independent of its binding partner, it seems more appealing to compute the structure of the interaction complex. Just as the folding problem with pseudoknots [20], finding the energetically optimal interaction structure is NP-complete [21]. It is, however, not even desirable to solve

the general "RIP" problem, because too highly entangled structures typically are not formed in nature. Practical approaches therefore restrict the set of interaction structures that are searched. So far, four classes of structures have been investigated in some detail:

1. Only base-pairs between the interacting RNAs are considered, no base pairs are allowed within each structure. As argued above, disregarding the internal structure of the interaction partners may be too crude an approximation.
2. Interactions between the two molecules are restricted to the external bases of the two partners. Such structures can be computed by means of a straightforward generalization of the usual pseudoknot-free folding algorithm [22, 23]. This class of structures, however, is still too restrictive as it rules out frequent motifs such as kissing-hairpins [24].
3. The other extreme is to consider all "tangle-free" interaction structures. This leads to a rather expensive algorithm with a runtime $\mathcal{O}(m^3 \cdot n^3)$, where m and n are the lengths of the interacting sequences, and quartic memory consumption [25, 21, 26, 27], which is prohibitive for many large-scale applications. Another problem is that the interaction structures contain many types of complex loops for which energy parameters are unknown.
4. The `RNAup` approach [28] restricts the region of interaction to a single interval on each of the interaction partners, while arbitrary pseudoknot free structures are allowed elsewhere, see Fig. 1. This model is sufficient for most but not all known RNA-RNA interactions. For example, the OxyS–fhlA interaction [29] contains two separate kissing complexes and therefore can not be predicted using `RNAup`. Most bacterial sRNAs however show one well defined interaction with a typical interaction length from 9 bp up to 60 bp and variable degrees of complementarity between ncRNAs and target sequence [30, 31]. In [28], only the target molecule was assumed to be structured, while the ncRNA partner was assumed to be a miRNA or siRNA without internal structure. Here we will drop this restriction.

Instead of directly computing the interaction structure, `RNAup` decomposes the problem into three steps: For each subsequence (with bounds i and j) of an RNA, we compute the probability $P[i, j]$ that it is unpaired. This probability is equivalent to the free energy of *making* the binding regions accessible. The optimal interaction structure is then computed by assessing all possible combinations of binding sites of both partners.

This conceptual decomposition of RNA/RNA binding into an unfolding and an interaction contribution has most recently been adopted by several groups. Long *et al.* [32] developed a model for modeling the interaction between a miRNA and a target as a two-step hybridization reaction: nucleation at an accessible target site, followed by hybrid elongation to disrupt local target secondary structure and formation of the complete miRNA-target duplex. Lu & Mathews [33] predicted the cost of opening base pairs in the mRNA for hybridization to siRNA by calculating the structure once without constraints and then once with the constraint that the nucleotides in the hybridization site are forced to be single-stranded. A similar approach is taken in Tafer et al. [34] where accessibility

is computed using the RNAplfold program [35]. Kertesz *et al.* [19] devised a parameter-free model for microRNA-target interaction that computes the difference between the free energy gained from the formation of the microRNA-target duplex and the energetic cost of unpairing the target to make it accessible to the microRNA.

In the following sections we first describe an algorithmic improvement in the computation of $P[i,j]$ that leads to a significant speed-up of `RNAup`. Then we show how to include secondary structure information of both interaction partners in the computation of the free energy of binding. In the results section, we report how these improvements allow us to more precisely describe translational control by bacterial sRNA.

2 Algorithm

`RNAup` calculates the energetics of RNA-RNA interactions in a stepwise process. The free energy of binding ΔG consists of the "breaking energies" ΔG_u that are necessary to render the binding site on each molecule accessible and a contribution ΔG_h that describes the energy gain due to hybridization:

$$\Delta G = \Delta G_u^A + \Delta G_u^B + \Delta G_h, \tag{1}$$

where A and B denote the two interacting molecules. In principle, Eq. 1 has to be evaluated for every possible combination of interacting regions in molecule A and B. In practice, our algorithm first computes the accessibilities ΔG_u for all regions up to a maximum size w and then combines these regions to compute the hybridization energies ΔG_h.

