Systematic Computational Analysis Of Potential Rnai Regulation In *Toxoplasma gondii*

Mehmet Volkan Çakır
Department of Molecular Biology and Genetics
Izmir Institute of Technology
Izmir, Turkey
mehmetcakir@iyte.edu.tr

Jens Allmer
Department of Molecular Biology and Genetics
Izmir Institute of Technology
Izmir, Turkey
jens@allmer.de

Abstract— **RNA interference (RNAi) is the mechanism through which RNA interferes with the production of other RNAs in a sequence specific manner. Micro RNA (miRNA) is a type of RNA which is transcribed as pri-miRNAs and processed to pre-miRNAs in the nucleus. These pre-miRNAs are then exported from the nucleus and processed in the cytoplasm to double stranded RNA with one strand providing target specificity..*Toxoplasma gondii* is a parasitic apicomplexan which causes several diseases. *T. gondii* is a good candidate for computational efforts with its small and publicly available genome files and extensive information about its gene structure. Although the existence of RNA interference in *T. gondii* is being debated, establishment of its complete potential RNAi regulatory network may be beneficial for further investigations into the topic.**

Keywords-RNAi, miRNA, RNAi regulation

I. INTRODUCTION

The term RNA interference (RNAi) refers to a cellular process where one strand of a double-stranded RNA (dsRNA) inhibits the expression of one or multiple target genes in a sequence specific manner. RNAi is a field of enormous current interest. The RNAi mechanism is initiated by dsRNA precursors that vary in length, origin and three dimensional structures. Short dsRNA ultimately become active as short single strands which guide recognition, cleavage or translational repression of complementary single-stranded RNAs (ssRNA), such as messenger RNAs (mRNA) or viral genomic/antigenomic RNAs [1]. It has been reported that RNAi mechanism may interfere with chromatin modification as well [2]. Based on the origin of the interfering RNA two major types of interfering RNAs can be differentiated, miRNAs (endogenous, micro RNAs) and siRNAs (exogenous, small interfering RNAs) [3,4]. Both endogenous and exogenous RNAi mechanisms require stepwise endonucleolytic cleavage of precursor dsRNA by specific RNase-III-type endonucleases, such as Drosha and Dicer, to generate approximately 20 to 30 base pairs (bp) long dsRNA with two nucleotide overhangs at both ends of the double strand [5]. The miRNA precursors are usually stem-loop structure forming noncoding transcripts with characteristic bulges and mismatches within the folded molecule which are thought to destabilize miRNA precursors and provide important features for processing [6,7]. Drosha cleaves next to the lower stem matches approximately 11 nucleotides from the SD (Single stranded RNA : Double stranded RNA) junction of the hairpin structure. SD junction is the border where single stranded RNA meets double stranded RNA. Dicer cleaves near the loop about 22 nucleotides from the Drosha cleavage site to generate a miRNA:miRNA* duplex [8,9,10,11,12,13]. After Dicer processing, one strand of the resulting short dsRNA duplex is incorporated into the RISC complex (RNA-induced silencing complex), which is a multi protein complex with the ability to incorporate ssRNA and with a slicer function, for targeting mRNAs by base pairing with the incorporated ssRNA [5]. In animal miRNA mechanism partial complementarity between miRNA and 3' untranslated region (UTR) of target mRNA, like those of *Caenorhabditis elegans*, often leads to transcriptional inhibition while in plants, miRNAs mostly interfere via cleavage of sequence-complementary mRNAs [1,14]. Recent studies suggest that miRNA mediated repression is accompanied by mRNA deadenylation, destabilization and mRNA decay [15,16,17,18,19].

Unfortunately, there are a limited number of experimental studies on RNAi regulation in apicomplexan parasites. However, studies indicate presence of RNAi regulation in apicomplexans. In spite of the debates on RNAi metabolism in *T. gondii* there are studies which suggest an RNAi regulation in *T. gondii* that resembles the one of eukaryotes [20]. Moreover, it has been reported that the genome of *T. gondii* contains candidate sequences with convincing similarity to RNAi genes [21]. Also existence of an inducible RNAi system in lower eukaryotes has been proposed by several studies [22,23,24]. Studies on RNAi regulation in *T. gondii* do not show much divergence from efforts on other apicomplexans. One encouraging study has shown downregulation of uracil phosphoribosyltransferase (UPRT) via introduction of dsRNAs [25]. A comprehensive study showing fully functional RNAi regulation in *T. gondii* remains elusive. Thus, computational efforts on RNAi regulation may provide beneficial information to the scientific community of these organisms.

In essence our study is to acquire a collection of potential endogenous miRNA interferences within *T. gondii*. Their

978-1-4244-5969-8/10/$26.00 ©2009 IEEE

sources and their potential targets as well as all potential miRNA and target sequences involved. Type of origin (whether they are exon, intron or non-coding sequences) and their locations in the genome of *T. gondii* is also of prime interest. By this way, the type and location of sequences which can potentially act as miRNAs and their potential can be inferred. The derived information enables the Toxoplasma community and scientists within the field of RNAi to use a potential RNAi regulation map of *T. gondii*. Furthermore, the software system developed in this study can be adjusted to be used on other genomes since it provides enough flexibility via its switches and commands. Almost all of the system thresholds and parameters can be fine tuned by the user. Additionally, implemented classes can be used separately in other systems of miRNA regulation analysis or the overall system can be used by just enriching its implementation with custom classes, methods or filtering steps since a genome-wide RNAi analysis entails a great amount of time and coding effort.

