RNAplex: a fast and flexible RNA-RNA interaction search tool

Hakim Tafer, Ivo L. Hofacker

htafer@tbi.univie.ac.at, ivo@tbi.univie.ac.at

Abstract:

Regulatory RNAs often unfold their action via RNA-RNA interaction. Transcriptional gene silencing by means of siRNAs and miRNA as well as snoRNA directed RNA editing rely on this mechanism. ncRNA regulation in bacteria is mainly based upon RNA duplex formation. Finding putative target sites for newly discovered ncRNAs is a lengthy task as tools for cofolding RNA molecules like `RNAcofold` and `RNAup` have a run time proportional to $\mathcal{O}((n+m)^3)$ which makes them unpractical for whole genome search. We present a new program, `RNAplex`, especially designed to quickly find possible hybridization sites for a query RNA in large RNA databases. In contrast to earlier approaches, `RNAplex` uses a slightly different energy model which reduces the computational time by a factor 65 compared to RNAhybrid without loss of sensitivity.

1 Introduction

For decades, RNA molecules were dismissed as simple cell servants quietly transmitting genetic information from DNA and converting it into proteins. However the discovery that double stranded non-coding RNAs (dsRNAs) can efficiently inhibit gene expression by hybridizing to a target mRNA aroused strong interest in the scientific community. Recent studies have shown that many RNA-RNA interactions play a crucial role in different cellular processes. RNA-RNA interactions mediate pseudouridylation and methylation of rRNA [BCH02], splicing of pre-mRNA [ZB97], nucleotide insertion into mRNAs [Ben92], transcription and translation control (siRNA, miRNA, stRNA) [FXM$^+$, KG07, BS02] or plasmid replication control [ET90]. While siRNAs are often fully complementary to their targets, most of the ncRNAs interact in a more intricate manner which does not involve perfect hybridization. For example in *E.Coli*, *OxyS*, which is involved in oxidative stress response, interacts with its target mRNA, *fhlA*, through a two sites kissing complex formation [AA00].

Systematic target prediction for the plethora of genomic information brought by ncRNA detection programs and high throughput sequencing is a challenging problem and different kinds of tools are available to solve it. On one hand, BLAST [AGM$^+$90] or FASTA [PL88] search for long stretches of perfect complementarity between a query and a target sequence. GUUGle [GG06] can efficiently locate potential complementary regions and, in contrast to BLAST, also allows for G·U pairs.A typical application for these programs is

for example siRNA target search. Their main drawback is that they do not give information about the thermodynamics of the interaction between the query and the target RNA. Moreover their lack of sensitivity is a real issue when looking for more complex interactions found for example between miRNA and their targets.
On the other hand, RNA folding algorithms based on free energy minimization are at present among the most accurate and most generally applicable approaches for RNA folding [TS88, Zuk00, ZS81]. They are based upon a large number of measurements performed on small RNAs and the assumption that stacking base pairs and loop entropies contribute additively to the free energy of RNA secondary structures [MSZT99, Mat04].
A straightforward approach to folding two RNA molecules is to concatenate the two sequences and apply a slightly modified RNA folding algorithm. This approach is used for example by the `RNAcofold` [HFS$^+$94, BTM$^+$06] and `pairfold` [AZC05] programs. However, the restriction to pseudo-knot free structures in standard folding algorithms is a more serious issue when dealing with RNA duplexes, as many known RNA-RNA interactions are mediated e.g.by "kissing hairpins" or other structure motifs that appear as pseudo-knots when the sequences are concatenated.

As in the case of single sequences [Aku00] inclusion of pseudo-knots makes the problem NP-complete [AKN$^+$06] in the unrestricted case. Polynomial time complexity can be achieved like in Alkan [AKN$^+$06] and Pervouchine [Per04], where intramolecular structures of each molecule are pseudoknot free and intermolecular binding pairs are not allowed to cross. While these algorithms can predict complicated interaction motifs, such as the bacterial `OxyS-fhlA` system, they run in $\mathcal{O}(n^3 \cdot m^3)$ making them prohibitively expensive for most applications. Moreover, these algorithms suffer from a lack of good parameters: Little is known about the energetics of more complicated loop-types, so that predicted optimal structures will often not correspond to reality.

