Towards conditional RNAi: shape and sequence transduction with small conditional RNAs

Thesis by

Lisa Marie Hochrein

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2013

(Defended May 31, 2013)

Acknowledgments

I want to begin with thanking and acknowledging my husband Cambrian Liu. I would have never made it to this point without all his love, support, and encouragement. Thanks, Cambrian, for always believing that I would finish my PhD and convincing me that this day would eventually arrive—honestly, I'm not sure I believe it yet. He is my best friend, supporter, and confidant (plus an amazing cook). I'm grateful that Caltech brought us together.

My time at Caltech has not matched my initial expectations when I began in 2004. There have been many detours and unexpected events. Regardless, during my time at Caltech, I have grown tremendously as a scientist and an individual. Over the years, there have been numerous people willing to support my goals, provide guidance, and give me space to explore.

My advisor, Niles Pierce, has been a tremendous source of support. As a pioneer in the field of nucleic acid engineering, he has very innovative ideas with the goal of building useful tools and therapeutics. To make these goals a reality, he provided me with both the resources and freedom to explore these ideas. His persistence and high standards have taught me to be a better, and more independent, scientist. While we discovered that engineering nucleic acids *in vivo* is extremely challenging, I have no doubt that the Pierce Lab will continue to make large contributions to the field, and perhaps, someday build a truly smart therapeutic. His dedication to careful and thorough research should serve as a model for the entire research community.

I have had the privilege of working alongside many talented individuals over my years at Caltech. Being a member of a lab that draws members from diverse backgrounds, I have been exposed to many different ways of thinking and approaching scientific questions. While this occasionally causes conflict and some confusion at group meetings, my thinking and approach has been challenged and improved.

Ma'ayan Schwarzkopf has been a great co-author and friend. Together, we developed five conditional RNAi mechanisms and traded ideas, approaches, and insights that helped push our projects forward. We also shared the frustration of slowly realizing that conditional RNAi *in vivo*

was a much more challenging project than we originally envisioned. She has always been willing to lend a helping hand, share her knowledge of biology, and offer advice—an overall great experimental partner.

Jessie Ge, a terrific undergraduate researcher, was another great team member. She worked tirelessly to develop the cell lysate system. None of the results presented in Chapter 3 would have been possible without her hard work. Ma'ayan and I were extremely lucky to have her work with us for a year and half. We were continually impressed with her ability to quickly learn new techniques and ideas and to work independently.

All of the results presented in this thesis were possible because of the NUPACK software suite, which was built through the hard work of the NUPACK team. Without the design algorithms and analysis tools that comprise NUPACK, designing conditional RNAi mechanisms would have been a monumental task. Joe Zadeh, Brian Wolfe, Robert Dirks, Justin Bois, and Conrad Steenberg, along with others, have built a tool for scientists and engineers that predicts nucleic acid secondary structure with confidence. In addition, Brian Wolfe provided his expertise to analyze our next-generation sequencing results and produced the beautiful heat maps shown in Appendix A.

Other lab members, including Jeff Vieregg, Victor Beck, Harry Choi, Jonathan Sternberg, and Naeem Husain, have been helpful colleagues and have always been willing to share their expertise. Melinda Kirk and Colby Calvert have kept the lab running smoothly and spoiled me by ensuring my orders always arrive the next day. I also need to acknowledge my colleagues Jonathan Sternberg, Naeem Husain, Tobias Heinen, and Colby Calvert for working alongside me through an experience which no scientist ever wishes to know firsthand. I sincerely hope that each of us can move beyond the wasted years and establish successful scientific careers.

John Rossi has been a thoughtful committee member who provided valuable insight into the unexpected silencing behavior shown in Chapter 4 and provided generous financial and technical resources for our preliminary deep sequencing experiments. My other committee members, David Tirrell and Mark Davis, have offered good advice on exploring the various open-ended questions presented in Chapter 4. I also appreciate John Burnett's willingness to serve as a committee member at the last minute. Additionally, I want to thank the funding agencies that made this research possible: the NSF, which supported me with a graduate research fellowship, and the NIH.

During my first two years at Caltech, I was a member of the Arnold lab, where I became well versed in directed evolution for protein engineering. This background in directed evolution complements my current work in rational design. Frances is an admirable advisor who set up her

lab wisely and trained her students well. I learned many invaluable lessons as a member of the Arnold lab. I am especially thankful for the training provided by Michelle Meyer.

Of course, I would have never made it to this point without the help and support of numerous teachers and mentors along the way. These are people who encouraged me to believe in myself and gave me opportunities to explore exciting ideas beyond the classroom. This includes many teachers, especially my junior high school math and science teacher Mrs. McMorrow, who was wise enough to understand that my reading novels in class was a result of boredom and made the extra effort to provide challenging material for me. I am grateful to my high school teachers who assured me that the world was much bigger than my small town made it feel and encouraged me to attend U.C. Berkeley.

My time at Berkeley was exciting, full of new friends, adventures, and ideas. I especially enjoyed being an undergraduate researcher in the Schaffer Lab, working with Ana O'Neill. I loved being part of a lab and, for the first time, realized that biology was an interesting puzzle and not a long list of words to memorize. This experience exposed me to research, graduate school, and the challenge of working on projects that bridge various disciplines.

One cannot survive the challenge of a school like Caltech without friends, and I am thankful for my many friends, both at Caltech and elsewhere. The chemical engineering students who started with me in 2004 were a great source of support. Without them, qualifying and candidacy exams would have been much more daunting and overwhelming. I especially need to thank the four other women in my class for their friendship—we shared the frustrations that only other graduate students can understand, and we maintained a weekly lunch date for many years. Many of my best friends at Caltech were the housemates I lived with at 860 S. Los Robles Ave. They were always a great source of entertainment and company, and I miss my built-in friends.

And finally, I must thank my family. My family has been a great source of support and love. My parents, Julie and Peter Hochrein, instilled in me a great love for exploring, either through traveling, the outdoors, or reading. They helped me with my first scientific endeavors and taught me the scientific method. Through my dad, I learned the art of optimizing and the joy of numbers. My mom fed my love of learning, read to me many, many books, and has been a reliable source of hugs. My sisters, Michelle and Heather Hochrein, have always been a big part of my life. We had fun playing as we grew up, and we still enjoy having adventures together. My entire family has supported my work at Caltech, but I am most grateful that they have always trusted me to follow my own dreams and encouraged me to pursue whatever I found interesting and meaningful.

Abstract

RNA interference (RNAi) is a powerful biological pathway allowing for sequence-specific knockdown of any gene of interest. While RNAi is a proven tool for probing gene function in biological circuits, it is limited by being constitutively ON and executes the logical operation: *silence gene Y.* To provide greater control over post-transcriptional gene silencing, we propose engineering a biological logic gate to implement "conditional RNAi." Such a logic gate would silence gene Y only upon the expression of gene X, a completely unrelated gene, executing the logic: *if gene X is transcribed, silence independent gene Y.* Silencing of gene Y could be confined to a specific time and/or tissue by appropriately selecting gene X.

To implement the logic of conditional RNAi, we present the design and experimental validation of three nucleic acid self-assembly mechanisms which detect a sub-sequence of mRNA X and produce a Dicer substrate specific to gene Y. We introduce small conditional RNAs (scRNAs) to execute the signal transduction under isothermal conditions. scRNAs are small RNAs which change conformation, leading to both shape and sequence signal transduction, in response to hybridization to an input nucleic acid target. While all three conditional RNAi mechanisms execute the same logical operation, they explore various design alternatives for nucleic acid self-assembly pathways, including the use of duplex and monomer scRNAs, stable versus metastable reactants, multiple methods of nucleation, and 3-way and 4-way branch migration.

We demonstrate the isothermal execution of the conditional RNAi mechanisms in a test tube with recombinant Dicer. These mechanisms execute the logic: *if mRNA X is detected, produce a Dicer substrate targeting independent mRNA Y.* Only the final Dicer substrate, not the scRNA reactants or intermediates, is efficiently processed by Dicer. Additional work in human whole-cell extracts and a model tissue-culture system delves into both the promise and challenge of implementing conditional RNAi *in vivo.*

Contents

List of Figures

List of Tables

Common abbreviations

Common abbreviations	
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
ssRNA	single-stranded RNA
dsRNA	double-stranded RNA
nt	nucleotide
bp	base pair
RNAi	RNA interference
siRNA	small (or short) interfering RNA
DsiRNA	Dicer-substrate RNAs
AS	antisense strand
GS	guide strand
SS	sense strand
PS	passenger strand
scRNA	small conditional RNA
2′-O-Me	2′-O-Methyl RNA
RT	reverse transcription
PCR	polymerase chain reaction
Nomenclature	
Drosha	endoribonuclease, cleaves pri-miRNAs into pre-miRNAs in the nucleus
Dicer	endoribonuclease, cleaves dsRNA into siRNAs and pre-miRNAs into miRNAs
RISC	RNA-induced silencing complex
CRISPR	clustered regularly interspaced short palindromic repeats
deep sequencing	next-generation sequencing
ligation adapter	ligation linker

Chapter 1

Introduction

In this modern information age, the unprecedented amount of computational power at our fingertips has changed the way we live. While this computational power, built from basic logic gates, allows for hours of entertainment playing "Angry Birds" on our iPhones, life-threatening diseases are still treated with relatively unsophisticated therapeutics—sadly, our phones are smarter than our pharmaceuticals. Most drugs lack the ability to perform even basic logic. This is especially lamentable since biological systems function via complex circuits made up of logical operations. The goal of the body of work presented here is to create a framework for engineering smart molecules capable of performing calculations inside a living organism by sensing and responding to the external and internal environment. Such therapeutics would detect a biological input signal, execute logical computations, and respond according to the computational result.

A smart therapeutic could be built with various materials to interact with many biological pathways. The one we envision creates an artificial link between the expression of two unrelated genes. We chose to build our smart therapeutic from nucleic acids, specifically RNA and modified RNA, due to their inherent programmability and biocompatibility. RNA is a versatile molecule responsible for many processes, such as protein translation, RNA processing, and gene regulation, within a cell. We chose to interface with biology through the RNA interference (RNAi) pathway due its ability to selectively knock down a specific gene based on its sequence.[1–3] Using nucleic acids to interface with the RNAi pathway, we seek to engineer an artificial link within a biological circuit that modulates the expression of one gene based on the expression of a completely unrelated gene. Thus, our smart therapeutic would conditionally control gene expression through RNAi, performing "conditional RNAi."

Conditional RNAi executes the logic:

if gene X is transcribed, silence independent gene Y.

By contrast, conventional RNAi is constitutively active. When an siRNA targeted to gene Y is present, it executes the logic:

silence gene Y.

This highlights the lack of controllability—it is difficult to confine the knockdown to a specific locus and time—of the RNAi pathway through conventional siRNAs. In contrast, conditional RNAi would allow spatiotemporal control by appropriately choosing gene X to restrict the knockdown of gene Y in a tissue- and/or time-specific manner.

Our conditional RNAi logic gates are built with small conditional RNAs (scRNAs) which implement isothermal signal transduction through triggered nucleic acid hybridization cascades. scRNAs are RNAs or 2′-O-Methyl (2′-O-Me) modified RNAs[4] that change conformation in response to the detection of an input target, leading to shape and sequence transduction to produce a final product capable of interfacing with the RNAi pathway.

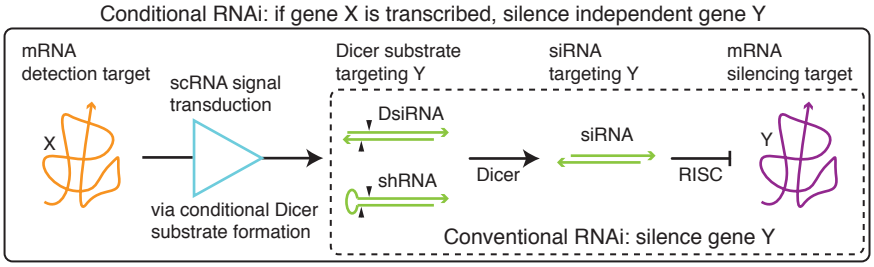


Figure 1.1: Conditional RNAi turns ON in the presence of the detection target X. Otherwise, it remains OFF. The ON state produces a Dicer substrate that feeds into the conventional RNAi pathway (inset box). For both conventional and conditional RNAi, Dicer cleaves a DsiRNA or shRNA to produce an siRNA, the effector molecule of RNAi. Gene Y is silenced through mRNA degradation mediated by RISC.

1.1 Nucleic acid engineering

Nucleic acid engineering is a growing field that utilizes specific base pairing rules to accurately predict and design nucleic acid secondary structure. Through rational design, nucleic acids were designed to form various nanostructures, including smiley faces,[5] nanocages,[6,7] and tubes.[8] Dynamic behavior can be programmed through the use of strand displacement via toehold-mediated

branch migration.[9–12] Various logic gates (AND and OR) have been programmed in DNA and allow for basic DNA computing.[13–15] A polymerization reaction that initiates upon detection of a input nucleic acid has been used to detect and amplify endogenous mRNA *in situ*.[16,17]

Rational design of nucleic acid structures is possible because of nucleic acid secondary structure prediction and design algorithms. These algorithms utilize experimentally derived energy parameters to calculate thermodynamic properties of nucleic acid strands in dilution solutions. Painstaking experiments have generated energy parameters for both DNA and RNA.[18–20] These experimental parameters allowed our lab to developed NUPACK, a software suite for the analysis and design of nucleic acid systems.[21] NUPACK analysis is able to calculate the thermodynamic properties of a dilute solution of interacting nucleic acid strands.[22–24] Sequences can be designed in NUPACK through an algorithm that minimizes the ensemble defect, employing both positive and negative design, to optimize sequences according to specified structural goals and sequence constraints.[25] Upon designing the sequences *in silico*, the short nucleic acid sequences (up to 200-nt) are chemically synthesized. The ability to easily and precisely specify the molecular composition of nucleic acids makes them a suitable engineering material.

The inherent programmability of nucleic acids is a key attribute that allows us to design conditional RNAi mechanisms. The choice of the input and output mRNA sequences significantly constrains the design space. Once these targets are specified, the NUPACK design algorithm can predict sub-sequences that will function together to execute the conditional RNAi mechanism. In our experience, the NUPACK design algorithm is highly successful, and often the first set of sequences tested functions well in test-tube studies. With the NUPACK design tools, reprogramming our mechanisms to reverse the input and output targets, or more interestingly, to detect any gene of choice and silence any other unrelated gene, is relatively straightforward task. Through the programmability and biocompatibility of nucleic acids and the NUPACK design tools, engineering nucleic acids presents the opportunity to interface with endogenous biological circuits.

1.2 RNAi pathway

Nucleic acids are an ideal material with which to construct biologic logic gates because they are natural components of the cellular environment, and there are many biological pathways that detect DNA and RNA. RNA interference (RNAi) is one such pathway. It is a powerful mechanism of sequence specific, post-transcriptional gene knockdown in eukaryotes.[1,2] RNAi is activated by short

RNA duplexes termed small interfering RNAs (siRNA), which direct enzymatic cleavage of mRNA containing the complementary sequence through the RNA-induced silencing complex (RISC). This sequence-specific cleavage leads to mRNA degradation and prevents protein translation.[26,27] The power of RNAi exists in its programmability; siRNAs can be designed to target any gene through sequence alone. This programmability has led RNAi to become an extremely popular tool to probe gene function. RNAi can even discriminate between sequences that differ by a single nucleotide if the mutation is properly positioned within the siRNA.[28,29]

The RNAi pathway machinery is comprised of RISC and the upstream processing enzymes Dicer and Drosha.[3] Gene knockdown with RISC follows four steps: (1) siRNA loading into RISC, (2) removal of one of the siRNA strands, (3) target mRNA recruitment, and (4) target mRNA cleavage.[30] siRNAs are the effector molecules of RNAi and are short dsRNAs with 2-nt 3′-overhangs.[31,32] The guide strand of the siRNA is incorporated into RISC, while the passenger strand is removed either through degradation or unwinding.[26,33–36] RISC, through the "slicing" function of Ago2, endonucleolytically cleaves mRNA (which is complementary to the guide strand via Watson-Crick base pairing).[27,32] Once activated, RISC can cleave multiple copies of the complementary mRNA, leading to robust knockdown of target gene expression.[37]

The upstream processing enzymes Dicer and Drosha are both RNases, which function to trim long dsRNA strands into siRNAs.[38–40] Drosha primarily processes pri-miRNAs, hairpins with long overhangs on either end, into pre-miRNAs in the nucleus. Dicer trims pre-miRNA and other dsRNA substrates into siRNAs. While some organisms have separate Dicer enzymes to cleave miRNAs and siRNAs, in humans a single Dicer enzyme generates both miRNAs and siRNAs. An siRNA is a short RNA duplex of approximately 19–21-bp in length with a 2-nt overhang at both 3′ ends.[31,32,41,42] Dicer cleavage produces an siRNA with the characteristic 5′-phosphate(PO_4) and 3′-hydroxyl(OH) end groups. These end groups help differentiate siRNAs from degradation products.

A synthetic Dicer substrate can be used to generate an siRNA. Dicer substrates, short RNA duplexes and short hairpin RNAs (shRNAs), generally have a single 2-nt 3′-overhang to direct Dicer cleavage. Dicer functions as a molecular ruler and measures ∼21-nt from the 2-nt 3′-overhang to cleave a standard siRNA.[43,44] Dicer-substrate RNAs (DsiRNAs) were introduced by Kim et al. as a strategy to improve RNAi gene knockdown. The improved silencing was hypothesized to result from improved loading of RISC through direct transfer of the siRNA from Dicer to RISC.[45,46] Current reports, however, show that DsiRNAs are not necessarily more effective than siRNAs.[47] A standard DsiRNA is a 25-nt dsRNA duplex with asymmetric ends: a 2-nt 3′-overhang and a

blunt end. The asymmetry helps direct Dicer cleavage.[43,44] An alternative Dicer substrate, an shRNA, was designed as a single molecule to allow for cellular expression.[48] shRNAs mimic the hairpin structure of pre-miRNAs, with the exception that shRNAs are designed to have perfect base pairing in the stem (pre-miRNAs generally have some mismatches). Broadly, an shRNA has a 19–29 base pair stem, 4–23 nucleotide loop, and a 2-nt 3′-overhang, but many functional variations have been demonstrated.[49,50] While shRNAs are primarily expressed, shRNAs also function when delivered.[45,51]

An siRNA can be designed to target any 21-nt sequence, but not all siRNAs are effective. An effective siRNA can achieve greater than 90% gene knockdown, but many siRNAs are less effective and some do not function at all.[52] While many researchers have suggested rules to design potent siRNAs based on sequence and structural properties,[53,54] many of the rules conflict or are dispensable.[52] Generally, a pool of siRNAs targeting the same gene is tested and the best siRNAs are determined empirically. Interestingly, shifting an siRNA a few bases along the mRNA can completely alter the strength of gene silencing.[55] The effectiveness of gene silencing is also influenced by the accessibility of the mRNA region, and less structured regions yield more gene knockdown.[56] Unfortunately, it is difficult to predict mRNA secondary structure because of the length and pseudoknotted structure of mRNAs.[57]

It has been hypothesized that RNAi evolved to silence viral RNAs and transposable elements, and to regulate gene expression.[58] siRNAs and miRNAs both post-transcriptionally silence genes. siRNAs induce gene knockdown through degradation of the complementary mRNA. miRNAs bind to the 3′-UTR of mRNA, which prevents mRNA translation.[59] Unlike siRNAs, miRNAs do not require perfect Watson-Crick base paring with the mRNA. Thus, a single miRNA is able to silence many different mRNA targets. Only recently has the importance of miRNA gene regulation become apparent. While the details are still being unraveled, thousands of miRNAs have been identified, and it is believed that over half of human transcripts are regulated by miRNAs.[60]

While RNAi has become a standard research tool, its power has proven more difficult to harness for therapeutic purposes due to challenges with delivery, stability, and off-target effects.[61] RNA is unstable in the cellular environment and the blood stream. Chemical modifications increase stability and improve the half-life of siRNAs, and work has been done to determine chemical modifications tolerated by the RNAi pathway.[62–64] Delivery of siRNAs is still a major hurdle, but advances have been made in delivery via the use of biodegradable nanoparticles, lipids, bacteria and attenuated viruses.[61]

1.3 Alternative implementations of conditional RNAi

The RNAi pathway is constitutively on: if an siRNA is present and capable of silencing, the gene will be silenced. Some level of spatiotemporal control can be achieved through targeted siRNA delivery[65] or controlled shRNA expression with tissue specific and/or switchable promoters.[48,66–69] Multiple groups have implemented a form of conditional RNAi which detects a small molecule and silences a gene through production of a Dicer or Drosha substrate.[70–74] These nucleic acid logic gates consist of an aptamer bound to a Dicer or Drosha substrate. An aptamer, an oligonucleotide that specifically binds a target molecule, detects the small molecule input. Aptamer binding leads to a conformational change of the oligonucleotide, which is the source of the switching behavior. The RNAi pathway can be switched either ON by producing a Dicer substrate or OFF by sequestering a Dicer substrate. The silencing observed is a function of the small molecule input concentration and can be tuned by changing the gate design. In implementations to date, these conditional RNAi mechanisms have been expressed, and the small molecule has been added exogenously.

Aptamer-based conditional RNAi gates are difficult to reprogram because a new aptamer needs to be incorporated. Currently, aptamers cannot be rationally designed and are generated through directed evolution and selection.[75,76] Changing the input is time consuming and limits the input signals to those for which a robust aptamer already exists or can be selected. Another drawback is that these mechanisms rely on the expression of the gate within the cell. While this is convenient in tissue culture, it becomes equivalent to gene therapy when applied on a organism level. Because of the pitfalls of gene therapy,[77] it is likely that a therapeutic which can be transiently delivered would be more promising. There is no fundamental reason why these gates could not be delivered; however, delivery is a major challenge and the gates would likely require significant reengineering to allow for successful delivery. Also, it is possible that the gates only function due to the high concentration levels achieved through continuous expression.

Our vision of conditional RNAi implements logic similar to the gates discussed above, but our goal is to detect an endogenously expressed gene rather than an exogenous small molecule. These gates will exploit the natural strength of nucleic acid base pairing to detect a sub-sequence of an endogenous mRNA. The use of nucleic acids as both the input and output molecules allows the logic gates to be completely programmable and designed rationally *in silico* by tools such as NUPACK.

The specific design goals of our conditional RNAi gates are detailed in Chapter 2. A brief overview is given in Table 1.1. Several groups have achieved subsets of these goals. Masu et

Table 1.1: Conditional RNAi mechanism design goals.

Conditional RNAi design goals:
1) Mechanism interfaces with the RNAi pathway by producing a Dicer substrate.
2) Nucleic acid input target is a sub-sequence of a full-length mRNA.
3) Input and output targets are 100% sequence independent.
4) Strong ON state: robust input detection, signal transduction, and Dicer substrate formation.
5) Strong OFF state: no silencing in the absence of the input target.
6) Dicer cleavage of only the final product; reactants and intermediates are not Dicer substrates.

al. demonstrated a simple mechanism that detects a short RNA target.[78] In tissue culture, they showed that the initial substrate was OFF and that the final product induced gene knockdown, but they failed to show signal transduction. Xie et al. demonstrated a gate capable of detecting mRNA fragments in cell-free *Drosophila* embryo lysate.[79] However, the input target is constrained by the output target (the input target is a slight variant of output target antisense). Thus, while they demonstrated logical computation in lysate, the choice of input target sequence dictated the output target sequence. Kumar et al. expressed a gate in tissue culture which produced a Drosha substrate, leading to gene knockdown, in response to the transfection of a short chemically modified target.[80] Detection of an endogenous mRNA was not demonstrated. None of the three previous examples achieved all six of our conditional RNAi design goals.

1.4 Future applications of conditional RNAi

In the future, a robust implementation of conditional RNAi *in vivo* has the potential for profound value as both a therapeutic and a research tool. As a therapeutic, conditional RNAi could determine if a cell is diseased by detecting a signature mRNA target, and then either specifically kill the cell or up/down-regulate an unrelated gene of choice. The most obvious target diseases are those where the expression of a somatic mutation differentiates the diseased cells from the healthy cells, as is the case for cancer.[81] For example, a conditional RNAi therapeutic could detect an oncogenic mRNA marker and silence an essential housekeeping gene, thereby killing the cancer cell. Conditional RNAi could act as a "smart" therapeutic that kills only cancerous cells, minimizing side-effects and improving patient outcomes.

Another ambitious therapeutic application of conditional RNAi is the treatment of autoimmune diseases through selective killing of the oligoclonal population of immune cells which mischaracterize the self as an intruder.[82,83] Selectively killing the immune cells responsible for creating the autoim-

mune response could treat the disease at the source and hopefully lead to remission. Treatment of either cancer or autoimmune diseases with conditional RNAi likely requires personalized medicine. The signature mutation(s) in each patient would need to be determined individually in order to create a custom drug. While this process is certainly burdensome, it is not unrealistic as next-generation sequencing becomes cheaper and part of standard clinical practice. Since conditional RNAi is reprogrammable, it can be personalized to make a custom drug for each individual.

As a research tool, conditional RNAi mechanisms could allow gene Y to be silenced in a specific tissue or at a specific developmental stage by appropriately selecting gene X. Alternatively, the spatiotemporal expression of any gene could be reported visually by specifying gene Y as a fluorescent protein or a regulator of a fluorescent protein. Thus, a network of genes can be investigated and modulated by conditional RNAi mechanisms. A tool that artificially links any two genes of interest would open a whole new avenue of exploring biological systems.

While the mechanism designs presented in this body of work focus on implementing the logic IF/THEN, scRNAs can be used to program other logic operations such as AND, OR, NOT and NOR. These various logic gates could then be combined to implement more complex logical computations.

In addition to endogenous mRNAs, scRNAs could be designed to detect small RNAs or potentially DNA. Detection of a single-nucleotide polymorphism requires careful balancing of conditional RNAi energetics. Mechanisms will need to be optimized to achieve single nucleotide discrimination, but work in our lab has indicated it is possible.[84] Additionally, a conditional RNAi mechanism could be altered to produce either an siRNA or Drosha substrate instead of a Dicer substrate.

While we have focused on the RNAi pathway, there are many other cellular pathways which interact with nucleic acids. The output of an scRNA logic gate could be designed to interact with a specific cellular pathway. For example, an scRNA logic gate could generate a final product with immunostimulatory effects via interaction with receptors such as RIG-I and TLRs.[85–87] Alternatively, an scRNA logic gate could produce an output that is simultaneously immunostimulatory and an RNAi substrate.[88] Other pathways for post-transcriptional gene regulation could be harnessed, such as the nucleic acid based adaptive immune system CRISPR found in bacteria and archaea. This system relies on small RNAs for sequence specific silencing of foreign nucleic acids.[89,90] As understanding of this system grows, it is becoming evident that CRISPR can be programmed in a similar fashion to RNAi.[91,92] The conditional RNAi mechanisms presented here are only the first steps towards programmable biological logic gates.

1.5 Thesis overview

This thesis covers the work I have done towards implementing conditional RNAi logic gates. Chapter 2 introduces three conditional RNAi mechanisms. Each mechanism design and initial implementation in a test tube, including experiments with recombinant Dicer, are shown in detail. Chapter 3 introduces our work towards developing a mammalian cell lysate system to test our conditional RNAi mechanisms. Chapter 4 discusses work done towards engineering functional conditional RNAi mechanisms in tissue culture. While we have yet to demonstrate conditional RNAi in tissue culture, our efforts have uncovered many details about the fate of scRNAs in a cellular environment.

Additional information is provided in the appendices. Appendix A describes work towards improving the protocol used to create small RNA libraries for next-generation sequencing in an attempt to reduce the bias that currently plagues small RNA libraries.[93–95] Appendix B describes the materials and methods used for experiments presented in this thesis. Appendix C includes supplementary figures.

1.6 Bibliography

[1] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669), 806–11 (1998).

[2] Hannon, G. J. RNA interference. *Nature* **418**(6894), 244–51 (2002).

[3] Kurreck, J. RNA interference: from basic research to therapeutic applications. *Angewandte Chemie* **48**(8), 1378–98 (2009).

[4] Uhlmann, E. and Peyman, A. Antisense oligonucleotides—a new therapeutic principle. *Chemical Reviews* **90**(4), 543–584 (1990).

[5] Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**(7082), 297–302 (2006).

[6] Chen, J. H. and Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**(6319), 631–3 (1991).

[7] Goodman, R. P., Schaap, I. A., Tardin, C. F., Erben, C. M., Berry, R. M., Schmidt, C. F., and Turberfield, A. J. Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**(5754), 1661–5 (2005).

[8] Rothemund, P. W., Ekani-Nkodo, A., Papadakis, N., Kumar, A., Fygenson, D. K., and Winfree, E. Design and characterization of programmable DNA nanotubes. *Journal of the American Chemical Society* **126**(50), 16344–52 (2004).

[9] Yurke, B., Turberfield, A. J., Mills, Jr., A. P., Simmel, F. C., and Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**(6796), 605–8 (2000).

[10] Yin, P., Choi, H. M., Calvert, C. R., and Pierce, N. A. Programming biomolecular self-assembly pathways. *Nature* **451**(7176), 318–22 (2008).

[11] Zhang, D. Y. and Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131**(47), 17303–14 (2009).

[12] Zhang, D. Y. and Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry* **3**(2), 103–13 (2011).

[13] Seelig, G., Soloveichik, D., Zhang, D. Y., and Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**(5805), 1585–8 (2006).

[14] Qian, L. and Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**(6034), 1196–201 (2011).

[15] Qian, L. and Winfree, E. A simple DNA gate motif for synthesizing large-scale circuits. *Journal of the Royal Society, Interface / the Royal Society* **8**(62), 1281–97 (2011).

[16] Dirks, R. M. and Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America* **101**(43), 15275–8 (2004).

[17] Choi, H. M., Chang, J. Y., Trinh le, A., Padilla, J. E., Fraser, S. E., and Pierce, N. A. Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nature Biotechnology* **28**(11), 1208–12 (2010).

[18] SantaLucia, J., J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95**(4), 1460–5 (1998).

[19] Serra, M. J. and Turner, D. H. Predicting thermodynamic properties of RNA. *Methods in Enzymology* **259**, 242–61 (1995).

[20] Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* **288**(5), 911–40 (1999).

[21] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32**(1), 170–3 (2011).

[22] Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., and Pierce, N. A. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review* **49**(1), 65–88 (2007).

[23] Dirks, R. M. and Pierce, N. A. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *Journal of Computational Chemistry* **24**(13), 1664–77 (2003).

[24] Dirks, R. M. and Pierce, N. A. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *Journal of Computational Chemistry* **25**(10), 1295–304 (2004).

[25] Zadeh, J. N., Wolfe, B. R., and Pierce, N. A. Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry* **32**(3), 439–52 (2011).

[26] Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell* **15**(2), 185–97 (2004).

[27] Pratt, A. J. and MacRae, I. J. The RNA-induced silencing complex: a versatile gene-silencing machine. *The Journal of Biological Chemistry* **284**(27), 17897–901 (2009).

[28] Schwarz, D. S., Ding, H., Kennington, L., Moore, J. T., Schelter, J., Burchard, J., Linsley, P. S., Aronin, N., Xu, Z., and Zamore, P. D. Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLOS Genetics* **2**(9), e140 (2006).

[29] Du, Q., Thonberg, H., Wang, J., Wahlestedt, C., and Liang, Z. A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Research* **33**(5), 1671–7 (2005).

[30] Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**(5), 563–74 (2002).

[31] Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**(1), 25–33 (2000).

[32] Elbashir, S. M., Lendeckel, W., and Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development* **15**(2), 188–200 (2001).

[33] Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**(5689), 1437–41 (2004).

[34] Rand, T. A., Petersen, S., Du, F., and Wang, X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**(4), 621–9 (2005).

[35] Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J., Hannon, G. J., and Joshua-Tor, L. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Structural & Molecular Biology* **12**(4), 340–9 (2005).

[36] Leuschner, P. J., Ameres, S. L., Kueng, S., and Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Reports* **7**(3), 314–20 (2006).

[37] Hutvagner, G. and Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**(5589), 2056–60 (2002).

[38] Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. Ribonuclease activity and RNA binding of recombinant human Dicer. *The EMBO Journal* **21**(21), 5864–74 (2002).

[39] Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**(6818), 363–6 (2001).

[40] Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V. N. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**(6956), 415–9 (2003).

[41] Hamilton, A. J. and Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**(5441), 950–2 (1999).

[42] Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**(6775), 293–6 (2000).

[43] Rose, S. D., Kim, D. H., Amarzguioui, M., Heidel, J. D., Collingwood, M. A., Davis, M. E., Rossi, J. J., and Behlke, M. A. Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Research* **33**(13), 4140–56 (2005).

[44] Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D., and Doudna, J. A. Structural basis for double-stranded RNA processing by Dicer. *Science* **311**(5758), 195–8 (2006).

[45] Kim, D. H., Behlke, M. A., Rose, S. D., Chang, M. S., Choi, S., and Rossi, J. J. Synthetic

dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nature Biotechnology* **23**(2), 222–6 (2005).

[46] Amarzguioui, M., Lundberg, P., Cantin, E., Hagstrom, J., Behlke, M. A., and Rossi, J. J. Rational design and in vitro and in vivo delivery of Dicer substrate siRNA. *Nature Protocols* **1**(2), 508–17 (2006).

[47] Foster, D. J., Barros, S., Duncan, R., Shaikh, S., Cantley, W., Dell, A., Bulgakova, E., O'Shea, J., Taneja, N., Kuchimanchi, S., Sherrill, C. B., Akinc, A., Hinkle, G., Seila White, A. C., Pang, B., Charisse, K., Meyers, R., Manoharan, M., and Elbashir, S. M. Comprehensive evaluation of canonical versus Dicer-substrate siRNA in vitro and in vivo. *RNA* **18**(3), 557–68 (2012).

[48] Gupta, S., Schoer, R. A., Egan, J. E., Hannon, G. J., and Mittal, V. Inducible, reversible, and stable RNA interference in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **101**(7), 1927–32 (2004).

[49] Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J., and Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes & Development* **16**(8), 948–58 (2002).

[50] Li, L., Lin, X., Khvorova, A., Fesik, S. W., and Shen, Y. Defining the optimal parameters for hairpin-based knockdown constructs. *RNA* **13**(10), 1765–74 (2007).

[51] Siolas, D., Lerner, C., Burchard, J., Ge, W., Linsley, P., Paddison, P., Hannon, G., and Cleary, M. Synthetic shRNAs as potent RNAi triggers. *Nature Biotechnology* **23**(2), 227–231 (2005).

[52] Pei, Y. and Tuschl, T. On the art of identifying effective and specific siRNAs. *Nature Methods* **3**(9), 670–6 (2006).

[53] Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., and Khvorova, A. Rational siRNA design for RNA interference. *Nature Biotechnology* **22**(3), 326–30 (2004).

[54] Yuan, B., Latek, R., Hossbach, M., Tuschl, T., and Lewitter, F. siRNA selection server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Research* **32**(Web Server issue), W130–4 (2004).

[55] Holen, T., Amarzguioui, M., Wiiger, M. T., Babaie, E., and Prydz, H. Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Research* **30**(8), 1757–66 (2002).

[56] Gredell, J. A., Berger, A. K., and Walton, S. P. Impact of target mRNA structure on siRNA silencing efficiency: A large-scale study. *Biotechnology and Bioengineering* **100**(4), 744–55 (2008).

[57] Zhang, Y., Zhang, J., and Wang, W. Atomistic analysis of pseudoknotted RNA unfolding. *Journal of the American Chemical Society* **133**(18), 6882–5 (2011).

[58] Cerutti, H. and Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics* **50**(2), 81–99 (2006).

[59] Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2), 215–33 (2009).

[60] Pasquinelli, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature Reviews Genetics* **13**(4), 271–82 (2012).

[61] Burnett, J. C., Rossi, J. J., and Tiemann, K. Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnology Journal* **6**(9), 1130–46 (2011).

[62] Amarzguioui, M., Holen, T., Babaie, E., and Prydz, H. Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Research* **31**(2), 589–95 (2003).

[63] Chiu, Y. L. and Rana, T. M. siRNA function in RNAi: a chemical modification analysis. *RNA* **9**(9), 1034–48 (2003).

[64] Deleavey, G. F., Watts, J. K., Alain, T., Robert, F., Kalota, A., Aishwarya, V., Pelletier, J., Gewirtz, A. M., Sonenberg, N., and Damha, M. J. Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing. *Nucleic Acids Research* **38**(13), 4547–57 (2010).

[65] Oliveira, S., Storm, G., and Schiffelers, R. M. Targeted delivery of siRNA. *Journal of Biomedicine & Biotechnology* **2006**(4), 63675 (2006).

[66] Tiscornia, G., Tergaonkar, V., Galimi, F., and Verma, I. M. CRE recombinase-inducible RNA interference mediated by lentiviral vectors. *Proceedings of the National Academy of Sciences of the United States of America* **101**(19), 7347–51 (2004).

[67] Wiznerowicz, M., Szulc, J., and Trono, D. Tuning silence: conditional systems for RNA interference. *Nature Methods* **3**(9), 682–8 (2006).

[68] Cullere, X., Lauterbach, M., Tsuboi, N., and Mayadas, T. N. Neutrophil-selective CD18 silencing using RNA interference in vivo. *Blood* **111**(7), 3591–8 (2008).

[69] Lee, S. K. and Kumar, P. Conditional RNAi: towards a silent gene therapy. *Advanced Drug Delivery Reviews* **61**(7-8), 650–64 (2009).

[70] An, C. I., Trinh, V. B., and Yokobayashi, Y. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA* **12**(5), 710–6 (2006).

[71] Beisel, C. L., Bayer, T. S., Hoff, K. G., and Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology* **4**, 224 (2008).

[72] Tuleuova, N., An, C. I., Ramanculov, E., Revzin, A., and Yokobayashi, Y. Modulating endogenous gene expression of mammalian cells via RNA-small molecule interaction. *Biochemical and Biophysical Research Communications* **376**(1), 169–73 (2008).

[73] Kumar, D., An, C. I., and Yokobayashi, Y. Conditional RNA interference mediated by allosteric ribozyme. *Journal of the American Chemical Society* **131**(39), 13906–7 (2009).

[74] Beisel, C. L., Chen, Y. Y., Culler, S. J., Hoff, K. G., and Smolke, C. D. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Research* **39**(7), 2981–94 (2011).

[75] Ellington, A. D. and Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**(6287), 818–22 (1990).

[76] Bouchard, P. R., Hutabarat, R. M., and Thompson, K. M. Discovery and development of therapeutic aptamers. *Annual Review of Pharmacology and Toxicology* **50**, 237–57 (2010).

[77] Verma, I. M. and Weitzman, M. D. Gene therapy: twenty-first century medicine. *Annual Review of Biochemistry* **74**, 711–38 (2005).

[78] Masu, H., Narita, A., Tokunaga, T., Ohashi, M., Aoyama, Y., and Sando, S. An activatable siRNA probe: trigger-RNA-dependent activation of RNAi function. *Angewandte Chemie* **48**(50), 9481–3 (2009).

[79] Xie, Z., Liu, S. J., Bleris, L., and Benenson, Y. Logic integration of mRNA signals by an RNAi-based molecular computer. *Nucleic Acids Research* **38**(8), 2692–701 (2010).

[80] Kumar, D., Kim, S. H., and Yokobayashi, Y. Combinatorially inducible RNA interference triggered by chemically modified oligonucleotides. *Journal of the American Chemical Society* **133**(8), 2783–8 (2011).

[81] Pleasance, E. D., Cheetham, R. K., Stephens, P. J., McBride, D. J., Humphray, S. J., Greenman, C. D., Varela, I., Lin, M. L., Ordonez, G. R., Bignell, G. R., Ye, K., Alipaz, J., Bauer, M. J., Beare, D., Butler, A., Carter, R. J., Chen, L., Cox, A. J., Edkins, S., Kokko-Gonzales, P. I., Gormley, N. A., Grocock, R. J., Haudenschild, C. D., Hims, M. M., James, T., Jia, M., Kingsbury, Z., Leroy, C., Marshall, J., Menzies, A., Mudie, L. J., Ning, Z., Royce, T., Schulz-Trieglaff, O. B., Spiridou, A., Stebbings, L. A., Szajkowski, L., Teague, J., Williamson, D., Chin, L., Ross, M. T., Campbell, P. J., Bentley, D. R., Futreal, P. A., and Stratton, M. R. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**(7278), 191–6 (2010).

[82] Skulina, C., Schmidt, S., Dornmair, K., Babbe, H., Roers, A., Rajewsky, K., Wekerle, H., Hohlfeld, R., and Goebels, N. Multiple sclerosis: brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. *Proceedings of the National Academy of Sciences of the United States of America* **101**(8), 2428–33 (2004).

[83] Hohlfeld, R. and Wekerle, H. Autoimmune concepts of multiple sclerosis as a basis for selective immunotherapy: from pipe dreams to (therapeutic) pipelines. *Proceedings of the National Academy of Sciences of the United States of America* **101 Suppl 2**, 14599–606 (2004).

[84] Sternberg, J. B.-Z. *Signal Transduction with Hybridization Chain Reactions.* PhD thesis, California Institute of Technology, (2013).

