

University of Massachusetts Medical School

eScholarship@UMMS

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

2012-07-09

Understanding Small RNA Formation in *Drosophila Melanogaster*: A Dissertation

Elif Sarinay Cenik

University of Massachusetts Medical School

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss



Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Animal Experimentation and Research Commons](#), [Biochemistry, Biophysics, and Structural Biology Commons](#), [Enzymes and Coenzymes Commons](#), and the [Nucleic Acids, Nucleotides, and Nucleosides Commons](#)

Repository Citation

Cenik ES. (2012). Understanding Small RNA Formation in *Drosophila Melanogaster*: A Dissertation. GSBS Dissertations and Theses. <https://doi.org/10.13028/bw3j-x330>. Retrieved from https://escholarship.umassmed.edu/gsbs_diss/615

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

**UNDERSTANDING SMALL RNA FORMATION IN
*DROSOPHILA MELANOGASTER***

A Dissertation Presented

By

Elif Sarinay Cenic

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

July 9th, 2012

Biochemistry and Molecular Pharmacology

UNDERSTANDING SMALL RNA FORMATION IN *DROSOPHILA*
MELANOGASTER

A Dissertation Presented By
ELIF SARINAY CENIK

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

Phillip D. Zamore, Thesis Advisor

Sean Ryder, Member of Committee

William Theurkauf, Member of Committee

Traci M.T. Hall, Member of Committee

Andrei Korostelev, Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee.

Melissa J. Moore, Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

Anthony Carruthers, Ph.D., Dean of the Graduate School of Biomedical
Sciences

Biochemistry and Molecular Pharmacology
July 9th 2012

ACKNOWLEDGEMENTS

I would like to thank all the people who supported me during my doctoral study.

I would like to thank my advisor, Phillip Zamore, for giving me the opportunity to be a part of the exciting research going on in Zamore lab, and for teaching me the essentials of high quality science. I always admired his never-ending passion and rigorous attention to details. He thought me the importance of quantification, being critical, designing the right experimental controls, and thinking about the essential problems in the small RNA field. With his help, I learned a great deal about scientific writing, communicating with figures and how to present my science in a simple, accurate and exciting way.

I would like to thank our former and present laboratory members for creating a friendly and exciting atmosphere to work and for always being available for experimental help, scientific discussions and social outings.

I would like to thank my thesis committee, especially to my thesis chair Melissa Moore for her great ideas that pushed my project further and her insightful advices for my career. I would like to thank Sean Ryder for his valuable input and Traci Tanaka Hall for being a supportive collaborator for our Dicer-2 manuscript.

I would like to thank my husband Can Cenik for his wisdom and being always there for me during my good and bad times; and my family back in Turkey who allowed me pursue my dream to be a scientist.

ABSTRACT

Drosophila Dicer-2 generates small interfering RNAs (siRNAs) from long double-stranded RNA (dsRNA), whereas Dicer-1 produces microRNAs from pre-microRNA. My thesis focuses on the functional characteristics of two *Drosophila* Dicers that makes them specific for their biological substrates. We found that RNA binding protein partners of Dicers and two small molecules, ATP and phosphate are key in regulating *Drosophila* Dicers' specificity. Without any additional factor, recombinant Dicer-2 cleaves pre-miRNA, but its product is shorter than the authentic miRNA. However, the protein R2D2 and inorganic phosphate block pre-miRNA processing by Dicer-2. In contrast, Dicer-1 is inherently capable of processing the substrates of Dicer, long dsRNAs. Yet, partner protein of Dicer-1, Loqs-PB and ATP increase the efficiency of miRNA production from pre-miRNAs by Dicer-1, therefore enhance substrate specificity of Dicer-1. Our data highlight the role of ATP and regulatory dsRNA-binding partner proteins to achieve substrate specificity in *Drosophila* RNA silencing.

Our study also sheds light onto the function of the helicase domain in *Drosophila* Dicers. Although Dicer-1 doesn't hydrolyze ATP, ATP enhances miRNA production by increasing Dicer-1's substrate specificity through lowering its K_M . On the other hand, Dicer-2 is a dsRNA-stimulated ATPase that hydrolyzes ATP to ADP, and ATP hydrolysis is required for Dicer-2 to process long dsRNA. Wild-type Dicer-2, but not a mutant defective in ATP hydrolysis, is processive; generating siRNAs faster than it can dissociate from a long dsRNA

substrate. We propose that the Dicer-2 helicase domain uses ATP to generate many siRNAs from a single molecule of dsRNA before dissociating from its substrate.

Piwi-dependent small RNAs, namely piRNAs, are a third class of small RNAs that are distinct from miRNAs and siRNAs. Their primary function is to repress transposons in the animal germline. piRNAs are Dicer-independent, and require Piwi family proteins for their biogenesis and function. Recently in addition to their presence in animal germlines, the presence and function of piRNA-like RNAs in the somatic tissues have been suggested (Yan et al. 2011; Morazzani et al. 2012; Rajasethupathy et al. 2012). We have investigated whether the piRNA-like reads in our many *Drosophila* head libraries come from the germline as a contaminant or are soma-specific. Most of the piRNA reads in our published head libraries show high similarity to germline piRNAs. However, piRNA-like reads from manually dissected heads are distinct from germline piRNAs, proving the presence of somatic piRNA-like small RNAs. We are currently asking the question whether these distinct piRNA-like reads in the heads are dependent on the Piwi family proteins, like the germline piRNAs.

TABLE OF CONTENTS

| | |
|--|------------|
| Acknowledgements | iii |
| Abstract | iv |
| Table of Contents | vi |
| List of Tables | ix |
| List of Figures | xi |
| List of Copyrighted material | xiv |
| | |
| Chapter I: introduction | 15 |
| Discovery and Importance of Small RNAs | 15 |
| Small RNA Pathways in <i>Drosophila melanogaster</i> | 19 |
| The Role of Dicer in Small RNA Biogenesis | 22 |
| Piwi Dependent Small RNAs in <i>Drosophila melanogaster</i> | 34 |
| | |
| Chapter II: Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-driven ribonuclease. | 39 |
| Preface | 40 |
| Abstract | 41 |
| Introduction | 42 |
| Results | 46 |
| Discussion | 86 |
| Materials and Methods | 90 |

| | |
|---|------------|
| Acknowledgements | 94 |
| Chapter III: Processing of pre-miRNAs by <i>Drosophila</i> Dicer-1 | 95 |
| Preface | 96 |
| Abstract | 97 |
| Introduction | 98 |
| Results | 101 |
| Discussion | 113 |
| Materials and Methods | 115 |
| Acknowledgements | 118 |
| Chapter IV: Presence of piRNAs in <i>Drosophila</i> heads | 119 |
| Preface | 120 |
| Abstract | 121 |
| Introduction | 122 |
| Results | 126 |
| Discussion | 156 |
| Materials and Methods | 158 |
| Acknowledgement | 161 |
| Chapter V: Final Summary and Conclusions | 162 |
| Bibliography | 170 |

Appendices**191**

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., Zamore, P.D. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell.* 42(2), 172-84.

Cenik, E.S., Zamore, P.D. (2011). Argonaute proteins. *Curr. Biol.* 21(12), R446-9.

LIST OF TABLES

CHAPTER II: Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease.

Table 2.1. Michaelis-Menten Analysis of Dicer-2 and Partner Proteins. **56**

Table 2.2. Michaelis-Menten Analysis of siRNA Production and ATP Consumption by Dicer-2. **63**

Table 2.S1. List of Synthetic DNA Oligonucleotides. **80**

Table 2.S2. List of Synthetic RNA Oligonucleotides. **83**

Table 2.S3. DNA Templates for Transcription and DNA Splints for RNA Ligation, Used in Figures 2.3 and 2.4. **84**

Table 2.S4. List of Synthetic RNA Oligonucleotides Assembled by Splinted Ligation into Site-Specifically Labeled Sense-Strand RNA and Used in Figure 2.4, 2.5 and S4. **85**

CHAPTER III: Processing of pre-miRNAs by *Drosophila* Dicer-1

Table 3.1. Michaelis-Menten Analysis of Dicer-1. **112**

CHAPTER IV: Presence of piRNAs in *Drosophila melanogaster* heads.

Table 4.1. Pearson and Spearman Correlations between Transposon Mapping piRNA Reads from Several Published Datasets. **132**

Table 4.2. Correlations between Transposon Mapping piRNA Reads from Several Published Datasets. **137**

| | |
|--|------------|
| Table 4.3. Gene mapping piRNA are Concentrated in a Small Subset of Coding Transcripts. | 145 |
| Table 4.4. Gene Ontology (GO) Enrichment of Transcripts with the Most Mapped piRNAs in dcr-2 Heads. | 147 |
| Table 4.5. Many transcripts with 3'UTR Mapping piRNAs have Neurological Function. | 148 |
| Table 4.S1. Summary of Manually Dissected piwi Mutant, ago2 Heterozygous Male Head Small RNA Library Reads. | 155 |

LIST OF FIGURES

CHAPTER I: Introduction

- Figure 1.1.** The siRNA and the miRNA Pathway in *Drosophila melanogaster*. 17
- Figure 1.2.** Molecular Mechanism of dsRNA Cleavage by *Giardia intestinalis* Dicer. 24
- Figure 1.3.** Domain Structure of *Drosophila* Dicers and their RNA Binding Protein Partners. 25
- Figure 1.4.** The Transcript Variants and Domain Structures of Loquacious (Loqs) Isoforms. 31
- Figure 2.5.** Structures of two pre-miRNAs; pre-*let-7* and pre-miR-1. 33

CHAPTER II: Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease.