Pseudo-knot free hybrid structures as in the case of `RNAcofold` can be computed in $\mathcal{O}((n+m)^3)$ time. However, the exclusion of pseudo-knots essentially means that interactions can happen only in the exterior loop of the concatenated sequences. Mückstein [MTH$^+$06] recently considered an asymmetric model in which the base pairing is unrestricted in a large target RNA, while the interaction partner is restricted to intermolecular base pairs. `RNAup` works by modeling the total binding energy as a sum of two contributions, the energy needed to make the target site accessible (by breaking intermolecular pairs) and the energy gained through the RNA-RNA interaction. In contrast to `RNAcofold`, `RNAup` allows binding to an unpaired region in any kind of loop, the main limitation is that the interaction is confined to a single such binding site.

A further reduction in computational complexity is achieved by omitting the computation of secondary structures within the monomers. This is implemented by RNAhybrid [RSHG04] and `RNAduplex` from the Vienna RNA package. It is the simplest and fastest approach with a theoretical time complexity scaling as $\mathcal{O}(m \cdot n \cdot \max(n,m))$ which can be reduced to $\mathcal{O}(m \cdot n \cdot L^2)$ by restricting the maximum loop length to L.

Although these programs are fast enough e.g. to predict possible targets of a microRNA, they are still cumbersome for large scale applications comprising many small RNAs and genome-wide searches. Here we present a new version of `RNAduplex`, `RNAplex`, which

is based on a slightly simplified energy model. In this model the loop energy is an affine function of the loop size instead of a log-function. This approach reduces the time complexity to $\mathcal{O}(m \cdot n)$ resulting in a speedup factor of 65 when compared to RNAhybrid, without loss of sensitivity. In particular, the relative energy difference between both energy models remains smaller than 7% for all known miRNA/mRNA interactions. We compare the performance of `RNAplex` and BLAST in finding experimentally verified off-target effects published by Jackson [JBS+03]. We further apply `RNAplex` to study the propensity of ncRNAs predicted by RNAz [WHS05, WHL+05] to bind to known mRNAs.

2 Method

2.1 Energy model

`RNAduplex/RNAhybrid` are essentially equivalent to the classic RNA folding algorithm of Zuker & Stiegler [ZS81] when only interior loops are allowed. As such they have a time complexity of $\mathcal{O}((n+m)^4)$ in the naive implementation , where n and m represents the length of the interacting nucleotide sequences. It is a common practice to speed up these algorithms by restricting the loop size to L leading to $\mathcal{O}((n \cdot m \cdot L^2)$, where $L = 16$ in the case of RNAhybrid. Here we use a simplified energy model that allows us to get rid of the constant but fairly large prefactor L^2.

Since we are neglecting intra-molecular structure here, the only loop types that can appear are stacked pairs, bulge loops, and interior loops. The Turner energy parameters provide look-up tables for the free energies of stacked pairs as well as for small interior loops (1x1, 2x1, and 2x2 loops). These look-up tables are used in `RNAplex` without change. Likewise, bulge loops of length 1 are treated exactly as in the full energy model, namely by adding the stacking energy of the two pairs closing the loop plus a sequence independent penalty. Larger bulge loops are normally assigned a length dependent penalty that grows logarithmically for large loops. In `RNAplex` this bulge energy is approximated by an affine function. Similarly, large interior loops are normally modeled by a size dependent term, an asymmetry penalty, and sequence dependent "terminal mismatches". Here again, we replace the size dependent loop energy by an affine function and neglect the asymmetry. The resulting energy model is exact for small loops and slightly overestimates the loop energies of large interior and bulge loops.