[85] Janeway, C. A., J. and Medzhitov, R. Innate immune recognition. *Annual Review of Immunology* **20**, 197–216 (2002).

[86] Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunology* **5**(7), 730–7 (2004).

[87] Takeda, K. and Akira, S. Toll-like receptors in innate immunity. *International Immunology* **17**(1), 1–14 (2005).

[88] Schlee, M., Hornung, V., and Hartmann, G. siRNA and isRNA: two edges of one sword. *Molecular Therapy: The Journal of the American Society of Gene Therapy* **14**(4), 463–70 (2006).

[89] Sorek, R., Kunin, V., and Hugenholtz, P. CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology* **6**(3), 181–6 (2008).

[90] Horvath, P. and Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**(5962), 167–70 (2010).

[91] Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, 3rd, C. V., Graveley, B. R., Terns, R. M., and Terns, M. P. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular Cell* **45**(3), 292–302 (2012).

[92] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**(6096), 816–21 (2012).

[93] Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J. T., Nusbaum, J. D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., and Tuschl, T. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**(9), 1697–712 (2011).

[94] Linsen, S. E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., Fritz, B., Wyman, S. K., de Bruijn, E., Voest, E. E., Kuersten, S., Tewari, M., and Cuppen, E. Limi-

tations and possibilities of small RNA digital gene expression profiling. *Nature Methods* **6**(7), 474–6 (2009).

[95] Jayaprakash, A. D., Jabado, O., Brown, B. D., and Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research* **39**(21), e141 (2011).

Chapter 2

Engineering diverse nucleic acid mechanisms for conditional Dicer substrate formation

2.1 Introduction

In this chapter, we introduce three mechanisms that implement the logic of conditional RNAi and describe the design process and test-tube validation of these mechanisms.* The mechanisms are constructed from small conditional RNAs (scRNAs), which are small RNA or 2′-O-Me-modified RNA structures that change conformation in response to an input signal. The designs were inspired by previous work in designing initiated, isothermal hybridization cascades built from DNA.[1,2] However, working with RNA, instead of DNA, and accommodating biological sequence constraints and enzymatic processing rules required new design approaches. The engineering of multiple designs allowed for exploration of the permissible design parameters. As a result, these mechanism designs advance the field of nucleic acid nanotechnology. In addition, they provide a foundation for constructing synthetic biological logic gates from nucleic acids.

Conditional RNAi mechanisms require that scRNAs transduce both sequence and Dicer cleavage susceptibility, which we refer to as shape. Sequence transduction is required the meet the design goal of sequence independence between input X and output Y. Shape switching is necessary to allow for good OFF and ON states. scRNA reactants must be resistant to Dicer cleavage to maintain the OFF state, and the final product must be a Dicer substrate to achieve an ON state. The shape

*In total, our lab has developed five conditional RNAi mechanisms, which are presented as a cohesive unit in the paper we submitted for publication under the title "Conditional Dicer substrate formation via shape and sequence transduction with small conditional RNAs." The three mechanisms covered in this thesis are the conditional RNAi mechanisms that I directly developed. The original design for Mechanism 1: Conditional DsiRNA formation using stable scRNAs was done by Niles Pierce and Peng Yin. I did all of the experimental work and mechanism optimization.

switch is achieved through either a change in secondary structure or the location of Dicer-resistant modifications.

2.1.1 Conditional RNAi design goals

Successful conditional RNAi mechanisms achieve the design goals specified in Table 1.1. The first design goal specifies that the final product is a substrate of the RNAi pathway and can induce selective knockdown of the output target mRNA Y. This final product can be an siRNA or a substrate for either of the upstream processing enzymes Dicer and Drosha. We focused on producing Dicer substrates rather than siRNAs directly because Dicer cleavage can remove extraneous sequences that are the byproduct of signal transduction, which allowed increased flexibility in designing the final product structure and sequence. The mechanism design space was greatly increased by allowing this flexibility.

Mechanisms 1 and 3 produce a dsRNA duplex final product, while Mechanism 2 produces an shRNA final product. Traditional Dicer substrates served as the templates for the final products, but alterations were required for Mechanisms 1 and 3. Based on literature reports that Dicer acts as a ruler which measures from the 2-nt overhang on the $3'$ end, we directed cleavage by ensuring that all the final products had one 2-nt $3'$-overhang, followed by a perfect duplex of $\geq$19 base pairs.[3,4] We assumed that altering the duplex end opposing the 2-nt $3'$-overhang should not interfere with Dicer cleavage, allowing non-canonical Dicer substrates to produce canonical siRNAs.

The second design goal in Table 1.1 specifies that the input signal is a nucleic acid sequence, specifically a sub-sequence of an mRNA. (By definition, interfacing with the RNAi pathway requires that the output target is also a nucleic acid sequence.) Detecting an endogenous sequence, in this case an mRNA, allows the logic gate to assess the state of a cell through detection of both sequence information and expression levels. The third design goal specifies that the input and output nucleic acid sequences (mRNA sub-sequences) are completely independent. This constraint increases the power of the logic gates because the input and output sequences can be reprogrammed independently and the selection of one does not limit the selection of the other. NUPACK design algorithms help select suitable mRNA sub-sequences for signal transduction.

A conditional RNAi mechanism must accurately report the state of the cell. This leads to the fourth and fifth design goals: a mechanism must have strong ON and OFF states. These requirements are essential to prevent both false negatives (failure to silence the output target gene in the presence of the input target) and false positives (silencing of the output target gene in the

absence of the input target). A mechanism with a good ON state maximizes the formation of the final Dicer substrate when the input sequence is present. Conversion is a measurement of the ON-pathway reaction that converts scRNA reactants to products in the presence of the input target. A mechanism with a good OFF state minimizes the formation of the final Dicer substrate when the input sequence is absent. The Dicer substrate formed without an input target is referred to as "leakage" and the leakage reaction is the pathway that generates leakage. While perfect ON and OFF states are ideal, they are difficult to achieve.

		Final Product Designation
input target	+ conditional RNAi mechanism	conversion (ON state)
no input target	+ conditional RNAi mechanism	leakage (OFF state)

Table 2.1: Defining the quality of the ON and OFF states for conditional RNAi mechanisms. Conversion is the (desired) production of the final Dicer substrate when the input target is present. Leakage is the (undesired) production of the final Dicer substrate when the input target is absent and is caused by leaky signal transduction. Our goal is to maximize conversion while minimizing leakage.

The final design goal is that only the final product is a Dicer substrate. This requirement is necessary to maintain the mechanism OFF state. The starting structures and intermediates cannot be Dicer substrates in order to prevent both non-conditional silencing and degradation of the components. This goal was achieved by making the molecules smaller than a standard Dicer substrate and/or using 2′-O-Me modifications to prevent enzyme cleavage. Previous reports indicated that modifying an scRNA with 2′-O-Me could prevent Dicer processing of a substrate that would otherwise be a Dicer substrate.[5]

2.2 Results and discussion

2.2.1 Mechanism design

The design process began with a sketch of the mechanism pathway. The sketch depicted the structures of the reactants, intermediates, and products, and mapped the desired pathway interactions (similar to the mechanism schematics shown in panel (a) of Figures 2.1, 2.5, and 2.8). We believe that simpler mechanisms are preferable since they present fewer opportunities for non-ideal behavior. The number of base pairs in each structure was used as a rough estimate of the free energy of the structure. (This estimate falsely assumes all base pairs are energetically equivalent.) The

ON-pathway reaction is driven forward through energetic gains at each step (each step must increase the number of base pairs). The OFF-pathway reaction is minimized by ensuring that either (1) the reactants are energetically favored over the leakage products (the reactants contain more base pairs than the reactants) or (2) the reactants are kinetically trapped in structures that have no traversable pathway connecting the reactants with the leakage products.

Each structure was segmented into sequence domains. A domain (e.g., 'a' or 'b') is a short sequence which appears in multiple locations. Complementary domains are indicated by an '*.' For example, 'a' and 'a^*' are complementary sequences and base pair to form the duplex 'a/a^*.' A well designed set of sequences has orthogonal domains. Thus, 'a' will only base pair to 'a^*' and not to 'b', 'b^*', etc. In reality, it can be difficult to design domains which are completely orthogonal, especially for complex mechanisms with sequence constraints.

Assigning domains to the structures programs the conformational changes in the hybridization cascade. Conformational changes occur either through base pairing between two unpaired complementary domains (e.g., 'a' + 'a^*' $\rightarrow$ 'a/a^*') or through branch migration where two domains compete with each other for base pairing to a complementary domain (e.g., 'a_1/a^*' + 'a_2' $\rightarrow$ 'a_1' + 'a_2/a^*' describes a 3-way branch migration where 'a_1' and 'a_2' are two instances of the same domain). Because branch migration is slow to initiate, strand exchange can be sped up through the use of a 'toehold' (an unpaired domain) adjacent to the duplex region. The invading strand will first base pair with the toehold and then initiate branch migration.[6]

2.2.2 Sequence design

Strand sequences were designed by NUPACK once the mechanism pathway was fully defined. Conditional RNAi mechanisms are highly constrained by the choice of input and output mRNAs, and the standard sequence design algorithm previously utilized by NUPACK proved insufficient for this design problem. An 'External Sequence Constraint' design algorithm was developed to scan the input and output mRNA sequences for sub-sequences that function together to produce behavior that best matches the pathway prescribed in the mechanism schematic.[7,8] The NUPACK design algorithm solves the constrained multistate sequence design problem by minimizing the sum of the ensemble defect for the initial and final states plus key intermediates specified in an objective function. The ensemble defect is defined as the average number of incorrectly paired nucleotides at equilibrium evaluated over the ensemble of (unpseudoknotted) secondary structures in the objective function.[9,10] Minimizing the ensemble defect provides both positive and negative design, resulting

in sequences where the desired structures are favored and all others are disfavored.[9,10] Any domain unconstrained by the input or output target was designed by NUPACK to minimize the ensemble defect. The constraints imposed by the external sequences drastically reduced the design space. Given this challenge, automatic sequence design was followed by NUPACK test-tube analysis to assess the tradeoffs between different sets of sequences. Sequences were handpicked based on NUPACK analysis results for subsequent test-tube validation. After experimental evaluation, the objective function was altered (e.g., by adding/removing target structures or modifying component dimensions) and the mechanism was iteratively optimized.

All three mechanisms were designed to produce a Dicer substrate specific to the silencing target d2EGFP mRNA upon the detection of a sub-sequence of detection target DsRed2 mRNA. We chose these mRNA sequences with tissue culture experiments in mind (refer to Chapter 4). While the two mRNAs were held constant between the three mechanisms, the exact mRNA sub-sequences used varied for each mechanism. The mRNA regions used are listed in Table 2.2. Since the mechanisms are completely (or nearly completely) constrained by the d2EGFP and DsRed2 mRNA sequences, the selection of the mRNA sub-sequences was an important design variable.

Following mechanism and sequence design, the mechanisms were validated in the test tube. These studies verified signal transduction, quantified the OFF/ON response, and ensured that only the final product was processed by recombinant Dicer. The ON state was verified by introducing either a short detection target X_s (a synthetic RNA strand corresponding to the appropriate mRNA sub-sequence) or the full-length mRNA detection target X (*in vitro* transcribed DsRed2). The OFF state for each mechanism was verified for three conditions: the absence of an input target, the addition of full-length mRNA silencing target Y (*in vitro* transcribed d2EGFP) and the addition of full-length *in vitro* transcribed GAPDH mRNA (Z). The output target Y is included to ensure that no unexpected interactions occur between the scRNAs and the output target. By definition, the scRNAs have domains complementary to Y. Thus, the silencing target Y could potentially induce unintentional triggering for a poorly conceived mechanism. Full-length *in vitro* transcribed GAPDH mRNA tests for off-target triggering with a non-related mRNA. While GAPDH mRNA does not encompass all endogenous sequences, we felt that it offered more insight into off-target effects than reactions with total RNA. It is difficult to use sufficient amounts of total RNA *in vitro* to produce meaningful results (i.e., both on- and off-target effects would likely be undetectable due to the low concentration of any given mRNA in total RNA). For experiments with recombinant Dicer, we show triggering with only X_s to eliminate the possibility of Dicer cleaving the mRNA,

which would make the results more difficult to interpret.

Mechanism	Detection Target: DsRed2 mRNA location	Silencing Target: d2EGFP mRNA location
Mechanism 1 (M1)	598–615	542 –562
Mechanism 2 (M2)	277–305	137–157
Mechanism 3 (M3)	9–46	70–92

Table 2.2: Source of the detection and silencing target sequences used for each conditional Dicer substrate formation mechanism. The detection target is a sub-sequence of DsRed2 mRNA, while the silencing target is a sub-sequence of d2EGFP mRNA. Numbering begins at the translation start site of the corresponding mRNA. Full mRNA sequences are listed in Appendix B.2.

2.2.3 Computational stepping analyses

Equilibrium test-tube calculations were performed using the analysis feature of the NUPACK web application to step through the molecular assembly and disassembly operations for each mechanism (depicted in the mechanism schematics). These calculations were used to check that the desired reactants, intermediates, and products were predicted to form with high yield in test tubes containing different subsets of strands. Typically, sequence domains that were intended to be completely unstructured were predicted to have some degree of base pairing at equilibrium, reflecting the challenge of designing scRNA hybridization cascades using sequences that are constrained by external sequences. Analysis calculations were performed using nearest-neighbor free energy parameters for RNA at 37° in 1M Na^+ and 0.5 μM per strand.[11] Chemical modifications (2′-O-Me RNA) were not accounted for in the physical model.

2.2.4 Stability versus metastability

Conditional hybridization cascades use discrete steps to travel isothermally through the free-energy landscape. Ideally, the scRNA reactants are maintained in the OFF state until the addition of the detection target initiates the pathway. The OFF state is achieved by utilizing (1) metastable reactants which are trapped in a local minimum free-energy state, (2) stable reactants which reside in the global minimum free-energy state, or (3) reactants which partition between multiple states at equilibrium. For a metastable mechanism, the equilibrium state favors the products regardless of the presence of the detection target. For a stable mechanism, the equilibrium state favors the reactants when the detection target is absent, but favors the products when the detection target is present. A stable OFF state is considered more resilient to perturbations than a metastable

OFF state. We classify a mechanism as stable only if the reactants dominate the OFF state at equilibrium. An OFF state that partitions between the reactants and products at equilibrium is classified as metastable because a kinetic trap is still required.

Mechanism	Final Product	Reactants only Concentration (μM)	Reactants + X_s Concentration (μM)	Computational Prediction	Experimental Result
M1	B·C	1×10^{-3}	0.5	stable	stable
M2	B	2×10^{-7}	0.5	stable	stable
M3	B·D	0.5	0.5	not stable	metastable

Table 2.3: NUPACK test-tube analysis: stable versus metastable mechanisms. Each mechanism was analyzed in NUPACK in the absence or presence of X_s at 37°C and 0.5 μM/strand using RNA nearest-neighbor energy parameters.[11] The predicted concentration of the final product is reported. A mechanism is considered stable if the predicted final product concentration is near zero in a test tube containing only the reactant scRNAs.

Table 2.3 classifies the three conditional RNAi mechanisms as stable or not stable based on NUPACK analysis of the equilibrium distribution of the reactions with and without X_s at 37°C and 0.5 μM/strand. The calculations use RNA nearest-neighbor energy parameters, despite the fact that the mechanisms tested experimentally were extensively modified with 2′-O-Me, for which the energy parameters are not fully determined.[12,13] While these predictions serve as useful guides for understanding the mechanisms, the classification of stable or metastable must be verified experimentally. (Refer to Figures 2.4, 2.7, and 2.8.)

It can be difficult to determine experimentally if reactants are stable or metastable. Annealing (heating the components together to a high temperature and cooling slowly to room temperature) is a method frequently used in structural nucleic acid nanotechnology to access the minimum free-energy state and relax the system to equilibrium.[14] In a stable system, the reactants are at the thermodynamic minimum and the system should return to the initial state when the structures are disrupted. In a metastable system, the system should find the true thermodynamic minimum when the initial structures are disrupted and the trap is removed. Unfortunately, if hairpins or molecules with significant internal secondary structure are present, annealing may not bring the system to the global minimum free-energy state.[2,15] As the system cools, the internal base pairs form at a higher temperature than the intermolecular base pairs, potentially trapping the system in a local minimum. Anneals are shown for most steps of the three mechanisms; however, caution should be used when drawing conclusions from anneals about reactant stability versus metastability.

2.2.5 Mechanism 1: Conditional DsiRNA formation using stable scRNAs

Our first implementation of conditional RNAi, "Conditional DsiRNA formation using stable scRNAs," is shown in Figure 2.1(a). It is a two-step reaction using two stable scRNAs, one duplex and one hairpin, to generate a DsiRNA. The duplex A·B detects X and B is released. B is now available to invade the loop of hairpin C, leading to the formation of the duplex B·C with the canonical 2-nt 3′ overhang of a Dicer substrate. Strand A and a portion of C are modified with 2′-O-Me to prevent Dicer cleavage of the scRNA reactants and intermediates. Functionally, A·B detects X, leading to the production of the DsiRNA B·C targeting Y.

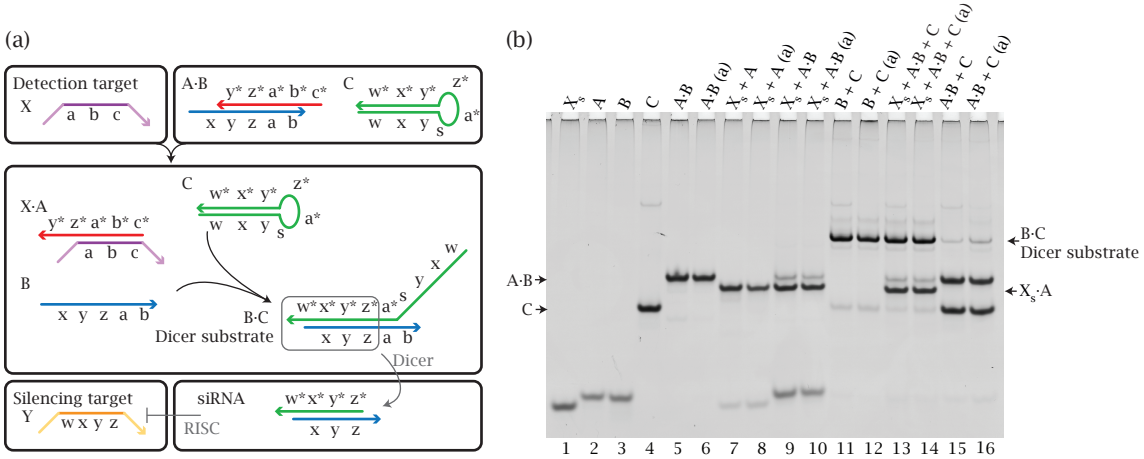


Figure 2.1: Mechanism 1: Conditional DsiRNA formation using stable scRNAs. **(a)** Mechanism. scRNA A·B detects the detection target X_s (corresponding to the mRNA sub-sequence '$a-b-c$') and releases B, which opens hairpin C. B·C is a Dicer substrate. Domain lengths: $|a| = 6, |b| = 4, |c| = 8, |s| = 5, |w| = 2, |x| = 12, |y| = 4, |z| = 3$. **(b)** Conditional Dicer substrate formation *in vitro*. In the OFF state, a minimal amount of B·C is produced (lane 15, anneal: lane 16). In the ON state, the Dicer substrate B·C is produced (lane 13, anneal: lane 14). (a) indicates an anneal.

The mechanism quickly achieves signal transduction through two steps, a simplification over the three steps required for a mechanism composed of Yin et al. style hairpins.[16] The two-step signal transduction is achieved through the use of a duplex scRNA in the first step and a loop invasion in the second step. As depicted in Figure 2.1, the duplex A·B detects the input target X through base pairing between domain 'c' in X and the toehold 'c^*' in A. Branch migration occurs through the duplex domain 'b/b^*' and 'a/a^*' with X competing with B for base pairing with A. The remaining base pairs (7-bp) holding A·B together spontaneously disassociate due to entropic forces. B, now single stranded, binds to the loop of hairpin C through the toeholds 'a' and 'z'. Loop invasion, followed by branch migration, opens C and forms the final product B·C.

The sequences, which are completely constrained by the sequences of DsRed2 mRNA (X) and d2EGFP mRNA (Y), were designed using NUPACK. The design objective function contained the structures X_s, A, B, C, A·B, A·B·C, X_s·A, and B·C. This objective function is likely over-constrained and can be pared down for future designs. The target secondary structures, except those for A and A·B·C, are depicted in Figure 2.1(a). The target secondary structure for A was the single-stranded monomer, and the target secondary structure for the 'trimer' A·B·C was comprised of the duplex A·B and the hairpin C. Two design modifications were made to allow visual separation of the components by 20% native PAGE. First, X_s was lengthened to be a slightly larger sub-sequence of DsRed2 mRNA than the minimum required (the sub-sequence '$a-b-c$'). This modification allows separation of X_s·A from the siRNAs. Second, five non-reactive nucleotides were added to the loop of C to separate it from the siRNA bands. These 5-nt are the only unconstrained nucleotides in the scRNA sequence design.

NUPACK test-tube analysis verified that these modifications did not significantly alter the equilibrium properties of the mechanism. Figure 2.2 shows NUPACK predictions for test tubes containing either the reactants, intermediates, or products. NUPACK predicts that increasing the length of X_s introduces undesired internal secondary structure and reduces conversion at step 1. However, the full reaction is predicted to fully convert. Hence, the strong second step compensates for the relatively weak first step.

Strand A is modified with 2′-O-Me to prevent duplex A·B from acting as a non-standard Dicer substrate. Hairpin C is an RNA:2′-O-Me hybrid to prevent unintended cleavage of C in the hairpin state. The majority of the region in C that binds to B is RNA to allow Dicer cleavage of the final product (B·C). The remaining nucleotides are 2′-O-Me RNA. In the future, B may need to be partially modified to protect it during the short time it exists as an ssRNA. Any modifications to B, however, cannot interfere with Dicer cleavage of B·C or the subsequent functioning of the siRNA.

Figure 2.1(b) demonstrates each step of the mechanism using native polyacrylamide gel electrophoresis (PAGE). An anneal of each step is included. Both the A·B duplex (lane 5) and C hairpin (lane 4) run as single bands, indicating that both are well formed. The first step of the mechanism leads to X_s·A and B (lane 9). In the second step, B opens C to produce B·C (lane 11). The full reaction produces Dicer substrate B·C (lane 13). When A·B and C are incubated in the absence of a detection target (lane 15), a small amount of leakage is visible.

The OFF and ON states of Mechanism 1 (M1) are demonstrated in Figure 2.3(a) and quantified

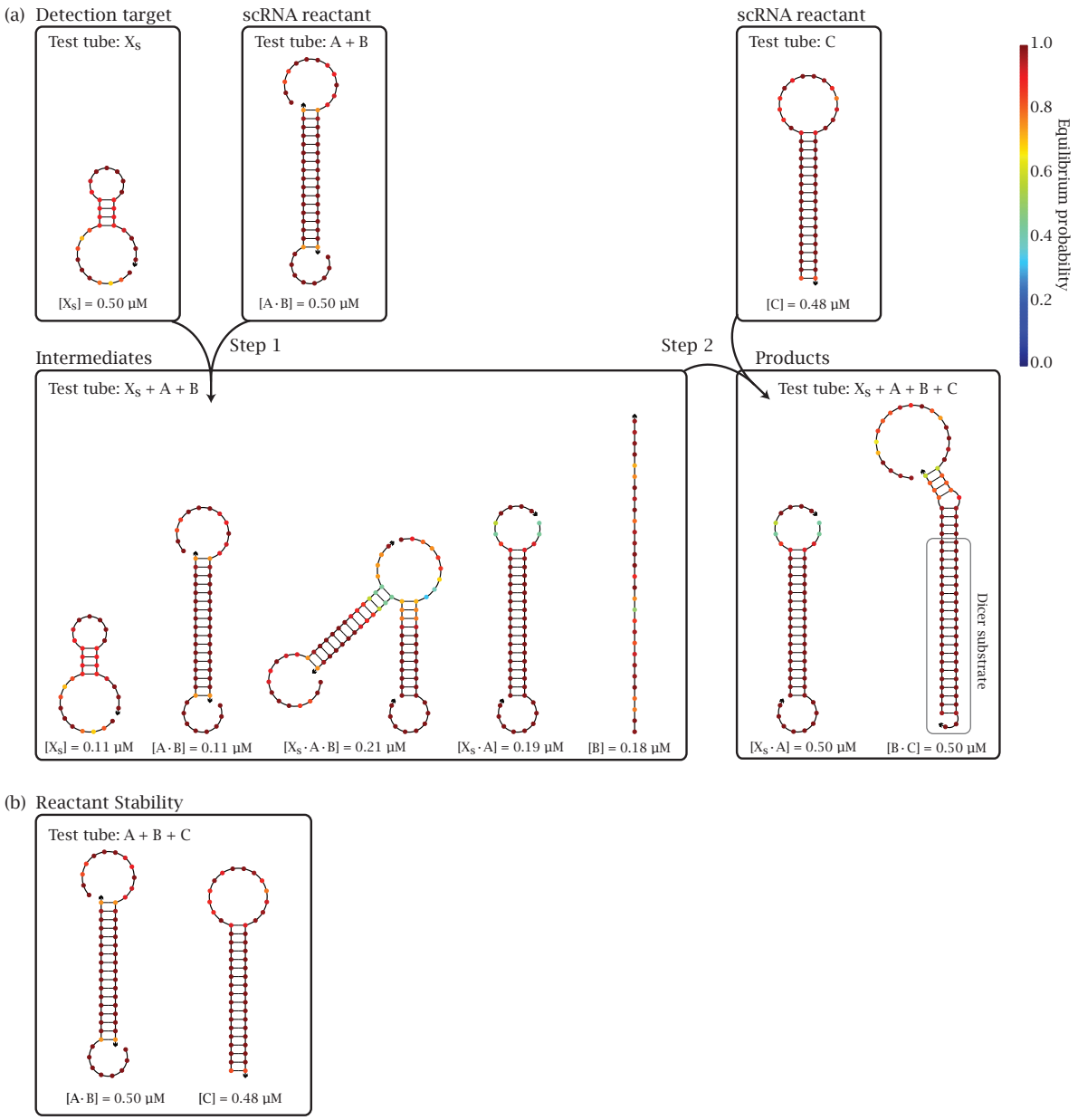


Figure 2.2: Computational stepping analysis for Mechanism 1. **(a)** Equilibrium test-tube calculations showing the predicted concentration and base-pairing properties for X_s, individual scRNA reactants, and the pathway intermediates and products. Step 1 yields the desired complexes, X_s·A and B, in addition to unreacted scRNAs and the trimer X_s·A·B. This indicates that step 1 is relatively weak, likely due to the internal secondary structure in X_s. Step 2 yields the expected products with C driving the reaction to completion. **(b)** Equilibrium test-tube calculations predict that the scRNAs A·B and C are stable. A test tube containing A, B and C predominantly form the duplex A·B and hairpin C. **(a, b)** Each box represents a test tube containing the strands listed at 0.5 μM each. The predominate complexes for each test tube, up to maximum complex size of four, are shown as MFE structures where the base shading represents the probability of that base adopting the depicted state at equilibrium. The predicted concentration of each complex in the specified test tube is listed below the corresponding MFE structure.[7]

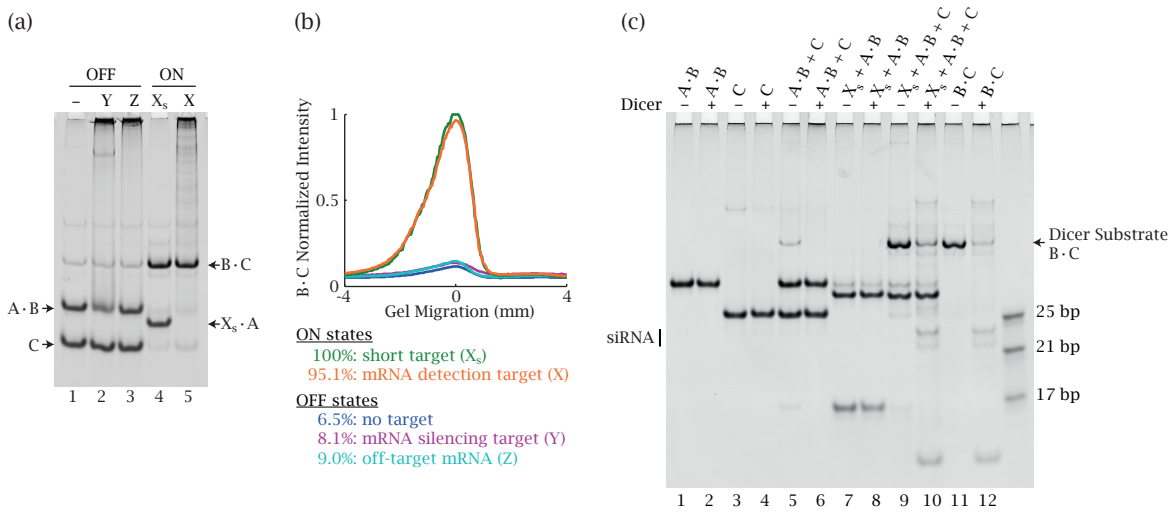


Figure 2.3: Mechanism 1: ON/OFF state quantification and Dicer processing. **(a)** OFF and ON state demonstration. OFF state: minimal production of Dicer substrate B·C in the absence of detection target X ('−'), the presence of mRNA silencing target Y, or the presence of mRNA off-target Z. ON state: strong production of B·C in the presence of short RNA detection target X_s ('$a - b - c$') or full-length mRNA detection target X. **(b)** Quantification of the Dicer substrate band (B·C) in (a). Plot of the normalized SYBR Gold intensity of the Dicer substrate B·C band versus gel migration distance centered around the B·C band. **(c)** Conditional siRNA production *in vitro*. The reactants and intermediates are not cleaved by Dicer. Dicer substrate B·C is efficiently processed by Dicer to produce siRNAs (lanes 10 and 12). (−/+) indicates absence/presence of Dicer.

in Figure 2.3(b). The ON state is strong when triggered with either a short RNA detection target X_s (lane 4) or a full-length mRNA detection target X (lane 5). All three OFF states produce less than 10% leakage, giving an order of magnitude discrimination between the OFF and ON states. See Figure C.1 in Appendix C for quantification of the ON/OFF states in three independent gels.

Efficient Dicer processing of the final product B·C, but not the reactants or intermediates, is shown in Figure 2.3(c). In the presence of recombinant Dicer, the reactants A·B and C are not cleaved (lanes 2 and 4 respectively). When the reaction pathway is initiated with X_s, B·C forms and is subsequently cleaved to various length siRNAs (lane 10). This cleavage patten appears identical to the cleavage pattern of B·C alone (lane 12). The small amount of B·C produced in the absence of an input target is cleaved (lane 6). This is expected because B·C is a Dicer substrate regardless of its origin. The multiple bands formed upon Dicer cleavage correspond to both the siRNA and the portions of B and C which are removed.

This mechanism was designed to be stable. Based on the anneal, it appears stable because the production of B·C is similar for the two hour incubation and the anneal of A·B and C (refer to Figure 2.1(b), lanes 15 and 16). To provide an even more convincing demonstration of scRNA stability, we incubated the products A and B·C together. For a stable mechanism, the scRNA reactants will dominate at equilibrium. Thus, incubating the products of a stable mechanism together will lead to the formation of the stable scRNA reactants. When the products are incubated together, they slowly convert to the initial reactants A·B and C through A + B·C $\rightarrow$ A·B + C as shown in Figure 2.4. Even though this reaction is highly favorable and nearly all the strands return to the scRNA reactants A·B and C, there will always be a small amount of the leakage products present due to thermodynamic partitioning between the reactants and products.

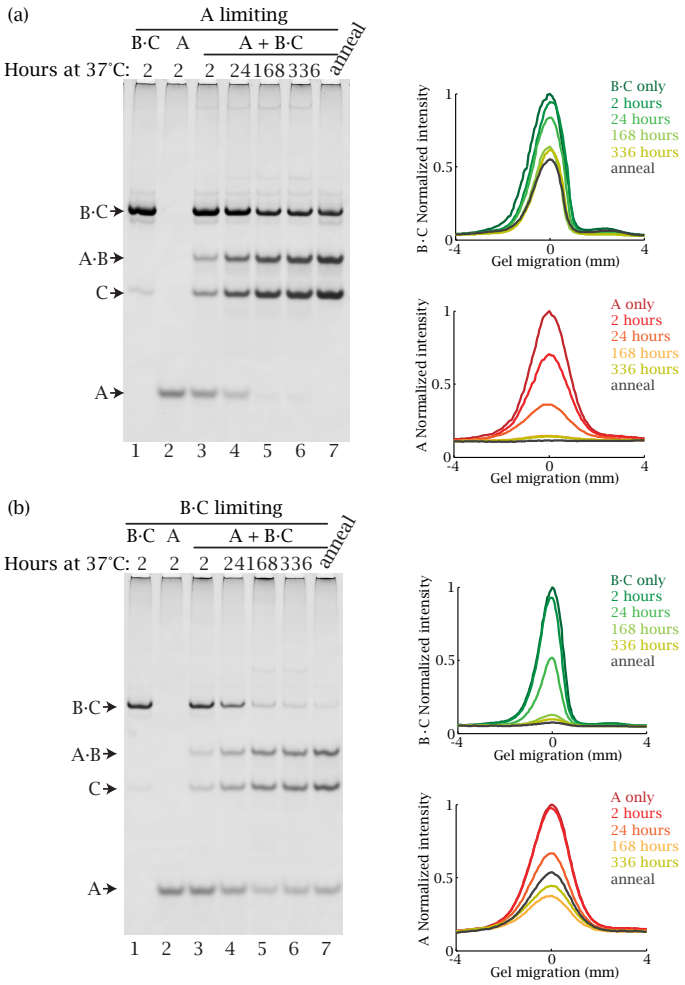


Figure 2.4: Mechanism 1 uses stable scRNA reactants. A·B and C are stable scRNAs based on native PAGE analysis of the reverse reaction, A + B·C → A·B + C, after 2, 24, 168 (1 week), or 336 (2 weeks) hours of incubation at 37°C. After 336 hours (2 weeks) of incubation at 37°C, the limiting reactant is nearly depleted. The distribution of the annealed reaction appears very similar to the distribution for the reaction after 336 hours of incubation at 37°C. The plots depict the relative SYBR Gold intensity of either the B·C or A band. **(a)** A is limiting. **(b)** B·C is limiting.

2.2.6 Mechanism 2: Conditional shRNA formation using a single stable scRNA

Our second implementation of conditional RNAi simplifies the signal transduction to a single step. The schematic of Mechanism 2, "Conditional shRNA formation using a single stable scRNA," is shown in Figure 2.5(a). Signal transduction is achieved in a single step through the use of one duplex scRNA, A·B, which disassociates upon detecting the input signal to produce the final substrate B, a monomer shRNA. Functionally, A·B detects X, leading to the production of the shRNA B targeting Y.

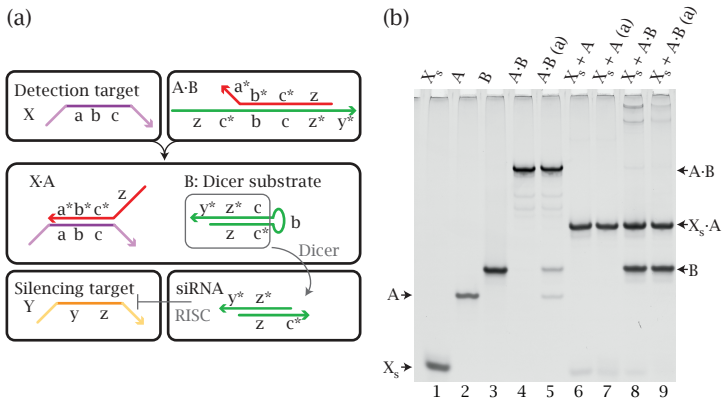


Figure 2.5: Mechanism 2: Conditional shRNA formation using a single stable scRNA. **(a)** Mechanism schematic. The input target X hybridizes to A·B to form X·A and release shRNA B. Domain lengths: $|a| = 12, |b| = 14, |c| = 3, |y| = 2, |z| = 19$. **(b)** Conditional Dicer substrate formation *in vitro*. In the OFF state, a minimal amount of B is produced (lane 4, anneal: lane 5). In the ON state, the shRNA B is produced (lane 8, anneal: lane 9).

Figure 2.5(a) illustrates the mechanism. In the initial configuration, A binds to the loop and one half of the hairpin stem of B to form the complex A·B. Input target X initiates the hybridization cascade by binding to the exposed toehold 'a^*' on A in A·B and branch migrates through domains 'b^*' and 'c^*.' The release of domain 'c' on B allows it to form intramolecular base pairs with the domain 'c^*,' also in B. The subsequent branch migration completely displaces A from B. The final products are X·A and shRNA B.

Mechanism 2 (M2) was inspired by the fact that a Dicer substrate can be a hairpin, an shRNA, instead of a duplex RNA. This point was clearly illustrated by the observation that the unmodified RNA hairpin C from M1 silenced Y after transfection in tissue culture (refer to Chapter 4 for a discussion). M2 is essentially a subset of M1 in reverse, with slightly different dimensions. As demonstrated earlier, the leakage products of M1 revert to the reactant scRNAs after incubation at 37°C: A + B·C → A·B + C. M2 is exactly this reaction where A is the input target X and B·C

is renamed to A·B yielding X + A·B → X·A + B. The reaction rate is increased by lengthening the toehold/toehold nucleation domain between X and A (increased from 4-bp in M1 to 12-bp in M2).

To repurpose the reverse leakage reaction of M1 to a functional conditional RNAi mechanism, the shape switching needed to be reversed. This required changes to both the 2′-O-Me modification pattern and domain lengths. In M2, the hairpin must be a Dicer substrate and the duplex immune to Dicer processing (the exact opposite of the desired processing properties in M1). The hairpin, B, was changed to unmodified RNA with a 23-bp stem and a 2-nt 3′ overhang to promote Dicer processing. The kissing strand, A, which disrupts the hairpin structure to form the duplex, is completely 2′-O-Me RNA. This prevents the duplex A·B from serving as a Dicer substrate.

NUPACK was used to design sequences for the mechanism. As for M1, the sequences for M2 are completely constrained by the sequences of DsRed2 mRNA (X) and d2EGFP mRNA (Y). The design objective function contained the structures X_s, A, B, A·B, and X_s·A. The target secondary structures, except for the one for A, are depicted in Figure 2.5(a). The target secondary structure for A was the single-stranded monomer. The sequences were verified to have the desired equilibrium properties by NUPACK test-tube analysis. Figure 2.6 shows the NUPACK predictions for test tubes containing either the reactants or products and visually demonstrates that the one-step mechanism is extremely simple.

The stepping properties of the mechanism are demonstrated with native PAGE (Figure 2.5(b)). A·B is well-formed and there are no detectable monomer bands for A and B (lane 4). The scRNA reactant is also the OFF state of this simple mechanism. Annealing A·B produces a small amount of A and B because the internal base pairs of B are favored during the annealing process (lane 5). The full reaction, X_s + A·B, produces predominately X_s·A and B (lane 8). Since B alone silences, it is critical to carefully prepare the starting A·B duplex through native PAGE purification.