- Figure 2.1.** Dicer-1 and Dicer-2 Produce Different Products from Pre-*let-7*. 47
- Figure 2.2.** R2D2 and Phosphate Inhibit Dicer-2 Processing of Short Substrates. 49
- Figure 2.S1.** R2D2 Inhibits Dicer-2 Processing of a 25 bp dsRNA Substrate and Creatine Phosphate Inhibits Dicer-2 Processing of a Long dsRNA Bearing 5 Triphosphate termini. 51
- Figure 2.S2.** Michaelis-Menten Analysis of Dicer-2, Dicer-2 + Loqs-PD, and Dicer-2/R2D2 Using a 515 bp dsRNA Substrate. 54
- Figure 2.3.** Dicer-2 Requires ATP Hydrolysis to Process Long dsRNA. 60

| | |
|---|------------|
| Figure 2.S3. ATP Hydrolysis is Required for Efficient Long dsRNA Processing by Dicer-2. | 65 |
| Figure 2.4. Evidence that Dicer-2 is Processive. | 71 |
| Figure 2.S4. The Rate of siRNA Production for Different dsRNA Termini: Two 5 Deoxynucleotides Block Entry of Dicer-2, but Blunt and 3' Overhanging Ends are Processed Similarly. | 73 |
| Figure 2.5. Dicer-2 Remains Associated with its Substrate in the Presence of ATP. | 77 |
| Figure 2.6. A Model for <i>Drosophila</i> Dicer-2. | 79 |
| CHAPTER III: Processing of pre-miRNAs by <i>Drosophila</i> Dicer-1 | |
| Figure 3.S1. Alignment of the Predicted ATP-Binding Site of Dicer-1 with Known or Putative ATPase Proteins. | 102 |
| Figure 3.1. ATP and Loqs-PB Increase Multiple-Turnover Pre-miRNA Processing by Dicer-1. | 104 |
| Figure 3.S2. Non-hydrolyzable β - γ ATP Analogs do not Replace ATP for Pre-miRNA Processing by Dicer-1. | 107 |
| Figure 3.2. Dicer-1, but not Dicer-2, Generates Intermediates when Processing Long dsRNA. | 109 |
| CHAPTER IV: Presence of piRNAs in <i>Drosophila Melanogaster</i> heads. | |
| Figure 4.1. Summary of Sequencing Reads. | 127 |

| | |
|---|------------|
| Figure 4.2. Transposon Mapping piRNAs Suggest a High Probability of Contamination. | 130 |
| Figure 4.3. piRNAs are More Concentrated in the 3'UTR Region of Coding Transcripts. | 135 |
| Figure 4.4. Gene Mapping piRNA Abundance and Gene Expression Correlate Weakly. | 138 |
| Figure 4.5. Representative coding genes with mapping piRNAs. | 141 |
| Figure 4.6. piRNAs Isolated from Manually Dissected Male Heads do not Originate from Germline Contamination. | 151 |
| Figure 4.7. Coding Region Mapping piRNAs are Found in Manually Dissected Head Small RNA Library. | 153 |

LIST OF COPYRIGHTED MATERIAL

The chapters of the dissertation have appeared in whole or part in publications below:

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., Zamore, P.D. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell.* 42(2), 172-84.

Cenik, E.S., Zamore, P.D. (2011). Argonaute proteins. *Curr .Biol.* 21(12), R446-9.

CHAPTER I: INTRODUCTION

Discovery and Importance of Small RNAs

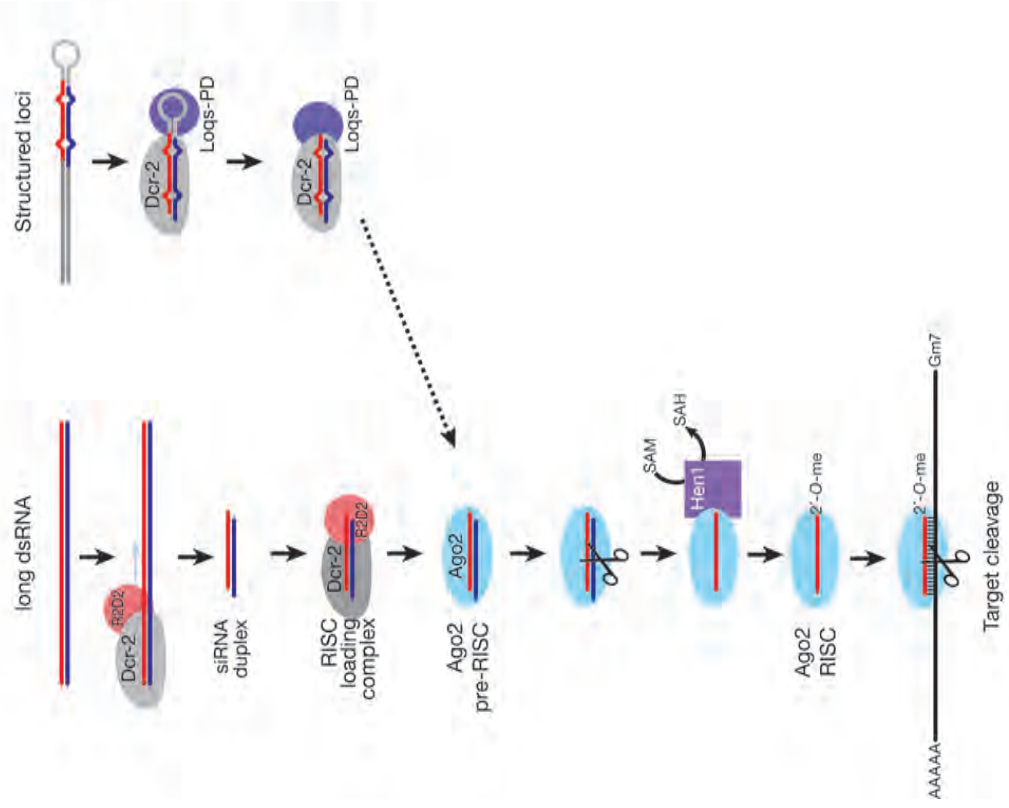
Small RNAs are small regulatory molecules comprised of ribonucleotides. They are generally smaller than 30 nucleotides and do not encode for a protein. They function as RNA molecules to regulate gene expression in a sequence specific way at both the transcriptional and posttranscriptional levels (Baulcombe et al. 1996). Small interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-dependent small RNAs (piRNAs) are perhaps the three most well known classes of small RNAs.

Victor Ambros and his colleagues discovered small RNAs as gene expression regulators for the first time in 1993. The *lin-4* gene surprisingly did not encode for a protein but rather 22 (mature miRNA) -60 nucleotide (pre-miRNA) transcripts that are complementary to the 3'UTR of *lin-14* mRNA (Lee et al. 1993). This seminal work established this antisense relationship as a negative regulation mechanism controlling *lin-14* protein levels. What was a curious observation in worms later turned out to be a genome wide gene regulation mechanism that is conserved from plants to humans (Hamilton et al. 1999; Grishok et al. 2001, Brennecke; Hipfner et al. 2003). In 1998, Andrew Fire, Craig Mello and their colleagues identified, double stranded RNA as a potent genetic interference molecule (Fire et al. 1998). Double stranded or siRNAs are used as an important tool to modulate gene expression in laboratory studies and are being tested as a cure for incurable diseases such as Huntington's disease

(Montgomery et al. 1998; Kennerdell et al. 1998; Billy et al. 2001; Boutla et al. 2001; Chiu et al. 2002; Elbashir et al. 2001).

Since these two important discoveries, our knowledge about small RNA function and biogenesis has greatly expanded. Small RNAs have been implicated in diverse functions ranging from regulating development to repressing transposon activity (Hamilton et al. 1999; Reinhart et al. 2000; Lai et al. 2002, Aravin et al. 2001; Aukerman et al. 2003; Brennecke; Hipfner et al. 2003; Catalanotto et al. 2002; Chen et al. 2004; Grishok et al. 2001; Stark et al. 2003; Johnston et al. 2003; Kanellopoulou et al. 2005; Lau et al. 2001; Mallory et al. 2004; Noma et al. 2004; Vagin et al. 2006).