2.2 Recursion

The structure of RNA duplexes predicted by our model can be decomposed in stacking pairs, interior loops and bulges. Our dynamic programming algorithm therefore employs four tables representing sub-structures that end in a base pair C, interior loop I and bulge on the first or second sequence, B^x,B^y, respectively. The central quantity $C_{i,j}$ stores the best energy of interaction between sub-sequence $x_1..x_i$ and $y_j..y_m$. Similarly $B^{x,y}_{i,j}$ store

the best energy of interaction given that residue y_j, respectively x_i, is aligned to a bulge. Finally $I_{i,j}$ store the best energy of interaction given that x_i and y_j are in an interior loop.

Based on these matrices the recursion relation can be written as:

$$C(i,j) = \min \begin{cases} C(i-1,j+1) + \mathcal{S}(i,j;i-1,j+1) \\ C(i-1,j+2) + \mathcal{S}(i,j;i-1,j+2) + P_{\text{bulge}} \\ C(i-2,j+1) + \mathcal{S}(i,j;i-2,j+1) + P_{\text{bulge}} \\ C(i-2,j+2) + \mathcal{I}(i,j;i-2,j+2) \\ C(i-3,j+2) + \mathcal{I}(i,j;i-3,j+2) \\ C(i-2,j+3) + \mathcal{I}(i,j;i-2,j+3) \\ C(i-3,j+3) + \mathcal{I}(i,j;i-3,j+3) \\ I(i-1,j+1) + \mathcal{M}(i,j;i-1,j+1) \\ B^x(i-1,j+1) \\ B^y(i-1,j+1) \end{cases} \quad (1)$$

$$I(i,j) = \min \begin{cases} C(i-1,j+1) + \mathcal{M}(i-1,j+1;i,j) + g^I_{\text{open}} + 2g^I_{\text{ext}} \\ I(i-1,j) + g^I_{\text{ext}} \\ I(i,j+1) + g^I_{\text{ext}} \end{cases} \quad (2)$$

$$B^x(i,j) = \min \begin{cases} C(i-1,j) + g^B_{\text{open}} + g^B_{\text{ext}} \\ B^x(i-1,j) + g^B_{\text{ext}} \end{cases} \quad (3)$$

$$B^y(i,j) = \min \begin{cases} C(i,j+1) + g^B_{\text{open}} + g^B_{\text{ext}} \\ B^y(i,j+1) + g^B_{\text{ext}} \end{cases} \quad (4)$$

where $\mathcal{S}(i,j,k,l)$ represents the energy gained by stacking the $x_i \cdot y_j$ base pair onto the $x_k \cdot y_l$ base pair. As usual, bulges of length 1 are modeled as the sum of a bulge penalty P_{bulge} plus the stacking energy of the adjacent base pairs. $\mathcal{M}(i,j;i-1,j+1)$ represents the "mismatch" energy of the unpaired nucleotides $i-1, j+1$ adjacent to the pair (i,j). $\mathcal{I}$ represents the energy contribution of the small interior loops. Finally $g^{B,I}_{\text{open}}$ and $g^{B,I}_{\text{ext}}$ represent the parameters of the affine loop energy function that approximates the conventional Turner loop energies. These parameters were gained by linearly fitting the loop energy model.

In our model a duplex starts with 2 stacked pairs $(i,j) \cdot (i-1,j+1)$. The hybridization of the recursion matrices should ensure that all structural element has to start and end inside the recursion matrices. This means that no interior loops and no bulges on the target sequence may be closed before $i = 3$. Moreover no bulge and no interior loop on the query sequence may be closed before $j = m-2$. Finally $C_{1,0}$ is set to 0. As a consequence the matrix are initialized in the following way

$$\begin{aligned} I(1,j) &= I(2,j) &&= \infty \; \forall j \\ B^x(1,j) &= B^x(2,j) &&= \infty \; \forall j \\ I(i,m) &= I(i,m-1) &&= \infty \; \forall i \\ B^y(i,m) &= B^y(i,m-1) &&= \infty \; \forall i \end{aligned}$$

The above recursion is graphically represented in figure 1.

