

IRIS: Intermolecular RNA Interaction Search

Dmitri D. Pervouchine^{1,2}
dp@bu.edu

¹ Center for BioDynamics, Boston University, Boston, MA 02215, USA

² Department of Bioengineering and Bioinformatics, Moscow State University, Vorobiovy gory, 119899 Moscow, Russia

Abstract

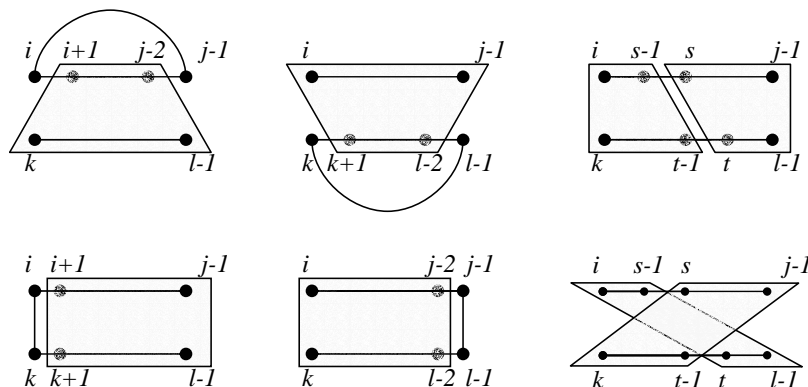
Here we present IRIS, a method for prediction of RNA-RNA interactions that is based on dynamic programming and extends current RNA secondary structure prediction approaches. Using this method we have found a number of interesting refinements to the structures of RNA-RNA complexes that have been studied previously and predicted novel targets for several known regulatory RNAs in *E. coli*. The computational time and memory usage of IRIS are $O(n^3m^3)$ and $O(n^2m^2)$, respectively, where n and m are the lengths of the input sequences. IRIS can be used for analysis of antisense regulatory systems in sequenced organisms and for the design of artificial riboregulators such as antisense drugs.

Keywords: riboregulator, riboswitch, micro RNA, small interfering RNA, secondary structure, dynamic programming, pseudoknot.

1 Introduction

In recent years, the class of non-coding RNAs has acquired many new members in prokaryotes, eukaryotes and archaea [7, 24, 29]. Non-coding RNAs are involved in catalysis and metabolite-sensing, but more typically they are employed by the cell in sequence-specific recognition and processing of other RNA molecules [27]. In particular, they can silence or repress genes at the posttranscriptional level by using base complementarity to hybridize with mRNAs. In animals and plants they are known as small interfering RNAs (siRNAs) and microRNAs (miRNAs), whereas in bacteria the commonly used term is riboregulator [15]. While siRNAs are often fully complementary to their targets, miRNAs and riboregulators interact with mRNAs in a more intricate manner, one which does not involve perfect duplexing [2]. Riboregulators in *E. coli* demonstrate that this interaction can be rather complex, such that intermolecular helices alternate with unpaired regions and intramolecular secondary structure. For instance, the small RNA *oxyS*, which is involved in the oxidative stress response system, interacts with its target, *flhA* mRNA at two sites that reside in the loops of two hairpins [3, 4]. The *dsrA* RNA, which activates and represses, respectively, two other transcriptional regulators *hn-S* and *rpoS*, pairs to their mRNAs by different parts of its three stem-loops [12, 20]. The antisense RNA *copA* binds to the leader region of *copT* mRNA and forms an asymmetrical X-shaped conformation with a four-way junction [8]. The small RNA *ryhB* interferes with translation and transcription of the polycistronic gene *sdhCDAB* by hybridization with the ribosome binding site [14]. In *C. elegans*, the small temporal RNAs *lin-4* and *let-7* pair to the 3' untranslated regions of their target genes in multiple copies, forming bulged double-stranded structures [9, 10, 11].

The wealth of genomic information that was brought by high throughput sequencing poses a challenging problem to systematically search for new targets of known riboregulators and miRNAs. The first step in this direction is to develop an approach for predicting RNA-RNA interactions that extends beyond usage of standard sequence alignment tools. The most straightforward way to find out whether two RNA sequences are able to hybridize with each other is to search for a reverse complement

Figure 1: Recursive calculation of M_{kl}^{ij} .

of one of them in another using, for instance, nucleotide BLAST [1]. This method is applicable for locating long stretches of complementarity and is useful in, for instance, siRNA target search [21, 23]. Another approach, which seems to be more suitable for finding miRNA targets, consists in adopting the RNA structure prediction programs to treat two input sequences [5]. One could concatenate two RNA sequences, input the result to MFOLD, and obtain the intra and intermolecular pairings from the MFOLD output [30]. However, this approach has a serious problem because MFOLD can only predict pseudoknot-free structures, that is, it will erroneously miss the optimal intermolecular interaction in favor of intramolecular secondary structure [19]. The secondary structure prediction method that is tolerant to pseudoknots is potentially applicable but not practicable because of high computational complexity [22].