The ON and OFF states of the mechanism are shown in Figure 2.7(a) and are quantified in Figure 2.7(b). The ON state with X_s is very strong and consumes all of A·B, leading to over two orders of magnitude discrimination between the ON and OFF states. Full-length mRNA X is less successful at initiating the reaction and produces only a quarter the amount of B produced with X_s. This is still, however, more than one order of magnitude greater than the OFF states. Changing the location in DsRed2 mRNA that serves as the detection target will likely increase the conversion with X. For all three OFF states, there is no detectable production of B (<0.5% relative to B produced in the ON state with X_s). See Figure C.2 in Appendix C for quantification of the

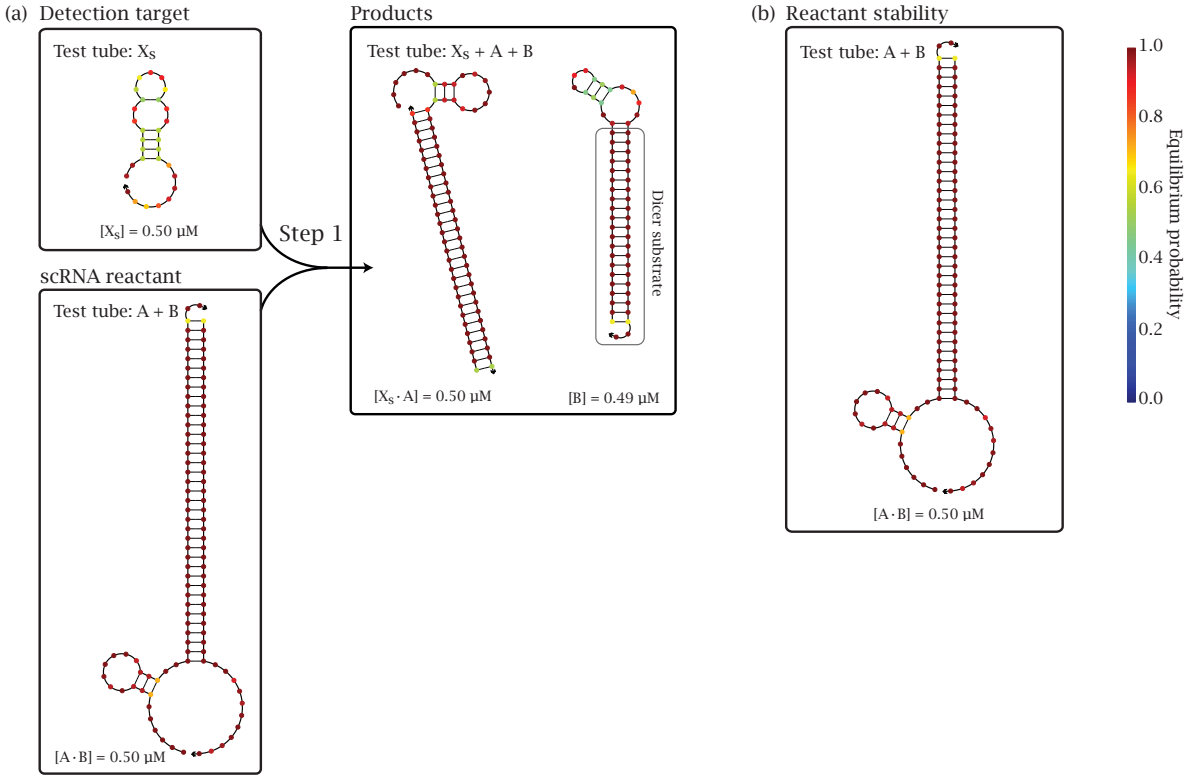


Figure 2.6: Computational stepping analysis for Mechanism 2. **(a)** Equilibrium test-tube calculations showing the predicted concentration and base-pairing properties for hypothetical test tubes corresponding to X_s, scRNA reactant A·B, and the final products. There is some internal base pairing in domains intended to be single stranded. The products are expected to dominate the full reaction. **(b)** Equilibrium test-tube calculations predict that scRNA A·B is stable. A test tube containing A and B predominantly forms the duplex A·B. **(a, b)** Each box represents a test tube containing the strands listed at 0.5 μM each. The predominate complexes for each test tube, up to maximum complex size of three, are shown as MFE structures where the base shading represents the probability of that base adopting the depicted state at equilibrium. The predicted concentration of each complex in the specified test tube is listed below the corresponding MFE structure.[7]

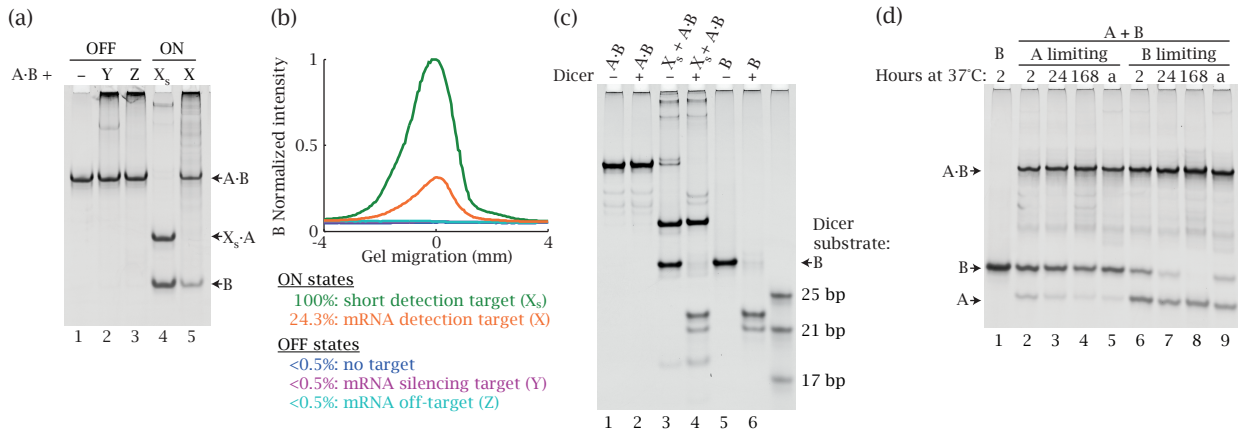


Figure 2.7: Mechanism 2: ON/OFF state quantification, Dicer processing, and stability demonstration. **(a)** OFF and ON state demonstration. OFF states: non-detectable production of Dicer substrate B in the absence of detection target X ('−'), the presence of mRNA silencing target Y (full-length d2EGFP mRNA), or the presence of mRNA off-target Z (full-length GAPDH mRNA). ON states: strong production of B in the presence of short RNA detection target X_s ('$a-b-c$') or full-length mRNA detection target X (full-length DsRed2 mRNA). **(b)** Quantification of the final product band B in (a). Plot of the normalized SYBR Gold intensity of the shRNA B band versus gel migration distance centered around the B band. **(c)** Conditional siRNA production *in vitro*. The reactants and intermediates are not cleaved by Dicer. Dicer substrate B is efficiently processed by Dicer to produce siRNAs (lanes 4 and 6). (−/+) indicates absence/presence of Dicer. **(d)** scRNA reactant stability. Demonstration that A·B is a stable scRNA through analysis of the reverse leakage reaction, A + B → A·B, after 2, 24, or 168 hours of incubation at 37°C. The annealed reaction is denoted by 'a.'

ON/OFF states in three independent gels.

In the presence of recombinant Dicer, only the shRNA B is cleaved to produce siRNAs (Figure 2.7(c)). Addition of X_s triggers the release of shRNA B (lane 3), which is cleaved by Dicer (lane 4). Interestingly, B is cleaved into two distinct bands which run as 21-bp and 22- or 23-bp (lane 6). The two bands potentially correspond to two distinct siRNAs of different lengths or a single-length siRNA and the hairpin loop fragment, which is released upon Dicer cleavage. The duplex A·B is not a Dicer substrate (lane 2), likely because A is modified with 2′-O-Me RNA. The small amount of A·B remaining in the '−' Dicer full reaction (lane 3) disappears when Dicer is added (lane 4). This disappearance could be due to a small amount of Dicer cleavage of A·B. Alternatively, the cleavage of B by Dicer could shift the equilibrium towards the products and allow complete consumption of A·B.

To demonstrate the stability of A·B, we show that the cognate reaction occurs spontaneously in reverse (i.e., A + B $\rightarrow$ A·B). A and B were incubated for 2, 24, or 168 hours at 37°C in reactions where either A or B were the limiting reactant (Figure 2.7(d)). An anneal was included for comparison. After 168 hours (1 week) of incubation at 37°C, the limiting reactant is no longer visible as a monomer and is entirely in the A·B duplex. Annealing A and B does not capture the equilibrium behavior, and the monomers are favored compared to an isothermal incubation for 168 hours. This is consistent with the fact that annealing reactions favor hairpins. The reaction is relatively slow, perhaps due to undesired secondary structure in the A or B monomers.

2.2.7 Mechanism 3: Conditional DsiRNA formation via template-mediated 4-way branch migration

Our third mechanism was designed to explore the alternate design principle of 4-way branch migration. In Mechanism 3, "Conditional DsiRNA formation via template-mediated 4-way branch migration," two scRNA duplexes are brought in close proximity through the use of a template (Figure 2.8). This facilitates strand exchange through 4-way branch migration between the two duplexes and leads to the production of a Dicer substrate. The dual binding event localizes the two scRNAs and releases two short toeholds, which initiate the 4-way branch migration and increase the rate of exchange. Functionally, A·B and C·D detect X, leading to the production of the DsiRNA B·D targeting Y.

This mechanism consists of two duplexes, A·B and C·D, with interior bulges and relatively long toeholds (Figure 2.8(a)). The input target X is detected through hybridization between the long

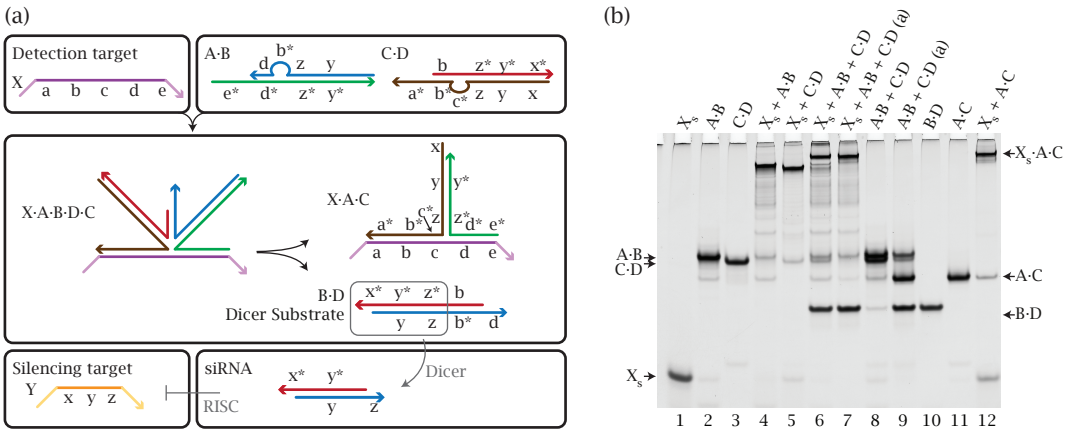


Figure 2.8: Mechanism 3: Conditional DsiRNA formation via template-mediated 4-way branch migration. **(a)** Mechanism. The input target X is dually detected by complexes A·B and C·D. Toehold binding and subsequent branch migration allows a five-strand complex to form (X·A·B·D·C). A 4-way branch migration releases the Dicer substrate B·D. Domain lengths: $|a| = 8, |b| = 6, |c| = 6, |d| = 7, |e| = 11, |x| = 2, |y| = 19, |z| = 2$. **(b)** Conditional Dicer substrate formation *in vitro*. In the OFF state, a minimal amount of B·D is produced (lane 8, anneal: lane 9). In the ON state, the Dicer substrate B·D is produced (lane 6, anneal: lane 7).

toeholds ('e^*' on A·B and 'a^*' on C·D) and X. Subsequent branch migration between X and the duplexes releases both of the interior bulges and exposes 'b' and 'b^*.' These two short toeholds initiate the 4-way branch migration between A·B and C·D, increasing the overall speed of strand exchange.[17] Two different methods were used to sequester these short toeholds. In A·B, 'b^*' is sequestered in an interior bulge. In C·D, 'b' is sequestered in a duplex. Sequestering the toehold in a duplex is likely more effective at preventing leakage than the interior bulge. Unfortunately, using a duplex to sequester both short toeholds would have constrained the detection target sequence to contain both 'b' and 'b^*'. This would greatly limit the number of possible endogenous input sequences and was avoided. The interior bulge on C·D ensures that the 3-way branch migration between X and C·D progresses to completion so that 'b' is single stranded and available for the next step. The 4-way branch migration resolves to the final Dicer substrate (B·D) and a waste by-product bound to the input target (X·A·C). In the absence of the input target X, the duplexes do not interact and the system is OFF.

To date, nearly all nucleic acid hybridization cascades have used strand displacement mediated by 3-way branch migration. In 3-way branch migration, the invading unpaired strand replaces one strand in a duplex.[6] In 4-way branch migration, two duplexes exchange strands.[18] This method of strand exchange has not been significantly explored in nucleic acid engineering.[19] Initially, we attempted to produce an siRNA as the final substrate to avoid requiring Dicer processing. However,

the need to initiate the 5-way junction with short auxiliary toeholds introduces a small amount of sequence dependence between the template and the final duplex. If the final duplex is a Dicer substrate instead of an siRNA, these regions of sequence carry-over can be removed through Dicer cleavage to produce an siRNA that is completely sequence independent from the template. Therefore, we chose to focus on producing a final Dicer product instead of an siRNA.

This conditional RNAi mechanism introduces the novel concept of template-mediated nucleation. Templated nucleation brings two nucleic acids in close proximity through base pairing with a third nucleic acid. In this example, the scRNA duplexes A·B and C·D nucleate with the detection target X. Thus, the nucleation does not require sequence complementarity between the scRNAs, as is the case for either toehold/toehold or loop/toehold nucleation. Templated nucleation allows the mechanism to proceed in one step because, by construction, the template sequence is independent from the final substrate. The short toeholds used to increase the 4-way structure initiation rate introduce a small amount of sequence dependence between the template and the final Dicer substrate, but these domains are removed by Dicer cleavage. We believe this is the first demonstration of templated nucleation to mediate conditional strand displacement via either 3-way or 4-way branch migration.

Mechanism 3 (M3) sequences were designed by NUPACK. Similar to the other two mechanisms, the scRNA sequences were entirely constrained by the sequences of the two mRNAs chosen as the detection and silencing targets. The objective function included the structures X_s, A·B, C·D, and B·D. The target secondary structures are depicted in Figure 2.8(a). We did not include the trimer X_s·A·C in the objective function because its addition would have over-emphasized less critical base-pairing requirements during the design process. As shown in Figure 2.9, the trimer is well formed even though it was not included in the objective function. Despite this success, the objective function is likely not ideal. For future iterations, it may be necessary to specify the individual monomers A, B, C, and D as single stranded to ensure that undesired internal secondary structure in the monomers does not compete with the desired duplex structures. Since both scRNA reactants A·B and C·D can potentially function as non-canonical Dicer substrates, Dicer cleavage was prevented by modifying A and C with 2′-O-Me.

The stepping properties of the mechanism are demonstrated with native PAGE in Figure 2.8(b). A·B and C·D exist primarily as duplexes (lanes 2 and 3). Individually, the duplexes bind to X_s, showing that the two target detection events can occur independently (lanes 4 and 5). When both duplexes are incubated with X_s in the full reaction, the mechanism behaves as designed and the

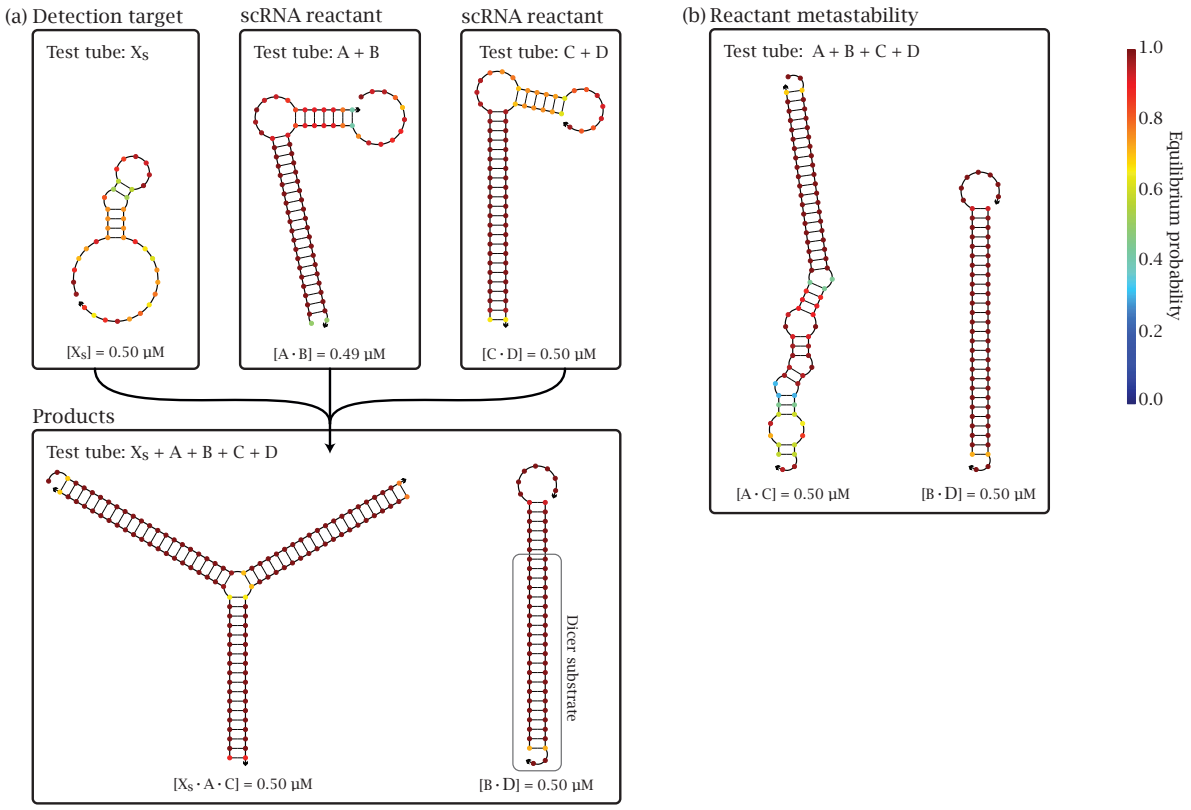


Figure 2.9: Computational stepping analysis for Mechanism 3. **(a)** Equilibrium test-tube calculations showing the predicted concentration and base-pairing properties for hypothetical test tubes corresponding to X_s, individual scRNA reactants, and the final products. There is some internal base pairing in X_s. The final products form with near quantitative yield for the full reaction. **(b)** Equilibrium test-tube calculations predict that scRNAs A·B and C·D are not stable. A test tube containing A, B, C, and D predominantly forms the duplexes A·C and B·D, the leakage products, at equilibrium. **(a,b)** Each box represents a test tube containing the strands listed at 0.5 μM each. The predominate complexes for each test tube, up to maximum complex size of five, are shown as MFE structures where the base shading represents the probability of that base adopting the depicted state at equilibrium. The predicted concentration of each complex in the specified test tube is listed below the corresponding MFE structure.[7]

strands resolve into the final product B·D and the by-product X_s·A·C (lane 6). When the duplexes are incubated together without a target, a small amount of the leakage products B·D and A·C form (lane 8). For this system, it is difficult to ensure that all of the strands are present in the correct stoichiometry. Duplexes A·B and C·D were purified by native PAGE to ensure correct stoichiometry between the two individual strands in the duplexes. Titration gels were used to determine the correct stoichiometry between the duplexes.

The ON and OFF states of the mechanism are shown in Figure 2.10(a) and are quantified in Figure 2.10(b). The ON state with both X_s and full-length X show strong production of the final product B·D. The strong ON state for X is likely due to a combination of selecting an accessible region in X and using long toeholds to detect this region. The three OFF states produce under 10% leakage. The amount of leakage produced is variable between the three OFF states, likely due to pipetting variation and, potentially, binding of B·D to the mRNA silencing target Y and/or mRNA off-target Z. See Figure C.3 in Appendix C for quantification of the ON/OFF states in three independent gels.

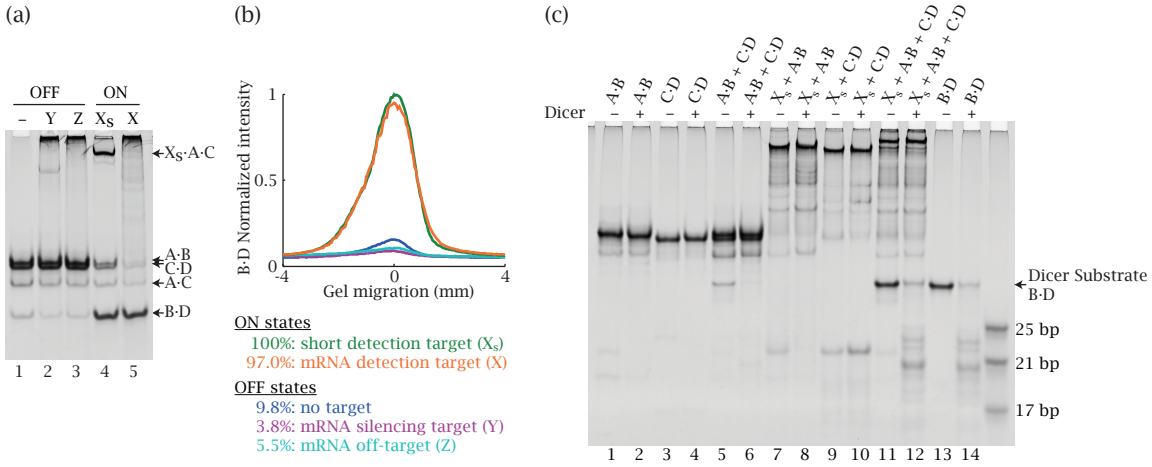


Figure 2.10: Mechanism 3: ON/OFF state quantification and Dicer processing. **(a)** OFF and ON state demonstration. OFF states: minimal production of Dicer substrate B·D in the absence of detection target X ('−'), the presence of mRNA silencing target Y, or the presence of mRNA off-target Z. ON states: strong production of B·D in the presence of short RNA detection target X_s ('$a-b-c-d-e$') or full-length mRNA detection target X. **(b)** Quantification of the final product band in (a). Plot of the normalized SYBR Gold intensity of the B·D band versus gel migration distance centered around the B·D band. **(c)** Conditional siRNA production *in vitro*. The reactants are not cleaved by Dicer. Dicer substrate B·D is cleaved to produce siRNAs (lanes 12 and 14). (−/+) indicates absence/presence of Dicer.

In the presence of recombinant Dicer (Figure 2.10(c)), the mechanism converts with a short input target and produces B·D. Dicer cleaves B·D to produce a variety of siRNAs (lanes 12 and 14). The

predominate band appears to be slightly shorter than 21-bp. Neither A·B nor C·D is cleaved by Dicer because A and C are modified with 2′-O-Me (lanes 2 and 4). B and D are RNA. Despite the fact that B·D is fairly similar to a canonical DsiRNA, it required twice as much recombinant Dicer as the other two mechanisms. Even with the increased amount of Dicer enzyme, B·D is not fully cleaved. Of course, recombinant Dicer reactions *in vitro* are not necessarily indicative of *in vivo* silencing, but it is curious that this final substrate is difficult to cleave.

Mechanism 3 was envisioned as a stable mechanism, but it is metastable for the set of sequences tested. NUPACK predicts that the reactants are not stable and will fully convert to the leakage products at equilibrium (i.e., NUPACK predicts that a test tube containing the strands A, B, C, and D will form the leakage duplexes A·C and B·D as shown in Figure 2.9). The leakage results from the unintended base pairing between domains 'e^*' on A and 'a^*' on C. Due to the imposed sequence constraints, these domains are not completely orthogonal. Potentially, stability could be restored by using sequences with orthogonal domains, but this may be unattainable due to sequence constraints. In which case, the length of 'e' and 'a' could be reduced, though this may limit the overall mechanism conversion.

Experimentally, an anneal of the scRNA duplexes A·B and C·D partitions the strands between the four possible duplexes (Figure 2.8(b), lane 9). An anneal of this mechanism should approximate equilibrium partitioning between the reactants and products because the four duplexes are equally favored during the cooling process (no substantial internal secondary structure). The partitioning demonstrates that the scRNA reactants A·B and C·D are more stable than predicted by NUPACK, presumably due to the 2′-O-Me modifications that are not accounted for in the NUPACK calculations. The substantial presence of the final duplexes A·C and B·D in the annealed reaction indicates that the OFF state is maintained to some degree by metastable scRNAs; hence, the mechanism is classified as metastable. (A mechanism is only classified as stable if the reactants dominate at equilibrium.)

2.3 Conclusions

We demonstrate in this chapter the engineering of three nucleic acid mechanisms that execute the molecular logic:

If mRNA detection target X is present,

produce a Dicer substrate targeting independent mRNA silencing target Y.

Test-tube studies of the three mechanisms demonstrate that the six design goals have been achieved. All three mechanisms demonstrate a strong conditional ON/OFF response, with more than an order of magnitude increase in Dicer substrate formation in the presence of the full-length mRNA detection target X. Additionally, studies with recombinant Dicer show that only the final substrate is a Dicer product, achieved through the prudent use of domain dimensioning and/or chemical modification of the scRNA reactants.

Table 2.4: Summary of the design features for three conditional RNAi mechanisms.

Design features	M1	M2	M3
number of steps	2	1	1
monomer scRNAs	1	0	0
duplex scRNAs	1	1	2
DsiRNA production	✓		✓
shRNA production		✓	
stable reactants	✓	✓	
metastable reactants			✓
toehold/toehold nucleation	✓	✓	✓
loop/toehole nucleation	✓		
template/toehold nucleation			✓
3-way branch migration	✓	✓	✓
4-way branch migration			✓
spontaneous dissociation	✓		✓
ON/OFF ratio (orders of magnitude with X_s)	1	2	1

Each conditional RNAi mechanism relies on a combination of notable properties to meet the design goals. The properties are summarized in Table 2.4. There are many possible pathways to achieve conditional RNAi—many more than those demonstrated here. These three mechanisms function through the use of monomer and/or duplex scRNA reactants to conditionally produce a Dicer substrate using, at most, two signal transduction steps. The final Dicer substrate is either a duplex (a DsiRNA) or a hairpin (an shRNA). This final product is processed by recombinant Dicer to produce an siRNA. The scRNA reactants are held in their initial conformations either by a kinetic trap (metastable reactants) or thermodynamic partitioning (stable reactants), leading to strong OFF states. Various methods of nucleation and strand exchange were explored through these three mechanisms. Overall, each mechanism detects the input target and produces an ON state at least an order of magnitude greater than the OFF state. M2 demonstrates the greatest dynamic range between the ON and OFF states (the leakage is non-detectable, leading to an ON:OFF ratio $\geq$200). M1, being a stable mechanism, can be redesigned to reduce the leakage and potentially

yield the same dynamic range as M2.

The remaining challenge is to implement these mechanisms *in vivo*. It is likely that additional, unanticipated design requirements will emerge during the engineering and validation process. By designing three unique mechanisms for conditionally forming a Dicer substrate, we demonstrate many design features to draw upon in future logic gates.

2.4 Bibliography

[1] Seelig, G., Soloveichik, D., Zhang, D. Y., and Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**(5805), 1585–8 (2006).

[2] Yin, P., Choi, H. M., Calvert, C. R., and Pierce, N. A. Programming biomolecular self-assembly pathways. *Nature* **451**(7176), 318–22 (2008).

[3] Rose, S. D., Kim, D. H., Amarzguioui, M., Heidel, J. D., Collingwood, M. A., Davis, M. E., Rossi, J. J., and Behlke, M. A. Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Research* **33**(13), 4140–56 (2005).

[4] Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D., and Doudna, J. A. Structural basis for double-stranded RNA processing by Dicer. *Science* **311**(5758), 195–8 (2006).

[5] Masu, H., Narita, A., Tokunaga, T., Ohashi, M., Aoyama, Y., and Sando, S. An activatable siRNA probe: trigger-RNA-dependent activation of RNAi function. *Angewandte Chemie* **48**(50), 9481–3 (2009).

[6] Zhang, D. Y. and Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry* **3**(2), 103–13 (2011).

[7] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32**(1), 170–3 (2011).

[8] Wolfe, B. R. Personal communication (2009).

[9] Dirks, R. M. and Pierce, N. A. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *Journal of Computational Chemistry* **25**(10), 1295–304 (2004).

[10] Zadeh, J. N., Wolfe, B. R., and Pierce, N. A. Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry* **32**(3), 439–52 (2011).

[11] Serra, M. J. and Turner, D. H. Predicting thermodynamic properties of RNA. *Methods in Enzymology* **259**, 242–61 (1995).

[12] Kierzek, E., Ciesielska, A., Pasternak, K., Mathews, D. H., Turner, D. H., and Kierzek, R. The influence of locked nucleic acid residues on the thermodynamic properties of 2'-O-methyl RNA/RNA heteroduplexes. *Nucleic Acids Research* **33**(16), 5082–93 (2005).

[13] Kierzek, E., Mathews, D. H., Ciesielska, A., Turner, D. H., and Kierzek, R. Nearest neighbor parameters for Watson-Crick complementary heteroduplexes formed between 2'-O-methyl RNA and RNA oligonucleotides. *Nucleic Acids Research* **34**(13), 3609–14 (2006).

[14] Feldkamp, U. and Niemeyer, C. M. Rational design of DNA nanoarchitectures. *Angewandte Chemie* **45**(12), 1856–76 (2006).

[15] Bois, J. S. *Analysis of Interacting Nucleic Acids in Dilute Solutions.* PhD thesis, California Institute of Technology, (2007).

[16] Schwarzkopf, M. *Engineering Nucleic Acid Mechanisms for Regulation and Readout of Gene Expression: Conditional Dicer Substrate Formation and Sensitive Multiplexed Northern Blots.* PhD thesis, California Institute of Technology, (2013).

[17] Dabby, N., Chen, H., Schaeffer, J., and Winfree, E. The kinetics of toehold-mediated four-way branch migration. In revision.

[18] Karymov, M., Daniel, D., Sankey, O. F., and Lyubchenko, Y. L. Holliday junction dynamics and branch migration: single-molecule analysis. *Proceedings of the National Academy of Sciences of the United States of America* **102**(23), 8186–91 (2005).

[19] Bois, J. S., Venkataraman, S., Choi, H. M., Spakowitz, A. J., Wang, Z. G., and Pierce, N. A. Topological constraints in nucleic acid hybridization kinetics. *Nucleic Acids Research* **33**(13), 4090–5 (2005).

Chapter 3

Conditional siRNA production in cell lysate

3.1 Introduction

Programming conditional RNAi logic in a living organism presents many challenges. Clearly, the test-tube environment and the cellular environment are quite different, and a conditional RNAi mechanism that fulfills our design goals in the test tube may fail for numerous reasons in the cellular environment. Unfortunately, it is difficult to tease apart the source of the intracellular mechanism failure since it is nearly impossible to isolate each variable (refer to Chapter 4 for a more detailed discussion of efforts in this direction). To bridge the disconnect between the test tube and the cell, we developed a functional Dicer cleavage assay in a whole-cell lysate system derived from a human cell line. The goal is to use this system to provide meaningful insight into our conditional RNAi mechanisms by focusing on a smaller set of variables in a relatively controlled environment.

Whole-cell lysate systems are proven research tools. They have played a role in elucidating the fine details of both Dicer cleavage and RISC functionality.[1–5] Extracts from *Drosophila* embryos are the most commonly used cell lysate system for studying RNAi, and experiments in *Drosophila* cell lysate were the first to demonstrate that Dicer products are 21–22-nt in length.[1,6] Mammalian whole-cell extracts have been used less frequently to study RNAi, but there are a few examples. To study translational repression mediated by miRNA let-7, Wakiyama et al. used whole-cell extracts derived from a human cell line which overexpressed components of the miRNA pathway.[7] Sakurai et al. used whole-cell extracts derived from a human cell line to investigate Dicer binding.[8]

A cell lysate system allows each step of a conditional RNAi mechanism to be verified inde-

pendently. These steps include scRNA detection of the mRNA target, scRNA signal transduction to product a Dicer substrate, and subsequent cleavage of the Dicer substrate. This step-by-step analysis is currently not possible in a living cell (e.g., when a mechanism is transfected into a cell line as in Chapter 4). With our current experimental setup, transfection of an scRNA into a cell only supplies information about the ability of the conditional RNAi mechanism to knock down the silencing target mRNA. Assaying only the last step provides information about the functionality of the entire mechanism. If the mechanism functions as desired, this is all we need to know. But if it does not (which has been the case to date), we do not know why or where the logic is failing. A cell lysate can provide more information by employing the techniques used in the development of test-tube conditional RNAi mechanisms (Chapter 2). Thus, the cell lysate allows fine tuning of signal transduction in conditions similar to the cellular environment but without many of the other challenges.

A cell lysate system removes the major challenge of delivering scRNAs. RNA delivery is a challenging problem even for canonical siRNAs or Dicer substrates.[9] Delivering scRNAs is even more demanding because scRNAs must be delivered in sufficient concentrations to the correct location inside a cell. For mechanisms built from multiple scRNAs (M1 and M3), it is necessary that the components are in close proximity at appropriate concentrations to execute the pathway. For mechanisms with metastable components (M3), and likely even for those with stable components (M1 and M2), it is necessary to maintain the proper secondary structure during delivery to maintain a strong OFF state.

The scRNA concentration can be controlled in a cell lysate system. Multicomponent scRNA mechanisms are concentration dependent, and it is unknown how they will function at low concentrations. We designed and tested the three conditional RNAi mechanisms in the test tube at 0.5 μM. Likely the scRNA intracellular concentration will be much lower because common RNA delivery methods are extremely inefficient.[10] The low scRNA concentrations prove problematic because nucleic acid hybridization pathways depend on intermolecular interactions to provide the forward energetic driving force. As the concentration of the pathway components decrease, the forward driving force may not be sufficient to generate the desired switching behavior. While NUPACK predicts that all three mechanisms will function properly at low concentrations, a cell lysate system allows us to verify this experimentally. The problem then lies in determining a concentration that corresponds to the relevant cellular concentration. The concentration of endogenous mRNA or transfected scRNAs in the cell lysate cannot be used to determine intracellular concentration

because lysing the cells destroys the true concentration information.

3.2 Results and discussion

3.2.1 Development of a cell lysate system to test conditional RNAi mechanisms

Initial cell lysate experiments were based on the mammalian whole-cell extract system previously used to study Dicer binding.[8] Whole-cell extracts were prepared from HEK 293 cell lines. The cells were lysed by sonication, and the extracts were flash frozen until use. scRNA strands were labeled with ^{32}P prior to addition to the cell lysate. After a two-hour incubation at 37°C, the RNAs were extracted from the lysate, separated by PAGE, and ^{32}P was visualized using phosphorimaging plates. Refer to Appendix B, sections B.9–B.11 for method details. Visualization of the scRNAs with SYBR Gold staining, as was done exclusively in Chapter 2, is not feasible in a cell lysate due to the abundance of cellular RNAs. Additionally, the increased sensitivity of ^{32}P visualization over SYBR Gold staining allowed investigation of the mechanisms at scRNA concentrations up to two times lower than the concentrations used in Chapter 2. Testing the mechanisms at low concentration is a necessary step towards a functional *in vivo* mechanism because, while we have not yet determined the relevant intracellular concentration of scRNAs (i.e., what is the concentration of delivered scRNAs in the cytoplasm?), we expect that it is low.

In preliminary studies, the protocol from Sakurai et al. reliably produced a cell lysate system with Dicer activity. The DsiRNA was nearly completely cleaved into the corresponding siRNA after two hours as judged by separation using native PAGE (data not shown). However, the labeled scRNAs disappeared after incubation in the cell lysate. After one to two hours of incubation in the cell lysate at 37°C, a band corresponding to the full-length ^{32}P 5′-end labeled scRNAs (both RNA and 2′-O-Me RNA strands) was, at best, faintly visible. A strong ^{32}P signal was found at the bottom of the gel in a band corresponding to approximately a single nucleotide. Initially, the scRNA disappearance in the cell lysate was believed to be due to RNA degradation, despite the addition of an RNase inhibitor. Further investigation, however, indicated that the scRNAs were not degraded, but rather the radioactive 5′ phosphate was removed in the cell lysate. Various phosphatase inhibitors were tested, but none prevented the removal of the 5′-phosphate. Interestingly, 5′-phosphate removal does not appear to occur for the DsiRNA. Perhaps the recessed 5′-end protects the phosphate from enzymatic removal.

An internal labeling strategy was devised to prevent removal of ^{32}P from scRNAs in the cell

lysate. Moving the ^{32}P from the 5′-end to an internal location in the backbone restored a band corresponding approximately to the full-length scRNA. This confirmed that scRNAs are not significantly degraded in the cell lysate. Internal labeling is achieved by repairing a nicked double-stranded scRNA through ligation. A hairpin is the easiest scRNA to internally label, but with more effort, ^{32}P can be incorporated into the backbone of any scRNA. Briefly, an scRNA is internally labeled through a two-step process. First, the scRNA is divided into two fragments. These fragments are designed to base pair to form a double-stranded structure with a nick. The short fragment is 5′-end labeled with ^{32}P. Then the two RNA fragments are ligated using T4 RNA ligase 2 to create the complete scRNA. The nick point is positioned within the predicted siRNA to ensure that the Dicer cleavage product is labeled with ^{32}P. With this protocol change, we could then proceed to investigate conditional RNAi mechanisms in the cell lysate system.

Through depletion of Dicer, we demonstrate that Dicer is responsible for cleaving Dicer substrates into siRNAs. Dicer substrates are not cleaved when Dicer is removed from the lysate through RNAi knockdown of Dicer in cell culture, followed by immunodepletion after cell lysis. In addition to demonstrating Dicer-mediated cleavage, depleting Dicer from the cell lysate allows the scRNA signal transduction step to be separated from the Dicer cleavage step. Dicer depletion is especially valuable when troubleshooting the conditional RNAi mechanism; the depleted Dicer lanes are similar to the '−' Dicer lanes in the recombinant Dicer processing stepping gels from Chapter 2.

Supplementing the cell lysate with ATP was not necessary for Dicer cleavage. In fact, attempts to deplete ATP from the lysate, in hopes of preventing Dicer cleavage, proved unsuccessful. Other reports confirm our finding that Dicer cleavage does not require ATP *in vitro*.[11,12] However, there have been reports that the Dicer complex, as opposed to purified Dicer, is significantly enhanced by the addition of ATP.[13] Previous studies also suggest that ATP is required for Dicer's helicase activity, which is not necessary for cleaving dsRNA with 3′ overhangs, and that ATP stimulates the release of the substrate.[5]

3.2.2 Mechanism 1: Conditional DsiRNA formation using stable scRNAs

Initial cell lysate tests of Mechanism 1 (M1), described in detail in Chapter 2, showed very little production of the final Dicer substrate B·C in response to addition of X_s (data not shown). The robust OFF to ON switching behavior triggered by X_s or X, as observed in Chapter 2, appears to be lost in the cell lysate tests. Likely, the reduced scRNA concentrations (500 nM in the stepping

gels shown in Chapter 2 versus 5–25 nM in the cell lysate) is partially responsible for the lower yields of B·C.

Mechanism 1 behaves as two bimolecular reactions when the components are present in approximately equimolar concentrations.

$$X + A \cdot B \rightleftharpoons X \cdot A + B \qquad \text{(step 1)}$$

$$B + C \rightleftharpoons B \cdot C \qquad \text{(step 2)}$$

$$X + A \cdot B + C \rightleftharpoons X \cdot A + B \cdot C \qquad \text{(full reaction)}$$

As the scRNA and detection target concentrations are reduced, the probability of two reactants colliding and interacting decreases. The forward rate of production for step 1 is described by $k_{f1}[X][A \cdot B]$. Hence, reducing the concentration of X and/or A·B will reduce the rate of production of the intermediates X·A and B. Similarly, the forward rate of production for step 2 is described by $k_{f2}[B][C]$, thus the rate of production of product B·C will decrease if the concentrations of B (dependent through step 1 on the concentration of X and A·B) and/or C are decreased. If product formation is limited by kinetics, increasing the incubation time should improve B·C production.

If equilibrium is achieved during the two-hour incubation, the distribution of the reactants and products can be described by the reaction equilibrium constant K_{rxn}. K_{rxn} is related to the ΔG_{rxn} and concentrations by:

$$K_{rxn} = K_{step1} \cdot K_{step2} = e^{-\frac{\Delta G_{rxn}}{RT}} = \frac{[X \cdot A][B \cdot C]}{[X][A \cdot B][C]}$$

assuming a dilute solution and

$$\Delta G_{rxn} = \Delta G_{step1} + \Delta G_{step2} \qquad K_{step1} = e^{-\frac{\Delta G_{step1}}{RT}} = \frac{k_{f1}}{k_{r1}} \qquad K_{step2} = e^{-\frac{\Delta G_{step2}}{RT}} = \frac{k_{f2}}{k_{r2}}$$

For example, at equilibrium, a 90% yield of B·C at 500 nM will be reduced to a 63% yield at 5 nM. Regardless of the cause (kinetic or thermodynamic) of the reduced B·C yield at lower reactant concentrations, the mechanism needs to be redesigned to operate optimally in this lower concentration regime. Making ΔG_{rxn} more favorable will shift the equilibrium distribution towards the products and achieve faster kinetics.

NUPACK was used to calculated the ΔG_{rxn} for various versions of mechanism 1 (Table 3.1). For version 1 of mechanism 1 (the version shown in Chapter 2), NUPACK predicts complete conversion

Mechanism 1 version	ΔG_{step1} (kcal/mol)	ΔG_{step2} (kcal/mol)	ΔG_{rxn} (kcal/mol)	$\Delta G_{leakage}$ (kcal/mol)
V1: $\|b\| = 4, \|c\| = 8, \|x\| = 12, \|y\| = 4$	−0.69	−19.94	−20.63	7.55
V2: $\|b\| = 7, \|c\| = 14, \|x\| = 14, \|y\| = 2$	−13.84	−19.91	−33.75	8.63
V3: $\|b\| = 7, \|c\| = 19, \|x\| = 14, \|y\| = 2$	−21.38	−19.91	−41.29	8.21

Table 3.1: Calculated ΔG values for variations of M1 at 37°C using RNA nearest-neighbor energy parameters.[14] NUPACK was used to calculate the ΔG of step 1 (X_s + A·B ⇌ X_s·A + B), step 2 (B + C ⇌ B·C), and the full reaction (X_s + A·B + C ⇌ X_s·A + B·C) for each mechanism version. In addition, the ΔG of the leakage reaction (A·B + C ⇌ A + B·C) was calculated. ΔG_{step2} and ΔG_{rxn} were adjusted to account for the unequal number of reactants and products. When calculated with the values given above, K_{step2} and K_{rxn} will have the dimensions of L/mol. K_{step1} and $K_{leakage}$ are dimensionless. NUPACK calculations do not account for the 2′-O-Me RNA modifications used experimentally. The domains not listed in the table are constant for all three mechanism versions (i.e., $|a| = 6, |s| = 5, |w| = 2, |z| = 3$).

of reactants to products at the concentrations tested, but experimental results show minimal B·C production (data not shown). This disagreement can be attributed to multiple factors. First, the cell lysate is a complex environment compared to conditions used to determine the energy parameters used in NUPACK. Second, the scRNA molecules are partially modified with 2′-O-Me RNA. RNA and 2′-O-Me RNA have different energy parameters, but unfortunately the 2′-O-Me energy parameters have not been fully determined experimentally[15,16] and so are not used in NUPACK.