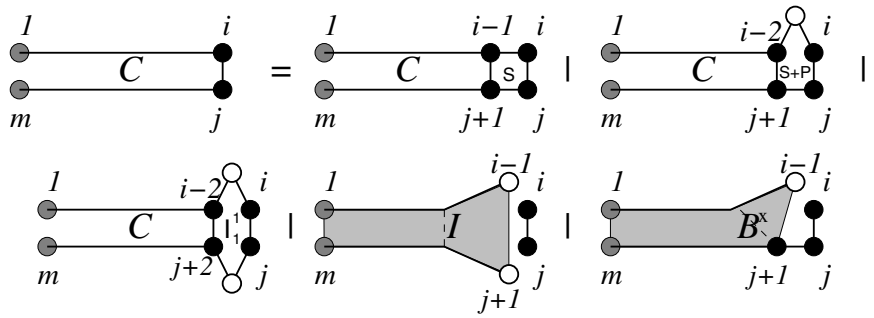


Figure 1: Simplified representation of the structure decomposition used in `RNAplex`. For clarity only the decomposition of the closed structure terms (see equation 1) is shown. Black dots represent paired bases. White dots denote unpaired bases. Given that x_i and y_j are paired, C stores the best energy of interaction between $x_1..x_i$ and $y_j..y_m$. $\mathcal{S}$ is the stacking energy of two pairs of nucleotides. P is the bulge penalty to add to 1x0 bulges. I is the matrix holding the best energy of interaction given that x_i and y_j are in an interior loop. I_1^1 is the destabilizing energy of a 1x1 interior loop (1x2, 2x1 and 2x2 cases not shown) and B^x represents the matrix storing the best energy of interaction given that residue y_j is aligned to a bulge. The cases where x_i and y_j do not pair (interior loop and bulge extension and/or creation) are not shown

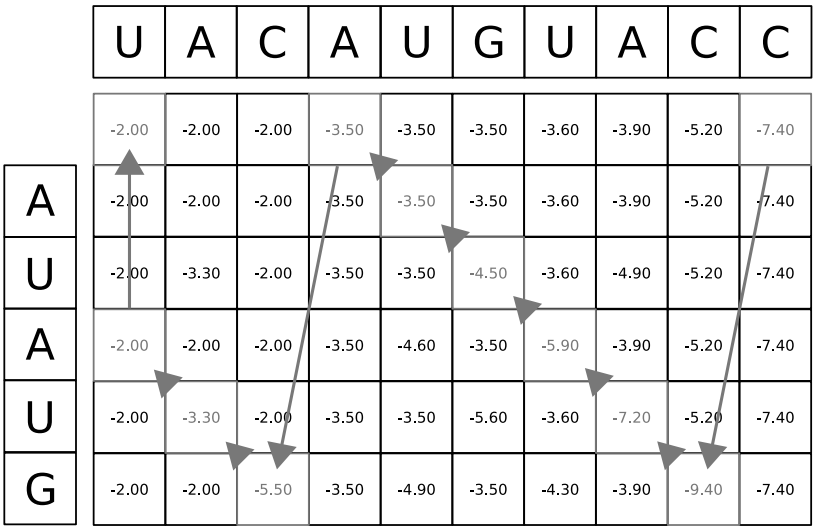
2.3 Suboptimal Hits

In case of long sequences it is possible that many hybridization sites reach a significant interaction energy. `RNAduplex` handles these by backtracking all position in the recursion matrix which have an energy higher than a given threshold. The main problem of this method is that the majority of hits returned are overlapping and do not contain biological relevant information. RNAhybrid reports suboptimal hits by masking previously reported sites and recomputing the whole recursion matrix, which is very time consuming. `RNAplex` uses an idea developed by Durbin [DEKM98] which allows to recover all non overlapping suboptimal hybrids above a given threshold in one pass. This is achieved on one hand by storing the score of the alignment minus the energy threshold, the so-called match score, in an additional line of the C table and on the other hand by setting unpaired region in the C array to the current match score (see figure 2). The recurrence relation for the C matrix is then changed to

$$C(i,0) = \min \begin{cases} C(i-1,0) \\ \min_j C(i-1,j) - T \end{cases} \quad (5)$$

$$C(i,j) = \min \begin{cases} C(i,0) \\ C(i,j) \text{ from eq.(1)} \end{cases} \quad (6)$$

a)



b)

(((&))) 1,3 : 1,3 (-3.50)
(((((&))))) 5,9 : 1,5 (-5.90)

Figure 2: a) The repeat dynamic programming matrix generated by `RNAplex` for the two RNA sequences "UACAUGUACC"and "GUAUA" and a threshold set to $-2\ kCal/mole$. In red the backtracking path returning two separate match regions, with energy -5.90 and -3.50 $kCal/mole$. b) The duplexes in dot-bracket format as returned by `RNAplex`.

