

Phase Transitions in Gene Knockdown Networks of Transitive RNAi

Shibin Qiu and Terran Lane

Dept. of Computer Science, University of New Mexico, Albuquerque, NM, 87131
{squi, terran}@cs.unm.edu

Abstract. Since gene silencing by RNA interference (RNAi) has been observed when inexact matches exist in siRNA-mRNA binding, the number of mismatched nucleotides allowed by nature becomes an important quantity in characterizing RNAi specificity. We use scale-free graphs to model the knockdown interactions among different genes and estimate the allowable flexibility by examining transitive RNAi, which amplifies siRNA and cyclically silences targets. Simulation results in *S. pombe* indicate that continually increasing the number of mismatches risks transcriptome-wide knockdown and eventually turns RNAi from defensive to self-destructive. At the phase transition, the number of mismatches indicates a critical value beyond which tRNAi would cause an organism unstable. This critical value suggests an upper limit of no more than 6 nt mismatches in the binding.

1 Introduction

RNAi becomes unspecific when close or exact identity between siRNA and non-target mRNA produces “off-target” degradation. Empirical studies have demonstrated that such off-target effects occur in the presence of single and multiple nucleotide mismatches between siRNA and off-target mRNA [1, 2, 3, 4]. It is conceivable that an siRNA-mRNA binding that is too permissive might knock down a substantial portion of genes in an organism and cause loss of function of too many genes, which would in turn make an organism unstable. Since such instability is not observed, there must be an upper limit on the flexibility allowed by nature. To date, it is unclear how much flexibility is actually allowed in this binding function. Here we estimate the allowable mismatches by examining the knockdown interactions of transitive RNAi using a network model.

In transitive RNAi (tRNAi), gene knockdown is mediated by ~ 21 nucleotide (nt) siRNAs produced by enzyme Dicer from an initiating dsRNA. These siRNAs also serve as primers for RNA-directed RNA polymerase (RdRP) to generate elongated dsRNA from the target mRNA, which are cleaved again by Dicer to produce secondary siRNAs, yielding degradation of secondary targets [5]. tRNAi (or post-transcriptional gene silencing, PTGS) has been observed in plants, fungi, flies, and worms [5, 6, 7]. tRNAi is highly efficient for gene silencing and has been termed as “nature’s own PCR machine” due to its siRNA amplification [6]. But

siRNA proliferation in tRNAi also increases the chances of undesirable knock-downs, especially when the siRNA-mRNA binding is not stringent.

In *D. melanogaster*, where tRNAi has been observed [6], mismatches of 3 or 4 nt between siRNA and mRNA were tolerated [1]. Although tRNAi has not been confirmed in mammals, mutations in siRNA of 3 to 4 nt were allowed [2]. Moreover, in some cases siRNAs with 9 nt mismatches have been shown to directly silence non-target genes in cultured human cells [3]. In RNAi experiments, mismatches can also be intentionally introduced during siRNA synthesis to improve silencing efficacy, as RISC prefers lower binding energy at one of the termini of siRNA [8]. However, excessive flexibility is dangerous since it may silence too many unintended targets and, more seriously, may cause instability.

Since tRNAi may cause chain reactions, and its effects are difficult to fully observe [5, 9], it is expensive, if feasible, to experimentally investigate tRNAi at a genomic scale by silencing each gene as primary target. However, a computational simulation is able to model the knockdown interactions at a transcriptomic scale. In this work, we construct *gene knockdown networks* based on principles and parameters observed in biological experiments to investigate the impact of mismatches on genome stability.

We find that the knockdown network became dense when less stringent RNAi bindings were employed in our simulation and a giant component (a large group of connected genes) emerged. In a knockdown network, genes connected to a common node can be silenced by a single primary siRNA and the giant component can be silenced with only a few primary siRNAs. This phase transition with respect to binding permissiveness will turn RNAi from defensive into self-destructive and imply a loss of genomic stability. Such instability is not observed in nature, implying that the biological siRNA-mRNA binding function does not support a large number of mismatches in general. This phase transition suggests a critical value for the permitted flexibility of RNAi binding. Due to evolutionary conservations, the results can be used as a references for allowable mismatches in organisms not having tRNAi.

2 Network Model of Transitive RNAi

In this section, we describe the biological mechanism of tRNAi and the construction of our network model.

2.1 Transitive RNAi

We can outline dsRNA-mediated tRNAi in the following steps [5].