There have been several programs and tools designed for computational RNAi analysis. Some programs predict folding and secondary structure of RNA sequences by different approaches such as abstract shapes or evolutionary approaches such as RNAshapes [26,27,28,29], CONTRAfold [30], Mfold [31], pknotsRG [32], RNAfold [33,34,35], Sfold [36,37], Pfold [38,39], RNAcast [27], RNAforester [40,41], RNASampler [42], RNAalifold [43] to name only a few. Other programs predict RNA:RNA interactions like MicroTar [44], miTarget [45], RNAhybrid [46] and others. Finally, programs that accomplish other specific goals like guide strand prediction (e.g.: RISCbinder [47]) or identifying miRNA homologs (e.g.: miRNAminer [48]) need to be mentioned. However, all of the programs are designed to handle limited number of steps if not only one within the large number of steps necessary to achieve both RNAi source and target prediction. Most of the programs reach a good level of success in their predictions; nevertheless they certainly are not capable of carrying out a stepwise genome-wide RNAi analysis. The software developed in this study incorporates some of the tools and amends them with further steps to achieve fully automated genome wide prediction of potential RNA interference.

II. Materials and Methods

A. Programming Language

In order to run a genome wide analysis it is essential to have an automated system. Using a pre-implemented system or a pre-designed program is not a judicious way of fulfilling the step by step RNAi regulation analysis on the about 80 mb length *T. gondii* genome. Besides, there is no standart genome wide analysis program or system for *T. gondii*. In addition to programs which we incorporated into our system for many steps and for the overall automation we developed our own code. While all our implementations are written in Java™ programming language, incorporated programs may have been developed using other languages.

B. File Types

Our analysis starts with two types of files regarding *T. gondii* genome. One of these files is in FASTA format and the other is in GFF format which was defined by Sanger Institute (http://www.sanger.ac.uk/resources/software/gff/spec.html). Both files were obtained from ToxoDB – Toxoplasma gondii Genome Resource [49]. ToxoDB is a single organism database which contains publicly available annotated and raw data concerning *T. gondii*. FASTA file (TgondiiME49Genomic_ToxoDB-5.0.fasta) is the source for sequences while GFF (TgondiiME49_ToxoDB-5.0.gff) file is the source of sequence features that were mapped to the genome such as exons.

C. Included Programs in the System

Although we designed and implemented many new methods, we used some external and auxiliary scientific programs to achieve some steps in our system. While this reduces the flexibility of the system it ensures proper working sub systems and faster development. Pre-implemented programs in our system for the three steps are RNAshapes for folding RNA sequences into stem-loop and hairpin structures [26,27,28, 29], RNAhybrid for calculating free energy of two associated RNA sequences [50], and Basic Local Alignment Search Tool (BLAST, blastn) for searching the whole genome for antisense strands [51].

RNAshapes is freely available software package which folds single stranded RNA sequences and integrates three RNA analysis tools, namely the analysis of shape representatives, the calculation of shape probabilities and the consensus shapes approach [28]. For Microsoft Windows a graphical user interface and structure graph output are also included but are not mandatory for the system.

Most of the current RNA folding algorithms base their prediction on calculating a minimum free energy or a large number of potential suboptimal structures. Most such structures are redundant and therefore expensive in regard to algorithmic space and time demands. Current algorithms are collected and classified in a recent work by Gardner and Giegerich [52]. RNAshapes program uses shape representatives (shreps) which is the structure with the minimum free energy inside a shape class [28]. This is how it minimizes space and time requirements for the calculations. RNAshapes is a good candidate for our purposes allowing folding of pri-miRNAs to pre-miRNAs and calculation of minimum free energies as well as for obtaining stable hairpins for further analysis.

RNAhybrid, is a tool for finding the minimum free hybridisation energy between two RNA sequences. The tool is primarily meant as a program for evaluation of miRNA:target duplexes. In our case it was used to calculate the stability of the ends of Dicer products in regard to minimum free energy. This information was used to determine the less stable end which allows to decide which strand of the *miRNA:miRNA duplex will be degraded which one will be assembled into RISC. However, instead of selecting one strand of mature miRNA duplex, both strands were considered in our analysis as potential antisense strands. The calculated minimum free energy (mfe) values were recorded for further analysis and discrimination among results in the generated database. Both strands can be incorporated into RISC with different

Identify applicable sponsor/s here. If no sponsors, delete this text box. (sponsors)

probabilities according to their mfe values. Thus it is reasonable to investigate potential interactions of both strands.

Finally, BLAST, one of the most widely used bioinformatics algorithms for approximate comparison of sequences was used to determine possible interference targets. Basically, BLAST offers the ability to align a query sequence with a database of sequences, and to identify similar sequences within the database that reflect a desired local similarity to the query sequence above a designated threshold. Other miRNA:target recognition software come with restrictions they may for example consider interactions according to prior knowledge obtained from a specific target organism. However, *T. gondii* does not have any identified miRNAs which prohibits the use of such approaches in our system. We chose to use BLASTN for nucleotide searches since using more targeted tools for miRNA:target prediction, which have been trained on a different organism, may cause low sensitivity and low accuracy.

D. Step by Step Process

Overall system starts with the parsing of genomic files, both fasta and gff filesof *T. gondii*. Whole system containts a certain number of processing and filtering steps. Thresholds for filtering are calculated from 75-85 nucleotides long hairpin sequences from miRBase [53,54,55,56].

Initiative sequences of the system are 80 nucleotides long potential pri-miRNA sequences. Each and every 80 nucleotides sequence is considered as potential pri-miRNA source since there is no identified source or source related feature in *T. gondii*.

