Specificity and Mechanism of microRNAs in the Regulation of Gene Expression

by

John Gerard Doench

B.A. Hamilton College
Clinton, NY, 2000

Submitted to the
Department of Biology
in partial fulfillment of the
requirements for the degree of
Doctorate of Philosophy
at the
Massachusetts Institute of Technology
June 2005

Signature of Author.....

John G. Doench
Department of Biology
May 18, 2005

Certified by.....

Phillip A. Sharp
Institute Professor of Biology
Thesis Supervisor

Accepted by.....

Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

Specificity and Mechanism of microRNAs in the Regulation of Gene Expression
by
John Gerard Doench

Submitted to the Department of Biology on May 18, 2005
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

ABSTRACT

Originally thought of as a molecular inferior to its nucleic acid cousin DNA, RNA has more recently been appreciated as an equal partner in biology, a molecule at the heart of many crucial cellular reactions and perhaps the first molecule of life itself. The discovery of RNA interference (RNAi) has further emphasized the importance of RNA-based processes in the regulation of gene expression. One arm of the RNAi response uses a large class of endogenous, small RNA species termed microRNAs (miRNAs). The establishment of a mammalian tissue culture system has allowed for investigation of both the mechanism and specificity of miRNA-directed translational repression.

The term RNAi can be used to encompass a wide variety of gene silencing phenomena. The canonical RNAi pathway, as first described by Fire and colleagues in *C. elegans* and studied biochemically in *Drosophila* by Tuschl, Zamore, and colleagues, is a post-transcriptional mechanism of gene silencing, in which short, interfering RNAs (siRNAs) guide the cleavage of complementary mRNAs. Endogenous miRNAs are similar to siRNAs, and the two pathways, siRNA-directed mRNA cleavage and miRNA-directed translational repression, share common protein components yet lead to different outcomes. Our results indicate that the distinct outcome of these pathways is largely determined by the interaction of the small RNA species with its mRNA target. Additionally, variation in the number of miRNA binding sites shows that miRNAs can act synergistically to enhance repression activity.

Further experimentation into the specificity of miRNAs revealed that the miRNA does not simply basepair to its target mRNA but rather that regions of the miRNA contribute differently to translational repression activity. The 5' region of the miRNA, the first ~8 nucleotides, is necessary and sufficient for target recognition. The 3' region can contribute significantly to activity, however, in cases where the 5' region has less-than-optimal complementarity. Multiple miRNAs can regulate a single mRNA, and the degree of translational repression is dependent on the expression level of both the miRNA and the mRNA. These results indicate that miRNAs are capable of regulating a substantial percentage of the genome and thus are integral factors in the control of gene expression. Finally, the observation that miRNAs can direct mRNA cleavage, albeit inefficiently, offers promise for finding endogenous miRNA targets and understanding the scope of miRNA-directed regulation of gene expression.

Thesis Supervisor: Phillip A. Sharp
Title: Institute Professor of Biology

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

My mentors and colleagues

I'd like to thank Dave Bartel and Frank Solomon for serving on my committee for four years, and for being consistently helpful and encouraging. Also, the entire fifth floor of the Cancer Center has been an incredible place to work, socially and scientifically, and I'd especially like to thank Dave Housman and Tyler Jacks for their leadership in encouraging that atmosphere. In the Sharp lab, I thank Mauro, Amanda, Amy, Dean, Hristo, Alla, Joel, Will, Mike Murray, Chonghui, KB, Derek, and Carl for their help and tolerance. Finally, Margarita has always been there with a kind word and a smile, and there's no way to thank her enough.

Chris Petersen joined the lab the same time as I did, and it is not hyperbole to say that I would not have made it through my Ph.D without him. Being thrown to the wolves is much less scary if you're with someone else, and Chris has helped me handle inevitable setbacks and frustration with humor, intelligence, and camaraderie. I am truly lucky to have him as a colleague and friend.

Phil has been an excellent mentor. Grad school can be hellishly frustrating, but I can honestly say that never once did I blame any lack of results or sense of failure on him; I think that says something about how indebted I feel to him. He encouraged independence, but he also inspired with his obvious continued love of science. Never one to be labeled warm-and-fuzzy, Phil nevertheless clearly cared about my work and my progression as a scientist, always putting what was best for me ahead of what got results the fastest. I thank him for his continued commitment to training, and for restraining from smiting me with a thunderbolt for all the many jokes and insults I hurled his way.

My friends

My classmates Phil, Erik, Andy, Seth, and Chris, for Guys' Nights that kept me sane even as they made me drunk. Y'argh! The Pits, for summers of fun on the diamond and in the Muddy. Two championships in four years, let's get one more. Sean, for lazy summer afternoons as they should be, with a BBQ fired up and a beer in hand. Matt, Brian, Chris, and Ian, for weekends in Saratoga. Keara, my rock in lab and out, in the hallways or at the Thirsty and the Burren; you have no idea how much you've helped. Janet, my innocence; Lynne, my compassion; Lizzy, my perspective; Beth, my conscience; Amy, my inspiration.

My family

Whatever my parents couldn't teach me, or, more likely, whenever I was too stubborn to listen, I could rely on everyone else to point me in the right direction. Aunt Jenny, for being a safe haven during those hard days of middle school, and again during those hard days of grad school. Uncle Fran, the ultimate warrior, for never taking the easy way out, for teaching me to finish what you start, and always, always, give it your all. Uncle Sean, the coach, for showing me how shut up and do your job, and teaching me that being an adult means never blaming anyone else for the choices you've made. Aunt Mary, for never letting me take myself too seriously, and for Claire, Tommy, Petie, Timmy, and Billy. Uncle Tim, for forever making me laugh, and further emphasizing that I shouldn't take myself too seriously. Jane, for being the first to tell me when I'm being a jerk, and for always making me proud. Grandpa Sheehan, the founder of this feast.

Jed, maybe now, equipped with a Ph.D., I can figure out how to stop the onsides kicks of those '85 Bears. Dad, after four years of college and five years at MIT, I can still say you're the most thoughtful person I know. I think I'm still more of a game show host than a scientist, though, and unfortunately for Dawson and Downey, I don't do anything better than I play softball. Mom, can you believe they won?

And Catherine…

It ain't no sin to be glad you're alive

-- Bruce Springsteen

Chapter One

RNA interference & microRNAs

RNA interference

Seminal Early Discoveries

The RNAi pathway was first characterized in the nematode worm *C. elegans*. Attempts at using antisense RNA to silence genes of interest were complicated by the observation that sense RNA preparations oftentimes also mimicked loss-of-function alleles (Guo and Kemphues 1996). This contradiction was resolved when Fire, Mello, and colleagues reasoned that a small amount of contaminating double-strand RNA (dsRNA) may be triggering an endogenous silencing pathway. They tested this hypothesis experimentally and showed the existence of an endogenous pathway in which genes homologous to introduced dsRNA are silenced post-transcriptionally, a process they dubbed RNA interference (Fire et al. 1998).

RNAi was quickly shown to be an evolutionarily conserved response in *Neurospora* (Cogoni and Macino 1999), *Drosophila* (Kennerdell and Carthew 1998; Tuschl et al. 1999), and mammals (Wianny and Zernicka-Goetz 2000; Elbashir et al. 2001a). Insights into the mechanism of RNAi was first obtained in *Drosophila* embryo lysates. Using a dual-luciferase reporter assay, Tuschl, Zamore, and colleagues demonstrated an *in vitro* system that recapitulated the key aspects of RNAi, namely the sequence-specific degradation of mRNA triggered by input dsRNA (Tuschl et al. 1999). When they radiolabeled their input dsRNA, they saw that it was first processed into smaller fragments of 20 – 25 nucleotides (Zamore et al. 2000). These small RNAs were true intermediates in the pathway, as they could be excised from a gel and used to trigger the degradation of mRNA. This small RNA species immediately called to mind recent work from the plant literature, where it was shown that the process of co-suppression, or the silencing of multiple copies of transgenes inserted into the genome, correlated with the appearance of small RNA species (Hamilton and Baulcombe 1999). Chemical characterization of these small RNAs generated by the *Drosophila* lysate revealed that

they were double-stranded with 5' phosphates and 3' hydroxyls, and they were duplexed in a staggered fashion, such that each 3' end had a two nucleotide overhang (Zamore et al. 2000; Elbashir et al. 2001b). This species was named a short, interfering RNA (siRNA).

The Hannon laboratory performed a great deal of early work characterizing the protein components of the RNAi pathway. Using classical column chromatography and biochemical fractionation in cultured *Drosophila* S2 cells, Hammond and colleagues purified the enzymatic activity of the effecter complex of RNAi, which they named the RNA-induced Silencing Complex (RISC) (Hammond et al. 2000). Further purification identified Argonaute-2 (Ago-2) as a component of RISC (Hammond et al. 2001), a finding that complemented a genetic screen performed in *C. elegans* that had found members of this family of proteins to be important for a productive RNAi response (Tabara et al. 1999). To find the activity that processed the input dsRNA to smaller RNAs, the Hannon lab took a candidate gene approach, focusing on RNase III enzymes in the *Drosophila* genome, as the structure of siRNAs suggested that they arose from RNase III processing. Tagged constructs were assayed for the ability to process long, dsRNA into siRNAs, followed by an RNAi-on-RNAi experiment, in which the candidate protein was knocked down with RNAi and the ability of the cells to perform RNAi was then assayed. These experiments confirmed the central role of an RNase III enzyme named Dicer in the RNAi pathway (Bernstein et al. 2001).

These seminal experiments provide the basic framework for the canonical RNAi pathway (Figure 1a, p. 31; Table 1, p. 32). Long, dsRNA is processed by Dicer into siRNAs. One strand is chosen for incorporation into RISC, and the siRNA is unwound. RISC then uses the small RNA as a guide to find mRNAs with perfect complementarity and effects their cleavage, resulting in a rapid degradation of the mRNA and silencing of the gene. This section will discuss

the details of the mechanisms and components of RNAi pathway, largely as defined in the systems presented above.

* * *

Dicer

A central protein in the RNAi pathway is the RNase III enzyme Dicer, necessary for processing long dsRNA into siRNAs, as well as processing mature miRNAs from their precursor hairpins (Bernstein et al. 2001; Hutvagner et al. 2001; Zhang et al. 2004). The human and *C. elegans* genome encodes a single Dicer protein while organisms such as *Arabidopsis* and *Drosophila* have more than one Dicer. In these cases, the different Dicers have evolved specialized functions. For example, Dcr-1 in *Drosophila* is generally reserved for the processing of microRNAs, while Dcr-2 is involved in the canonical RNAi pathway (Lee et al. 2004c). Knockout models reveal that Dicer is essential. In *C. elegans,* homozygous null alleles of Dicer are sterile and survive to adulthood only because of maternal rescue (Grishok et al. 2001; Knight and Bass 2001). Additionally, a Dicer knockout mouse dies during early embryonic development (Bernstein et al. 2003), and mouse embryonic stem cells lacking Dicer proliferate more slowly compared to wild type cells, and cannot differentiate *in vitro* (Kanellopoulou et al. 2005).

Dicer is necessary and sufficient to cleave long dsRNA into siRNAs, although Dicer interacts with several other proteins *in vivo*. On a molecular level, the best-studied Dicer is mammalian Dicer, which has two RNase III domains, a dsRNA binding motif, and a PAZ domain (Bernstein et al. 2001). Site-directed mutagenesis studies have made clear the mechanism of the RNase III domains in siRNA and miRNA processing (Zhang et al. 2004). The two RNase III domains are positioned ~21 bp upstream of one end of the dsRNA substrate, with

one on each strand of the dsRNA. Conserved aspartic and glutamic acid residues in each domain catalyze cleavage of the phosphodiester bond to produce a 3' hydroxyl and a 5' phosphate. The two RNase III domains are positioned in a staggered fashion such that the product of Dicer cleavage is not a blunt RNA but rather a dsRNA with a two nucleotide overhang on the 3' end.

siRNAs

The products of Dicer cleavage, siRNAs, are 21-23 nucleotide RNAs, originally identified as dsRNA species with 3' overhangs of two nucleotides and a phosphate on each 5' end (Zamore et al. 2000; Elbashir et al. 2001b). Although these molecules are geometrically symmetric, they are not symmetric in terms of sequence, and only one strand of an siRNA becomes incorporated into RISC (Nykanen et al. 2001; Martinez et al. 2002). This functional asymmetry is largely predicted by the basepairing interactions of the first few nucleotides on each side of the duplex, as the strand with its 5' end on the side with less thermal stability (more A-U pairs than G-C pairs) will be chosen as the guide strand of RISC (Figure 1b, p. 31) (Khvorova et al. 2003; Schwarz et al. 2003). siRNAs themselves can be introduced into cells to trigger the RNAi pathway, a technique now widely used to study gene function in mammalian cells (Elbashir et al. 2001a). This strand preference should be followed in designing siRNAs, but there are certainly more criteria, both known and unknown, that further contribute to the functionality of siRNAs (Reynolds et al. 2004).

RISC: components and assembly

Assembly of the RISC has been best characterized in *Drosophila* lysates. Dicer and the protein R2D2 (homologous to rde-4 in *C. elegans)* bind the siRNA, with R2D2 binding nearest

the 5' end of the passenger strand and Dicer nearest the 5' end of the guide strand (Tabara et al. 2002; Liu et al. 2003; Tomari et al. 2004b). Recognition by R2D2 is dependent on a 5' phosphate, thus ensuring that only authentic siRNAs can assemble into RISC, although exogenous, unphosphorylated siRNAs introduced into cells are rapidly phosphorylated by an endogenous kinase activity (Nykanen et al. 2001; Tomari et al. 2004b). The next step in RISC assembly requires recruitment of several more proteins, one of which is a member of the Argonaute family. Finally, the siRNA must be unwound to create a RISC programmed for cleavage of mRNA, a step that appears to require the helicase *armitage,* at least in *Drosophila* ovaries (Tomari et al. 2004a). Various RISCs have been described depending on the fractionation scheme (Hammond et al. 2000; Nykanen et al. 2001; Martinez et al. 2002; Pham et al. 2004). The largest, sedimenting at ~80S, appears to associate with ribosomes (Pham et al. 2004); indeed, components of RISC have been found associated with the L5 and L11 ribosomal proteins as well as the 5S rRNA (Ishizuka et al. 2002). The smallest purified complex, known as "minimal RISC," sediments as if it contains a single Argonaute protein with few if any accessory proteins (Martinez and Tuschl 2004). Recently, RISC activity has been reconstituted from only recombinant Ago-2 expressed in *E. coli* and a single-stranded guide RNA (Rivas et al. 2005).

At the core of RISC is a member of the Argonaute family of proteins, which is further subdivided into the Ago subfamily and the Piwi subfamily (Carmell et al. 2002). The number of Argonautes varies between organisms, ranging from one in *S. Pombe* to ~27 in *C. elegans;* humans have four from the Ago subfamily and four from the Piwi subfamily. The Ago subfamily has been directly implicated in RNAi, as human Ago-2 is the core component of RISC (Hammond et al. 2001; Martinez et al. 2002; Liu et al. 2004; Song et al. 2004). hAgo-1, 3, and 4 have been shown to bind exogenous siRNAs, but do not have cleavage activity (Liu et al. 2004;

Meister et al. 2004); it is likely that these Agos are involved in related pathways, such as miRNA-directed translational repression.

Argonaute proteins are characterized by two domains, the PAZ and PIWI domains (and are thus sometimes known as 'PPD' proteins). The crystal structure of the PAZ domain shows that it contacts 3' overhangs, implicating this domain in recognizing and loading siRNAs into RISC (Lingel et al. 2003; Yan et al. 2003). Crystal structures have revealed that Argonaute itself performs the endonucleolytic cleavage of target mRNAs. The PIWI domain contacts the 5' region of the siRNA guide strand bound to target mRNA, an A-form helix, and positions catalytic aspartate residues in a RNase H-like fold near the scissile phosphate of the target mRNA (Liu et al. 2004; Song et al. 2004). Furthermore, contacts with the 5' phosphate of the guide strand of the siRNA are essential for function (Ma et al. 2005; Parker et al. 2005). These crystal structures also make clear why human Ago-1 and Ago-4 cannot cleave target mRNAs, as both have a mutation in the DDH catalytic triad; the lack of cleavage activity by Ago-3, however, remains unexplained on a structural level (Rivas et al. 2005).

Although RISC activity only requires Ago-2 and a small RNA *in vitro,* less-stringent purifications and co-immunoprecipitation approaches have revealed other proteins that associate with cleavage activity, suggesting that these proteins are not directly involved in mRNA cleavage, but perhaps assist in RISC assembly and regulate RISC activity (Table 1, p. 32). Studies of RISC assembly in *Drosophila* lysates indicate that Dicer itself is a part of RISC (Pham et al. 2004), although Dicer is not needed for exogenous siRNA-mediated cleavage in mammalian cells (Kanellopoulou et al. 2005). Other components of RISC include the Vasa-intronic gene (VIG), identified via biochemical purification of RISC activity from *Drosophila* S2 cells (Caudy et al. 2002). Knockdown of this protein via RNAi reduces RNAi activity, although

no particular activity has been ascribed to VIG. Likewise, the micrococcal nuclease protein Tudor-SN associates with RISC activity in *Drosophila* lysates, as well as small RNAs in both *C. elegans* and cultured human cells (Caudy et al. 2003). Tudor-SN is not the enzyme involved in the endonucleolytic cleavage of the target mRNA, as this nuclease family produces 2' 3' cyclic phosphates, but may instead play a role in degrading the cleaved transcript.

A particularly intriguing component of RISC is the *Drosophila* homolog of the Fragile X Mental Retardation protein (FMRP), found to associate with Ago-2 (Caudy et al. 2002; Ishizuka et al. 2002). FMRP, an RNA binding protein with some sequence specificity, is known to associate with polyribosomes (Corbin et al. 1997; Darnell et al. 2001; Ceman et al. 2003; Stefani et al. 2004), and the mental retardation associated with loss of this protein's activity in humans is consistent with its role in regulating translation of particular mRNAs at neuronal synapses (Antar and Bassell 2003). Analysis of RNAi *in vitro* has shown that FMRP is not required for mRNA cleavage activity, and knockdown of FMRP has little-if-any effect on RNAi activity in cell culture (Caudy et al. 2002; Ishizuka et al. 2002). It has been hypothesized that FMRP helps provide specificity for localization of RISC or RISC-like complexes to specific mRNAs for other activities, such as translational repression. Accordingly, FMRP has been shown to repress translation of mRNAs, and mice lacking FMRP have altered polysome distributions of a large number of mRNAs (Brown et al. 2001).

Target Cleavage and Degradation

Biochemical systems have allowed detailed analysis of the mechanism of RISC cleavage. The guide strand of the siRNA pairs to the mRNA and directs an endonucleolytic cleavage between the nucleotides opposite the 10^{th} and 11^{th} nucleotides of the guide strand of the siRNA,

as measured from the 5' end (Llave et al. 2002; Martinez et al. 2002; Martinez and Tuschl 2004). This cleavage produces a 3' hydroxyl on the 5' end of the cleaved mRNA and a 5' phosphate on the 3' end (Martinez and Tuschl 2004; Schwarz et al. 2004). Studies of siRNAs mismatched to their targets have shown that RISC can tolerate some degree of mismatch in the 5' region of the siRNA, and slightly more mismatch in the 3' region, although these changes lead to decreased cleavage kinetics (Haley and Zamore 2004). These results obtained *in vitro* correlate well with large scale microarray analyses of "off-target" effects of siRNAs (Jackson et al. 2003). Transfected siRNA have been shown to down-regulate, on average, a few dozen mRNAs, with many of these transcripts bearing complementarity to the siRNA, especially the 5' region. This off-target effect at the mRNA level is generally less than two fold, and the degree of effect at the protein level has not been well-studied (Lim et al. 2005). For experiments using siRNAs to study gene function, an siRNA of a different sequence should be used to confirm that a phenotype is due to knockdown of the gene of interest rather than an off-target effect; because the siRNA has a different sequence, the off-target profile should also be different. Importantly, these off-target effects are tolerated in transgenic animals expressing shRNAs against genes of interest (Rubinson et al. 2003; Tiscornia et al. 2004; Ventura et al. 2004; Grimm et al. 2005).

The products of RISC cleavage are mRNAs no longer capable of productive translation, and are rapidly degraded. The 5' cleavage product is degraded 3'-to-5' by the exosome (Orban and Izaurralde 2005). There is one report that the targets of miRNA-directed cleavage in plants have several uridines added to the 3' end of the 5' cleavage product, although the significance of this observation is unclear (Shen and Goodman 2004). The 3' cleavage product, which contains a 5' phosphate, is subject to degradation by the XRN family of 5'-to-3' exonucleases (Souret et al. 2004; Orban and Izaurralde 2005). Both the exosome and XRN exonucleases have been

implicated in other RNA surveillance pathways, including the nonsense-mediate decay pathway (NMD) and the nonstop decay pathway (NSD) (van Hoof et al. 2002).

RNA-directed RNA polymerase

A significant difference between RNAi in plants, worms, and fungi from that in flies and mammals is the requirement for an RNA-dependent RNA polymerase (RdRP) activity in the former (Cogoni and Macino 1999; Smardon et al. 2000). RdRPs amplify the RNAi signal, and RdRP proteins are required for productive RNAi in these organisms (Fagard et al. 2000; Mourrain et al. 2000). siRNAs generated from Dicer cleavage could perform two functions in the RNAi pathway. First, the guide strand of the siRNA can enter RISC and perform mRNA target cleavage. Additionally, the guide strand of an siRNA may function as a primer, binding to the mRNA target and allowing polymerization of an RNA strand antisense to the mRNA. This now-dsRNA could serve as a substrate for Dicer, thereby amplifying the RNAi effect. RdRP activity was formally proven to be a part of the RNAi pathway in an experiment in *C. elegans* utilizing an *pha-4*:GFP fusion transgene, such that the mRNA would code for *pha-4* in its 5' half and GFP in its 3' half (Sijen et al. 2001). As expected, dsRNA homologous only to GFP silenced this transgene. This same dsRNA also silenced endogenous *pha-4*, implying that siRNAs generated against the GFP were then extended by an RdRP activity to make dsRNA homologous to *pha-4*. These siRNAs could now target the endogenously-encoded *pha-4* mRNAs, thus silencing *pha-4* in *trans* and leading to the term "transitive RNAi."

RdRP homologs have not been identified in either the mammalian or *Drosophila* genomes. Biochemical analysis has shown that siRNAs with modifications on their 3' ends, such that the 3' end could not serve as a primer for extending an RNA transcript, were still able

to support RNAi (Schwarz et al. 2002). The ability to amplify the RNAi signal may be important for RNAi both to persist for several generations in worms and to spread to other cells and tissues in worm and plant systems, phenomena not observed in flies and vertebrates (Fagard et al. 2000; Grishok et al. 2000; Winston et al. 2002).

Endogenous RNAi pathways

The overriding theme for the general function of RNAi is the silencing of potentially harmful genetic elements. For example, several mutants in the RNAi pathway in *C. elegans* were also identified in a screen for genes involved in silencing transposons (Ketting et al. 1999). Furthermore, in plant systems, it is clear that RNAi has evolved as a defense against viruses (Marathe et al. 2000). When a plant is infected with a virus, it uses the RNAi machinery first to silence the virus in the infected cells, and small RNAs homologous to the virus can be readily detected (Dalmay et al. 2000; Mourrain et al. 2000; Vance and Vaucheret 2001). Plants also mount a second response, sending a signal to other cells of the plant to render them immune to a second attack by the same virus. As would be expected in the evolutionary arms race between pathogen and host, some viruses have evolved mechanisms to defend against the endogenous RNAi response. For example, tombusvirus encodes the P19 protein that binds tightly to siRNAs, thus titrating them out of the cell and allowing viral replication (Vargason et al. 2003; Ye et al. 2003; Lakatos et al. 2004).

While RNAi was originally characterized as a post-transcriptional pathway, it has become clear that RNAi can also silence genes on the transcriptional level. *S. pombe* contains a single homolog each of Dicer, Argonaute, and RdRP, and cloning efforts revealed siRNAs that correspond to heterochromatic centromere repeats (Reinhart and Bartel 2002; Volpe et al. 2002).

In this system, the small RNA guides the RITS complex (RNA-induced initiation of transcriptional gene silencing) to homologous regions of the genome and induces transcriptional silencing via methylation of histone 3 lysine 9, which is then recognized by Swi6 (Volpe et al. 2003; Verdel et al. 2004). It has been hypothesized that the small RNA does not interact with DNA but instead with nascent transcripts, and that this interaction localizes a complex containing the RdRP (Motamedi et al. 2004). Work in *Drosophila* has reinforced the role of RNAi in transcriptional silencing, as *piwi* and *aubergine,* both Argonaute proteins, are required for HP1 (the Swi6 homolog) localization to heterochromatic regions and recruitment of the Polycomb repressive complex (Pal-Bhadra et al. 2002; Pal-Bhadra et al. 2004). The role for transcriptional silencing has been demonstrated in mammalian systems, as mouse ES cells deficient for Dicer show accumulation of major and minor satellite repeats (Kanellopoulou et al. 2005). Thus, in a wide-variety of organisms, RNAi appears to play a central role in organizing chromatin structure.

RNAi Technology

The ability to knockdown the expression of any gene of interest has made RNAi an important new technique in the biologist's toolkit. Long dsRNA can be used to trigger RNAi in the model organisms *C. elegans* and *Drosophila*, and can be delivered in several ways. In *C. elegans,* dsRNA can be injected directly into the worm, although this approach does not lend itself to large scale analysis. Instead, the worm can either be soaked in a solution of *in vitro* transcribed dsRNA or fed bacteria with plasmids expressing dsRNA; likewise, *Drosophila* embryos can be injected with dsRNA, or cultured S2 cells can be soaked in dsRNA or transfected with dsRNA-expressing plasmids. Most mammalian cells are refractory to long

dsRNA, as this induces the interferon response, leading to a global arrest of translation and apoptosis. siRNAs, however, are short enough to avoid this response, and can be introduced into mammalian cells via lipid-mediated transfection or electroporation. Additionally, the potential for siRNA-based therapeutics has been demonstrated in mouse models. Unmodified siRNAs targeted to viral genes could be administered nasally to inhibit respiratory virus infection (Bitko et al. 2005), while chemically modified siRNAs were introduced into the blood stream to target endogenous apoB (Soutschek et al. 2004).