1. A dsRNA is introduced into the cell to initiate gene silencing;
2. Dicer cleaves the dsRNA into siRNAs of length 21-25 nt;
3. RISC unwinds the siRNA to help target the appropriate mRNA;
4. RISC-siRNA complex recognizes the primary target mRNA;
5. RISC cleaves the siRNA/mRNA hybrid and the primary target is degraded;
6. RdRP produces elongated dsRNA from the target site on the primary target;

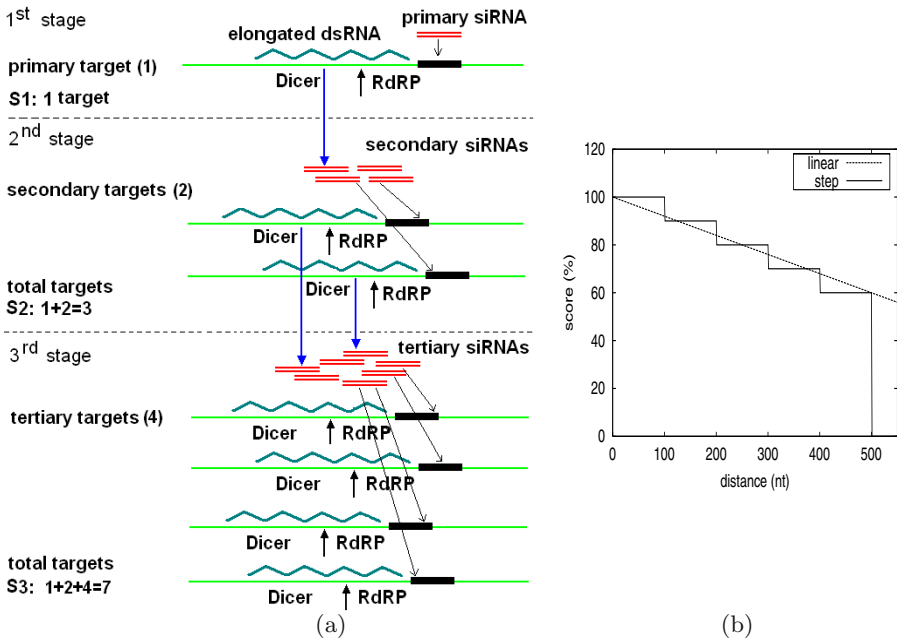


Fig. 1. Transitive RNAi. (a) First three stages of tRNAi, silencing 7 targets with a single primary siRNA. (b) Scoring schemes for secondary siRNAs.

7. Dicer cleaves the elongated dsRNA into secondary siRNAs;
8. RISC destroys the secondary targets using the secondary siRNAs. Now the secondary targets become primary targets, and tRNAi cycles to step 3.

In addition to degrading the primary target mRNA as in ordinary RNAi, the siRNA hybridized with its complimentary target can be used by RdRP to produce an elongated dsRNA upstream of the target site (step 6). Dicer then generates more siRNAs, the secondary siRNAs, using the elongated dsRNA (step 7). These secondary siRNAs associate with RISC which knocks down secondary targets (step 8). The secondary siRNAs are in turn elongated to generate tertiary siRNAs, which silence tertiary targets. Since tRNAi is cyclic [5], it may cause a chain reaction of gene knockdown [5]. Figure 1 (a) shows tRNAi of the first three stages, where seven targets are silenced with a single primary siRNA.

Even though G-U wobbles and bulges may also be permitted in siRNA-target binding [2], mismatches contribute the most to off-targeting effects [10]. We used contiguous mismatches, because they are frequently employed experimentally and are computationally easy to implement [1, 2, 3].

2.2 Simulation of Transitive RNAi

We randomly pick an siRNA from a gene g_i , $i = 1, 2, \dots, N$, where N is the number of genes in the transcriptome of an organism. And find its target genes. We introduce variable mismatches in the siRNA-mRNA binding as a parameter.

This siRNA is the primary siRNA and its targets are primary targets (including off-targets). We denote the collection of primary targets as T_i^1 . We then pick a 500 nt elongated dsRNA upstream of the binding site from each gene in T_i^1 (see explanations later in this section). Overlapping secondary siRNAs are generated at all possible positions from these elongated dsRNAs. We assign differential silencing scores to the secondary siRNAs, with higher scores being assigned to siRNAs closer to the binding site on the primary target and lower scores to those farther away from the binding site [5]. We employ two scoring schemes, i.e. linear and the step-wise, as shown in Figure 1(b).