siRNA pathway



miRNA pathway

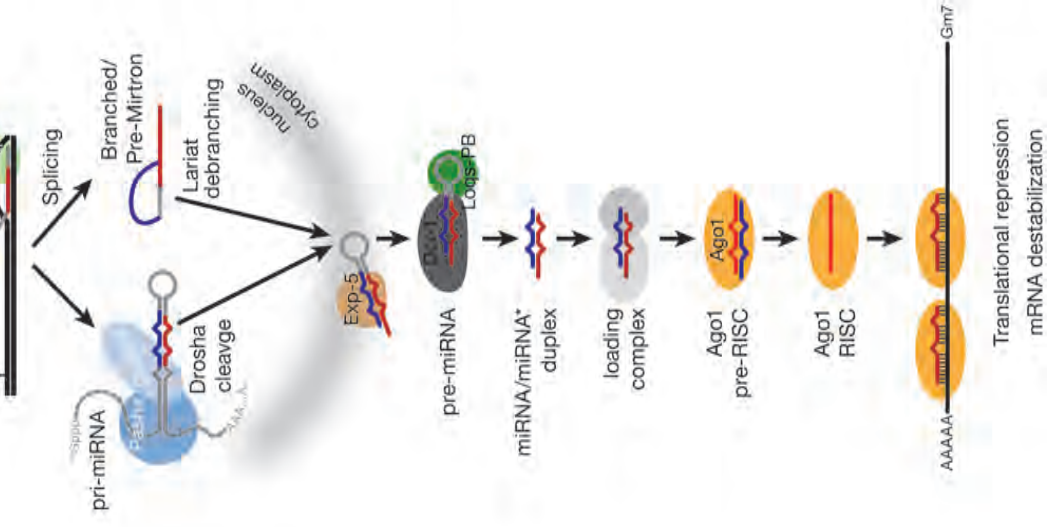


Figure 1.1. The siRNA and the miRNA Pathway in *Drosophila melanogaster* (adapted from Ghildiyal & Zamore, 2009). The siRNA pathway originates from long double stranded RNAs whereas the miRNA pathway starts with very long endogenous precursor transcripts called primary miRNAs (pri-miRNAs). The siRNA precursors are cleaved by Dicer-2 to form siRNA duplexes that are subsequently loaded into an Argonaute-2 (Ago2) RNA induced silencing complex (RISC). After cleavage of the passenger strand, single stranded and methylated mature siRNA bound Argonaute-2 cleaves target mRNAs. Drosha/Pasha complex processes pri-miRNAs to ~60 nt short hairpins called pre-miRNAs in the nucleus. After being exported to cytoplasm, Dicer1/ Loqs PB complex cleaves the pre-miRNAs into miRNA-miRNA* duplexes. miRNA-miRNA* duplex is loaded into RISC comprised of Argonaute-1 (Ago1). miRNA bound RISC functions in mRNA destabilization or translational repression of the antisense target mRNA

Small RNA Pathways in *Drosophila melanogaster*

siRNAs and miRNAs are functionally distinct and their biogeneses require different protein complexes in *Drosophila melanogaster* (Lee et al. 2004, Figure 1.1). miRNAs are mainly responsible for repressing gene expression through mRNA destabilization and translational repression (Lewis et al. 2003; Bartel et al. 2004; Selbach et al. 2008; Baek et al. 2008; Brodersen et al. 2008). miRNAs are essential in many core processes such as development and stem cell maintenance (Bohmert et al., 1998; Grishok et al. 2001; Knight and Bass 2001; Wienholds et al., 2003; Giraldez et al. 2005). On the other hand, the main function of siRNAs is to protect *Drosophila* genome from invasion by somatic transposons and defend against viruses (Galiana-Arnoux et al., 2006; Chung et al., 2008; Deddouche et al., 2008; Flynt et al., 2009; Saleh et al., 2009; van Rij et al., 2009).

The siRNA pathway starts with long double stranded RNAs that might be originated from viral replication intermediates (Flynt et al., 2009), bidirectionally transcribed overlapping transcripts (Okamura et al., 2008a; Watanabe et al., 2008); hybridized transcripts to antisense pseudogene derived transcripts (Tam et al., 2008), and long hairpins formed by inverted transcripts (Yang and Kazazian, 2006; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008b). These precursors are cleaved by Dicer-2 to form siRNA duplexes that are subsequently loaded into silencing complexes, namely Argonaute-2 (Flynt et al., 2009; Tomari et al. 2004; Matranga et al. 2005; Pham et al. 2005;

Miyoshi et al. 2005; Rand et al. 2004; Rand et al. 2005; Leuschner et al. 2006).

After cleavage of the passenger strand, single stranded and methylated mature siRNA bound Argonaute-2 cleaves target mRNAs upon perfect base pairing to mature siRNA sequences (Zamore et al. 2000; Hammond et al. 2001; Schwarz et al. 2003; Schwarz et al. 2004; Liu, Carmell et al. 2004; Haley et al. 2004; Pham et al. 2005; Kim, Lee et al. 2006; Wang et al. 2009; Martinez et al. 2002; Martinez et al. 2004; Yang et al. 2006; Ameres et al. 2007; Horwich et al. 2007; Chung et al. 2008).

The miRNA pathway starts with very long endogenous precursor transcripts called primary miRNAs (pri-miRNAs) (Cai et al. 2004; Lee, Kim et al. 2004). Drosha/Pasha complex processes these pri-miRNAs to ~60 nt short hairpins called pre-miRNAs in the nucleus (Papp et al. 2003; Lee et al. 2002; Lee et al., 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2006). Pre-miRNAs are then exported to cytoplasm in an Exportin-5 dependent manner (Yi et al. 2003; Bohnsack 2004; Lund et al. 2004). Dicer1/Loqs PB complex cleaves the pre-miRNAs into miRNA-miRNA* duplexes in the cytoplasm (Forstemann et al. 2005). miRNA-miRNA* duplex is loaded into RNA induced silencing complex (RISC) comprised of Argonaute-1 (Ago1) (Forsteman et al. 2007; Tomari et al. 2007). The miRNA is unwound from miRNA* strand after binding Ago-1 (Kawamata et al. 2009; Okamura et al. 2009). Mature miRNA bound RISC is called mature RISC; it functions in mRNA destabilization or translational

repression of the antisense target mRNA (Lewis et al. 2003; Bartel et al. 2004; Selbach et al. 2008; Baek et al. 2008; Brodersen et al. 2008).

The Role of Dicer in Small RNA Biogenesis

Dicer was first identified and named for its cleavage of double stranded RNA in 2001 and discovered to be the key protein required for both miRNA and siRNA generation (Bass et al. 2000; Bernstein et al 2001; Hutvanger et al. 2001; Jiang et al. 2005).

In *Drosophila*, two Dicer proteins lie at the center of both siRNA and microRNA biogenesis pathways (Figure 1.1). Both Dicer 1 and Dicer 2 cut structurally very similar double stranded RNAs and release 21/22 nucleotide products (Zamore et al. 2000). Nonetheless, they are remarkably specific to their respective substrates. Previous genetic studies have revealed this specificity in vivo (Lee et al. 2004). Dicer-1 is required for microRNA maturation (Jiang et al., 2005;). When dicer-2 is absent, siRNA formation is abolished but miRNAs remain unaffected. Similarly dicer-1 mutation is lethal in flies, meaning that Dicer-2 cannot rescue miRNA processing.

The domain architectures of *Drosophila* Dicers are identical except for the N terminus. Both Dicers have two RNase III domains that are required for RNA cleavage (Bernstein et al. 2001; Bass et al. 2000). Crystal structure of Dicer from a unicellular eukaryote, *Giardia* Dicer, illuminated how the core domains of Dicer relate to its function. (Figure 1.2; (Macrae et al., 2006; Macrae et al., 2008; Jinek et al., 2009). The double stranded RNA binding domain and the PAZ domain are required for binding to RNA. PAZ domain binds to the 5' and 3' end of the double stranded RNA. The distance between the open helical end of the RNA to the

active sites is exactly 25 nucleotides in *Giardia* Dicer determining the product size. The catalytically active RNase III domains are 2 nucleotides apart from each other, therefore upon cleavage the product has a 2 nucleotide 3' overhang (Bass, 2000; Cerutti et al., 2000; Lingel et al., 2003; Song et al., 2003; Yan et al., 2003; Lingel et al., 2004; Ma et al., 2004; Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006; MacRae et al., 2007).

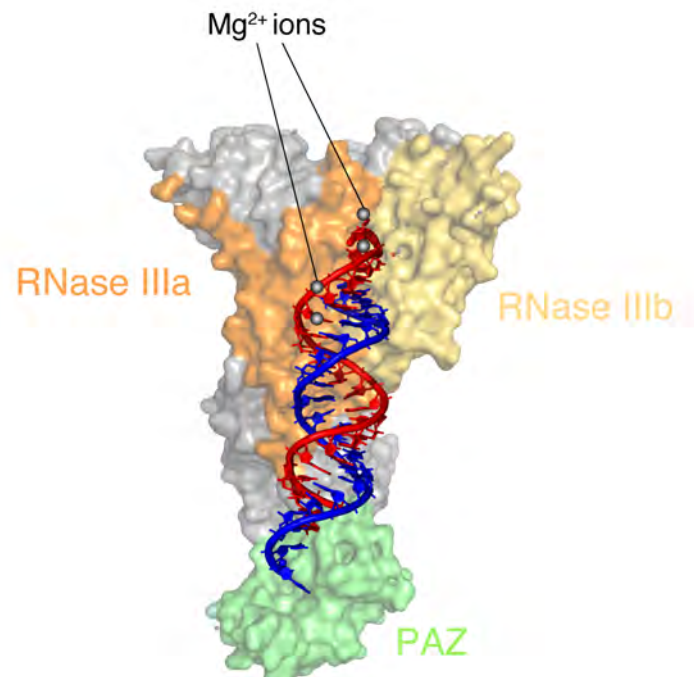
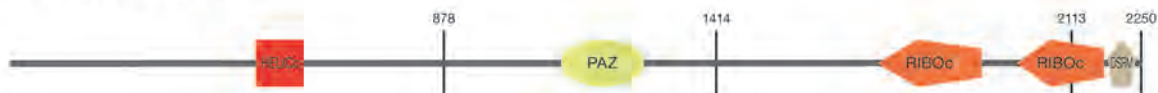


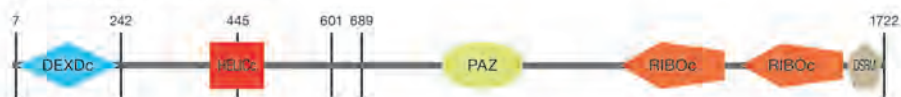
Figure 1.2. Molecular Mechanism of dsRNA Cleavage by *Giardia*

***intestinalis* Dicer.** The crystal structure of *Giardia intestinalis* Dicer (PDB ID: 2FFL, Macrae et al. 2006) with overlaid structure of an siRNA (PDB ID: 2F8S, Yuan et al. 2006).