Despite the disagreement between NUPACK predictions and experimental results, NUPACK calculations can still provide useful insight. As discussed earlier, improving the thermodynamic driving force of the overall reaction should increase the production of B·C. NUPACK analysis shows that the first step in this two-step mechanism is especially weak. As shown in Table 3.1, the ΔG of the first step in the initial version (V1) of M1 is only −0.69 kcal/mol, while the second step is significantly stronger with a ΔG of −19.94 kcal/mol.

In an attempt to improve B·C production, the first step in M1 was made stronger in version 2 (V2). The mechanism schematic is shown in Figure 3.1(a). The gold dot indicates the location of ^{32}P in the backbone of C. The length of the toehold that initiates base pairing between A and the detection target X (domain '*c*') was increased from 8-nt to 14-nt and the lengths of other dimensions were altered. These changes shift step 1 towards the products (ΔG_{step1}: −0.69 kcal/mol (V1) → −13.84 kcal/mol (V2)) without significantly changing the leakage reaction (as judged by comparable $\Delta G_{leakage}$ of 7.55 kcal/mol (V1) and 8.63 kcal/mol (V2)).

Lysate experiments with this new set of sequences (M1 V2) show significantly more B·C pro-

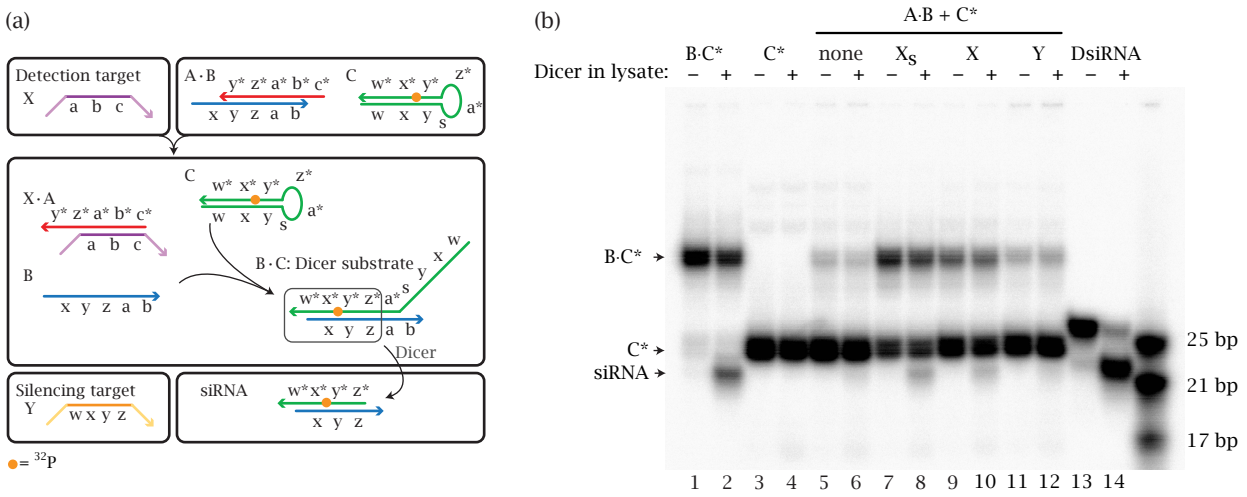


Figure 3.1: Mechanism 1, version 2, in cell lysate. **(a)** Schematic illustrating the mechanism (V2 domain lengths: $|a| = 6, |b| = 7, |c| = 14, |s| = 5, |w| = 2, |x| = 14, |y| = 2, |z| = 3$). Internal labeling with ^{32}P is denoted by a gold dot. **(b)** Each step of the mechanism was investigated in wild-type cell lysate ('+' lanes) or in cell lysate where Dicer was selectively depleted ('−' lanes). Final Dicer substrate B·C* is partially cleaved by Dicer to produce an siRNA (lane 2), while the standard DsiRNA* is nearly completely cleaved (lane 14). Initial scRNA reactant C* is not cleaved. As desired, significantly more B·C* is produced for the ON states with X_s (lane 7) and X (lane 9) than for the OFF states: no target (lane 3) or silencing target mRNA Y (lane 12). However, only a small amount of B·C* is cleaved by Dicer to produce an siRNA (lanes 8 and 10). The reactions were carried out at 37°C for 2 hours with 5 nM scRNAs and 1.25 μg/μL cell lysate. '*' denotes the strand labeled with ^{32}P.

duction at low concentrations with both X_s and X (Figure 3.1(b), lanes 7 and 9). However, the subsequence cleavage of B·C into an siRNA is still poor (lanes 8 and 10). C alone is not cleaved by Dicer (lane 4), and there is minimal formation of B·C in the absence of a target or in the presence of target mRNA Y.

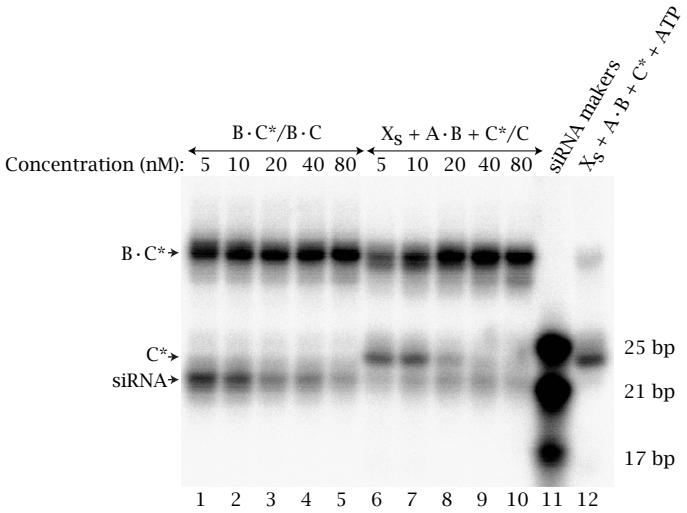


Figure 3.2: Concentration dependence of final product B·C cleavage and production for Mechanism 1, version 2. The percentage of B·C* cleaved to produce an siRNA decreased with increasing concentrations. Conversely, the percentage of C* found in the final product B·C* and the final siRNA increased with increasing concentration. Wild-type cell lysate was used for all lanes ('+' Dicer). All components were present at equimolar concentrations. The concentration of radioactively labeled C* was kept constant (5 nM) for all lanes, and conditions with higher scRNA concentrations were supplemented with unlabeled C. Lane 12 is identical to lane 6 (X_s + A·B + C*) except for the addition of 0.5 mM ATP. The reactions were carried out at 37°C for 2 hours with the scRNA concentration indicated and 1.25 μg/μL cell lysate. '*' denotes the strand labeled with ^{32}P.

The concentration dependence of M1 V2 was tested in cell lysate (Figure 3.2). For straightforward analysis, ^{32}P labeled C was kept constant (5 nM) for all lanes and non-radioactively labeled C was supplemented as needed to achieve the desired scRNA concentrations (5–80 nM). Thus, a fixed distribution of ^{32}P signal across lanes indicates concentration independence. Figure 3.2 shows this is not the case, and both conversion and Dicer cleavage are concentration dependent. As the concentration increases, the percentage of B·C that is cleaved by Dicer to produce an siRNA decreases (lanes 1–5). In contrast, B·C production from X_s + A·B + C increases as the concentration increases (lanes 6–10). This is expected based on the discussion earlier. Thus, for this set of sequences, it is impossible to find an optimal concentration for both strong B·C production and cleavage. Interestingly, the addition of ATP completely eliminates the conversion of hairpin C to the final duplex B·C when tested at 5 nM (lane 12). Perhaps the ATP supplementation activates

an alternate pathway, which either binds or degrades the scRNAs and prevents signal transduction.

The inverse correlation between concentration and Dicer cleavage of B·C in Figure 3.2, lanes 1–5, suggests that the RNAi pathway is saturated and decreasing the scRNA concentration will increase B·C cleavage. Pathway saturation can occur from Dicer binding of B·C, the scRNA reactants A·B and C, or intermediates. While hairpin C does not appear to be cleaved by Dicer, it could bind to the active site of Dicer and compete with B·C. Since neither A nor B are labeled with ^{32}P, it is not possible to investigate their fate in the cell lysate. To further compound the effect, B·C could be a poor Dicer substrate because it has non-canonical dimensions. The canonical DsiRNA is nearly completely cleaved at all concentrations tested. This is consistent with our findings in Chapter 2 that B·C requires more recombinant Dicer than a DsiRNA to be fully cleaved (1 U for B·C vs. 0.25 U for a DsiRNA).

M1 needs to be tested at concentrations below 5 nM to determine if it is possible to achieve full processing of B·C. Since M1 V2 does not fully convert at 5 nM, the mechanism needs to be redesigned. Better conversion at sub-nanomolar concentrations could be achieved by increasing the magnitude of ΔG_{rxn} through further lengthening the toehold '*c*' (refer to Table 3.1). M1 V3 has not yet been tested, but follow-up experiments should help determine if poor B·C production is due to insufficient energetic driving forces.

It is important to note that the reactions in M1 will show bimolecular reaction behavior only if all the components are present in similar concentrations. In a situation where X is limiting and is present at very low concentrations relative to A·B, the yield from the first step would depend only on the concentration of A·B. Depending on delivery efficiency, this could be the situation observed *in vivo*.

3.2.3 Mechanism 2: Conditional shRNA formation using a single stable scRNA

The conditional shRNA formation mechanism (M2), detailed in Chapter 2, was also tested in the cell lysate system. The final product, B, is a canonical shRNA, and a large fraction of B is cleaved by Dicer into an siRNA (Figure 3.3, lane 2). Interestingly, this particular shRNA does not silence Y (d2EGFP) when transfected into HEK 293 d2EGFP cells, despite producing an siRNA in a cell lysate. Likely the mRNA target region is inaccessible, preventing RISC cleavage of the mRNA. Alternatively, the sense strand, instead of the antisense strand, may be loaded into RISC to serve as the guide strand.

In lysate, the overall mechanism behaves as expected using a short detection target X_s (Figure

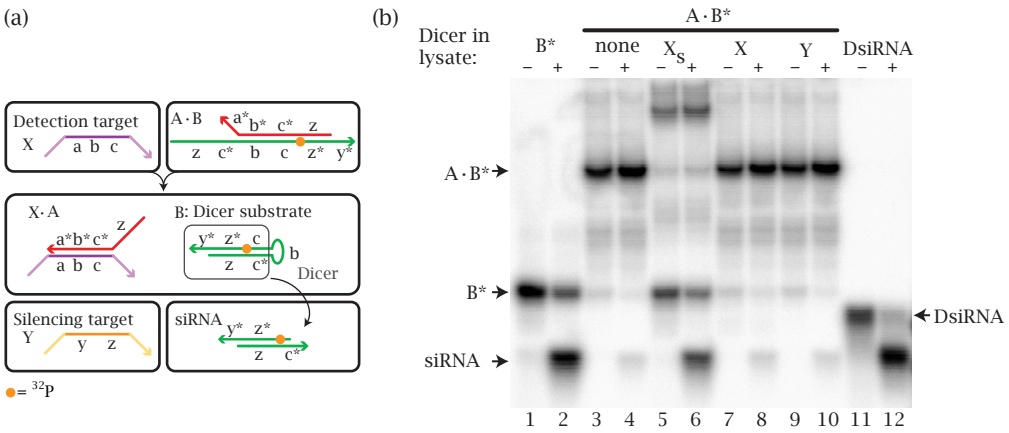


Figure 3.3: Mechanism 2 in cell lysate. **(a)** Schematic illustrating the mechanism (domain lengths: $|a| = 12, |b| = 14, |c| = 3, |y| = 2, |z| = 19$). Internal labeling with ^{32}P is denoted by a gold dot. **(b)** Each step of the mechanism was investigated in wild-type cell lysate ('+' lanes) or in cell lysate where Dicer was selectively depleted ('−' lanes). Final Dicer substrate shRNA B* and the DsiRNA are cleaved by Dicer to produce an siRNA (lanes 2 and 12, respectively). Initial scRNA reactant A·B* is not cleaved. As desired, significantly more B* is produced for the ON state with X_s (lane 5) than for the OFF states: no target (lane 3) and silencing target mRNA Y (lane 9). B* is cleaved by Dicer to produce an siRNA (lane 6). Full-length mRNA X does not produce B* (lane 7) and appears similar to the OFF states (lanes 3 and 9). The reactions were carried out at 37°C for 2 hours with 2.5 nM scRNAs and 1.25 μg/μL cell lysate. '*' denotes the strand labeled with ^{32}P.

3.3). Since signal transduction is achieved through a single scRNA with equal number of reactants and products, the equilibrium distribution is not concentration dependent. The forward rate of production, however, is still bimolecular and is dependent on the concentration of both X and A·B. Dicer cleavage of the full reaction (X_s + A·B → X_s·A + B) produces only half the amount of siRNA as cleavage of the final hairpin B (lane 6 vs. lane 2). This difference cannot be attributed to incorrect concentration measurements of B and A·B since the same amount of radioactive material was loaded in each lane. Radioactive signal corresponds to the amount of B present since it is the only molecule labeled with ^{32}P.

While A·B is completely consumed in the full reaction, a significant portion of the signal appears in a band which runs above A·B (lanes 5 and 6 in Figure 3.3). This band could be the trimer X_s·A·B or the homodimer B·B. Signal transduction proceeds through the trimer X_s·A·B intermediate. The trimer forms through base pairing between X_s and A, initiated by the toehold 'a/a^*.' Strand exchange through branch migration must occur to dissociate the trimer into B and X_s·A. If the higher band corresponds to the trimer X_s·A·B, it suggests that B is not released because of an insufficient driving force. Perhaps the short intramolecular toehold is not strong enough to close the hairpin loop. The contrast between the strong ON state with the short detection target X_s and

the non-existent ON state for the full-length mRNA X could indicate unintentional binding of the ssRNA regions in B ('z', 'c^*', and potentially 'b') with the mRNA, which prevents the release of B. Alternatively, the particular mRNA region targeted may not be accessible for scRNA binding in lysate.

Under cell lysate conditions, the ON and OFF states for the short detection target X_s are separated by only one order of magnitude. In comparison, the gels in Chapter 2 showed two orders of magnitude discrimination. This may be due to the reduced scRNA concentration, which shifts the equilibrium between the scRNA duplex and monomers towards the monomers (i.e., A·B $\rightleftharpoons$ A + B shifts towards the right). Because B is the final Dicer substrate, shifting the reaction towards the monomers leads to more leakage. To further reduce leakage, the strength of the A·B duplex could be increased relative to B. This can be achieved by increasing the size of the hairpin loop, since the base pairs in the stem of B are present in both A·B and B.

In order to improve conversion with full-length X mRNA, we tested helper strands. Helper strands are sequences complementary to the mRNA regions on either side of the scRNA binding location. The binding of helper strands reduce mRNA target secondary structure near the scRNA binding location. This will improve conversion if mRNA secondary structure makes the scRNA binding location inaccessible or if there are undesired interactions between the single-stranded regions of open B and the mRNA. NUPACK sequence design used short X_s, not the full-length mRNA X, during sequence optimization. Hence, the interactions between the single-stranded segments of the scRNAs and the mRNA sequence near the binding site were not considered in the NUPACK design. Neither 50-nt nor 25-nt DNA helper strands, positioned on both sides of the scRNA binding site, improved production of B when A·B was incubated with full-length mRNA X (data not shown). It is possible that the DNA was degraded in the cell lysate. Future experiments should explore the potential of 2′-O-Me RNA helper strands to improve B production.

3.2.4 Mechanism 3: Conditional DsiRNA formation via template-mediated 4-way branch migration

Mechanism 3: Conditional DsiRNA formation via template-mediated 4-way branch migration has not yet been tested in cell lysate, but it is expected to be sensitive to concentration changes. The mechanism requires the co-localization of two duplex scRNAs with a target, which becomes less energetically favorable as the concentrations are decreased. Entropic forces are more significant at lower concentrations and disassociation of the duplexes into their respective monomers becomes

more favorable (i.e., A·B $\rightarrow$ A + B and C·D $\rightarrow$ C + D). Since the duplex scRNAs are fully constrained by the detection and silencing mRNA targets, they are prone to disassociation at low concentrations because the monomers have significant internal secondary structure. The presence of monomers is expected to increase leakage, though the extent could vary based on the secondary structure of the monomers. While designing this mechanism, the objective function did not attempt to minimize the internal secondary structure of the monomers, which likely means the design is not optimal for low concentrations. In the future, the monomeric components of duplex scRNAs should be specified to be single stranded in the objective function.

3.3 Conclusions

We have developed a reliable tool to study conditional RNAi mechanisms *in vitro* under conditions similar to the cellular environment. The cell lysate system allows for step-wise interrogation of the conditional RNAi mechanisms and has led to important insights. It has shown the importance of designing mechanisms which function at the appropriate concentrations. Additionally, it appears possible that the scRNAs competitively inhibit Dicer cleavage. More work is needed to design variations of the mechanisms presented in Chapter 2 that function robustly, including detection of full-length mRNA targets, in the cell lysate system. Hopefully, these studies will bring us closer to *in vivo* conditional RNAi. In the future, a cell lysate system could replace test-tube studies for the initial testing of new scRNA logic gates.

Despite the obvious advantages of studying conditional RNAi mechanisms in cell lysate instead of the test tube, it is unknown if the results obtained from studying conditional RNAi mechanisms in a cell lysate system will transfer to *in vivo* systems. While many challenging variables have been removed, the cell lysate also introduces new variables. It is unknown if the cellular components, including salts and metal co-factors, are present at relevant concentrations. Non-biologically relevant salt and metal co-factor concentrations may deactivate important enzymatic pathways. Additionally, the homogenization of the cell compartments may lead to interactions, and potentially scRNA degradation, that would not be possible in a compartmentalized cell.

To produce a functional cell lysate, we added both a protease inhibitor cocktail and an RNase inhibitor. The addition of these may compensate for the disruption of the cell and homogenization of the cellular compartments. However, the RNase inhibitor could prevent degradation of the scRNAs, providing a false sense of faith in the integrity of scRNAs *in vivo*.

The current cell lysate system achieves robust Dicer cleavage. Further development is needed to recapitulate the entire RNAi pathway and complete our lysate system. To execute the full RNAi pathway, we need to add RISC functionality and cleave an mRNA in a sequence-specific manner. However, it may be sufficient to show Dicer cleavage of the final Dicer substrates because we can test our putative siRNAs in living cells (see Chapter 4). There still may be value in developing a cell lysate system that includes a functional RISC with Ago2 slicing activity. Such a system could be used to study other aspects of the RNAi pathway, such as determining if the unexpected silencing observed in Chapter 4 is mediated through Ago2.

3.4 Bibliography

[1] Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**(1), 25–33 (2000).

[2] Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., and Sharp, P. A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & Development* **13**(24), 3191–7 (1999).

[3] Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *The EMBO Journal* **20**(23), 6877–88 (2001).

[4] Wakiyama, M., Kaitsu, Y., and Yokoyama, S. Cell-free translation system from Drosophila S2 cells that recapitulates RNAi. *Biochemical and Biophysical Research Communications* **343**(4), 1067–71 (2006).

[5] Welker, N. C., Maity, T. S., Ye, X., Aruscavage, P. J., Krauchuk, A. A., Liu, Q., and Bass, B. L. Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Molecular Cell* **41**(5), 589–99 (2011).

[6] Elbashir, S. M., Lendeckel, W., and Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development* **15**(2), 188–200 (2001).

[7] Wakiyama, M., Takimoto, K., Ohara, O., and Yokoyama, S. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes & Development* **21**(15), 1857–62 (2007).

[8] Sakurai, K., Amarzguioui, M., Kim, D. H., Alluin, J., Heale, B., Song, M. S., Gatignol, A., Behlke, M. A., and Rossi, J. J. A role for human Dicer in pre-RISC loading of siRNAs. *Nucleic Acids Research* **39**(4), 1510–25 (2011).

[9] Burnett, J. C., Rossi, J. J., and Tiemann, K. Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnology Journal* **6**(9), 1130–46 (2011).

[10] Whitehead, K. A., Langer, R., and Anderson, D. G. Knocking down barriers: advances in siRNA delivery. *Nature Reviews Drug Discovery* **8**(2), 129–38 (2009).

[11] Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. Ribonuclease activity and RNA binding of recombinant human Dicer. *The EMBO Journal* **21**(21), 5864–74 (2002).

[12] Zhang, H., Kolb, F. A., Brondani, V., Billy, E., and Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *The EMBO Journal* **21**(21), 5875–85 (2002).

[13] Pellino, J. L., Jaskiewicz, L., Filipowicz, W., and Sontheimer, E. J. ATP modulates siRNA interactions with an endogenous human Dicer complex. *RNA* **11**(11), 1719–24 (2005).

[14] Serra, M. J. and Turner, D. H. Predicting thermodynamic properties of RNA. *Methods in Enzymology* **259**, 242–61 (1995).

[15] Kierzek, E., Ciesielska, A., Pasternak, K., Mathews, D. H., Turner, D. H., and Kierzek, R. The influence of locked nucleic acid residues on the thermodynamic properties of 2'-O-methyl RNA/RNA heteroduplexes. *Nucleic Acids Research* **33**(16), 5082–93 (2005).

[16] Kierzek, E., Mathews, D. H., Ciesielska, A., Turner, D. H., and Kierzek, R. Nearest neighbor parameters for Watson-Crick complementary heteroduplexes formed between 2'-O-methyl RNA and RNA oligonucleotides. *Nucleic Acids Research* **34**(13), 3609–14 (2006).

Chapter 4

Experiments towards conditional RNA interference in tissue culture

4.1 Introduction

To achieve the dream of a smart therapeutic via conditional RNAi logic, a conditional RNAi mechanism must function robustly inside a cell and interface with the RNAi pathway. This chapter presents work towards engineering scRNAs to perform logic in tissue culture. We devised a model tissue culture experimental system for initial studies towards engineering *in vivo* conditional RNAi mechanisms. These experiments allowed us to investigate variables introduced in the transition from the test-tube and cell-lysate environments to the cellular environment. Unfortunately, engineering nucleic acid pathways inside a cell has proven to be challenging and unsuccessful to date.

The experimental system devised to test our mechanisms consists of two related HEK 293 cell lines expressing one or two fluorescent proteins. One cell line expresses both DsRed2 and d2EGFP (X + Y); the other cell line expresses only d2EGFP (Y). As described in Chapter 2, our conditional RNAi mechanisms are designed to detect DsRed2 (X) and silence d2EGFP (Y). Thus, in the cell line expressing X + Y, the conditional RNAi mechanisms should detect X and silence Y through the production of a Dicer substrate specific to Y. In the cell line expressing only Y, there should be no silencing of Y because the mechanism should remain OFF in the absence of detection target X. A traditional RNAi substrate (an siRNA, DsiRNA or shRNA) targeting Y will silence Y in both cell lines. The d2EGFP mRNA has a destabilized tail to allow for faster protein turnover.[1] Thus, reduction of the d2EGFP fluorescent signal, monitored through flow cytometry or fluorescence microscopy, provides a readout of Y silencing. In future applications, the diagnosis target (Y) would likely be an endogenous gene. For all the subsequent graphs, expression level is reported

as the average across multiple wells of the mean fluorescent signal and the error bars show the standard deviation of the mean fluorescent signal.

4.1.1 Challenges of implementing nucleic acid hybridization pathways *in vivo*

There are many challenges to implementing nucleic acid hybridization pathways *in vivo*. We believe the most significant challenges are the following: (1) delivery, (2) branch migration, (3) RNA stability, (4) unintended enzymatic processing, and (5) protein binding. Likely, even more challenges await us. An advantage of designing a diverse set of conditional RNAi mechanisms in Chapter 2 is that we have multiple design features to study while exploring the cellular design space.

To date, all functional intracellular nucleic acid mechanisms which rely on secondary structure conformational changes have been expressed.[2–6] As far as we are aware, there have been no demonstrations of a functional nucleic acid logic gate with delivered gates. The conditional RNAi mechanisms demonstrated in Chapter 2 cannot be expressed due to the use of scRNA dimers. Alternate conditional RNAi mechanisms comprised entirely of hairpins are better candidates for expression. Regardless, we believe that conditional RNAi mechanisms hold more promise as a deliverable, rather than expressed, therapeutic. Transfection reagents have been optimized for the delivery of siRNAs and achieve high levels of gene knockdown in tissue culture. These transfection reagents will likely also deliver scRNAs, although it is unclear if they will be able to deliver scRNAs in sufficient concentrations to the proper intracellular location. Good delivery is especially critical for mechanisms with two or more scRNAs (M1 and M3). While the most common method of siRNA delivery in tissue culture is lipid transfection, other methods include electroporation, microinjection, and peptide delivery.

Molecular beacons offer many parallels to scRNAs and have been reported to be successfully delivered into cells.[7] Molecular beacons are oligonucleotide hairpins with a short stem and a large loop. Generally, molecular beacons are dual-labeled probes, with a fluorophore on one end and a quencher on the opposite end.[8] When the hairpin is closed, the fluorophore is in close proximity to the quencher and there is minimal fluorescent signal. When the loop region binds to its complementary sequence in the cell, the hairpin is torn apart (hence, the necessity for a short stem) and the fluorophore is located further from the quencher and is able to emit signal. Alternative approaches have used FRET in place of the fluorophore/quencher pair. Delivery of quenched molecular beacons demonstrates that the secondary structure of the hairpin was maintained during delivery.[9] However, it appears that the most successful delivery techniques are microporation or

attachment of a ligand to allow receptor-mediated delivery.[10–12] Additionally, the presence of a quencher/fluorophore or FRET pair could substantially alter the molecular properties, potentially impacting delivery and secondary structure stability.

There has been little study of toehold-mediated branch migration inside cells. 4-way branch migration occurs naturally in the resolution of Holliday junctions, indicating that branch migration is possible inside a cell. However, resolution of the 4-way junction is facilitated by ATP-fueled molecular motors.[13] For our purposes, it is highly important that enzyme-free, toehold-mediated branch migration is possible inside a cell, since our mechanisms cannot function without strand exchange. If branch migration is possible inside a cell, the optimal toehold length for nucleation will likely be different from the optimal length determined experimentally *in vitro*.[14]

The cellular environment could potentially sequester or alter the scRNAs through unintended protein binding and/or degradation. There are many RNA binding proteins which could potentially bind to scRNAs, making them unavailable to execute the logic pathway. If these enzymes also process or degrade the scRNAs, the pathway could be turned on accidentally (i.e., in the absence of X). Because conditional RNAi pathways are energetically tuned, removing even a few nucleotides from an scRNA through degradation or cleavage could alter the logic operation. Degradation is partially prevented through chemical modifications. Modifications are necessary to increase scRNA stability and prevent unintended silencing of the initial scRNA components. The current mechanisms are partially modified with 2′-O-Me to prevent unintended Dicer cleavage. Unfortunately, we were unable to do a systematic analysis of the effect of 2′-O-Me modifications. Likely, additional modifications will be required.

4.2 Results and discussion

4.2.1 scRNA delivery

The scRNA reactants were delivered to the HEK 293 cell lines primarily by commercial lipid-based transfection reagents. Lipofectamine RNAiMAX (Life Technologies, Inc.) consistently produced the greatest gene knockdown and was used for most transfections. RNAiMAX was designed specifically for siRNA delivery, and we believe that scRNAs are sufficiently similar in shape and size to canonical siRNAs that RNAiMAX can efficiently deliver our components. The alternative delivery methods of microinjection and Neon electroporation were explored briefly but did not produce significantly different results to warrant their continued use. Previous reports demonstrate strong

d2EGFP (Y) knockdown with a DsiRNA referred to here as DsiRNA 1.[3] DsiRNA 1 was used as a positive transfection and silencing control in most experiments.

Unlike traditional siRNA-mediated gene knockdown, our conditional RNAi mechanisms M1 and M3 require co-delivery of two scRNAs, which then must interact inside the cell after detecting the mRNA target X. (M2 consists of a single scRNA and does not have a co-delivery requirement.) We attempted to verify co-delivery by transfecting two DsiRNAs—DsiRNA 1 targeted to Y and a DsiRNA targeted to X (Figure 4.1). The DsiRNAs were transfected with RNAiMAX at 20 nM each and assayed by flow cytometry 48 hours post-transfection. Both mRNAs were silenced: ≥85% silencing was observed for Y and 55–60% silencing was observed for X. Therefore, about half the cells showed knockdown of both proteins. Measuring protein levels, instead of mRNA levels, through fluorescent signal underestimates X knockdown because of the relatively long half-life of X. (Y turnover is more rapid due to the addition of the destabilizing tail.)

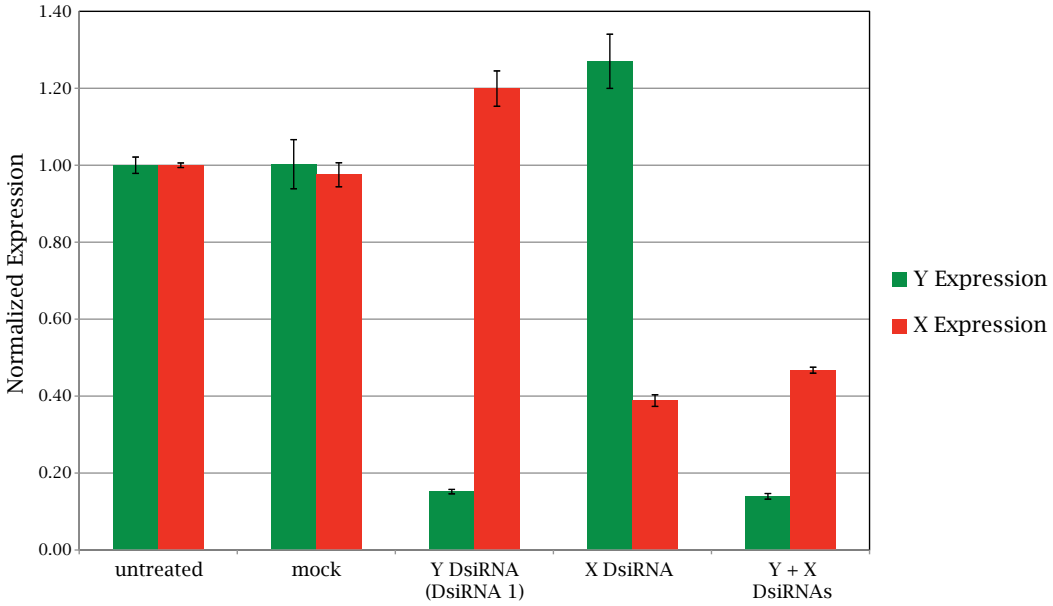


Figure 4.1: Co-transfection of two DsiRNAs, one (DsiRNA 1) specific to d2EGFP (Y) and the other specific to DsRed2 (X). Transfection of the Y-specific DsiRNA alone or in a split transfection with the X-specific DsiRNA leads to Y knockdown. Transfection of the X-specific DsiRNA alone or in a split transfection with the Y-specific DsiRNA leads to X knockdown. X knockdown is not as large as Y knockdown due to the lack of a destabilizing tail on X. Interestingly, for samples where only one fluorescent protein is silenced, expression of the fluorescent protein which is not targeted increases slightly. This may be due to the use of the CMV promoter to drive X and Y expression. The transfection was carried out in the X + Y cell line as described in Section B.8 of Appendix B using a split transfection and a final RNA concentration of 20 nM per strand. Gene knockdown was assayed 48 hours post-transfection. Error bars were calculated using four replicate wells.

To demonstrate the interaction of two nucleic acid strands inside a cell, DsiRNA 1 was separated

into its two ssRNA components and delivered through a split transfection (Figure 4.2). A split transfection is designed to prevent the strands from interacting and forming the duplex prior to entering the cell. (In a split transfection, the transfection complexes are generated in two separate reactions, each using half the standard volumes and containing either DsiRNA 1 PS or GS.) Initially, both transfection complexes were added simultaneously to the cell culture. The split transfection of DsiRNA 1 PS and GS with the HiPerFect transfection reagent showed no knockdown of Y. Interestingly, when RNAiMAX was used instead of HiPerFect for the split transfection of PS and GS, strong silencing of Y was observed, nearly equivalent to the silencing observed for the annealed duplex. Potentially, these results demonstrate that PS and GS interacted inside the cell to produce a functional DsiRNA, but only when transfected with RNAiMAX. Alternatively, the RNAiMAX transfection complexes may have fused outside of the cell, allowing the ssRNAs to form the duplex DsiRNA before endocytosis. This is more consistent with the fact that ssRNA is very unstable intracellularly. Further experiments showed that spacing the addition of the two RNAiMAX transfection complexes by 5.5 hours reduced the silencing to 25% knockdown of Y after 20 hours and 55% knockdown of Y after 46 hours (data not shown).

4.2.2 Non-canonical Dicer substrates

To ensure that our non-canonical Dicer substrates are functional and can silence Y, we compared the silencing ability of two non-canonical Dicer substrates to two canonical substrates. M1 and M3 signal transduction leads to a final product that is a non-canonical Dicer substrate. M2 produces a canonical shRNA. In Chapter 2, we demonstrated that recombinant Dicer cleaves the non-canonical final substrates to produce siRNAs. Additionally, Chapter 3 demonstrated that the non-canonical final substrates are cleaved in cell lysate to produce siRNAs. However, neither demonstrated that the siRNAs are functional and capable of silencing *in vivo*.

We verified that non-canonical Dicer substrates are able to selectively knock down gene expression in our experimental system. The silencing ability of various substrates was tested in the cell line expressing only Y. Figure 4.3(a) shows excellent silencing of Y for both the canonical and non-canonical substrates. The canonical substrates include a DsiRNA and an shRNA. DsiRNA 1 has the traditional DsiRNA dimensions: a 25-nt sense strand and a 27-nt antisense strand, leading to a 2-nt 3′-overhang. The shRNA, with a 23-bp stem, a 7-nt loop, and a 2-nt 3′-overhang, falls within the range of traditional shRNAs. The non-canonical substrates are the final products of M1 and M3. As seen in Figure 4.3(a), both silence Y very well and demonstrate that replacing the tradi-

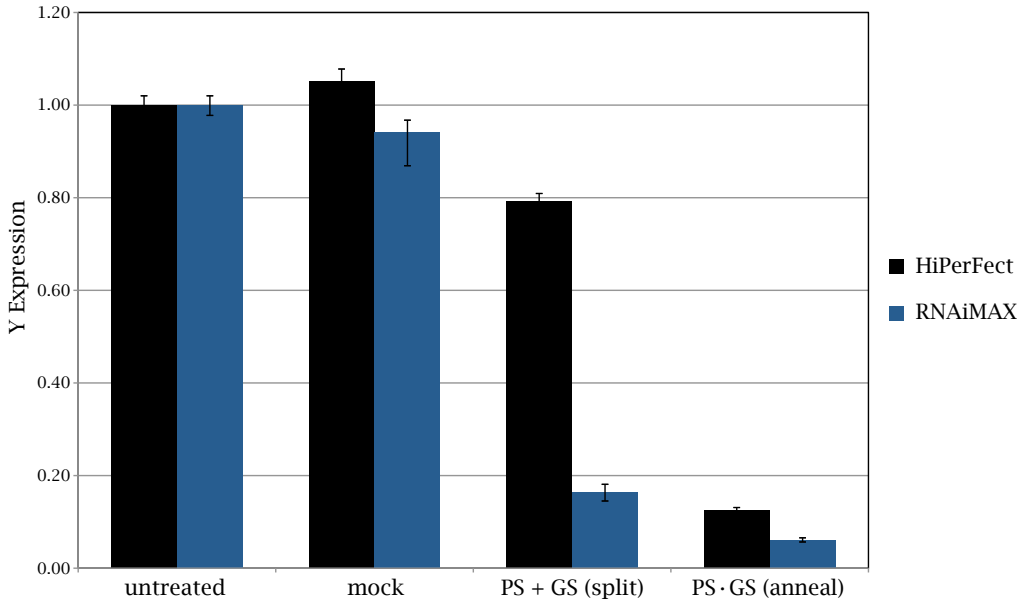


Figure 4.2: Co-transfection of ssRNA strands DsiRNA 1 PS and DsiRNA 1 GS. The two strands are complementary and should base pair to form DsiRNA 1, a potent silencer of d2EGFP (Y). When the ssRNA strands are divided, prepared in separate transfection complexes (split transfection), and added to the cells at the same time, silencing is observed for transfections prepared with the RNAiMAX transfection reagent but not the HiPerFect transfection reagent. When the duplex DsiRNA 1 is formed before transfection (PS·GS, anneal), silencing is observed for transfections with both transfection reagents. RNAiMAX transfections were carried out as described in Section B.8 of Appendix B with a final RNA concentration of 25 nM per strand. HiPerFect transfections were carried out according to instructions provided by the manufacturer (Qiagen). 3 μL HiPerFect per well was used to transfect RNA strands at a final concentration of 20 nM. Gene knockdown was assayed 22–24 hours post-transfection. Error bars were calculated using three replicate wells for HiPerFect and two replicate wells for RNAiMAX (four replicate wells for untreated).

tional blunt end of a DsiRNA with overhanging sequences does not interfere with Dicer processing. In fact, these non-canonical substrates function nearly as well as the canonical substrates.

The concentration-dependent silencing ability of a canonical substrate, DsiRNA 1, was compared to a non-canonical substrate, M3 B·D. Figure 4.3(b) illustrates the gene silencing induced by transfecting the substrates over a concentration range of three-orders of magnitude (0.2 nM to 200 nM). Both duplexes silence Y well, even below 1 nM, and produce nearly identical dose responses. DsiRNA 1 reaches maximum silencing at 10 nM, while M3 B·D reaches maximum silencing at 5 nM. RNAiMAX, the transfection reagent used for this experiment, recommends using 30 nM siRNA as a starting point for optimizing transfection (Life Technologies, Inc., refer to product data sheet). Since both substrates reach maximum silencing at concentrations lower than 30 nM, it is clear that these are both potent Dicer substrates and HEK 293 is an easily transfected cell line. These data show that certain non-canonical substrates function as well as canonical substrates.

A drawback, however, to Dicer's ability to cleave many non-canonical substrates is that certain scRNA reactants also silence Y. This was unexpected because the scRNAs either lack the 2-nt 3′-overhang or have a duplex region shorter than 19-bp. Both of these criteria were believed to be necessary for a structure to be processed by Dicer,[15,16] but apparently the rules are not very strict. Y knockdown does not necessarily imply that Dicer cleaved the scRNA. An alternative pathway may lead to the observed silencing or the secondary structure of the scRNA may be disrupted during delivery or intracellularly to produce a structure more amenable to Dicer cleavage.