We search for the secondary targets T_i^2 matched by the secondary siRNAs using a filter based on a silencing score threshold of $f_s=90\%$ (see explanations later in this section). Since tRNAi is cyclic [5], we continue the target search by using the secondary targets to produce tertiary siRNAs and finding the tertiary targets T_i^3 . Figure 1(a) shows 4 secondary siRNAs and 7 tertiary siRNAs. The induced targets, when gene g_i is aimed primarily, is the collection of the targets at each stage. If we observe tRNAi up to the t^{th} stage, then the total targets are in $S_i^t = \bigcup_{j=1}^t T_i^j$. For example, if $t = 3$, then S_i^3 contains primary, secondary, and tertiary targets when g_i is the primary target of the siRNA.

Through siRNA amplification and binding flexibility, silencing g_i with a primary siRNA will likely knock down other genes. The knockdown relationships between genes can be represented by a graph, as described below. As every gene in S_i^t will be degraded as a consequence of a primary siRNA when g_i is intended to be silenced, we add an edge from g_i to each gene in S_i^t . We conduct a target search for each gene of an organism and construct a directed graph G_t as the knockdown network of t -stage tRNAi. The graph G_t is parameterized by m , length of mismatch allowed by the siRNA-target binding, and t , the number of stages. We then study the statistics of G_t and compute the largest connected component of G_t . The phase transition of $G - t$ is used to estimate the allowable mismatches in the binding. We average the statistics over 100 networks, each of which is initiated by a random siRNA primarily aiming each gene.

To simulate the general principles of tRNAi, we ignore some of the subtle variations reported in the literature since they do not yield significant differences. Particularly, we need to point the following specifics in our computation.

1. The length of the elongated region is restricted to 500 nt, which is more consistent with the model proposed by Sijen et al. [5], although Lipardi and colleagues have proposed that the full length of the target mRNA be used [6].

2. We use the region upstream of the binding site to generate elongated dsRNA, although some experiments suggest both upstream and downstream produce secondary siRNAs [11, 12].

3. If an siRNA matches multiple sites on a target, all upstream regions are used for elongated dsRNA for the next stage. This is consistent with the extension (ligation) model of Lipardi and colleagues [6].

4. We do not allow self-cycles: once a gene has been included as a target in S^{t-1} covering stage $t - 1$, it will not be included in S^t .

5. Though tRNAi may contain infinitely many stages, we only consider up to 9 stages. This consideration is based on the threshold model [5, 13], which suggests that an abundance of target mRNA is required to maintain tRNAi. Since siRNA amplification leads to acceleration in RNA degradation, which reduces the mRNA level dramatically, the possibility of very long chain is unlikely.

6. Although every siRNA in the elongated region of 500 nt is possibly functional, and is scored according to the model of Sijen [5] (Figure 1(b)), we used a cutoff threshold of $f_s=90\%$. We make this simplification since we found that lowering f_s to increase the length of effective region did not have a significant effect on the graphs. However, adjusting f_s provides flexibility in our simulation.

2.3 Algorithms and Graph Statistics

We use string kernels and binary search trees for efficient searches [4]. For a random graph, the relative largest component size ρ is the number of nodes in the largest connected cluster divided by the total number of nodes in the graph. When ρ is greater than a certain level, the connected genes form a giant component [14]. To ensure the largest component is indeed a giant component, we use $\rho \geq 90\%$ as the criterion. When a graph increases its density to eventually contain a giant component as m gets larger, it is said to go through a phase transition. To quantify the density of a graph, we also measure the average distance between any two reachable vertices in the graph. We use breadth first-search algorithms to calculate the component sizes.

3 Results

Since evidence of tRNAi has been observed in *S. pombe* [11], we use its 5401 cDNA sequences (obtained from Sanger Institute <ftp://ftp.sanger.ac.uk>) to test our network model. Due to large amounts of genomic scale searches and graphical computations, data of this small organism is tractable. Results are shown below.

3.1 Degree Distributions

We evaluated degree distributions in the gene knockdown networks of *S. pombe* using 21 nt siRNAs, simulating tRNAi for three stages, $t = 3$, and allowing for two mismatch lengths, $m = 4$ and 6. Figure 2 (a) shows the log-log plot of the cumulative frequencies $P_{cum}(k) = \sum_{d \geq k} P(d)$, where $P(d)$ is the frequency of vertices having d edges. As indicated, the degrees roughly follow power distributions. If expressed as $P(k) \sim k^{-\gamma}$, then $\gamma = 3.28$ and 3.59 for the two curves (statistically significant, p-value $< 10^{-3}$). Since a power law distribution is not modal, the connectivity of a knockdown network is characterized by a continuum hierarchy with most of the genes connecting to a small number of genes, and a small number of genes connecting to a large number of genes. A power law distribution also suggests that the knockdown network is scale free [14].