Transfection of siRNAs results in only a transient silencing of gene expression, as the siRNAs become diluted as the cell divides, yet many prospective applications of RNAi require a stable knockdown. For this, short, hairpin RNAs (shRNAs) can be expressed from DNA vectors, which are then processed into siRNAs and mediate silencing. The first generation of shRNAs were expressed from RNA pol III promoters, usually the U6 or H1 promoters, as these are compact, are transcribed to high levels in a wide-variety of cells, and have a defined transcription stop site, a run of 5 thymidines (Brummelkamp et al. 2002; Paddison et al. 2002). A second generation of shRNAs utilizes the miRNA pathway to produce siRNAs, as the small RNA sequence of interest is placed into the context of mir-30 (Zeng et al. 2002). It is known that sequences outside the small RNA are important for proper processing, and this method preserves those signals (Chen et al. 2004; Zeng et al. 2005). Furthermore, this approach allows the siRNA to be expressed from RNA pol II, thus allowing the use of temporally and spatially regulated promoters. These shRNA expression cassettes can be placed into viral vectors to expand the range of cells that can be studied, as standard DNA transfection techniques are robust for only a narrow range of cells; for example, lentiviral vectors transduce almost every cell type, including non-dividing, primary cells, thereby allowing for analysis of gene knockdown in a

more physiologically-relevant setting (Rubinson et al. 2003; Stewart et al. 2003). Viral-based shRNAs have also been used to create mice silenced for genes of interest, and this technology provides an alternative to standard mouse knockout approaches (Rubinson et al. 2003; Tiscornia et al. 2004; Ventura et al. 2004; Grimm et al. 2005).

microRNAs

Discovery of a large class of genes

It was not long after the discovery of siRNAs in the original *Drosophila* lysate experiments that an obvious question set in: are there endogenous siRNAs? These experiments led to the discovery of a large class of endogenously encoded small RNAs that were termed microRNAs (miRNAs) (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). MicroRNAs have the interesting distinction of essentially being discovered twice, as work in worms had previously identified a small RNA that remained an isolated observation for some time (Lee et al. 1993; Wightman et al. 1993). The *lin-4* RNA would turn out, however, to be the founding member of a large class of interesting genes.

Original cloning efforts were undertaken by the Bartel, Tuschl, and Ambros labs. (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). Although the original protocols varied, current cloning efforts select for small RNAs by two key properties. First, siRNAs are a defined size, so one step involved size-fractionation of RNA on a denaturing polyacrylamide gel, to isolate RNA of approximately 21 nucleotides. Second, siRNAs contain a 5' phosphate and a 3' hydroxyl group, whereas most RNA degradation products do not. To take advantage of this property, a linker is ligated onto the 3' end of the RNA population, in an ATP-independent reaction to prevent circularization of the small RNA molecules. A second linker is ligated onto the 5' end, dependent on the presence of a 5' phosphate. The resulting products are amplified by reverse transcription and polymerase chain reaction, then restriction digested, concatamerized, and cloned into bacteria and sequenced.

The results of initial cloning experiments were not, as might have been expected, large numbers of RNA sequences that could form perfect complements with two nucleotide 3'

overhangs with other RNA sequences in the population, i.e. endogenous siRNAs. Instead, the researchers found sequences that appeared to exist predominately as single-stranded RNA in the cell. Interestingly, when these RNA sequences were found in their genomic context, the transcript that would contain these sequences would invariably form a hairpin structure. Northern blots for these species showed not only the single-stranded ~22 nucleotide form which was cloned, but also a band at ~60-70 nucleotides, corresponding to the hairpin structure. As further evidence that these cloned products were biologically relevant and not simply random degradation products, several of the sequences were conserved from worms to flies to mammals. The first cloning efforts yielded several dozen of these small RNA species in each organism, and this new class of RNA species was dubbed microRNAs (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001).

Concurrent with these cloning efforts, researchers in the RNAi field began to appreciate some previous work on a very interesting small, endogenous RNA. In the early 1980s, a screen was performed in *C. elegans* for genes involved in lineage defects (*lin* mutants) (Horvitz and Sulston 1980; Chalfie et al. 1981). The *lin-4 (lf)* mutant showed an interesting phenotype, whereby certain cells were retarded, meaning that instead of differentiating during the life of the worm, they continually divided into the same cell type (Ambros and Horvitz 1984). Another mutant, the *lin-14 (lf)* mutant, showed the opposite phenotype; a precocious mutant, it skipped over early developmental stages (Wightman et al. 1991). Due to their opposite phenotypes, it was hypothesized that the *lin-4* gene product regulated *lin-14*. Mapping of the *lin-4* mutation was incredibly laborious, and the *lin-4* locus became winnowed down to a smaller and smaller region, making it less and less likely that *lin-4* could encode a protein (Lee et al. 2004a; Ruvkun et al. 2004). Finally, it was determined that the *lin-4* gene did not code for a protein but rather

for a small RNA species (Lee et al. 1993). Furthermore, when the sequence of the ~22 nucleotide *lin-4* RNA was determined, it was immediately noticed that it shared a great deal of complementarity to several regulatory regions in the *lin-14* 3' UTR (Wightman et al. 1993). Further experimentation proved the importance of this RNA:RNA interaction, and *lin-4* was later called a small, temporal RNA (stRNA) for its important role in proper developmental timing.

For almost a decade, *lin-4* remained an isolated case. In the late 1990s, however, when RNAi was discovered, this "other" small RNA started to receive more notice, and researchers also found another example in the worm of a small RNA gene controlling development, *let-7* (Pasquinelli et al. 2000; Reinhart et al. 2000). To complete the connection to RNAi, genetic and biochemical studies implicated Dicer in processing the hairpin precursor to the ~22 nucleotide mature RNA (Hutvagner et al. 2001; Knight and Bass 2001), and several Argonaute proteins were also needed for their biogenesis and function (Grishok et al. 2001). Not surprisingly, cloning efforts from *C. elegans* for small RNAs identified both *let*-7 and *lin-4,* and thus *lin-4* became the founding member of this new class of genes, the microRNAs.

* * *

The Scope of miRNAs

MicroRNAs are defined as ~22 nucleotide noncoding RNAs processed by Dicer from a hairpin precursor (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). In contrast to siRNAs, which could arise from theoretically anywhere in the genome, miRNAs are discrete genes, and criteria have been established for identifying an RNA species as a bona fide miRNA (Ambros et al. 2003). MicroRNAs have been found in all metazoans that have been investigated (Bartel 2004), and recent work has shown that DNA viruses such as Epstein-Barr also encode miRNAs (Pfeffer et al. 2004).

The sequences of cloned miRNAs show a preference for a uridine as the first nucleotide, followed by a purine, as well as a weaker preference for an adenosine at the eighth position (Lau et al. 2001). Based on both cloning data as well as computational predictions, estimates for the number of miRNAs in the human genome range from 250 – 1000 (Lim et al. 2003a; Lim et al. 2003b; Bartel 2004; Berezikov et al. 2005). For comparison, there are approximately 500 kinases in the human genome, comprising 2% of protein-coding genes. Some miRNAs, such as *let-7* and *mir-1,* are highly conserved across metazoan evolution. It is clear that miRNAs have undergone duplication, as there are closely-related miRNAs found at various locations in the genome; for example, there are five *mir-30* sequences (designated a through e) in the human genome that differ from each other by only a few nucleotides.

MicroRNAs are found in diverse places in the genome, with some located far from any known protein-coding genes, while about one-third are found within introns (Bartel 2004). The majority of miRNAs are transcribed by RNA polymerase II (Cai et al. 2004); some are expressed independently, while others are expressed in clusters (Lau et al. 2001; Lee et al. 2004b). MicroRNAs have varied expression patterns, both spatially and temporally. For example, mammalian *mir-1* is expressed primarily in muscle while *mir-124* is mostly restricted to brain (Lagos-Quintana et al. 2002; Lim et al. 2005); likewise, the *mir-290-295* cluster is expressed only during the earliest stages of mammalian development (Houbaviy et al. 2003).

Mechanism

Much that is known about the mechanism of metazoan microRNAs comes from studies performed on the original *lin-4:lin-14* interaction in *C. elegans*. The *lin-4* mediated repression of *lin-14*, unlike RNAi, does not occur via targeting the *lin-14* mRNA for cleavage and degradation.

Northern blot analysis showed that *lin-14* levels remain the same when *lin-4* expression increases during the L1 to L2 developmental transition (Olsen and Ambros 1999). There are central bulges in all of the several *lin-4:lin-14* interactions (Ha et al. 1996), and RNAi requires essentially perfect duplexes for full cleavage activity (Elbashir et al. 2001c). Because mRNA levels do not appear to be changed, this is not a post-transcriptional phenomenon like RNAi, but instead has been dubbed "translational repression." It is possible that *lin-4* does not affect the process of translation per se, but instead destabilizes the newly-made protein; nevertheless, this silencing effect is often described as translational repression.

Biochemical analysis of the *lin-4:lin-14* interaction revealed that *lin-4* does not cause *lin-14* to lose ribosomes, as the mRNA remains in the polysome fraction of a sucrose gradient (Olsen and Ambros 1999). This result is interpreted as indicating that translation initiates properly on the repressed mRNA; otherwise, the *lin-14* mRNA would be associated with fewer ribosomes, and thus would fractionate differently. This same result has been seen with the *lin-28* mRNA, another target of *lin-4* (Seggerson et al. 2002). These are but two studies, and both were performed in *C. elegans*. Future work in this system and in other organisms will be needed to reveal a deeper understanding of how miRNAs actually lead to a decrease in steady-state protein levels. Indeed, an immediate alternative to this pathway is found in plants. Plant microRNAs, unlike metazoan miRNAs, oftentimes have extensive complementarity to known mRNAs, and function in an RNAi-like fashion, cleaving the target mRNA (Llave et al. 2002; Rhoades et al. 2002; Tang et al. 2003). And while most animal miRNAs have limited complementarity to any mRNA, miR-196 bears almost-exact complementarity to the HOXB8 gene, which appears to be cleaved in an RNAi-like fashion (Yekta et al. 2004).

Biogenesis

MicroRNAs begin as long transcripts known as the primary-miRNA (pri-miRNA) (Figure 2, p. 33) (Lee et al. 2002; Lee et al. 2003). An RNase III enzyme, Drosha, recognizes the hairpin that contains the miRNA and cleaves the primary transcript to produce a 60-70 nucleotide hairpin known as the precursor-miRNA (pre-miRNA) (Lee et al. 2002; Lee et al. 2003). Drosha interacts with a protein with a dsRBM, known as Pasha/DGCR8, to form the 'microprocessor' complex (Denli et al. 2004; Gregory et al. 2004; Han et al. 2004). Drosha cleavage requires at least 6 nucleotides of dsRNA below the base of the stem, as well as a relatively unstructured hairpin loop (Zeng et al. 2005). The Drosha cleavage defines one end of the miRNA, leaving a 5' phosphate and two nucleotide, 3' overhang. This hairpin structure is then recognized by Exportin 5, and the pre-miRNA is exported into the cytoplasm in a Ran-GTP dependent process (Yi et al. 2003; Lund et al. 2004).

In the cytoplasm, the pre-miRNA encounters Dicer. In worms and mammals, it appears that the same Dicer is responsible for both RNAi and miRNA processing, but in *Drosophila*, there are two Dicers, with Dcr-1 utilized in miRNA biogenesis (Grishok et al. 2001; Hutvagner et al. 2001; Lee et al. 2004c). The PAZ domain of Dicer recognizes the 5' phosphate and two nucleotide 3' overhang and positions the two RNase III domains, one on each strand of the hairpin stem, in the same way that Dicer functions in siRNA processing (Tahbaz et al. 2004; Zhang et al. 2004). Dicer then cleaves off the hairpin loop, to produce a molecule that bears much similarity to a canonical siRNA. As mentioned before, however, miRNAs are found largely as single-stranded species, and thus the strand analogous to the passenger strand of an siRNA is sometimes referred to as the miRNA* strand. Correlative evidence suggests that the same rules underlying siRNA strand selection apply to miRNAs (Schwarz et al. 2003); this

siRNA-like intermediate is quickly unwound, with one strand being stabilized in a microRNA ribonucleoprotein complex (miRNP) whereas the other strand is rapidly degraded.

miRNP

The ribonucleoprotein complex that contains microRNAs, known simply as the miRNP (or sometimes 'miRISC'), is not as well-defined biochemically as RISC, largely because the exact biochemical activity of miRNAs is largely unknown and thus the miRNP does not yet lend itself to *in vitro* assays. It is known that the miRNP contains an Argonaute protein, as an antibody against human Ago-2 immunoprecipitates many miRNAs (Mourelatos et al. 2002). Furthermore, over-expressed, tagged versions of human Ago-1, 2, 3, and 4 all associate with endogenous microRNAs, although only Ago-2, as the central component of RISC, has a known function (Liu et al. 2004; Meister et al. 2004). In *Drosophila,* Ago-1 associates preferentially with miRNAs while Ago-2 associates with siRNAs, although unlike in mammals, both Ago proteins can catalyze mRNA cleavage (Lee et al. 2004c).

In human cells, Gemin-3 and Gemin-4 are found associated with Ago-2 and with microRNAs (Mourelatos et al. 2002). Gemin-3, a DEAD-box RNA helicase, and Gemin-4, a novel protein, were previously found associated with the Survival of Motor Neuron (SMN) complex, which is predominantly nuclear and is important for the biogenesis of many ribonucleoproteins (Paushkin et al. 2002). SMN is a complex distinct from the miRNP, as several other Gemins associate with SMN that are not found as part of the miRNP; likewise, Ago-2 does not associate with these other Gemins. Purification of the minimal protein components of RISC activity does not yield Gemin-3 or -4, arguing that they are not necessary for mRNA cleavage and instead may function predominantly in the miRNA pathway (Martinez

et al. 2002; Martinez and Tuschl 2004). As mentioned above, studies of RISC have identified several proteins that do not as yet have a defined role in canonical RNAi. It is possible that some of these proteins are involved instead in the miRNA pathway. FMRP, for example, is a known modulator of translation and may be involved in conferring additional specificity to miRNAs.

Overlap and distinctions between RNAi and miRNAs

The miRNA pathway and RNAi share a great deal in common, such as the need for Dicer processing to generate the small RNA, as well as the central role of the Argonaute family of proteins in the effecter complex. Furthermore, it has been shown that a fraction of miRNAs reside in RISC, as a miRNA will cleave a target mRNA with sufficient complementarity (Hutvagner and Zamore 2002). Conversely, exogenous siRNAs are capable of translationally repressing a target mRNA with imperfectly-complementary binding sites (Doench et al. 2003). The decision between mRNA cleavage and translational repression appears to depend on the extent of complementarity between small RNA and target mRNA, although additional work will be needed to see if additional factors also modulate this decision.

Despite the great deal of overlap between RNAi and the microRNA pathway, their end results are qualitatively different, the former resulting in RNA cleavage, the latter in translational repression. Additionally, RNAi is a much more robust response, whereby target mRNAs are degraded in a catalytic cycle, while evidence suggests that miRNAs remain stably bound to their target mRNAs and modulate gene expression on a finer scale (Hutvagner and Zamore 2002; Doench et al. 2003; Haley and Zamore 2004). A knockout mouse of Ago-2 is embryonic lethal, but mouse embryonic fibroblasts (MEFs) can be derived (Liu et al. 2004). As expected, these MEFs are no longer capable of performing RNAi with a perfectly-complementary RNA:RNA

interaction, but translational-repression activity is still intact (Liu et al. 2004). This result implies that, in mammals, at least one of the other Argonautes can perform translational repression.

miRNA Targets

The discovery of miRNAs as an abundant class of regulatory RNAs immediately prompted investigation into the potential target mRNAs. Due to the high degree of complementarity to their targets, plant miRNA targets were quickly identified (Rhoades et al. 2002), but the search for animal miRNA targets has been more arduous. The first examples of regulation were found via classical genetic approaches, whereby mutagenesis happened to hit the microRNA (or its binding sites in the 3' UTR) and thus disrupted its function. In addition to the *C. elegans* examples of *lin-4* regulating *lin-14* and *lin-28*, and *let-7* regulating *lin-41*, several other interactions have been found (Lee et al. 1993; Wightman et al. 1993; Moss et al. 1997; Pasquinelli et al. 2000; Reinhart et al. 2000). For example, the *bantam* locus in *Drosophila* was simply a mutant locus, and there was no clear understanding of how this particular mutation caused the phenotype of premature apoptosis. With the discovery of miRNAs, it was soon appreciated that the *bantam* locus contained a miRNA, and experiments showed that *bantam* repressed the expression of the pro-apoptotic gene *hid* (Brennecke et al. 2003). Classical genetic approaches are likely of limited use for finding most miRNA targets, however, as a miRNA is a relatively small target for random mutagenesis as compared to a protein-coding gene. Furthermore, evidence suggests that miRNA family members can function redundantly. An alternative approach to this problem is computational biology.

A few general rules have been developed for miRNA:mRNA target interactions, derived both from observations of known targets and from direct experimentation. In general, it is assumed that the miRNA targets the 3' UTR of the mRNA, as is the case for *lin-14*, *lin-41*, *lin-28*, and *hid* (Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000; Slack et al. 2000; Brennecke et al. 2003; Lin et al. 2003; Vella et al. 2004). More recent work suggests that miRNAs may also target the coding sequence, although none of these targets have been validated (Lewis et al. 2005). Second, a miRNA will likely bind to the target mRNA multiple times; cell culture experiments with reporter genes have shown that multiple binding sites are needed for productive translational repression (Ha et al. 1996; Doench et al. 2003). Third, basepairing to the 5' region of the miRNA, sometimes referred to as the "seed" region, is a critical determinant of activity, and there is ample computational and experimental evidence to support this assumption (Lai 2002; Lewis et al. 2003; Doench and Sharp 2004). Lastly, most computational approaches rely on conservation of a binding site across several species (Lewis et al. 2003; Stark et al. 2003; John et al. 2004; Lewis et al. 2005). Certainly, this criterion helps to eliminate false positive predictions, but also loses miRNA targets that are specific for a given species.

Several attempts have been made to define miRNA targets in this computational approach, and the results are, perhaps expectedly, quite varied, depending on the exact nature of the algorithm in terms of the relative weight of the different assumptions and parameters discussed above (compare, for example, Lewis et al. 2003; John et al. 2004). What is clear, however, is that miRNAs have the potential to regulate many genes, a conclusion bolstered by direct experimentation showing that as few as ~8 nucleotides of complementarity between miRNA and target mRNA have translational repression activity (Doench and Sharp 2004). The

most recent computational predictions estimate that more than one-third of human mRNAs are conserved targets of miRNAs (Brennecke et al. 2005; Lewis et al. 2005; Xie et al. 2005).

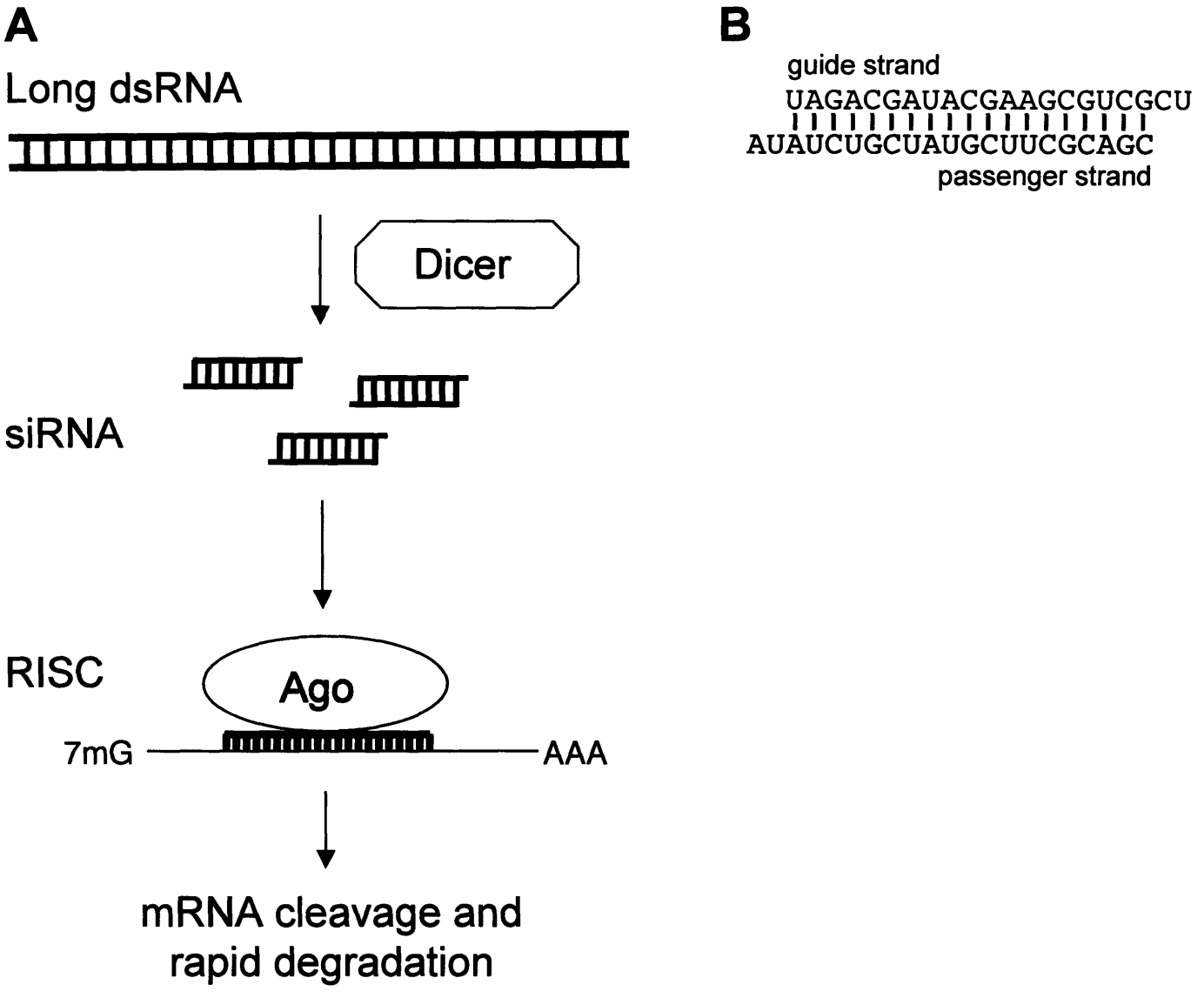

Figure 1. RNAi pathway

A. The canonical RNAi pathway is triggered by long, double-stranded RNA. This RNA is a substrate for the RNase III enzyme Dicer, which processively cleaves the dsRNA to produce short, interfering RNAs (siRNAs). One strand of the siRNA is chosen for incorporation into the RNA-induced Silencing Complex (RISC). At the heart of RISC is a member of the Argonaute family of proteins. RISC uses this guide strand to find perfectly complementary target RNAs and cleave them, resulting in their rapid degradation.

B. Asymmetry of siRNAs. siRNAs consist of a guide strand and a passenger strand, with the former incorporated into RISC, the latter rapidly degraded. This asymmetry is predicted by the basepairing of the first few nucleotides at the end of the duplex. The strand with less thermal stability at the 5' end (more A:U pairing than G:C pairing) is chosen for incorporation into RISC. In this example, the top, red strand of the duplex would be chosen; it is known as the guide strand, while the other strand is known as the passenger strand.

C. elegans	Function (citation)
alg-1, alg-2	Ago proteins; required for maturation and function of miRNAs (Grishok et al. 2001)
dcr-1	Dicer; required for RNAi and miRNA processing (Knight and Bass 2001)
drh-1	DEAD box helicase; associates with Dicer/rde-1/rde-4 to produce siRNAs (Tabara et al. 2002)
drsh-1	Drosha; processes primary miRNA transcript (Denli et al. 2004)
ego-1	germline RdRP (Smardon et al. 2000)
eri-1	negatively regulates RNAi pathway by degrading siRNAs (Kim et al. 2005)
mut-14	DEAD box helicase; functions downstream of siRNA production (Tijsterman et al. 2002a)
mut-7	putative 3'-5' exonuclease; required for siRNA accumulation (Ketting et al. 1999)
pash-1	partner of Drosha; required for miRNA biogenesis (Denli et al. 2004)
ppw-1	Ago protein; required for germline RNAi (Tijsterman et al. 2002b)
rde-1	Ago protein; required for siRNA production (Tabara et al. 1999)
rde-2/mut-8	associates with mut-7; required for siRNA accumulation (Tops et al. 2005)
rde-3/mut-2	polymerase β nucleotidyltransferase; required for siRNA accumulation (Chen et al. 2004)
rde-4	dsRNA binding protein; required for siRNA production (Tabara et al. 1999)
rrf-1	somatic RdRP (Sijen et al. 2001)
rrf-3	putative RdRP; loss of activity increases RNAi activity (Simmer et al. 2002)
sid-1	transmembrane protein required for systemic RNAi (Winston et al. 2002)

Drosophila	Function (citation)
Ago-1	Ago protein; binds miRNAs (Okamura et al. 2004)
Ago-2	Ago protein; catalytic core of RISC (Hammond et al. 2001)
Armitage	RNA helicase; required for siRNA unwinding in ovaries (Tomari et al. 2004a)
Aubergine	Ago protein; required for RNAi in oocytes and transcriptional gene silencing (Kennerdell et al. 2002)
Dcr-1	Dicer involved in miRNA biogenesis (Lee et al. 2004c)
Dcr-2	Dicer involved in siRNA production (Bernstein et al. 2001)
Drosha	processes primary miRNA transcript (Lee et al. 2003)
FMR1	Fragile X Mental Retardation homolog; component of RISC (Ishizuka et al. 2002)
Pasha/DGCR8	complexes with Drosha to process miRNAs (Gregory et al. 2004)
Piwi	Ago protein; required for RNAi and transcriptional gene silencing (Pal-Bhadra et al. 1999)
R2D2	dsRNA binding protein; associates with Dicer to sense siRNA asymmetry (Liu et al. 2003)
Spindle-E	RNA helicase; required for RNAi in oocytes and TGS (Aravin et al. 2001)
Tudor SN	micrococcal nuclease homolog; component of RISC (Caudy et al. 2003)
VIG	putative RNA binding protein; component of RISC (Caudy et al. 2002)

Table 1. Genes associated with RNAi activity in *C. elegans* and *Drosophila.*
Genes identified by phenotype in *C. elegans* to be important for RNAi are listed. The functions are inferred from either structural similarity to other proteins or direct biochemical analysis. The genes listed for *Drosophila* have been shown to be involved in RNAi-related activities by genetic and biochemical experiments, or have been associated with RISC activity. Counterparts for most of the list of *Drosophila* genes have been readily identified in human and mouse.