Drosophila Dicer-1



Drosophila Dicer-2



Human Dicer



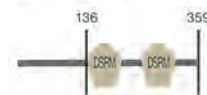
Giardia Dicer



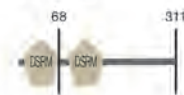
Loqs-PB



Loqs-PD



R2D2



-  DEXDc DEAD-like helicase superfamily
-  HELICc Helicase superfamily C-terminal domain
-  PAZ PAZ domain
-  RIBOc Ribonuclease III family
-  DSRM Double-stranded RNA binding motif

Figure 1.3. Domain Structure of *Drosophila* Dicers and their RNA

Binding Protein Partners. Domain structure of *Drosophila* Dicer-1, *Drosophila* Dicer-2, Human Dicer, Giardia Dicer, Loqs PB, LoqsPD, and R2D2 were adapted from the SMART protein domain analysis (Schultz et al., 1998; Ponting et al., 1999).

Function of Dicer is coupled to additional RNA binding partner proteins in metazoans (Figure 1.3). In *Drosophila melanogaster*, miRNA and siRNA pathways associate with specific Dicers, and specific RNA binding partners. These partner proteins are relatively small and have only double stranded RNA binding domains. The presence of the N-terminal DexDc domain and the differences in partner proteins are potential sources of specificity for these proteins (Jiang et al. 2005; Saito et al. 2005). Loquacious transcript is alternatively spliced to produce four variant transcripts (Figure 1.4). Loqs-PB, translated from one of these variants is in complex with Dicer-1, and functions in the miRNA pathway (Forstemann et al., 2005; Jiang et al., 2005). Loqs-PB enhances miRNA maturation, but not loading of mature miRNAs into RISC (Forstemann et al. 2005, Liu et al. 2007). On the other hand, R2D2, which is in complex with Dicer-2, is required to load siRNAs into Argonaute 2 containing RNA silencing complex (Liu et al., 2003; Tomari et al., 2004; Liu et al., 2006). Loqs-PD, an alternatively spliced isoform of Loquacious gene, also forms a complex with Dicer-2 and is involved in the endogenous siRNA pathway (Figure 1.1, 1.3, 1.4) (Zhou et al., 2009; Okamura et al., 2008b; Hartig et al., 2009; Zhou et al., 2009; Miyoshi et al., 2010; Hartig and Forstemann, 2011).

Dicer1/ Loqs PB complex cleaves the pre-miRNAs into miRNA-miRNA* duplexes in the cytoplasm. Pre-miRNAs are ~60 nt short hairpins that are exported after pri-miRNA processing (Cai et al. 2004; Lee, Kim et al. 2004) by Drosha/Pasha complex (Papp et al. 2003; Lee et al. 2002; Lee et al., 2003; Denli

et al., 2004; Gregory et al., 2004; Han et al., 2006). Pre-miRNAs are composed of a partially double stranded region called stem, a loop structure at the one end of the stem, and an open helical two nucleotide 3' overhang at the other end of the stem (Figure 1.5). Dicer-1/Loqs-PB complex cleaves the double stranded stem from the loop side, releasing a mature ~22 nucleotide miRNA/miRNA* duplex. Depending on where the mature miRNA strand comes from, pre-miRNAs can be divided into two classes. In the first class of pre-miRNAs, mature miRNA is derived from the 5' arm. Therefore, Dicer-1 cleavage occurs at the 3' end of the mature miRNA. In the second class of pre-miRNAs, mature miRNA strand is derived from the 3' arm of the hairpin. In this case, Dicer-1 cleaves to form 5' end of the mature miRNA.

Dicer-2 processes long dsRNA and long hairpins into ~21 nucleotide siRNAs. Dicer cleaves dsRNAs starting from the open double-stranded ends (Vermeulen et al., 2005; Macrae et al., 2006; Weinberg et al., 2011). Dicer-2 seems to be efficient in processing long double stranded RNA molecules, since never an intermediary product is observed. Its efficiency is furthermore increased in the presence of ATP (Zamore et al., 2000; Nykanen et al., 2001).

RNAse III and PAZ domains are sufficient for dicing activity, yet all metazoan dicers contain helicase domains (Figure 1.3). The studies about the function of the helicase domain of Dicer have been puzzling. Human Dicer was reported to be ATP-independent, and the removal of the helicase domain increased its efficiency if the helicase domain was removed (Zhang et al., 2002;

Ma et al., 2008). A different study implicated the helicase domain in processing thermodynamically unstable hairpin RNAs (Soifer et al., 2008). Helicase domain of Dicer was also implicated in siRNA pathway independent functions such as interaction with HIV-I Tat protein (Bennasser & Jeang, 2006) and mediating interferon induction (Deddouche et al., 2008).

The helicase domain of Dicer-2 belongs to the RIG-I (retinoic acid-inducible gene) family of RNA helicases. RIG-I recognizes viral nucleotides and activates interferon response in vertebrates (Yoneyama & Fujita, 2007; Zou et al., 2009). RIG-I was reported to sense triphosphates at the 5' end and to translocate along viral dsRNA. Triphosphate and dsRNA structure are required to induce ATPase activity, and subsequent translocation of RIG-I (Myong et al., 2009).

siRNA formation from long dsRNAs requires ATP in flies and fission yeast (Zamore et al., 2000; Colmenares et al., 2007). Dicer-2 is responsible for production of siRNAs and its helicase domain has a predicted ATP binding site. Moreover, a mutation in the predicted ATP binding site (walker A motif) of Dicer-2 was found to block inverted repeat derived siRNA silencing in *Drosophila* (Lee et al., 2004).

The following two chapters of my thesis work examined the effects of ATP and RNA binding protein partners on the miRNA and siRNA formation by Dicer-1 and Dicer-2. I used recombinantly expressed purified proteins for in vitro dicing assays and to measure the effect of partner proteins on the activity of *Drosophila*

Dicers. In our work, we tried to identify the ATP requiring step in dicing and explore the function of the helicase domain of *Drosophila* Dicer-2. In addition, we also explored the mechanistic basis of substrate specificities of Dicer-1 and Dicer-2. We found that RNA binding protein partners of Dicers and two small molecules, ATP and phosphate are key in regulating *Drosophila* Dicers' specificity.

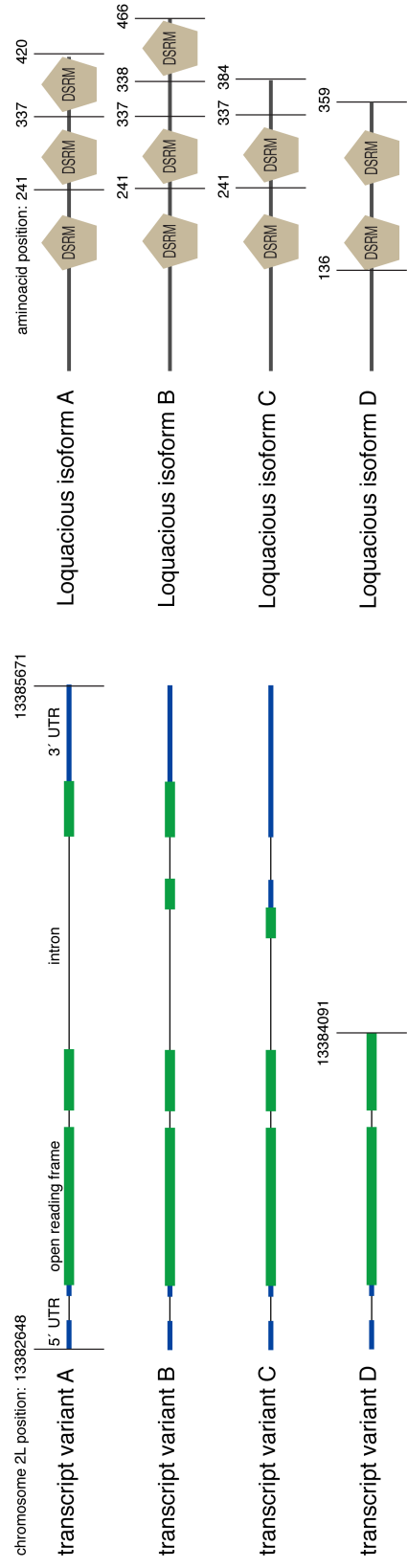


Figure 1.4. The Transcript Variants and Domain Structures of Loquacious (Loqs) Isoforms. Transcript variants of four different Loqs isoforms, which are transcribed from the same location and alternatively spliced, are shown on the left. On the right, domain architectures of four different Loqs isoforms are shown. The transcript and the domain structures Loqs isoforms were adapted from UCSC genome browser and smart domain protein analysis (Kent et al., 2002; Fujita et al., 2010; Schultz et al., 1998; Ponting et al., 1999).