The method introduced here is designed for intermolecular RNA interaction search (IRIS) [34]. Essentially, it is a product of sequence alignment and two MFOLD-type secondary structure prediction algorithms, implemented as four-dimensional dynamic programming. It shares many common features with secondary structure prediction method with pseudoknots, but is less computationally intensive. The input consists of two RNA sequences. Each of the sequences is allowed to form its own nested secondary structure and to hybridize to the other molecule. The computational time and storage are $O(n^3m^3)$ and $O(n^2m^2)$, respectively, where n and m are the lengths of the input sequences. Although the total degree of the algorithm is six (as in [22]), it is more practical than the pseudoknotted algorithm when one of the sequences is much shorter than the other. This, indeed, is the case for riboregulators and miRNAs. The computational time needed to obtain the optimal secondary structure with pseudoknots by [22] would be $(n+m)^6$, whereas with IRIS it only takes n^3m^3 . This facilitates more than 10^6 -fold increase in speed for $n = 200$ and $m = 20$.

2 Methods

The principle of secondary structure prediction methods that are based on the dynamic programming consists of the recursive derivation of the secondary structure for all segments of the sequence, proceeding progressively from short segments to the entire molecule [33]. A fitness function (typically, the lowest equilibrium free energy) is optimized on each step of the recursion. The fitness function used in MFOLD is the sum of energy impacts from stacking interactions, dangling bases, hairpins, bulges, internal loops and multiloops [16]. It is clear that the set of thermodynamic parameters that is used for RNA secondary structure prediction can be also used for structure prediction of RNA-RNA complexes, as they are based on energies of the same structural elements. For simplicity of notation, here we describe the algorithm, in which the fitness function is the maximum number of base pairs, although the same principle works for thermodynamic calculations.

First, we recall the formal language of secondary structures. A secondary structure on a sequence (x_1, \dots, x_n) is a set of pairs (i, j) such that for every two pairs (i, j) and (i', j') the condition $i = i'$ implies $j = j'$ and vice versa. In other words, each position in can be paired to at most one other position. A secondary structure is called nested, if for every two pairs (i, j) and (i', j') , either $i \leq i' \leq j' \leq j$ or $i' \leq i \leq j \leq j'$ holds.

Denote by X_{ij} the segment (x_i, \dots, x_{j-1}) and let M_{ij} be the maximum number of base pairs in it. It is important to note that the last nucleotide x_j is not included. This way of indexing is different from one of Nussinov [17] because later we will deal with segments of zero length. For instance, $X_{i+1} = \{x_i\}$, and X_{ii} is an empty set. The matrix M_{ij} can be calculated recursively by formula

$$M_{ij} = \max \left\{ M_{i+1j-1} + \delta_{ij-1}, \max_{i < k < j} \{M_{ik} + M_{kj}\} \right\} \quad (1)$$

with the initial conditions $M_{ii} = M_{i+1i} = 0$, $i = 1 \dots n$. Here $\delta_{ij} = 2$, if x_i and x_j are complementary nucleotides, and $\delta_{ij} = 0$ otherwise.

Now we extend this formalism to the case of two sequences. Denote them by $X = (x^1, \dots, x^n)$ and $Y = (y_1, \dots, y_m)$. In what follows, we assume that X is written in 5'-3' direction, Y is written in 3'-5' direction, the superscripts always refer to X , and the subscripts always refer to Y . Consider the following sets of pairings: (1) pairings of x^i with x^j , (2) pairings of y_k with y_l , and (3) cross-pairings of x^i with y_k . We require that

1. Every position in X and Y participates in at most one pairing.
2. The pairings of x^i with x^j form nested secondary structure.
3. The pairings of y_k with y_l form nested secondary structure.
4. If x^i is paired to y_k , and x^j is paired to y_l , then $i < j$ implies $k < l$ and vice versa.

If the conditions 1-3 are met then the set of pairings is said to be the *joint secondary structure*. If, in addition, the condition 4 is met then the joint secondary structure is said to contain *no generalized pseudoknots*. Presence of generalized pseudoknots is not critical for the algorithm's performance. However, if generalized pseudoknots are not present (which is the case most of the times), the joint secondary structure admits representation that is analogous to parenthesis representation for nested secondary structures. This provides a simple and convenient output of the program in text format [34].