2.4 Memory usage

In order to reduce the memory and time consumption of `RNAplex` a slightly different recursion has been used. It should first be noted that based on the recursion presented in section 2.2 each position (i,j) in the matrices $C_{i,j}, I_{i,j}, B^x_{i,j}$ and $B^y_{i,j}$ can be computed from the previous two columns of those matrices. So in a first step, instead of keeping the whole recursion matrices in memory, only 3 columns of each matrix are needed to locate the position in the alignment which scores higher than a given threshold T. The memory usage of this step reduces to $\mathcal{O}(3 \cdot m)$ with m the length of the query sequence. Storing these positions allows one then to recompute the local alignment of the query sequence to the substring of the target sequence around the stored region. The memory consumption of this step reduces to $\mathcal{O}(lm)$ where l represents the maximum hybridization length. The reduced memory usage permits to reduce the computation time by a factor two, mainly because most of the computation are done inside the faster processor cache memory.

We also tried to reduce the computation time of RNAplex by identifying stretches of complementarity before attempting the more time consuming dynamic programming procedure. GUUGle, which locates potential helical regions under RNA base pairing rules with the help of suffix arrays, was used to find these highly complementary regions. The trade-off between speed and sensitivity is controlled by the ktup parameter, which specifies the size of complementarity to search for. We compared the CPU time and sensitivity of RNAplex and GUUGle+RNAplex when searching for experimentally verified miRNA targets. Up to a word size of 7 RNAplex is faster than GUUGle+RNAplex,while the sensitivities for both programs are the same. GUUGle+RNAplex performs better than RNAplex from a word size of 8 nt however at the cost of a reduced sensitivity. RNAplex+GUUGle may prove to be useful for searching of gapped interactions with complementary regions longer than 7 nt.

2.5 Accuracy

RNAplex uses an affine function instead of a log function to model the destabilizing loop energy. This energy model also neglects destabilizing energy coming from loop asymmetry. As a result the hybridization energies as computed by RNAplex and RNAduplex may differ. Table 1 reports the energy difference between RNAduplex/RNAhybrid and RNAplex for loop size up to 14 and loop asymmetry up to 7. As expected, the difference between RNAduplex/RNAhybrid and RNAplex results correlate with the loop asymmetry. On the other hand the energy model of RNAplex deviates only by 0.4 $kCal/mole$ from the standard one for symmetrical loops up to size 14 and bulges up to size 7. To further test our energy model we used RNAplex to recover the experimentally verified human miRNA target sites. All the sites were predicted correctly by RNAplex. Moreover, we compared the energy returned by RNAplex with the energies computed by RNAhybrid on the same data set. The maximal relative energy difference between RNAplex and RNAhybrid was 6.8% or 1.2 $kCal/mol$ for a predicted interaction energy of -17.40 $kCal/mole$.

2.6 Performance

The run time performance of RNAplex has been compared to RNAhybrid by searching for possible hybridization sites for 10 randomly generated 19-mers against a set of 10, 20, 50 and 100 sequences of length 2000 nt each. Both programs were compiled with "-O2 -g" and ran on an Intel(R) Core(TM)2 CPU 6600 @ 2.40GHz. In all cases RNAplex is about 65 times faster than RNAhybrid without loss of sensitivity. Moreover in contrast to RNAhybrid, RNAplex can search for hybridization sites in sequence longer than $2KB$, which makes it a fast tool for genome wide hybridization studies. At this velocity RNAplex can search for all possible hybridization sites of a 19 nt miRNA against the whole human 3'UTR database in $80sec$, instead of $90min$ for RNAhybrid.