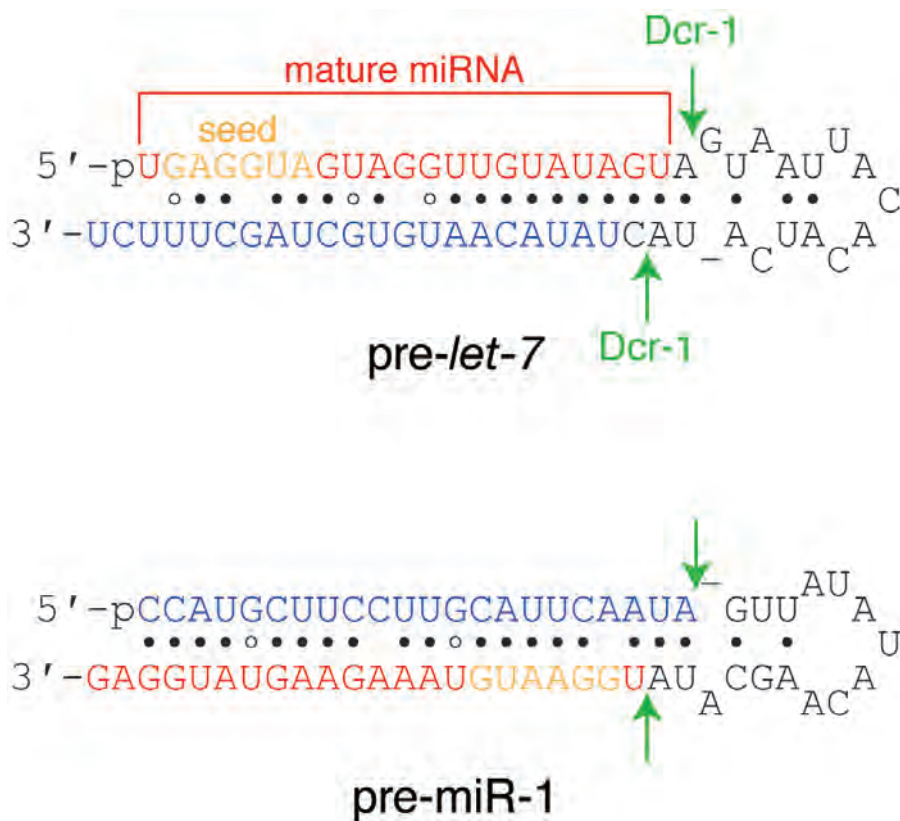


Figure 1.5. Structures of two pre-miRNAs; *pre-let-7* and *pre-miR-1*. Mature miRNA strand is in red and miRNA* strand is in blue. Seven nucleotides in yellow denotes the seed sequence, which is important for target mRNA recognition. Green arrows point at Dicer-1 cleavage site.

Piwi Dependent Small RNAs in *Drosophila melanogaster*

The second part of my thesis is about the somatic Piwi-dependent small RNA family. While siRNAs and miRNAs are found throughout the whole organism, piRNAs were discovered as a distinct class of germ line specific small RNAs (Vagin et al. 2006; Grivna et al. 2006). piRNAs are conserved throughout the animal kingdom, and protect the germline from transposon invasion (Grivna et al. 2006; Brennecke et al. 2007; Houwing et al. 2007; Das et al. 2008; Grimson et al. 2008). Animals without functional piRNAs are sterile (Schupbach et al 1991; Klattenhoff et al. 2007; Batista et al. 2008). piRNAs guide PIWI proteins to silence transposons in the germ line of animals (Cox et al. 2000; Aravin et al. 2001; Aravin et al. 2003; Grivna et al. 2006; Brennecke et al. 2007; Batista et al. 2008; Das et al. 2008; Chamberyron et al. 2008; Desset et al. 2008).

piRNA biogenesis and function diverge significantly from miRNAs and siRNAs (Brennecke et al. 2007; Li et al. 2008). piRNAs are longer, generally 24 to 30 nucleotides long, their 3' termini 2'O methylated , and they don't have double stranded RNA precursors (Aravin et al. 2001; Aravin et al. 2004; Vagin et al. 2006; Saito et al. 2007; Kirino et al. 2007; Ohara et al. 2007). Unlike siRNAs and miRNAs, piRNA production does not require Dicer and likely involves only single-stranded precursor RNAs (Vagin et al. 2006; Girard et al. 2006; Saito et al. 2007; Kirino et al. 2007; Ohara et al. 2007; Brennecke et al. 2007; Batista et al. 2008). The biogenesis of piRNAs and their role in transposon silencing have been elucidated mainly from studies of *Drosophila* ovaries and testes (Aravin et

al. 2001; Brennecke et al. 2007; Li et al. 2009). piRNA biogenesis utilize piwi-family proteins Aubergine, Argonaute 3, and Piwi but none of the Dicers (Vagin et al. 2006; Aravin et al. 2006; Grivna et al. 2006; Brennecke et al. 2007; Gunawardane et al. 2007; Klattenhoff et al. 2008; Khurana et al. 2010).

Drosophila gonads express three PIWI proteins: Piwi, Aubergine, and Ago3 (Ghildiyal et al. 2009; Cox et al. 1998; Cox et al. 2000; Harris et al. 2001; Vagin et al. 2006; Saito et al. 2006; Nishida et al. 2007; Yin et al. 2007; Klattenhoff et al. 2007; Dasset et al. 2008; Malone et al. 2009). Piwi localizes to the nucleus in both the germline and the surrounding somatic follicle cells, and helps regulate differentiation and patterning of the germline nurse cells as well as the oocyte (Cox et al. 1998; Cox et al. 2000).

Our lab and others have extensively used small RNA sequencing approaches to study this pathway in the germline (Vagin et al. 2006; Brennecke et al. 2007; Malone et al. 2009; Li et al. 2009). High throughput sequencing of piRNAs bound to Piwi, Aubergine, and Ago3 suggested a model called the “Ping-Pong” mechanism for the production and subsequent amplification of piRNAs in response to transcription of the transposons they target. The Ping-Pong model proposes that Aubergine and Ago3 collaborate both to increase the abundance of piRNAs and to bias piRNAs toward the antisense strand (Brennecke et al. 2007; Gunawarde et al. 2007). The detailed mechanism by which Aubergine and Piwi acquire primary piRNAs is unknown, but recent results suggest that they derive from long RNAs transcribed from “piRNA clusters”. These regions of the

genome are specifically transcribed in the gonads and are characterized by transposon-richness, and a dearth of genes (Olivieri et al. 2010). The Ping-Pong model postulates that Aubergine, bound to an anti-sense primary piRNA pairs with the mRNA transcript of an active transposon, resulting in cleavage of the target and generation of a 3' cleavage product bearing a 5' monophosphate. The 5' end of the 3' cleavage fragment then becomes the 5' end of a sense piRNA bound to Ago3. This "secondary piRNA" can then direct cleavage of a primary piRNA transcript derived from a piRNA cluster. The next step reverses this process. The Ago3:sense piRNA complex cleaves the long transcript from a piRNA cluster, generating antisense RNA fragments that bind to Aub. These fragments are envisioned to be trimmed to mature piRNAs (Brennecke et al. 2007; Gunawarde et al. 2007; Li et al. 2009).

Transposon derived piRNAs bound to Piwi silence transposon expression in the nucleus by a poorly understood mechanism. Although the mechanism is not clearly known and controversial, they are thought to transcriptionally silence genes and repeat sequences by DNA methylation or heterochromatin formation (Lees-Murdock et al. 2003; Saito et al. 2006; Kato et al. 2007). Heterochromatin dependent transcriptional silencing was detected at piRNA clusters yet the heterochromatin formation was suggested to be independent from the piRNA pathway (Moshkovich et al., 2010). Piwi was reported to interact with the histone protein HP1a and the HP1a interacting motif is required for the silencing function of Piwi (Brower-Toland et al., 2007). Moreover, Piwi and a piRNA species was

suggested to promote euchromatic histone modification in subtelomeric heterochromatin (Yin & Lin, 2007). Another observation related to the function of piRNA pathway came from positional effect variegation studies in *Drosophila* (Todeschini et al., 2010). In this study, authors took advantage of a transsilencing effect where a P transgene such as P-lacZ in subtelomeric heterochromatin is able to repress another P-lacZ copy in trans. This transsilencing effect, which is maternal and epigenetically transmitted through meiosis, was found to be fully or partially abolished by piRNA pathway genes aubergine and piwi (Todeschini et al., 2010).

In addition to their presence in animal germlines, the presence and function of piRNA-like small RNAs in the somatic tissues have been suggested (Yan et al. 2011; Lee et al. 2011; Morazzani et al. 2012) yet it was not clear whether they were methylated. Recently, methylated Piwi bound RNAs were found in the *Aplysia* central nervous system. piRNAs that are antisense to the promoter region of CREB2 are shown to be upregulated by serotonin. Upon upregulation, these piRNA species facilitates methylation of a conserved CpG island in the promoter region of CREB2, regulating transcription of CREB2. This event promotes long-term changes in neurons upon stimulation (Rajasethupathy et al. 2012).