A pair of segments (x^i, \dots, x^{j-1}) and (y_k, \dots, y_{l-1}) is called the *bisegment* and is denoted by XY_{kl}^{ij} . Here, the last positions x^j and y_l and are also not included. In this notation, M_{kl}^{ij} is a bisegment with trivial first component. Let M_{kl}^{ij} be the maximum number of base pairs in XY_{kl}^{ij} . Similarly to (1), we can calculate M_{kl}^{ij} recursively by formula

$$M_{kl}^{ij} = \max \left\{ M_{kl}^{i+1j-1} + \delta^{ij-1}, M_{k+1l-1}^{ij} + \delta_{kl-1}, M_{kl-1}^{ij-1} + \delta_{l-1}^{j-1}, M_{k+1l}^{i+1j} + \delta_k^i, \max_{s,t} \{M_{kt}^{is} + M_{tl}^{sj}\} \right\} \quad (2)$$

In other words, either two terminal bases of the first sequence pair, or two terminal bases of the second sequence pair, or a terminal base of the first and a terminal base of the second sequences pair, or the bisegment XY_{kl}^{ij} is splitted into two previously processed bisegments. The last term in equation (2), $M_{kt}^{is} + M_{tl}^{sj}$ does not account for generalized pseudoknots. In order to treat generalized pseudoknots, one should replace it with $\max\{M_{kt}^{is} + M_{tl}^{sj}, M_{tl}^{is} + M_{kt}^{sj}\}$. This procedure is illustrated in Figure 1.

The recursion (2) stops when $i = j$ or $k = l$, that is, when one of the sequences in the bisegment becomes empty. Therefore, the initial conditions for (2) are $M_{kk}^{ij} = M_{kk+1}^{ij} = M^{ij}$ and $M_{kl}^{ii} = M_{kl}^{ii+1} = M_{kl}$ for all i, j, k , and l , where the matrices M^{ij} and M_{kl} are calculated by equation (1):

$$M^{ij} = \max \left\{ M^{i+1j-1} + \delta^{ij-1}, \max_{i < s < j} \{M^{is} + M^{sj}\} \right\}, \quad (3)$$

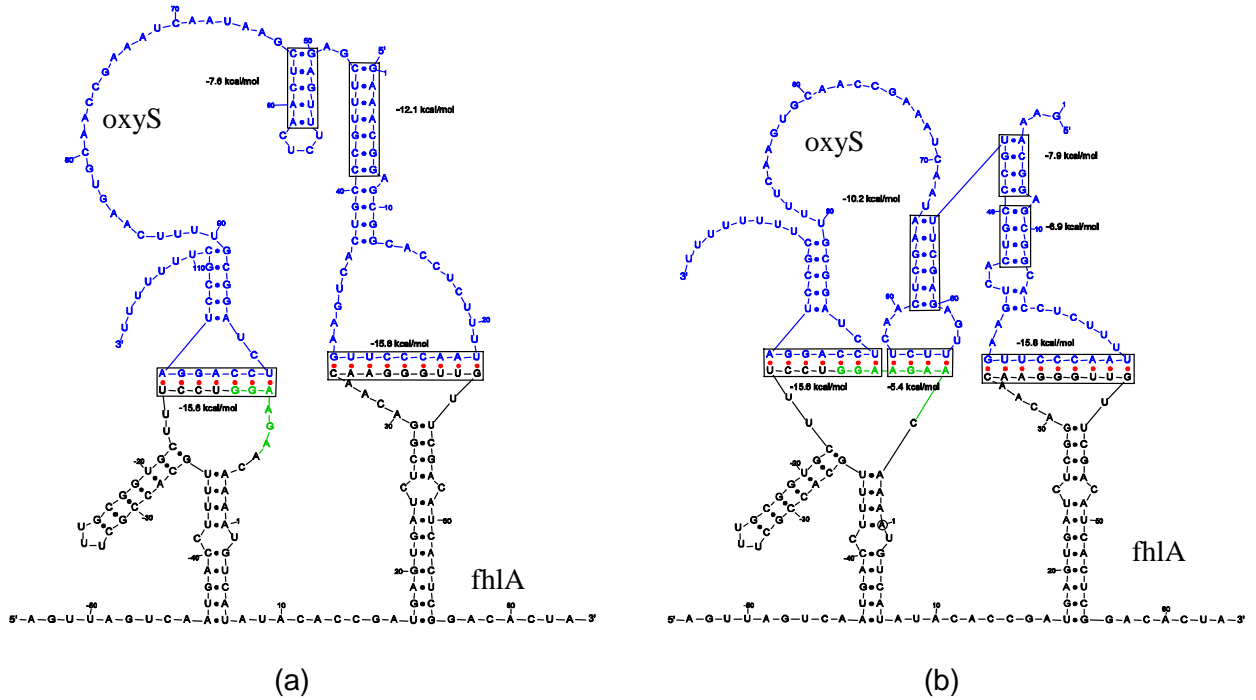


Figure 2: The structure of *oxyS*(blue)-*fhlA*(black) complex proposed by Argaman *et al.* [4] (left) and the structure predicted by IRIS (right). The Shine-Dalgarno sequence is shown in green.