Potential pri-miRNA sequences folded by RNAshapes program using standard shape folding mode (-a). There are certain filtering steps regarding folded RNAs after running RNAshapes. Firstly, hairpins are filtered due to their Gibbs free energy value (kcal/mol) which is an output of RNAshapes program. Hairpins with a greater minimum free energy(mfe) value then -30 kcal/mol are filtered out in this step (see Fig.1). The hairpins are also filtered loosely according to their length of longest stretches, length of shapes, mismatches in longest stretches and mismatches in shapes(see Fig.2,3). Hairpins which have shapes shorter than 32 nucleotides and hairpins which have longest stretches shorter than 30 nucleotides are filtered out. Maximum number of mismatches allowed in shapes is 10 while maximum number of mismatches allowed in longest stretches is 7. Length of flanking ends are calculated as supplementary values (see Fig.4).

After folding and certain filtering we apply two artificial cleavages: Drosha and Dicer cleavage. All of the folded hairpins are subject to cleavages. Pinpointing the SD junction is fundamental to locate Drosha cleavage cite since the cleavage location of Drosha enzyme denotes a certain distance from SD junction. The distance is specified as 11 nucleotides from SD junction in coherence with the study of Han et al.[9]. To pinpoint the SD junction, lower stem matches of hairpin is located. The minimum length of the matching double stranded RNA stretch to be considered as SD junction is specified as 3(see Fig.5). Following Drosha cleavage, artificial Dicer cleavage is carried out at the location which is 22 nucleotides away from Drosha cleavage cite in consistence with several studies [57,58,59,60].

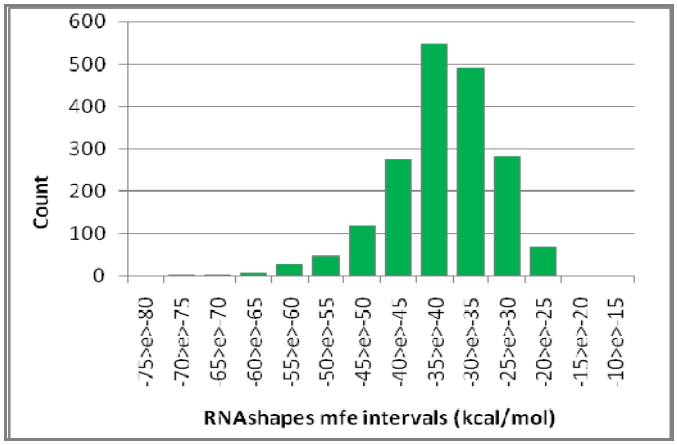

Figure 1. Mfe values of folded miRBase hairpins. Hairpins are folded by RNAshapes program as in our system. Horizontal axis denotes intervals of mfe (kcal/mol) while vertical axis denotes number of elements in each interval.

Dicer cleavage produces approximately 22 nucleotides long RNA duplexes with 2 nucleotide 3' end overhangs. Then RNAhybrid program is hired to calculate mfe values of three parts of duplexes which are arbitrarily named left, middle and right just to designate three regions of duplexes with same length. To be able to decide which strand of RNA duplex(Dicer product) will be considered as antisense strand and which strand of duplex will be considered as guide strand it is fundamental to know mfe distribution since the strand which possess less stability on its 5' end has more chance to be incorporated into RISC complex [7,61,62]. Also regional mfe values of miRBase hairpins are calculated (see Fig.6).

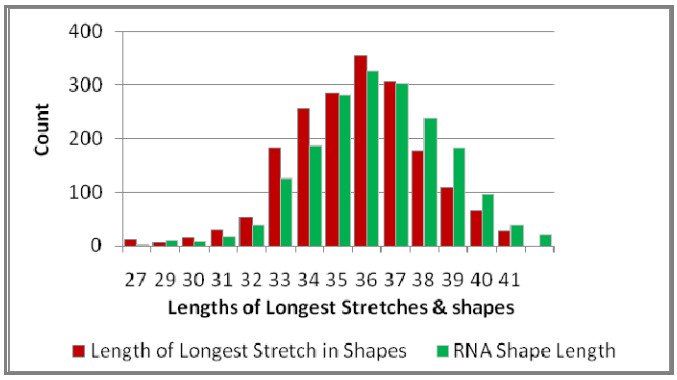

Figure 2. Lengths of shapes and longest stretches of miRBase hairpins after folding by RNAshapes. Horizontal axis denotes lengths of shapes/longest stretches while vertical axis denotes number of hairpins with corresponding length.

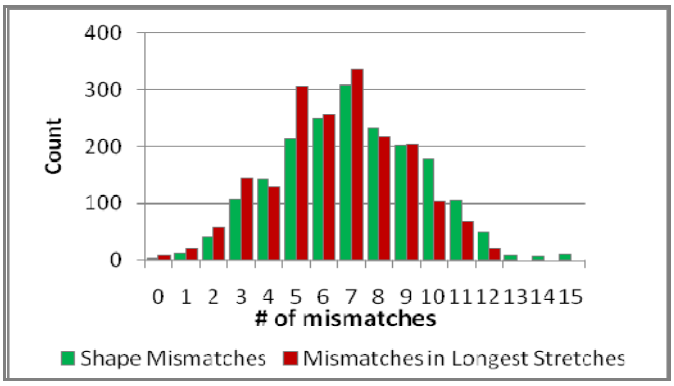

Figure 3. Mismatches in shapes of miRBase hairpins after folding by RNAshapes. Horizontal axis denotes the number of mismatches in shape representations while vertical axis denotes number of hairpins with corresponding number of mismatches.