Figure 4.4 compares the silencing of three hairpins that are expected to have identical secondary structures. Hairpin C1 is 100% RNA and was tested initially as scRNA C for M1. However, it became clear that in cells, the hairpin alone silences Y. The Y knockdown is less than DsiRNA 1 or the final product B·C of the mechanism (compare Y knockdown in Figures 4.3(a) and 4.4), but C1 still silences well and significantly compromises the OFF state of the mechanism. In an attempt to reduce this unintended silencing, we added 2′-O-Me modifications to the portion of the hairpin removed during Dicer cleavage (i.e., the 24-nt on the 5′-end; see C1m in Figure 4.4). While the 2′-O-Me modification reduced the silencing ability of the C1m hairpin, it was not sufficient to completely prevent hairpin C1m from silencing. Interestingly, changing the sequence and slightly altering the modification pattern abolished hairpin C silencing (Figure 4.4). The sequence of C was completely changed to target a different region of Y. Additionally, the region complementary to Y was reduced from 27-nt to 21-nt, and the region modified with 2′-O-Me was reduced from 24-nt to 21-nt. Likely, the sequence alterations, not the change in 2′-O-Me modification, is responsible for

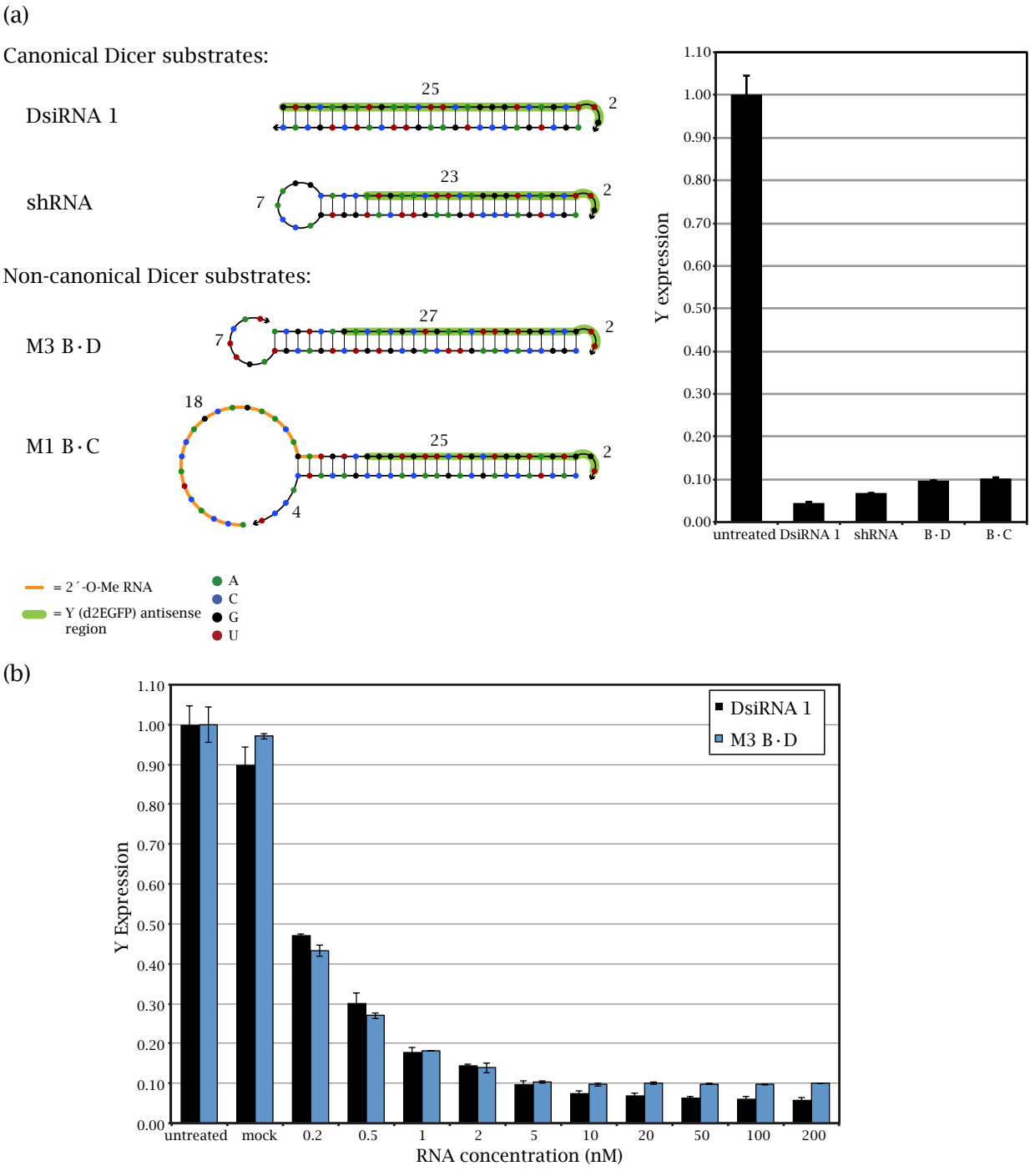


Figure 4.3: Both canonical and non-canonical Dicer substrates targeting Y achieve ≥90% gene knockdown. **(a)** Secondary structure representation of the four Dicer substrates and the corresponding reduction in Y expression following transfection at 100 nM (50 nM for M1 B·C) assayed 24–26 hours post-transfection. Green indicates the region complementary to Y, and an orange backbone indicates 2′-O-Me modification. Error bars were calculated using two replicate wells for untreated and B·D and three replicate wells for DsiRNA 1, shRNA, and B·C. **(b)** Dose response of Y expression to varying concentrations of either DsiRNA 1 or M3 B·D. Silencing was assayed 24 hours post-transfection. Both substrates show similar dose responses, but DsiRNA 1 achieves greater maximum silencing. Error bars were calculated using two replicate wells.

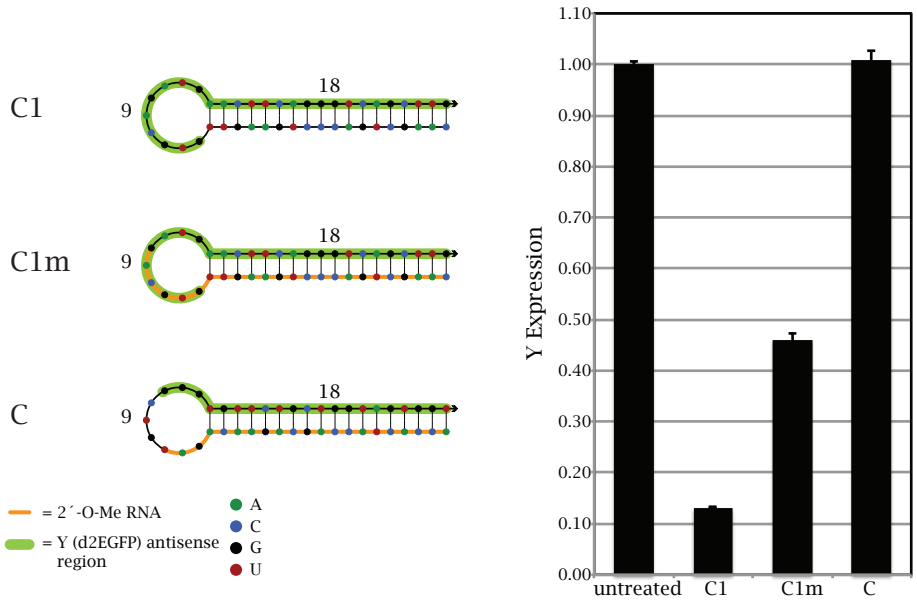


Figure 4.4: scRNA hairpin C shows various levels of Y expression knockdown with three sequence and 2′-O-Me variants. Secondary structure representation of the three hairpins and the corresponding reduction in Y expression following transfection at 100 nM for C1 and C1m and 50 nM for C assayed 24–26 hours post-transfection. Green indicates the region complementary to Y, and orange indicates 2′-O-Me modification. Error bars were calculated using two replicate wells for untreated, C1, and C1m and three replicate wells for C.

abolishing the silencing. A version of C1 with only 21-nt of Y complementarity still needs to be tested. All three hairpins functional equally well in the B·C duplex at the concentrations tested.

These examples highlight the possibility that many of our initial components are Dicer substrates, but only a few happen to have enough sequence complementary to Y to produce the observable readout of Y knockdown. Since the rules of Dicer cleavage are not as rigid as we initially believed, in the future, all scRNA reactants will need to be tested to confirm they are not capable of silencing.

4.2.3 scRNA saturation of the RNAi pathway

Designing the appropriate scRNA shape, either through dimensioning or chemical modifications, to prevent Dicer cleavage does not necessarily prevent scRNA binding to Dicer or RISC. scRNA binding to the active site of Dicer or RISC could saturate the RNAi pathway and/or competitively inhibit the final Dicer substrate silencing.[17] Saturation of the RNAi pathway has been observed *in vivo* and can lead to toxicity due to the competitive inhibition of endogenous miRNA.[18,19] While we observe that 2′-O-Me modifications prevent scRNA silencing, it is unknown if the modifications also

prevent scRNA binding in the active site of Dicer or RISC. Simultaneous transfection of a DsiRNA-like substrate, where the PS strand was completely modified with 2′-O-Me, with a DsiRNA reduced the observed gene knockdown. (Refer to Figure C.4 in Appendix C.) This implies that modified scRNAs compete with Dicer substrates for binding in the RNAi pathway. scRNAs bound in the active site of either Dicer or RISC present two challenges for our mechanisms: (1) sequestration of the scRNAs, making them unavailable for signal transductions, and (2) competitive inhibition of the final Dicer substrate processing. Additional experiments are required to fully investigate scRNA saturation of the RNAi pathway and competitive inhibition.

4.2.4 Testing conditional RNAi mechanisms in tissue culture

Based on the successful demonstration of three conditional Dicer substrate formation mechanisms in a test tube (Chapter 2), we proceeded to test the mechanisms in tissue culture. We hoped that mechanisms which met our design goals in a test tube would also function in our experimental tissue culture system. Unfortunately, it became clear that a functional test-tube mechanism does not guarantee a functional mechanism in a living cell. The cellular environment presents many additional challenges, and these must be addressed systematically before it will be possible to demonstrate successful conditional RNAi mechanisms in tissue culture. A brief summary of the experimental results obtained from testing our conditional RNAi mechanisms in tissue culture are presented here to help build the foundation for continued studies.

As described earlier, the mechanism OFF state was tested using a cell line expressing only the silencing target Y. The mechanism ON state was tested using a cell line expressing both the detection target X and the silencing target Y. A co-culture of the two cell lines allowed simultaneous testing of both the ON and OFF states (i.e., the two cell lines were mixed ∼50%/50% in the well before adding the transfection complexes). The cell lines were differentiated during flow-cytometry analysis based on the expression of X (DsRed2 fluorescence). Co-cultures guarantee that both cell populations are exposed to identical conditions. Knockdown resulting from individual scRNA reactants or uninitiated (i.e., leakage) formation of the final Dicer substrate will occur in both cell lines. Thus, unintentional silencing can be distinguished from silencing produced by a functional conditional RNAi mechanism.

Our experimental setup requires detection of full-length mRNA X intracellularly through base pairing with one or two scRNAs. In case this detection step was the reason the mechanisms were unable to function, we also explored transfecting the short detection target X_s. Unfortunately,

transfection of X_s in the same transfection complexes as the scRNAs leads to premature triggering of the reaction. To overcome this problem, we developed a transfection method termed "split transfections" (introduced earlier for co-delivery experiments). Multiple tubes of transfection complexes were prepared for each well so that X_s was isolated from the scRNAs. We generally added the split transfection complexes to the cells at the same time. To rule out transfection complex fusion, it is better to transfect X_s a few hours after scRNA transfection. However, most cells lines do not tolerate multiple transfections well. Compounding the problem is the fact that ssRNA (i.e., X_s) is degraded quickly inside cells. A 2′-O-Me-modified version of X_s could possibly provide more useful results.

4.2.5 Mechanism 1: Conditional DsiRNA formation using stable scRNAs

Mechanism 1, "Conditional DsiRNA formation using stable scRNAs," was tested in a co-culture of cells expressing Y or X + Y. Results are shown in Figure 4.5. The final Dicer substrate (annealed B·C) acts as a positive control and silences Y by nearly 80% in both cell lines at 50 nM. Neither of the scRNA reactants, A·B nor C, silence Y compared to untreated and mock transfected cells. The full reaction was tested by transfecting both A·B and C. If the conditional RNAi mechanism functioned as expected, Y should be silenced in the X + Y cell line, but not in the Y cell line. (Conditions for which silencing is expected are marked by a red 'x.') Unfortunately, Y expression remained identical between the two cell lines 22 hours post-transfection.

Interestingly, this mechanism does not appear to produce leakage. Test-tube studies in Chapter 2 produced ∼5–10% leakage, but there is no observed silencing due to leakage upon transfection of the full mechanism. A·B and C were prepared in separate transfection complexes to prevent any leakage from occurring before transfection. The lack of silencing due to leakage indicates either that the scRNAs did not get delivered to the cytoplasm, the scRNAs did not interact inside the cell, the OFF state is very strong intracellularly, or not enough leakage product was formed to produce detectable gene knockdown.

The scRNA C used for the mechanism shown in Figure 4.5 was modified with a 5′-inverted dideoxy-T (5′-InvddT). The addition of the 5′-InvddT was originally tested as a method to prevent C silencing. This concern was based on the fact that C1m silenced Y, and while C1m has a different sequence to that of C, it has identical dimensions and a nearly identical 2′-O-Me modification pattern (Figure 4.4). As shown in Figure 4.5, transfection of C alone does not silence Y; however, further investigation showed that C with an unmodified 5′-end also does not silence Y (C in Figure

4.4). Thus, the addition of a 5′-InvddT appears to be neutral with regard to Y silencing.

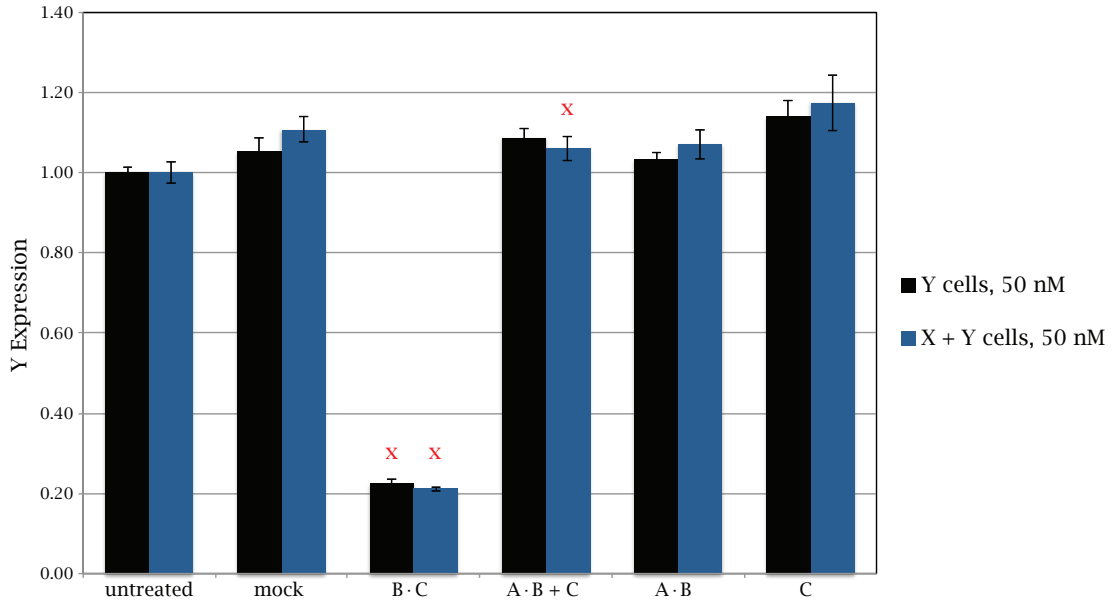


Figure 4.5: Transfection of M1 scRNAs into a co-culture of the Y and X + Y cell lines shows strong gene knockdown for the final Dicer substrate (B·C), but no knockdown is observed for the full mechanism (A·B + C). A red 'x' indicates a condition for which Y silencing was expected. A split transfection divided the two scRNA reactants into different transfection complexes to avoid interaction prior to transfection. The transfection was carried out as described in Section B.8 of Appendix B with a final RNA concentration of 50 nM per strand. Gene knockdown was assayed ∼22 hours post-transfection. Error bars were calculated using three replicate wells.

4.2.6 Mechanism 2: Conditional shRNA formation using a single stable scRNA

The sub-sequence of d2EGFP mRNA chosen as the silencing target for M2 was found to be a poor selection. The DsiRNA targeting this region silences d2EGFP poorly (∼20% Y knockdown). This poor silencing may result from an inaccessible mRNA target region or a poorly designed siRNA. For this DsiRNA, the predicted siRNA has energetic asymmetry between the siRNA ends that would likely favor RISC loading of the sense strand, not the antisense strand as desired.[20] Regardless of the reason for the poor silencing, the full mechanism cannot be properly tested if the final Dicer substrate does not silence. Once the mechanism is redesigned such that the corresponding DsiRNA is functional, we expect that B, as a canonical shRNA, will silence Y. We do not know whether the mechanism will function inside a cell. We will first demonstrate detection of a full-length mRNA target in lysate before attempting further cell studies. Because this mechanism uses only one scRNA, it has a significant advantage over the other mechanisms. Refer to Supplementary Figure C.5 in Appendix C for the silencing data.

4.2.7 Mechanism 3: Conditional DsiRNA formation via template-mediated 4-way branch migration

Mechanism 3, "Conditional DsiRNA formation via template-mediated 4-way branch migration," was tested in a co-culture of the Y and X + Y cell lines (Figure 4.6(a)). Two scRNA concentrations (20 nM and 100 nM) were tested. The final Dicer substrate, B·D, served as a positive control. In all conditions, Y knockdown with B·D was greater than 80%. The duplexes were transfected individually to verify that the scRNA reactants do not silence significantly. The duplexes were transfected at twice the concentration of A·B + C·D to keep the total amount of RNA per well constant. A·B alone silences approximately 20%. This is unexpected since A is modified with 2′-O-Me (all but one base), which is believed to prevent Dicer cleavage. Based on this result and others, it appears that while 2′-O-Me modification greatly reduces scRNA silencing, it does not entirely prevent it. It may merely slow down the Dicer cleavage rate, since the knockdown from 2′-O-Me-modified scRNAs increases over time. C·D has less unintended silencing than A·B. The error bars shown are the standard deviation between the means of replicate wells. However, the true variation in the experiment is evident by comparing Y expression in the mock transfected cells. The mock transfection of X + Y cells at 20 nM shows ∼20% silencing when normalized by the corresponding untreated cells. For the other three conditions, Y expression in the untreated and mock transfected cells is approximately the same.

When the full mechanism, A·B + C·D, is tested, we hope to see silencing in the cells containing the detection target X (marked by a red 'x'). The duplexes were prepared as two different transfection complexes in an attempt to avoid interaction outside of the cell. Unfortunately, there is no silencing observed above that induced by A·B alone. This result was confirmed multiple times while varying parameters such as concentration, transfection reagent, and incubation time. Changing these variables did not find a set of conditions where the full reaction was even marginally functional. The toeholds on A and C, which nucleate with X, were increased to 25-nt each with the hope that this would improve conversion. The silencing results from this mechanism (data not shown) gave qualitatively the same results as with the shorter toeholds.

After failing to see any evidence that M3 functioned in living cells, we designed a non-conditional mechanism (Figure 4.6(b)). This mechanism does not require detection of a target to initiate the 4-way branch migration and should convert spontaneously. The duplexes exchange strands as follows: J·K + L·M → J·L + K·M. The strands J and L are modified with 2′-O-Me to prevent the duplexes

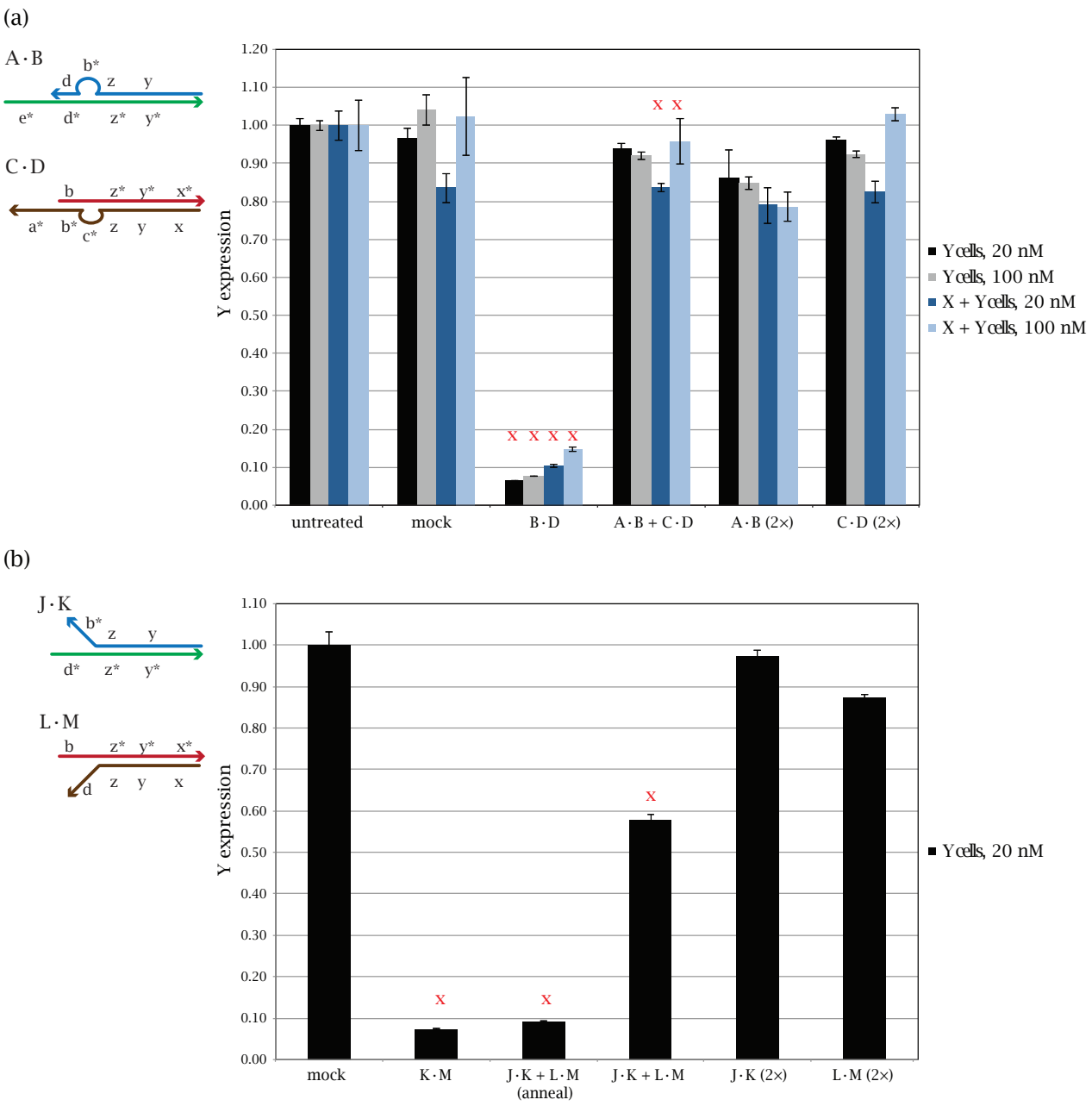


Figure 4.6: Mechanism 3 cell study results. **(a)** Transfection of M3 scRNAs into a co-culture of the Y and X + Y cell lines shows strong gene knockdown with the final Dicer substrate (B·D), but no knockdown is observed for the full mechanism (A·B + C·D) above that observed for the individual components. Error bars were calculated using four replicate wells at 20 nM and three replicate wells at 100 nM. **(b)** Transfection of a non-conditional mechanism requires duplex strand exchange through 4-way branch migration to produce the Dicer substrate K·M. Annealing the full mechanism leads to >90% Y knockdown, while only ∼40% Y knockdown is observed when the duplexes are prepared in separate transfection complexes. Error bars were calculated using two replicate wells. **(a, b)** A split transfection was used to divide the two duplexes into different transfection complexes to avoid interaction prior to transfection. The transfection was carried out as described in Section B.8 of Appendix B with a final RNA concentration of 20 nM and 100 nM per strand (a) and 20 nM per stand (b). Gene knockdown was assayed 24–26 hours post-transfection.

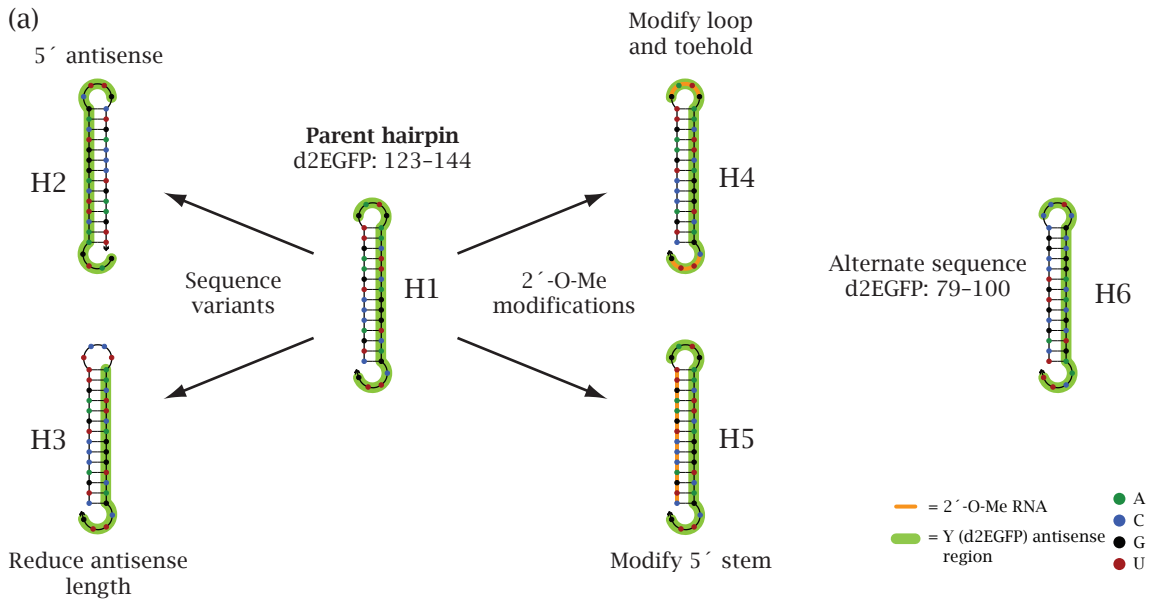
J·K and L·M from silencing. In this non-conditional design, J and L have complementary toeholds in addition to the complementary toeholds on K and M (Figure 4.6(b)). Neither set of toeholds are sequestered. When the two duplexes are mixed, they should immediately interact and exchange strands to produce the final products J·L and the Dicer substrate K·M. The duplexes were annealed separately and prepared in different transfection complexes. The final product K·M and an anneal of J·K + L·M were included as positive controls. Both show excellent knockdown of Y (greater than 90% knockdown, Figure 4.6(b)). When the full system was tested, ∼40% knockdown was observed. The individual components silence less than 15%. It is unknown if the duplexes interacted inside the cells or in the surrounding medium (through fusion of the transfection complexes). However, this provides a best case estimate of the range of this mechanism—approximately 30% difference between the ON and OFF states. While this dynamic range is much lower than desired, this may actually be a very encouraging result. It could be evidence that toehold-mediated 4-way branch migration of transfected RNAs is occurring inside the cell; however, interactions in the medium must be ruled out before definitively stating that branch migration occurs inside the cell.

4.2.8 Short hairpin silencing

The unintended silencing of various scRNA reactants led us to investigate the gene knockdown induced by hairpins of different sequences and structures. To our surprise, the best gene knockdown was produced by a very small hairpin (Figure 4.7). Knockdown of Y with this hairpin rivals that achieved with our best DsiRNA. The 36-nt hairpin has a 4-nt toehold, a 14-bp stem, and 4-nt loop (4–14–4 hairpin) and is significantly smaller than canonical shRNAs. The silencing of other non-canonical short hairpins with various 2′-O-Me modifications is further investigated in Supplementary Figure C.6. Other groups have also confirmed that short shRNAs are capable of silencing inside a cell.[21–23]

Additionally, the hairpin has a 4-nt toehold instead of the standard 2-nt toehold. Vermeulen et al. found that for toeholds equal to or greater than 4-nt, the 3′-end was not used to determine the Dicer cleavage position.[24] They suggest, instead, that the single-stranded to double-stranded junction positions Dicer using the three innermost nucleotides of the 3′-overhang.

We investigated the 4–14–4 hairpin further in an attempt to understand the mechanism of silencing. The 4–14–4 hairpin structure was found to be extremely accommodating to change. The variations tested and their relationship relative to H1 is shown in Figure 4.7(a). The parent hairpin H1 targets the 123–144 region of Y, matching the region targeted by DsiRNA 1. H2 and H3 are



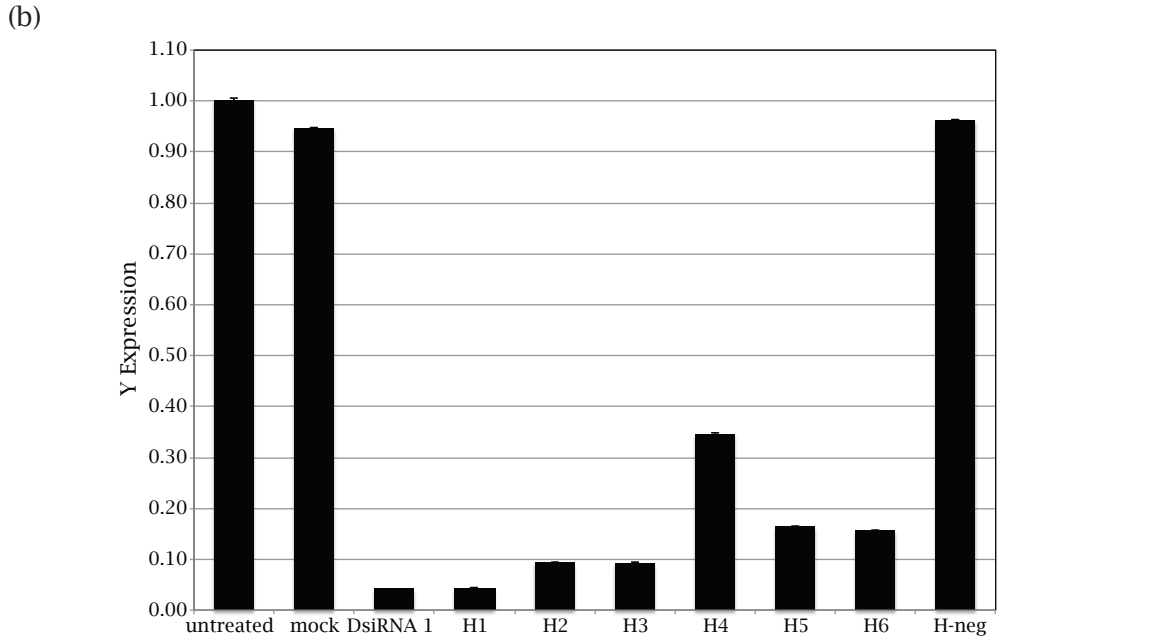


Figure 4.7: 4–14–4 (4-nt toehold–14-bp stem–4-nt loop) RNA hairpins induce strong Y silencing. **(a)** Hairpin structures depicting the parent hairpin and the related sequence and 2′-O-Me variants that were tested to help elucidate the silencing mechanism. Green highlighting indicates the sequence complementary to Y, and an orange backbone indicates 2′-O-Me modification. **(b)** Y knockdown for the hairpins shown in (a) transfected at 100 nM as described in Section B.8 of Appendix B and assayed by flow cytometry 24 hours post-transfection. DsiRNA 1 serves as a positive silencing control. H-neg is a 4–14–4 hairpin complementary to β-actin and serves as a negative control. Error bars were calculated using two replicate wells.

sequence variants of H1. H4 and H5 are 2′-O-Me-modified variants of H1. H6 has the same design as H1 but targets a different region of Y (79–100). Figure 4.7(b) shows the relative silencing ability of each hairpin when transfected at 100 nM using the RNAiMAX transfection reagent and assayed after 24 hours. H-neg is a hairpin designed to target β-actin (ACTB). H-neg may silence β-actin, but it should not reduce Y expression.

For hairpin H2, the antisense region is shifted from the 3′-end in H1 to the 5′-end in H2. This leads to a 4-nt 5′-toehold instead of a 4-nt 3′-toehold. The relationship between H1 and H2 can be visualized by realizing that they form the same circular sequence. The only difference between the two hairpins is the location of the strand break. Both H1 and H2 silence very well and achieve greater than 90% Y knockdown (Figure 4.7(b)). It is surprising that neither the polarity of the toehold nor the location of the antisense sequence appears to significantly influence silencing. Perhaps this indicates multiple processing pathways. Alternatively, the 4-nt loop could be functionally equivalent to the 4-nt toehold. If a 4-nt loop is indistinguishable from a 4-nt toehold, the hairpins appear identical, and it is no longer surprising that the hairpins induce similar levels of gene knockdown.

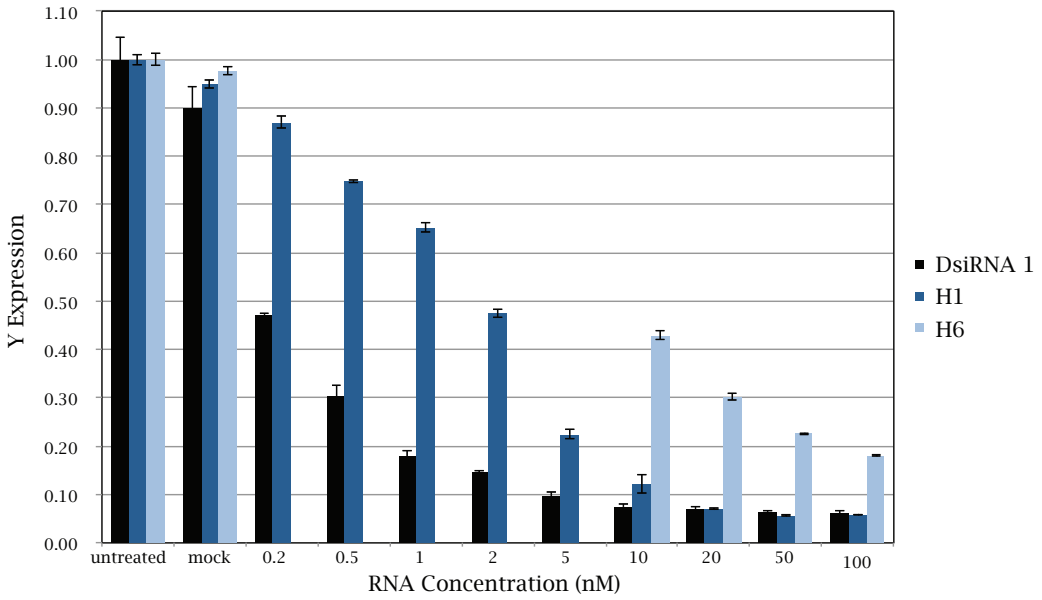


Figure 4.8: Dose response of Y expression to varying concentrations of either DsiRNA 1 or 4–14–4 hairpins H1 or H6. The transfection was carried out as described in Section B.8 of Appendix B, and silencing was assayed 24–25 hours post-transfection. DsiRNA 1 is significantly more potent at lower concentrations than H1 or H6. At 20 nM and above, DsiRNA 1 and H1 show similar levels of gene knockdown. Error bars were calculated using two replicate wells.

For hairpin H3, the nucleotides in the loop of H1 are replaced with a random sequence. This

reduces the antisense region from 22-nt to 18-nt. This change does not significantly reduce the silencing ability of H3 compared to H1. Hairpin H4 has the same sequence as H1, but the 4-nt toehold and the 4-nt loop are 2′-O-Me-modified RNA. This does not entirely prevent the hairpin from silencing, but it does reduce the silencing of H4 to 65%. This implies that the mechanism of hairpin silencing is partially disrupted when 2′-O-Me modifications are placed at the toehold and loop. For hairpin H5, modifying the 5′-stem, which contains a portion of the sense strand, with 2′-O-Me-modified RNA reduces the silencing slightly compared to H1. It is possible that the reduced silencing ability of hairpins containing 2′-O-Me-modified RNA is due to reduced delivery.

Hairpin H6 demonstrates that 4–14–4 hairpin silencing is not a property of a particular sequence. Figure 4.8 compares the silencing induced by DsiRNA 1, H1, and H6 at various RNA concentrations. While H1 silences as well as DsiRNA 1 at concentrations $\geq$ 20 nM, H1 is a less potent silencer than DsiRNA 1 at lower concentrations. It makes sense that H1 and DsiRNA 1 behave similarly since they both target the same region in Y. H6 is able to achieve greater than 80% gene knockdown at 100 nM, but its silencing ability is reduced at lower concentrations. We observed that DsiRNA 1, as well as other DsiRNAs, exhibit toxicity when transfected at 10 nM and kill a significant fraction of the cells after three days (refer to Figure C.7 in Appendix C). The cause of the toxicity is unknown. The RNAi pathway could be saturated or the DsiRNAs may have off-target effects which lead to toxicity.[18] Interestingly, the 4–14–4 hairpins H1 and H6 did not show toxicity after three days.

Figure 4.9 demonstrates that the observed short hairpin silencing is independent of the transfection reagent. The hairpins C1 and H1 induce strong gene knockdown when transfected with either the RNAiMAX or the HiPerFect transfection reagents. This eliminates the possibility that the observed silencing is an artifact of a particular transfection reagent. However, both RNAiMAX and HiPerFect are lipid-based transfection reagents and deliver RNA using similar cellular pathways. As observed previously, RNAiMAX produces greater Y knockdown. Since the hairpin stems are only 14-bp, it is possible that the transfection reagents disrupt the hairpin secondary structure, and this disruption leads to silencing. However, it is unlikely that the silencing is due to antisense since RNase-H independent antisense methods only function well when targeted to either the 5′-UTR or splice sites.[25] The hairpin silencing still needs to be confirmed for an alternate method of delivery to verify that it is not a function of lipid-based transfection. If the silencing proves to be dependent on the transfection method, it will be an important clue for understanding the mechanism by which these hairpins silence.

Recombinant Dicer does not cleave the 4–14–4 hairpins, as seen in Figure 4.10. DsiRNA 1 and

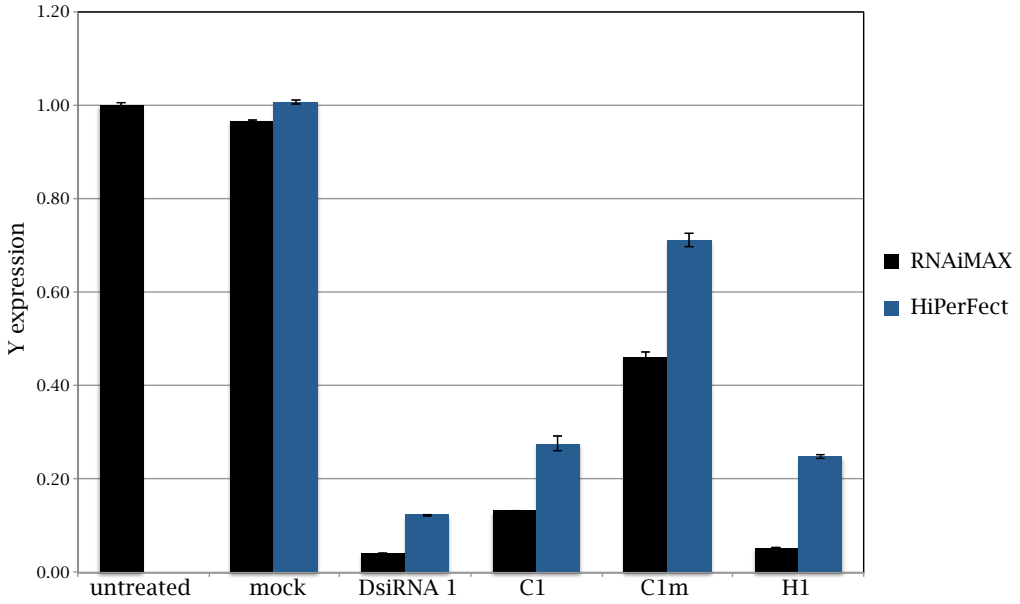


Figure 4.9: Y is silenced by various short hairpins transfected with either RNAiMAX or HiPerFect. RNAiMAX uniformly produces greater gene knockdown. The relative Y expression is shown for transfections carried out as described in Section B.8 of Appendix B with a final RNA concentration of 100 nM and assayed by flow cytometry 26 hours post-transfection. DsiRNA 1 serves as a positive silencing control. Error bars were calculated using two replicate wells.

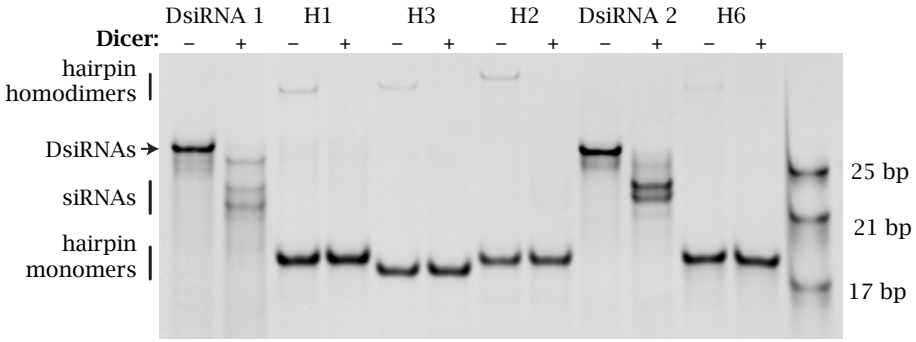


Figure 4.10: Recombinant Dicer cleaves 4–14–4 hairpin homodimers (faint bands near the top of the gel) but not the hairpin monomers. DsiRNA 1, DsiRNA 2, and the 4–14–4 hairpins H1, H3, H2, and H6 were incubated with 0.25 U Recombinant Turbo Human Dicer for 2 hours at 37°C and separated by native PAGE. (−/+) indicates absence/presence of Dicer.