$$M_{kl} = \max \left\{ M^{k+l-1} + \delta^{kl-1}, \max_{k < t < l} \{ M^{kt} + M^{tl} \} \right\} \quad (4)$$

with the initial conditions $M_{kk} = M_{k k+1} = M^{ii} = M^{i i+1} = 0$ for all i and k . As always, superscripts refer to the sequence X , and subscripts refer to the sequence Y .

Calculation of M_{kl}^{ij} is organized as follows. First, we initialize the matrices M^{ij} and M_{kl} , and then compute M^{ij} and M_{kl} by equations (3) and (4). Next, we initialize the four-dimensional matrix M_{kl}^{ij} using M^{ij} and M_{kl} , and then compute M_{kl}^{ij} by equation (2). The number $M_{1 m+1}^{1 n+1}$ is the desired maximum number of base pairs, and the optimal joint secondary structure is obtained from the matrix M_{kl}^{ij} by traceback. In this setup, the computational time and space are $O(n^3 m^3)$ and $O(n^2 m^2)$, respectively.

The algorithm described in this section can be modified and extended to a more realistic schema, one which is based on thermodynamic parameters rather than on scoring matrix. The reader is referred to the supplementary material for the description of the thermodynamic algorithm [34].

3 Results

In this section we apply IRIS to several RNA-RNA complexes that have been described in the literature. The annotated genomic sequences of *E. coli* K12 (NC_000913), *S. flexneri* 2457T (NC_004741), *S. typhi* Ty2 (NC_004631), and *S. typhimurium* LT-2 (NC_003197) were obtained from NCBI website. The calculations were performed with a temperature parameter setting of 37°C for all sequences. Comparative sequence analysis was performed using CLUSTALW [26] (alignments not shown). The following pairs of regulatory RNA vs. target mRNA were analysed: *oxyS* and *fhlA*, *dsrA* and *rpoS/hns* (results not shown), *gcvB* and *dppA/oppA*, *dicF* and *ftsZ/ftsA*, and *ryhB* and *sdhC/bfr/sodB* (results not shown). The predicted structures have been translated from text format to a more friendly, pictorial representation (Figures 2-3) using JAVA-based software [18].