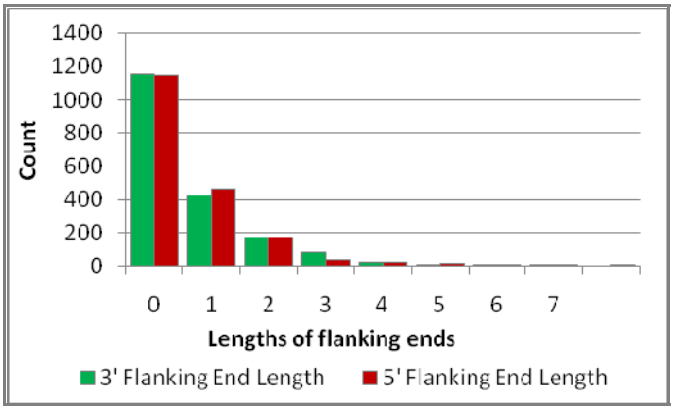

Figure 4. Length of the 5' and 3' flanking ends of miRBase hairpins. Horizontal axis denotes the length of flanking ends while vertical axis denotes number of hairpins with corresponding flanking end length.

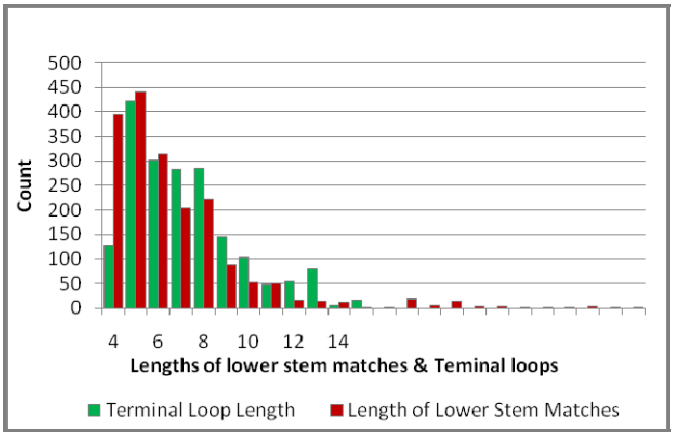

Figure 5. Length of lower stem matches and terminal loops in folded hairpins. Horizontal axis denotes the number of matches in the lower stem and terminal loop. Lower stem is the region from SD junction to closes mismatch. Vertical axis denotes the number of hairpins with corresponding length of lower stem matches and terminal loops.

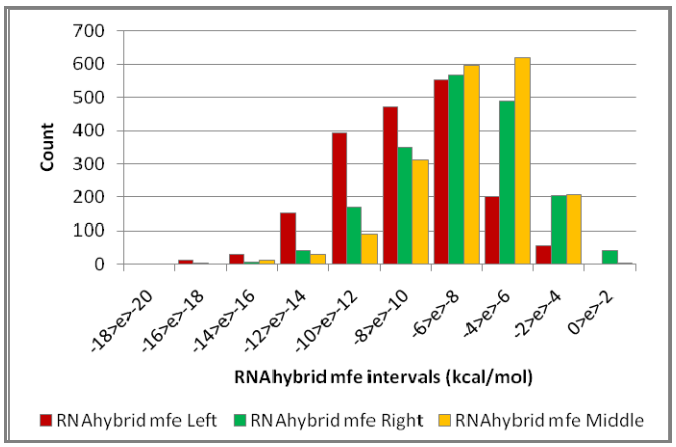

Figure 6. Regional mfe values of miRBase hairpins calculated by RNAhybrid. Horizontal axis denotes the energy intervals of regions while vertical axis denotes the number of hairpins in corresponding interval.

Resultant RNA duplexes of Dicer cleavage contain two potential mature miRNAs. Potential interactions of these mature miRNAs are investigated in the whole genome of *T. gondii* by BLAST. An e-value threshold of 0.03 is hired since a 22 nucleotide long miRNA with 17 nucleotide partial match can end up with hits which may possess e-value up to 0.023. So we expect at least 17 matches between potential miRNAs and their targets. Also we controlled mismatches an gaps seperately in order to limit results. We let 4 maximum mismatches and 4 maximum gaps in miRNA:target interaction. However gaps and mismathces can be at most 5 at total. At the end we have potential miRNA sources, potential miRNA targets and their potential interactions with certain values.

III. Results and Discussion

The aim of our study is to identify all potential miRNA sources and all their potential interaction targets. The significance of the hairpins identified in this study are evaluated by the values obtained from previously, experimentally identified hairpin sequences. The hairpin sequences are obtained from the file hairpin.fa from miRBase [53,54,55,56]. For a number of measurable properties distributions have been calculated which are used assessment of potential miRNA interactions of *T.gondii*(see Section Step by Step Process).

Since there is no identified Drosha homologues in *T.gondii* our first interest as miRNA source is intronic sequences which can bypass certain enzymatic cleavages [63]. Nevertheless, we have considered all possible interactions which are; intron→3'UTR, exon→3'UTR, intergenic region→3'UTR, intron→exon, exon→exon, intergenic region→exon, intron→intron, exon→intron and intergenic region→intron. We did not expect all of these interactions to be biologically prospective. Moreover we did not get results from all of the interactions. We have results for only four of the interactions that are intron→3'UTR, intron→exon, exon→exon and exon→3'UTR. Intron→exon and exon→exon interactions could be possible in the case of alternative splicing. We have identified 89532 total interactions between intronic sources and 3'UTR targets, 61376 total interactions between exonic sources

and 3'UTR targets, 9895 total interactions between exonic sources and exonic targets, 13731 total interactions between intronic sources and exonic targets.

We have two peaks for minimum free energy(mfe) values of our hairpins from different sources, between intervals -40kcal/mole to -35kcal/mole and -35kcal/mole to -30kcal/mole as expected. However our data gives the highest number of hairpins between -35kcal/mole and -30kcal/mole interval while miRBase hairpins exert highest number of hairpins between -40kcal/mole and -35kcal/mole (see Fig.1,7).