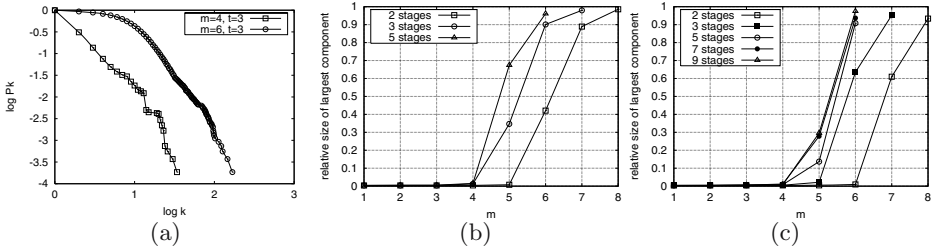


Fig. 2. Knockdown network properties. (a) Log-log plots of degree distributions. X-axis is the degree in logarithmic scale; Y-axis is the cumulative frequency $P_{cum}(k) = \sum_{d \geq k} P(d)$ in logarithmic scale. (b) Phase transitions: m , on X-axis, is the number of mismatches in siRNA-mRNA bindings. Y-axis is ρ , the largest component size divided by the number of genes. (c) Convergence of phase transitions up to $t = 9$ stages.

3.2 Phase Transitions

We computed the relative size ρ of the largest component of the knockdown networks. Figure 2 (b) shows ρ for transitive RNAi covering two, three, and five stages in the chain reaction, when increasing flexibility was allowed by siRNA-target bindings ($m=1-8$) using 21-mers as siRNA. As Figure 2 (b) suggests, when siRNA-target binding was stringent ($m \leq 4$ nt), the graphs were sparse and the largest components were small ($\rho < 1\%$). The average distance between any two vertices in the graph was about 12 hops. But the graphs experienced very sharp phase transitions to have giant components when increasing flexibility was allowed ($m \geq 5$). If we consider 5 transitive stages, a 6 nt mismatch yielded a giant component containing 95% of the nodes in the network. In such a dense network, the average distance was about 3 hops. Figure 2 (b) also indicates that more stages increased the chance of phase transitions when m was small.

In the knockdown network, genes connected to a common gene can be knocked down by a single primary siRNA. And the giant component can be knocked down with only a few primary siRNAs, since the average distance is short. This phase transition will turn RNAi from defensive into self-destructive and suggests a loss of stability in an organism. Therefore, mismatches of 6 nt or longer would cause instability in an organism, and thus are not biologically plausible.

Though the degree distributions, $P(k)$, both followed a power law before and after the phase transition, they had slightly different slopes. $\gamma = 3.28$ for the graph corresponding to $m = 4$ before the phase transition, and $\gamma = 3.59$ for $m = 6$ after the phase transition. The network having a giant component also had a larger maximum degree and average degree.

3.3 Convergence of Phase Transitions

Previously, we examined phase transitions of tRNAi up to 5 stages. To further investigate the effect of stages, we simulated 9 stages and display the results in Figure 2 (c). As indicated in the figure, phase transition patterns of 7 and 9 stages did not differ much from that of 5 stages and the patterns

of stages 7 and 9 were almost indistinguishable. The curves tend to converge with stages. Even though we did not simulate more stages than 10 due to computational intensity of the simulation, we expect a convergence of phase transition w.r.t stages for even more stages. Our results were solely derived from sequence data in the transcriptome and did not consider mRNA concentration. Nonetheless, they are consistent with the threshold model of tRNAi [5, 13].

4 Discussions and Conclusions

We modelled gene knockdown interactions as scale free graphs, which demonstrated that increasing the flexibility of siRNA-target binding increased the density of the graph. Statistics indicated that giant components existed if mismatches of more than 6 nt were allowed. This phase transition would turn RNAi from defensive to self-destructive. Therefore, the critical mismatch allowed in RNAi, especially in tRNAi, is unlikely to be higher than 6 nt.