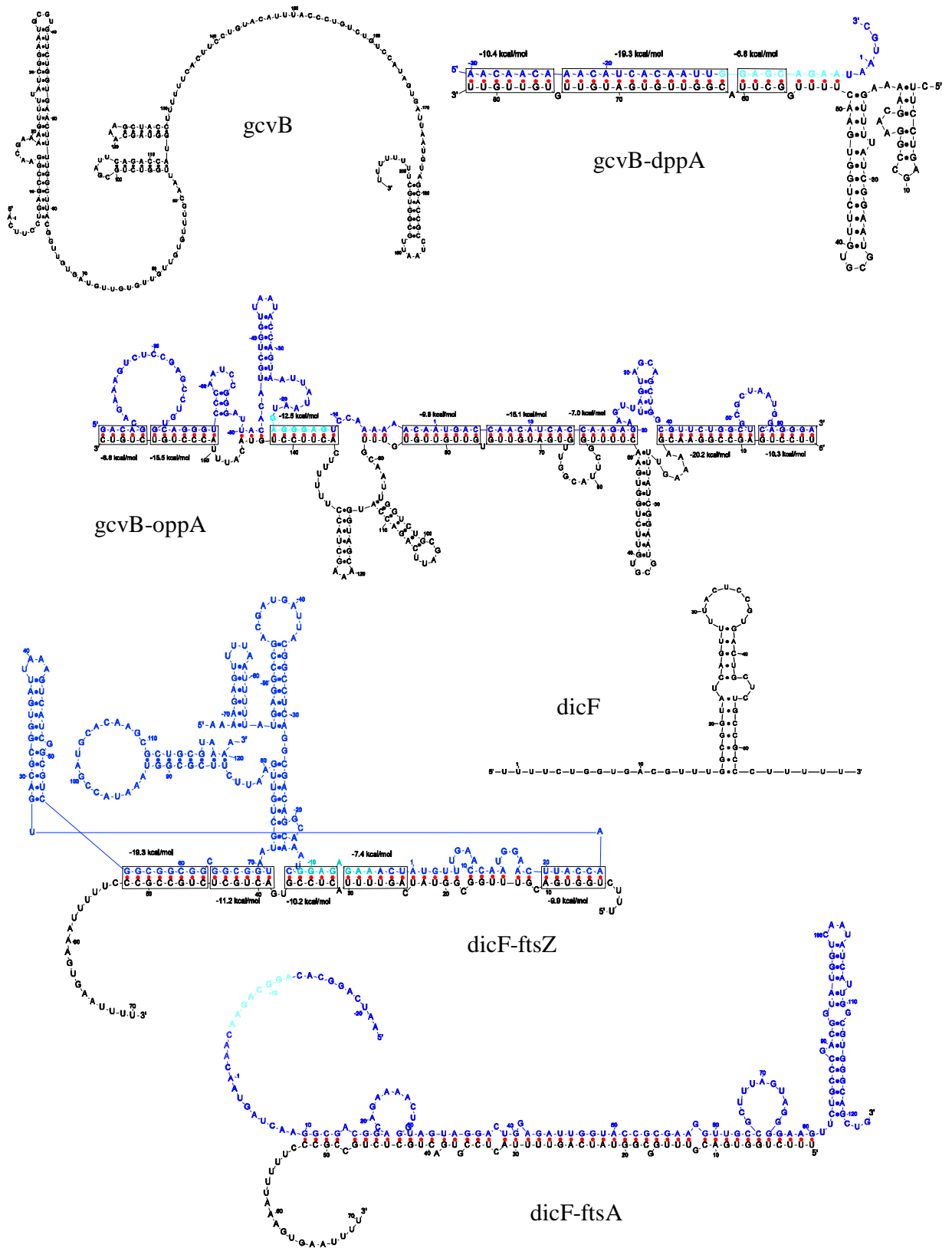


Figure 3: The predicted structures for *gcvB* RNA, *gcvB*(black)-*dppA*(blue), *gcvB*(black)-*oppA*(blue) complexes, *dicF* RNA, *dicF*(black)-*ftsZ*(blue), and *dicF*(black)-*ftsA*(blue) complexes. The ribosome binding site is shown in light blue.

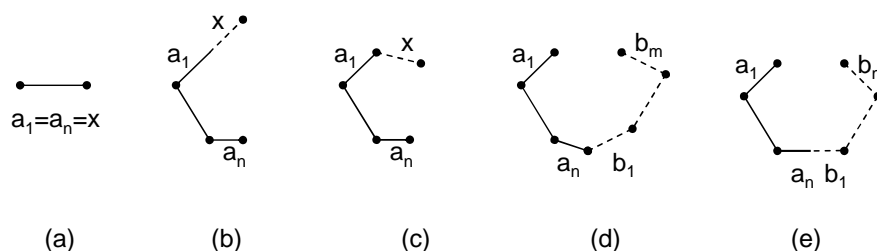


Figure 4: Transformations of polygons (see text).

4 Discussion

Oxys and fhIA. The *oxyS* RNA is expressed in *E. coli* in response to oxidative stress and is known to repress the translation of *fhIA* gene by blocking ribosome binding. It was found in [3] that *oxyS* operates by pairing with a short sequence overlapping the Shine-Dalgarno sequence. Later on, deletion and mutation studies revealed that *oxyS-fhIA* interaction involves a second site residing further downstream, within the coding region of *fhIA* [4]. The structure of *oxyS-fhIA* complex that was proposed in [4] consists of four adjacent stem-loops, two in each of the interacting molecules, which form two stable kissing complexes (Figure 2a). We examine this interaction and find that, in fact, it is not the minimum free energy structure. Moreover, the Shine-Dalgarno sequence (shown in green) is only partially obstructed. The optimal structure predicted by IRIS is shown in Figure 2b and has slightly different arrangement of hairpins. According to IRIS prediction, there exist the third site of interaction that is located between the two other sites, which is able to sequester the Shine-Dalgarno sequence completely. Comparative sequence analysis in other enteric bacteria shows that this structure is conserved in *S. flexneri*, *S. typhi*, and *S. typhimurium*.