DsiRNA 2 were included as positive controls and both appear to be cleaved to standard siRNAs. DsiRNA 1 targets the same region in Y as H1, H2, and H3. DsiRNA 2 targets the same region in Y as H6. While the hairpin monomers are not cleaved by Dicer, the homodimers, seen as the faint bands near the top of the gel, disappear in the Dicer '+' lanes. This suggests the possibility that the active silencing structure is the homodimer. If this is true, it is surprising we observe such strong silencing since only a very small fraction of the hairpins are in the homodimer structure. Regardless, it is unknown if the behavior of recombinant Dicer is representative of the behavior of Dicer inside a cell. While it seems unlikely that the only silencing structure is the homodimer, efficient silencing with a homodimer has been previously reported.[26]

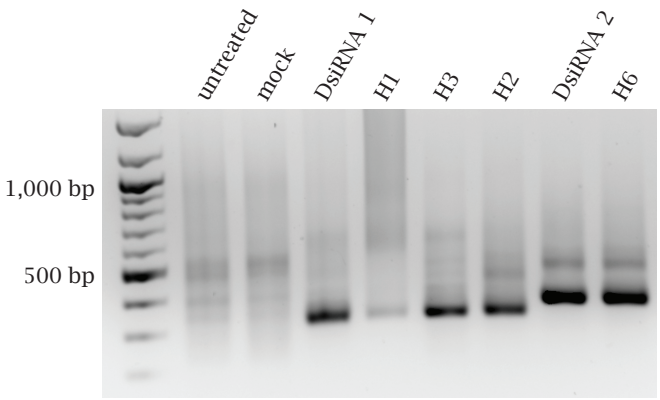


Figure 4.11: 5' RACE verifies that Y (d2EGFP mRNA) is cleaved in a sequence-specific manner after transfection of DsiRNAs or 4–14–4 hairpins. The strong bands in the DsiRNA and hairpin lanes correspond to amplification of the predicted Y fragment. There is a small amount of non-specific amplification in the untreated and mock samples.

5′ RACE (rapid identification of cDNA ends)[27] was used to verify that the observed silencing is due to specific cleavage of Y (d2EGFP mRNA) rather than a global halt in protein expression. An RNA oligonucleotide was ligated to the 5′-end of a total RNA sample collected after transfection of a DsiRNA or a 4–14–4 hairpin. The linker should selectively ligate to RNA with a 5′-PO_4, which includes mRNA that has been cleaved by RISC. Ago2 cleavage produces mRNA fragments with a 3′-OH and a 5′-PO_4.[28,29] Using primers specific for Y and the ligated RNA linker, the cleavage product was amplified (Figure 4.11). It was found that Y was cleaved in the region corresponding to the sequence of the transfected DsiRNA or hairpin (Table 4.1), indicating sequence-specific cleavage. However, the cleavage site appears to be heterogenous within this region. Interestingly, DsiRNA 1 produced some mRNA cleavage products that were not consistent with cleavage directed by the predicted siRNA. Perhaps, the asymmetric ends of the DsiRNA were not sufficient to completely

Strand	Predicted cleavage location	Experimental cleavage location PCR	Individual clones
DsiRNA 1	133–134 (21mer siRNA) 134–135 (22mer siRNA)	133–134, 134–135, 139–140	139–140
H1	between 123 and 144	134–135, 135–136, 137–138	135–136
H3	between 123 and 140	130–131	130–131, 131–132
H2	between 123 and 144	127–128	129–130
DsiRNA 2	87–88 (21mer siRNA) 88–89 (22mer siRNA)	88–89	88–89
H6	between 79 and 100	88–89	88–89

Table 4.1: 5' RACE determination of the exact mRNA cleavage location due to sequence-specific gene knockdown induced by the transfection of the listed RNA strands. The DsiRNAs and 4–14–4 hairpins indicated were transfected into the Y cell line, and total RNA was collected ∼24 hours post-transfection. The mRNA cleavage location was determined by sequencing either the 5' RACE-PCR reaction shown in Figure 4.11 or individual clones of the 5′ RACE-PCR. The numbering is relative to the translational start site. The cleavage occurred between the two nucleotides listed (e.g., 133–134 indicates cleavage between the 133rd and 134th nucleotides).

direct Dicer cleavage.

We attempted to determine the fate of our molecules inside the cell through next-generation sequencing. The cells were transfected with individual scRNAs, and total RNA was collected at 8 and 24 hours post-transfection. The total RNA was then used to create small RNA libraries which were sequenced with an Illumina HiSeq 2000. We encountered various problems with our results. First, very few full-length product reads were found for any scRNA, even for scRNAs which were spiked into the total RNA. Likely, the library generation protocol is unable to capture RNA species that contain significant secondary structure and/or are longer than a standard miRNA. Second, we noticed that hairpins with identical structures but different sequences produced varying cleavage patterns (refer to Figure A.1 in Appendix A). This could indicate sequence dependent cleavage but, more likely, this indicates bias in the library preparation. After further research, it became evident that small RNA library protocols exhibit significant bias that is not easily explained, and hence, not easily corrected.[30,31] This bias is less critical for experiments that seek to explore the differences in miRNA expression between samples. However, in our experiments, we attempted to compare the abundance of full-length (i.e., intact) to fragmented scRNAs in the same biological sample. The unreliable raw read-counts prevented accurate cross-sequence comparison. Theoretically, it is possible to generate a synthetic reference sample, but it is extremely difficult since the reference sample must include all scRNA fragments ≥10-nt in known concentrations. Thus, before next-generation sequencing will be useful to determine the fate of our scRNAs inside a cell, the bias in

small RNA libraries will need to be addressed. Work presented in Appendix A further explores small RNA library bias and introduces a library generation protocol that could potentially reduce library bias.

4.3 Conclusions and future directions

We have succeeded in engineering good ON and OFF states for both M1 and M3 in our tissue culture experimental setup. The initial components do not silence, maintaining a good OFF state. The final components, when annealed before delivery, silence Y well and demonstrate a strong ON state. Unfortunately, signal transduction has proved difficult, and we have yet to successfully turn a mechanism from OFF to ON with a nucleic acid trigger. Without signal transduction, we are unable to execute conditional logic. A greater understanding of the fate of scRNAs inside a cell is necessary to someday achieve signal transduction.

Unintended silencing by scRNAs, potentially through enzymatic cleavage, has been identified as a challenge. Chemical modifications help prevent, or at least slow down, this unintended silencing. But scRNAs cannot be fully modified because the final product must be a Dicer substrate. The shape transition from substrates that are immune to Dicer cleavage to one which is susceptible proved harder than originally imagined, partially because Dicer and/or another RNA processing enzyme are fairly lenient in their substrate requirements. Perhaps better use of chemical modifications could solve these problems. The final Dicer substrate may also need to be chemically modified. Luckily, there are many chemical modification patterns known to be compatible with Dicer cleavage and RISC activity.[32]

Unintended protein binding is a related source of difficulties for scRNAs in a cellular environment. While 2′-O-Me modifications may prevent Dicer cleavage of the scRNAs, they may be insufficient to prevent scRNA interactions with RNA binding proteins. More extreme modifications, such as attaching peptides, cholesterol, or other large molecules to the scRNAs, may be required to ensure that only the final product is recognized as a nucleic acid. Hopefully, the cell lysate system developed in Chapter 3 will be useful for studying protein:scRNA interactions and competitive inhibition of Dicer cleavage of the final Dicer substrate.

Another explanation for the lack of signal transduction is that the hybridization cascade does not function at the scRNA intracellular concentration levels. The mechanisms were tested at 0.5 μM in test-tube experiments in Chapter 2. The same mechanisms were then tested in tissue

culture. Subsequent cell lysate studies demonstrated that M1 is sensitive to concentration effects and significant changes were required to produce a functional mechanism at 5 nM. This experience illustrates the importance of determining the intracellular concentration of scRNAs and optimizing mechanisms for the appropriate concentration.

Low intracellular concentrations may result from poor scRNA delivery. Lipid-based transfection reagents work well for the delivery of siRNAs (and DsiRNAs) to cell lines, but they may not be the best delivery method for our scRNAs. scRNAs are similar in size to siRNAs, but scRNAs have a more stringent requirement of maintaining the original secondary structure and likely require a higher intracellular concentration. Lipid-based transfection reagents are notoriously inefficient and release only a small fraction of the transfected RNA into the cytoplasm (most of the RNA remains trapped in lysosomes).[33] While enough of the final Dicer substrate is delivered to produce strong gene knockdown by RNAi, it is likely that our mechanisms require a higher concentration in order to execute hybridization cascade pathways which require co-localization of scRNAs, nucleation, and branch migration.

We have identified a few roadblocks for conditional RNAi mechanisms and likely others exist. The challenge remains to overcome these roadblocks. Assays need to be designed to study each variable independently. The problem with measuring only gene knockdown is that if we solve one problem (e.g., delivery) without solving the remaining problems, we would not know we solved that problem. Hence, creativity must be employed to separate the system into its various components.

4.3.1 RNAi gene silencing via very small RNA structures

The silencing of various short hairpins, especially versions of the 4–14–4 structure, show that a great range of structures are capable of inducing sequence-specific gene knockdown. Work published around the time of our experiments demonstrates that short shRNAs (sshRNAs) can induce gene silencing in a sequence-specific manner.[21,34] These short shRNAs have 16–19-bp stems with loops of 0–10-nt. The discovery of these short shRNAs leads to many interesting questions. Is the observed silencing due to Dicer cleavage of the hairpins? The "rules" governing Dicer cleavage are well characterized, and Dicer is not believed to be extremely promiscuous. Instead, could Ago2, the protein responsible for cleaving mRNA in RISC, cleave the hairpins? Or does RISC unwind the hairpins? RISC is known to have helicase activity.[35] Or is this evidence of a completely different pathway capable of post-transcriptional gene silencing? Unfortunately, these questions remain to be answered.

In 2010, three papers reported the discovery of a Dicer-independent miRNA processing pathway for miRNA-451.[36–38] Unlike other pre-miRNA, the region of pre-miRNA-451 that is complementary to its target is located in the stem and loop of the hairpin. This caused confusion for many years because Dicer cleavage removes the hairpin loop. These three papers solved the mystery and report that the pre-miRNA is not cleaved by Dicer. Instead, it is cleaved by a component of RISC, Ago2, known as Slicer. Based on deep-sequencing experiments, they show that Ago2 cleaves approximately in the middle of the 3′-stem of the hairpin. The cleavage product is then trimmed to produce the mature miR-451. At this time, miRNA-451 is the only known Dicer-independent miRNA.

Our next step is to determine the scRNA cleavage patterns, so as to learn the new design "rules." Next-generation sequencing has the potential to identify scRNA cleavage patterns. However, this method has significant drawbacks, as described earlier and in more detail in Appendix A, which include bias and difficulty sequencing structures with strong secondary structure and 2′-O-Me modifications. Microarray analysis, small RNA Northerns and functional silencing assays can also help elucidate the scRNA cleavage patterns. In addition, enzymes responsible for scRNA cleavage could be identified through pull-down of scRNA:protein complexes.[39]

4.3.2 Future directions for achieving conditional RNAi *in vivo*

If we look beyond the goal of producing a Dicer substrate, there are likely many biological pathways well-suited to interact with scRNA-mediated logic gates. For example, miRNAs are effectively knocked down through base pairing to an anti-miR, generally a chemically modified complementary nucleic acid strand.[40,41] A conditional mechanism could be designed to execute the logic: *if gene X is expressed, prevent miRNA Y activity.* Because the final product does not require enzymatic processing, this mechanism, in theory, should be easier to execute *in vivo* than the mechanisms discussed here.

Despite the numerous challenges and uncertainties discussed here, engineering conditional RNAi mechanisms holds great promise. We have presented the conceptual framework to design *in vivo* conditional logical gates constructed from nucleic acids. The three conditional RNAi mechanisms introduce design versatility that will provide flexibility to modify the mechanisms as our understanding of scRNAs' interactions with the cellular environment improves. The first steps summarized here will hopefully lead to a time when the power of scRNA-mediated logical operations can be successfully integrated into a biological organism. Implementing conditional RNAi as a therapeutic or research tool has the potential to improve health and to lead to a better understanding of

biological systems.

4.4 Bibliography

[1] Li, X. Q., Zhao, X. N., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C. C., and Kain, S. R. Generation of destabilized green fluorescent protein transcription reporter. *Journal of Biological Chemistry* **273**(52), 34970–34975 (1998).

[2] An, C. I., Trinh, V. B., and Yokobayashi, Y. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA* **12**(5), 710–6 (2006).

[3] Beisel, C. L., Bayer, T. S., Hoff, K. G., and Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology* **4**, 224 (2008).

[4] Tuleuova, N., An, C. I., Ramanculov, E., Revzin, A., and Yokobayashi, Y. Modulating endogenous gene expression of mammalian cells via RNA-small molecule interaction. *Biochemical and Biophysical Research Communications* **376**(1), 169–73 (2008).

[5] Kumar, D., An, C. I., and Yokobayashi, Y. Conditional RNA interference mediated by allosteric ribozyme. *Journal of the American Chemical Society* **131**(39), 13906–7 (2009).

[6] Beisel, C. L., Chen, Y. Y., Culler, S. J., Hoff, K. G., and Smolke, C. D. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Research* **39**(7), 2981–94 (2011).

[7] Santangelo, P., Nitin, N., and Bao, G. Nanostructured probes for RNA detection in living cells. *Annals of Biomedical Engineering* **34**(1), 39–50 (2006).

[8] Goel, G., Kumar, A., Puniya, A. K., Chen, W., and Singh, K. Molecular beacon: a multitask probe. *Journal of Applied Microbiology* **99**(3), 435–42 (2005).

[9] Tan, W., Wang, K., and Drake, T. J. Molecular beacons. *Current Opinion in Chemical Biology* **8**(5), 547–53 (2004).

[10] Chen, A. K., Behlke, M. A., and Tsourkas, A. Efficient cytosolic delivery of molecular beacon conjugates and flow cytometric analysis of target RNA. *Nucleic Acids Research* **36**(12), e69 (2008).

[11] Nitin, N., Santangelo, P. J., Kim, G., Nie, S., and Bao, G. Peptide-linked molecular beacons for efficient delivery and rapid mRNA detection in living cells. *Nucleic Acids Research* **32**(6), e58 (2004).

[12] Nitin, N. and Bao, G. NLS peptide conjugated molecular beacons for visualizing nuclear RNA in living cells. *Bioconjugate Chemistry* **19**(11), 2205–11 (2008).

[13] Constantinou, A., Davies, A. A., and West, S. C. Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. *Cell* **104**(2), 259–68 (2001).

[14] Zhang, D. Y. and Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131**(47), 17303–14 (2009).

[15] Kim, D. H., Behlke, M. A., Rose, S. D., Chang, M. S., Choi, S., and Rossi, J. J. Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nature Biotechnology* **23**(2), 222–6 (2005).

[16] Kim, D. H. and Rossi, J. J. Strategies for silencing human disease using RNA interference. *Nature Reviews Genetics* **8**(3), 173–84 (2007).

[17] Castanotto, D., Sakurai, K., Lingeman, R., Li, H., Shively, L., Aagaard, L., Soifer, H., Gatignol, A., Riggs, A., and Rossi, J. J. Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. *Nucleic Acids Research* **35**(15), 5154–64 (2007).

[18] Grimm, D., Streetz, K. L., Jopling, C. L., Storm, T. A., Pandey, K., Davis, C. R., Marion, P., Salazar, F., and Kay, M. A. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**(7092), 537–41 (2006).

[19] Snove, O., J. and Rossi, J. J. Toxicity in mice expressing short hairpin RNAs gives new insight into RNAi. *Genome Biology* **7**(8), 231 (2006).

[20] Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**(2), 199–208 (2003).

[21] Ge, Q., Ilves, H., Dallas, A., Kumar, P., Shorenstein, J., Kazakov, S. A., and Johnston, B. H. Minimal-length short hairpin RNAs: the relationship of structure and RNAi activity. *RNA* **16**(1), 106–17 (2010).

[22] Dallas, A., Ilves, H., Ge, Q., Kumar, P., Shorenstein, J., Kazakov, S. A., Cuellar, T. L., McManus, M. T., Behlke, M. A., and Johnston, B. H. Right- and left-loop short shRNAs have distinct and unusual mechanisms of gene silencing. *Nucleic Acids Research* **40**(18), 9255–71 (2012).

[23] Dallas, A. and Johnston, B. H. Design and chemical modification of synthetic short shRNAs as potent RNAi triggers. *Methods in Molecular Biology* **942**, 279–90 (2013).

[24] Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W. S., Karpilow, J., and Khvorova, A. The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* **11**(5), 674–82 (2005).

[25] Takyar, S., Hickerson, R. P., and Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120**(1), 49–58 (2005).

[26] Lapierre, J., Salomon, W., Cardia, J., Bulock, K., Lam, J. T., Stanney, W. J., Ford, G., Smith-Anzures, B., Woolf, T., Kamens, J., Khvorova, A., and Samarsky, D. Potent and systematic RNAi mediated silencing with single oligonucleotide compounds. *RNA* **17**(6), 1032–7 (2011).

[27] Frohman, M. A., Dush, M. K., and Martin, G. R. Rapid production of full-length cDNAs from rare transcripts—amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America* **85**(23), 8998–9002 (1988).

[28] Martinez, J. and Tuschl, T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes & Development* **18**(9), 975–980 (2004).

[29] Zhuang, F., Fuchs, R. T., and Robb, G. B. Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. *Journal of Nucleic Acids* **2012**, 360358 (2012).

[30] Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J. T., Nusbaum, J. D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., and Tuschl, T. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**(9), 1697–712 (2011).

[31] Linsen, S. E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., Fritz, B., Wyman, S. K., de Bruijn, E., Voest, E. E., Kuersten, S., Tewari, M., and Cuppen, E. Limi-

tations and possibilities of small RNA digital gene expression profiling. *Nature Methods* **6**(7), 474–6 (2009).

[32] Collingwood, M. A., Rose, S. D., Huang, L., Hillier, C., Amarzguioui, M., Wiiger, M. T., Soifer, H. S., Rossi, J. J., and Behlke, M. A. Chemical modification patterns compatible with high potency Dicer-substrate small interfering RNAs. *Oligonucleotides* **18**(2), 187–200 (2008).

[33] Dominska, M. and Dykxhoorn, D. M. Breaking down the barriers: siRNA delivery and endosome escape. *Journal of Cell Science* **123**(Pt 8), 1183–9 (2010).

[34] Ge, Q., Dallas, A., Ilves, H., Shorenstein, J., Behlke, M. A., and Johnston, B. H. Effects of chemical modification on the potency, serum stability, and immunostimulatory properties of short shRNAs. *RNA* **16**(1), 118–30 (2010).

[35] Robb, G. B. and Rana, T. M. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Molecular Cell* **26**(4), 523–37 (2007).

[36] Cheloufi, S., Dos Santos, C. O., Chong, M. M., and Hannon, G. J. A Dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**(7298), 584–9 (2010).

[37] Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G. J., Lawson, N. D., Wolfe, S. A., and Giraldez, A. J. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**(5986), 1694–8 (2010).

[38] Yang, J. S., Maurin, T., Robine, N., Rasmussen, K. D., Jeffrey, K. L., Chandwani, R., Papapetrou, E. P., Sadelain, M., O'Carroll, D., and Lai, E. C. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **107**(34), 15163–8 (2010).

[39] Ule, J., Jensen, K., Mele, A., and Darnell, R. B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**(4), 376–86 (2005).

[40] Meister, G., Landthaler, M., Dorsett, Y., and Tuschl, T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10**(3), 544–50 (2004).

[41] Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M., and Stoffel, M. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**(7068), 685–9 (2005).

Appendix A

Novel method to reduce bias in small RNA libraries for next-generation sequencing

A.1 Abstract

Next-generation sequencing of small RNAs has the potential to generate large quantities of previously unavailable data and to improve our understanding of small RNA function.[1] However, the current protocols used to prepare small RNA libraries are significantly biased.[2–4] This bias prevents determination of absolute small RNA numbers—only relative changes in small RNA numbers can be determined accurately. The bias is primarily introduced by the single-stranded ligation of adapters to small RNAs.[2,4] This step is highly influenced by the sequence and structure of both the small RNAs and the adapters.[5,6] The work presented here illustrates the ligation bias and introduces methods we are developing to reduce ligation bias in preparing small RNA libraries. Our ultimate goal is to develop a less-biased protocol to generate small RNA libraries for next-generation sequencing.

A.2 Introduction

Next-generation sequencing has the potential to both discover new small RNAs and provide quantitative information about the tissue-specific expression of small RNAs under various conditions.[1] Small RNAs are important gene regulators, controlling many cellular pathways.[7] Next-generation sequencing (also known as deep sequencing due to the high number of reads and sequence space coverage) of miRNAs has shown miRNA dysregulation in cancer,[8–10] miRNA control of develop-

ment,[11] and extensive intracellular 3′-end modification of small RNAs.[12]

Most small RNA library preparation protocols consist of three steps: (1) ligation of adapters to both ends of the small RNA, (2) reverse transcription (RT) of the ligation product, and (3) amplification of the cDNA by PCR. The adapter ligation is selective for Dicer products because only small RNAs that display the characteristic 3′-hydroxyl and 5′-phosphate of Dicer cleavage are ligated.[13,14] To selectively amplify other small RNAs, the protocols can be altered to ligate small RNAs with other characteristic end-group chemistries, indicative of the enzyme responsible for cleavage.[14] Sample multiplexing is achieved through barcoding. The barcodes are short variable sequences, generally six nucleotides, that are located in the adapter. After sequencing, the barcodes are used during data analysis to assign reads to their original sample.[15] Barcodes are introduced either during the ligation step (the adapter contains the barcode) or during the PCR step (the primer contains the barcode).

While small RNA library preparation protocols are fairly fast and easy, they produce a biased library. This is demonstrated clearly when a pool of equimolar synthetic miRNAs are sequenced. The expected sequencing read distribution for a non-biased library preparation of an equimolar miRNA pool is Poisson.[6,16] A few groups have published results of these control experiments and have shown that the frequency of individual miRNA sequence reads varies by more than three or four orders of magnitude,[2,3] far outside the expected distribution. The bias helps explain the contradictory results produced when investigating the same biological sample using next-generation sequencing, microarray analysis, and quantitative PCR.[3,17]

The small RNA library bias is primarily due to adapter ligation.[2,4] Single-stranded RNA ligation methods have low yield and slow kinetics, with certain sequences ligating more efficiently to the adapters than others.[5] Additionally, RNA ligases have strong small RNA preferences, and the choice of RNA ligase influences the library distribution.[2] An evolved variant of RNA ligase 2 is often used to ligate a pre-adenylated RNA linker to the 3′-end of the small RNA to prevent unintended concatemerization and ring formation.[18]

The most effective way to reduce small RNA library bias is to improve the representation of each and every small RNA in the library. Various ligation protocols have been proposed to increase ligation yields and, hence, reduce the bias.[5,19] Unfortunately, none have been able to significantly reduce the bias. Scientists at New England Biolabs (NEB) investigated the ligation bias by isolating the 3′ ligation step.[5] They showed strong evidence of ligation bias and developed adapters with randomized regions to increase ligation efficiency and reduce ligation bias. They

report an improvement of 10%. Sorefan et al. introduced a "high definition" protocol to reduce RNA ligase-dependent cloning bias.[19] This new protocol identified previously unidentified miRNAs. They noted that miRBase, a database of miRNAs,[20–23] is composed of miRNAs favored by the current adapter ligation protocols, leading to the mischaracterization of the relevant importance of various miRNAs.

An alternate approach to improving small RNA representation in small RNA libraries is to replace single-stranded RNA ligation with double-stranded ligation using splints. Ambion small RNA library preparation protocols utilize a method they term "Ligase Enhanced Genome Detection" (LEGenD^{TM}). This protocol uses simultaneous ligation of the 3′ and 5′ adapters to small RNAs using DNA oligonucleotides as splints. Another group has improved ssDNA ligation and reduced ligation bias by using DNA hairpins that function as both the adapter and the splint.[24]

A.3 Results

A.3.1 Results from next-generation sequencing of 4–14–4 RNA hairpins

As described in Chapter 4, we attempted to determine the fate of scRNAs inside the cell using next-generation sequencing. Unfortunately, the results were inconclusive and caution is required while interpreting the results because of insufficient controls. Figure A.1 illustrates the deep sequencing results for five different 4–14–4 scRNA hairpins at 8 hours and 24 hours post-transfection. The five 4–14–4 hairpins include three with a 3′-toehold and two with a 5′-toehold. Identical scRNAs structures show different cleavage patterns. In Figure A.1, each nucleotide is represented by a shaded dot. The color shading corresponds to the fraction of mapped sequence reads which include that particular nucleotide (i.e., the probability of that nucleotide existing in a sequencing read that maps to the hairpin). A non-processed hairpin in a non-biased library would appear uniformly dark red because every base would exist with a probability near one.

Based on the unexpected silencing of scRNAs shown in Chapter 4, we expect that the scRNAs are enzymatically processed but are not certain of the cleavage rules. If the enzymatic processing of 4–14–4 hairpins is sequence independent, hairpins with identical structures are expected to produce very similar cleavage patterns. This is not the case for these hairpins, and it is important to determine if this is a biological result or an artifact of the sequencing protocol.

The three 3′-toehold 4–14–4 hairpins in Figure A.1 do appear to have similar coloration patterns, though not similar enough to deduce a cleavage "rule." The probability of a base existing in the

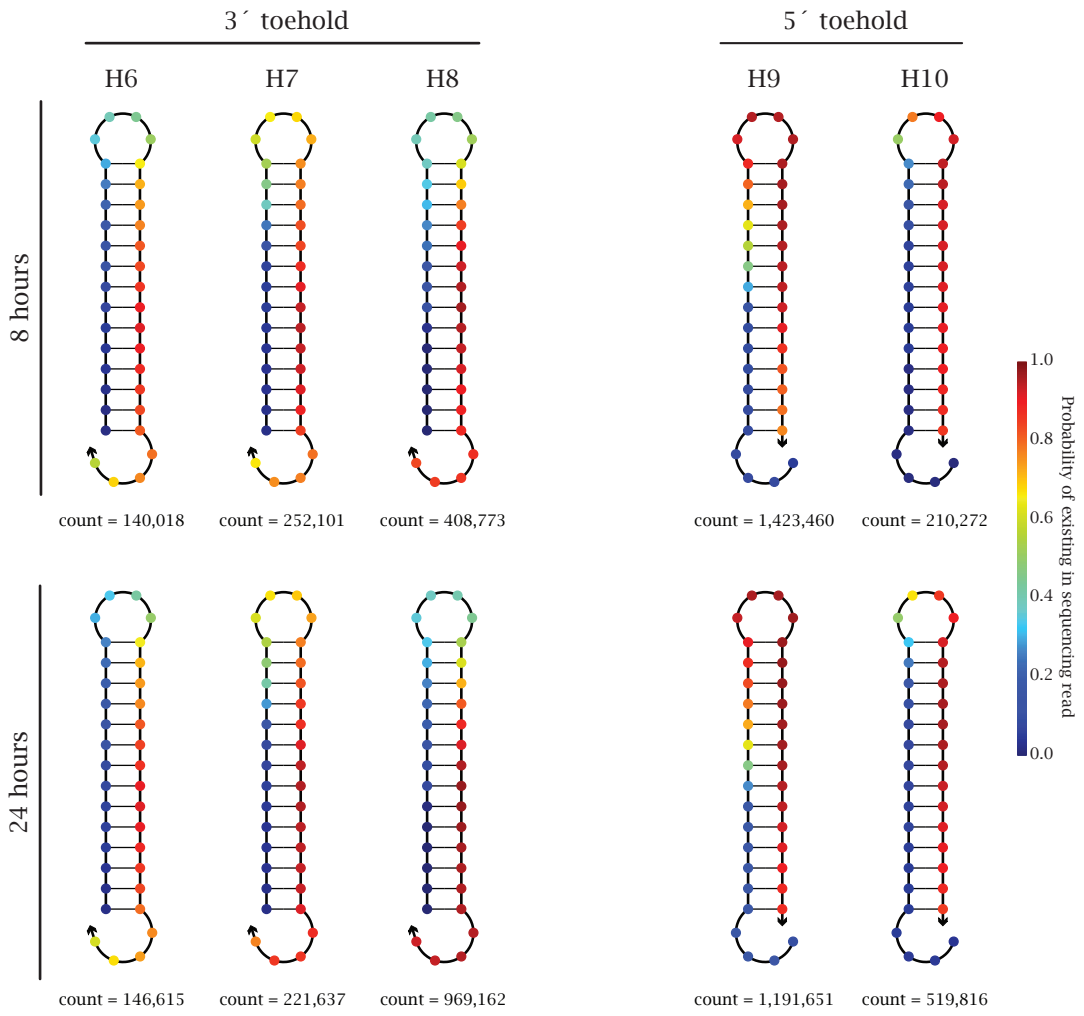


Figure A.1: Next-generation sequencing results for five transfected 4–14–4 RNA hairpins. The coloring represents the probability that a base is present in the sequencing reads that map to the hairpin (sequence alignments of fragments <10-nt were not included). The count is the total number of reads that map to the hairpin. Total RNA was collected at both 8 hours and 24 hours post-transfection. H6 was transfected in HEK 293 d2EGFP cells, H7 and H9 were transfected in U87MG Delta cells, while H8 and H10 were transfected in TC71 cells.

sequencing reads does not appear to change significantly between 8 and 24 hours, implying that neither further cleavage nor strand degradation occurs in this time. The 3′-stem is well represented in the sequencing reads while the 5′-stem is essentially never found. The greatest variation in the probability of a nucleotide existing between the three hairpins is seen in the hairpin loop and toehold regions. With the 5′-toehold hairpins, the toehold region rarely appears in the sequencing reads, while the loop is well represented. A hypothesis consistent with these results is that enzymatic cleavage functions by measuring ∼16-nt to 18-nt from the 3′-end of the hairpin. Perhaps the predicted hairpin structure is disrupted during delivery and the location of the toehold is not relevant.

Alternatively, these results indicate a strong preference in the library generation protocol towards retaining fragments that originated from the 3′-end of the hairpin. This preference is plausible since the 3′ adapter was attached before the 5′ adapter. The possibility of bias is further supported by experimental results that show moving the AS sequence from the 3′-stem to the 5′-stem does not prevent Y (d2EGFP) silencing, despite the fact that the 5′-stem rarely appears in the sequencing reads. While we cannot produce conclusive evidence that our deep sequencing results are biased, it appears likely that the results are strongly influenced by a sequencing artifact.

A.3.2 Primary source of bias: linker ligation

Figure A.2 provides a striking example of small RNA ligation bias. Four closely related small RNAs were ligated to the universal miRNA cloning linker sold by NEB for small RNA cloning and miRNA library construction. As seen, none of the ligation reactions fully ligate the small RNA to the universal miRNA cloning linker. Even more troubling is the large variation in the ligation product yield between the four reactions. For these ligations, the universal miRNA cloning linker, also referred to as the 3′ linker or 3′ adapter, is ligated to the 3′-end of the small RNA. The 3′ linker is 5′-adenylated to allow the ligation to occur without ATP. Ligation reactions performed with T4 RNA ligase 2, truncated, K227Q, in the absence of ATP should avoid unintended ligation reactions, including self-ligation of small RNAs into rings or polymers.[25] The 3′ linker also contains a 3′-amine blocking group to prevent the 3′-end from participating in ligation reactions. For the reactions shown in Figure A.2, the linker was used in approximately threefold excess relative to the small RNA strand.

None of the small RNA strands tested in Figure A.2 are fully ligated to the universal miRNA cloning linker. PS-D and GS-s show especially poor ligation to the 3′ linker. GS-D and PS-D are

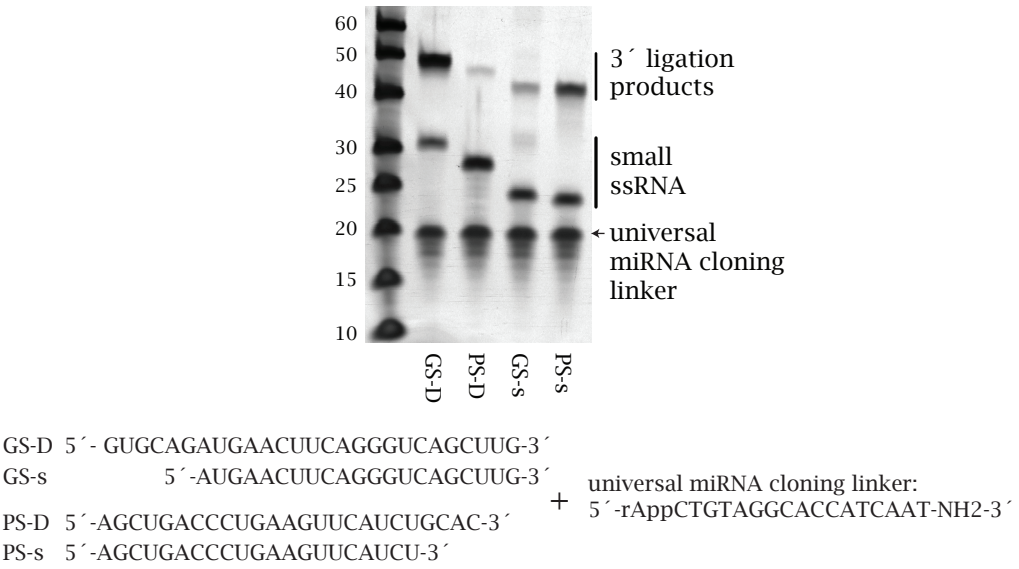


Figure A.2: Ligation bias is evident from the significantly different yields from ligating a universal miRNA cloning linker to small RNAs. The universal miRNA cloning linker (NEB: New England Biolabs) was ligated to the guide strand or passenger strand of DsiRNA 1 (GS-D, PS-D) or the corresponding siRNA (GS-s, PS-s). 3.75 pmol RNA was ligated to 10 pmol universal miRNA cloning linker in T4 RNA ligase reaction buffer with 10% PEG8000 using T4 RNA ligase 2, truncated, K227Q (NEB). The reactions were incubated at 25°C for 2 hours and separated by 15% denaturing PAGE. A 10/60 DNA length standard (IDT) was used to provide approximate size information.

the guide and passenger strands, respectively, of DsiRNA 1. GS-s and PS-s correspond to the guide and passenger strands, respectively, of the siRNA predicted to be the product of Dicer cleavage of DsiRNA 1. Thus, GS-D and GS-s are identical except for the last 6-nt at the 5′-end, yet GS-D is ligated to the 3′ linker much more efficiently than GS-s. This difference in ligation yield is especially worrisome since the ligation occurs at the 3′-end. This shows that six bases located more than 20-nt from the ligation site can significantly change the ligation yield and illustrates the complexity of the ligation bias. The bias, therefore, is not solely due to the primary sequence and is likely also caused by structural effects. This makes it extremely difficult to predict the ligation bias *a priori*.[5]

While Figure A.2 demonstrates the effect of small RNA structure on ligation, it follows that the adapter structure also influences the ligation reaction. Other groups have demonstrated that the linker ligation bias is dependent on both the adapter and small RNA structure. Adapter dependent bias is a significant concern for samples where the barcode is introduced during the ligation step. To achieve barcoding, each sample uses a slightly different adapter. This influences the ligation reaction, and different sequencing read frequencies are obtained for identical small RNA samples.[6] Therefore, the current best practice is to introduce the barcodes during the PCR step despite the

extra work of preparing the samples in parallel during the ligation and RT steps.[6]

A.3.3 Improving single-stranded RNA ligation via splinted T4 DNA ligation

In theory, ligation bias will vanish if 100% of the small RNA strands are ligated to linkers. As demonstrated in Figure A.2, the current ligation reactions are far from achieving complete ligation. In hopes of improving the ligated substrate yield, we tested splinted ligation, which is a common technique used to enhance single-stranded ligation. A nucleic acid splint bridges the ligation junction by base pairing to both ligation substrates to form a nicked double-stranded structure that co-localizes the 5′- and 3′-ends of the single-stranded nucleic acids. This greatly increases the ligation rate because ligases generally prefer double-stranded substrates.

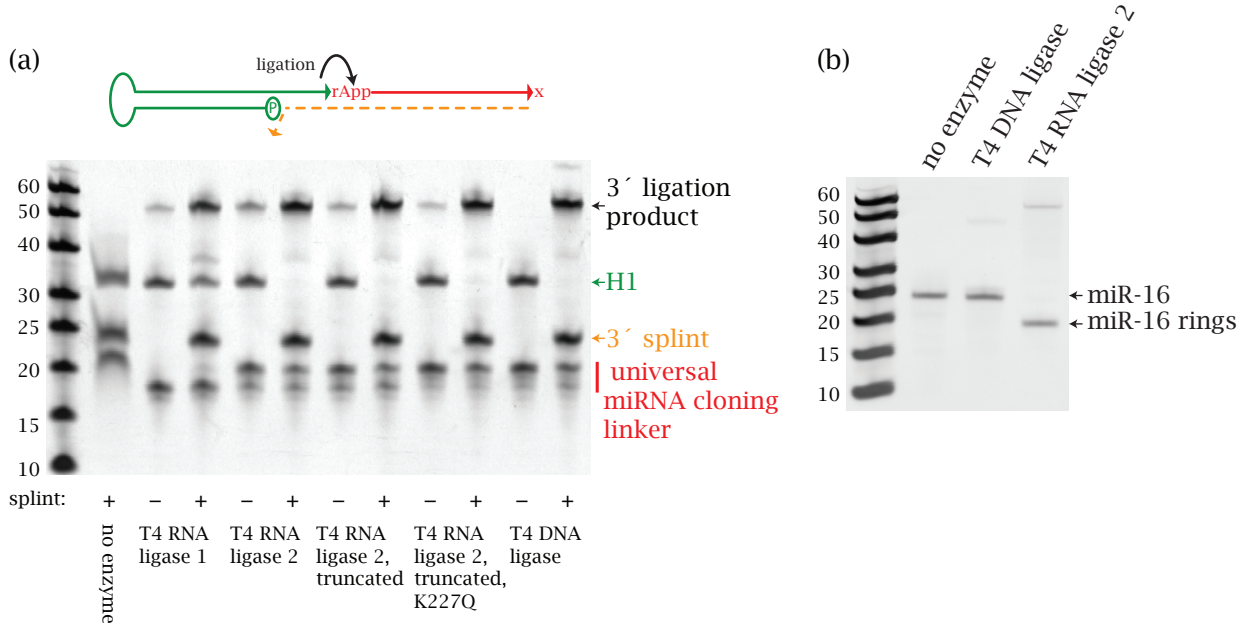


Figure A.3: **(a)** Ligation reactions for five T4 ligases joining the 3′-end of H1 to the universal miRNA cloning linker (NEB). A splint complementary to the universal miRNA cloning linker and 6-nt of H1 increases ligation efficiency for all enzymes tested. 3.75 pmol RNA was ligated to 10 pmol universal miRNA cloning linker with or without 10 pmol 3′ splint in T4 RNA ligase reaction buffer with 10% PEG8000 using T4 RNA ligase 1; T4 RNA ligase 2; T4 RNA ligase 2, truncated; T4 RNA ligase 2, truncated, K227Q; or T4 DNA ligase (NEB). **(b)** miR-16 self-ligates to form rings in the presence of T4 RNA ligase 2 but not T4 DNA ligase. 7.5 pmol miR-16 was incubated with either no enzyme, T4 DNA ligase, or T4 RNA ligase 2 in T4 RNA ligase 2 buffer with 20% PEG8000. **(a),(b)** The reactions were incubated at 25°C for 2 hours and separated by 15% denaturing PAGE. A 10/60 DNA length standard (IDT) was used to provide approximate size information.

Figure A.3(a) shows ligation of H1 to the universal miRNA cloning linker in the absence and presence of a DNA splint for five commonly used T4 ligases commercially available from NEB. The

addition of a splint greatly increases the ligation yield for all five ligases tested. Adding a splint to the ligation reactions leads to nearly complete ligation of H1 and the 3′ linker (as judged by visual depletion of H1) for all ligases tested except for T4 RNA ligase 1. It is not surprising that T4 RNA ligase 1, known as an ssRNA ligase, is a poor dsRNA ligase. (T4 RNA ligase 2 is known as a dsRNA ligase.) However, both T4 RNA ligase 1 and T4 RNA ligase 2 appear to be equally poor ssRNA ligases (Figure A.3(a), compare lanes without splint), demonstrating the difficulties of single-stranded ligation.

In Figure A.3(a), the universal miRNA cloning linker runs as two distinct bands in the gel. The upper band is the original pre-adenylated universal miRNA cloning linker. The lower band is the de-adenylated universal miRNA cloning linker. De-adenylation occurs through the transfer of AMP from the linker to the ligase active site. The AMP in the ligase active site can now be transferred to any 5′-PO_4, which defeats the purpose of performing the reactions without ATP and leads to unintended ligations. T4 RNA ligase 2, truncated, K227Q was engineered specifically to lack de-adenylation activity.[25] Figure A.3(a) confirms that the linker is not de-adenylated with T4 RNA ligase 2, truncated, K227Q.