The fourth chapter of my thesis addresses the question whether somatic piwi dependent small RNAs are present in *Drosophila* head tissues. We have investigated whether the piRNA-like reads in our many *Drosophila* head libraries

come from the germline as a contaminant or are soma-specific. We found that most of the piRNA reads in our published head libraries show high similarity to germline piRNAs. However, piRNA-like reads from manually dissected heads are distinct from germline piRNAs, proving the presence of somatic piRNA-like small RNAs.

**CHAPTER II: PHOSPHATE AND R2D2 RESTRICT THE
SUBSTRATE SPECIFICITY OF DICER-2, AN ATP-DRIVEN
RIBONUCLEASE.**

Preface

The work in this chapter was the result of a collaborative work. Dr. Gang Lu, Dr. Robert Dutcher, and Dr. Yeming Wang from Dr. Traci M Tanaka Hall's laboratory recombinantly expressed Dicer-2 and Dicer-2^{G31R}; designed, recombinantly expressed and purified Dicer-2^{D1217,1614N}. Dr. Ryuya Fukunaga performed dicing experiments in Figure 2.2 and 2.S1B, 2.S3A. Dr. Ryuya Fukunaga purified recombinantly expressed Dicer-2 and Dicer-2^{G31R}. I carried out the dicing and ATP consumption assays and northern hybridization in Figures 2.1, 2.S1, 2.3, 2.S3B, 2.S3C, 2.4, 2.S4, 2.5, Table 2.1, and 2.2. I purified recombinantly expressed Dicer-2 and Dicer-2/R2D2.

Abstract

Drosophila Dicer-2 generates small interfering RNAs (siRNAs) from long double-stranded RNA (dsRNA), whereas Dicer-1 produces microRNAs from pre-microRNA. What makes the two Dicers specific for their biological substrates? We find that purified Dicer-2 can efficiently cleave pre-miRNA, but that inorganic phosphate and the Dicer-2 partner protein R2D2 inhibit pre-miRNA cleavage. Dicer-2 contains C-terminal RNase III domains that mediate RNA cleavage, and an N-terminal helicase motif whose function is unclear. We show that Dicer-2 is a dsRNA-stimulated ATPase that hydrolyzes ATP to ADP; ATP hydrolysis is required for Dicer-2 to process long dsRNA, but not pre-miRNA. Wild-type Dicer-2, but not a mutant defective in ATP hydrolysis, can generate siRNAs faster than it can dissociate from a long dsRNA substrate. We propose that the Dicer-2 helicase domain uses ATP to generate many siRNAs from a single molecule of dsRNA before dissociating from its substrate.

Introduction

In *Drosophila melanogaster*, distinct pathways produce 21 nt small interfering RNAs (siRNAs) and ~22 nt microRNAs (miRNAs). The RNase III enzyme Drosha, aided by its partner protein, Pasha, cleaves primary miRNAs to release pre-miRNAs, ~70 nt long stem-loop structures that contain a mature miRNA within their stems (Lee et al., 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Han et al., 2006). The pre-miRNA is then cleaved by Dicer-1, acting with its dsRNA-binding domain (dsRBD) protein partner, Loquacious-PB (Loqs-PB), to liberate a duplex comprising the mature miRNA bound to its miRNA*, a partially complementary small RNA derived from the opposite arm of the pre-miRNA stem (Förstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005; Ye et al., 2007). Mature miRNAs can derive from either the 5' or 3' arm of the pre-miRNA stem.

In contrast to miRNAs, *Drosophila* siRNAs are generated by Dicer-2 (Lee et al., 2004), which forms a stable complex with the dsRNA-binding protein R2D2 (Liu et al., 2003). In vitro, Dicer-2 can produce siRNAs in the absence of R2D2, but both Dicer-2 and R2D2 are required to load siRNAs into Ago2 (Liu et al., 2003; Tomari et al., 2004; Pham and Sontheimer, 2005; Liu et al., 2006; Tomari et al., 2007). Exogenous siRNAs derive from long dsRNA molecules that are generated experimentally, from viral RNA genomes or intermediates of replication, whereas endo-siRNAs derive from convergent transcription of mRNAs or from RNA from mobile genetic elements (Yang and Kazazian, 2006;

Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a; Okamura et al., 2008b; Tam et al., 2008; Watanabe et al., 2008). A special class of endogenous siRNAs, hp-esiRNAs derive from partially self-complementary hairpin transcripts. Production of esiRNAs by Dicer-2 requires an alternative partner protein, Loqs-PD. This Loqs isoform contains only two of the three dsRBDs found in the Dicer-1 partner protein, Loqs-PB (Okamura et al., 2008b; Hartig et al., 2009; Zhou et al., 2009; Miyoshi et al., 2010; Hartig and Forstemann, 2011).

Dicer-1 and Dicer-2 each contain two RNase III domains, which form an intramolecular heterodimer whose dimer interface creates two active sites (Zhang et al., 2004; Macrae et al., 2006; Ye et al., 2007). Like other members of the Dicer family, Dicer-1 and Dicer-2 each contain a C-terminal dsRBD and a central PAZ domain, an RNA-binding motif specialized to recognize the two-nucleotide, 3' single-stranded tails of Drosha and Dicer products (Bass, 2000; Cerutti et al., 2000; Lingel et al., 2003; Song et al., 2003; Yan et al., 2003; Lingel et al., 2004; Ma et al., 2004; Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006; MacRae et al., 2007). The structure of *Giardia intestinalis* Dicer and functional studies using human Dicer suggest that the distance between the PAZ domain and the active sites of the RNase III domains establishes the length of the small RNA product (Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006; MacRae et al., 2007; Takeshita et al., 2007).

Differences in the domain architecture of Dicer-1 and Dicer-2 are unlikely to

explain their distinct substrate specificities. *Drosophila* Dicer-2 shares its domain architecture—including N-terminal DExDc, Helicase C, PAZ, RNase IIIa, RNase IIIb, dsRBD domains—with human Dicer, which produces both siRNAs and miRNAs. In contrast, Dicer-1 lacks a DExDc domain, yet has a Helicase C domain. DExDc/H and DEAD box domains are found in a wide range of RNA “helicases,” proteins that couple ATP hydrolysis to RNA binding or unwinding (Pyle, 2008). In addition to unwinding nucleic acids, helicase domains can couple ATP hydrolysis to translocation along nucleic acid molecules, rearrange RNA:protein or protein: protein interactions, or act as RNA chaperones (Bianco and Kowalczykowski, 2000; Beran et al., 2006; Bowers et al., 2006; Dumont et al., 2006; Halls et al., 2007; Pyle, 2008; Franks et al., 2010).

In *Drosophila* and *Caenorhabditis elegans*, siRNA production from long dsRNAs requires ATP (Zamore et al., 2000; Bernstein et al., 2001; Ketting et al., 2001; Nykanen et al., 2001). Moreover, a point mutation (G31R) in the *Drosophila* Dicer-2 helicase domain blocks siRNA production in vivo, although the protein retains the ability to collaborate with R2D2 to load synthetic siRNAs (Lee et al., 2004; Pham et al., 2004) and highly paired miRNA/miRNA* duplexes (Forstemann et al., 2007) into Ago2. In contrast, a mutation predicted to inhibit nucleotide binding by the human Dicer helicase domain does not affect dicing (Zhang et al., 2002).

What restricts a given Dicer to a specific dsRNA substrate? We find that purified, recombinant Dicer-2 can cleave pre-miRNA, but that R2D2 inhibits

processing of pre-miRNA by Dicer-2, while promoting use of its biologically relevant substrate by reducing the K_M of Dicer-2 for long dsRNA. Moreover, physiological concentrations of inorganic phosphate block pre-miRNA processing by Dicer-2, but do not inhibit processing of long dsRNA by Dicer-2 or pre-miRNA processing by Dicer-1. Thus, the characteristic specificity of Dicer-2 for long dsRNA is not intrinsic to the enzyme, but rather emerges in the presence of inorganic phosphate and R2D2. We also find that Dicer-2 is a dsRNA-stimulated ATPase that hydrolyzes ATP to ADP; ATP hydrolysis is required for Dicer-2 to process long dsRNA but not pre-miRNA. Wild-type Dicer-2, but not a mutant defective in ATP hydrolysis, can generate siRNAs faster than it dissociates from its long dsRNA substrate. We envision that the Dicer-2 helicase domain uses ATP to drive the movement of Dicer-2 along dsRNA, enabling it to generate many siRNAs from a single molecule of substrate before dissociating from the dsRNA.

Results

Dicer-2 Processes Pre-miRNA Inaccurately

In vivo, Dicer-1 but not Dicer-2 is required to produce miRNAs from the stems of pre-miRNA. Surprisingly, purified, recombinant Dicer-2 cleaved pre-miRNA (Figure 2.1). However, the size of the miRNA and miRNA* products generated by Dicer-2 differed from those produced by Dicer-1: the predominant Dicer-2 product was one nucleotide shorter than that produced by Dicer-1. For miRNAs residing on the 5' arm of their pre-miRNA, such a difference in size would not alter the miRNA seed sequence, but could promote their inappropriate loading into Argonaute2, which favors 21 nt RNAs, rather than Argonaute1, which prefers 22mers (Ameres et al., 2011). For the ~60% of *D. melanogaster* miRNAs derived from the 3' arm of their pre-miRNA, the seed sequence of the Dicer-2 product would differ from the authentic miRNA and would therefore regulate a repertoire of mRNAs different from that controlled by the authentic miRNA. The biological consequences of such misregulation are predicted to be dramatic, suggesting that processing of pre-miRNA by Dicer-2 is suppressed in vivo.