Gcvb and dppA/oppA. Transcription of *gcvB* RNA in *E. coli* is controlled by transcriptional regulators of the *gcvTHP* operon, which encodes enzymes of the glycine cleavage system [28]. It has been shown to repress the translation of *OppA* and *DppA* genes, the periplasmic-binding protein components of the two major peptide transport systems. We propose the following structures that could be responsible for inhibition of translation of *OppA* and *DppA* by *gcvB* (Figure 3), although the detailed mechanism involving *gcvB* in the repression of these two genes has not been studied yet. The structure of the *gcvB* RNA (200-bp long) contains three distinguished stem-loops, two of which are followed by polyuridine track ($n = 5 - 7$) and, therefore, may act as transcriptional terminators. The structure of the *gcvB-oppA* complex involves two intermolecular helices, which precede and follow the putative terminator, respectively, while the interaction between *gcvB* and *dppA* mRNA only contains helices that precede the putative terminator. This difference might impose another level of control or differential influence on regulation of these two genes. The Shine-Dalgarno sequence in *gcvB-oppA* complex is obstructed, while the structure in *gcvB-dppA* complex is located in the upstream region. This correlates very well with the fact that *oppA* regulation appears to be at the translational level, whereas *dppA* regulation occurs at the mRNA level [28]. The structure is conserved in *S. flexneri*, *S. typhi*, and *S. typhimurium*.

Dicf and ftsZ. A 190 nt RNA *dicF* is processed from a polycistronic transcript (*dicB* operon) by RNase III and RNase E [6]. A cell-division gene, *ftsZ*, has been identified as the target of *dicF* RNA by a genetic screen for suppressors of *dicF*-dependent inhibition of cell division [25]. We confirm that *dicF* RNA has significant complementarity to the *ftsZ* mRNA in the region surrounding the Shine-Dalgarno sequence, which is consistent with the result that *dicF* regulates *ftsZ* by interfering with ribosome binding. We also find that *dicF-ftsZ* complex admits another region of complementarity (Figure 3), which gives rise to a generalized pseudoknot. The structure of the *dicF-ftsZ* complex is similar to the one of *dsrA-hns*, in which *dsrA* RNA interacts with both 5' and 3' ends of the *hns* mRNA [12]. We

also report that *ftsA*, another gene that is involved in cell division, has significant complementarity to *dicF* RNA (Figure 3). Interestingly, *dicF-ftsA* interaction appear to be downstream of the start codon. As in the case of *gcvB-dppA*, it suggests that *dicF*-mediated regulation of *ftsA* occurs at the mRNA level, while the regulation of *ftsA* occurs at the translational level.

5 Backbone Geometry

The structure of *oxyS-fhlA* complex (Figure 2) has one interesting property. The problem is that the rightmost intermolecular helix is surrounded by AA and UU nucleotides, which could have been paired. However, the helix has to be "straight" because it makes a full turn over the distance of 10 base pairs. Thus, the distance between two terminal nucleotides of the helix is approximately 3.4 nm [31]. On the other hand, the phosphodiester bond is approximately 0.7 nm long. Therefore, the bases A and U that are adjacent to the helix cannot reach each to form a base pair because of backbone constrains. In the case of single RNA molecule we didn't have this problem because on each step of the recursion the paired nucleotides were independent of previously built helices. Now we need to keep track of the geometry of the growing structure, as the multiloop energy is not a simple function of number of interior bases and loop's index. The same sort of artifacts is peculiar to the RNA structure prediction algorithm with pseudoknots [22].

Geometrically, this problem comes down to the following. We need to arrange a number of straight segments in 3D space and connect them by freely-joint chains such that natural spatial constraints are met. The straight segments and the freely-joint chains are helices and unpaired fragments, respectively. In particular, the triangle inequality must hold, that is, for any closed chain of segments in the structure, the length of each segment is not greater than the sum of lengths of the other segments. Although the triangle inequality filters out the configurations that *a priori* contradict Euclidian geometry, it is necessary but not sufficient condition for the given arrangement of helices to exist. In general, this problem can be as hard as the problem of graph embeddings [32].

We use the triangle inequality to detect structures that violate backbone constraints as follows. For each bisegment we consider four chains of straight segments that correspond to its four faces: top, bottom, left, and right. They account for four possibilities of pairing one of the two 5' ends with one of the two 3' ends. Here by a_1, \dots, a_n we denote the lengths of the segments, including both helices and single-stranded regions. Then, the ends of the chain can touch each other if and only if all of the

$$\lambda_j = a_j - \sum_{i \neq j} a_i \quad (5)$$

are negative, or, equivalently, $\lambda = \max\{\lambda_j\}$ is negative. We call λ_j the *discrepancy of the j -th segment*. Denote by π the perimeter of the chain a_1, \dots, a_n . Heuristically, λ and π are the minimum and the maximum distance between the ends of the freely-joint chain. On each step of the dynamic programming we either make a base pair at one of the ends of the bisegment or split it into smaller parts. The inequality $\lambda < 0$ can be used to determine whether a base pair is possible.