In order to compute free energies of binding we cannot rely on finding a single optimal structure only. Instead, we have to compute the partition functions associated with these three free energy terms. This can be done with (suitably modified) variants of the algorithm introduced by McCaskill [36] and implemented in the `Vienna RNA package` [37]. Recall that the equilibrium partition function is defined as

$$Z = \sum_S \exp(-\beta F(S)), \tag{2}$$

where $F(S)$ is the free energy of a secondary structure S, and $\beta = 1/(RT)$ is the inverse of the temperature times Boltzmann's constant (here expressed as the gas constant, i.e. for energies per mol). Note that individual secondary structures are assigned temperature dependent free energies with entropic contributions arising from the ensemble of microscopic conformations that are assigned to a single secondary structure as macro state. Energy parameters used here are taken from [38]

2.1 Calculation of Accessibility

Partition functions for subsequences contain the information necessary to compute the frequency of structural motifs, in the simplest case individual unpaired bases or base pairs [36].

Here, we are interested in the probability $P_u[i,j]$ that the sequence interval $[i,j]$ is unpaired, which is equivalent to the energy $\Delta G_u[i,j] = -RT\ln(P_u[i,j])$ necessary to make the subsequence from i to j single-stranded. An unpaired interval $[i,j]$ is either "exterior", i.e. not enclosed by a basepair, or there exists an enclosing base pair (p,q) such that $p < i < j < q$ and there is no other pair (s,t) such that $p < s < i < j < t < q$. We can therefore express $P_u[i,j]$ in terms of restricted partition functions for these two cases:

$$P_u[i,j] = \frac{Z(1,i-1)Z(j+1,n) + \sum_{p<i}\sum_{j<q}\hat{Z}(p,q)Z_{pq}[i,j]}{Z(1,n)} \tag{3}$$

where $\hat{Z}(p,q)$ is the partition function outside base pair (p,q), and $Z_{pq}[i,j]$ the partition function inside a base pair (p,q) given that the interval $[i,j]$ is unpaired. Here we introduce an improved recursion for $\hat{Z}(p,q)Z_{pq}[i,j]$ that reduces the CPU requirements of the previous implementation of `RNAup` [39] from $\mathcal{O}(n^3 \cdot w)$ to $\mathcal{O}(n^3)$, where n is the length of the sequence and w is the maximal size of the unstructured region $[i,j]$.

As in [39], we start from the observation that $Z_{pq}[i,j]$ consists of three contributions, of which the summation of all multi-loop energies is the most complex one. This multi-loop part is again split into three parts, depending on whether the unpaired region is to the left or to the right of all components of a multi-loop or in between them, Fig. 2:

$$\begin{aligned} Z^{mult}[i,j] = & \sum_{p<i<j<q} \hat{Z}(p,q)\times \\ & \Bigg(\underbrace{Z^{M2}(p+1,i-1)e^{-\beta c(q-i)}}_{\text{left}} + \underbrace{Z^{M2}(j+1,q-1)e^{-\beta c(j-p)}}_{\text{right}} \\ & + \underbrace{Z^{M}(p+1,i-1)e^{-\beta c(j-i+1)}Z^{M}(j+1,q-1)}_{\text{in-between}} \Bigg) \end{aligned} \tag{4}$$

The crucial improvement is obtained by replacing the double sum in Eq. 3 by two separate summation steps. For the last, "in-between", summand we use the auxiliary variables

$$Z^{MM}(q)[i] = \sum_{1\le p<i} \hat{Z}(pq)Z^{M}(p+1,i-1) \tag{5}$$

For $Z_l^M(q)[i]$ where the unpaired region $[i,j]$ is to the left of all multi-loop components, we introduce

$$Z_l^M(q)[i] = \sum_{1\le p<i} \hat{Z}(p,q)Z^{M2}(p+1,i-1)e^{-\beta c(q-i)} \tag{6}$$