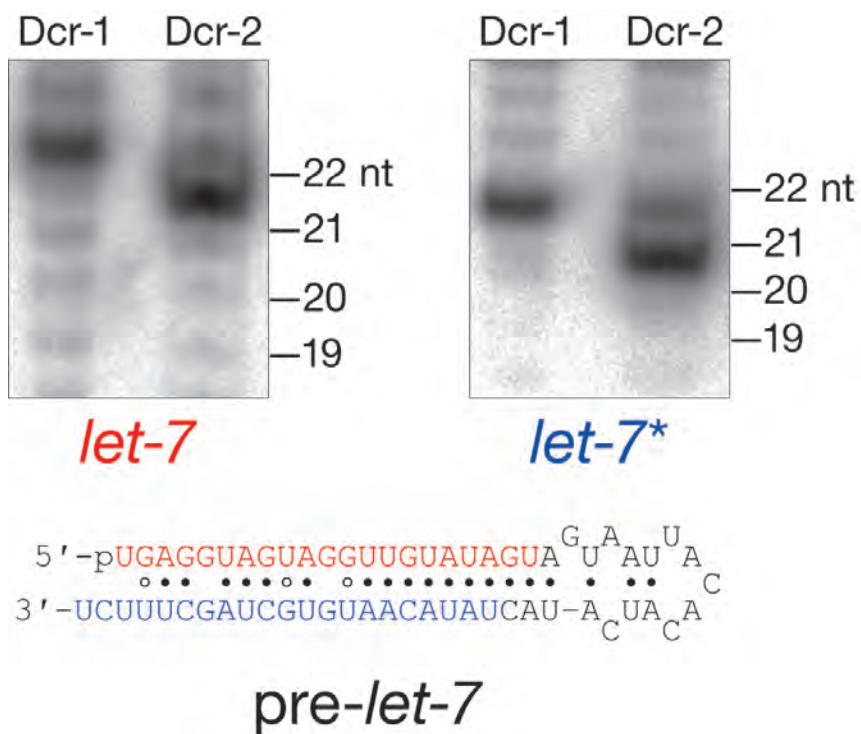


Figure 2.1. Dicer-1 and Dicer-2 Produce Different Products from Pre-let-7.

Synthetic, 5 monophosphorylated pre-let-7 (1.0 μ M) was incubated with Dicer-1 (6.0 nM) or Dicer-2 (16.2 nM) for 1 h. Products were resolved by electrophoresis. *let-7* and *let-7** were detected by Northern hybridization.

R2D2 Inhibits Dicing of Pre-miRNA by Dicer-2

We compared the rate of pre-let-7 processing by Dicer-2 alone to the rate of purified Dicer-2/R2D2 heterodimer, Dicer-2 supplemented with equimolar Loqs-PD, and Dicer-2/R2D2 heterodimer supplemented with Loqs-PD. R2D2 significantly inhibited pre-let-7 processing by Dicer-2 when either enzyme (data not shown) or substrate was in excess (Figure 2.2A; p-value for excess substrate= 0.0009). Similar inhibition of Dcr-2 processing by R2D2 was observed for a 25 bp RNA duplex (Figure 2.S1A), suggesting that R2D2 suppresses processing of short, double-stranded substrates irrespective of the extent of complementarity or the presence of a loop. In contrast, we did not detect any inhibition of pre-let-7 processing when Loqs-PD was added to Dicer-2, even though the same preparation of Loqs-PD lowered the K_M of Dicer-2 for long dsRNA 10-fold (Table 2.1). These data suggest that, R2D2, but not Loqs-PD, helps suppress pre-miRNA processing by Dicer-2 in vivo.

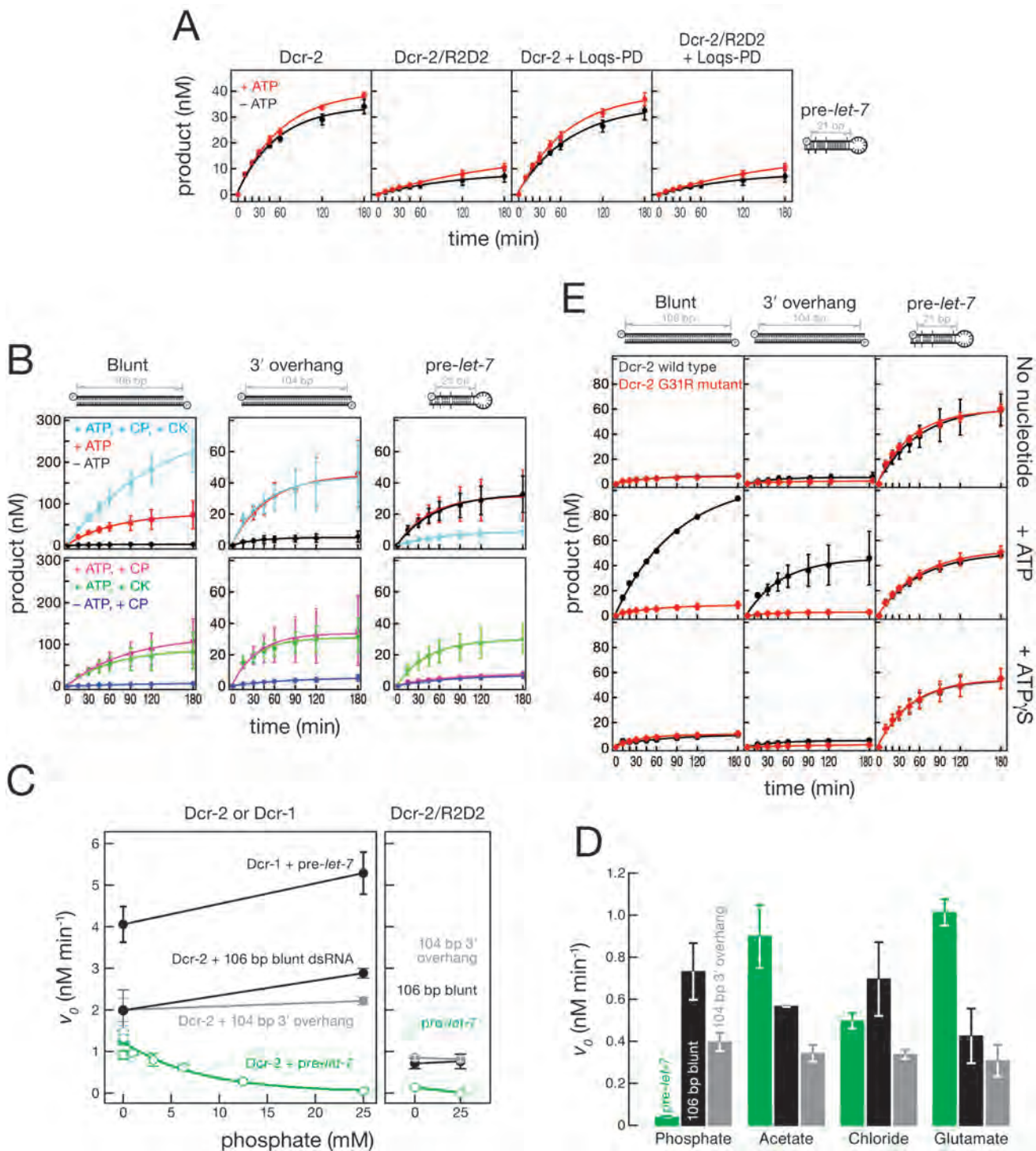


Figure 2.2. R2D2 and Phosphate Inhibit Dicer-2 Processing of Short

Substrates. (A) let-7 production was monitored using 5 ³²P-radiolabeled pre-let-7 (100 nM) with or without ATP for Dicer-2 alone (8 nM), Dicer-2/R2D2 (8 nM), Dicer-2 + Loqs-PD (8 nM + 8 nM) or Dicer-2/R2D2 + Loqs-PD (8 nM + 8 nM). (B) Processing of internally ³²P-radiolabeled long dsRNA or 5 ³²P-radiolabeled pre-let-7 (100 nM) by Dicer-2 (8 nM). CK, creatine kinase; CP, creatine phosphate. (C) Initial velocities for the processing of 100 nM pre-let-7 or long dsRNA (106 bp blunt ended or 104 bp with 2 nt, 3 overhanging ends) by 8 nM Dicer-2 or Dicer-2/R2D2 in the presence of increasing concentrations of potassium phosphate. Total potassium in the reaction was kept constant. (D) Initial velocities for Dicer-2 processing pre-let-7, 106 bp blunt ended dsRNA or 104 bp dsRNA with 2 nt, 3 overhanging ends in the presence of 25 mM potassium phosphate, acetate, chloride, or glutamate. (E) Processing of internally ³²P-radiolabeled long dsRNA or 5 ³²P-radiolabeled pre-let-7 substrate (100 nM) by wild-type or G31R mutant Dicer-2 (8 nM) with or without ATP or with ATPS (1 mM). Values are mean ± standard deviation for three independent experiments.