Direct calculation of λ requires checking all the subsegments of the given segment and, therefore, results in non-polynomial computational time. However, it is not necessary to keep track of λ_j for each segment, if we only extend but not change the structures that have been built before. These extensions are summarized in Figure 4. They start with a dinucleotide (a). A new base pair can elongate the existing straight segment (b) or be connected to it freely-jointly (c). Similarly, when bisegment is split into parts, the corresponding polygons are concatenated. The concatenation can be soft (d), if the terminal segments are connected freely-jointly, or rigid (e), if they merge into a new straight segment. It turns out that we only need λ , λ_1 , λ_n , and π in order to determine whether the chain can close to form a polygon. These parameters are recalculated on every step of the dynamic programming for each of the four chains according to Table 1. By α and β in Table 1 we denote the

discrepancies of the initial and the terminal segment, that is, λ_1 and λ_2 , respectively. The subscripts refer to the first and the second polygon, which are being connected, concatenated, or closed.

Although the triangle inequality doesn't give a sufficient condition for structure to exist, it allows to get rid of simple structural artifacts. In principle, the theory of graph embeddings can be used here to obtain not only necessary but also the sufficient conditions, but it would give rise to a very complicated non-polynomial algorithm.

Table 1: Transformations of polygons (see text).

Transformation		α	β	λ	π
Initiation		x	x	x	x
Elongation	left	$\alpha + x$	$\beta - x$	$\max\{\lambda - x, \alpha + x\}$	$\pi + x$
	right	$\alpha - x$	$\beta + x$	$\max\{\lambda - x, \beta + x\}$	$\pi + x$
Connection	left	$x - \pi$	$\beta - x$	$\max\{\lambda - x, x - \pi\}$	$\pi + x$
	right	$\alpha - x$	$x - \pi$	$\max\{\lambda - x, x - \pi\}$	$\pi + x$
Concatenation	soft	$\alpha_1 - \pi_2$	$\beta_2 - \pi_1$	$\max\{\lambda_1 - \pi_2, \lambda_2 - \pi_1\}$	$\pi_1 + \pi_2$
	rigid	$\alpha_1 - \pi_2$	$\beta_2 - \pi_1$	$\max\{\beta_1 + \alpha_2, \lambda_1 - \pi_2, \lambda_2 - \pi_1\}$	$\pi_1 + \pi_2$
Closing	soft/soft	$\max\{\lambda_1 - \pi_2, \lambda_2 - \pi_1\}$			
	soft/rigid	$\max\{\beta_1 + \alpha_2, \lambda_1 - \pi_2, \lambda_2 - \pi_1\}$			
	rigid/rigid	$\max\{\alpha_1 + \beta_2, \beta_1 + \alpha_2, \lambda_1 - \pi_2, \lambda_2 - \pi_1\}$			

6 Conclusion

Riboregulators and micro RNAs represent one of those few cases in molecular biology where functional relationship between genes can indeed be established from sequence data. Prediction of targets of regulatory non-coding RNAs is a logical and pertinent task at the current state of art. The method developed in this work provides a generic framework for this problem. It has been shown to agree with several known examples of RNA regulation, yielded a number of interesting refinements to their structures, and allowed to predict novel targets. Although the inherent computational complexity precludes applications of this method on genome-wide scale, it still can be used for the analysis of antisense regulatory systems in sequenced organisms and for the design of artificial riboregulators such as antisense drugs.

Acknowledgments

I thank James Collins, Farren Isaacs, Daniel Dwyer for their support and inspiring discussions, and Maxim Frank-Kamenetskii for his valuable remarks on the thermodynamic model. This work was supported by NIH grant 1 R01 NS46058.

References

- [1] Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J., Basic local alignment search tool, *J. Mol. Biol.*, 215:403–410, 1990.
- [2] Altuvia, S., Gerhart, E., and Wagner, H., Switching on and off with RNA, *Proc. Natl. Acad. Sci. USA*, 97(18):9824–9826, 2000.