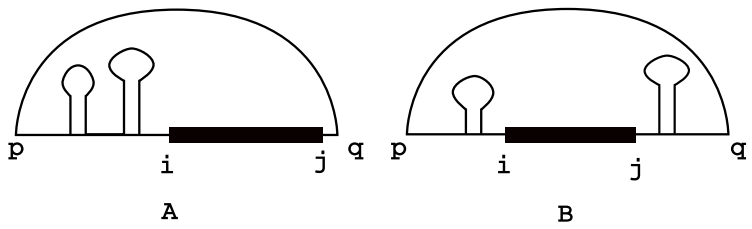

Fig. 2. Decomposition for calculating multiloop contributions: Base pair $[p, q]$ that includes the unpaired region $[i, j]$ is drawn as an arc connecting bases p and q. The unpaired region $[i, j]$ is drawn as a bold black line. In the one-sided multiloop case (A) a structured region containing *at least* two structure components is on one side of the unpaired region. In case (B) the unpaired region $[i, j]$ is between two structured regions. In case (B) we have to take care to make a unique decomposition of the multiloop into a 3' part that contains exactly one component and a 5' part with at least one component.

and an analogous term is used for the "right" contribution. Computing these values costs $\mathcal{O}(n^3)$. By using them, we can compute

$$\begin{aligned} Z^{mult}[i,j] = &\sum_{j<q} Z^{MM}(q)[i] e^{-\beta c(j-i+1)} Z^{M}(j+1, q-1) \\ &+ \sum_{p<i} Z_r^{M}(p)[j] \\ &+ \sum_{j<q} Z^{M}(j+1, q-1) + Z_l^{M}(q)[i] \end{aligned} \tag{7}$$

in $\mathcal{O}(n^2 \cdot w)$ time, i.e., the entire algorithm is $\mathcal{O}(n^3)$. The computations for hairpin and interior loop contributions are handled in the same way.

In comparison to McCaskill's partition function algorithm, `RNAup` needs to store five additional matrices (Z^{M2}, Z^{MM}, Z_l, Z_r and one additional matrix for the interior loop case). Hence we buy the speed-up by $\mathcal{O}(w)$ by increasing the memory requirements by only about a factor of 2. A comparison of the execution times of the old and the new version of `RNAup` shows that the new version is 20 times faster for the default settings ($w = 25$) and sequence lengths below 400 nucleotides. For sequence lengths between 400 and 2000 nucleotides the speed up decreases with increasing sequence length, but the new version is at least 12 times faster. This substantial performance gain considerably facilitates large-scale applications.

2.2 Free Energy of Interaction

In [39] we used $P_u[i, j]$ for the (long) target mRNA only, assuming that the siRNA or miRNA is unstructured due to its short length. This approximation cannot be justified for most bacterial small RNAs, however. Hence, we extended `RNAup` to take the secondary structure of both interacting molecules into account.

Suppose the interaction region covers the intervals $[i^*, j^*]$ and $[i, j]$ in the two RNAs. As in `RNAhybrid` and related programs, we allow interior loops and bulges in the interaction region. The partition function over all these binding conformations is obtained by the following recursion:

$$Z^I[i, j, i^*, j^*] = \sum_{\substack{i<k<j \\ i^*>k^*>j^*}} Z^I[i, k, i^*, k^*] e^{-\beta I(k,k^*;j,j^*)}. \tag{8}$$

where $I(k, k^*; j, j^*)$ is the energy contribution for the interior loop delimited by the base pairs (k, k^*) and (j, j^*).