The reason of the slight shift between miRBase hairpins and suggested hairpins by our data can be the difference of hairpin lengths between two miRBase hairpins and intronic hairpins. We analyzed hairpins which are longer than 75 nucleotides and shorter than 85 nucleotides from miRBase. However all of our hairpins from *T.gondii* genome are 80 nucleotides in length. The other reason can be the difference between previously identified RNAi mechanisms and the RNAi mechanism of *T.gondii*. Since there is not much known about RNAi regulation in T.gondii the details of RNAi regulation in T.gondii can not be accessed precisely. Another reason should be the number of sequences with tandem repeats identified by our system. Since repetitive sequences intrinsically hold potential of folding into "good" hairpins they arise as candidate sources for miRNAs. Nevertheless, repetitive sequences' sequence composition is a factor which decreases their specifity of interaction as well as it introduces them as potential sources. That is why we include nucleotide counts in our system. Sequences with certain composition can be easily filtered out due to their dinucleotide(see Fig.8,9) and trinucleotide(see Fig.8) counts.

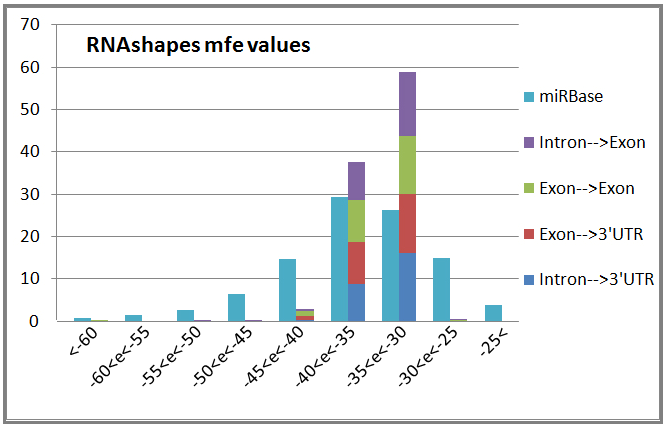

Figure 7. Mfe values of source hairpins calculated by RNAshapes. Horizontal axis denotes the energy intervals of regions while vertical axis denotes the number of hairpins in corresponding interval(values of vertical axis are weighted by the total number of hairpins).

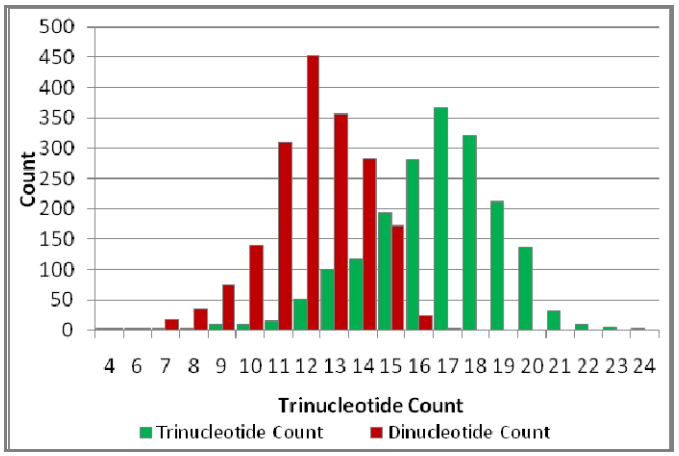

Figure 8. Number of different dinucleotides and trinucleotides of miRBase hairpins.

MirBase hairpins exert distributions around 12 for dinucleotide count and 17 for trinucleotide count. However our data reflects highest number of hairpins with 7 dinucleotide count(see Fig.9). This difference is because of repetitive sequences most possibly.

There is again a slight shift in shape lengths between miRBase hairpins and T.gondii potential hairpins (see Fig.2,10).

Repetitive sequences fold into hairpins with less mismatches and short terminal loops hence end up with longer hairpin shapes which are produced by RNAshapes. Data about mismatches in hairpins affirm this slight difference (see Fig.11).

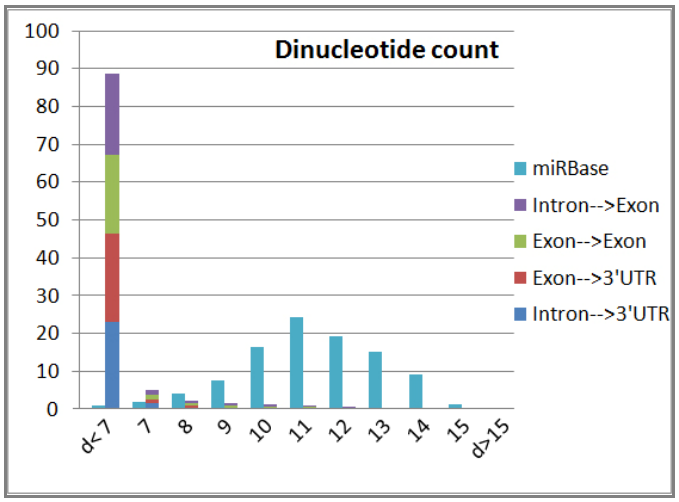

Figure 9. Number of different dinucleotides of potential miRNAs of T.gondii(values of vertical axis are weighted by the total number of hairpins).

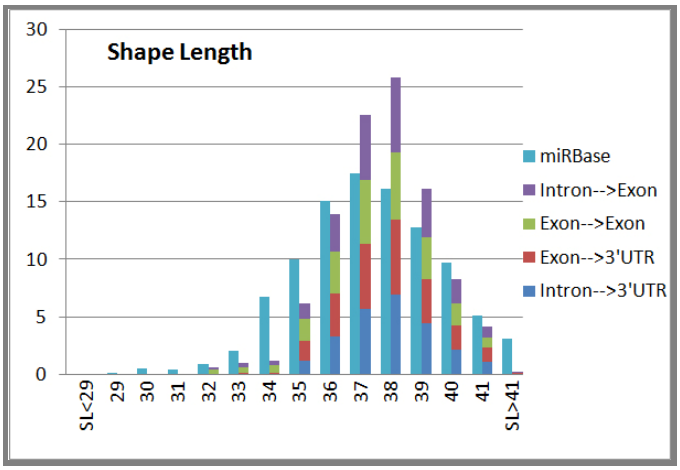

Figure 10. Shape length of intronic hairpins of *T.gondii*. Shapes are produced by RNAshapes program as mentioned above.