- [3] Altuvia, S., Zhang, A., Argaman, L., Tiwari, A., and Storz, G., The *Escherichia coli* OxyS regulatory RNA represses fhlA translation by blocking ribosome binding, *The EMBO Journal*, 17(20):6069–6075, 1998.
- [4] Argaman, L. and 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):1101–1112, 2000.
- [5] Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S., MicroRNA targets in *Drosophila*, *Genome Biol.*, 5(1):R1. Epub, 2003.
- [6] Faubladiere, M., Cam, K., and Bouche, J.P., *Escherichia coli* cell division inhibitor DicF-RNA of the dicB operon. Evidence for its generation *in vivo* by transcription termination and by RNase III and RNase E-dependent processing, *J. Mol. Biol.*, 212(3):461–471, 1990.
- [7] Gaspin, C., Cavallé, J., Erauso, G., and Bachellerie, J.-P., Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: Lessons from the *Pyrococcus* genomes, *J. Mol. Biol.*, 297:895–906, 2000.
- [8] Kolb, F.A., Malmgren, C., Westhof, E., Ehresmann, C., Ehresmann, B., Wagner, E.G.H., and Romby, P., An unusual structure formed by antisense-target rna binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment, *RNA*, 6:311–324, 2000.
- [9] Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T., Identification of novel genes coding for small expressed RNAs, *Science*, 294:853–857, 2001.
- [10] Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P., An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*, *Science*, 294:858–862, 2001.
- [11] Lee, R.C. and Ambros, V., An extensive class of small RNAs in *Caenorhabditis elegans*, *Science*, 294:862–864, 2001.
- [12] Lease, R.A. and Belfort, M., A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures, *Proc. Natl. Acad. Sci. USA*, 97(18):9919–9924, 2000.
- [13] Leydold, J. and Stadler, P.F., Minimal cycle bases of outerplanar graphs, *Elec. J. Comb.*, 5:209–222, 1998.
- [14] Massé, E. and Gottesman, S., A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA*, 99(7):4620–4625, 2002.
- [15] Massé, E., Majdalani, N., and Gottesman, S., Regulatory roles for small RNAs in bacteria, *Curr. Opin. Microbiol.*, 6(2):120–124, 2003.
- [16] Mathews, D.H., Sabina, J., Zuker, M., and Turner, H., Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure, *J. Mol. Biol.*, 288:911–940, 1999.
- [17] Nussinov, R., Pieczenik, G., Griggs, J.R., and Kleitman, D.J., Algorithms for loop matching, *J. Appl. Math.*, 35(1):68–82, 1978.
- [18] Pervouchine, D., Drawing RNA structures with pseudoknots, <http://math.bu.edu/people/dp/prj/rnadraw/>, in preparation.

- [19] Pervouchine, D.D., Graber, J.H., and Kasif, S., On the normalization of RNA equilibrium free energy to the length of the sequence, *Nuclear Acids Res.*, 31(9):e49, 2003.
- [20] Repoila, F., Majdalani, N., and Gottesman, S., Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: The RpoS paradigm, *Mol. Microbiol.*, 48(4):855, 2003.
- [21] Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P., Prediction of plant microRNA targets, *Cell*, 110(4):513–520, 2002.
- [22] Rivas, E. and Eddy, S., A dynamic programming algorithm for RNA structure prediction including pseudoknots, *J. Mol. Biol.*, 285:2053–2068, 1999.
- [23] Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M., Identification of Drosophila MicroRNA Targets, *PLoS Biol.*, 3:E60, 2003.
- [24] Sudarsan, N., Barrick, J.E., and Breaker, R.R., Metabolite-binding RNA domains are present in the genes of eukaryotes, *RNA*, 9:644–647, 2003.
- [25] Tetart, F. and Bouche, J.P., Regulation of the expression of the cell-cycle gene *ftsZ* by DicF antisense RNA. Division does not require a fixed number of FtsZ molecules, *Mol. Microbiol.*, 6(5):615–620, 1992.
- [26] Thompson, J.D., Higgins, D.G., and Gibson, T.J., CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, 22:4673–4680, 1994.
- [27] Tuschl, T., Functional genomics: RNA sets the standard, *Nature*, 421:220–221, 2003.
- [28] Urbanowski, M.L., Stauffer, L.T., and Stauffer, G.V., The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*, *Mol. Microbiol.*, 37(4):856, 2000.
- [29] Vitreschak, A.G., Rodionov, D.A, Mironov, A.A., and Gelfand, M.S., Riboswitches: The oldest mechanism for the regulation of gene expression?, *TRENDS in Genetics*, 20(1):44–50, 2004.
- [30] Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H., and Zuker, M., Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding, *Proc. Natl. Acad. Sci. USA*, 91:9218–9222, 1994.
- [31] Watson, J.D. and Crick, F.H.C., A structure for deoxyribose nucleic acid, *Nature*, 171:737, 1953.
- [32] Wilson, R.J., *Introduction to Graph Theory*, London, Longman, 1975.
- [33] Zucker, M. and Sankoff, D., RNA secondary structures and their prediction, *Bull. Math. Biol.*, 46:46, 1984.
- [34] <http://math.bu.edu/people/dp/prj/irisw/>