As we want to avoid having to keep track of a four dimensional array, we compute the partition function $Z^*[i, j]$ over all structures where region $[i, j]$ in the *longer* molecule is involved in the interaction. While doing this, we keep track of the region where $Z^I[i, j, i^*, j^*]$ is maximal. The recursion for the calculation of $Z^*[i, j]$ is shown in Eq 9.

$$Z^*[i, j] = P_u^A[i, j] \sum_{i^*>j^*} P_u^B[i^*, j^*] Z^I[i, j, i^*, j^*]. \tag{9}$$

From $Z^*[i, j]$ we can readily compute $\Delta G[ij]$, the free energy of binding given the binding site is in region $[i, j]$. For visual inspection, $\Delta G[ij]$ can be reduced to the optimal free energy of binding $\Delta G[i]$ at a given position i, see Eq 10. The memory requirement for these steps is $\mathcal{O}(n \cdot w^3)$, the required CPU time scales as $\mathcal{O}(n \cdot w^5)$, which, at least for long target RNAs, is dominated by the first step, i.e., the computation of the $P_u[i, j]$.

$$\begin{aligned} \Delta G[i, j] &= -RT \ln Z^*[i, j]. \\ \Delta G[i] &= \min_{k \le i \le l} \{\Delta G[k, l]\}. \end{aligned} \tag{10}$$

The positional free energy, $\Delta G[i]$, referring to position i in the target molecule, is written to a file. For the region with maximal $Z^I[i, j, i^*, j^*]$, we use `RNAduplex` to print out the optimal interaction structure.

3 Results

To test whether the changes in `RNAup` improve its applicability, we studied experimentally verified interactions between bacterial small RNAs (sRNAs) and their targets [30]. Bacterial sRNAs are ideally suited to examine the usefulness of the inclusion of the secondary structure of both interaction partners into the free energy calculations, since sRNAs are long enough to be highly structured. Furthermore the binding region usually spans only part of the sRNA binds. Therefore, the secondary structure of the sRNA will critically influence the exact location of the binding site.

As a first test we compared the binding sites predicted by the old version of `RNAup`, which neglects sRNA structure, with the predictions of the new version

that computes the contributions of both structures. As expected, when omitting the structure within the sRNA the binding energy was markedly higher (mean -24.97 ± 5.97) than in the new version (mean -15.54 ± 1.99).
When comparing binding site location with the location of experimentally verified binding sites, see Table 1, we found that the new version predicts binding sites more accurately than the old version. In the new version 3 binding sites were predicted with perfect accuracy (the predicted binding site did not deviate by more than one base pair from the binding site reported in literature), and 7 binding sites deviate by at most 17 base pairs, see Table 1. Neglecting sRNA structure, on the other hand, predicts no binding site with perfect accuracy, 9 binding sites show a deviation between 4 to 45 base pairs, $(4, 11, 12, 16, 27, 33, 39, 39, 45)$, and one binding site prediction was wrong, i.e. far away from the site reported in literature.
This comparison emphasises the importance of the inclusion of secondary structure information of both binding partners when predicting sRNA-mRNA interactions. Neglecting the structure of the sRNA results in an overestimation of the length of the predicted interaction and in most cases hinders the clear localization of the proper target site boundary.

In addition to the location of the binding site, the regulatory effects upon binding of the sRNA to the its target mRNA was studied. We used a data set consisting of 9 small regulatory RNAs from *E.Coli*, their 9 reported mRNA targets and the fold-change in protein concentration induced by all 81 possible mRNA-ncRNA interactions [30]. Among those interactions, 8 targets were downregulated, 2 were upregulated, and no or only marginal changes were detected for the others (see Table 1). Downregulation usually occurs when the hybridisation of the ncRNA with its cognate mRNA blocks the ribosome entry sites on the target (for a review see [40]). In contrast, upregulation typically takes place when the sRNA-mRNA hybridization disrupts intrinsic inhibitory structures that sequester the ribosome binding site and/or the start codon [41–43]. In many cases the sRNA-mRNA interactions are assisted by the RNA chaperone protein *Hfq* [44].

Target prediction was performed with the mRNA constructs (117-689 nts) described in [30] and the full length sRNAs (69-220 nts). The mRNA constructs included a long 5'UTR sequence (57-565 nts) and a comparably short fragment of the CDS (35-139 nts). Both the hybridisation energy and the target site position were computed with `RNAup` for all sRNA-mRNA combinations.

For each sRNA we tested which of the mRNA constructs was predicted to bind most strongly. To our satisfaction the most favorable binding energy for each sRNAs was found for its cognate target (see Table 1).