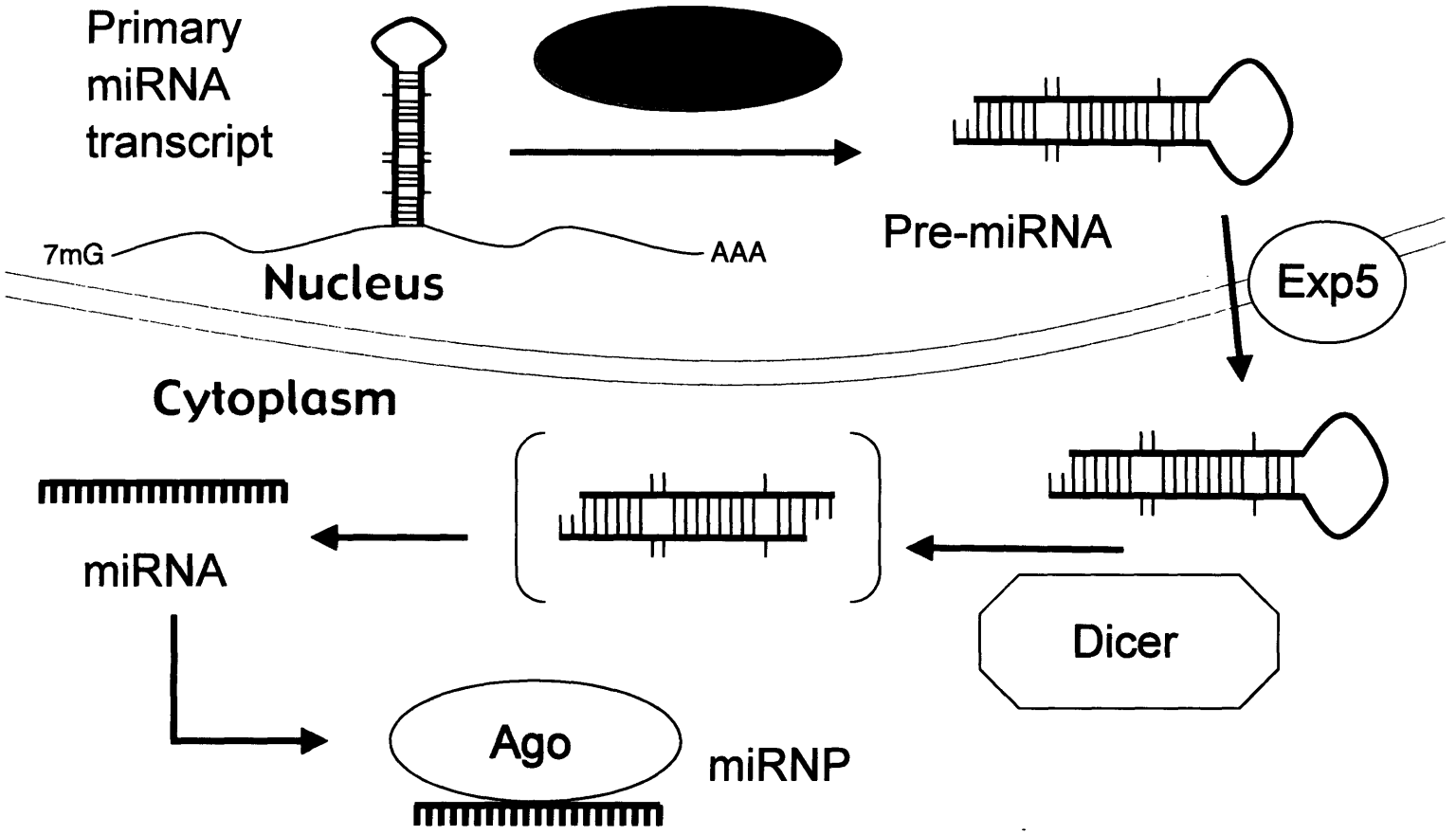

Figure 2. MicroRNA biogenesis

MicroRNAs are transcribed as a long primary transcript that is first processed by the RNase III enzyme Drosha to produce a pre-miRNA. This species is exported to the cytoplasm by Exportin 5, where it is further processed by Dicer, to give rise to a transient duplexed species remiscient of a canonical siRNA. One strand of this duplex is incorporated into the miRNP, a ribnonucleoprotein complex that contains an Argonaute family member.

References

Ambros, V., B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, and T. Tuschl. 2003. A uniform system for microRNA annotation. *RNA* **9**: 277-9.

Ambros, V. and H.R. Horvitz. 1984. Heterochronic mutants of the nematode Caenorhabditis elegans. *Science* **226**: 409-16.

Antar, L.N. and G.J. Bassell. 2003. Sunrise at the synapse: the FMRP mRNP shaping the synaptic interface. *Neuron* **37**: 555-8.

Aravin, A.A., N.M. Naumova, A.V. Tulin, V.V. Vagin, Y.M. Rozovsky, and V.A. Gvozdev. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* **11**: 1017-27.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-97.

Berezikov, E., V. Guryev, J. van de Belt, E. Wienholds, R.H. Plasterk, and E. Cuppen. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**: 21-4.

Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-6.

Bernstein, E., S.Y. Kim, M.A. Carmell, E.P. Murchison, H. Alcorn, M.Z. Li, A.A. Mills, S.J. Elledge, K.V. Anderson, and G.J. Hannon. 2003. Dicer is essential for mouse development. *Nat Genet* **35**: 215-7.

Bitko, V., A. Musiyenko, O. Shulyayeva, and S. Barik. 2005. Inhibition of respiratory viruses by nasally administered siRNA. *Nat Med* **11**: 50-5.

Brennecke, J., D.R. Hipfner, A. Stark, R.B. Russell, and S.M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**: 25-36.

Brennecke, J., A. Stark, R.B. Russell, and S.M. Cohen. 2005. Principles of microRNA-target recognition. *PLoS Biol* **3**: e85.

Brown, V., P. Jin, S. Ceman, J.C. Darnell, W.T. O'Donnell, S.A. Tenenbaum, X. Jin, Y. Feng, K.D. Wilkinson, J.D. Keene, R.B. Darnell, and S.T. Warren. 2001. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**: 477-87.

Brummelkamp, T.R., R. Bernards, and R. Agami. 2002. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**: 550-3.

Cai, X., C.H. Hagedorn, and B.R. Cullen. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**: 1957-66.

Carmell, M.A., Z. Xuan, M.Q. Zhang, and G.J. Hannon. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**: 2733-42.

Caudy, A.A., R.F. Ketting, S.M. Hammond, A.M. Denli, A.M. Bathoorn, B.B. Tops, J.M. Silva, M.M. Myers, G.J. Hannon, and R.H. Plasterk. 2003. A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**: 411-4.

Caudy, A.A., M. Myers, G.J. Hannon, and S.M. Hammond. 2002. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* **16**: 2491-6.

Ceman, S., W.T. O'Donnell, M. Reed, S. Patton, J. Pohl, and S.T. Warren. 2003. Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum Mol Genet* **12**: 3295-305.

Chalfie, M., H.R. Horvitz, and J.E. Sulston. 1981. Mutations that lead to reiterations in the cell lineages of C. elegans. *Cell* **24**: 59-69.

Chen, C.Z., L. Li, H.F. Lodish, and D.P. Bartel. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83-6.

Cogoni, C. and G. Macino. 1999. Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**: 166-9.

Corbin, F., M. Bouillon, A. Fortin, S. Morin, F. Rousseau, and E.W. Khandjian. 1997. The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. *Hum Mol Genet* **6**: 1465-72.

Dalmay, T., A. Hamilton, S. Rudd, S. Angell, and D.C. Baulcombe. 2000. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**: 543-53.

Darnell, J.C., K.B. Jensen, P. Jin, V. Brown, S.T. Warren, and R.B. Darnell. 2001. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**: 489-99.

Denli, A.M., B.B. Tops, R.H. Plasterk, R.F. Ketting, and G.J. Hannon. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231-5.

Doench, J.G., C.P. Petersen, and P.A. Sharp. 2003. siRNAs can function as miRNAs. *Genes Dev* **17**: 438-42.

Doench, J.G. and P.A. Sharp. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**: 504-11.

Elbashir, S.M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. 2001a. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494-8.

Elbashir, S.M., W. Lendeckel, and T. Tuschl. 2001b. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**: 188-200.

Elbashir, S.M., J. Martinez, A. Patkaniowska, W. Lendeckel, and T. Tuschl. 2001c. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J* **20**: 6877-88.

Fagard, M., S. Boutet, J.B. Morel, C. Bellini, and H. Vaucheret. 2000. AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc Natl Acad Sci U S A* **97**: 11650-4.

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

Gregory, R.I., K.P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235-40.

Grimm, D., K. Pandey, and M.A. Kay. 2005. Adeno-associated virus vectors for short hairpin RNA expression. *Methods Enzymol* **392**: 381-405.

Grishok, A., A.E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D.L. Baillie, A. Fire, G. Ruvkun, and C.C. Mello. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**: 23-34.

Grishok, A., H. Tabara, and C.C. Mello. 2000. Genetic requirements for inheritance of RNAi in C. elegans. *Science* **287**: 2494-7.

Guo, S. and K.J. Kemphues. 1996. A non-muscle myosin required for embryonic polarity in Caenorhabditis elegans. *Nature* **382**: 455-8.

Ha, I., B. Wightman, and G. Ruvkun. 1996. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. *Genes Dev* **10**: 3041-50.

Haley, B. and P.D. Zamore. 2004. Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol* **11**: 599-606.

Hamilton, A.J. and D.C. Baulcombe. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950-2.

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

Hammond, S.M., S. Boettcher, A.A. Caudy, R. Kobayashi, and G.J. Hannon. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**: 1146-50.

Han, J., Y. Lee, K.H. Yeom, Y.K. Kim, H. Jin, and V.N. Kim. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**: 3016-27.

Horvitz, H.R. and J.E. Sulston. 1980. Isolation and genetic characterization of cell-lineage mutants of the nematode Caenorhabditis elegans. *Genetics* **96**: 435-54.

Houbaviy, H.B., M.F. Murray, and P.A. Sharp. 2003. Embryonic stem cell-specific MicroRNAs. *Dev Cell* **5**: 351-8.

Hutvagner, G., J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834-8.

Hutvagner, G. and P.D. Zamore. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056-60.

Ishizuka, A., M.C. Siomi, and H. Siomi. 2002. A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* **16**: 2497-508.

Jackson, A.L., S.R. Bartz, J. Schelter, S.V. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, and P.S. Linsley. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**: 635-7.

John, B., A.J. Enright, A. Aravin, T. Tuschl, C. Sander, and D.S. Marks. 2004. Human MicroRNA targets. *PLoS Biol* **2**: e363.

Kanellopoulou, C., S.A. Muljo, A.L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D.M. Livingston, and K. Rajewsky. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**: 489-501.

Kennerdell, J.R. and R.W. Carthew. 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**: 1017-26.

Kennerdell, J.R., S. Yamaguchi, and R.W. Carthew. 2002. RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. *Genes Dev* **16**: 1884-9.

Ketting, R.F., T.H. Haverkamp, H.G. van Luenen, and R.H. Plasterk. 1999. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**: 133-41.

Khvorova, A., A. Reynolds, and S.D. Jayasena. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209-16.

Kim, J.K., H.W. Gabel, R.S. Kamath, M. Tewari, A. Pasquinelli, J.F. Rual, S. Kennedy, M. Dybbs, N. Bertin, J.M. Kaplan, M. Vidal, and G. Ruvkun. 2005. Functional Genomic Analysis of RNA Interference in C. elegans. *Science*.

Knight, S.W. and B.L. Bass. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science* **293**: 2269-71.

Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-8.

Lagos-Quintana, M., R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel, and T. Tuschl. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**: 735-9.

Lai, E.C. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**: 363-4.

Lakatos, L., G. Szittya, D. Silhavy, and J. Burgyan. 2004. Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *Embo J* **23**: 876-84.

Lau, N.C., L.P. Lim, E.G. Weinstein, and D.P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858-62.

Lee, R., R. Feinbaum, and V. Ambros. 2004a. A short history of a short RNA. *Cell* **116**: S89-92, 1 p following S96.

Lee, R.C. and V. Ambros. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**: 862-4.

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-54.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V.N. Kim. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415-9.

Lee, Y., K. Jeon, J.T. Lee, S. Kim, and V.N. Kim. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**: 4663-70.

Lee, Y., M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek, and V.N. Kim. 2004b. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**: 4051-60.

Lee, Y.S., K. Nakahara, J.W. Pham, K. Kim, Z. He, E.J. Sontheimer, and R.W. Carthew. 2004c. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**: 69-81.

Lewis, B.P., C.B. Burge, and D.P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.

Lewis, B.P., I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, and C.B. Burge. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787-98.

Lim, L.P., M.E. Glasner, S. Yekta, C.B. Burge, and D.P. Bartel. 2003a. Vertebrate microRNA genes. *Science* **299**: 1540.

Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769-73.

Lim, L.P., N.C. Lau, E.G. Weinstein, A. Abdelhakim, S. Yekta, M.W. Rhoades, C.B. Burge, and D.P. Bartel. 2003b. The microRNAs of Caenorhabditis elegans. *Genes Dev* **17**: 991-1008.

Lin, S.Y., S.M. Johnson, M. Abraham, M.C. Vella, A. Pasquinelli, C. Gamberi, E. Gottlieb, and F.J. Slack. 2003. The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev Cell* **4**: 639-50.

Lingel, A., B. Simon, E. Izaurralde, and M. Sattler. 2003. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**: 465-9.

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

Liu, Q., T.A. Rand, S. Kalidas, F. Du, H.E. Kim, D.P. Smith, and X. Wang. 2003. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**: 1921-5.

Llave, C., Z. Xie, K.D. Kasschau, and J.C. Carrington. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**: 2053-6.

Lund, E., S. Guttinger, A. Calado, J.E. Dahlberg, and U. Kutay. 2004. Nuclear export of microRNA precursors. *Science* **303**: 95-8.

Ma, J.B., Y.R. Yuan, G. Meister, Y. Pei, T. Tuschl, and D.J. Patel. 2005. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**: 666-70.

Marathe, R., R. Anandalakshmi, T.H. Smith, G.J. Pruss, and V.B. Vance. 2000. RNA viruses as inducers, suppressors and targets of post-transcriptional gene silencing. *Plant Mol Biol* **43**: 295-306.

Martinez, J., A. Patkaniowska, H. Urlaub, R. Luhrmann, and T. Tuschl. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**: 563-74.

Martinez, J. and T. Tuschl. 2004. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev* **18**: 975-80.

Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng, and T. Tuschl. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**: 185-97.

Moss, E.G., R.C. Lee, and V. Ambros. 1997. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**: 637-46.

Motamedi, M.R., A. Verdel, S.U. Colmenares, S.A. Gerber, S.P. Gygi, and D. Moazed. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**: 789-802.

Mourelatos, Z., J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**: 720-8.

Mourrain, P., C. Beclin, T. Elmayan, F. Feuerbach, C. Godon, J.B. Morel, D. Jouette, A.M. Lacombe, S. Nikic, N. Picault, K. Remoue, M. Sanial, T.A. Vo, and H. Vaucheret. 2000. Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**: 533-42.

Nykanen, A., B. Haley, and P.D. Zamore. 2001. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**: 309-21.

Okamura, K., A. Ishizuka, H. Siomi, and M.C. Siomi. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**: 1655-66.

Olsen, P.H. and V. Ambros. 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**: 671-80.

Orban, T.I. and E. Izaurralde. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**: 459-69.

Paddison, P.J., A.A. Caudy, E. Bernstein, G.J. Hannon, and D.S. Conklin. 2002. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev* **16**: 948-58.

Pal-Bhadra, M., U. Bhadra, and J.A. Birchler. 1999. Cosuppression of nonhomologous transgenes in Drosophila involves mutually related endogenous sequences. *Cell* **99**: 35-46.

Pal-Bhadra, M., U. Bhadra, and J.A. Birchler. 2002. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. *Mol Cell* **9**: 315-27.

Pal-Bhadra, M., B.A. Leibovitch, S.G. Gandhi, M. Rao, U. Bhadra, J.A. Birchler, and S.C. Elgin. 2004. Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303**: 669-72.

Parker, J.S., S.M. Roe, and D. Barford. 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**: 663-6.

Pasquinelli, A.E., B.J. Reinhart, F. Slack, M.Q. Martindale, M.I. Kuroda, B. Maller, D.C. Hayward, E.E. Ball, B. Degnan, P. Muller, J. Spring, A. Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**: 86-9.

Paushkin, S., A.K. Gubitz, S. Massenet, and G. Dreyfuss. 2002. The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol* **14**: 305-12.

Pfeffer, S., M. Zavolan, F.A. Grasser, M. Chien, J.J. Russo, J. Ju, B. John, A.J. Enright, D. Marks, C. Sander, and T. Tuschl. 2004. Identification of virus-encoded microRNAs. *Science* **304**: 734-6.

Pham, J.W., J.L. Pellino, Y.S. Lee, R.W. Carthew, and E.J. Sontheimer. 2004. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117**: 83-94.

Reinhart, B.J. and D.P. Bartel. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**: 1831.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-6.

Reynolds, A., D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, and A. Khvorova. 2004. Rational siRNA design for RNA interference. *Nat Biotechnol* **22**: 326-30.

Rhoades, M.W., B.J. Reinhart, L.P. Lim, C.B. Burge, B. Bartel, and D.P. Bartel. 2002. Prediction of plant microRNA targets. *Cell* **110**: 513-20.

Rivas, F.V., N.H. Tolia, J.J. Song, J.P. Aragon, J. Liu, G.J. Hannon, and L. Joshua-Tor. 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol.*

Rubinson, D.A., C.P. Dillon, A.V. Kwiatkowski, C. Sievers, L. Yang, J. Kopinja, D.L. Rooney, M.M. Ihrig, M.T. McManus, F.B. Gertler, M.L. Scott, and L. Van Parijs. 2003. A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat Genet* **33**: 401-6.

Ruvkun, G., B. Wightman, and I. Ha. 2004. The 20 years it took to recognize the importance of tiny RNAs. *Cell* **116**: S93-6, 2 p following S96.

Schwarz, D.S., G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P.D. Zamore. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199-208.

Schwarz, D.S., G. Hutvagner, B. Haley, and P.D. Zamore. 2002. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol Cell* **10**: 537-48.

Schwarz, D.S., Y. Tomari, and P.D. Zamore. 2004. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Curr Biol* **14**: 787-91.

Seggerson, K., L. Tang, and E.G. Moss. 2002. Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Dev Biol* **243**: 215-25.

Shen, B. and H.M. Goodman. 2004. Uridine addition after microRNA-directed cleavage. *Science* **306**: 997.

Sijen, T., J. Fleenor, F. Simmer, K.L. Thijssen, S. Parrish, L. Timmons, R.H. Plasterk, and A. Fire. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**: 465-76.

Simmer, F., M. Tijsterman, S. Parrish, S.P. Koushika, M.L. Nonet, A. Fire, J. Ahringer, and R.H. Plasterk. 2002. Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. *Curr Biol* **12**: 1317-9.

Slack, F.J., M. Basson, Z. Liu, V. Ambros, H.R. Horvitz, and G. Ruvkun. 2000. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell* **5**: 659-69.

Smardon, A., J.M. Spoerke, S.C. Stacey, M.E. Klein, N. Mackin, and E.M. Maine. 2000. EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Curr Biol* **10**: 169-78.

Song, J.J., S.K. Smith, G.J. Hannon, and L. Joshua-Tor. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434-7.

Souret, F.F., J.P. Kastenmayer, and P.J. Green. 2004. AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol Cell* **15**: 173-83.

Soutschek, J., A. Akinc, B. Bramlage, K. Charisse, R. Constien, M. Donoghue, S. Elbashir, A. Geick, P. Hadwiger, J. Harborth, M. John, V. Kesavan, G. Lavine, R.K. Pandey, T. Racie, K.G. Rajeev, I. Rohl, I. Toudjarska, G. Wang, S. Wuschko, D. Bumcrot, V. Koteliansky, S. Limmer, M. Manoharan, and H.P. Vornlocher. 2004. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**: 173-8.

Stark, A., J. Brennecke, R.B. Russell, and S.M. Cohen. 2003. Identfication of Drosophila MicroRNA Targets. *PLOS Biology* **1**: 1-13.

Stefani, G., C.E. Fraser, J.C. Darnell, and R.B. Darnell. 2004. Fragile X mental retardation protein is associated with translating polyribosomes in neuronal cells. *J Neurosci* **24**: 7272-6.

Stewart, S.A., D.M. Dykxhoorn, D. Palliser, H. Mizuno, E.Y. Yu, D.S. An, D.M. Sabatini, I.S. Chen, W.C. Hahn, P.A. Sharp, R.A. Weinberg, and C.D. Novina. 2003. Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**: 493-501.

Tabara, H., M. Sarkissian, W.G. Kelly, J. Fleenor, A. Grishok, L. Timmons, A. Fire, and C.C. Mello. 1999. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**: 123-32.

Tabara, H., E. Yigit, H. Siomi, and C.C. Mello. 2002. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. *Cell* **109**: 861-71.

Tahbaz, N., F.A. Kolb, H. Zhang, K. Jaronczyk, W. Filipowicz, and T.C. Hobman. 2004. Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. *EMBO Rep* **5**: 189-94.

Tang, G., B.J. Reinhart, D.P. Bartel, and P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. *Genes Dev* **17**: 49-63.

Tijsterman, M., R.F. Ketting, K.L. Okihara, T. Sijen, and R.H. Plasterk. 2002a. RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs. *Science* **295**: 694-7.

Tijsterman, M., K.L. Okihara, K. Thijssen, and R.H. Plasterk. 2002b. PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of C. elegans. *Curr Biol* **12**: 1535-40.

Tiscornia, G., V. Tergaonkar, F. Galimi, and I.M. Verma. 2004. CRE recombinase-inducible RNA interference mediated by lentiviral vectors. *Proc Natl Acad Sci U S A* **101**: 7347-51.

Tomari, Y., T. Du, B. Haley, D.S. Schwarz, R. Bennett, H.A. Cook, B.S. Koppetsch, W.E. Theurkauf, and P.D. Zamore. 2004a. RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell* **116**: 831-41.

Tomari, Y., C. Matranga, B. Haley, N. Martinez, and P.D. Zamore. 2004b. A protein sensor for siRNA asymmetry. *Science* **306**: 1377-80.

Tops, B.B., H. Tabara, T. Sijen, F. Simmer, C.C. Mello, R.H. Plasterk, and R.F. Ketting. 2005. RDE-2 interacts with MUT-7 to mediate RNA interference in Caenorhabditis elegans. *Nucleic Acids Res* **33**: 347-55.

Tuschl, T., P.D. Zamore, R. Lehmann, D.P. Bartel, and P.A. Sharp. 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev* **13**: 3191-7.

van Hoof, A., P.A. Frischmeyer, H.C. Dietz, and R. Parker. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**: 2262-4.

Vance, V. and H. Vaucheret. 2001. RNA silencing in plants--defense and counterdefense. *Science* **292**: 2277-80.

Vargason, J.M., G. Szittya, J. Burgyan, and T.M. Tanaka Hall. 2003. Size selective recognition of siRNA by an RNA silencing suppressor. *Cell* **115**: 799-811.

Vella, M.C., E.Y. Choi, S.Y. Lin, K. Reinert, and F.J. Slack. 2004. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev* **18**: 132-7.

Ventura, A., A. Meissner, C.P. Dillon, M. McManus, P.A. Sharp, L. Van Parijs, R. Jaenisch, and T. Jacks. 2004. Cre-lox-regulated conditional RNA interference from transgenes. *Proc Natl Acad Sci U S A* **101**: 10380-5.

Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi, S.I. Grewal, and D. Moazed. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672-6.

Volpe, T., V. Schramke, G.L. Hamilton, S.A. White, G. Teng, R.A. Martienssen, and R.C. Allshire. 2003. RNA interference is required for normal centromere function in fission yeast. *Chromosome Res* **11**: 137-46.

Volpe, T.A., C. Kidner, I.M. Hall, G. Teng, S.I. Grewal, and R.A. Martienssen. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833-7.

Wianny, F. and M. Zernicka-Goetz. 2000. Specific interference with gene function by double-stranded RNA in early mouse development. *Nat Cell Biol* **2**: 70-5.

Wightman, B., T.R. Burglin, J. Gatto, P. Arasu, and G. Ruvkun. 1991. Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. *Genes Dev* **5**: 1813-24.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855-62.

Winston, W.M., C. Molodowitch, and C.P. Hunter. 2002. Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. *Science* **295**: 2456-9.

Xie, X., J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, and M. Kellis. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338-45.

Yan, K.S., S. Yan, A. Farooq, A. Han, L. Zeng, and M.M. Zhou. 2003. Structure and conserved RNA binding of the PAZ domain. *Nature* **426**: 468-74.

Ye, K., L. Malinina, and D.J. Patel. 2003. Recognition of small interfering RNA by a viral suppressor of RNA silencing. *Nature* **426**: 874-8.

Yekta, S., I.H. Shih, and D.P. Bartel. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**: 594-6.

Yi, R., Y. Qin, I.G. Macara, and B.R. Cullen. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**: 3011-6.

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

Zeng, Y., E.J. Wagner, and B.R. Cullen. 2002. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9**: 1327-33.

Zeng, Y., R. Yi, and B.R. Cullen. 2005. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *Embo J* **24**: 138-48.

Zhang, H., F.A. Kolb, L. Jaskiewicz, E. Westhof, and W. Filipowicz. 2004. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**: 57-68.

Chapter Two

siRNAs Can Function as miRNAs

This chapter is presented in the context of its contemporary science, and originally appeared in

Genes and Development **17**: 438-42 (2003).

The experiments described here were performed in collaboration with Christian P. Petersen.

Abstract

With the discovery of RNA interference (RNAi) and related phenomena, new regulatory roles attributed to RNA continue to emerge. Here we show, in mammalian tissue culture, that a short interfering RNA (siRNA) can repress expression of a target mRNA with partially complementary binding sites in its 3' UTR, much like the demonstrated function of endogenously encoded microRNAs (miRNAs). The mechanism for this repression is cooperative, distinct from the catalytic mechanism of mRNA cleavage by siRNAs. The use of siRNAs to study translational repression holds promise for dissecting the sequence and structural determinants and general mechanism of gene repression by miRNAs.