With certain thresholds and restrictions potential miRNA sources and their targets can be accessed easily in our system.

An example of significant interactions seem to occur between third intron of TGME49_078070 gene which is a conserved hypothetical protein and 3'UTR region of TGME49_078440 gene which is a transcription regulatory protein SNF2(ATP-dependant helicase). Both of the genes are on chromosome XII. Minimum free energy value of the hairpin which derives from the TGME49_078070 gene is -30, dinucleotide count of miRNA which derives that hairpin is 12 while trinucleotide count is 17. Mfe value, dinucleotide and trinucleotide counts fit quite well to the miRBase assessment (see Fig.1,8). Shape length of potential hairpin of TGME49_078070 gene is(39 nucleotides long), regional mfe values are left (-9kcal/mole), middle(-6.4kcal/mole), right(-3.7 kcal/mole), length of potential miRNA(22 nucleotides long) also verifies miRBase assessment. MiRNA of TGME49_078070 gene and its target TGME49_078440 gene's 3'UTR region exerts a perfect match interaction with a BLAST e-value of 3e-005 that also confirms the specifity of interaction.

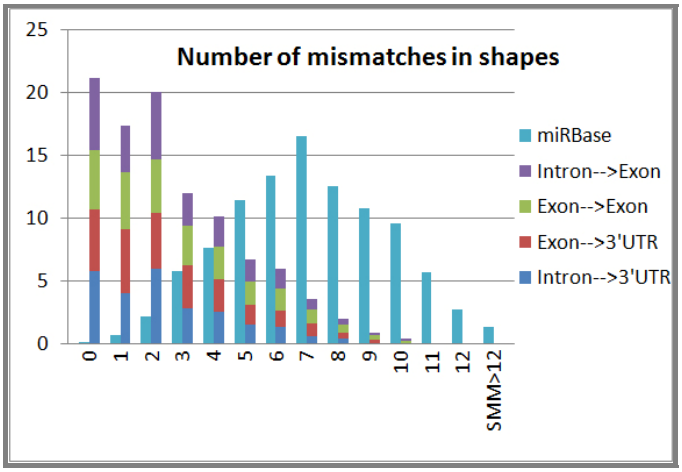

Figure 11. Mismatches in shapes of hairpins of *T.gondii*. (values of vertical axis are weighted by the total number of hairpins). Shapes are produced by RNAshapes program as mentioned above.

IV. Conclusion

We presented a fully automated system for the prediction of RNA interference given a FASTA file with the genome and an associated GFF file with the mapped features. The system was tested using *T. gondii* since it has a small practical genome size and in order to establish a complete RNA interference network so that the current discussion about whether RNA interference exists or not can be tested more easily in the future.

Analysis of the resulting regulatory network shows that its features are in accordance with expectations for established networks. This was better for the already established RNA interference pathways and significantly worse for interactions other than the expected ones. This underlines the possibility of RNAi regulation in *T. gondii* and the effectiveness of our system.

V. Outlook

In the future we intend to calculate the minimum free energy of all source-target RNA complexes found with our method to further filter the results . Also we plan to establish an easy to host database application which may be incorporated into existing databases like ToxoDB. The analysis of an organism with more experimental examples for RNA interference like for *Chlamydomonas reinhardtii* with our system is also planned for the near future.

Acknowledgment

We are indepted to the reviewers who informed us about an additional method to further scrutinize the results obtained in the last step of our program. We also would like to thank Bünyamin Akgül and Bilge Karaçalı for great discussion about the project.

References

[1] G. Meister, and T. Tuschl, "Mechanisms of gene silencing by double-stranded RNA," Nature, vol. 431, pp. 343-349, Sep 16. 2004.

[2] Z. Lippman, and R. Martienssen, "The role of RNA interference in heterochromatic silencing," Nature, vol. 431, pp. 364-370, Sep 16. 2004.

[3] E.J. Chapman, and J.C. Carrington, "Specialization and evolution of endogenous small RNA pathways," Nat Rev Genet, vol. 8, pp. 884-896, Nov. 2007.

[4] P.D. Zamore, and B. Haley, "Ribo-gnome: the big world of small RNAs," Science, vol. 309, pp. 1519-1524, Sep 2. 2005.

[5] M. Nowotny, and W. Yang, "Structural and functional modules in RNA interference," Curr Opin Struct Biol, vol. 19, pp. 286-293, Jun. 2009.

[6] G. Hutvagner, and P.D. Zamore, "A microRNA in a multiple-turnover RNAi enzyme complex," Science, vol. 297, pp. 2056-2060, Sep 20. 2002.

[7] A. Khvorova, A. Reynolds, and S.D. Jayasena, "Functional siRNAs and miRNAs exhibit strand bias," Cell, vol. 115, pp. 209-216, Oct 17. 2003.

[8] S.M. Elbashir, W. Lendeckel, and T. Tuschl, "RNA interference is mediated by 21- and 22-nucleotide RNAs," Genes & Development, vol. 15, pp. 188-200, January 15, 2001. 2001.

[9] J. Han, Y. Lee, K.H. Yeom, J.W. Nam, I. Heo, J.K. Rhee, S.Y. Sohn, Y. Cho, B.T. Zhang, and V.N. Kim, "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex," Cell, vol. 125, pp. 887-901, Jun 2. 2006.