loop size query	loop size target							
	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0.20
1	0	0	0	0.10	1.35	1.90	2.40	2.90
2	0	0	0	0.35	0.9	1.35	1.90	2.35
3	0	0.10	0.35	0.10	0.35	0.90	1.35	1.80
4	0	1.35	0.90	0.35	0.10	0.35	0.80	1.25
5	0	1.90	1.45	0.90	0.35	0.20	0.25	0.70
6	0	2.40	1.90	1.35	0.80	0.25	0.30	0.10
7	0.20	2.90	2.35	1.80	1.25	0.70	0.10	0.40

Table 1: Maximal energy difference between `RNAduplex` and `RNAplex` for common loops found in miRNA/mRNA interaction. The first row gives information about the size of the loop in the target strand while the first column reports the loop-size on the query strand. The total loop size can be retrieved by adding the coordinates of a given element in the table. The loop asymmetry is calculated by subtracting the coordinates of a given element in the table. The maximal energy difference between `RNAduplex` and `RNAplex` is found for 7x1 loops where the asymmetry is highest. Inversely the smallest energy differences are found for symmetrical loops and bulges. In our energy model g^I_{open} and g^I_{ext} were set to 1.20 $kCal/mole$ and 0.15 $kCal/mole$ respectively while g^B_{open} and g^B_{ext} were set to 2.00 $kCal/mole$ and 0.40 $kCal/mole$

3 Application

3.1 Off-target effect prediction

Successful application of post-transcriptional gene silencing depends on one hand on reliable selection of potent siRNAs and on the other hand on the ability to ensure specificity of the siRNA for its cognate mRNA. The most commonly used tool to control siRNA specificity is the basic local alignment search tool BLAST. However due its limited capacity to search for gapped alignment, BLAST may in many cases miss relevant hits.

This is examplified in table 2 where proved off-targeted genes from Jackson where blasted against the whole human mRNA database. Jackson showed that as few as 5 contiguous identical nucleotides could lead to off-target effects. While `RNAplex` could retrieve all 9 off-targets validated by Jackson, BLAST returned only 3 of them. Moreover the time needed by `RNAplex` to scan the whole human mRNA database for undesirable matches was 80sec , which is fast enough for considering `RNAplex` as a more sensitive alternative to BLAST.

Accession Number	Identity	BLAST	RNAplex
NM_002271	14/15	YES	YES
NM_021033	14/15	YES	YES
NM_017748	11/15	YES	YES
A133672	8/13	NO	YES
NM_004165	10/12	NO	YES
NM_002946	5/11	NO	YES
NM_018457	9/10	NO	YES
NM_013242	8/10	NO	YES
AW237459	8/11	NO	YES

Table 2: BLAST and RNAplex performance at retrieving experimentally confirmed off-targets from Jackson [JBS+03]. Accession numbers are from NCBI Genbank database. The max length of contiguous identical nucleotides as well as the total number of identical nucleotides is shown in column 3. The last column shows the ability of RNAplex and BLAST to recover the off-targeted transcripts. NCBI BLAST was run with E=10000,wordsize=7 and Number of hits = 2000. 80 sec were necessary for RNAplex on a Intel(R) Core(TM)2 CPU 6600 @ 2.40GHz for finding putative hits

3.2 ncRNAs interactions with mRNAs

Regulatory RNAs often function by means of direct RNA-RNA binding. In order to gain insight in the propensity of ncRNAs to bind to mRNAs, we investigated whether RNAz predictions show an increased tendency of interacting with known mRNAs compared to dinucleotide shuffle RNAz hits. RNAplex was used to find putative targets for 1500 highly structured RNAz predictions in known human mRNAs. For each RNAz and shuffled hit, the most stable interaction was kept. However interactions involving RNAz hits with overlapping mRNAs were discarded. Interaction free energy distributions of the true and shuffled RNAz hits were tested against the null hypothesis of a common distribution using the Kolmogorov-Smirnov method. The null hypothesis was rejected with $p < 10^{-4}$. The densities of interaction free energy are shown in figure 3. A similar work was done in a recently published paper [BB+07]. In contrast to our approach which only used RNAplex to find putative targets, *Bompfünewerer et al.* located possible interaction partners with NCBI BLAST. This filtering step, which allows to rapidly scan for possible interactions, may also cause a loss in sensitivity when compared to the results returned by RNAplex. Interestingly one of the strongest interactions results from the hybridization of pre-Mir-219 with *YAP-1* mRNA. Given the high interactions energy, one might consider a siRNA-like function of pre-Mir-219 against *YAP-1*.