Since the most common mechanism of translational control is to influence ribosome binding at the Shine-Dalgarno (SD) sequence, we checked the position and structural effects of the predicted interactions. For each of the 8 interactions that resulted in downregulation, we found the binding site to be at or close to the Shine-Dalgarno sequence. This type of inhibition can thus be predicted by

Table 1. Binding site summary for the 10 functional interactions published by Urban et.al [30]. Column $\Delta\Delta G$ shows the optimal binding energy calculated with `RNAup`. Column Position gives the binding position relative to the start codon. Column Position lit. gives the binding position found in the literature.

mRNA	sRNA	regulation	$\Delta\Delta G$	Position	Pos.lit.	cite
RyhB	sodB	-	-11.50	-18,+4	-4,+5	[45]
DsrA	hns	-	-14.60	-10,+11	+7,+19	[46]
MicA	ompA	-	-13.60	-21,-6	-21,-6	[47]
MicC	ompC	-	-15.80	-30,-15	-30,-15	[48]
MicF	ompF	-	-17.80	-11,+9	-11,+10	[48]
Spot42	galK	-	-17.00	-18,+30	-19,+21	[49]
SgrS	ptsG	-	-17.33	-28,-10	-28,+4	[50]
GcvB	dppA	-	-17.30	-30,-7	-31,-14	[31]
DsrA	rpoS	+	-14.52	-126,-97	-119,-97	[42]
RprA	rpoS	+	-15.90	-134,-94	-117,-94	[42]

comparing `RNAup` predictions with sequence features that are easy to recognize in bacterial genomic sequences.

Our data set contains only two examples of upregulation, namely binding of *DsrA* and *RprA* to *rpoS*. In both cases, binding leads to the disruption of a helix which normally sequesters the Shine-Dalgarno sequence as well as the start codon. We remark that this is an example of the modifier RNA mechanism that was proposed in [51, 52].

To assess the ability of `RNAup` to predict upregulating interactions we first compared the accessibility of the region around the start codon of all 9 mRNAs, with the mean accessibility of all 4463 genes in the E.Coli genome. Mean accessibility was computed for regions of 401 nts, centered at the start codon. For comparability we used the same 401 nts regions of our 9 target genes rather than the constructs used above. The accessibilities and corresponding opening energies were computed with RNAup for unpaired regions of length 4. The screen against the E.coli genome with all 9 sRNAs took 16 CPU days on one core of an Intel Core2 duo CPU with 2 GB RAM running at 2.40GHz.

With a local opening energy of 4.51 kcal/mol *rpoS* is the most inaccessible transcript among the 9 transcripts presented here. Genome-wide only 8.8% of the transcripts have a less accessible start codon than *rpoS*. In contrast, the eight downregulated transcripts showed a higher than average (2.23 kcal/mol) accessibility, ranging from 0.30 kcal/mol for ompA to a maximum of 1.27 kcal/mol for ryhB.

After binding *DsrA*, the accessibility of the *rpoS* start codon changes dramatically. With only 1.40 kcal/mol, bound *rpoS* is much more accessible than the average transcript and belongs to the 33% most accessible genes, see fig. 3. The same effect is seen upon binding with *RprA*, with a local accessibility after binding of 1.90 kcal/mol. Technically, accessibilities after binding can be com-