[10] G. Hutvagner, J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore, "A cellular function for the RNA-interference enzyme

Dicer in the maturation of the let-7 small temporal RNA," Science, vol. 293, pp. 834-838, Aug 3. 2001.

[11] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V.N. Kim, "The nuclear RNase III Drosha initiates microRNA processing," Nature, vol. 425, pp. 415-419, Sep 25. 2003.

[12] I.J. Macrae, K. Zhou, F. Li, A. Repic, A.N. Brooks, W.Z. Cande, P.D. Adams, and J.A. Doudna, "Structural basis for double-stranded RNA processing by Dicer," Science, vol. 311, pp. 195-198, Jan 13. 2006.

[13] Y. Tomari, and P.D. Zamore, "Perspective: machines for RNAi," Genes Dev, vol. 19, pp. 517-529, Mar 1. 2005.

[14] C. Lu, D.H. Jeong, K. Kulkarni, M. Pillay, K. Nobuta, R. German, S.R. Thatcher, C. Maher, L. Zhang, D. Ware, B. Liu, X. Cao, B.C. Meyers, and P.J. Green, "Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs)," Proc Natl Acad Sci U S A, vol. 105, pp. 4951-4956, Mar 25. 2008.

[15] S. Bagga, J. Bracht, S. Hunter, K. Massirer, J. Holtz, R. Eachus, and A.E. Pasquinelli, "Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation," Cell, vol. 122, pp. 553-563, Aug 26. 2005.

[16] I. Behm-Ansmant, J. Rehwinkel, T. Doerks, A. Stark, P. Bork, and E. Izaurralde, "mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes," Genes Dev, vol. 20, pp. 1885-1898, Jul 15. 2006.

[17] A.J. Giraldez, Y. Mishima, J. Rihel, R.J. Grocock, S. Van Dongen, K. Inoue, A.J. Enright, and A.F. Schier, "Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs," Science, vol. 312, pp. 75-79, Apr 7. 2006.

[18] L. Wu, and J.G. Belasco, "Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells," Mol Cell Biol, vol. 25, pp. 9198-9208, Nov. 2005.

[19] L. Wu, J. Fan, and J.G. Belasco, "MicroRNAs direct rapid deadenylation of mRNA," Proc Natl Acad Sci U S A, vol. 103, pp. 4034-4039, Mar 14. 2006.

[20] L. McRobert, and G.A. McConkey, "RNA interference (RNAi) inhibits growth of Plasmodium falciparum," Mol Biochem Parasitol, vol. 119, pp. 273-278, Feb. 2002.

[21] E. Ullu, C. Tschudi, and T. Chakraborty, "RNA interference in protozoan parasites," Cell Microbiol, vol. 6, pp. 509-519, Jun. 2004.

[22] P. Bastin, K. Ellis, L. Kohl, and K. Gull, "Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system," J Cell Sci, vol. 113 (Pt 18), pp. 3321-3328, Sep. 2000.

[23] H. Shi, A. Djikeng, T. Mark, E. Wirtz, C. Tschudi, and E. Ullu, "Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA," RNA, vol. 6, pp. 1069-1076, Jul. 2000.

[24] Z. Wang, J.C. Morris, M.E. Drew, and P.T. Englund, "Inhibition of Trypanosoma brucei Gene Expression by RNA Interference Using an Integratable Vector with Opposing T7 Promoters," Journal of Biological Chemistry, vol. 275, pp. 40174-40179, December 22, 2000. 2000.

[25] F. Al-Anouti, and S. Ananvoranich, "Comparative analysis of antisense RNA, double-stranded RNA, and delta ribozyme-mediated gene regulation in Toxoplasma gondii," Antisense Nucleic Acid Drug Dev, vol. 12, pp. 275-281, Aug. 2002.

[26] R. Giegerich, B. Voss, and M. Rehmsmeier, "Abstract shapes of RNA," Nucleic Acids Res, vol. 32, pp. 4843-4851. 2004.

[27] J. Reeder, and R. Giegerich, "Consensus shapes: an alternative to the Sankoff algorithm for RNA consensus structure prediction," Bioinformatics, vol. 21, pp. 3516-3523, Sep 1. 2005.

[28] P. Steffen, B. Voss, M. Rehmsmeier, J. Reeder, and R. Giegerich, "RNAshapes: an integrated RNA analysis package based on abstract shapes," Bioinformatics, vol. 22, pp. 500-503, Feb 15. 2006.

[29] B. Voss, R. Giegerich, and M. Rehmsmeier, "Complete probabilistic analysis of RNA shapes," BMC Biol, vol. 4, pp. 5. 2006.

[30] C.B. Do, D.A. Woods, and S. Batzoglou, "CONTRAfold: RNA secondary structure prediction without physics-based models," Bioinformatics, vol. 22, pp. e90-98, July 15, 2006. 2006.

[31] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," Nucl. Acids Res., vol. 31, pp. 3406-3415, July 1, 2003. 2003.

[32] J. Reeder, and R. Giegerich, "Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics," BMC Bioinformatics, vol. 5, pp. 104, Aug 4. 2004.

[33] I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast folding and comparison of RNA secondary structures," Monatshefte für Chemie / Chemical Monthly, vol. 125, pp. 167-188. 1994.

[34] J.S. McCaskill, "The equilibrium partition function and base pair binding probabilities for RNA secondary structure," Biopolymers, vol. 29, pp. 1105-1119. 1990.

[35] M. Zuker, and P. Stiegler, "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," Nucleic Acids Res, vol. 9, pp. 133-148, Jan 10. 1981.

[36] Y. Ding, C.Y. Chan, and C.E. Lawrence, "RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble," RNA, vol. 11, pp. 1157-1166, Aug. 2005.