Introduction

The RNA interference (RNAi) pathway was first recognized in *Caenorhabditis elegans* as a response to exogenously introduced long double stranded RNA (dsRNA) (Fire et al. 1998). An RNase III enzyme, Dicer, cleaves the dsRNA into duplexes of 21-23 nt termed short interfering RNAs (siRNAs), which then guide a multicomponent complex known as RISC (RNA Induced Silencing Complex) to mRNAs sharing perfect complementarity and target their cleavage (Hamilton and Baulcombe 1999; Tuschl et al. 1999; Zamore et al. 2000; Hammond et al. 2000; Bernstein et al. 2001; Elbashir et al. 2001a). The RNAi pathway has been implicated in silencing transposons in the *C. elegans* germline (Tabara et al. 1999; Ketting et al. 1999), silencing Stellate repeats in the *Drosophila* germline (Aravin et al. 2001), and serving as an immune response against invading viruses in plants (reviewed in Baulcombe 2001). Very little, however, is known about the intrinsic biological role of RNAi in mammalian systems; indeed, no endogenous siRNAs have been identified in mammals. Nevertheless, transfection of mammalian cells with exogenous siRNAs has rapidly been adopted as a technology for targeted gene silencing (Elbashir et al. 2001a).

A related short RNA species, microRNAs (miRNAs), has been identified in organisms ranging from plants to nematodes to mammals (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001; Reinhart et al. 2002). These endogenous RNA species are first transcribed as a long RNA and then processed to a pre-miRNA of ~ 70 nt (Lee et al. 2002). This pre-miRNA forms an imperfect hairpin structure which is processed by Dicer to produce the mature, single strand ~22 nt miRNA (Grishok et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001). Despite the large library of miRNAs now known in animals, only two have a known function; *lin-4* and *let-7* regulate endogenous genes involved in developmental timing in *C. elegans* by partially

basepairing to the 3' UTR of target mRNAs such as *lin-14* and *lin-41*, respectively (Lee et al. 1993; Wightman et al. 1993; Ha et al. 1996; Reinhart et al. 2000; Slack et al. 2000). This interaction does not affect the stability of the target mRNA but rather represses gene expression through an unknown mechanism known as translational repression (Olsen and Ambros 1999). The polysome profile of the target mRNA does not change upon gene silencing, suggesting that this repression occurs after initiation of translation, and potentially occurs post-translationally (Olsen and Ambros 1999). This form of regulation is likely to be conserved in mammalian cells since overexpression of miR-30 can repress a reporter gene with partially complementary miR-30 binding sites in its 3' UTR without affecting mRNA stability (Zeng et al. 2002).

The RNAi pathway of siRNA-directed mRNA cleavage and the miRNA-mediated translational repression pathway are genetically and biochemically distinct. In addition to different outcomes, the two pathways have differential requirements for Paz-Piwi domain (PPD) proteins in *C. elegans*. Translational repression by *lin-4* and *let-7* depends on *alg-1* and *alg-2* for miRNA processing and/or stability yet these genes are not required for RNAi (Grishok et al. 2001), while *rde-1* is needed in RNAi but is not necessary for translational repression (Tabara et al. 1999). In HeLa cells, Gemin 3 and 4 proteins immunoprecipitate with RISC activity (Hutvagner and Zamore 2002) and miRNAs (Mourelatos et al. 2002), but have not been detected as components of purified RISC activity from S100 extracts (Martinez et al. 2002).

In addition to requiring Dicer processing to generate the short RNA, RNAi and translational repression share common components. The PPD protein eIF2C2 both immunoprecipitates with miRNAs from HeLa cells (Mourelatos et al. 2002) and co-purifies with RISC activity (Martinez et al. 2002). Additionally, endogenous *let-7* in HeLa extracts is capable of directing cleavage of a perfectly complementary target mRNA, suggesting that RNAi and

translational repression share common entry points if not overlapping machinery (Hutvagner and Zamore 2002). Because of these similarities, we reasoned that siRNAs may be capable of repressing gene expression via the miRNA-mediated pathway.

Results and Discussion

To test the ability of siRNAs to function like miRNAs in repressing gene expression, we designed a binding site that would basepair to the antisense strand of a siRNA known to be active *in vivo* for cleavage of the cell-surface receptor CXCR4 mRNA (Fig. 1A). Notably, this binding site contains a central bulge, thereby precluding RISC-directed mRNA cleavage (Elbashir et al. 2001a; Holen et al. 2002). We introduced four of these binding sites as consecutive repeats separated by four nucleotides into the 3' UTR of the *Renilla reniformis* luciferase reporter gene (*Rr*-luc); we also made a similar 3' UTR construct with a single binding site with perfect complementarity, to serve as a positive control for RNAi activity. Transfection of HeLa cells and subsequent luciferase assays revealed that the CXCR4 siRNA induced at least ten fold silencing of both of these constructs (Fig. 1B). RT-PCR showed that the two constructs were suppressed by two different mechanisms, as the perfectly complementary antisense siRNA:mRNA interaction resulted in a significant decrease in the steady state mRNA level, while the bulged interaction did not significantly reduce the mRNA level (Fig. 1C). Trace radiolabeling of an independent RT-PCR experiment was also used to better quantitate RNA levels, normalizing first within a sample to the control *Photinus pyralis* luciferase (*Pp*-luc) and then across samples to the (-) siRNA transfection. Such quantitation revealed that the perfectly complementary construct, targeted for RNAi, showed a greater than ten fold decrease in RNA level, while the bulged construct showed only 1.2 fold reduction in RNA level (data not shown). Interestingly, the sense strand of the same CXCR4 siRNA was capable of repressing a mRNA with four bulged binding sites (Fig. 1D). However, in this case the level of repression was only four fold as compared to the ten fold repression observed above (data not shown). As an additional control, the four bulged CXCR4 binding sites (Fig. 1A) were introduced into the *Pp*-

luc vector. Luciferase assays showed six fold repression (data not shown). Northern analysis of cytoplasmic RNA confirmed that the bulged binding sites do not cause a decrease in mRNA levels, relative to the ß-actin control (Fig. 1E). Thus, we conclude that a siRNA can function like a miRNA, repressing gene expression without a concordant decrease in mRNA stability.

Cloning efforts in many labs have revealed a large library of miRNAs, yet *C. elegans lin-4* and *let-7* remain the only miRNAs with known mRNA targets for translational repression in animals, and no such interactions are known in mammals. Computational prediction of targets is difficult because the rules for miRNA:mRNA pairing which function in translational repression have not been determined. Systematic manipulation of genes encoding miRNAs to explore these rules is complicated because the mutant genes must be processed by Dicer and the rules for this cleavage are not known. However, the ability of a siRNA to function by a miRNA-type pathway allows direct investigation of sequence and structure requirements for translational repression in the absence of Dicer processing.

To begin to define these rules, different siRNA sequences were tested for their ability to repress reporters in the luciferase assay. Because both the more effective strand of the CXCR4 siRNA (Fig. 1A) and the only previously studied example of miRNA repression in mammalian cells (Zeng et al. 2002) had a 3'-AGG-5' bulge in the siRNA strand when paired to the target mRNA, we tested the importance of this sequence. Two constructs were designed which would basepair to the sense or antisense strand of a siRNA previously used to effectively target GFP mRNA for cleavage. The siRNA:mRNA interaction with the AGG bulge was two fold more effective than that with the ACC bulge (Fig. 2, comparison of A & B). By using a different siRNA, the AGG bulge of the siRNA:mRNA interaction in figure 2A was replaced with an ACC bulge, and the ACC bulge of the siRNA:mRNA interaction in figure 2B was replaced with an

AGG bulge. (We note that in Fig. 2A the two 3' bases of the siRNA were changed from UU to CC.) Surprisingly, none of these changes had an effect on the degree of repression. Therefore, by this assay the sequence of the bulge is not the major determinant of translational repression activity.

Since in *Drosophila* embryo extracts the antisense strand of the siRNA sets the ruler for cleavage of target mRNA, at the ninth nucleotide from its paired 5' end (Elbashir et al. 2001b), the position of the bulge may be a critical determinant of translational repression activity. However, both the most effective and least effective bulges tested (Fig. 1A and 2B, respectively) position the bulge eight basepairs from the 5' end of the siRNA. Furthermore, another active construct positioned the bulge nine basepairs from the 5' end (Fig. 2A). We speculate that a combination of these sequence and structural parameters govern the ability of a siRNA/miRNA to induce translational repression, but that an expanded study will be necessary to define them.

The number of miRNA binding sites in a target mRNA is a likely determinant of the effectiveness of translational repression. Indeed, the *lin-14* 3' UTR contains seven potential *lin-4* miRNA binding sites, and the *lin-41* 3' UTR contains one *lin-4* miRNA and two *let-7* miRNA binding sites (reviewed in Banerjee and Slack 2002). To investigate this possibility, a series of *Rr*-luc reporters with an increasing number of binding sites—0, 2, 4, and 6—were transfected into HeLa cells with increasing concentrations of CXCR4 siRNA. The level of repression increased with increasing number of binding sites and with increasing concentrations of siRNA (Fig. 3A). To compare the effectiveness of translational repression to mRNA cleavage by siRNAs, a series of *Pp*-luc reporters with an increasing number of binding sites—0,1, 2, and 3—perfectly complementary to the CXCR4 siRNA were transfected with increasing concentrations of siRNA. Like the translational repression effect observed above, the level of

gene silencing by RNAi increases with increasing number of perfectly complementary binding sites and with increasing concentration of siRNA (Fig 3B). As might be expected from a mechanism that results in cleavage of the mRNA, RNAi silences gene expression to a greater extent than translational repression.

The mechanism of mRNA cleavage in RNAi implies that each siRNA:binding site interaction will function independently of another interaction; once a mRNA is cleaved it is expected to be rapidly degraded, and thus a second cleavage event would have little if any effect on gene expression. To assess this, we divided the repression observed for each construct in figure 3B by the number of binding sites on that mRNA, at each concentration of siRNA. These values were then normalized to the repression observed for a single binding site to assess the relative contribution of each site (Fig. 3D). As expected, the relative effectiveness of each site remained the same as the number of binding sites increased. Addition of more binding sites likely only increases the probability of the single necessary cleavage event, and thus multiple binding sites function independently of one another. This same analysis was applied to the translational repression constructs in figure 3A, normalizing to the construct with two binding sites (Fig. 3C). Strikingly, the degree of repression achieved by increasing the number of sites is not simply additive, as each site in the construct with four binding sites conferred twice as much repression as each site in the construct with two binding sites. The effectiveness of each binding site in the construct with six sites was similar to that of the construct with four sites. These results suggest that the effects of binding multiple miRNA complexes to the 3' UTR are likely to be cooperative. Ribonucleoprotein complexes could either mutually stabilize one another or cooperatively interact to more effectively inhibit translation or both. As with other cooperative

interactions in gene regulation, this would allow a cell to fine-tune the expression of a mRNA by regulating the degree of binding of different miRNAs to the 3' UTR of the mRNA.

The discovery that siRNAs can function in translational repression as miRNAs, and that the sequence requirements for this interaction are less stringent than those for RNAi, may help to explain non-specific effects sometimes observed in experiments utilizing siRNAs for targeted gene silencing. Using an arbitrary 21 nt sequence, BLAST searches against the mRNAs predicted from the human genome identify multiple inexact matches with 16-18 nt complementarity. Combined with the potential of GU wobble basepairs, and depending on the overall sequence rules for translational repression, there may be translational repression of a number of off-target genes by the introduction of a siRNA intended to knock-down the expression of only the target gene. However, the mechanistic finding that several binding sites are needed to produce a significant effect on protein expression may make non-specific siRNA effects the exception rather than the rule, and to date siRNAs have certainly been used with ostensible specificity.

Materials and Methods

DNA constructs and siRNAs

3' UTR binding sites for the siRNAs were constructed by a multimerization of DNA oligonucleotides (IDT), gel purification, PCR amplification, and restriction digestion. The products were inserted into the XbaI site immediately downstream of the stop codon in either the pRL-TK vector coding for the *Renilla reniformis* luciferase (*Rr*-luc) or the pGL3 control vector coding for the *Photinus pyralis* luciferase (*Pp*-luc) (Promega). siRNAs were purchased as single strands, deprotected, and annealed according to the manufacturer (Dharmacon). All sequences for siRNAs and 3' UTR constructs used in this study are available on the Sharp Lab website at http://web.mit.edu/sharplab/RNAi/sequences.html

Cell culture and transfections

Adherent HeLa cells were grown in 10% IFS in DMEM, supplemented with glutamine in the presence of antibiotics. For all transfections, except those noted below, cells were transfected with Lipofectin and the PLUS reagent (Invitrogen). On the day before transfection, exponentially growing cells were trypsinized and plated into 24-well plates at a density of $3x10^4$ cells/well in antibiotic-free media. The next day the cells were transfected with 0.2 μg DNA and 25 nM siRNA in a final volume of 250 μL. For Fig. 1E and Fig. 3, cells were transfected with Lipofectamine 2000, as during the course of this study we found that this reagent delivers effective doses of siRNAs at lower concentrations. On the day before transfection, cells were trypsinized and plated into 24-well plates at a density of $8x10^4$ cells/well in antibiotic-free media. The next day cells were transfected with 0.8 μg DNA and 5 nM siRNA, unless noted, in a final volume of 500 μL.

Luciferase assays

Dual-Luciferase assays (Promega) were performed 24 hours post-transfection according to the manufacturer's protocol and detected with an Optocomp I Luminometer (MGM Instruments). *Rr*-luc target vectors were co-transfected with control pGL3, and *Pp*-luc target vectors were co-transfected with a pRL-CMV control (Promega). Transfections were harvested 24 hours post-transfection, and the two luciferase activities consecutively assayed.

RT-PCR

Total RNA was harvested from transfected HeLa cells using the RNAeasy kit (Qiagen). Total RNA was DNase treated twice with DNase-Free (Ambion) and reverse transcribed using Omniscript reverse transcriptase (Qiagen) with a DNA primer complementary to a region near the SV40 polyadenylation sequence found in both the *Pp*-luc and *Rr*-luc reporter vectors (5'-GCATTCTAGTTGTGGTTTGTCC). Trace radiolabeled PCR products were detected via autoradiography, and quantitated with ImageQuant software v. 1.2 (Molecular Dynamics).

Northern Analysis

Cytoplasmic RNA was harvested by hypotonic lysis without detergent and subsequent needle homogenization of HeLa cells 24h after transfected using Lipofectamine 2000. Nuclei were pelleted at 1500 x g for 15 min and the supernatant treated with proteinase K, extracted in phenol:chloroform and again in chloroform, precipitated with isopropanol and washed with 70% ethanol. Samples were then treated with DNase-Free (Ambion). Northern analysis was performed using the NorthernMax kit (Ambion). 10 μg of RNA from the (+) siRNA or (-) siRNA samples were separated by electrophoresis on a 1% formaldehyde agarose gel and transferred onto Hybond N+ nitrocellulose by downward transfer (Amersham Pharmacia). The 1.5 kb ORF of the *Pp*-luc cDNA was generated by restriction digest of the pGL3 control vector with XbaI and HindIII (New England Biolabs), and used with DECA-Prime II (Ambion) in the

presence of ^{32}P-dATP to generate a random-primed DNA used to probe the membrane. The membrane was stripped and reprobed with ß-actin probe, generated from DECAtemplate-ß-actin-mouse (Ambion).

Acknowledgements

The authors thank A. Grishok for helpful discussion, and C. Novina, D. M. Dykxhoorn, H. Houbaviy, D. Tantin, and R. Bodner for comments on the manuscript. J.G.D. is a Howard Hughes Medical Institute Predoctoral Fellow. C.P.P. is a National Science Foundation Predoctoral Fellow.

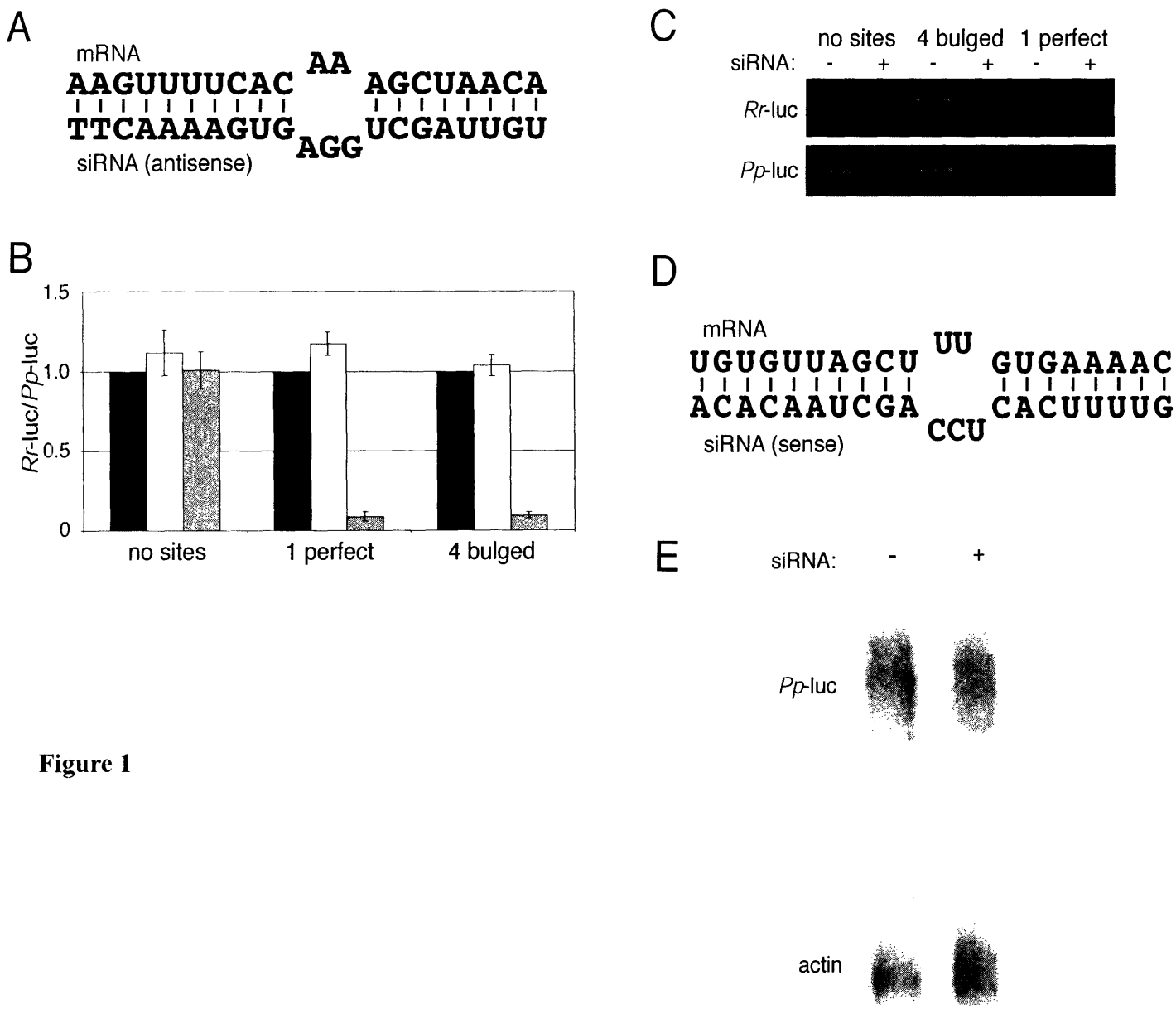
A
mRNA
AAGUUUUCAC AA AGCUAACA
TTCAAAAGUG AGG UCGAUUGU
siRNA (antisense)
B
Rr-luc/Pp-luc
1.5
1.0
0.5
0
no sites
1 perfect
4 bulged
C
no sites 4 bulged 1 perfect
siRNA: - + - + - +
Rr-luc
Pp-luc
D
mRNA
UGUGUUAGCU UU GUGAAAAC
ACACAAUCGA CCU CACUUUUG
siRNA (sense)
E
siRNA: - +
Pp-luc
actin

Figure 1

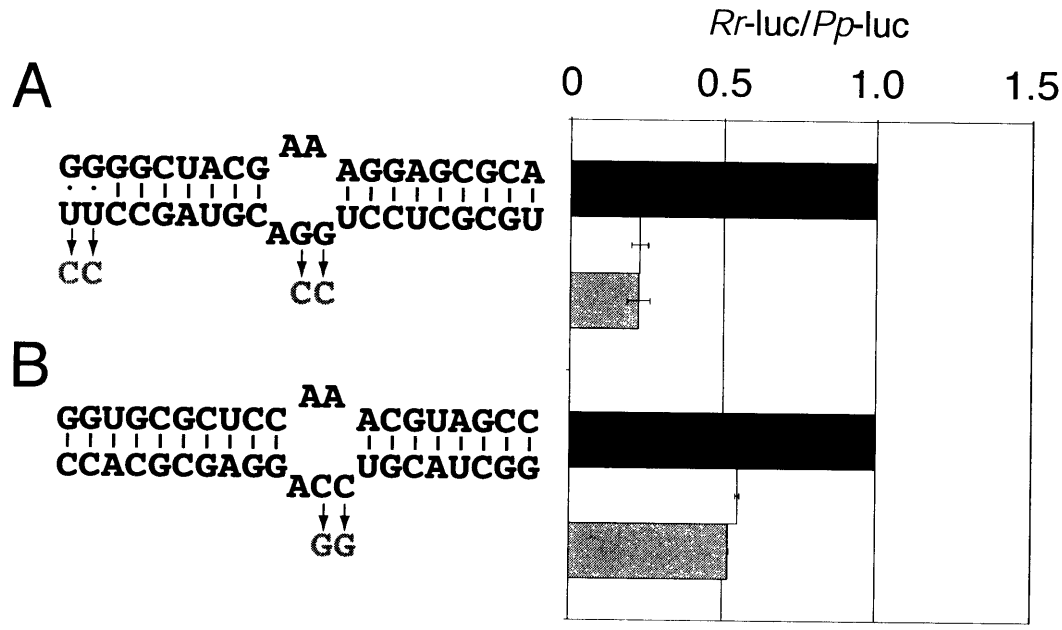
A
GGGGCUACG
AA
AGGAGCGCA
UUCCGAUGC
AGG
UCCUCGCGU
CC
CC
B
GGUGCGCUCC
AA
ACGUAGCC
CCACGCGAGG
ACC
UGCAUCGG
GG
Rr-luc/Pp-luc
0
0.5
1.0
1.5

Figure 2

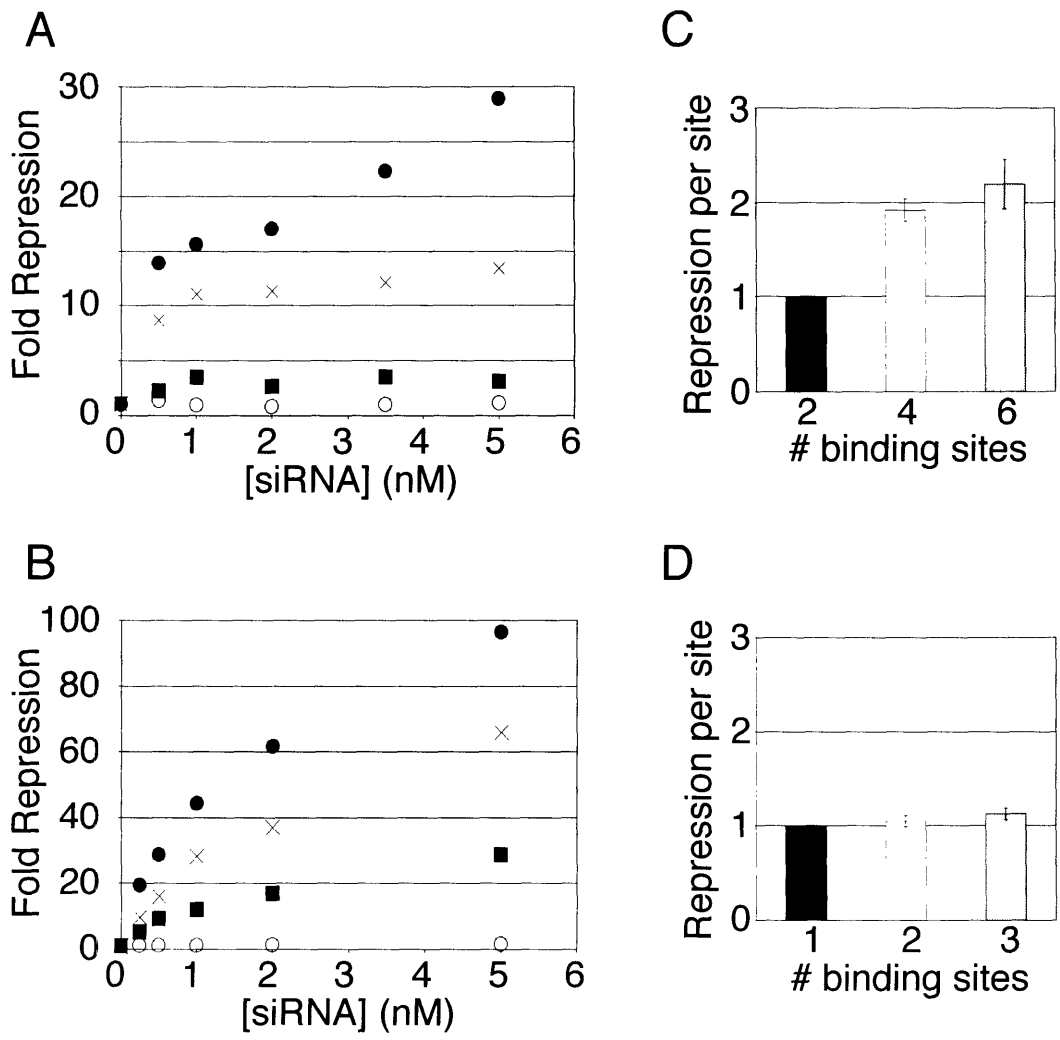
A
Fold Repression
30
20
10
0
0 1 2 3 4 5 6
[siRNA] (nM)
B
Fold Repression
100
80
60
40
20
0
0 1 2 3 4 5 6
[siRNA] (nM)
C
Repression per site
3
2
1
0
2 4 6
binding sites
D
Repression per site
3
2
1
0
1 2 3
binding sites

Figure 3

Figure Legends

Figure 1. siRNAs translationally represses a target mRNA. **(A)** Schematic of the proposed interaction between a binding site engineered into the 3' UTR of the target mRNA and the antisense strand of the CXCR4 siRNA. The thymidines at the 3' end of the siRNA are deoxynucleotides. **(B)** Dual Luciferase assay of transfected HeLa cells. Three *Renilla reniformis* luciferase (*Rr*-luc) constructs were used in this assay. One was unmodified ("no sites"), one contained a binding site perfectly complementary to the siRNA strand shown in (A) ("1 perfect"), and one contained four of the binding sites shown in (A) in tandem repeat ("4 bulged"). A *Photinus pyralis* luciferase (*Pp*-luc) served as an internal transfection control. The cells were transfected with no siRNA (black bars), a non-specific (targeting GFP) siRNA (white bars), or the CXCR4 siRNA (gray bars). The ratios of *Rr*-luc to *Pp*-luc expression were normalized to the no siRNA transfections, +/- S.E. from three independent experiments. **(C)** RT-PCR of harvested RNA. Total RNA was harvested from cells transfected with the constructs described in (B), transfected with or without the CXCR4 siRNA. Control experiments demonstrate that DNA was successfully removed from the RNA preparation and that the PCR was in the linear range of amplification (data not shown). **(D)** Schematic of the proposed interaction between the sense strand of the CXCR4 siRNA and a designed binding site. **(E)** RNA analysis of *Pp*-luc with four bulged CXCR4 binding sites (shown in A), targeted for translational repression, transfected either with the CXCR4 siRNA (+) or no siRNA (-). RNA was detected by Northern analysis, probing for either *Pp*-luc or ß-actin.