We have identified here a large number of evolutionary conserved structure ncRNA candidate genes that interact with mRNAs significantly stronger than random sequences.

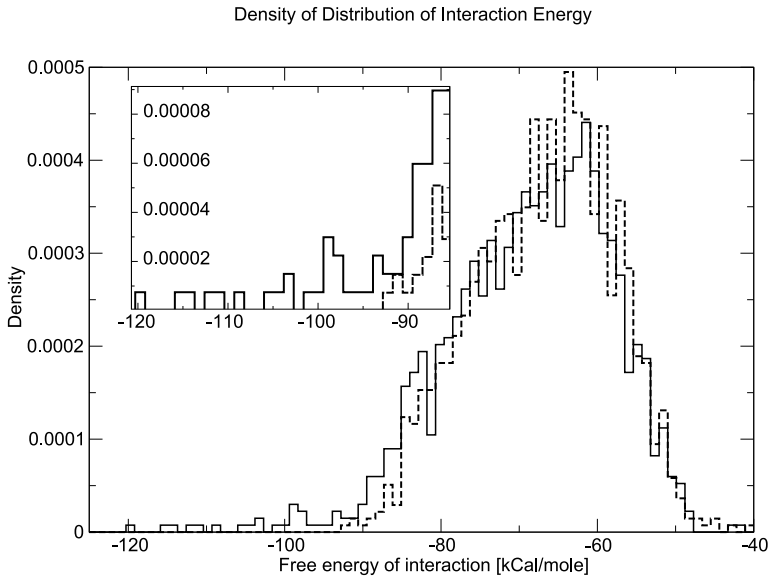


Figure 3: Density of distribution of interaction energy for true RNAz hits (solid line) and shuffled RNAz hits (broken line) against the human mRNA. The inset shows the tail of the distribution.

4 Discussion

The folding problem of more than one RNA strand can be treated at different levels of complexity. Because of the high computational cost of many algorithms for prediction of RNA-RNA interactions, target search may be best performed by a hierarchical search strategy, employing a series of filters that balance speed versus accuracy. Here we have introduced the program,`RNAplex`, which drastically reduces the search time of possible hybridization partners, mainly by neglecting intramolecular interactions. In the context of RNA-RNA interaction-search strategy, RNAplfold [BBB$^+$] can be used to filter out inaccessible target sites predicted by RNAplex. The remaining target sites can further be analyzed using tools such as `RNAup`, `RNAcofold` or `IRIS`.

5 Acknowledgments

This work was supported by `Siemens` and by the Wiener Wissenschafts-, Technologie- und Forschungsfonds.

References

[AA00] L Argaman and S Altuvia. 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, Jul 2000.

[AGM+90] SF. Altschul, W. Gish, W. Miller, EW. Myers, and DJ. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, Oct 1990.

[AKN+06] C. Alkan, E. Karakoc, JH. Nadeau, SC. Sahinalp, and K. Zhang. RNA-RNA interaction prediction and antisense RNA target search. *J Comput Biol*, 13(2):267–282, Mar 2006.

[Aku00] Y. Akutsu. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Discrete Applied Mathematics*, 104:45–62, 2000.

[AZC05] M. Andronescu, ZC. Zhang, and A. Condon. Secondary structure prediction of interacting RNA molecules. *J Mol Biol*, 345(5):987–1001, Feb 2005.