The power-law-like degree distribution of the knockdown network in *S. pombe* implies that the knockdown network is scale free, suggesting that larger genomes exhibit similar patterns of knockdown interactions. Therefore, our results on *S. pombe* may be generalizable to other organisms. In particular, excessive mismatches in tRNAi will cause instability and there must be an upper limit to the allowable flexibility in all organisms. Since *S. pombe* has a small transcriptome, we are not sure whether the limits in other species are the same, or shorter due to higher chance of knockdown interactions.

Our results about the convergence of phase transition were solely derived from transcript sequence and did not consider mRNA concentration. Nonetheless, they are consistent with the threshold model of tRNAi [5, 13]. The threshold model suggests that an abundance of target mRNA is required to maintain tRNAi. Since siRNA amplification leads to acceleration in RNA degradation, which in turn reduces the mRNA level dramatically, the existence of very long chains are unlikely. Our results implies that sequence usage and concentration are consistent and somewhat related.

When we simulated tRNAi, we made conservative assumptions about the length of elongated dsRNA, only considered upstream of the target site, and limited the number of stages. If the assumptions are violated in nature, the graphs would be denser and the permissive flexibility even more stringent. Therefore our estimation is an optimistic upper bound. The similarity model used to characterize siRNA-target binding only considered contiguous mismatches. If we relax the contiguity restriction, the graph might get denser and the critical mismatch might decrease slightly. If we consider variable silencing potencies of different siRNAs, some siRNAs may be non-functional and the graph would be somewhat sparser and the critical mismatch might be a little longer. But if siRNAs are uniformly picked, then the knockdown networks are still scale-free. If both upstream and downstream sequences were used for elongated dsRNA, then the graph would get denser and the critical mismatch might decrease slightly.

Additionally, dsRNA may induce DNA methylation, which degrades more targets at transcription level and causes smaller critical mismatches [12].

Although we only studied tRNAi, high degree of evolutionary conservations of the RNAi mechanism implies that similar behaviors may exist in organisms not having tRNAi, such as *H. sapiens*. Furthermore, it is possible for larger mammalian genomes to have smaller critical mismatches. Therefore, our estimations are more consistent with the results in [1, 2]. However, mismatches longer than 6 nt, such as 9 nt in [3], are generally infrequent, and would probably cause extensive undesirable knockdowns and even instabilities.

Acknowledgement

This work is supported by NIH grant P20RR18754 from the Institutional Development Award Program of the National Center for Research Resource. We thank Coenraad M. Adema for helpful suggestions.

References

1. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T.: Functional anatomy of siRNA for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *The EMBO Journal* **20** (2001) 6877–6888
2. Saxena, S., Jonsson, Z.O., Dutta, A.: Small RNAs with imperfect match to endogenous mRNA repress translation. *Journal of Biological Chemistry* **278** (2003) 44312–44319
3. Jackson, A., Bartz, S., Schelter, J., Kobayashi, S., Burchard, J., Mao, M., Li, B., Cavet, G., Linsley, P.S.: Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnology* **21** (2003) 635–637
4. Qiu, S., Adema, C.M., Lane, T.: A computational study of off-target effects of RNA interference. *Nucleic Acids Research* **33** (2005) 1834–1847
5. Sijen, T., Fleenor, J., Simmer, F., Thijssen, K., Parrish, S., Timmons, L., Plasterk, R., Fire, A.: On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107** (2001) 465–476
6. Lipardi, C., Wei, Q., Paterson, B.: RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNA that are degraded to generate new siRNA. *Cell* **107** (2001) 297–307
7. Dalmay, T., Hamilton, A., Rudd, S., Angell, S., Baulcombe, D.: An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101** (2000) 543–553
8. Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., et al.: Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Research* **32** (2004) 936–948
9. Dillin, A.: The specifics of small interfering RNA specificity. *Proc. Natl. Acad. Sci. USA* **100** (2003) 6289–6291
10. Qiu, S., Lane, T.: String kernels of imperfect matches for off-target detection in RNA interference. In Sunderam, V., et al., eds.: *Proc. 1st IWBRA and 5th ICCS. LNCS vol. 3515*, Springer-Verlag (2005) 894–902
11. Allshire, R.: RNAi and heterochromatin: a hushed-up affair. *Science* **297** (2002) 1818–1819

12. Vaistij, F., Jones, L., Baulcombe, D.: Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *The Plant Cell* **14** (2002) 857–867
13. Metzloff, M., O'Dell, M., Cluster, P., Flavell, R.: RNA-mediated RNA degradation and chalcone synthase A silencing in *Petunia*. *Cell* **88** (1997) 845–854
14. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* **286** (1999) 509–512