Figure 2. Analysis of sequence and structure rules for siRNA:mRNA interaction. HeLa cells were transfected with constructs containing four binding sites in tandem repeat with imperfect

complementarity to either the antisense (A) or sense (B) strand of a GFP siRNA. The effect on luciferase expression is shown by the white bars, +/- S.E. from two independent experiments, normalized to cells transfected with no siRNA (black bars). A different siRNA was then used to produce different bulges, shown in gray with arrows. These new interactions were assayed and are depicted with gray bars.

Figure 3. Comparison of RNAi and translational repression. **(A)** Titration of *Rr*-luc constructs containing 0 (O), 2 (■), 4 (X), or 6 (●) of the bulged binding sites, for pairing with the antisense strand of the CXCR4 siRNA, as depicted in Fig. 1A. The level of repression achieved is plotted, normalized to cells transfected with no siRNA. **(B)** Titration of *Pp*-luc constructs containing 0 (O), 1 (■), 2 (X), or 3 (●) binding sites perfectly complementary to the antisense strand of the CXCR4 siRNA (see Fig. 1A). **(C)** Analysis of the relative repression each site contributes for the data presented in (A), normalized to the construct with two binding sites, +/- S.E. **(D)** Analysis of the relative repression each site contributes for the data presented in (B), normalized to the construct with one binding site, +/- S.E.

References

Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M. and Gvozdev, V. A. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* **11:** 1017-27.

Banerjee, D. and Slack, F. 2002. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays* **24:** 119-29.

Baulcombe, D. 2001. RNA silencing. Diced defence. *Nature* **409:** 295-6.

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409:** 363-6.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. 2001a. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494-8.

Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. and Tuschl, T. 2001b. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J* **20:** 6877-88.

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

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G. and Mello, C. C. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106:** 23-34.

Ha, I., Wightman, B. and Ruvkun, G. 1996. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. *Genes Dev* **10:** 3041-50.

Hamilton, A. J. and Baulcombe, D. C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286:** 950-2.

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

Holen, T., Amarzguioui, M., Wiiger, M. T., Babaie, E. and Prydz, H. 2002. Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res* **30:** 1757-66.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T. and Zamore, P. D. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293:** 834-8.

Hutvagner, G. and Zamore, P. D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297:** 2056-60.

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J. and Plasterk, R. H. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15:** 2654-9.

Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. and Plasterk, R. H. 1999. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99:** 133-41.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294:** 853-8.

Lau, N. C., Lim, L. P., Weinstein, E. G. and Bartel, D. P. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294:** 858-62.

Lee, R. C. and Ambros, V. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294:** 862-4.

Lee, R. C., Feinbaum, R. L. and Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75:** 843-54.

Lee, Y., Jeon, K., Lee, J. T., Kim, S. and Kim, V. N. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21:** 4663-70.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. 2002. Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi. *Cell* **110:** 563.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16:** 720-8.

Olsen, P. H. and Ambros, V. 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216:** 671-80.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R. and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403:** 901-6.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. and Bartel, D. P. 2002. MicroRNAs in plants. *Genes Dev* **16:** 1616-26.

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R. and Ruvkun, G. 2000. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell* **5:** 659-69.

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A. and Mello, C. C. 1999. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99:** 123-32.

Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. and Sharp, P. A. 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev* **13:** 3191-7.

Wightman, B., Ha, I. and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75:** 855-62.

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

Zeng, Y., Wagner, E. J. and Cullen, B. R. 2002. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9:** 1327-33.

Chapter Three

Specificity of microRNA target selection in translational repression

This chapter is presented in the context of its contemporary science, and originally appeared in

Genes and Development **18**: 504 – 11 (2004)

Abstract

MicroRNAs (miRNAs) are a class of non-coding RNAs found in organisms as evolutionary distant as plants and mammals, yet most of the mRNAs they regulate are unknown. Here we show that the ability of a miRNA to translationally repress a target mRNA is largely dictated by the free energy of binding of the first 8 nucleotides in the 5' region of the miRNA. However, G:U wobble basepairing in this region interferes with activity beyond that predicted on the basis of thermodynamic stability. Furthermore, a mRNA can be simultaneously repressed by more than one miRNA species. The level of repression achieved is dependent on both the amount of mRNA and the amount of available miRNA complexes. Thus, predicted miRNA:mRNA interactions must be viewed in the context of other potential interactions and cellular conditions.

Introduction

The canonical RNA interference (RNAi) pathway begins with the cleavage of long, double-stranded RNA into an intermediate RNA species of ~21 nucleotides (nt) known as short, interfering RNA (siRNA)(reviewed in Zamore 2002; Dykxhoorn et al. 2003). These siRNA are double-stranded, with 5' phosphates and 2 nt 3' overhangs, indicators of RNaseIII cleavage, and indeed, the enzyme Dicer was identified as responsible for their generation (Bernstein et al. 2001). One of the two strands of the siRNA is incorporated into the RNA Induced Silencing Complex (RISC) (Hammond et al. 2000; Martinez et al. 2002; Khvorova et al. 2003; Schwarz et al. 2003). This strand then guides RISC to perfectly complementary mRNAs and cleaves them, resulting in their degradation. Several labs cloned short RNA species in order to find endogenous siRNAs, and these efforts led to the discovery of miRNAs as a large class of non-coding RNA (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001).

MicroRNAs are ~22 nt single-stranded RNA species found in a wide variety of organisms, ranging from plants to worms to humans (reviewed in Lai 2003; Bartel 2004). The founding member of the miRNA class, the *C. elegans* gene *lin-4*, as well as its target, the nuclear protein *lin-14,* were first identified in a screen for worms with defects in cell lineage progression (Horvitz and Sulston 1980; Chalfie et al. 1981). After over a decade of research, it was determined that *lin-4* did not code for a protein, but rather a small RNA species with imperfect complimentarity to several sites in the 3' untranslated region (UTR) of *lin-14* (Lee et al. 1993). Because expression of *lin-4* led to a decrease in *lin-14* protein level without a decrease in mRNA level, this phenomenon was dubbed translational repression (Wightman et al. 1991; Wightman et al. 1993). Biochemical analysis revealed that the repressed mRNAs remain in polysomes,

suggesting that the block in expression occurs after translation initiation, though little is known about the mechanism (Olsen and Ambros 1999; Seggerson et al. 2002).

While the mechanism of miRNA action remains elusive, their biogenesis is rapidly becoming clear. Primary miRNA transcripts are first processed in the nucleus by the RNaseIII enzyme Drosha to produce a hairpin RNA of ~70 nt (Lee et al. 2003). In a pathway dependent on Exportin-5, this pre-miRNA is then exported into the cytoplasm (Yi et al. 2003; Lund et al. 2004), where Dicer then cuts the hairpin (Grishok et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001; Knight and Bass 2001; Lee et al. 2002). Correlative evidence suggests that the same rules governing siRNA strand choice also hold for determining which side of the hairpin becomes the mature strand of the miRNA (Schwarz et al. 2003). The complex containing active miRNAs and the RISC involved in RNAi are similar if not identical, as endogenous miRNAs can cleave mRNAs with perfect complementarity (Hutvagner and Zamore 2002), and exogenously introduced siRNAs can translationally repress mRNAs bearing imperfectly complementary binding sites (Doench et al. 2003; Saxena et al. 2003; Zeng et al. 2003).

In addition to *lin-4* regulation of *lin-14*, there are now several other miRNAs with known targets. In *C. elegans, let-7* regulates both *lin-41* (Reinhart et al. 2000; Slack et al. 2000) and *hbl-1* (Abrahante et al. 2003; Lin et al. 2003), and *lin-4* also regulates *lin-28* (Moss et al. 1997). In *Drosophila,* the *bantam* gene was found to encode a miRNA that regulates the proapoptotic gene *hid* (Brennecke et al. 2003). *miR-2* and *miR-13* were predicted to regulate genes containing the K box motif (Lai 2002), and recent experimental work has validated this prediction (Boutla et al. 2003). MicroRNAs have also been implicated in fat metabolism (Xu et al. 2003) and hematopoietic lineage differentiation (Chen et al. 2004), although no targets were confirmed in these studies. Of note, these mRNAs tend to contain several binding sites for the miRNA,

emphasizing the potential importance of synergistic binding of the miRNA to the target. This synergism has been directly demonstrated, as addition of multiple binding sites into a 3' UTR resulted in more efficient inhibition of translation than that expected from the sum of the effect of each binding site individually (Doench et al. 2003).

Computational approaches have recently been used to identify potential miRNA targets (Enright et al. 2003; Lewis et al. 2003; Stark et al. 2003). The methods employed by Lewis et al. and Stark et al. incoporated conservation of the mRNA target site in related organisms to separate signal from noise. Additionally, the studies by Enright et al. and Stark et al. relied on inferences from known miRNA:mRNA interactions, a relatively small dataset. There are hundreds of identified miRNAs, with the vast majority of their potential targets unknown, and we thus decided to experimentally investigate the miRNA:mRNA pairing rules.

Results

As we and others have previously demonstrated, a siRNA can translationally repress a target mRNA with imperfectly complementary binding sites in its 3' UTR, and thus the siRNA functions as a miRNA (Doench et al. 2003; Saxena et al. 2003; Zeng et al. 2003). To determine if any region of the miRNA:mRNA interaction was of primary importance, 3' UTR constructs were designed to contain two base mismatches to the miRNA, tiled across the length of the binding site (Fig. 1A). Two identical mutant binding sites, separated by 4 nucleotides, were flanked by two of the original binding sites, each 11 nucleotides away, and cloned in the 3' UTR of the *Renilla* luciferase gene. This arrangement mimics known miRNA target mRNAs, which tend to have several binding sites, and potentially allows synergetic interactions for translational repression (Ha et al. 1996; Doench et al. 2003; Lewis et al. 2003). These constructs were co-transfected into HeLa cells with a control plasmid encoding firefly luciferase, either with or without the CXCR4 siRNA. Luciferase assays revealed that mutations creating mismatches with the 5' region of the miRNA inactivated the repression while the other mutations had no effect (Fig. 1B). For example, mutant H, mismatched at positions 3 and 4, and mutant G, mismatched at positions 5 and 6, do not silence reporter expression beyond the threshold of ~5 fold repression which is contributed by the two flanking, original sites. The other mutants silence expression ~12 fold, which is equivalent to that observed with four original sites. As determined by Ribonuclease Protection Assay (RPA), the CXCR4 siRNA did not have a significant effect on steady-state luciferase mRNA levels (Fig. 1C).

To test if positions 3 through 6 of the miRNA were uniquely important for repression, additional 3' UTR mutants were constructed, creating individual mismatches between the miRNA and mRNA or bulges in the miRNA or mRNA (Fig. 1D). For all these constructs,

interactions in the 3' region of the miRNA were held constant and two mutant sites were flanked by two original sites, as in Figure 1A. Luciferase assays revealed that some mutations hindered repression more than others, and that most mutations were neither fully active nor fully inactive for repression. As a means of quantifying the potential interaction, the free energy of first eight nucleotides of the miRNA binding to the various UTR constructs was calculated, using the mFold server (Zuker 2003). Plotting the calculated ΔG against the fold repression revealed a strong correlation (Fig. 1E). Interactions with a free energy less than approximately -5 kcal/mol were not active in repression beyond the 5 fold repression contributed by the two flanking sites, while those greater than -6 kcal/mol were optimally active, yielding 12 fold repression. Under these conditions, there appears to be a critical free energy required for effective repression.

The importance of interactions with the 3' region of the miRNA were investigated in constructs where the binding site for the 5' region of the miRNA was held constant. Three additional mutant binding sites were made, mismatching four nucleotides at a time, and a fourth mutant mismatching the entire 3' region of the miRNA (Fig. 2A). ΔG was then calculated, again using mFold and introducing a small loop to simulate the binding of the 5' region of the miRNA (see Matherial and Methods), and plotted against fold repression (Fig. 2B). Unlike the 5' region of the miRNA, interactions in the 3' region were of minimal importance, as all mutants generated approximately 12 fold repression, with a single exception; this construct was repressed only 6.7 fold. In this case, the introduced mutations probably allowed the mRNA to form a stable hairpin, as revealed by mFold, potentially leading to decreased accessibility for the miRNA.

In the above examples, in which interactions in the 3' region were not important, the stability of the miRNA:mRNA interaction in the 5' region was high (-9.1 kcal/mol). If this

interaction was energetically weaker but still fully effective, mutations in the 3' region might become more important. Thus, two 5' region mutants were combined with a 3' region mutant, again flanked by two original CXCR4 binding sites (Fig. 2C). Whereas the 5' region mutants each give full repression with a perfectly complementary 3' region (11.2 and 12.1 fold repression), they yielded no repression above baseline (4.1 and 4.1 fold repression) when basepairing in the 3' region was very weak. We conclude that the 5' region of the miRNA is the more important determinant of repression, but that the 3' region can also modulate this effect.

The role of G:U wobble basepairs, which are thermodynamically favorable and are common in RNA secondary structure, was investigated in the context of miRNA:mRNA interactions. Three mutant UTRs were constructed with single G:U wobbles, and one mutant was constructed with G:U wobble at three positions. Surprisingly, a single G:U wobble was detrimental to translational repression despite having a favorable ΔG value, and three G:U wobble pairings eliminated activity entirely (Fig. 3). A G:U wobble at position 3 in the 5' region reduced repression from 12 fold to 6 fold in spite of the fact that this pairing was not predicted to reduce the stability of the miRNA:mRNA interaction. Similarly, the mutant with three G:U basepairings had a theoretical stability of –6.3 kcal/mol in the 5' region, a value consistent with full repression with previous mutants, but was inactive in this assay.

To confirm that many of the above observations were also true for an endogenous miRNA, nine 3' UTRs were constructed, containing two binding sites each, that are predicted to basepair to endogenous *let-7a* miRNA with varying ΔG values in the 5' region (Fig. 4A). We note that, unlike in previous experiments, these constructs do not contain flanking binding sites. *let-7a* was chosen because it is known to be highly expressed in HeLa cells and paralogs expressed in HeLa cells share the same 8 nucleotides in the 5' region (Lagos-Quintana et al.

2001; Lim et al. 2003). Again, the degree of repression correlated with the ΔG values (Fig. 4B, gray bars). However, under conditions of pairing with endogenous *let-7a*, construct D, with a free energy value of –6.3 kcal/mol, was essentially inactive for repression. This contrasts with previous results with transfected siRNAs where values of –5 to –6 kcal/mol were active. To determine if this difference could be due to the concentration of miRNA, the experiment was repeated with additional *let-7a* introduced as a siRNA (Fig. 4B, white bars). As expected, additional *let-7a* did not lead to any repression of constructs with weak ΔG values (constructs B, C, and E). Interestingly, only a modest increase in repression (38%) was observed for construct A, with the strongest ΔG value (-11.0 kcal/mol), yet for construct D, with a near-threshold ΔG value of –6.3 kcal/mol, additional *let-7a* miRNA greatly increased repression (189%). Thus, miRNAs likely exist in a concentration dependent association with their binding sites, and the presence of more miRNAs increases these interactions, resulting in more repression. This model predicts that increasing the amount of mRNA would have the opposite effect. Indeed, exchanging the weak herpes virus thymidine kinase promoter for the strong CMV promoter in the construct with four original CXCR4 sites led to a dramatic decrease in repression, from 12 fold to less than 4 fold (data not shown).

The activity of the *let-7a* constructs also confirmed the detrimental effect of G:U wobble pairing (Fig. 4, constructs G, H, and I). A construct with a strong ΔG value, but with a G:U wobble at position 5, was not repressed with endogenous *let-7a* (construct G). Only upon addition of more *let-7a* could this construct be repressed. Furthermore, constructs with two G:U wobbles (constructs H and I) were not repressed by endogenous *let-7a*, nor did they significantly respond to additional *let-7a*.

We next examined the spacing requirements on the mRNA for miRNA interaction. Constructs with four original CXCR4 sites were used, and the distance between the two internal sites was varied. 3' UTRs with the two internal CXCR4 sites spaced by 4 or 0 nucleotides showed similar repression (Fig. 5, constructs A and B). To investigate possible steric hindrance between binding sites, constructs were designed such that the binding site for the 3' region of one CXCR4 siRNA would overlap with the binding site for the first four 5' nucleotides of another CXCR4 siRNA. To ensure that each internal site had a similar affinity for the miRNA, the binding site for the 3' region was disrupted in both sites. Perhaps surprisingly, this construct showed no decrease in repression (Fig. 5, construct C). However, if this overlap between the two sites was increased to nine nucleotides, the construct gave the same amount of repression as only one internal site (Fig. 5, compare constructs D and E). Because a binding site can prevent acces to a sufficiently nearby binding site, these results suggest that a factor stably associates with the mRNA. Indeed, miRNAs are thought to act by binding to their target mRNAs rather than by a catalytic mechanism requiring only a transient association between the miRNA and mRNA.

Combinatorial regulation, where two factors simultaneously regulate a single gene, is a common feature of eukaryotic cells. To test if a single mRNA could be repressed by more than one miRNA, two 3' UTR constructs were made, each of which contained two sites for the CXCR4 siRNA and two sites for a GFP siRNA (Fig. 6A). In order to avoid possible competition between the two siRNAs for access to protein assembly factors, the siRNAs were transfected at a less than saturating concentration (1 nM). The results indicate that two miRNAs can indeed simultaneously translationally repress a single mRNA (Fig. 6B). When either construct, GFP-CXCR4-CXCR4-GFP or CXCR4-GFP-GFP-CXCR4, was transfected with either siRNA alone, the degree of repression was approximately 3 fold. In contrast, cotransfection with both siRNAs

resulted in approximately 8 fold repression. Clearly, these reporters are being regulated by both siRNAs.

Discussion

We can draw several conclusions about miRNA:mRNA interactions from this study. First, the pairing of the miRNA 5' region to the mRNA is sufficient to cause repression, and the ΔG value of this interaction is an important determinant of activity. The 3' region of the miRNA is less critical, but can modulate activity in certain circumstances. Interestingly, G:U wobble pairing is highly detrimental to miRNA function despite its favorable contribution to RNA:RNA duplexes. These results support conclusions of recent computational investigations into miRNA target selection (Enright et al. 2003; Lewis et al. 2003; Stark et al. 2003), but also point towards potential improvements on the various methods. For example, the study by Lewis et al. required exact complementarity between 7 of the first 8 nucleotides of a miRNA and its target. However, our results suggest that a model based on a free energy of interactions is likely to better capture the possible targets of a miRNA. Stark et al. used thermodynamic parameters to generate their list of targets, but ranked their targets by the overall stability of the miRNA:mRNA interaction; our data show that the 5' region contributes more to specificty and activity. The study by Enright et al. allowed for G:U wobble pairing, but our results indicate that these interactions are strongly selected against in translational repression, perhaps as a means of preserving target specificty. Furthermore, the computational predictions allowed the possibility that a given mRNA can be regulated by more than one miRNA species, and our experiments validate this assumption.

Our studies on an endogenous miRNA, *let-7a,* indicate that a potential target must be evaluated in its cellular context. We demonstrate that a binding site which is not repressed by endogenous levels of miRNA becomes repressed upon addition of exogenous miRNA. Thus, the level of expression of both the mRNA and the miRNA, as well as potential competing binding

sites on other mRNAs, need be taken into account to determine whether the mRNA is endogenously regulated by the miRNA. For example, in one hypothetical scenario, a miRNA could be repressing a mRNA in a given cell type, but differentiation and subsequent expression of another mRNA, at higher levels and/or with stronger binding sites, could relieve the repression of the first mRNA. Validation of predicted miRNA:mRNA interactions by ectopic expression of either the mRNA target at artificially low levels, or the miRNA at artificially high levels, may "confirm" an interaction that does not exist *in vivo*. It is well-established that many miRNAs are limited in their expression to certain stages in development or to certain tissues and cell types (Bartel 2004). Computational prediction would be aided by taking into consideration expression profiling of both miRNA and mRNA levels, and biochemical methods or genetic analysis may be needed for definitive proof of a miRNA:mRNA interaction.

This study brings into focus the question of miRNA specificity. Indeed, miRNAs are an abundant species of RNA both in terms of the sheer number of miRNAs in the genome, currently estimated at 200 to 255 for the human genome (Lai 2003) and in terms of their expression levels, as some miRNAs are expressed at over 1,000 copies per cell (Lim et al. 2003). Additional factors may also be important for determining *in vivo* targets of miRNAs, such as the FMRP protein, a known regulator of mRNA expression that has been implicated in RNA silencing complexes (Caudy et al. 2002; Ishizuka et al. 2002). Alternatively, specificity may be entirely dictated by the sequence of the miRNA itself. That the thermodynamic stability of a region spanning only 8 nucleotides, a surprisingly low information content, is sufficient for miRNA activity may indicate a broad role for miRNAs in the regulation of gene expression.

Materials and Methods

Plasmid Construction

Two original CXCR4 sites, with XhoI and SpeI restriction sites between them, were inserted into the XbaI site in the 3' UTR of the pRL-TK plasmid (Promega). The mutant binding sites were then inserted by ligating annealied oligos into the XhoI and SpeI sites. Oligos were purchased from Qiagen, and all constructs were confirmed by sequencing. The *let-7a* and GFP constructs were made with the same strategy.

Cell Culture and Transfections

HeLa cells were maintained in DMEM with 5% calf serum and 5% inactivated fetal bovine serum, supplemented with glutamine and penicillin/streptomycin. The day before transfection, cells were seeded at 10^5 cells/well in a 24-well plate in antibiotic-free media, such that they would be 95% confluent at the time of transfection. Transfections were done with Lipofectamine 2000 according to the manufacturer's protocol (Invitrogen). 0.7 μg of pRL-TK plasmid and 0.1 μg of pGL3-Control plasmid (Promega) were used per well, and each sample was transfected in duplicate or triplicate. Transfections were done in a final volume of 0.5 ml, using siRNA at a final concentration of 5 nM (~0.03 μg). siRNAs were purchased from Dharmacon and prepared according to the manufacturer's instructions. Luciferase assays were performed 24 hours after transfection using the Dual-Glo luciferase kit (Promega).

Ribonuclease Protection Assay

HeLa cells were transfected in 6-well plates by scaling up the 24-well plate protocol by a factor of 5. 24 hours after transfection, total RNA was collected with the RNAeasy kit, including an

on-column DNase treatment (Qiagen). RNA probes were constructed by cloning PCR products into TOPO vectors (Invitrogen). The pGL3 probe corresponds to nucleotides 1142-1429 and was cloned into pCRII-TOPO, and the pRL-TK probe corresponds to nucleotides 1068-1297 and was cloned into pCR2.1-TOPO (position 1 of the plasmid as defined by the manufacturer). Transcription templates were linearized by SpeI restriction digestion (New England Biolabs) and transcribed in the presence of radiolabeled CTP (Perkin Elmer) using the T7 MAXIscript kit (Ambion). To allow for equivalent signals from the two mRNAs, the firefly luciferase probe was made with a five-fold lower specific activity. Ribonuclease Protection Assays were then performed with the RPA III kit, using 10 μg of RNA (Ambion). Gels were visualized on a Molecular Dynamics Storm 860 Phosphorimager, and quantitated with ImageQuant software version 1.2.

mFold Analysis

To determine ΔG values for the binding of the 5' region of the miRNA, the various mRNA binding sites were entered followed by "LLL" and then the first 8 nucleotides of the miRNA. The "LLL" tells mFold to treat the sequence as two separate RNA strands, and thus the initiation free energy, ΔI, is properly incorporated into the ΔG value (Zuker 2003). To determine ΔG values for the 3' region, the mRNA binding sites were entered followed by a loop of sequence "nnnGGGnnnnCCCnnn" and then the 3' region of the miRNA. The ΔG value of the loop alone is –1 kcal/mol, and this is included in the data shown. Because the siRNA used had two deoxythymidines at the 3' end, these were omitted from the free energy calculations, as indicated in the figures.

Acknowledgements

The authors thank C. Petersen for consistently helpful discussions and R. Bodner and D. Bartel for a critical reading of the manuscript. We also thank the Burge and Bartel labs for sharing data prior to its publication. J.G.D. is a Howard Hughes Medical Institute Predoctoral Fellow. This work was supported by United States Public Health Service MERIT Award R37-GM34277 from the National Institutes of Health and PO1-CA42063 from the National Cancer Institute to P.A.S. and partially by a Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute.