[BB+07] AF. , Bompfünewerer, R. Backofen, SH. Bernhart, C. Flamm, C. Fried, G. Fritzsch, J. Hackermüller, J. Hertel, IL. Hofacker, K. Missal, A. Mosig, SJ. Prohaska, D. Rose, PF. Stadler, A. Tanzer, S. Washietl, and S. Will. RNAs everywhere: genome-wide annotation of structured RNAs. *J Exp Zoolog B Mol Dev Evol*, 308(1):1–25, Jan 2007.

[BBB+] AF. Bompfunewerer, R. Backofen, SH. Bernhart, J. Hertel, IL Hofacker, PF. Stadler, and Sebastian Will. Variation on RNA Folding and Alignment. *J. Math. Biol.*

[BCH02] Jean Pierre Bachellerie, Jérôme Cavaillé, and Alexander Hüttenhofer. The expanding snoRNA world. *Biochimie*, 84(8):775–790, Aug 2002.

[Ben92] R. Benne. RNA editing in trypanosomes. The us(e) of guide RNAs. *Mol Biol Rep*, 16(4):217–227, Sep 1992.

[BS02] D. Banerjee and F. Slack. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays*, 24(2):119–129, Feb 2002.

[BTM+06] SH. Bernhart, H. Tafer, U. Mückstein, C. Flamm, PF. Stadler, and IL. Hofacker. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol*, 1(1):3–3, 2006.

[DEKM98] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis*. Cambridge University Press, 1998.

[ET90] Y. Eguchi and J. Tomizawa. Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of CoIE1 Rom protein. *Cell*, 60(2):199–209, Jan 1990.

[FXM+] A. Fire, S. Xu, MK. Montgomery, SA. Kostas, SE. Driver, and CC. Mello.

[GG06] W. Gerlach and R. Giegerich. GUUGle: a utility for fast exact matching under RNA complementary rules including G-U base pairing. *Bioinformatics*, 22(6):762–764, Mar 2006.

[HFS+94] IL. Hofacker, W. Fontana, PF. Stadler, SL. Bonhoeffer, M. Tacker, and P Schuster. Fast Folding and Comparison of RNA Secondary Structures. *Monatsh. Chem.*, 125:167–188, 1994.

[JBS+03] AL. Jackson, SR. Bartz, J. Schelter, SV. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, and PS. Linsley. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol*, 21(6):635–637, Jun 2003.

[KG07] JF. Kugel and JA. Goodrich. An RNA transcriptional regulator templates its own regulatory RNA. *Nat Chem Biol*, 3(2):89–90, Feb 2007.

[Mat04] DH. Mathews. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, 10(8):1178–1190, Aug 2004.

[MSZT99] DH. Mathews, J. Sabina, M. Zuker, and DH. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, 288(5):911–940, May 1999.

[MTH+06] U Mückstein, H Tafer, J Hackermüller, S H Bernhart, P F Stadler, and I L Hofacker. Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22(10):1177–1182, May 2006.

[Per04] D. D. Pervouchine. IRIS: intermolecular RNA interaction search. *Genome Inform*, 15(2):92–101, 2004.

[PL88] WR. Pearson and DJ. Lipman. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, 85(8):2444–2448, Apr 1988.

[RSHG04] M. Rehmsmeier, P. Steffen, M. Hochsmann, and R. Giegerich. Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10):1507–1517, Oct 2004.

[TS88] DH. Turner and N. Sugimoto. RNA structure prediction. *Annu Rev Biophys Biophys Chem*, 17:167–192, 1988.

[WHL+05] S Washietl, I L Hofacker, M Lukasser, A Hüttenhofer, and P F Stadler. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol*, 23(11):1383–1390, Nov 2005.

[WHS05] S Washietl, I L Hofacker, and P F Stadler. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A*, 102(7):2454–2459, Feb 2005.

[ZB97] K. Zorio, DAand Lea and T. Blumenthal. Cloning of Caenorhabditis U2AF65: an alternatively spliced RNA containing a novel exon. *Mol Cell Biol*, 17(2):946–953, Feb 1997.

[ZS81] M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9(1):133–148, Jan 1981.

[Zuk00] M. Zuker. Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, 10(3):303–310, Jun 2000.