Perhaps the most interesting result from Figure A.3(a) is that the ligation of H1 to the 3′ linker by T4 DNA ligase can be switched from ∼0% to nearly 100% by the addition of a splint. In the absence of a splint, T4 DNA ligase produces essentially no ligation products. In the presence of a splint, T4 DNA ligase ligates nearly 100% of H1 to the 3′ linker. This demonstrates that T4 DNA ligase can accept RNA as the 3′-OH end. Additionally, this demonstrates that T4 DNA ligase requires a double-stranded substrate. The switch-like behavior induced by the addition of a splint suggests that T4 DNA ligase could be used to achieve complete ligation of the desired substrates while reducing off-target ligations. In fact, this property has been exploited to selectively detect miRNAs.[26]

Because T4 DNA ligase appears to have no single-stranded ligase ability, it is much less likely to ligate small RNAs into rings or polymers. Ring formation is especially problematic and leads to biased libraries. If a small RNA substrate self-ligates and forms a ring instead of ligating to both linkers, it will not be amplified and will be missing in the final library. Biased libraries also occur due to concatemerization, though it is less common at low concentrations. An example of hsa-miR-16-5p (miR-16) ring formation with T4 RNA ligase 2, but not T4 DNA ligase, is shown in Figure A.3(b). miR-16 was incubated alone in ligation conditions with either no enzyme, T4 DNA ligase, or T4 RNA ligase 2. The mobility of miR-16 shifts in the presence of T4 RNA ligase

2, indicating that the miRNA has been circularized by ligation of the 5′-PO_4 to the 3′-OH on the same strand. T4 DNA ligase does not generate any miR-16 rings.

The properties of T4 DNA ligase were further investigated and are summarized in Table A.1. While T4 DNA ligase is able to ligate RNA to RNA, it has difficulty ligating a DNA 3′-end to an RNA 5′-end. Thus, the 5′ linker must be RNA at the 3′-end to use T4 DNA ligase.

Table A.1: T4 DNA ligase preference for DNA and RNA substrates. 3′-end refers to the oligonucleotide strand providing the 3′-OH for the ligation reaction. 5′-end refers to the oligonucleotide strand providing either the 5′-PO_4 (5′-adenyl group for pre-adenylated strands) for the ligation reaction.

3′-end	5′-end	Splint	Ligation with T4 DNA ligase
DNA	DNA	DNA	yes
RNA	DNA: pre-adenylated	none	none
RNA	DNA: pre-adenylated	DNA	yes
DNA	RNA	DNA	slight
RNA	RNA	DNA	yes

As discussed earlier, ligating 100% of the small RNA strands to linkers removes ligation bias. A small RNA may not be ligated to a linker for two reasons: (1) incomplete ligation of the linkers to the small RNA substrate or (2) self-ligation of the small RNA to form rings which prevent linker ligation. Splinted ligations using T4 DNA ligase could address both of these problems. Figure A.4 demonstrates the improved ligation protocol, increasing the ligation of both adapters to the small RNA to nearly 100%. The new protocol uses T4 DNA ligase for splinted ligation of both the 3′ and 5′ linkers to a small RNA. The splints base pair with the entire linker and six nucleotides of the small RNA. These splints will only facilitate the ligation of the strands that are complementary to their sequence. Because of this sequence complementarity requirement, they are referred to as "specific" splints. To extend this protocol to libraries of small RNAs, it will be necessary to use a pool of specific splints capable of binding to all small RNAs in a given library.

The improved ligation protocol was tested for four small RNAs in Figure A.4. Each small RNA was 5′-end labeled with ^{32}P and ligated sequentially to the 3′ and 5′ linkers. GS-D, GS-s and H1 were used earlier to illustrate ligation bias. H1 is a 4–14–4 hairpin that is representative of the substrates that we would like to sequence to help dissect the unexpected silencing discussed in Chapter 4. miR-16 was included as a representative miRNA. The ligation product with both linkers attached is the predominate product for GS-D, GS-s, and miR-16. Unfortunately, the predominate

band for H1 corresponds to the hairpin ligated to only the 3′ linker. The 5′ linker was unable to ligate to the hairpin (data not shown). Likely, the strong secondary structure of H1 (the 5′-end is recessed) prevents the splint from binding and prevents ligation. While this protocol appears to be a significant improvement over single-stranded ligation, it is unable to attach an adapter to a recessed end. This likely does not cause a problem for many miRNAs, which generally lack significant secondary structure. However, it appears that any small RNA with significant secondary structure will not be included in the library. Additional steps to allow for linker ligation to a recessed end through self-splinting are currently in development. Even if both linkers are successfully ligated to the small RNA, the secondary structure may lead to bias in the subsequent RT and PCR steps.

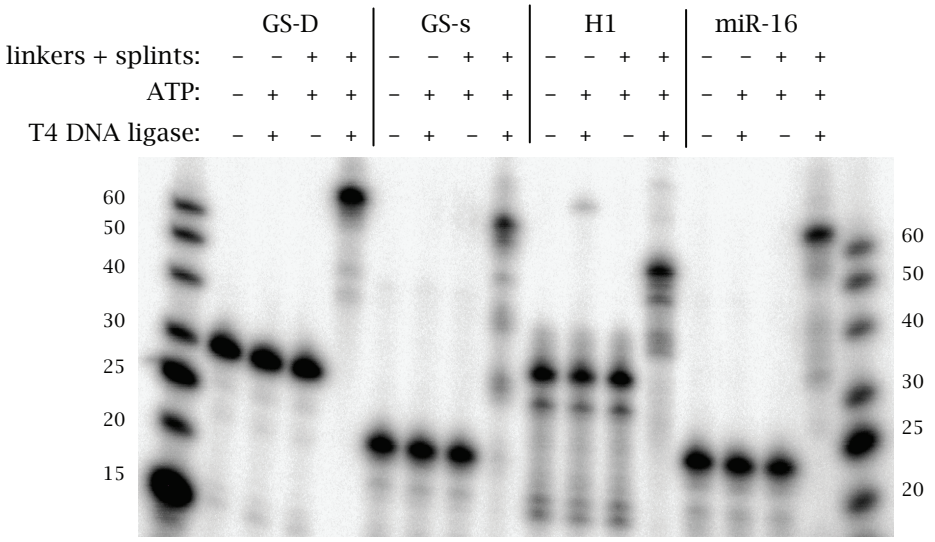


Figure A.4: Verification of 3′ and 5′ linker ligation to radiolabeled GS-D, GS-s, H1, and miR-16. The small RNA strand (0.3–0.6 pmol), both linkers and both splints (1.8 pmol each) were heated to 65°C for 10 min and cooled at 16°C for 5 min in 0.4× hybridization solution. The small RNA and 3′ linker were ligated in 20 μL reactions with T4 DNA ligase in T4 DNA ligase buffer without ATP, supplemented with 20% PEG8000, for 1 hour at 25°C. ATP (1 nM) was added to allow 5′ linker ligation, and the reaction was incubated for an additional 2.5 hours at 25°C. The strands were separated by 15% denaturing PAGE with formamide, exposed to a phosphorimager plate (Fujifilm type BAS-MS) and scanned using an FLA-5100 imaging system (Fuji Photo Film). Linkers and splints: 3′ linker: universal miRNA cloning linker; 5′ linker: 2.RNA acceptor oligo; and sequence-specific 3′ and 5′ splints.

A.3.4 Proposed protocol to reduce bias in small RNA cloning

To overcome the current bias in preparing small RNA libraries for next-generation sequencing, we propose a new small RNA library cloning protocol (Figure A.5). The full protocol includes adapter

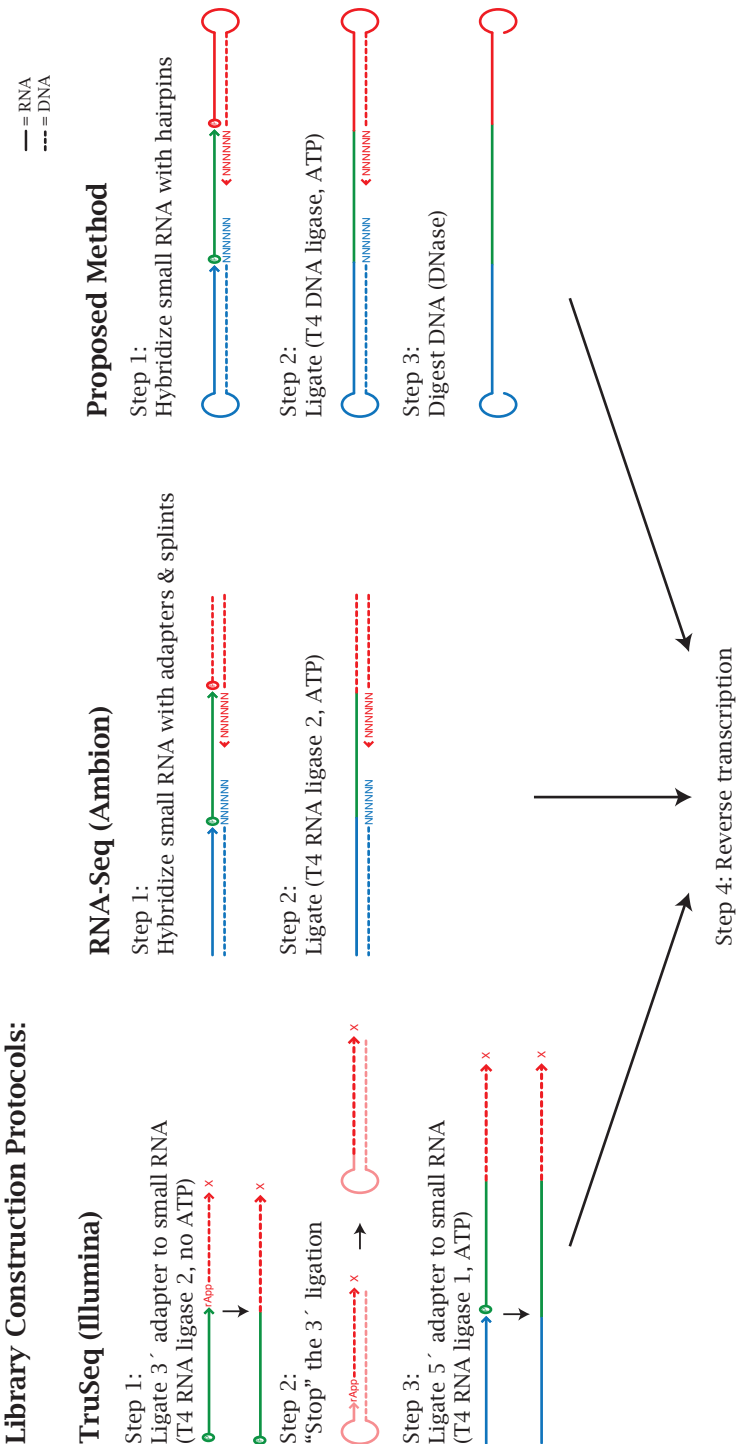


Figure A.5: Schematic depicting the construction of small RNA libraries using three methods to attach the 5′ and 3′ linkers. TruSeq (Illumina) ligates the linkers sequentially through single-stranded ligation. The 3′ linker is pre-adenylated and ligated without ATP to prevent spurious ligations. RNA-Seq (Ambion) simultaneously ligates both linkers using DNA splints and T4 RNA ligase 2. The proposed method simultaneously ligates both linkers, in the form of RNA:DNA hybrid hairpins that serve as both the linker and splint, with T4 DNA ligase. The splints are removed through DNA digestion.

ligation, which has been discussed in detail earlier, and adds the RT and PCR steps. The ligation step uses T4 DNA ligase, instead of a variant of T4 RNA ligase, with "universal" splints. Both adapters are ligated simultaneously to the small RNA in the presence of ATP. Rings are prevented through the use of T4 DNA ligase. To ligate any sequence with splinted ligation, we must transition from "specific" splints to "universal" splints.

A universal splint is not a single splint; instead, it is a pool of splints that has every possible six-nucleotide sequence, synthesized through the use of degenerate bases, in the region that base pairs with the small RNA. The portion of the splint that base pairs to the small RNA should be the minimal length required to achieve 100% ligation so as to reduce the size of the splint pool. Based on comparisons of specific splints that contained either 4-nt or 6-nt complementary to the small RNA, it appears that 4-nt is too short to achieve full ligation. A 6-nt complementarity region between the splint and the small RNA was needed to achieve 100% ligation. This creates a fairly large universal splint pool. A universal splint pool that contains every possible sequence will have $4^6 = 4096$ (4 possible bases at 6 locations) unique splints. For the small RNA to be the limiting strand, the universal splint pool, assuming equimolar concentration of each sequence, will need to be present in concentrations greater than 4096× that of an individual small RNA.

Unlike the linker:splint duplexes used in the ligations shown in Figure A.4, the proposed universal splints are hairpins that contain both the linker and the splint. The hairpins are RNA:DNA hybrids where the RNA portion of the hairpin is the linker, and the DNA portion is the splint. The RNA linker sequences are compatible with Illumina sequencing platforms. Combining the linker and splint into one oligonucleotide ensures perfect linker:splint stoichiometry. After ligation, DNase 1 is introduced to degrade the DNA splints, leaving a single-stranded RNA final product. DNase 1 is removed before reverse transcription by the Zymo RNA Clean and Concentrator column.

The protocol outlined in Figure A.5 is currently being optimized. To date, we can amplify 5 fmol of miR-16 using specific splints. Introducing universal splints leads to significant ligation of the 3′ linker to the 5′ linker. The fraction of linkers that ligate to another linker is very small and cannot be observed before PCR amplification. Unfortunately, the amplification of linker-linker ligation products overwhelms the signal of the amplified linker-small RNA-linker ligation products. Amplification of the linker-linker ligation products can deplete reagents in the PCR step and introduce bias in the amplification of the desired small RNA products.

The current experimental setup (5 fmol of miRNA ligated to 100 pmol of the universal splint pool) represents the most challenging situation because only one small RNA is present in a very

small quantity relative to the universal splints. All but one sequence in the universal 3′ splint pool and one sequence in the universal 5′ splint pool lack a complementary small RNA partner and have more opportunity to interact with other universal splints. Beginning with a pool of small RNAs, instead of just one, is more representative of a biological sample and may generate less linker-linker ligation products. The linker-linker ligation is possibly due to synthesis impurities that are apparent due to the large excess of universal splints. Further purification of the hairpins may reduce linker-linker ligation. Alternatively, we could use an LNA strand complementary to the linker-linker ligation junction to prevent PCR amplification.[27] This is a less desirable solution because amplification of desired sequences could also be blocked.

Likely, multiple protocol modifications, in addition to preventing linker-linker ligation, will be required to fully optimize this protocol. Upon completing protocol optimization, we plan to prepare a small RNA library from an equimolar pool of miRNAs and sequence the library using a standard next-generation sequencing platform. Our hope is that this new protocol will produce small RNA libraries with significantly less bias.

A.4 Discussion and conclusions

Next-generation sequencing technologies allow scientists to carry out experiments that only a few years ago were inconceivable. Unfortunately, small RNA sequencing is limited by biases introduced in the preparation of the library for sequencing. Our hope is that, upon optimization, the new small RNA library preparation protocol described here will produce sequencing results that accurately capture the nature of the underlying small RNA sample. By using universal splints and T4 DNA ligase to improve the linker ligation yield to small RNAs, we believe it is possible to significantly reduce the current ligation bias.

As expected, many of the most highly studied miRNAs are the ones that appear most frequently in small RNA sequencing results.[19] Since small RNA sequencing results are highly skewed, perhaps the most studied miRNAs are not actually the most abundant or biologically relevant but are simply the miRNAs most compatible with the current ligation protocols. This ligation bias is likely present in any protocol which relies on single-stranded ligation of a linker to a nucleic acid, such as the 5′ RACE assay in Chapter 4. In the future, it would be ideal to skip the linker ligation steps altogether. Perhaps new technologies or methods will emerge that dispense with the ligation step and directly sequence the small RNAs.

This chapter, along with the work of other groups, demonstrates that the current methods of generating small RNA libraries introduces large bias when ligating the $3'$ and $5'$ linkers to small RNAs. Unfortunately, many scientists remain unaware of the systemic biases introduced in small RNA library preparation. The quantitative nature of the sequencing results presents a false sense of certainty. The lack of awareness likely persists because deep sequencing is relatively expensive and complicated. To save money, many research groups attempt to maximize the information gained using the fewest number of sequencing runs and leave out comprehensive control experiments. To save time, many research groups outsource library preparation and have little direct knowledge of the process. As the use of next-generation sequencing grows, it is important for scientists to remember the potential pitfalls and work hard to overcome the limitations.

A.5 Bibliography

[1] Creighton, C. J., Reid, J. G., and Gunaratne, P. H. Expression profiling of microRNAs by deep sequencing. *Briefings in Bioinformatics* **10**(5), 490–7 (2009).

[2] Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J. T., Nusbaum, J. D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., and Tuschl, T. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**(9), 1697–712 (2011).

[3] Linsen, S. E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., Fritz, B., Wyman, S. K., de Bruijn, E., Voest, E. E., Kuersten, S., Tewari, M., and Cuppen, E. Limitations and possibilities of small RNA digital gene expression profiling. *Nature Methods* **6**(7), 474–6 (2009).

[4] Jayaprakash, A. D., Jabado, O., Brown, B. D., and Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research* **39**(21), e141 (2011).

[5] Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y., and Robb, G. B. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Research* **40**(7), e54 (2012).

[6] Alon, S., Vigneault, F., Eminaga, S., Christodoulou, D. C., Seidman, J. G., Church, G. M., and Eisenberg, E. Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Research* **21**(9), 1506–11 (2011).

[7] Kloosterman, W. P. and Plasterk, R. H. The diverse functions of microRNAs in animal development and disease. *Developmental Cell* **11**(4), 441–50 (2006).

[8] Schulte, J. H., Marschall, T., Martin, M., Rosenstiel, P., Mestdagh, P., Schlierf, S., Thor, T., Vandesompele, J., Eggert, A., Schreiber, S., Rahmann, S., and Schramm, A. Deep sequencing reveals differential expression of microRNAs in favorable versus unfavorable neuroblastoma. *Nucleic Acids Research* **38**(17), 5919–28 (2010).

[9] Farazi, T. A., Horlings, H. M., Ten Hoeve, J. J., Mihailovic, A., Halfwerk, H., Morozov, P., Brown, M., Hafner, M., Reyal, F., van Kouwenhove, M., Kreike, B., Sie, D., Hovestadt, V., Wessels, L. F., van de Vijver, M. J., and Tuschl, T. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer Research* **71**(13), 4443–53 (2011).

[10] Han, Y., Chen, J., Zhao, X., Liang, C., Wang, Y., Sun, L., Jiang, Z., Zhang, Z., Yang, R., Li, Z., Tang, A., Li, X., Ye, J., Guan, Z., Gui, Y., and Cai, Z. MicroRNA expression signatures of bladder cancer revealed by deep sequencing. *PLOS One* **6**(3), e18286 (2011).

[11] Kato, M., de Lencastre, A., Pincus, Z., and Slack, F. J. Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development. *Genome Biology* **10**(5), R54 (2009).

[12] Newman, M. A., Mani, V., and Hammond, S. M. Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA* **17**(10), 1795–803 (2011).

[13] Martinez, J. and Tuschl, T. RISC is a $5'$ phosphomonoester-producing RNA endonuclease. *Genes & Development* **18**(9), 975–980 (2004).

[14] Zhuang, F., Fuchs, R. T., and Robb, G. B. Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. *Journal of Nucleic Acids* **2012**, 360358 (2012).

[15] Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., Pawlowski, T. L., Laub, T., Nunn, G., Stephan, D. A., Homer, N., and Huentelman, M. J. Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods* **5**(10), 887–93 (2008).

[16] Srivastava, S. and Chen, L. A two-parameter generalized Poisson model to improve the analysis of RNA-seq data. *Nucleic Acids Research* **38**(17), e170 (2010).

[17] Willenbrock, H., Salomon, J., Sokilde, R., Barken, K. B., Hansen, T. N., Nielsen, F. C., Moller, S., and Litman, T. Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. *RNA* **15**(11), 2028–34 (2009).

[18] Yin, S., Ho, C. K., and Shuman, S. Structure-function analysis of T4 RNA ligase 2. *The Journal of Biological Chemistry* **278**(20), 17601–8 (2003).

[19] Sorefan, K., Pais, H., Hall, A. E., Kozomara, A., Griffiths-Jones, S., Moulton, V., and Dalmay, T. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**(1), 4 (2012).

[20] Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Research* **32**(Database issue), D109–11 (2004).

[21] Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* **34**(Database issue), D140–4 (2006).

[22] Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36**(Database issue), D154–8 (2008).

[23] Kozomara, A. and Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* **39**(Suppl 1: Database issue), D152–7 (2011).

[24] Kwok, C. K., Ding, Y., Sherlock, M. E., Assmann, S. M., and Bevilacqua, P. C. A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation. *Analytical Biochemistry* **435**(2), 181–6 (2013).

[25] Viollet, S., Fuchs, R. T., Munafo, D. B., Zhuang, F., and Robb, G. B. T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnology* **11**, 72 (2011).

[26] Maroney, P. A., Chamnongpol, S., Souret, F., and Nilsen, T. W. A rapid, quantitative assay for direct detection of microRNAs and other small RNAs using splinted ligation. *RNA* **13**(6), 930–6 (2007).

[27] Kawano, M., Kawazu, C., Lizio, M., Kawaji, H., Carninci, P., Suzuki, H., and Hayashizaki, Y. Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing. *BioTechniques* **49**(4), 751–5 (2010).

Appendix B

Methods

B.1 Oligonucleotides

Oligonucleotides were synthesized by Integrated DNA Technologies (IDT) and purified by either IDT (RNase-Free HPLC purification) or in-house HPLC or denaturing PAGE followed by an ethanol precipitation. Native PAGE purification was used to isolate duplexes for mechanisms requiring a duplex scRNA reactant.

Native PAGE duplex purification: duplexes were annealed (90°C for 3 minutes followed by a controlled cooling of −1°C per minute to 23°C in a PCR block) and run on a 15% native polyacrylamide gel. The duplex band was excised from the gel using UV shadowing. For oligonucleotides labeled with ^{32}P, the duplex band was located by a short exposure on an image plate (Fujifilm type BAS-MS) and scanned using an FLA-5100 imaging system (Fujifilm). A full-size printout was used as a map to excise the appropriate bands. Duplexes were eluted in 1× duplex buffer (100 mM potassium acetate, 20 mM HEPES, pH 7.5) overnight by gentle rotation, filtered, and frozen. For M2, it was critical to maintain the proper secondary structure and the duplexes were stored at 4°C.

Oligonucleotide concentrations were determined using A_{260} absorbance on a NanoDrop 8000 (Thermo Scientific) and extinction coefficients provided by IDT. For duplexes, the sum of the extinction coefficients was used. Prior to each reaction, hairpins were snap cooled (95°C for 90 seconds, 30-second incubation on ice, and room temperature incubation of at least 30 minutes). Complexes, except for M2 A·B, were annealed by heating to 90°C for 3 minutes followed by a controlled cooling of −1°C per minute to 23°C in a PCR block. DsRed2, GAPDH, and d2EGFP mRNAs were generated by *in vitro* transcription. Since mRNA concentrations were approximate, they were used at a concentration two times that of the short target. The mRNAs were heated

to 65°C for 5 minutes and cooled at room temperature for at least 30 minutes prior to use in the reactions.

Table B.1: Oligonucleotide sequences for Chapter 2

Mechanism 1: conditional DsiRNA formation using stable scRNAs

Strand	Domains	Sequence
M1 X_s	s1-a-b-c-s2	5′-CUGGACAUCACCUCCCACAACGAGGACUA-3′
M1 A	c*-b*-a*-z*-y*	5′-GUUGUGGGAGGUGAUGUCGGGUGUU-3′
M1 B	x-y-z-a-b	5′-CACUACCAGCAGAACACCCGACAUCACCU-3′
M1 C	w-x-y-s-a*-z*-y*-x*-w*	5′-ACCACUACCAGCAGAACAAGGUAGAUGUCGGGUGUUCUGCUGGUAGUGGU-3′

Mechanism 2: conditional shRNA formation using a single stable scRNA

Strand	Domains	Sequence
M2 X_s	a-b-c	5′-UGGGAGCGCGUGAUGAACUUCGAGGACGG-3′
M2 A	z-c*-b*-a*	5′-UUCAUCUGCACCACCGGCACCGUCCUCGAAGUUCAUCACGCGCUCCCA-3′
M2 B	z-c*-b-c-z*-y*	5′-UUCAUCUGCACCACCGGCACCGAUGAACUUCGAGGACGGUGCCGGUGGUGCAGAUGAACU-3′

Mechanism 3: conditional DsiRNA formation via template-mediated 4-way branch migration

Strand	Domains	Sequence
M3 X_s	a-b-c-d-e	5′-CUCCGAGAACGUCAUCACCGAGUUCAUGCGCUUCAAGG-3′
M3 A	e*-d*-z*-y*	5′-CCUUGAAGCGCAUGAACUGACACGCUGAACUUGUGGCCG-3′
M3 B	y-z-b*-d	5′-CGGCCACAAGUUCAGCGUGUCUGACGUAGUUCAU-3′
M3 C	x-y-z-c*-b*-a*	5′-AACGGCCACAAGUUCAGCGUGUCCGGUGAUGACGUUCUCGGAG-3′
M3 D	b-z*-y*-x*	5′-ACGUCAGACACGCUGAACUUGUGGCCGUU-3′

Table B.2: Oligonucleotide sequences for Chapter 3

Strand	Sequence
DsiRNA GS (AS)	5′-GUCAGACACGCUGAACUUGUGGCCGUU-3′
DsiRNA PS (SS)	5′-CGGCCACAAGUUCAGCGUGUCUGAC-3′
M1-V2 X_s	5′-GACAUCACCUCCCACAACGAGGACUACACCAU-3′
M1-V2 A	5′-GUAGUCCUCGUUGUGGGAGGUGAUGUCGGGUG-3′
M1-V2 B	5′-CACUACCAGCAGAACACCCGACAUCACCUCCC-3′
M1-V2 C	see M1 C in Table B.1
M2 X_s	see M2 X_s in Table B.1
M2 A	see M2 A in Table B.1
M2 B	see M2 B in Table B.1

Table B.3: Oligonucleotide sequences for Chapter 4

Strand	Sequence
DsiRNA 1: SS	5′-AGCUGACCCUGAAGUUCAUCUGCAC-3′
DsiRNA 1: AS	5′-GUGCAGAUGAACUUCAGGGUCAGCUUG-3′
DsiRNA 2: SS	5′-AAGUUCAGCGUGUCCGGCGAGGGCG-3′
DsiRNA 2: AS	5′-CGCCCUCGCCGGACACGCUGAACUUAC-3′
DsiRNA 3: SS	5′-UUCAUCUGCACCACCGGCAAGCUdGdC-3′
DsiRNA 3: AS	5′-GCAGCUUGCCGGUGGUGCAGAUGAACU-3′
DsiRNA 4: SS	5′-CGGCCACAAGUUCAGCGUGUCUGAC-3′
DsiRNA 4: AS	5′-GUCAGACACGCUGAACUUGUGGCCGUU-3′
DsiRNA DsRed: SS	5′-CCGAGAACGUCAUCACCGAGUUCdAdT-3′
DsiRNA DsRed: AS	5′-AUGAACUCGGUGAUGACGUUCUCGGAG-3′
shRNA	5′-AGCUGACCCUGAAGUUCAUredGGUGACCAAGGCACCAUGAACUUCAGGGUCAGCUUG-3′
Hairpin C1	5′-CAAGCUGACCCUGAAGUUGUGCAGAUGAACUUCAGGGUCAGCUUG-3′
Hairpin C1m	5′-CAAGCUGACCCUGAAGUUGUGCAGAUGAACUUCAGGGUCAGCUUG-3′
Hairpin C	5′-ACCACUACCAGCAGAACAGAUGUCGGGUGUUCUGCUGGUAGUGGU-3′
Hairpin C3	5′-AGCUGACCCUGAAGUUGGACAGAGAUGAACUUCAGGGUCAGCU-3′
Hairpin C3: 100% 2′-O-Me	5′-AGCUGACCCUGAAGUUGGACAGAGAUGAACUUCAGGGUCAGCU-3′
Hairpin C4	5′-dTdTCUGACCCUGAAGUUGGACAGAGAUGAACUUCAGGGUCAGCU-3′
Hairpin C5	5′-CAAGCUGACCCUGAAGGGACAGAUGAACUUCAGGGUCAGCUUG-3′
CI GS	5′-ACUCCCAUCCACUAUAUGAACUUCAGGGUCAGCUUG-3′
CI PSmod	5′-AGCUGACCCUGAAGUUCAU-3′
CI PS	5′-AGCUGACCCUGAAGUUCAUAUAGUGGAUGGGAGU-3′
M1 A	5′-GUUGUGGGAGGUGAUGUCGGGUG-3′
M1 B	5′-CACUACCAGCAGAACACCCGACAUCACCU-3′
M1 C	5′-InvddT-ACCACUACCAGCAGAACAGAUGUCGGGUGUUCUGCUGGUAGUGGU-3′
M3 A	see M3 A in Table B.1
M3 B	see M3 B in Table B.1
M3 C	see M3 C in Table B.1
M3 D	see M3 D in Table B.1
M3 J	5′-CUUCAGAUGAACUUCAGGGUCAGCU-3′
M3 K	5′-AGCUGACCCUGAAGUUCAUCUCGCGUGA-3′
M3 L	5′-CAAGCUGACCCUGAAGUUCAUCUGAAG-3′
M3 M	5′-UCACGCGAGAUGAACUUCAGGGUCAGCUUG-3′
H1	5′-CUGACCCUGAAGUUGAUGAACUUCAGGGUCAGCUUG-3′
H2	5′-GAUGAACUUCAGGGUCAGCUUGCUGACCCUGAAGUU-3′
H3	5′-CUGACCCUGAAGUUUCCUAACUUCAGGGUCAGCUUG-3′
H4	5′-CUGACCCUGAAGUUGAUGAACUUCAGGGUCAGCUUG-3′
H5	5′-CUGACCCUGAAGUUGAUGAACUUCAGGGUCAGCUUG-3′
H6	5′-UCAGCGUGUCCGGCCCUCGCCGGACACGCUGAACUU-3′
H-neg	5′-GGUUGCAAUGAUCUUGAUGAAGAUCAAGAUCAUUGC-3′
GeneRacer Oligo	5′-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3′
GeneRacer 5′ primer (DNA)	5′-CGACTGGAGCACGAGGACACTGA-3′
GeneRacer 5′ nested primer (DNA)	5′-GGACACTGACATGGACTGAAGGAGTA-3′
RT primer (DNA)	5′-CCGTCGCCGATGGGGGTGTTC-3′
Reverse primer (DNA)	5′-GCACGCTGCCGTCCTCGATGTTG-3′
Reverse nested primer (DNA)	5′-TCCAGCTTGTGCCCCAGGATGTT-3′

Table B.4: Oligonucleotide sequences for Appendix A.

Strand	Material	Sequence
H7	RNA	5′-AAAAGAAAGGUAAUUAUGAUUACCUUUCUUUUGCCC-3′
H8	RNA	5′-GAACCCUUCUUAUGUCUGCAUAAGAAGGGUUCUGCU-3′
H9	RNA	5′-CAUAAUUACCUUUCUUUUGGGCAAAAGAAAGGUAAU-3′
H10	RNA	5′-CAGACAUAAGAAGGGUUCAGCAGAACCCUUCUUAUG-3′
GS-D	RNA	see DsiRNA 1:AS in Table B.3
PS-D	RNA	see DsiRNA 1:SS in Table B.3
GS-s	RNA	5′-AUGAACUUCAGGGUCAGCUUG-3′
PS-s	RNA	5′-AGCUGACCCUGAAGUUCAUCU-3′
H1	RNA	see H1 in Table B.3
hsa-miR-16-5p	RNA	5′-UAGCAGCACGUAAAUAUUGGCG-3′
universal miRNA cloning linker (NEB)	DNA	5′-rAppCTGTAGGCACCATCAAT-NH_2-3′
3′ splint: GS-D, GS-s, H1	DNA	5′-ATTGATGGTGCCTACAGCAAGCT-3′
3′ splint: miR-16	DNA	5′-ATTGATGGTGCCTACAGCGCCAA-3′
5′ linker: 2.RNA acceptor oligo	RNA	5′-CAAUCCCUACCCUACCACUUCACCC-3′
5′ splint: GS-D	DNA	5′-CTGCACGGGTGAAGTGGTAGGGTAGGGATTG-3′
5′ splint: GS-s	DNA	5′-GTTCATGGGTGAAGTGGTAGGGTAGGGATTG-3′
5′ splint: H1	DNA	5′-GGTCAGGGGTGAAGTGGTAGGGTAGGGATTG-3′
5′ splint: miR-16	DNA	5′-CTGCTAGGGTGAAGTGGTAGGGTAGGGATTG-3′
5′ universal linker:splint pool[1]	DNA:RNA	5′-PO_4-NNNNNNGATCGTCGGACTGTAGAACTCrGrUrUrCrA rGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC-3′
3′ universal linker:splint pool[1]	RNA:DNA	5′-rArUrArGrCrArUrUrCrGrUrArUrGrCrCrGrUrCrUrUrC rUrGrCrUrUrGAGAAGACGGCATACGAATGCTATNNNNNN-3′
RT primer[1]	DNA	5′-CAAGCAGAAGACGGCATACGA-3′
PCR primer 2[1]	DNA	5′-CAAGCAGAAGACGGCATACGA-3′

The 5′ and 3′ universal linker:splint pools and primers used in the method proposed in Appendix A to reduce bias in small RNA libraries for next-generation sequencing are compatible with Illumina instruments.[1]

B.2 mRNA plasmids and sequences

DsRed2

pTnT-DsRed2 construction: the DsRed2 mRNA coding sequence was amplified from pDsRed2-C1 (Clontech, catalog number 632407) and directionally cloned into pTnT vector (Promega, catalog number L5610).

DsRed2 mRNA

```
  1 AUGGCCUCCU CCGAGAACGU CAUCACCGAG UUCAUGCGCU UCAAGGUGCG CAUGGAGGGC ACCGUGAACG
 71 GCCACGAGUU CGAGAUCGAG GGCGAGGGCG AGGGCCGCCC CUACGAGGGC CACAACACCG UGAAGCUGAA
```

[1]Oligonucleotide sequences

```
141 GGUGACCAAG GGCGGCCCCC UGCCCUUCGC CUGGGACAUC CUGUCCCCCC AGUUCCAGUA CGGCUCCAAG
211 GUGUACGUGA AGCACCCCGC CGACAUCCCC GACUACAAGA AGCUGUCCUU CCCCGAGGGC UUCAAGUGGG
281 AGCGCGUGAU GAACUUCGAG GACGGCGGCG UGGCGACCGU GACCCAGGAC UCCUCCCUGC AGGACGGCUG
351 CUUCAUCUAC AAGGUGAAGU UCAUCGGCGU GAACUUCCCC UCCGACGGCC CCGUGAUGCA GAAGAAGACC
421 AUGGGCUGGG AGGCCUCCAC CGAGCGCCUG UACCCCCGCG ACGGCGUGCU GAAGGGCGAG ACCCACAAGG
491 CCCUGAAGCU GAAGGACGGC GGCCACUACC UGGUGGAGUU CAAGUCCAUC UACAUGGCCA AGAAGCCCGU
561 GCAGCUGCCC GGCUACUACU ACGUGGACGC CAAGCUGGAC AUCACCUCCC ACAACGAGGA CUACACCAUC
631 GUGGAGCAGU ACGAGCGCAC CGAGGGCCGC CACCACCUGU UCCUGAGAUC UCGAGCUCAA GCUUCGAAUU
701 CUGCAGUCGA CGGUACCGCG GGCCCGGGAU CCACCGGAUC UAGAUAA
```

d2EGFP The d2EGFP mRNA transcript used for all gels consisted of EGFP fused to a destabilizing tail (d2EGFP).

pGEM-T easy-d2EGFP construction: the d2EGFP mRNA coding sequence was cloned from cells expressing d2EGFP (generous gift from Dr. Beisel) based on the pd2EGFP-1 (Clontech, catalog number 6008-1) sequence and cloned into pGEM-T easy vector (Promega, catalog number A1360).

d2EGFP mRNA

```
  1 AUGGUGAGCA AGGGCGAGGA GCUGUUCACC GGGGUGGUGC CCAUCCUGGU CGAGCUGGAC GGCGACGUAA
 71 ACGGCCACAA GUUCAGCGUG UCCGGCGAGG GCGAGGGCGA UGCCACCUAC GGCAAGCUGA CCCUGAAGUU
141 CAUCUGCACC ACCGGCAAGC UGCCCGUGCC CUGGCCCACC CUCGUGACCA CCCUGACCUA CGGCGUGCAG
211 UGCUUCAGCC GCUACCCCGA CCACAUGAAG CAGCACGACU UCUUCAAGUC CGCCAUGCCC GAAGGCUACG
281 UCCAGGAGCG CACCAUCUUC UUCAAGGACG ACGGCAACUA CAAGACCCGC GCCGAGGUGA AGUUCGAGGG
351 CGACACCCUG GUGAACCGCA UCGAGCUGAA GGGCAUCGAC UUCAAGGAGG ACGGCAACAU CCUGGGGCAC
421 AAGCUGGAGU ACAACUACAA CAGCCACAAC GUCUAUAUCA UGGCCGACAA GCAGAAGAAU GGCAUCAAGG
491 UGAACUUCAA GAUCCGCCAC AACAUCGAGG ACGGCAGCGU GCAGCUCGCC GACCACUACC AGCAGAACAC
561 CCCCAUCGGC GACGGCCCCG UGCUGCUGCC CGACAACCAC UACCUGAGCA CCCAGUCCGC CCUGAGCAAA
631 GACCCCAACG AGAAGCGCGA UCACAUGGUC CUGCUGGAGU UCGUGACCGC CGCCGGGAUC ACUCUCGGCA
701 UGGACGAGCU GUACAAGAAG CUUAGCCAUG GCUUCCCGCC GGAGGUGGAG GAGCAGGAUG AUGGCACGCU
771 GCCCAUGUCU UGUGCCCAGG AGAGCGGGAU GGACCGUCAC CCUGCAGCCU GUGCUUCUGC UAGGAUCAAU
841 GUGUAG
```

GAPDH

pGEM-T easy-GAPDH construction: the GAPDH mRNA coding sequence was cloned from HEK 293 cells and cloned into pGEM-T easy vector (Promega, catalog number A1360).

GAPDH mRNA

```
  1 AUGGGGAAGG UGAAGGUCGG AGUCAACGGA UUUGGUCGUA UUGGGCGCCU GGUCACCAGG GCUGCUUUUA
 71 ACUCUGGUAA AGUGGAUAUU GUUGCCAUCA AUGACCCCUU CAUUGACCUC AACUACAUGG UUUACAUGUU
141 CCAAUAUGAU UCCACCCAUG GCAAAUUCCA UGGCACCGUC AAGGCUGAGA ACGGGAAGCU UGUCAUCAAU
```

```
211 GGAAAUCCCA UCACCAUCUU CCAGGAGCGA GAUCCCUCCA AAAUCAAGUG GGGCGAUGCU GGCGCUGAGU
281 ACGUCGUGGA GUCCACUGGC GUCUUCACCA CCAUGGAGAA GGCUGGGGCU CAUUUGCAGG GGGGAGCCAA
351 AAGGGUCAUC AUCUCUGCCC CCUCUGCUGA UGCCCCCAUG UUCGUCAUGG GUGUGAACCA UGAGAAGUAU
421 GACAACAGCC UCAAGAUCAU CAGCAAUGCC UCCUGCACCA CCAACUGCUU AGCACCCCUG GCCAAGGUCA
491 UCCAUGACAA CUUUGGUAUC GUGGAAGGAC UCAUGACCAC AGUCCAUGCC AUCACUGCCA CCCAGAAGAC
561 UGUGGAUGGC CCCUCCGGGA AACUGUGGCG UGAUGGCCGC GGGGCUCUCC AGAACAUCAU CCCUGCCUCU
631 ACUGGCGCUG CCAAGGCUGU GGGCAAGGUC AUCCCUGAGC UGAACGGGAA GCUCACUGGC AUGGCCUUCC
701 GUGUCCCCAC UGCCAACGUG UCAGUGGUGG ACCUGACCUG CCGUCUAGAA AAACCUGCCA AAUAUGAUGA
771 CAUCAAGAAG GUGGUGAAGC AGGCGUCGGA GGGCCCCCUC AAGGGCAUCC UGGGCUACAC UGAGCACCAG
841 GUGGUCUCCU CUGACUUCAA CAGCGACACC CACUCCUCCA CCUUUGACGC UGGGGCUGGC AUUGCCCUCA
911 ACGACCACUU UGUCAAGCUC AUUUCCUGGU AUGACAACGA AUUUGGCUAC AGCAACAGGG UGGUGGACCU
981 CAUGGCCCAC AUGGCCUCCA AGGAGUAA
```

B.3 mRNA *in vitro* transcription

The plasmids were linearized by digestion with the following restriction enzymes (New England Biolabs): NotI (pTnT-DsRed), AatII (pGEM-T easy-d2EGFP), and SphI-HF (pGEM-T easy-GAPDH). The linearized plasmids were *in vitro* transcribed according to the manufacturer with incubation times ranging from 2 hours to overnight using either the T7-Scribe Standard RNA IVT kit (CELLSCRIPT, Inc.) for pTnT-DsRed or the SP6-Scribe Standard RNA IVT kit (CELLSCRIPT, Inc.) for pGEM-T easy-d2EGFP and pGEM-T easy-GAPDH. Transcribed mRNA was purified using RNeasy Protect Mini Kit (Qiagen) according to the manufacturer. mRNA concentration was determined using A_{260} absorbance on a NanoDrop 8000 (Thermo Scientific) and approximate molecular weight for each transcript. The transcripts are believed to be slightly longer than the sequences listed due to additional sequences at the transcription start and termination sites.