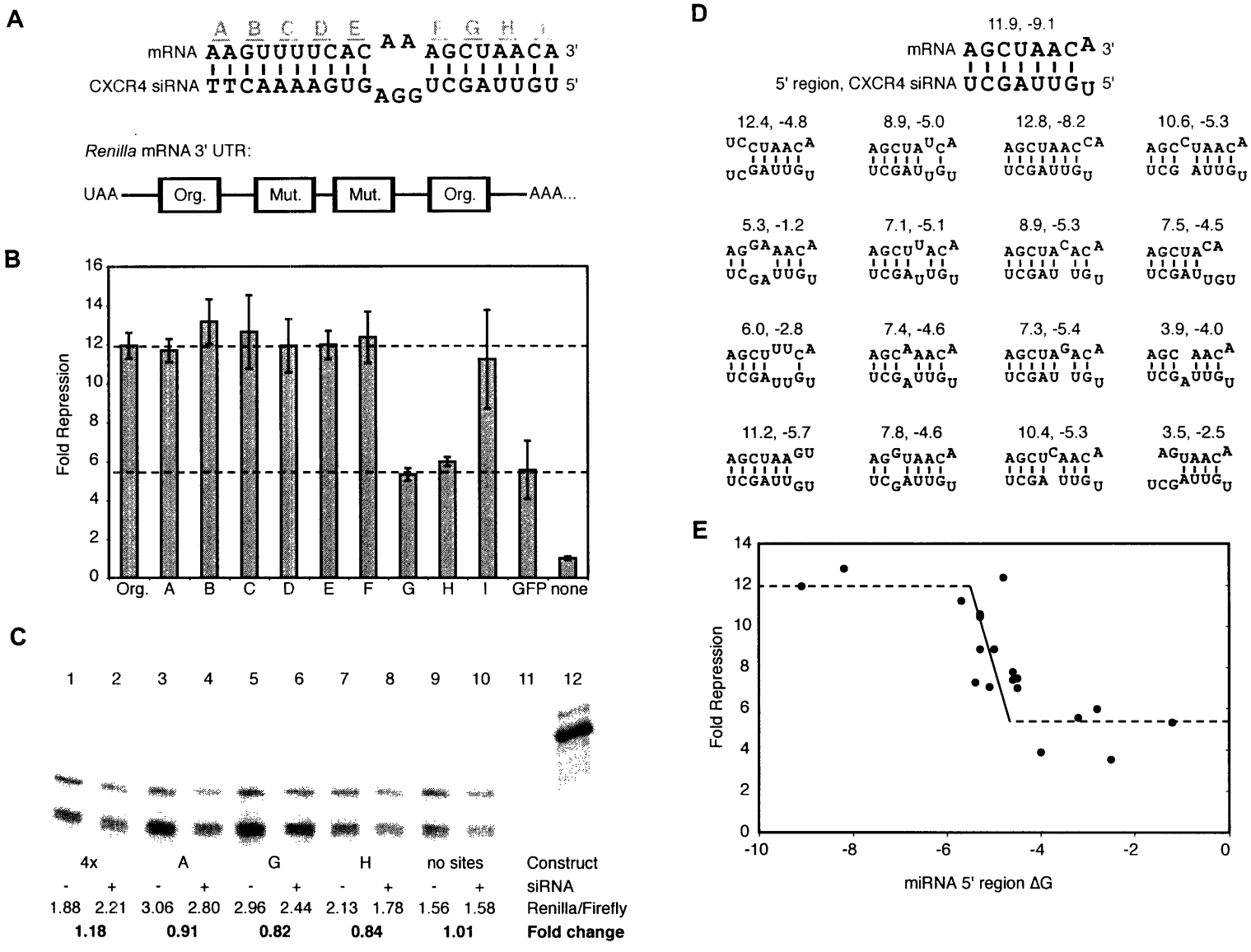
A
A B C D E F G H I
mRNA AAGUUUUCAC AA AGCUAACA 3'
CXCR4 siRNA TTCAAAAGUG AGG UCGAUUGU 5'
Renilla mRNA 3' UTR:
UAA
Org.
Mut.
Mut.
Org.
AAA...
B
Fold Repression
16
14
12
10
8
6
4
2
0
Org. A B C D E F G H I GFP none
C
1 2 3 4 5 6 7 8 9 10 11 12
4x A G H no sites Construct
- + - + - + - + - + siRNA
1.88 2.21 3.06 2.80 2.96 2.44 2.13 1.78 1.56 1.58 Renilla/Firefly
1.18 0.91 0.82 0.84 1.01 Fold change
D
11.9, -9.1
mRNA AGCUAACA 3'
5' region, CXCR4 siRNA UCGAUUGU 5'
12.4, -4.8
8.9, -5.0
12.8, -8.2
10.6, -5.3
5.3, -1.2
7.1, -5.1
8.9, -5.3
7.5, -4.5
6.0, -2.8
7.4, -4.6
7.3, -5.4
3.9, -4.0
11.2, -5.7
7.8, -4.6
10.4, -5.3
3.5, -2.5
E
Fold Repression
14
12
10
8
6
4
2
0
-10 -8 -6 -4 -2 0
miRNA 5' region ΔG

Figure 1

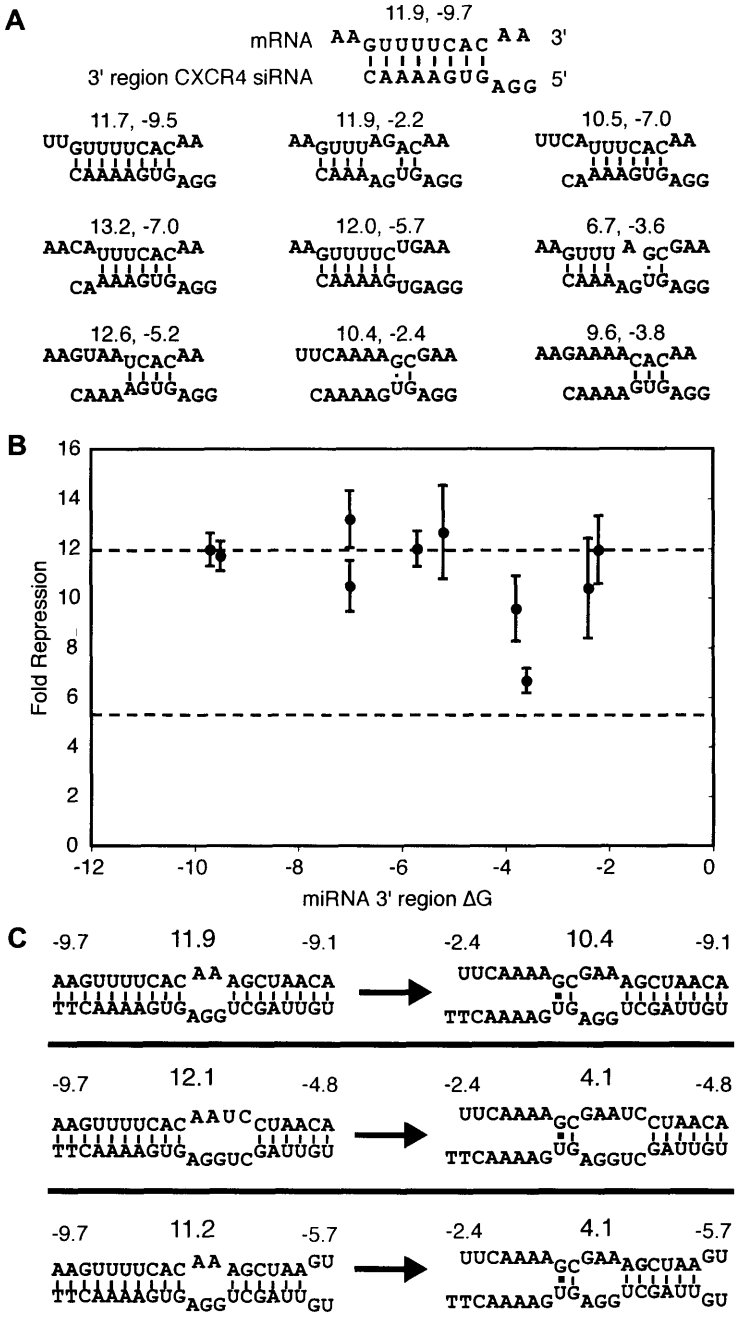
A
11.9, -9.7
mRNA AAGUUUUCAC AA 3'
3' region CXCR4 siRNA CAAAAGUG AGG 5'
11.7, -9.5
11.9, -2.2
10.5, -7.0
13.2, -7.0
12.0, -5.7
6.7, -3.6
12.6, -5.2
10.4, -2.4
9.6, -3.8
B
Fold Repression
miRNA 3' region ΔG
C
-9.7 11.9 -9.1
-2.4 10.4 -9.1
-9.7 12.1 -4.8
-2.4 4.1 -4.8
-9.7 11.2 -5.7
-2.4 4.1 -5.7

Figure 2

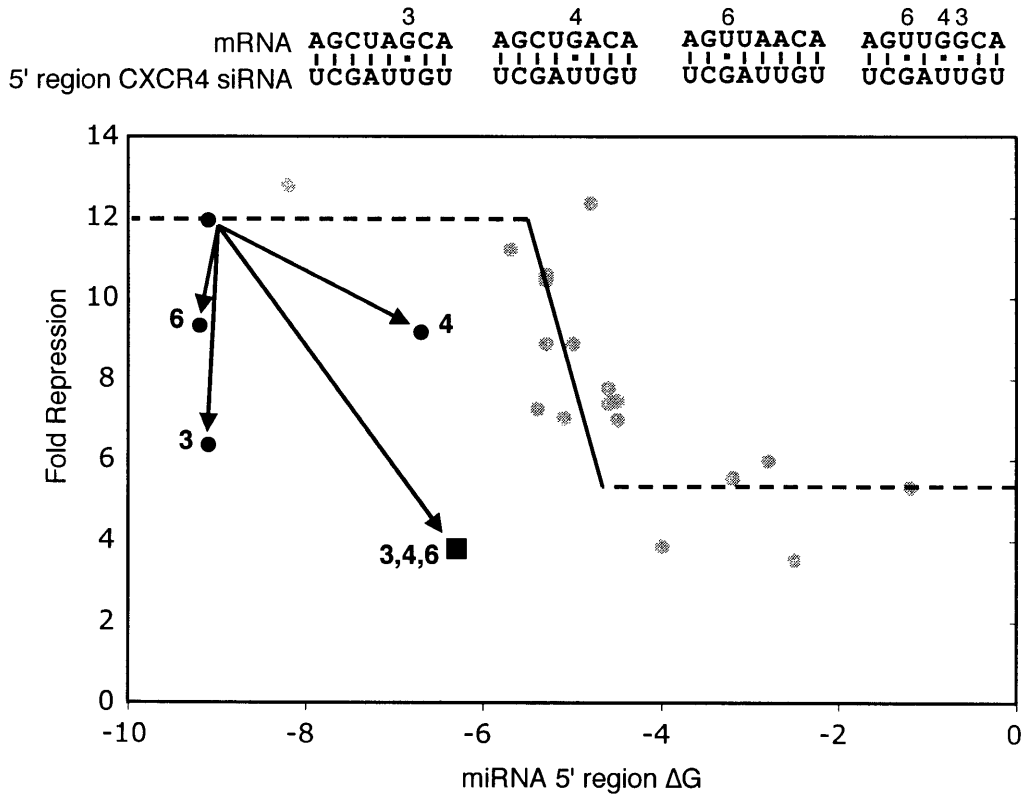
3
4
6
6 4 3
mRNA AGCUAGCA AGCUGACA AGUUAACA AGUUGGCA
5' region CXCR4 siRNA UCGAUUGU UCGAUUGU UCGAUUGU UCGAUUGU
Fold Repression
14
12
10
8
6
4
2
0
-10
-8
-6
-4
-2
0
miRNA 5' region ΔG
6
4
3
3,4,6

Figure 3

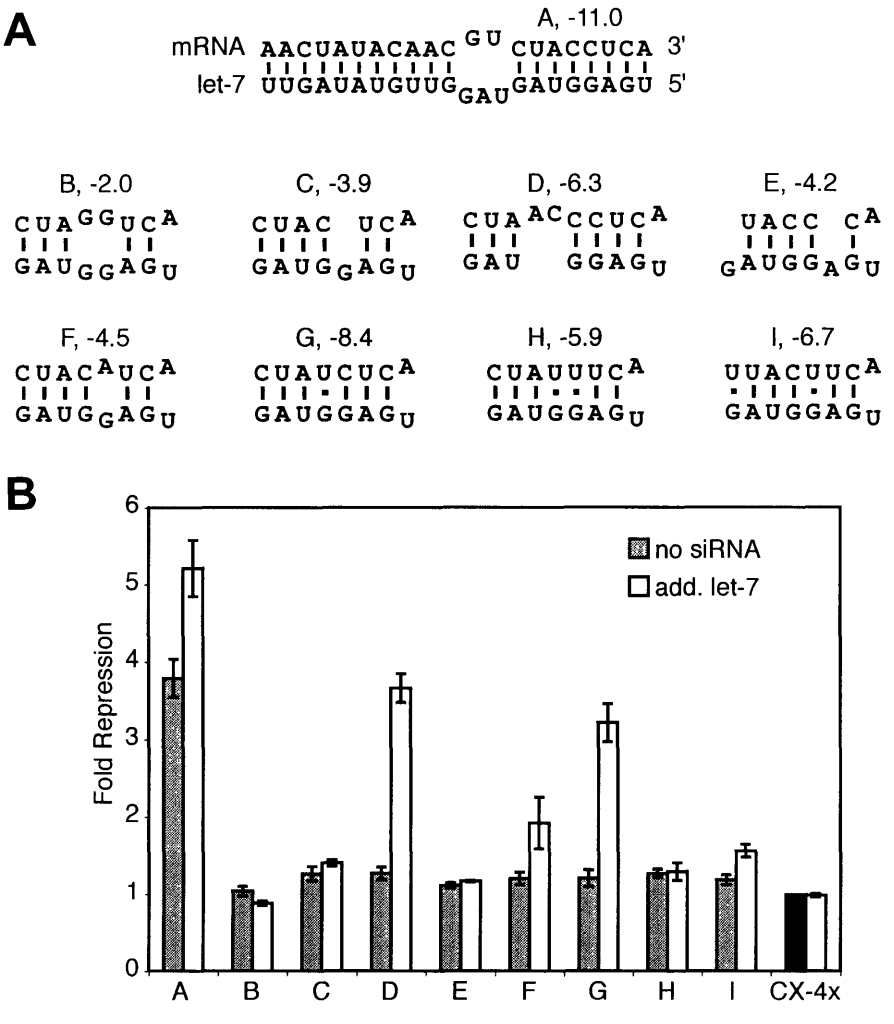
A
A, -11.0
mRNA AACUAUACAAC GU CUACCUCA 3'
let-7 UUGAUAUGUUG GAU GAUGGAGU 5'
B, -2.0
C, -3.9
D, -6.3
E, -4.2
F, -4.5
G, -8.4
H, -5.9
I, -6.7
B
Fold Repression
no siRNA
add. let-7
0
1
2
3
4
5
6
A
B
C
D
E
F
G
H
I
CX-4x

Figure 4

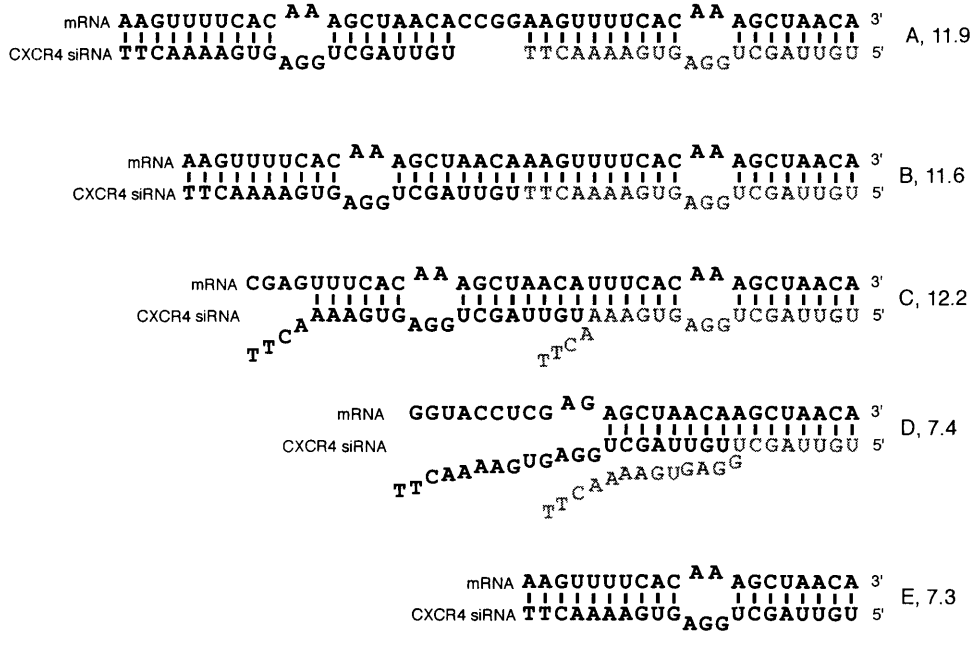
mRNA AAGUUUUCAC AA AGCUAACACCGGAAGUUUUCAC AA AGCUAACA 3'
CXCR4 siRNA TTCAAAAGUG AGG UCGAUUGU TTCAAAAGUG AGG UCGAUUGU 5'
A, 11.9
mRNA AAGUUUUCAC AA AGCUAACAAAGUUUUCAC AA AGCUAACA 3'
CXCR4 siRNA TTCAAAAGUG AGG UCGAUUGU TTCAAAAGUG AGG UCGAUUGU 5'
B, 11.6
mRNA CGAGUUUCAC AA AGCUAACAUUUCAC AA AGCUAACA 3'
CXCR4 siRNA AAAGUG AGG UCGAUUGU AAAGUG AGG UCGAUUGU 5'
T T C A
T T C A
C, 12.2
mRNA GGUACCUCG AG AGCUAACAAGCUAACA 3'
CXCR4 siRNA UCGAUUGU UCGAUUGU 5'
TT CAA AAG UG AGG
T T C A A AAG U GAG G
D, 7.4
mRNA AAGUUUUCAC AA AGCUAACA 3'
CXCR4 siRNA TTCAAAAGUG AGG UCGAUUGU 5'
E, 7.3

Figure 5

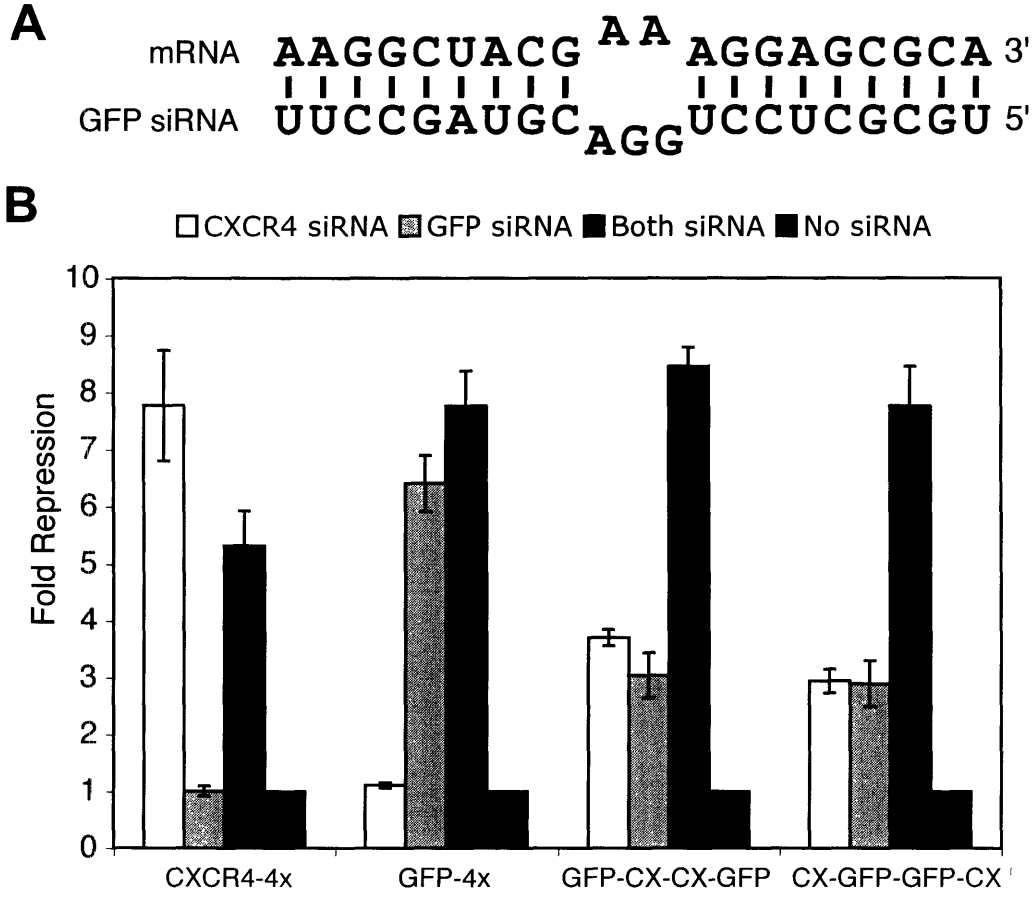
A
mRNA AAGGCUACG AA AGGAGCGCA 3'
GFP siRNA UUCCGAUGC AGG UCCUCGCGU 5'
B
CXCR4 siRNA
GFP siRNA
Both siRNA
No siRNA
Fold Repression
0
1
2
3
4
5
6
7
8
9
10
CXCR4-4x
GFP-4x
GFP-CX-CX-GFP
CX-GFP-GFP-CX

Figure 6

Figure Legends

Figure 1: 5' region of the miRNA determines translational repression.

(A) Schematic of the CXCR4 siRNA, antisense strand, basepairing to a designed 3' UTR binding site. The two 3'-most nucleotides are deoxythymidines. Mutations were made in the mRNA to form mismatches with the siRNA. In each case, the two nucleotide sequence of the mRNA was mutated to that of the siRNA. For example, mutant B contains a GU to CA mutation.

(B) Luciferase assay of mutant constructs. Constructs were transfected +/- siRNA, and fold repression determined. The upper dashed line corresponds to repression with four original sites, while the lower dashed line corresponds to repression with two original sites flanking two binding sites for an unrelated siRNA (targeting GFP), and thus serves as the lower bound for repression. The experiment was performed three times, and averages are presented +/- standard deviation.

(C) Ribonuclease Protection Assay of steady-state mRNA levels. The upper band corresponds to firefly luciferase mRNA (control), and the lower band to *Renilla* luciferase mRNA (targeted). Lane 12 is 5% of input probe, and lane 11 shows that no species are protected in untransfected HeLa cells. 4x is the construct with 4 original CXCR4 sites, while A, G, and H are described in (A). The *Renilla* mRNA level was normalized to the firefly, and then the fold change was calculated for each construct, dividing the +siRNA value into the -siRNA value; a value below 1 indicates a decrease in relative *Renilla* mRNA levels.

(D) Twelve additional mutants with alterations in the binding site for the first 8 nucleotides of the miRNA. The structure predicted by mFold is shown, and the original binding site is shown for comparison. The two numbers above each binding site correspond to the fold repression achieved and the calculated ΔG value.

(E) ΔG for the first 8 nucleotides of the miRNA binding to the mRNA, plotted against fold repression, for the mutants in (D) as well as mutants F through I from (A). The dashed lines correspond to the same bounds as in (B).

Figure 2: 3' region of the miRNA is rarely critical for repression.

(A) Nine mutants with alterations in the binding site for the 3' region of the miRNA. The structure predicted by mFold is shown, and the original binding site is shown for comparison. The nine sites shown are mutants A through E from Figure 1A, and four additional mutant constructs. The two numbers above each binding site correspond to the fold repression achieved and the calculated ΔG value.

(B) ΔG of the 3' region of the miRNA binding to the mRNA was calculated, and plotted against fold repression (+/- standard deviation from 3 independent experiments). The horizontal dashed lines are the same as in Figure 1.

(C) Effect of combined 5' and 3' binding site mutations. The left column shows the original binding site and two 5' binding site mutant constructs. The number centered above the binding site is the fold repression achieved, and the smaller numbers are the ΔG values for the binding of the 5' and 3' regions of the miRNA. Each construct on the left was then mutated in the 3' region binding site.

Figure 3: G:U wobble in the 5' region of the miRNA hinders repression. The 5' region of the CXCR4 siRNA binding to the mRNA is shown, as well as four mutant constructs which create G:U wobble pairing. These constructs were assayed and plotted on top of the data presented in Figure 1F. Arrows point from the original binding site to the 4 mutant constructs, and are

labeled with the position of the G:U wobble. Data points indicate the average of 3 independent experiments.

Figure 4: Endogenous *let-7a* confirms importance of miRNA 5' region.

(A) Schematic of a 3' UTR binding site, and its predicted interaction with endogenous *let-7a*, along with eight mutant binding sites for the 5' region of endogenous *let-7a*, together with the ΔG value. Constructs G, H, and I contain G:U wobble basepairs.

(B) Fold repression for the various constructs shown in (A). In gray is the fold repression achieved by endogenous *let-7a*. Expression values were first normalized internally to firefly luciferase expression, then across samples to the control construct, with 4 CXCR4 sites, shown in black. The constructs were then transfected with additional *let-7a*, and the fold repression in shown in white, again normalized to the expression of the control CXCR4-4x construct. Values are averages from 3 independent experiments, +/- standard deviation.

Figure 5: Distance requirements for miRNA accessibility. The binding sites inserted between two original CXCR4 sites are shown; for clarity, one of the CXCR4 siRNAs is shown in gray. The distance between the two sites was progressively reduced, until the 5' region of one site moved into the 3' region of the adjacent site. The fold repression achieved is indicated to the right of each schematic, the average of 3 independent experiments.

Figure 6: Two miRNAs can simultaneously repress a mRNA.

(A) Schematic of a binding site for a siRNA originally used to target GFP.

(B) Four constructs were transfected with either the GFP siRNA, the CXCR4 siRNA, both siRNAs, or no siRNA. One construct had 4 CXCR4 sites, one had 4 GFP sites, and two constructs had two of each, in the arrangement indicated. Fold repression was determined, normalized to the no siRNA transfection. The average of 3 independent experiments is shown, +/- standard deviation.

References

Abrahante, J.E., A.L. Daul, M. Li, M.L. Volk, J.M. Tennessen, E.A. Miller, and A.E. Rougvie. 2003. The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev Cell* **4**: 625-37.

Bartel, D. P. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism and Function. *Cell* **116**: 281-297.

Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-6.

Boutla, A., C. Delidakis, and M. Tabler. 2003. Developmental defects by antisense-mediated inactivation of micro-RNAs 2 and 13 in Drosophila and the identification of putative target genes. *Nucleic Acids Res* **31**: 4973-80.

Brennecke, J., D.R. Hipfner, A. Stark, R.B. Russell, and S.M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**: 25-36.

Caudy, A.A., M. Myers, G.J. Hannon, and S.M. Hammond. 2002. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* **16**: 2491-6.

Chalfie, M., H.R. Horvitz, and J.E. Sulston. 1981. Mutations that lead to reiterations in the cell lineages of C. elegans. *Cell* **24**: 59-69.

Chen, C.Z., L. Li, H.F. Lodish, and D.P. Bartel. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83-6.

Doench, J.G., C.P. Petersen, and P.A. Sharp. 2003. siRNAs can function as miRNAs. *Genes Dev* **17**: 438-42.

Dykxhoorn, D.M., C.D. Novina, and P.A. Sharp. 2003. Killing the messenger: short RNAs that silence gene expression. *Nat Rev Mol Cell Biol* **4**: 457-67.

Enright, A.J., B. John, U. Gaul, T. Tuschl, C. Sander, and D.S. Marks. 2003. MicroRNA targets in Drosophila. *Genome Biol* **5**: R1.