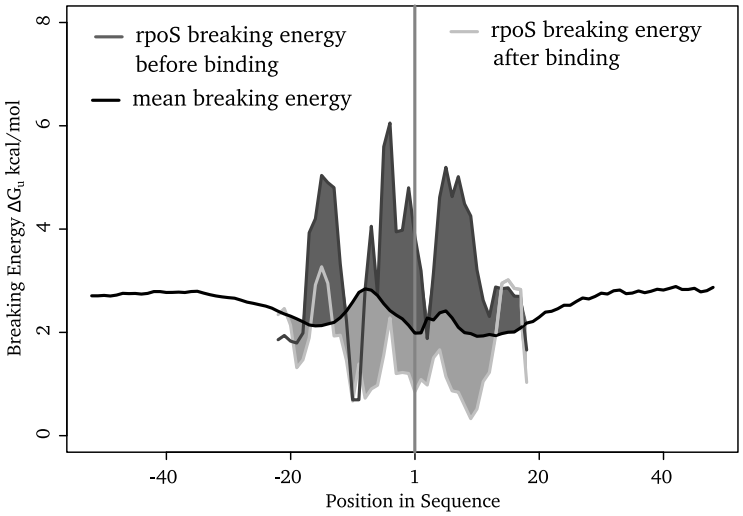

Fig. 3. Opening energy, ΔG_u plotted versus sequence position for the interaction of *DsrA* with textitrpoS. The vertical gray line marks the position of the start codon. The black line represents the average breaking energy for all *E. Coli* mRNAs. The dark gray line represents the opening energy of unbound *rpoS*, the light gray line the opening energy after binding *DsrA*. Unbound *rpoS* is less accessible than average (dark gray area), while bound *rpoS* is more accessible than average (light gray area).

puted easily by adding the constraint that nucleotides in the binding site remain single stranded.

4 Conclusion

Translational control by sRNAs is an important regulatory function throughout all bacteria. In contrast to e.g.micro RNAs, these regulatory RNAs are mostly structured. We have improved `RNAup` to take both target and sRNA structure into account. As we have also increased the speed of `RNAup`, it is now suitable for the computational identification of mRNA targets of bacterial sRNAs.

Furthermore, we find that `RNAup` can be used to predict the regulatory effect of sRNA binding by investigating the location of the binding site and the structural changes induced by binding in the vicinity of the start codon of the mRNA. A predicted binding close to the start codon or the Shine-Dalgarno sequence is a clear indicator for downregulation. While results look promising for upregulation, a bigger data set is needed to confirm that `RNAup` can also accurately predict it.

Our algorithm captures the most common types of interaction between regulatory RNAs and their targets, even though more complicated types of interactions, such as H/ACA snoRNA with their target rRNAs or OxyS–fhlA, are neglected. The speed of `RNAup` is clearly sufficient for genome wide searches for

sRNA–mRNA interactions in bacteria. In principle, the approach is equally applicable to interaction search in higher organisms. However, the larger genome size and longer UTR regions pose challenges both in terms of computation time and false positives.

5 Acknowledgments

This work has been funded, in part, by the Austrian GEN-AU projects bioinformatics integration network and non-coding RNA, the FP-6 EMBIO project, the *Deutsche Forschungsgemeinschaft* Proj No STA 850/7-1 as part of SPP-1258 "Sensory and Regulatory RNAs in Prokaryotes" and Siemens.