[37] Y. Ding, and C.E. Lawrence, "A statistical sampling algorithm for RNA secondary structure prediction," Nucleic Acids Res, vol. 31, pp. 7280-7301, Dec 15. 2003.

[38] B. Knudsen, and J. Hein, "RNA secondary structure prediction using stochastic context-free grammars and evolutionary history," Bioinformatics, vol. 15, pp. 446-454, June 1, 1999. 1999.

[39] B. Knudsen, and J. Hein, "Pfold: RNA secondary structure prediction using stochastic context-free grammars," Nucleic Acids Res, vol. 31, pp. 3423-3428, Jul 1. 2003.

[40] M. Hochsmann, T. Toller, R. Giegerich, and S. Kurtz, "Local similarity in RNA secondary structures," Proc IEEE Comput Soc Bioinform Conf, vol. 2, pp. 159-168. 2003.

[41] M. Hochsmann, B. Voss, and R. Giegerich, "Pure multiple RNA secondary structure alignments: a progressive profile approach," IEEE/ACM Trans Comput Biol Bioinform, vol. 1, pp. 53-62, Jan-Mar. 2004.

[42] X. Xu, Y. Ji, and G.D. Stormo, "RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment," Bioinformatics, vol. 23, pp. 1883-1891, Aug 1. 2007.

[43] S.H. Bernhart, I.L. Hofacker, S. Will, A.R. Gruber, and P.F. Stadler, "RNAalifold: improved consensus structure prediction for RNA alignments," BMC Bioinformatics, vol. 9, pp. 474. 2008.

[44] R. Thadani, and M.T. Tammi, "MicroTar: predicting microRNA targets from RNA duplexes," BMC Bioinformatics, vol. 7 Suppl 5, pp. S20. 2006.

[45] S.K. Kim, J.W. Nam, J.K. Rhee, W.J. Lee, and B.T. Zhang, "miTarget: microRNA target gene prediction using a support vector machine," BMC Bioinformatics, vol. 7, pp. 411. 2006.

[46] M. Rehmsmeier, P. Steffen, M. Hochsmann, and R. Giegerich, "Fast and effective prediction of microRNA/target duplexes," RNA, vol. 10, pp. 1507-1517, Oct. 2004.

[47] F. Ahmed, H.R. Ansari, and G.P. Raghava, "Prediction of guide strand of microRNAs from its sequence and secondary structure," BMC Bioinformatics, vol. 10, pp. 105. 2009.

[48] S. Artzi, A. Kiezun, and N. Shomron, "miRNAminer: a tool for homologous microRNA gene search," BMC Bioinformatics, vol. 9, pp. 39. 2008.

[49] J.C. Kissinger, B. Gajria, L. Li, I.T. Paulsen, and D.S. Roos, "ToxoDB: accessing the *Toxoplasma gondii* genome," Nucleic Acids Res, vol. 31, pp. 234-236, Jan 1. 2003.

[50] J. Kruger, and M. Rehmsmeier, "RNAhybrid: microRNA target prediction easy, fast and flexible," Nucleic Acids Res, vol. 34, pp. W451-454, Jul 1. 2006.

[51] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic local alignment search tool," J Mol Biol, vol. 215, pp. 403-410, Oct 5. 1990.

[52] P.P. Gardner, and R. Giegerich, "A comprehensive comparison of comparative RNA structure prediction approaches," BMC Bioinformatics, vol. 5, pp. 140, Sep 30. 2004.

[53] V. Ambros, B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones, M. Marshall, M.

Matzke, G. Ruvkun, and T. Tuschl, "A uniform system for microRNA annotation," RNA, vol. 9, pp. 277-279, Mar. 2003.

[54] S. Griffiths-Jones, "The microRNA Registry," Nucleic Acids Res, vol. 32, pp. D109-111, Jan 1. 2004.

[55] S. Griffiths-Jones, R.J. Grocock, S. van Dongen, A. Bateman, and A.J. Enright, "miRBase: microRNA sequences, targets and gene nomenclature," Nucleic Acids Res, vol. 34, pp. D140-144, Jan 1. 2006.

[56] S. Griffiths-Jones, H.K. Saini, S. van Dongen, and A.J. Enright, "miRBase: tools for microRNA genomics," Nucleic Acids Res, vol. 36, pp. D154-158, Jan. 2008.

[57] X. Ji, "The mechanism of RNase III action: how dicer dices," Curr Top Microbiol Immunol, vol. 320, pp. 99-116. 2008.

[58] R.F. Ketting, S.E.J. Fischer, E. Bernstein, T. Sijen, G.J. Hannon, and R.H.A. Plasterk, "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans," Genes & Development, vol. 15, pp. 2654-2659, October 15, 2001. 2001.

[59] S.W. Knight, and B.L. Bass, "A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans," Science, vol. 293, pp. 2269-2271, Sep 21. 2001.

[60] I.J. MacRae, K. Zhou, and J.A. Doudna, "Structural determinants of RNA recognition and cleavage by Dicer," Nat Struct Mol Biol, vol. 14, pp. 934-940, Oct. 2007.

[61] D.S. Schwarz, G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P.D. Zamore, "Asymmetry in the assembly of the RNAi enzyme complex," Cell, vol. 115, pp. 199-208, Oct 17. 2003.

[62] Y. Tomari, T. Du, and P.D. Zamore, "Sorting of Drosophila small silencing RNAs," Cell, vol. 130, pp. 299-308, Jul 27. 2007.

[63] J.G. Ruby, C.H. Jan, and D.P. Bartel, "Intronic microRNA precursors that bypass Drosha processing," Nature, vol. 448, pp. 83-86. 2007.

[64]