Grishok, A., A.E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D.L. Baillie, A. Fire, G. Ruvkun, and C.C. Mello. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**: 23-34.

Ha, I., B. Wightman, and G. Ruvkun. 1996. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. *Genes Dev* **10**: 3041-50.

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

Horvitz, H.R. and J.E. Sulston. 1980. Isolation and genetic characterization of cell-lineage mutants of the nematode Caenorhabditis elegans. *Genetics* **96**: 435-54.

Hutvagner, G., J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834-8.

Hutvagner, G. and P.D. Zamore. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056-60.

Ishizuka, A., M.C. Siomi, and H. Siomi. 2002. A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* **16**: 2497-508.

Ketting, R.F., S.E. Fischer, E. Bernstein, T. Sijen, G.J. Hannon, and R.H. Plasterk. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**: 2654-9.

Khvorova, A., A. Reynolds, and S.D. Jayasena. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209-16.

Knight, S.W. and B.L. Bass. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science* **293**: 2269-71.

Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-8.

Lai, E.C. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**: 363-4.

Lai, E.C. 2003. microRNAs: runts of the genome assert themselves. *Curr Biol* **13**: R925-36.

Lau, N.C., L.P. Lim, E.G. Weinstein, and D.P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858-62.

Lee, R.C. and V. Ambros. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**: 862-4.

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-54.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V.N. Kim. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415-9.

Lee, Y., K. Jeon, J.T. Lee, S. Kim, and V.N. Kim. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**: 4663-70.

Lewis, B.P., I. Shih, M.W. Jones-Rhoades, D.P. Bartel, and C.B. Burge. 2003. Prediction of mammalian microRNA targets. *Cell* **in press**.

Lim, L.P., N.C. Lau, E.G. Weinstein, A. Abdelhakim, S. Yekta, M.W. Rhoades, C.B. Burge, and D.P. Bartel. 2003. The microRNAs of Caenorhabditis elegans. *Genes Dev* **17**: 991-1008.

Lin, S.Y., S.M. Johnson, M. Abraham, M.C. Vella, A. Pasquinelli, C. Gamberi, E. Gottlieb, and F.J. Slack. 2003. The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev Cell* **4**: 639-50.

Lund, E., S. Guttinger, A. Calado, J.E. Dahlberg, and U. Kutay. 2004. Nuclear export of microRNA precursors. *Science* **303**: 95-8.

Martinez, J., A. Patkaniowska, H. Urlaub, R. Luhrmann, and T. Tuschl. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**: 563-74.

Moss, E.G., R.C. Lee, and V. Ambros. 1997. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**: 637-46.

Olsen, P.H. and V. Ambros. 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**: 671-80.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-6.

Saxena, S., Z.O. Jonsson, and A. Dutta. 2003. Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem* **278**: 44312-9.

Schwarz, D.S., G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P.D. Zamore. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199-208.

Seggerson, K., L. Tang, and E.G. Moss. 2002. Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Dev Biol* **243**: 215-25.

Slack, F.J., M. Basson, Z. Liu, V. Ambros, H.R. Horvitz, and G. Ruvkun. 2000. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell* **5**: 659-69.

Stark, A., J. Brennecke, R.B. Russell, and S.M. Cohen. 2003. Identfication of Drosophila MicroRNA Targets. *PLOS Biology* **1**: 1-13.

Wightman, B., T.R. Burglin, J. Gatto, P. Arasu, and G. Ruvkun. 1991. Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. *Genes Dev* **5**: 1813-24.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855-62.

Xu, P., S.Y. Vernooy, M. Guo, and B.A. Hay. 2003. The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol* **13**: 790-5.

Yi, R., Y. Qin, I.G. Macara, and B.R. Cullen. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**: 3011-6.

Zamore, P.D. 2002. Ancient pathways programmed by small RNAs. *Science* **296**: 1265-9.

Zeng, Y., R. Yi, and B.R. Cullen. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* **100**: 9779-84.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-15.

Appendix A

Establishment of a method to isolate endogenous miRNA targets

Introduction

MicroRNAs are a large class of genes, currently estimated to number 250 – 1000 in the human genome (Bartel 2004; Berezikov et al. 2005). These ~22 nucleotide RNAs have been shown to regulate gene expression at the level of mRNA stability and translation, yet the degree to which they shape the gene expression profile of the cell is not fully understood. The development of techniques that could facilitate experimental validation of miRNA targets would thus be quite useful for understanding the role these small RNAs play in orchestrating gene expression.

Currently, there are only a handful of fully validated miRNA:mRNA interactions, such as the regulation of *lin-14* by *lin-4, lin-41* by *let-7,* and *hid* by *bantam* (Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000; Brennecke et al. 2003). There are other examples where a miRNA has a known phenotype, but the target gene(s) is still unknown, such as the role of mir-181 in hematopoietic lineage differentiation (Chen et al. 2004). Because the specificity of a miRNA is largely conferred by only the first ~8 nucleotides, it is likely that miRNAs have many targets (Lewis et al. 2003; Doench and Sharp 2004). Indeed, computational approaches attempting to define miRNA interactions on a genome-wide scale have suggested that a substantial fraction of human mRNAs are conserved targets of miRNAs (Lewis et al. 2003; John et al. 2004; Lewis et al. 2005; Xie et al. 2005). One attempt at validating predicted targets confirmed 11 of 15 interactions, although it should be noted that these experiments were performed in a heterologous reporter system and it is not yet clear if this system faithfully recapitulates *in vivo* interactions (Lewis et al. 2003).

miRNAs are found associated with members of the Argonaute family of proteins (Mourelatos et al. 2002). Humans have four similar Ago proteins (Ago-1 to 4), and miRNAs

seem to associate equally well with each of them, at least when tagged versions are overexpressed (Meister et al. 2004). Ago-2 is the central component of the RISC in the RNAi pathway, using a single-stranded RNA to endonucleolytically cleave target mRNAs (Martinez et al. 2002; Liu et al. 2004). Thus, miRNAs that associate with Ago-2 are capable of cleaving their target mRNAs provided they have sufficient complementarity (Hutvagner and Zamore 2002; Yekta et al. 2004). The role of the other three Ago proteins is currently not known, but it is likely that at least one of them functions in the translational repression pathway. Mouse embryonic fibroblasts homozygous null for Ago-2 are no longer capable of cleaving mRNAs, but still translationally repress mRNAs with imperfect complementarity to exogenous small RNA (Liu et al. 2004).

Large-scale analysis of the effects of siRNA transfection into cultured cells has revealed that siRNAs exhibit a sequence-specific 'off-target' profile, with a few dozen genes down regulated (Jackson et al. 2003). Interestingly, most genes that are down regulated have complementary to the 5' region of the transfected miRNA, suggesting that these off-target effects are due to primary effects of the RNAi/miRNA pathways rather than secondary effects due to down regulation of the intended target gene. Additional microarray experiments have shown that transfection of the tissue-specific miRNAs *mir-1* and *mir-124,* specific to muscle and brain, respectively, shifts the mRNA profile of HeLa cells to a more muscle-like and brain-like signature (Lim et al. 2005). In these experiments, it is not clear if the mRNAs are down-regulated through RISC-mediated endocleolytic cleavage or if the small RNAs enter the translational repression pathway, and that activity leads to a modest steady-state decrease in mRNA levels.

The PIWI domain of Ago-2 bears much resemblence to the active site of RNase H, and biochemical analysis of RISC has shown that the mRNA is cleaved between the bases opposite the 10^{th} and 11^{th} nucleotides of the small RNA (Liu et al. 2004; Martinez and Tuschl 2004; Schwarz et al. 2004). The 5' cleavage product of the mRNA (the 7mG capped half) has a 3' hydroxyl, while the 3' cleavage product (the polyA half) has a 5' phosphate. Recent work from both plants and flies has shown that the 3' cleavage product is a substrate for degradation by the XRN family of nucleases (Souret et al. 2004; Orban and Izaurralde 2005). Human cells express two XRNs, XRN-1 and XRN-2. XRN-2 is known to localize to the nucleus, and has recently been shown to degrade cleaved, nascent transcripts and lead to transcription termination via a 'torpedo' mechanism (West et al. 2004). XRN-1 is cytoplasmic and has been implicated in degrading mRNAs in nonstop and nonsense mediate decay pathways, as well as decapped mRNAs (Parker and Song 2004).

Because the 3' products of Ago-2 mediated cleavage have a 5' phosphate, they are substrates for ligation by T4 RNA ligase. A modified 5' RACE assay can thus detect the exact site of mRNA cleavage (Llave et al. 2002; Yekta et al. 2004). Similarly, if Ago-2-associated miRNAs cleave their targets, even at low levels, a 5' RACE assay could detect these cleavage products and thus verify that a predicted miRNA:mRNA interaction occurs in a cellular context. Normally, RISC cleavage products are rapidly degraded and are detected at only low levels, if at all, on Northern blots, and thus may not be an abundant-enough substrate for reliable identification in a 5' RACE assay. However, the 3' cleavage product could be stabilized via siRNA-mediate knockdown of XRN-1. A combination of these techniques might allow for large-scale identification of miRNA targets.

Results

In order to formally implicate an XRN family member in the degradation of RNAi 3' cleavage products in mammalian cells, a ribonuclease protection assay (RPA) was optimized on the CXCR4 mRNA, as previous experiments have shown that the CXCR4 siRNA gives excellent reduction on the protein level. The probe was designed such that the siRNA would cleave near the middle of the protected sequence, thus allowing detection of both the full-length mRNA and any cleavage products in the same experiment (Figure 1a). This probe was optimized on control RNA from HeLa cells for probe:RNA ratio as well as digestion conditions.

siRNAs were designed against each of the two XRN family members in the human genome, using the siFinder Perl program (J.G.D. unpublished), based on siRNA design criteria (Reynolds et al. 2004). The XRN-1 and XRN-2 sequences were aligned, and the siRNAs were chosen in regions of minimal sequence homology to ensure specificity. These siRNAs were transfected into HeLa cells and RNA was harvested 48 hours post-transfection. Following reverse transcription, PCR was performed at several dilutions of cDNA to assay for mRNA knockdown, and all four siRNAs caused a reduction of their target XRN relative to control siRNAs; no reduction of the other, non-targeted XRN was observed. Finally, a real-time PCR assay was used to more accurately quantitate mRNA knockdown. The more active XRN-1 siRNA caused approximately five-fold reduction in mRNA level, while the more active XRN-2 siRNA caused approximately ten-fold reduction.

HeLa cells were transfected with siRNAs targeted to XRN-1, XRN-2, Ago-2, GFP, and *Renilla* luciferase, in addition to an untransfected control. 48 hours after the first transfection, the cells were transfected again, with the same siRNA as well as an additional siRNA, targeting

either GAPDH or CXCR4. 48 hours after the second transfection, RNA was harvested and an RPA performed (Figure 1b).

Transfection of the CXCR4 siRNA, as expected, led to a reduction in the level of full-length CXCR4 mRNA relative to the GAPDH siRNA control. Interestingly, a protected product of the expected size of the 3' cleavage product was detected in all of the samples transfected with the CXCR4 siRNA, but not in samples transfected with the GAPDH siRNA. This product is significantly more abundant in the sample that was first transfected with the XRN-1 siRNA, implicating XRN-1 in degrading RNAi 3' cleavage products. There was no stabilization of the 3' cleavage product in the sample first transfected with the XRN-2 siRNA; although this is a negative result, the real-time PCR data showing productive knockdown, as well as the known localization of XRN-2 to the nucleus, argues that XRN-2 does not degrade mRNAs that are RNAi cleavage products. Additionally, the RPA shows that reduction of Ago-2 via siRNA results in a modest stabilization of full-length CXCR4 mRNA as well as a reduction in the 3' cleavage product. Taken together, these data point towards the possibility of identifying endogeneous Ago-2 cleavage products in an unbiased manner.

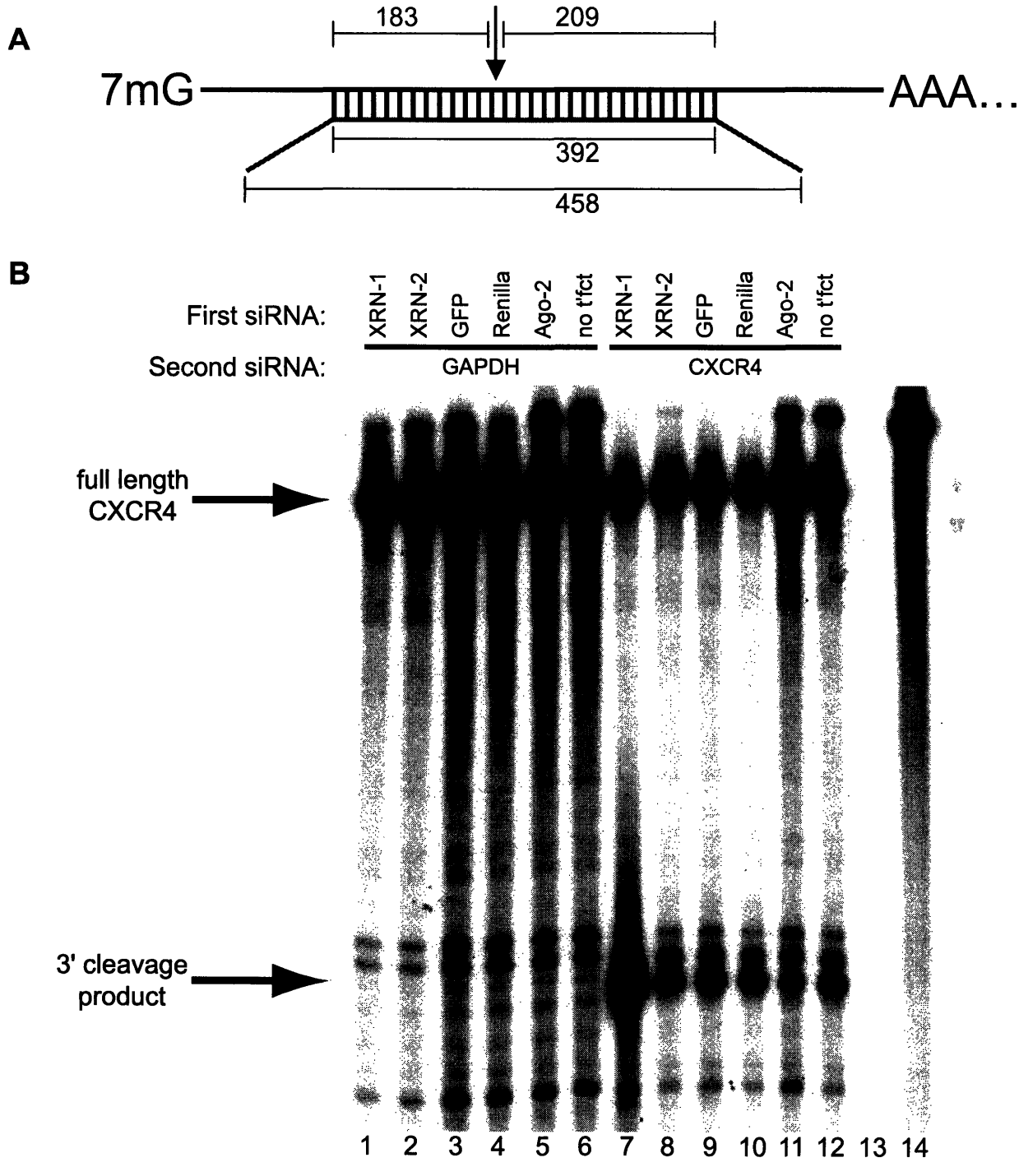

Figure 1

A. Schematic of probe used in the RPA to detect CXCR4 mRNA. The length of the undigested, full-length probe is 458 nucleotides; protected CXCR4 mRNA, 392 nt; 3' clevage product, 209 nt; 5' cleavage product, 183 nt.

B. Ribonuclease protection assay shows XRN-1 knockdown stabilizes RNAi 3' cleavage products. Cells were first transfected with the siRNA indicated, and then transfected with the first siRNA and either GAPDH or CXCR4 siRNA. Transfection with CXCR4 siRNA reduced steady-state levels of full length CXCR4 mRNA (compare lane 6 to 12); transfection with Ago-2 siRNA inhibited this reduction (lane 11). Furthermore, transfection of the CXCR4 siRNA also gave rise to a 3' cleavage product. This cleavage product was stabilized by knockdown of XRN-1 (lane 7), and less abundant with Ago-2 knockdown (lane 11). Lane 14 is undigested probe, and lane 13 is probe digested in yeast RNA.

Future Directions

The ability to down-modulate RNAi cleavage with an siRNA against Ago-2, and down-modulate cleavage product degradation with an siRNA against XRN-1, lends itself to large-scale identification of transcripts regulated by these pathways. Knockdown of Ago-2 is expected to cause an increase in steady-state levels of mRNA directly targeted by endonucleolytic cleavage, and these could be detected in high-throughput via microarray analysis or in a targeted approach by Northern blot. This knockdown would have pleiotropic effects on the RNA population in a cell, however, as the RNAi and miRNA pathways are likely to regulate a large number of genes. Thus, a second criteria is needed to eliminate false positives.

While knockdown of XRN-1 would not be expected to increase the steady-state levels of full-length mRNAs, any RNAi/miRNA 3' cleavage products would be stabilized. Ligation of an adaptor onto the 5' end of those products can allow for enrichment of these RNAs. Two general strategies are proposed, one involving exponential amplification through PCR, the other involing only linear amplification steps (Figure 2). In either approach, the starting RNA material is an important criteria. Total cellular RNA would be the least biased, but the high percentage of rRNA, which can serve as a substrate for T4 ligase due to its 5' phosphate, might cause unacceptable levels of background (Hannon et al. 1989). PolyA-selected RNA would reduce this background, although this selection procedure could bias the end results.

In the PCR-based, differential display approach, an RNA adaptor is ligated onto the 5' end of the RNA population. A reverse transcription step is performed, although unlike traditional differential display, the RT primer does not need to incorporate a primer site for future PCR. PCR is then employed using the ligated RNA adaptor as one primer, and one of a series of degenerate primers as the other primer; these primers are commercially available, and are

optmized for differential display analysis. The radiolabeled PCR products are run on a sequencing gel, and differences between the two RNA populations are visualized. Bands of interest can be excised and identified. This approach is reportedly more sensitive than microarray analysis, and, importantly, the exact site of the 5' phosphate and thus the cleavage site is revealed in sequencing.

A second approach also begins with ligation of an RNA adaptor onto the 5' phosphate of cleavage products. Reverse transcription is performed to make cDNA, followed by second-strand synthesis using a primer complementary to the ligated adaptor to make dsDNA. This dsDNA is then purified from the RNA in the sample. The sequence of the adaptor is the promoter for T7 RNA polyermase, and an *in vitro* transcription is performed. Now, all the RNA in the sample is of interest as it was a substrate for RNA ligation, and this pool of RNA can be analyzed via microarray. If large numbers of genes are RNAi/miRNA cleavage products, this technique may be more feasible and informative. However, unlike the differntial display approach, this technique does not give information as to the exact site of cleavage, and thus a modified 5' RACE assay would be needed to follow-up on cadidate genes.

Regardless of the method, the end result of these approaches will be lists of genes that may be endogenous substrate of Ago-2-mediated cleavage. However, other RNAs with 5' phosphates, such as those arising from RNase III or RNase H cleavage unrelated to the RNAi/miRNA pathway, would also appear in this analyis. Validation of real target genes could be performed in several ways. For example, if a candidate gene looks to be a target of a known miRNA, then 2'-O-Me inhibitors could be used to confirm that regulation (Hutvagner et al. 2004). Likewise, siRNA-mediated knockdown of Ago-2, Dicer, or other RNAi/miRNA pathway genes would be expected to increase the steady-state levels of the target mRNAs.

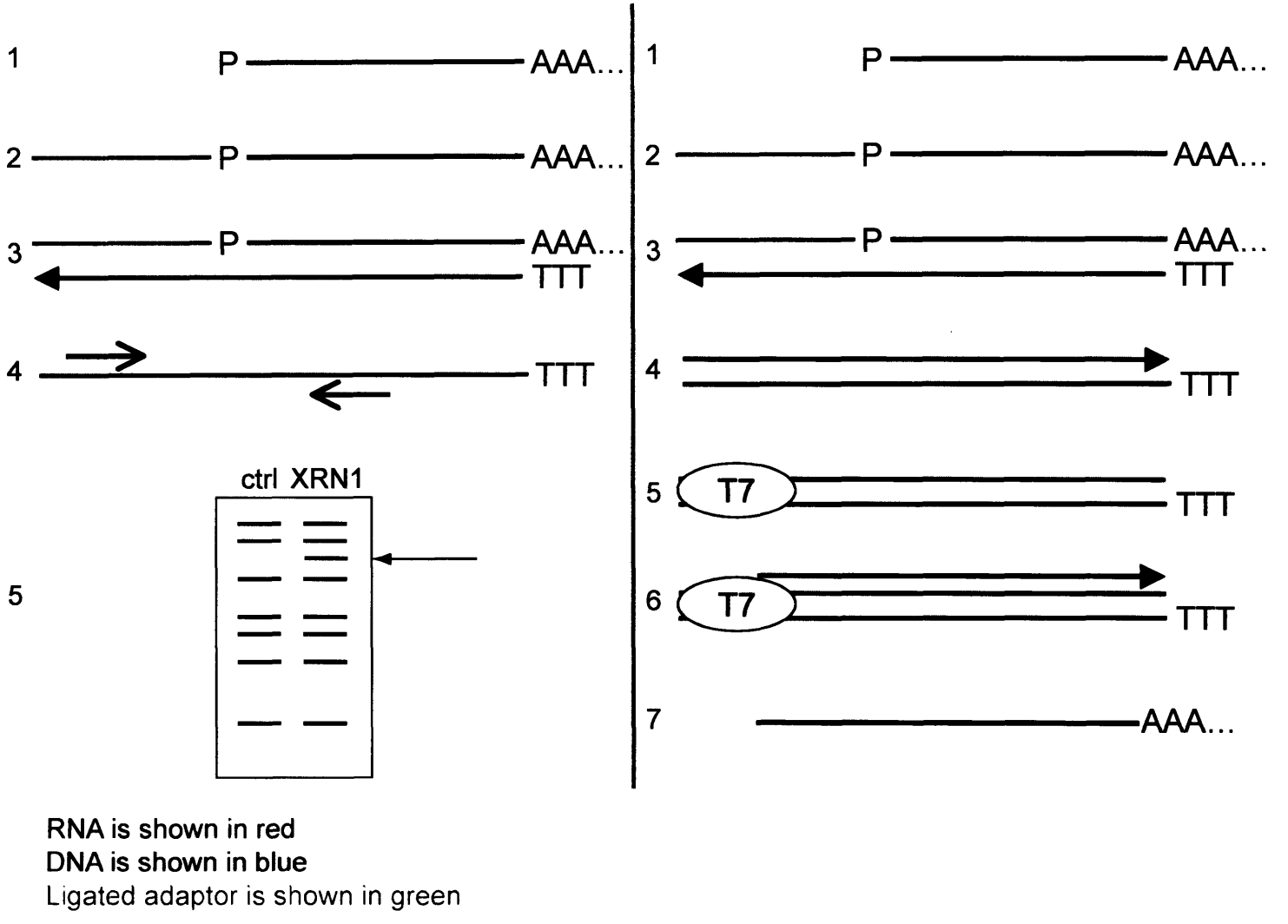

Figure 2. Comparison of methods to detect endogenous XRN-1 substrates.

Differential Display

1) RNA population contains RNA with a 5' phosphate.
2) This RNA serves as a substrate for ligation of an RNA adaptor with T4 RNA ligase.
3) Reverse transcriptase is used for synthesis of cDNA.
4) This cDNA serves as a substate for PCR, using one primer in the adaptor, and another primer binding to mRNA-derived sequence.
5) The radiolabeled PCR is then run on a gel, and compared to a control RNA population. Bands that appear in the XRN-1 siRNA-treated samples are excised and sequenced.

Microarray

1) RNA population contains RNA with a 5' phosphate.
2) This RNA serves as a substrate for ligation of an RNA adaptor with T4 RNA ligase.
3) Reverse transcriptase is used for synthesis of cDNA.
4) This cDNA is extended to dsDNA with a primer in the adaptor region and DNA polymerase.
5) The adaptor contains the binding site for T7 RNA polymerase.
6) An in vitro transcription is performed.
7) The RNA in the sample can serve as a substrate for microarray analysis.

References

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-97.

Berezikov, E., V. Guryev, J. van de Belt, E. Wienholds, R.H. Plasterk, and E. Cuppen. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**: 21-4.

Brennecke, J., D.R. Hipfner, A. Stark, R.B. Russell, and S.M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**: 25-36.

Chen, C.Z., L. Li, H.F. Lodish, and D.P. Bartel. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83-6.

Doench, J.G. and P.A. Sharp. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**: 504-11.

Hannon, G.J., P.A. Maroney, A. Branch, B.J. Benenfield, H.D. Robertson, and T.W. Nilsen. 1989. Accurate processing of human pre-rRNA in vitro. *Mol Cell Biol* **9**: 4422-31.

Hutvagner, G., M.J. Simard, C.C. Mello, and P.D. Zamore. 2004. Sequence-specific inhibition of small RNA function. *PLoS Biol* **2**: E98.

Hutvagner, G. and P.D. Zamore. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056-60.

Jackson, A.L., S.R. Bartz, J. Schelter, S.V. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, and P.S. Linsley. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**: 635-7.

John, B., A.J. Enright, A. Aravin, T. Tuschl, C. Sander, and D.S. Marks. 2004. Human MicroRNA targets. *PLoS Biol* **2**: e363.

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-54.

Lewis, B.P., C.B. Burge, and D.P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.

Lewis, B.P., I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, and C.B. Burge. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787-98.

Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769-73.

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

Llave, C., Z. Xie, K.D. Kasschau, and J.C. Carrington. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**: 2053-6.

Martinez, J., A. Patkaniowska, H. Urlaub, R. Luhrmann, and T. Tuschl. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**: 563-74.

Martinez, J. and T. Tuschl. 2004. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev* **18**: 975-80.

Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng, and T. Tuschl. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**: 185-97.

Mourelatos, Z., J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**: 720-8.

Orban, T.I. and E. Izaurralde. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**: 459-69.

Parker, R. and H. Song. 2004. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* **11**: 121-7.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-6.

Reynolds, A., D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, and A. Khvorova. 2004. Rational siRNA design for RNA interference. *Nat Biotechnol* **22**: 326-30.

Schwarz, D.S., Y. Tomari, and P.D. Zamore. 2004. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Curr Biol* **14**: 787-91.