References

1. The ENCODE Project Consortium: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447** (2007) 799–816
2. Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engström, P.G., Lenhard, B., Aturaliya, R.N., Batalov, S., Beisel, K.W., Bult, C.J., Fletcher, C.F., Forrest, A.R., Furuno, M., Hill, D., Itoh, M., Kanamori-Katayama, M., Katayama, S., Katoh, M., Kawashima, T., Quackenbush, J., Ravasi, T., Ring, B.Z., Shibata, K., Sugiura, K., Takenaka, Y., Teasdale, R.D., Wells, C.A., Zhu, Y., Kai, C., Kawai, J., Hume, D.A., Carninci, P., Hayashizaki, Y.: Transcript annotation in FANTOM3: Mouse gene catalog based on physical cdnas. PLoS Genetics **2** (2006) e62 doi:10.1371/journal.pgen.0020062.
3. Mattick, J.S., Makunin, I.V.: Non-coding RNA. Hum Mol Genet. **15** (2006) R17–29
4. Kapranov, P., Cheng, J., Dike, S., Nix, D., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Madhavan, G., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R.: RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science **316** (2007) 1484–1488
5. Schubert, S., Gruenweller, A., Erdmann, V.A., Kurreck, J.: Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. J Mol Biol **348**(4) (2005) 883–893
6. Vogel, J., Wagner, E.G.: Target identification of small noncoding RNAs in bacteria. Curr Opin Microbiol. **10** (2007) 262–270
7. The Athanasius F. Bompfünewerer RNA Consortium:, Backofen, R., Flamm, C., Fried, C., Fritzsch, G., Hackermüller, J., Hertel, J., Hofacker, I.L., Missal, K., Mosig, Axel Prohaska, S.J., Rose, D., Stadler, P.F., Tanzer, A., Washietl, S., Sebastian, W.: RNAs everywhere: Genome-wide annotation of structured RNAs. J. Exp. Zool. B: Mol. Dev. Evol. **308B** (2007) 1–25
8. Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., Stadler, P.F.: Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nature Biotech. **23** (2005) 1383–1390
9. Doran, J., Strauss, W.M.: Bio-informatic trends for the determination of miRNA-target interactions in mammals. DNA Cell Biol **26** (2007) 353–360
10. Maziére, P., Enright, A.J.: Prediction of microRNA targets. Drug Discov Today **12** (2007) 452–458
11. Tjaden, B., Goodwin, S.S., Opdyke, J.A., Guillier, M., Fu, D.X., Gottesman, S., Storz, G.: Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res. **34** (2006) 2791–2802
12. Bazeley, P.S., Shepelev, V., Talebizadeh, Z., Butler, M.G., Fedorova, L., Filatov, V., Fedorov, A.: `snoTARGET` shows that human orphan snoRNA targets locate close to alternative splice junctions. Gene **408** (2008) 172–179
13. Rehmsmeier, M., Steffen, P., Hochsmann, M., Giegerich, R.: Fast and effective prediction of microRNA/target duplexes. RNA **10**(10) (2004) 1507–1517
14. Zuker, M.: Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res **31**(13) (2003) 3406–3415
15. Dimitrov, R.A., Zuker, M.: Prediction of hybridization and melting for double-stranded nucleic acids. Biophys J **87**(1) (2004) 215–226

16. Hodas, N.O., Aalberts, D.P.: Efficient computation of optimal oligo-RNA binding. Nucleic Acids Res **32**(22) (2004) 6636–6642
17. Ding, Y., Lawrence, C.E.: Statistical prediction of single stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. Nucl. Acids Res. **29** (2001) 1034–1046
18. Ameres, S.L., Martinez, J., Schroeder, R.: Molecular basis for target RNA recognition and cleavage by human RISC. Cell **130**(1) (2007) 101–112
19. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., Segal, E.: The role of site accessibility in microRNA target recognition. Nat Genet **39**(10) (2007) 1278–1284
20. Akutsu, T.: Dynamic programming algorithms for RNA secondary structure with pseudoknots. Discrete Applied Mathematics **104** (2000) 45–62
21. Alkan, C., Karakoç, E., Nadeau, J.H., Sahinalp, S.C., Zhang, K.: RNARNA interaction prediction and antisense RNA target search. J. Comp. Biol. **13** (2006) 267–282
22. Andronescu, M., Zhang, Z.C., Condon, A.: Secondary structure prediction of interacting RNA molecules. J Mol Biol **345**(5) (2005) 987–1001
23. Bernhart, S.H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P.F., Hofacker, I.L.: Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol. Biol. **1** (2006) 3 [epub]
24. Wagner, E.G.H., Simons, R.W.: Antisense RNA control in bacteria, phage, and plasmids. Annu. Rev. Microbiol. **48** (1994) 713742
25. Pervouchine, D.D.: `IRIS`: Intermolecular RNA interaction search. Proc. Genome Informatics **15** (2004) 92–101
26. Aksay, C., Salari, R., Karakoc, E., Alkan, C., Sahinalp, S.C.: taveRNA: a web suite for RNA algorithms and applications. Nucleic Acids Res **35** (2007) W325–W329
27. Kato, Y., Akutsu, T., Seki, H.: A grammatical approach to RNA-RNA interaction prediction. AIP Conf. Proc. **952** (2007) 197–206 CMLS '07: 2007 International Symposium on Computational Models of Life Sciences.
28. Mückstein, U., Tafer, H., Hackermüller, J., Bernhard, S.B., Stadler, P.F., Hofacker, I.L.: Thermodynamics of RNA-RNA binding. Bioinformatics **22** (2006) 1177–1182
29. Argamana, L., Altuvia, S.: fhla repression by Oxys RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. J Mol Biol. **300**(5) (2000) 1101–12
30. Urban, J.H., Vogel, J.: Translational control and target recognition by Escherichia coli small RNAs in vivo. Nucleic Acids Res **35**(3) (2007) 1018–1037
31. Sharma, C.M., Darfeuille, F., Plantinga, T.H., Vogel, J.: A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev **21**(21) (2007) 2804–2817
32. Long, D., Chan, C.Y., Ding, Y.: Analysis of microRNA-target interactions by a target structure based hybridization model. Pac Symp Biocomput (2008) 64–74
33. Lu, Z.J., Mathews, D.H.: Efficient siRNA selection using hybridization thermodynamics. Nucleic Acids Res **36**(2) (2008) 640–647
34. Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J., Hofacker, I.L.: The impact of target site accessibility on the design of potent siRNAs. Nature Biotech. **26**(5) (2008) in press.
35. Bomfünewerer, A.F., Backofen, R., Bernhart, S.H., Hertel, J., Hofacker, I.L., Stadler, P.F., Will, S.: Variations on RNA folding and alignment: Lessons from benasque. J. Math. Biol. **56** (2008) 119–144
36. McCaskill, J.S.: The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers **29**(6-7) (1990) 1105–1119

37. Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, S., Tacker, M., Schuster, P.: Fast folding and comparison of RNA secondary structures. Monatsh. Chem. **125** (1994) 167–188
38. Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H.: Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol **288**(5) (1999) 911–940
39. Mueckstein, U., Tafer, H., Hackermueller, J., Bernhart, S.H., Stadler, P.F., Hofacker, I.L.: Thermodynamics of RNA-RNA binding. Bioinformatics **22**(10) (2006) 1177–1182
40. Gottesman, S.: Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet **21**(7) (2005) 399–404
41. Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., Gottesman, S.: DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc Natl Acad Sci U S A **95**(21) (1998) 12462–12467
42. Majdalani, N., Hernandez, D., Gottesman, S.: Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Mol Microbiol **46**(3) (2002) 813–826
43. Prévost, K., Salvail, H., Desnoyers, G., Jacques, J.F., Phaneuf, E., Massé, E.: The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Mol Microbiol **64**(5) (2007) 1260–1273
44. Valentin-Hansen, P., Eriksen, M., Udesen, C.: The bacterial Sm-like protein Hfq: a key player in RNA transactions. Mol Microbiol **51**(6) (2004) 1525–1533
45. Geissmann, T.A., Touati, D.: Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. EMBO J **23**(2) (2004) 396–405
46. Lease, R.A., Cusick, M.E., Belfort, M.: Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. Proc Natl Acad Sci U S A **95**(21) (1998) 12456–12461
47. Rasmussen, A.A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C., Valentin-Hansen, P.: Regulation of ompA mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. Mol Microbiol **58**(5) (2005) 1421–1429
48. Chen, S., Zhang, A., Blyn, L.B., Storz, G.: MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli. J Bacteriol **186**(20) (2004) 6689–6697
49. Moeller, T., Franch, T., Udesen, C., Gerdes, K., Valentin-Hansen, P.: Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. Genes Dev **16**(13) (2002) 1696–1706
50. Kawamoto, H., Koide, Y., Morita, T., Aiba, H.: Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol Microbiol **61**(4) (2006) 1013–1022
51. Meisner, N.C., Hackermüller, J., Uhl, V., Aszódi, A., Jaritz, M., Auer, M.: mRNA openers and closers: A methodology to modulate AU-rich element controlled mRNA stability by a molecular switch in mRNA conformation. Chembiochem. **5** (2004) 1432–1447
52. Hackermüller, J., Meisner, N.C., Auer, M., Jaritz, M., Stadler, P.F.: The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: A quantitative model. Gene **345** (2005) 3–12