B.4 Native polyacrylamide gel electrophoresis

Reactants were incubated at 0.5 μM each for two hours at 37°C in 1× duplex buffer (100 mM potassium acetate, 20 mM HEPES, pH 7.5). A reaction master mix was prepared for duplicates (e.g., isothermal incubation and anneal or '−' and '+' Dicer) and then split into two separate reactions. 2 pmol per strand (4 pmol mRNA) in 1× loading dye were separated by native PAGE. All gels were cast and run in 1× TBE. Native PAGE: 20% native polyacrylamide gels were run at 200 V for 8–10.5 hours unless otherwise specified. Gels were stained in 1× SYBR Gold (Life Technologies) for 10 minutes at room temperature and imaged using an FLA-5100 imaging system (Fujifilm). 5 μL siRNA markers (45 ng) were run as size markers in the Dicer processing gels. Prior

to testing the mechanisms, the concentration values were corrected relative to the appropriate X_s by varying the amount of each strand relative to X_s and annealing or incubating for 2 hours at 37°C. Separation by native PAGE allowed determination of the proper stoichiometry.

B.5 Quantification of ON and OFF states

Lanes used for quantification (ON and OFF states) were run on three separate days using different batches of snap-cooled and/or annealed reactants. Fujifilm Multi Gauge, version 3.0, software was used to calculate the SYBR Gold intensity lane profiles near the final product band. Plots were constructed in MATLAB using the profile data exported from Multi Gauge. For each profile, the peak was centered around the intensity maximum. The intensity values were normalized such that the highest profile intensity value for all lanes compared within a gel was set to one. The quantification percentages were calculated using either Multi Gauge (with auto detection of peaks and background) or using a MATLB script that subtracted the background. For the MATLAB script, the background was approximated by fitting a straight line between the outer 0.5 mm of the profile on both sides. The background area was then subtracted from the profile area. The signal was normalized to the ON state with X_s as the detection target. Based on quantifying the gels shown in the main text six times each, we believe that the error in the measurement of any gel is approximately 0.5%. This error should not be confused with the experimental variation between gels, which is significantly higher.

B.6 *In vitro* Dicer assay

Dicer reactions were performed using the Recombinant Human Turbo Dicer Enzyme kit (Genlantis, catalog number T520002) according to the manufacturer with modifications. The reactions were performed at 0.5 μM in 10 μL with enough Dicer enzyme to cut approximately all of the final substrate after 2 hours at 37°C. The following amounts of Turbo Dicer were used for the Dicer reactions: M1: 0.5 units; M2: 0.5 units; M3: 1 unit. Dicer, target and reactants were mixed simultaneously (i.e., the reactants were not pre-incubated with their target prior to addition of Dicer). Reactions were stopped by the addition of the appropriate loading dye and siRNA formation was verified by native PAGE separation.

B.7 Tissue culture

Human embryonic kidney (HEK) 293 cells and variants expressing fluorescent proteins were used for all tissue culture work. The HEK 293 cell line with stable d2EGFP expression was provided by Chase Beisel[1] and was generated using the pd2EGFP-1 plasmid (Clontech, catalog number 6008-1) with expression driven by the CMV promoter. The HEK 293 cell line with stable d2EGFP and DsRed2 expression was derived from the HEK 293 d2EGFP cell line through the addition of DsRed2 from the pDsRed2-C1 plasmid (Clontech, catalog number 632407) and clonal isolation. DsRed2 expression is also driven by the CMV promoter. HEK 293T cells were used for cell lysate experiments. Cells were grown as a monolayer in DMEM (Life Technologies, catalog number 11995-065) supplemented with 10% FBS (Life Technologies) in a humidified incubator at 37°C and 5% CO_2.

B.8 RNA transfection and d2EGFP knockdown analysis

Lipofectamine RNAiMAX (Life Technologies) was used for all transfections unless otherwise noted. Reverse transfections were performed in 24-well plates seeded with 100,000 cells in 0.5 mL DMEM with 10% FBS. For each well, 6–60 pmol RNA per strand were diluted in 50 μL Opti-MEM I medium (Life Technologies) and 1.5 μL RNAiMAX was diluted in 50 μL Opti-MEM I medium. Master mixes were made for all replicates. The two volumes were mixed gently and incubated at room temperature for 15 minutes. 100 μL transfection complexes were added per well and mixed gently. Mock transfection complexes contained buffer and the transfection reagent but no RNA. For split transfections, all volumes were split equally between the transfection complexes. Mock transfections contained no RNA. Cells were incubated at 37°C with 5% CO_2 for 24 hours unless otherwise noted. Dicer siRNA transfections were scaled for 10-cm plates.

HiPerFect (Qiagen) transfections were carried out as described above with the following modifications: 4.5 μL HiPerFect transfection reagent was used for each well, and the transfection complexes were mixed by vortexing before a $\sim$7 minutes incubation at room temperature.

d2EGFP knockdown was assayed using the Accuri C6 flow cytometer. Side- and forward-scatter was used to identify live cells. d2EGFP was excited with the 488 nm laser and fluorescence intensity at 530$\pm$15 nm was recorded. Reported Y expression is the mean d2EGFP signal normalized by the mean d2EGFP signal for either untreated or mock transfected cells. Error bars denoted the standard deviation between the means of replicates. While this method of data presentation fails

to adequately capture bimodal populations, bimodal populations were generally only noticeable for poor transfections. For experiments containing a co-culture of HEK 293 d2EGFP and HEK 293 d2EFP + DsRed2, the cells were assigned to their original population based on DsRed2 signal. DsRed2 was excited with the 488 nm laser excitation and fluorescence above 670 nm was recorded.

Despite our best attempts to make the two cell lines similar, the d2EGFP expression level in the cell line expressing both fluorescent proteins was significantly lower. Likely, the CMV promoter used to express the fluorescent proteins is responsible. As a result, the d2EGFP signal could not be directly compared between cell lines. By normalizing the d2EGFP signal, we saw equivalent levels of knockdown between the two cell lines. For example, DsiRNA 1 reduced the d2EGFP signal in both cells lines to less than 10% of the untreated/mock levels.

B.9 Cell lysate

The cell lysis protocol was previously described by Sakurai et al.[2] Briefly, HEK 293T cells were grown to confluency in 10-cm dishes, trypsinized, harvested, and washed with PBS. The pelleted cells were resuspended in 0.5 mL of buffer D (20 mM HEPES, 100 mM KCl, 0.2 mM EDTA, 5% glycerol, 0.5 mM DTT, and 1× HALT protease inhibitor cocktail (Thermo Scientific)) and sonicated on ice (Misonix S-4000). The cells were sonicated at an amplitude of 30, 10 seconds on, 30 seconds off, for a total on time of 1 minute and energy of 300–400 J. The lysate was cleared by centrifugation at 13,000g for 15 minutes at 4°C, aliquoted, frozen in a dry ice/ethanol bath, and stored at −80°C. Lysate protein concentration was determined by a Bradford assay using the a NanoDrop 8000 (Thermo Scientific). Lysate concentrations ranged from 1.8–5.25 μg/μL.

Dicer-minus lysate: Cells were transfected with a pool of human Dicer siRNAs (Santa Cruz Biotechnology, Inc.) for three days prior to lysis. After cell lysis, the cells were immunodepleted, as previously described,[2] by incubation with 15 μg rabbit Dicer antibody (sc-30226, Santa Cruz Biotechnology, Inc.) at 4°C for 2 hours with gentle agitation. 150 μL of Protein A/G plus agarose beads (sc-2003, Santa Cruz Biotechnology, Inc.) were added and incubated with gentle agitation at 4°C overnight. The beads were removed by collecting the supernatant after centrifugation at 1000g for 5 minutes at 4°C.

B.10 Radioactive labeling of oligonucleotides

To generate the internally labeled hairpins, each hairpin was ordered as two fragments designed to strongly form the hairpin structure while locating the nick point within the predicted final siRNA. For both hairpins, this meant using a short 3′ fragment. The short 3′ fragment was 5′ end labeled with ^{32}P by [$\gamma-^{32}$P] ATP (10mCi/ml, MP Biomedicals) using T4 polynucleotide kinase (New England Biolabs). Additionally, the DsiRNA passenger strand and siRNA marker (New England Biolabs) were 5′ end labeled with [$\gamma-^{32}$P] ATP. Unincorporated [$\gamma-^{32}$P] ATP was removed by spin column chromatography using Illustra MicroSpin G-25 columns (GE Healthcare) according to the manufacturer.

After labeling the 5′ end of the short fragment, both fragments were annealed together. The fragments were then ligated with an excess of the short fragment using T4 RNA ligase 2 (New England Biolabs) at 37°C for 1 hour. The ligase and any remaining unligated short fragment was removed using an Oligo Clean and Concentrator column (Zymo Research) which removes oligonucleotides shorter than 16-nt. Counts were measured on a Beckman LS-5000TD Liquid Scintillation Counter.

B.11 *In vitro* Dicer processing in cell lysate

RNA duplexes were annealed, unless previously purified, and stored at 4°C, and hairpins were snap cooled prior to incubation. For the DsiRNA control, the 5′ end-labeled passenger strand was annealed to the unlabeled guide strand. 25 μg of total protein in buffer D was incubated with 50–500 fmol each of the RNA strands in a 20 μL reaction supplemented with 3.2 mM Mg^{2+}, 1× protease inhibitor cocktail (Thermo Scientific), and 1 U/μL RNase Inhibitor (Applied Biosystems). The reactions were incubated for two hours at 37°C. The reactions were stopped by adding 1.5 volumes of Proteinase K buffer (200 mM Tris-HCl at pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% wt/vol SDS)[3] and 1 μL (20 μg/μL) Proteinase K (NEB) for 30 minutes at 37°C. RNA was purified using Oligo Clean and Concentrator columns (Zymo). For each gel, a volume corresponding to a fixed amount of radioactive signal (∼10,000 cpm) per lane was separated by native PAGE. Radioactive gels were exposed overnight onto an image plate (Fujifilm type BAS-MS) and scanned using an FLA-5100 imaging system (Fujifilm).

B.12 5′ RACE

RNA was transfected in the HEK 293 d2EGFP cells as described in Section B.8, and total RNA was collected ∼24 hour post-transfection using a Zymo RNA MiniPrep Kit with on-column DNase digestion. Total RNA was eluted in 25 μL RNase-free water. 20 pmol GeneRacer Oligo (IDT) was ligated to the 5′ end of ∼1.1 μg total RNA in a 20 μL reaction with 1 μL T4 RNA ligase 1 (NEB), 1 mM ATP, and 10% PEG8000 in 1× RNA ligase buffer. The reaction was incubated at 37°C for 1 hour and purified with an RNeasy column (Qiagen). ∼80 pg ligated total RNA was reverse transcribed using Superscript III First-Strand Synthesis SuperMix (Life Technologies, Inc.) per the manufacturer's instructions. The cDNA was amplified by nested PCR using AccuPrime Pfx SuperMix (Life Technologies, Inc.). For the first PCR amplification, GeneRacer 5′ primer and reverse primer were used with a touchdown PCR program: 94°C for 2 min., 5 cycles: 94°C for 30 sec., 72°C for 1 min., 5 cycles: 94°C for 30 sec., 70°C for 1 min., 20 cycles: 94°C for 30 sec., 65°C for 30 sec., 68°C for 1 min., followed by a 10 min. extension at 68°C. The nested PCR used the GeneRacer 5′ nested primer and the reverse nested primer with the same touchdown program. The amplified products were separated in an agarose gel. The PCR products were either sent directly for sequencing or were ligated in the pGEM-T Easy Vector after A-tailing according to the manufacturer's instructions. The plasmid was transformed into 5-α Competent *E. coli* cells and grown overnight. Individual clones were cultured, mini-prepped, and sequenced by Laragen.

B.13 Total RNA isolation for small RNA libraries

RNA was transfected into HEK 293 d2EGFP cells as described in Section B.8 with a final concentration of 100 nM for the 4–14–4 hairpins. For the TC71 and U87MG cell lines, the cells (initial cell seeding: ∼800,000 cells/well for TC71, ∼94,000 cells/well for U87MG) were grown overnight in 12-well plates before transfection. Cells were transfected according to the manufacturer's instructions with the following conditions: TC71 cells were transfected with 12 μL HiPerFect per well to give a final RNA concentration of 45.5 nM, and U87MG cells were transfected with 1.5 μL Oligofectamine per well to give a final RNA concentration of 100 nM. Cells were lysed directly on the plate using 0.5 mL TRIzol Reagent (Life Technologies, Inc.) by triturating approximately 10 times and then incubating at room temperature for 5 minutes. The lysate was transferred to 1.5 mL Eppendorf tubes and 0.1 mL chloroform was added. The tubes were agitated vigorously for 15 seconds and then incubated at room temperature for 2–3 minutes. The samples were centrifuged

at 12,000g for 15 minutes at 4°C, followed by removal of 0.2 mL of the aqueous phase. 0.2 mL H_2O and 0.4 mL chloroform were added for a second extraction where the samples were centrifuged at 12,000g for 10 minutes at 4°C. The aqueous phase was removed and the RNA was precipitated by the addition of 4 μg glycogen and 0.25 mL isopropanol. The samples were incubated at room temperature for 10 minutes before centrifugation at 12,000g for 10 minutes at 4°C. The liquid was decanted and the pellet was washed with 0.5 mL 90% ethanol by vortexing and centrifugation at 7,500g for 5 minutes at 4°C. The pellets were allowed to dry at room temperature and the total RNA was resuspended in 12 μL H_2O by heating to 55°C for 10 minutes.

B.14 Next-generation sequencing and data analysis

Small RNA libraries were prepared by the City of Hope DNA Sequencing/Solexa core. Small RNAs, ranging from 19–60-nt, were isolated; however, the upper limit was likely ∼57-nt to avoid including endogenous small non-coding RNAs. The samples were multiplexed with up to 16 samples per sequencing lane. The miRNA-seq protocol was run on the Illumina HiSeq 2000 and produced 60-nt + 7-nt single-end reads including a 6-nt 3′ barcode. The reads were trimmed to 57-nt by removing the barcode and 3-nt at the 5′-end.

The reads were mapped to the transfected RNA strands using code written in lab. Briefly, the 3′ adaptor was trimmed and the resulting sequences were searched for windows (minimum length of 10-nt) that mapped to the transfected RNA. Graphical representation of the results were provided by heat maps which depict the probability of a given nucleotide existing in sequences that map to at least 10-nt of the full-length transfected RNA oligonucleotide. Only sequences without mismatches were included in the analysis.

B.15 Small RNA ligation experiments

For the ligation experiments in Appendix A, the detailed ligation conditions are listed in the caption of each gel. For the radioactive gel in Figure A.4, refer to section B.10 for radioactive labeling. Reactions that included a hybridized prior to ligation were annealed in 0.25–1× hybridization solution. 1× hybridization solution: 300 nM NaCl, 20 nM Tris-HCl pH 8.0, 2 mM EDTA. For 5′ ligations, small RNA strands were phosphorylated with T4 PNK or ordered from IDT with a 5′-phosphate prior to addition to ligation reactions. T4 PNK reactions: 300 pmol RNA incubated with 50 units T4 PNK (NEB) in 1× T4 RNA ligase 2 reaction buffer in a 50 μL reaction at 37° for

1 hour and heat inactivated at 65° for 20 minutes.

The following is the current version of the proposed small RNA sequencing protocol shown in Figure A.5 in Appendix A. A 9 μL reaction containing 5 fmol hsa-miR-16-5p and 100 pmol each of the 5′ and 3′ universal splint pools in 0.33× hybridization solution was annealed. Upon cooling to room temperature, the ligation was initiated by the addition of 2,000 U T4 DNA ligase (NEB) in a 20 μL reaction with1× T4 DNA ligase buffer and 20% PEG8000. The reaction was carried out overnight at room temperature. The DNA portion of the universal splints was degraded at 37°C for 30 minutes by the addition of 2 U DNase 1 (NEB) to a 100 μL reaction supplemented with 0.5 mM $CaCl_2$ and 0.5 mM $MgCl_2$. The reaction was purified with an RNA Clean and Concentrator column (Zymo). Half of the purified reaction was reverse transcribed using SuperScript III First Strand SuperMix (Life Technologies, Inc.) per the manufacturer's instructions using 5 μM RT primer. One quarter of the cDNA was amplified using OneTaq 2× Master Mix (NEB) using 0.5 μM PCR primer 2 and 0.25 μM RT primer using a short PCR program: 94°C for 30 sec, 14 cycles: 94°C for 15 sec., 60°C for 30 sec., 68°C for 15 min., followed by a 5 min. extension at 68°C. The PCR products were separated by 10% native PAGE.

B.16 Bibliography

[1] Beisel, C. L., Bayer, T. S., Hoff, K. G., and Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology* **4**, 224 (2008).

[2] Sakurai, K., Amarzguioui, M., Kim, D. H., Alluin, J., Heale, B., Song, M. S., Gatignol, A., Behlke, M. A., and Rossi, J. J. A role for human Dicer in pre-RISC loading of siRNAs. *Nucleic Acids Research* **39**(4), 1510–25 (2011).

[3] Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., and Sharp, P. A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & Development* **13**(24), 3191–7 (1999).

Appendix C

Supplementary figures

C.1 Supplementary figures for Chapter 2

Quantification of conditional Dicer substrate formation

To characterize variability in scRNA signal transduction, gels used for the quantification of ON and OFF states were run on three separate days, preparing reactants each day as described above in Appendix B. ON states: short RNA detection target X_s, mRNA detection target X. OFF states: no target, mRNA silencing target Y, mRNA off-target Z. The calculated values for ON and OFF states were normalized to the ON state using the short detection target X_s. The uncertainty in quantifying any given gel is estimated to be less than 0.5%. This gel quantification uncertainty is significantly smaller than the variability observed between the three independent reaction replicates for each mechanism.

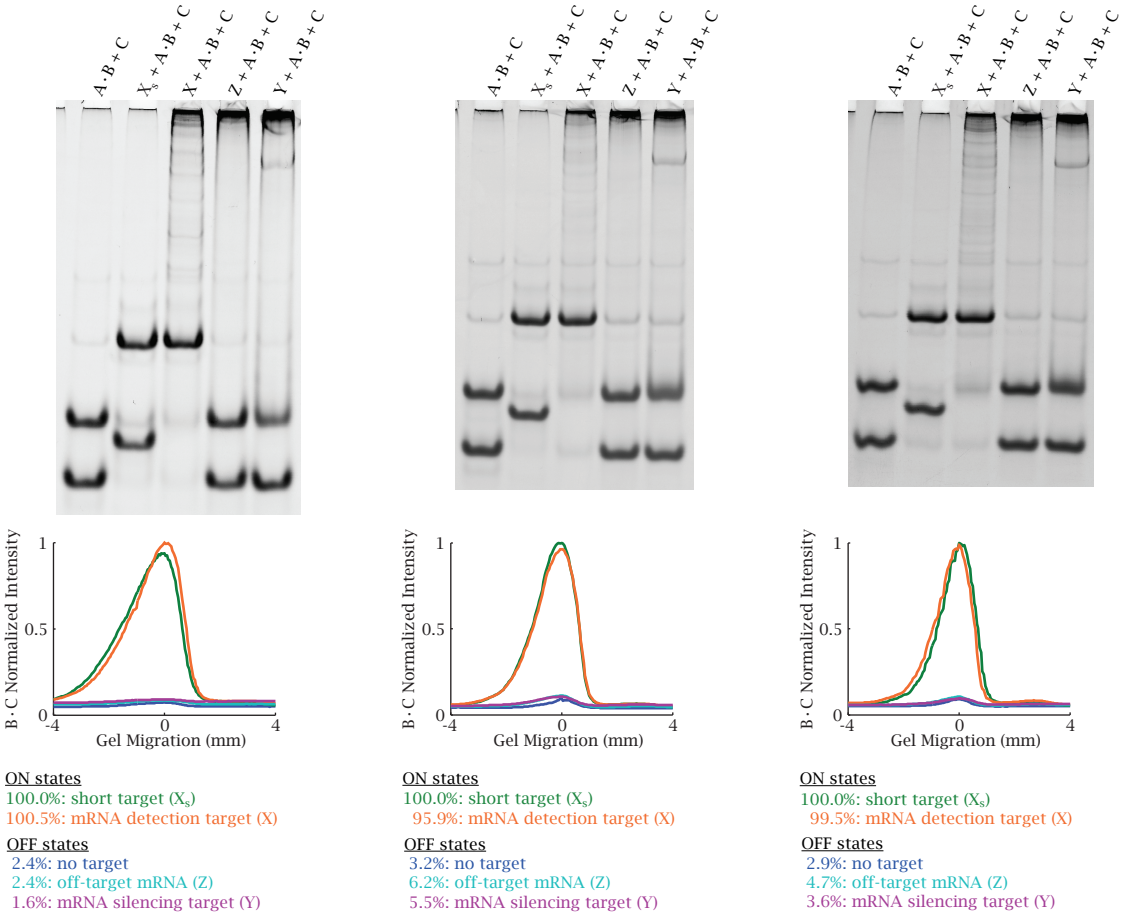


Figure C.1: Quantification of Mechanism 1 ON and OFF states. Three independent experiments examining the final product B·C formation in the ON and OFF states of conditional DsiRNA formation using stable scRNAs. ON states (X_s and DsRed2 mRNA target X) and OFF states (no target, GAPDH mRNA target Z and silencing d2EGFP mRNA target Y) of the mechanism. The OFF states show minimal B·C production. The amount of Dicer substrate B·C formed was quantified relative to the formation of B·C for the ON state with X_s.

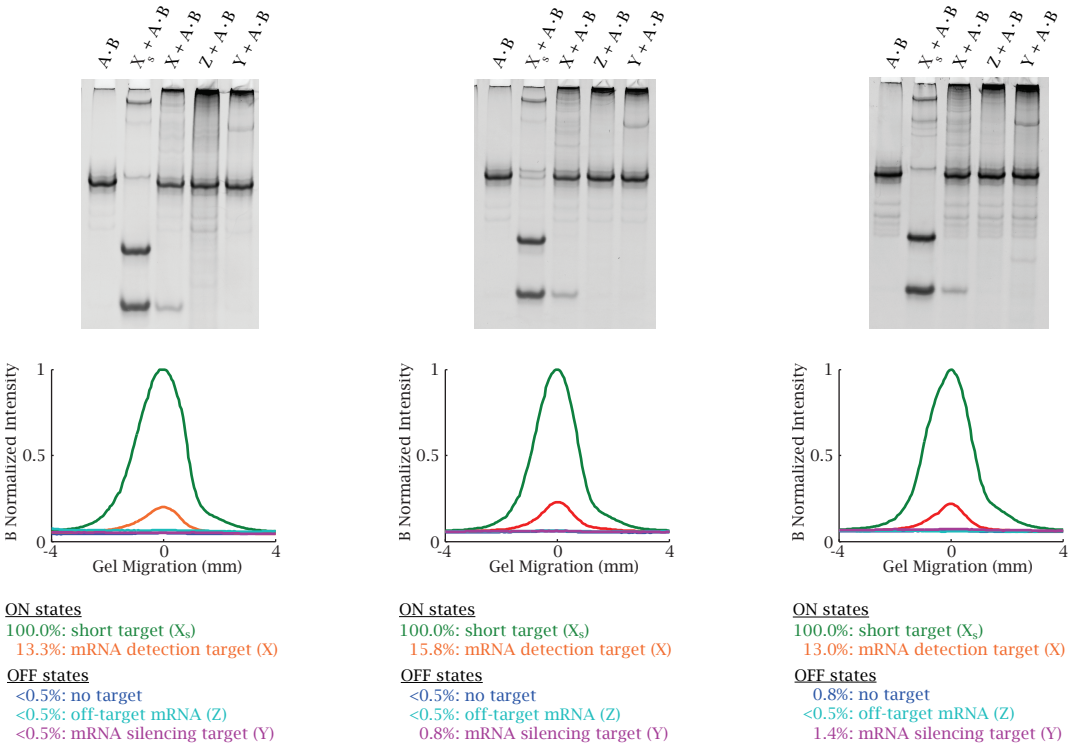


Figure C.2: Quantification of Mechanism 2 ON and OFF states. Three independent experiments examining the final product B formation in the ON and OFF states of conditional shRNA formation. ON states (X_s and DsRed2 mRNA target X) and OFF states (no target, GAPDH mRNA target Z and silencing d2EGFP mRNA target Y) of the mechanism. The OFF states show very low or undetectable levels of B production. The amount of Dicer substrate B formed was quantified relative to the formation of B for the ON state with X_s.

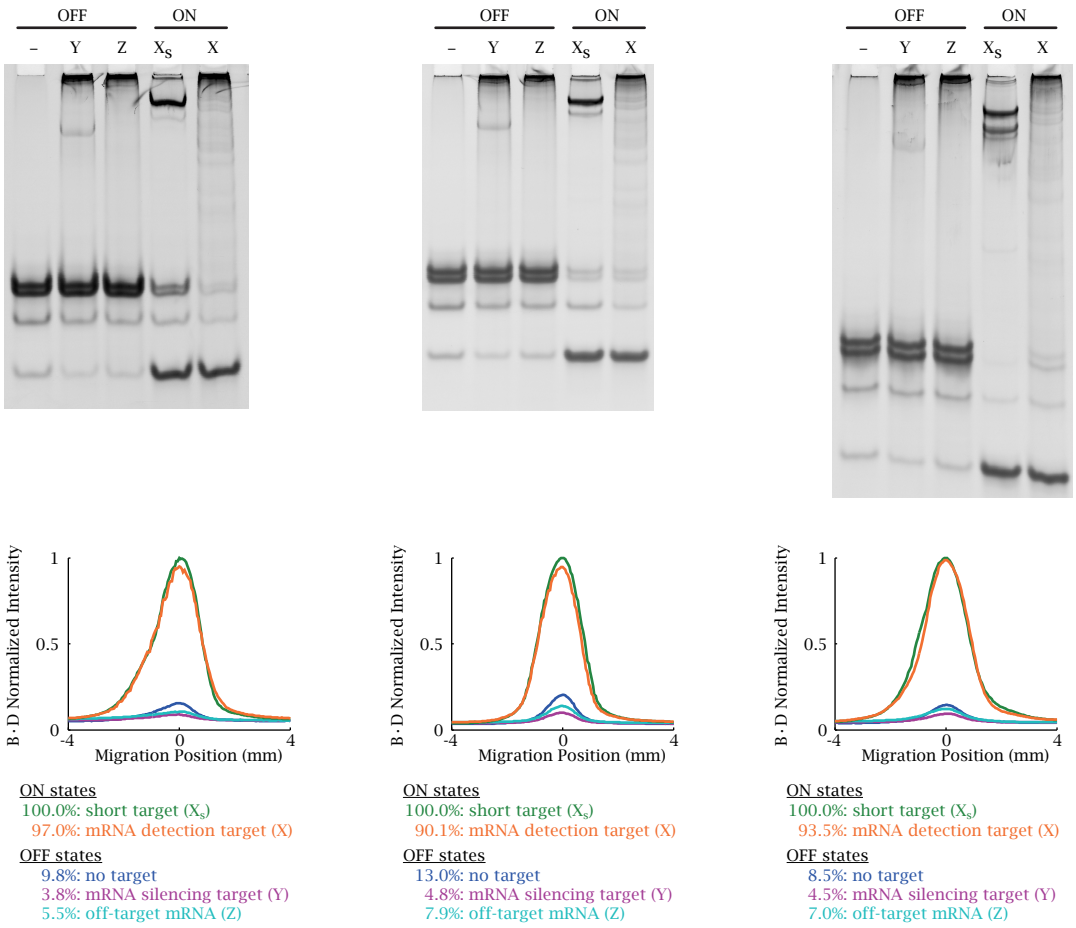


Figure C.3: Quantification of Mechanism 3 ON and OFF states. Three independent experiments examining the final product B·C formation in the ON and OFF states of conditional DsiRNA formation via template-mediated 4-way branch migration. ON states (X_s and DsRed2 mRNA target X) and OFF states (no target, GAPDH mRNA target Z and silencing d2EGFP mRNA target Y) of the mechanism. The OFF states show very low or undetectable levels of B·D production. The amount of Dicer substrate B·D formed was quantified relative to the maximum formation of B·D or the ON state with X_s.

C.2 Supplementary figures for Chapter 4

C.2.1 Investigation of scRNA saturation of RNAi pathway

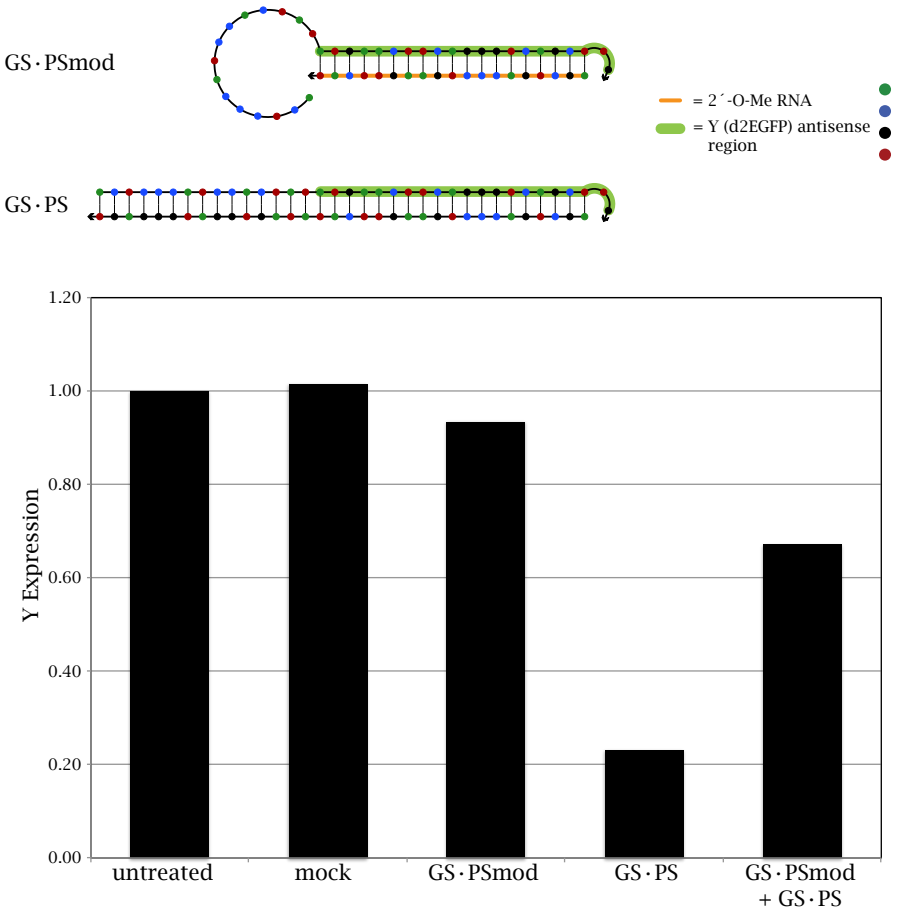


Figure C.4: Demonstration of competitive inhibition of the RNAi pathway using the duplexes GS·PSmod and GS·PS. Modifying one strand of the Dicer substrate duplex with 2′-O-Me prevents the duplex GS·PSmod from silencing Y. When a longer, RNA version of PS is used in the Dicer substrate GS·PS, the duplex silences Y efficiently. When both duplexes are transfected (GS·PSmod was transfected 3.5 hours prior to GS·PS transfection), the observed silencing is approximately the average of the silencing observed for each duplex individually. GS is the same for both duplexes. Green indicates the region complementary to Y, and an orange backbone indicates 2′-O-Me modification. Y expression is the normalized mean Y expression for 1 well. These results are representative of multiple experiments exploring competitive inhibition. The transfection was carried out in the Y cell lines as described in Section B.8 of Appendix B but with only 80,000 cells/well and a final RNA concentration of 20 nM per strand. Gene knockdown was assayed 24 hours post-transfection.

C.2.2 Mechanism 2: Conditional shRNA formation using a single stable scRNA

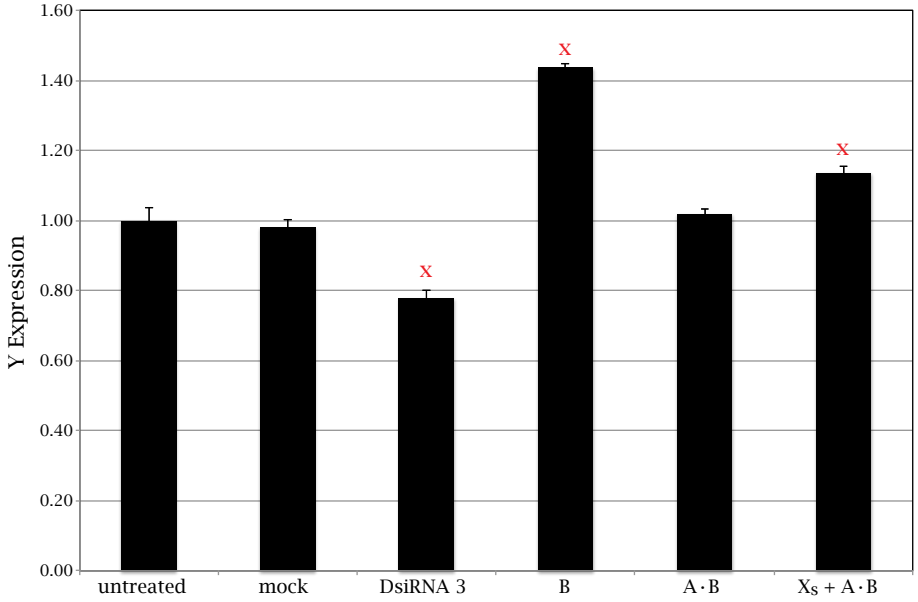


Figure C.5: Transfection of M2 final product B, scRNA A·B, or X_s + A·B into the Y cell line produced very little knockdown of Y expression. DsiRNA 3, a positive control that targets the same region in Y as shRNA B, produced little gene knockdown (∼20%), indicating that the region is a poor choice for gene silencing. A red 'x' indicates a condition for which Y silencing is expected. scRNA A·B should be OFF in this cell line because X is absent. The transfection was carried out as described in Section B.8 of Appendix B with a final RNA concentration of 20 nM per strand. Gene knockdown was assayed 24 hours post-transfection. Error bars were calculated using three replicate wells.

C.2.3 Short hairpin silencing

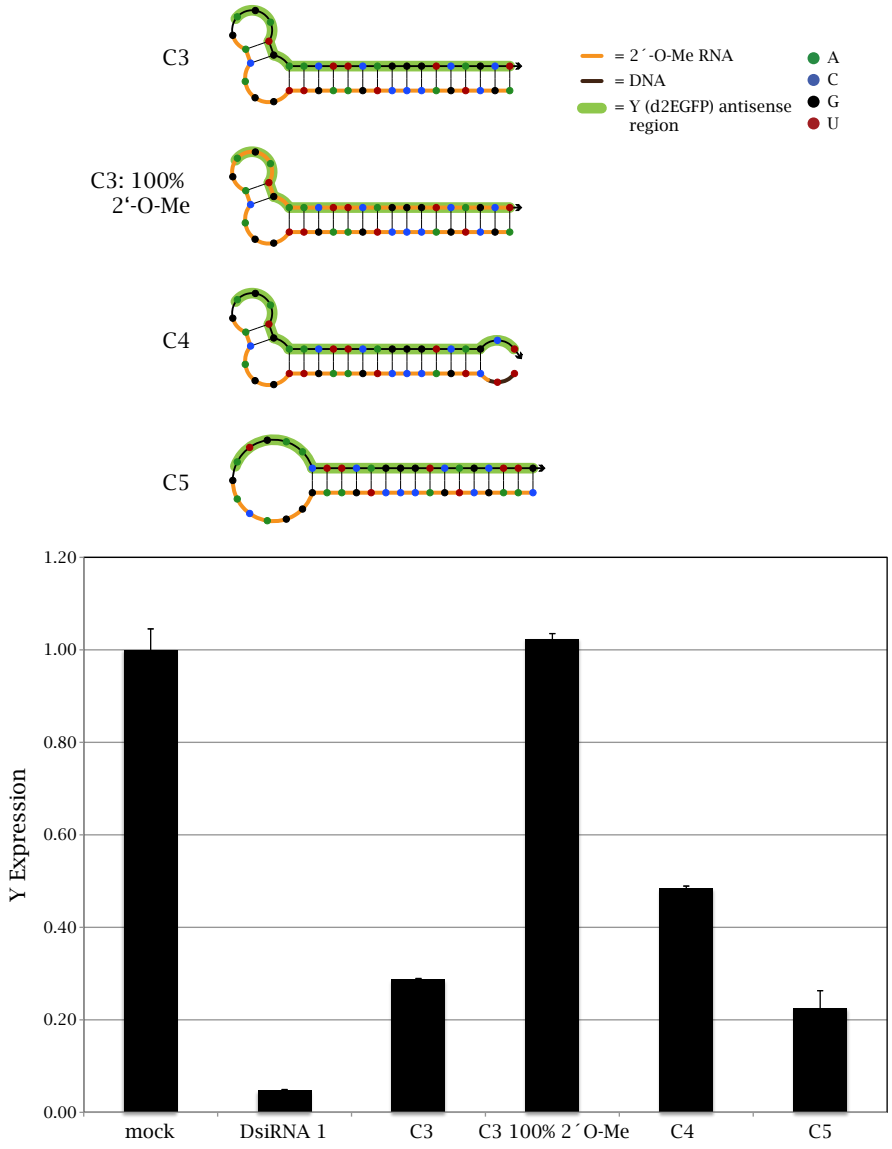


Figure C.6: Transfection of short hairpins produced varying degrees of Y knockdown. C3 and C3: 100% 2′-O-Me have identical sequences and vary only by the location of 2′-O-Me RNA nucleotides. Modifying the entire hairpin with 2′-O-Me prevents Y knockdown for at least 24 hours. C4 is nearly identical to C3 except that the end of the hairpin is forked instead of blunt and the last 2-nt of the 5′-end are DNA. This alteration reduces the silencing ability of the hairpin. C5 has the same dimensions and modification pattern as C3 but has an entirely different sequence. The sequence change slightly improves the silencing of hairpin C5 compared to C3. Green indicates the region complementary to Y, and an orange backbone indicates 2′-O-Me modification. The transfection was carried out as described in Section B.8 of Appendix B with a final RNA concentration of 100 nM per strand. Gene knockdown was assayed 24 hours post-transfection. Error bars were calculated using two replicate wells.

C.2.4 DsiRNA-induced toxicity

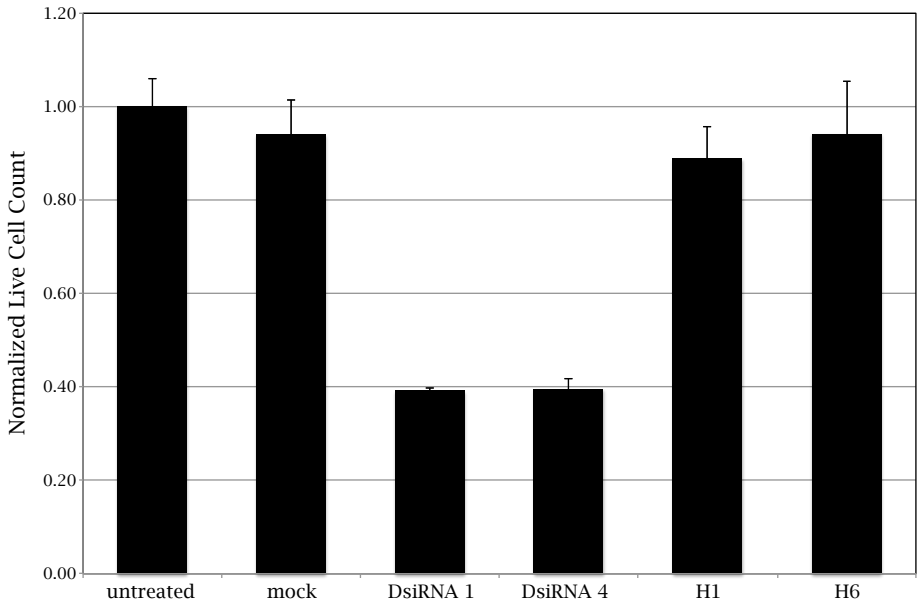


Figure C.7: DsiRNA-induced toxicity. RNAiMAX transfections of DsiRNA 1 or DsiRNA 4 show significantly fewer live cells 72 hours post-transfection than untreated or mock transfected cells. Interestingly, the 4–14–4 hairpins H1 and H6 do not show toxicity. DsiRNA 1 and H1 target the same region in Y. The Y cell line was transfected as described in Section B.8 of Appendix B with a final RNA concentration of 10 nM per strand. Normalized live cell count was determined by counting the live cells (identified by the forward- and side-scatter) in a fixed volume (30 μL) on the flow cytometer and normalizing to the untreated live-cell count. Error bars indicate the standard deviation of the normalized live-cell count in three replicate wells.