Souret, F.F., J.P. Kastenmayer, and P.J. Green. 2004. AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol Cell* **15**: 173-83.

West, S., N. Gromak, and N.J. Proudfoot. 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**: 522-5.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855-62.

Xie, X., J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, and M. Kellis. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338-45.

Yekta, S., I.H. Shih, and D.P. Bartel. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**: 594-6.

Appendix B

A cell-based reporter system to identify miRNA inhibitors

Introduction

It is predicted that microRNAs regulate approximately one-third of the human genome, yet very little is known about how miRNAs themselves are regulated. One way that miRNAs are known to be regulated is at the level of transcription, with many miRNAs showing strong temporal and spatial regulation (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001; Lagos-Quintana et al. 2002). Northern blot analysis also suggests that miRNAs can be regulated at the level of export and/or Dicer processing, as pre-miRNAs can sometimes be detected before the mature miRNA appears (Hutvagner et al. 2001).

Two of the first examples of miRNA-mediated translational repression, *lin-4* regulation of *lin-14* and *let-7* regulation of *lin-41,* were both cases where the miRNA is turned on at one stage in development and represses the target gene for the life of the worm (Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000). These observations are, at least theoretically, somewhat at odds with the known mechanism of miRNA action, translational repression. Why stably repress a gene at a step so far downstream in the flow of genetic information? In other words, why would a cell invest the energy needed to express a gene up through the initiation of translation, and risk many steps of potential mis-regulation, if that gene never needs to be expressed again? Indeed, translational regulation is generally thought to allow for a rapid response to stimuli, such as the local activation of translation at stimulated neuronal synapses, or IRES mediated-translation during a specific phase of the cell cycle (Holcik and Sonenberg 2005). Perhaps, *lin-4* and *let-7* are the exception rather than the rule in this regard. *Bantam,* for example, is a *Drosophila* miRNA that is known to repress the pro-apoptotic gene *hid* (Brennecke et al. 2003). In this case, the miRNA-mediated repression is likely to be turned off in a rapid fashion, when the appropriate stimuli signal for cell death. Furthermore, FMRP, a protein that

associates both with polysomes and with miRNAs, is known to be a substrate for phosphorylation (Antar and Bassell 2003; Ceman et al. 2003; Jin et al. 2004; Stefani et al. 2004).

Based on these observations, it is reasonable to hypothesize that there are potential signaling pathways that regulate miRNAs and the translational repression pathway. A system devised to screen for such pathways could thus uncover fundamental biology concerning the function of miRNAs.

Results

A system was devised whereby a reporter gene, in this case firefly luciferase, reports on the activity of an endogenous miRNA, *mir-21;* this miRNA was chosen because it is known to be expressed at high levels in HeLa cells. Six binding sites with imperfect complementarity to *mir-21* were cloned into the 3' UTR of luciferase, which should result in a persistent repression of luciferase expression. A control construct consisted of six binding sites to the CXCR4 siRNA, which should not be down-regulated. The plasmid constructs were linearized via restriction digestion, transfected into cells, and selected on G418. Approximately three weeks after the beginning of selection, individual cells were sorted via flow cytometry into 96-well plates, and individual colonies grown-up. Cells were then assayed for luciferase expression, and single clones chosen for further study.

A chemical genetics approach was used to conduct a primary screen on the *mir-21*-inhibited HeLa cells. The Annotated Chemical Library (ACL) was chosen for the first screen, as this library has been optimized for compounds with known biological activity and for commercial availability (Root et al. 2003). Chemicals were screened at 4 μg/mL, and cells assayed for luciferase expression 48 hours after application of the library. All manipulations were carried out with an automated robotics system. This screen yielded a list of 15 compounds that upregulated expression of the *mir-21*-repressed luciferase activity five fold or greater (Table 1). Like most chemical libraries, some compounds in the the ACL are arrayed redundantly (for example, the same chemical from different suppliers), and thus the apperance of azathioprine four times represents a particularly reliable hit.

The follow-up screen of hits obtained from this screen, as well as a second screen on a new library, are summarized in brief. It became clear that most if not all of the chemicals were

actually causing a transcriptional upregulation of the CMV promoter driving the expression of the luciferase reporter rather than modulating the miRNA pathway, as assayed by ribonuclease protection. Indeed, several of these compounds were quite toxic to the cells, and there are reports that cell stress, such as inhibition of translation or heat shock, activate the CMV promoter (Geelen et al. 1987).

The control CXCR4 cells should have also shown this response, and the question is why they did not. In retrospect, the use of single-cell clones may have been a poor choice, as clonal variation likely explains the relative insensitivity of the CXCR4 cells, with the same CMV promoter, to these transcriptional activators. The linearized plasmids insert into the genome at random and at different copy numbers, and it is therefore possible that the *mir-21* cassette inserted in a location more susceptible to transcriptional upregulation than the CXCR4 cassette. The use of only one test cell line and one experimental cell line further exacerbated this problem.

To move forward, new constructs were made, with four experimental and three control 3' UTRs regulating the expression of *Renilla* luciferase (Figure 1a). These constructs were driven by the herpes thymidine kinase promoter, as this promoter shows less responsiveness to cellular perturbations. Lastly, Invitrogen's Flp-In system was chosen for creating the cell lines. In this system, cell lines are purchased that contain a single FRT site inserted into the genome, and the gene of interest is cloned into a plasmid that also has a single FRT site. Co-transfection of this plasmid with a plasmid expressing FLP recombinase results in insertion of the gene of interest at a defined, single locus in the cell. This recombination also confers hygromycin resistance, and successfully recombined cells can thus be selected. All the cells that grow out are isogenic, and single-cell sorting of colonies is unnecessary.

Before creating the stable cell lines, the plasmid constructs were tested for activity in a transient transfection assay. Together with a firefly luciferase transfection control, the constructs were transfected into HeLa and 293 cells, and luciferase activity assayed at 24 hours (Figure 1b). The constructs with 3' UTRs that serve as binding sites for endogenous miRNAs were well-repressed in both cell lines, with *mir-20* showing the greatest activity. *mir-21* is known to be expressed in HeLa cells but is not detectable via Northern analysis in 293 cells (Zeng and Cullen 2003); *mir-16, -18,* and *-20* are known to be expressed in both HeLa and 293 cells (Lagos-Quintana et al. 2001; Nelson et al. 2004; Thomson et al. 2004). As expected, the *mir-21* UTR conferred repression only in HeLa cells, while mutated *mir-21* binding sites did not confer repression in HeLa cells or 293 cells. An additional validation utilized a 2'-O-Me oligonucleotide to inhibit *mir-21.* HeLa cells were transfected with constructs with no sites, *mir-21* sites, or mutated *mir-21* sites (Figure 1c). As expected, only the construct with *mir-21* sites showed a repression of luciferase expression. Co-transfection of an anti-*mir-21* 2'-O-Me oligonucleotide restored expression to the construct with *mir-21* sites, but did not alter the expression of the other two constructs.

These constructs were then transfected into Invitrogen's 293 Flp-In cell line and stable integrants selected. Four of these cell lines were chosen for further study, and the luciferase expression of these cells was assayed (Figure 1d). The lines with *mir-18* and *mir-20* binding sites showed decreased expression levels, approximately 10 fold and 130 fold, respectively, relative to the cell line with no binding sites.

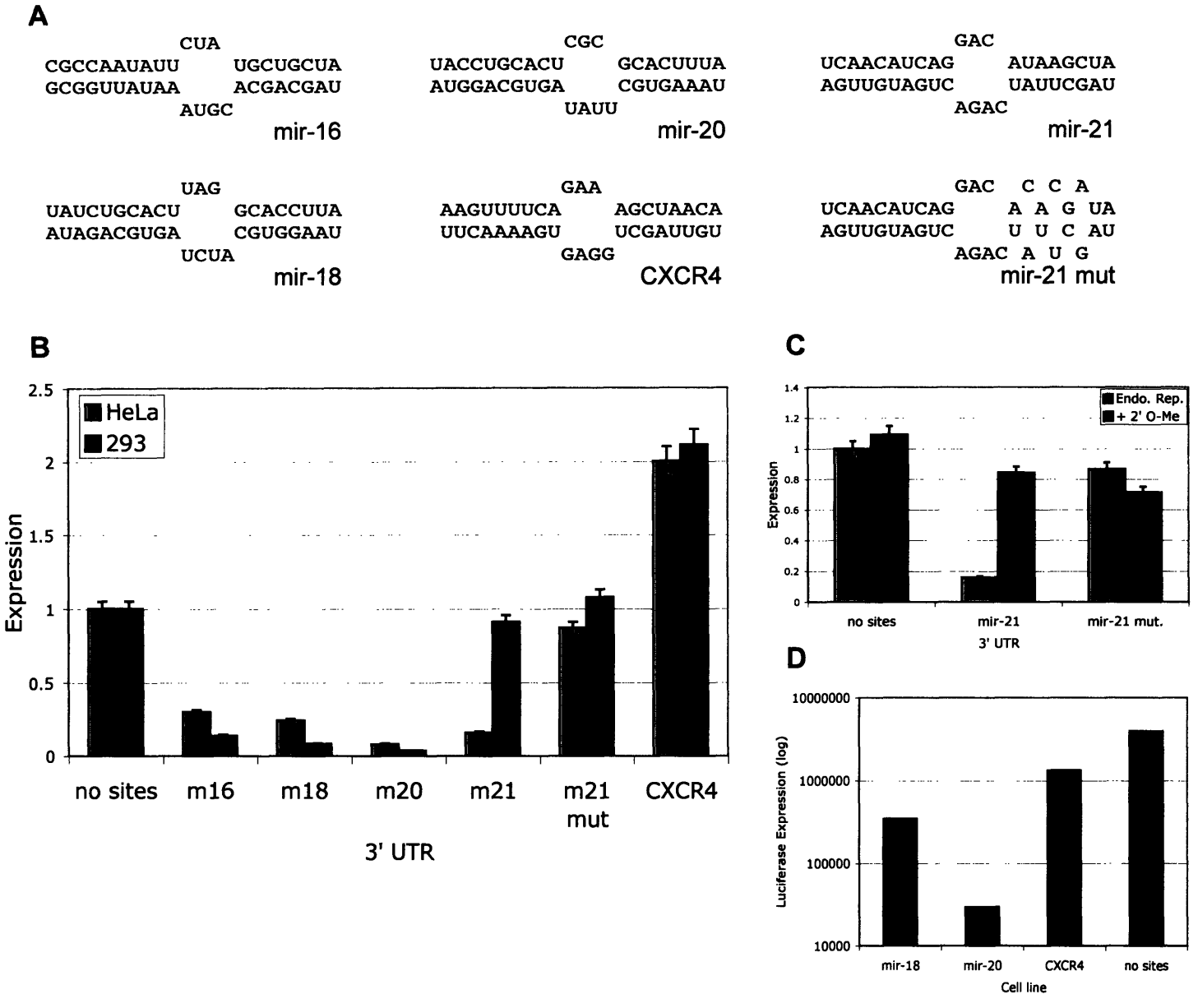

Figure 1. Characterization of miRNA reporter constructs.

A. Schematic of binding sites designed to bind to endogenous miRNAs. The mRNA binding site is shown 5' to 3' on the top, with the miRNA on the bottom. The four test UTRs have binding sites to miR-16, -18, -20, and -21, while three control UTRs have binding sites to the CXCR4 siRNA, mutated miR-21 binding sites, or no binding sites.

B. Transient transfection assay of these construct in HeLa and 293 cells. The test *Renilla* luciferase was normalized to the control firefly luciferase, and each value compared to the construct with no sites.

C. A 2'-O-Me inhibitor oligonucleotide with complementarity to endogenous *mir-21* was transfected into HeLa cells with the indicated constructs.

D. The reporter constructs were transfected into 293 cells, stable transformants selected, and luciferase activity determined. Lysates were also assayed for total protein content to assure that the luciferase values were comparable. Note that gene expression is plotted on a logarithmic scale.

Future Direction

The system described here, in which an endogneous miRNA is used to repress a reporter gene, allows for the screening of chemicals and genes that down-regulate the translational repression pathway. The initial design of the this system was not conducive to obtaining biologically relevant results, but several of the earlier problems have been identified and eliminated. Future screening experiments, whether using a chemical library or a library of siRNAs, should allow for identification of compounds and genes that modulate miRNA activity.

Fold Upregulation	Chemical
50	mercaptopurine
46	azathioprine
38	azathioprine
35	azathioprine
27	azathioprine
26	1,10-Phenanthroline
14	8-Bromo-cAMP
10	1-(2,3-Epoxy-b-D-Lyxofuranosyl)-Uracil
7	5-Thio-D-Glucose
6	8-Bromoadenosine 5'-diphosphate
6	Tyrphostin AG 1288
6	Zopiclone
6	Carbinoxamine Maleate Salt
5	Glipizide
5	Crystal Violet

Table 1. ACL hits for upregulation of *mir-21* repressed luciferase activity.

References

Antar, L.N. and G.J. Bassell. 2003. Sunrise at the synapse: the FMRP mRNP shaping the synaptic interface. *Neuron* **37**: 555-8.

Brennecke, J., D.R. Hipfner, A. Stark, R.B. Russell, and S.M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**: 25-36.

Ceman, S., W.T. O'Donnell, M. Reed, S. Patton, J. Pohl, and S.T. Warren. 2003. Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum Mol Genet* **12**: 3295-305.

Geelen, J.L., R. Boom, G.P. Klaver, R.P. Minnaar, M.C. Feltkamp, F.J. van Milligen, C.J. Sol, and J. van der Noordaa. 1987. Transcriptional activation of the major immediate early transcription unit of human cytomegalovirus by heat-shock, arsenite and protein synthesis inhibitors. *J Gen Virol* **68 (Pt 11)**: 2925-31.

Holcik, M. and N. Sonenberg. 2005. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol* **6**: 318-27.

Hutvagner, G., J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834-8.

Jin, P., R.S. Alisch, and S.T. Warren. 2004. RNA and microRNAs in fragile X mental retardation. *Nat Cell Biol* **6**: 1048-53.

Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-8.

Lagos-Quintana, M., R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel, and T. Tuschl. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**: 735-9.

Lau, N.C., L.P. Lim, E.G. Weinstein, and D.P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858-62.

Lee, R.C. and V. Ambros. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**: 862-4.

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-54.

Nelson, P.T., D.A. Baldwin, L.M. Scearce, J.C. Oberholtzer, J.W. Tobias, and Z. Mourelatos. 2004. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* **1**: 155-161.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-6.

Root, D.E., S.P. Flaherty, B.P. Kelley, and B.R. Stockwell. 2003. Biological mechanism profiling using an annotated compound library. *Chem Biol* **10**: 881-92.

Stefani, G., C.E. Fraser, J.C. Darnell, and R.B. Darnell. 2004. Fragile X mental retardation protein is associated with translating polyribosomes in neuronal cells. *J Neurosci* **24**: 7272-6.

Thomson, J.M., J. Parker, C.M. Perou, and S.M. Hammond. 2004. A custom microarray platform for analysis of microRNA gene expression. *Nat Methods* **1**: 47-53.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855-62.

Zeng, Y. and B.R. Cullen. 2003. Sequence requirements for micro RNA processing and function in human cells. *Rna* **9**: 112-23.

Conclusions

We have established a mammalian tissue culture system that has allowed us to ask fundamental questions about the specificity and mechanism of miRNAs, and the results presented herein give insight into the role of miRNAs in the regulation of gene expression. Many exciting questions await answers, and the importance of these small RNAs will continue to grow as we gain a better understanding of their biology.

The results presented in Chapter Two showed for the first time that siRNAs can function as miRNAs (Doench et al. 2003), a result later confirmed by other researchers (Saxena et al. 2003; Zeng et al. 2003). Perhaps the best outside support for the usefulness and relevance of this system has come from the Hannon lab. Using our constructs and MEFs knocked out for Argonaute 2, they found that RNAi cleavage activity was totally lost, but these cells still retained the ability to down-regulate luciferase expression via the translational repression pathway (Liu et al. 2004). Clearly, the RNAi pathway and the miRNA pathway are separable in mammalian cells. Our experiments also showed that miRNAs interact with their target mRNA in a synergistic fashion, a result not seen for the RNAi pathway. This observation may reflect mechanistic differences between the pathways, as mRNA cleavage is known to be catalytic, while miRNAs stably associate with their target mRNAs (C. Petersen, personal communication).

This CXCR4 system was further used, as described in Chapter Three, to examine the specificity of miRNAs (Doench and Sharp 2004). It had been assumed, mostly based on observations of *lin-4* and *let-7* targets in worms (Wightman et al. 1993; Reinhart et al. 2000) and 3' UTR motifs in flies that turned out to be likely miRNA targets in flies (Lai 2002), that the 5' region of the miRNA conferred most of the specificty, although this had never been tested experimentally. It was also reported that the exact nature of the miRNA:mRNA interaction was important for activity, as a bulged cytosine appeared to be important for *lin-4* regulation of *lin-14*

(Ha et al. 1996). Given the importance of determining, in a systematic fashion, how miRNAs find their targets, we turned to our luciferase-based cell culture system. We found that the 5' region of the miRNA was indeed necessary and sufficient for activity, and that the degree of repression correlated with the ΔG value of the miRNA:mRNA in that region. We also found that G:U wobble appeared to be specifically selected against, despite its favorable contributions to thermodynamic stability. We believe that the ΔG value simply serves as a surrogate reporter for the nature of the dsRNA helix in the 5' region, and that the Argonaute protein is actually selecting for an A-form helix. When that A-form helix is disturbed by, for example, a mismatch, the computed ΔG value reflects that disturbance; when a G:U wobble is introduced, however, the A-form helix is still disturbed, but the ΔG value is not. The recent crystal structure of the PIWI domain bound to dsRNA supports this conclusion (Ma et al. 2005; Parker et al. 2005). Satifyingly, the main conclusions of this work have since been shown to hold true in both flies (Brennecke et al. 2005) and zebrafish (Kloosterman et al. 2004).

These studies also indicated a potential relevance for the 3' region of the miRNA. When the binding of the 5' region of the miRNA was compromised, complementarity in the 3' region was needed for activity. Indeed, support for the relevance of this finding *in vivo* has come from *C. elegans*. The *lin-41* mRNA is expressed early in worm development, and is turned off by *let-7* during the L4 to adult transition. Other *let-7* family members, which have the same sequence in the 5' region, are expressed earlier than L4, and thus would be expected to regulate *lin-41* if 5' region complementarity were the only requirement for activity. The *let-7:lin-41* interaction is not a perfect match, however, and it appears that only *let-7,* and not the other family members, have sufficient complementarity in the 3' region to assist in target selection (Dave, Victor... is this accurate? Who/what should I cite?).

The potential importance of the 3' region raises questions regarding the current computational predictions of miRNA targets. Indeed, because of the mismatch in the 5' seed region, *let-7:lin-41* would not be predicted by current algorithms that have been used on the human genome (Lewis et al. 2005). Furthermore, some miRNAs are exquisitely conserved across their entire length (e.g. *mir-1)*, and thus it is likely that the 3' region of these miRNAs have an important function. An unbiased, experimental approach to uncover miRNA targets, such as that presented in Appendix A, could shed light on this issue.

In addition to questions of miRNA targets, a largely-unexplored field of miRNA biology is the mechanism of translational repression. This cell culture system has been used to begin to uncover details of the mechanism (C. Petersen, personal communication), but many questions still remain. If the results seen in worms hold true in other systems, namely that translation is repressed at some step after initiation (Olsen and Ambros 1999; Seggerson et al. 2002), then this is an unusual form of translational regulation. Perhaps translation itself is not affected, but rather the nascent protein is rapidly degraded. Another theory is that the mRNA is mislocalized, perhaps to a subcellular structure akin to stress granules, which are known to contain mRNAs and translation initiation factors but do not support productive translation (Kedersha and Anderson 2002).

Relatedly, while the activity of Argonuate 2 is established, the other three Argonautes in the human genome are not well understood, and it is likely that they do not act redundantly, as their expression levels vary from cell to cell (Sasaki et al. 2003; Meister et al. 2004). Of note, the fifth exon of Argonaute 1, as well as significant stretches of the flanking introns, constitute an 'ultra-conserved' element in the human genome, a stretch of at least 200 nucleotides that is absolutely invariant in mice, rats, and humans, and there is EST evidence that this exon is

alternatively spliced (Bejerano et al. 2004). Although the significance of this observation is unclear, this constitutes a tantalizing area of futher investigation. The discovery of chemicals and genes that modulate the miRNA pathway, such as through an approach outlined in Appendix B, could provide important insights into these questions.

Research into RNAi and miRNAs has given yet another example of the importance of RNA in modern biological systems. Additionally, this pathway has proven itself to be an important new technology as researchers try to untangle and understand the vast complexity that a sequenced genome presents.

References

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W.J. Kent, J.S. Mattick, and D. Haussler. 2004. Ultraconserved elements in the human genome. *Science* **304**: 1321-5.

Brennecke, J., A. Stark, R.B. Russell, and S.M. Cohen. 2005. Principles of microRNA-target recognition. *PLoS Biol* **3**: e85.

Doench, J.G., C.P. Petersen, and P.A. Sharp. 2003. siRNAs can function as miRNAs. *Genes Dev* **17**: 438-42.

Doench, J.G. and P.A. Sharp. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**: 504-11.

Ha, I., B. Wightman, and G. Ruvkun. 1996. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. *Genes Dev* **10**: 3041-50.

Kedersha, N. and P. Anderson. 2002. Stress granules: sites of mRNA triage that regulate mRNA stability and translatability. *Biochem Soc Trans* **30**: 963-9.

Kloosterman, W.P., E. Wienholds, R.F. Ketting, and R.H. Plasterk. 2004. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* **32**: 6284-91.

Lai, E.C. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**: 363-4.

Lewis, B.P., C.B. Burge, and D.P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.

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

Ma, J.B., Y.R. Yuan, G. Meister, Y. Pei, T. Tuschl, and D.J. Patel. 2005. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**: 666-70.

Meister, G., M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng, and T. Tuschl. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**: 185-97.

Olsen, P.H. and V. Ambros. 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**: 671-80.

Parker, J.S., S.M. Roe, and D. Barford. 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**: 663-6.

Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-6.

Sasaki, T., A. Shiohama, S. Minoshima, and N. Shimizu. 2003. Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics* **82**: 323-30.

Saxena, S., Z.O. Jonsson, and A. Dutta. 2003. Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem* **278**: 44312-9.

Seggerson, K., L. Tang, and E.G. Moss. 2002. Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Dev Biol* **243**: 215-25.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855-62.

Zeng, Y., R. Yi, and B.R. Cullen. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* **100**: 9779-84.

John Gerard Doench

Biographical Note

EDUCATION

- **Massachusetts Institute of Technology**, Cambridge, MA. 2000 – 2005
  Ph.D., Department of Biology.

- **Hamilton College**, Clinton, NY. 1996 – 2000
  B.A., Biochemistry; B.A. History.

RESEARCH EXPERIENCE

- **Graduate Research**, Massachusetts Institute of Technology. 2000 – 2005
  Advisor: Phillip A. Sharp, Ph.D. Department of Biology, Center for Cancer Research.
  Doctoral Thesis: Specificity and mechanism of microRNAs in the regulation of gene expression.

- **Undergraduate Research**, Hamilton College. 1999 – 2000
  Advisor: Timothy E. Elgren, Ph.D. Department of Chemistry.
  Summer research and senior thesis: Purification of E. coli Ribonucleotide Reductase and characterization of the pH dependence of peroxidase activity.

- **Undergraduate Research,** Hamilton College. 1999 – 2000
  Advisor: Douglas Ambrose, Ph.D. Department of History.
  Senior Thesis: Puritan sermons from 1700 – 1730.

- **Summer Research Fellowship**, Hamilton College. 1998
  Advisor: Dr. Kevin P. Grant, PhD. Department of History.
  John Locke and America.

TEACHING EXPERIENCE

- **Head Teaching Assistant**, General Biochemistry (7.05), MIT. Spring 2004

- **Teaching Assistant**, Introductory Biology (7.01), MIT. Fall 2001

- **Teaching Assistant**, General Chemistry, Hamilton College. 1998 – 2000

- **Writing Tutor**, Nesbitt–Johnston Writing Center, Hamilton College. 1997 – 2000

SCIENTIFIC PUBLICATIONS

- **Doench, J.G.** and Novina, C.D. "RNA interference" in *Encyclopedia of Molecular Cell Biology and Molecular Medicine.* Wiley, New York. *In press.*

- Petersen, C.P., **Doench, J.G.**, Grishok, A., and Sharp, P.A. "The Biology of Short RNAs" in *The RNA World, 3rd Edition.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. *In press.*

- **Doench, J. G.** and Sharp, P. A. (2004) Specificity of miRNA target selection in translational repression. Genes Dev *18,* 504–11.

- **Doench, J. G.**, Petersen, C. P., and Sharp, P. A. (2003) siRNAs can function as miRNAs. Genes Dev *17,* 438–42.

PRESENTATIONS

Oral Presentations

- **John G. Doench** and Phillip A. Sharp. *Specificity of microRNAs in translational repression.* Center for Cancer Research Fall Retreat, Waterville Valley, NH, November 2004.

- **John G. Doench**. *Killing the Messenger: The Discovery of RNA interference.* Symposium on Biotechnology Education, Museum of Science, Boston, MA, March 2004.

- **John G. Doench** and Phillip A. Sharp. *RNA interference and microRNAs: discoveries and advances.* Hobart and William Smith College, Geneva, NY, March 2004.

Poster Presentations

- **John G. Doench** and Phillip A. Sharp. *Specificity of microRNAs in translational repression.* Keystone symposium: siRNAs and miRNAs. Keystone, CO, April 2004.

- **John G. Doench**, Christian P. Petersen, and Phillip A. Sharp. *siRNAs can function as miRNAs.* Gordon Conference: Nucleic Acids. Newport, RI, June 2002.

HONORS AND AWARDS

- Teresa Keng Graduate Teaching Prize for excellence in teaching, 2004
- Howard Hughes Medical Institute Predoctoral Fellowship, 2000 – 2005
- Honorable Mention, NSF Graduate Fellowship Competition, 2000
- Salutatorian, Hamilton College, Class of 2000
- Departmental Honors in Biochemistry and History, Hamilton College, 2000
- Phi Beta Kappa, Hamilton College, 2000
- Dean's List, Hamilton College, 1996 – 2000
- Edgar Baldwin Graves Prize in History, Hamilton College, 2000
- Underwood Prize in Chemistry, Hamilton College, 2000
- Hans H. Schambach Scholarship, Hamilton College, 1996 – 2000