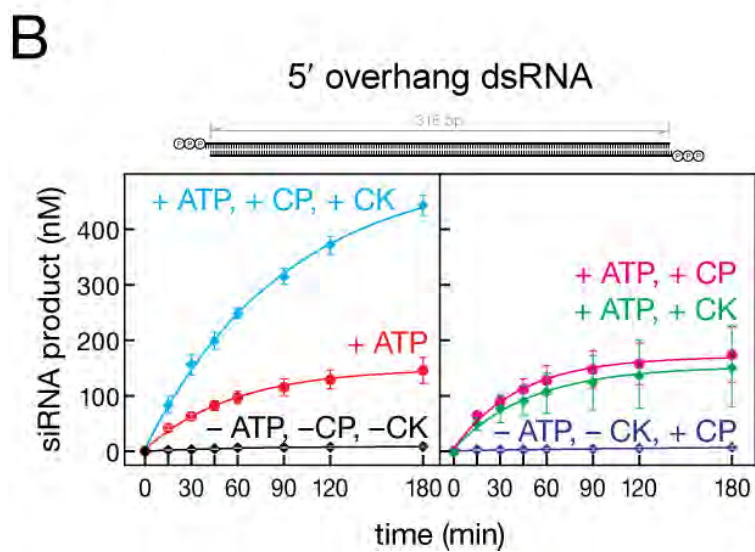
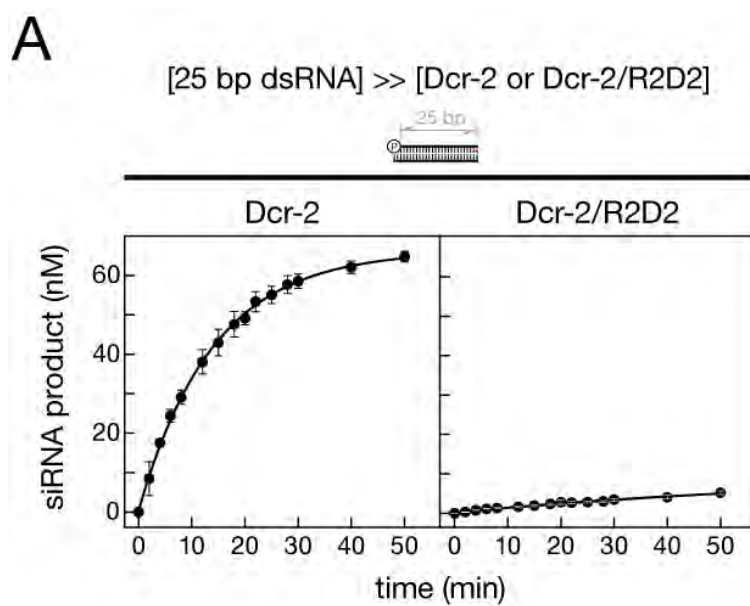
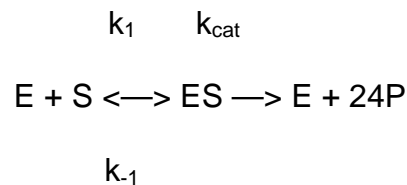


Figure 2.S1. R2D2 Inhibits Dicer-2 Processing of a 25 bp dsRNA Substrate and Creatine Phosphate Inhibits Dicer-2 Processing of a Long dsRNA Bearing 5 Triphosphate termini. (A) siRNA production by Dcr-2 (5.4 nM) or Dcr-2/R2D2 heterodimer (12 nM) was measured for a 25 bp dsRNA (100 nM). Red indicates deoxynucleotides. (B) A 5 triphosphorylated 316 bp dsRNA (100 nM) bearing 2 nt 5 overhanging ends was incubated with Dicer-2 (8 nM) in the presence or absence of ATP, creating kinase (CK), or creatine phosphate (CP).

R2D2 and Loqs-PD Decrease the K_M of Dicer-2 for Long dsRNA

Dicer-2 forms a stable heterodimer with R2D2 (Liu et al., 2003; Tomari et al., 2004; Liu et al., 2006), but it is unknown whether R2D2 modulates dicing rate. We measured the initial rate of dicing by Dicer-2 alone or by the Dicer-2/R2D2 heterodimer using increasing concentrations of long dsRNA substrate and saturating ATP (1 mM). For both Dicer-2 and Dicer-2/R2D2 the data fit well to the Michaelis-Menten kinetic scheme (Figure 2.S2)



where k_{cat} is the rate of complete conversion of substrate into siRNAs at saturating dsRNA concentration. The k_{cat} of Dicer-2/R2D2 processing a 515 bp dsRNA ($0.03 \pm 0.02 \text{ min}^{-1}$) was indistinguishable from that of Dicer-2 alone ($0.03 \pm 0.01 \text{ min}^{-1}$; Table 2.1). In contrast, the K_M for Dicer-2/R2D2 ($2 \pm 1 \text{ nM}$) was less than that of Dicer-2 alone ($6 \pm 2 \text{ nM}$, $p\text{-value} = 0.04$), suggesting that R2D2 increases the affinity of Dicer-2 for long dsRNA (Table 2.1). Similarly, supplementing Dicer-2 (2 nM) with purified recombinant Loqs-PD (2 nM) did not alter the k_{cat} , but did decrease K_M ($0.4 \pm 0.1 \text{ nM}$, $p\text{-value} = 0.02$), suggesting that Loqs-PD also increases the affinity of Dicer-2 for long dsRNA (Table 2.1).

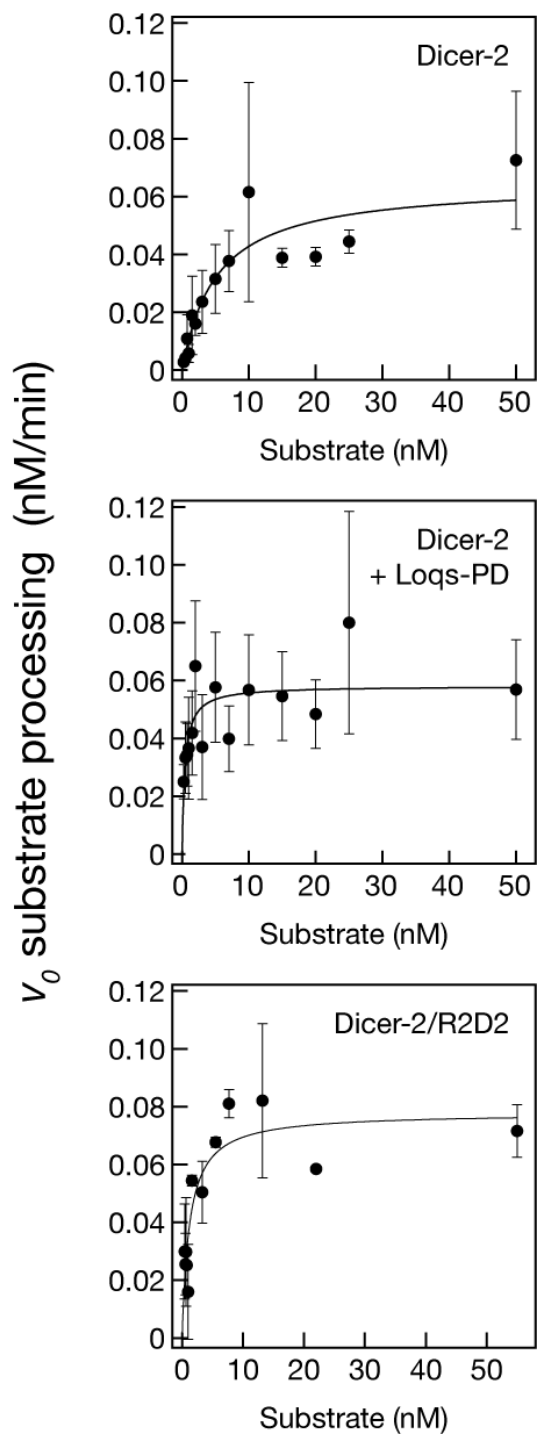


Figure 2.S2. Michaelis-Menten Analysis of Dicer-2, Dicer-2 + Loqs-PD, and Dicer-2/R2D2 Using a 515 bp dsRNA Substrate. The initial rate (velocity) of substrate processing by 2 nM Dicer-2 , 3 nM Dicer-2/R2D2, and 2 nM Dicer-2 + 2 nM Loqs-PD was measured for increasing concentrations of a 515 bp long dsRNA in the presence of 1 mM ATP. The data were fit to the Michaelis-Menten equation. V_{max} corresponds to the rate of complete conversion of a molecule of substrate into 24 siRNAs at saturating dsRNA concentration.

Table 2.1. Michaelis-Menten Analysis of Dicer-2 and Partner Proteins. Dicer-2, Dicer-2/R2D2, or Dicer-2 supplemented with equimolar Loqs-PD was incubated with a 515 bp dsRNA and saturating ATP (1 mM) ATP. The initial rates of converting dsRNA into siRNA for increasing amounts of substrate were measured and fit to the Michaelis-Menten equation (Figure S5). The table reports mean \pm standard deviation for three trials.

| | K_M (nM) | Change in K_M | V_{max} (nM min ⁻¹) | $[E_T]$ (nM) | k_{cat} (min ⁻¹) | Change in k_{cat} | k_{cat} / K_M (nM ⁻¹ min ⁻¹) | Change in k_{cat} / K_M |
|-------------------|---------------|--------------------|--------------------------------------|-----------------|-----------------------------------|------------------------|--|------------------------------|
| Dicer-2 | 6 \pm 2.0 | 1 | 0.07 \pm 0.01 | 2 | 0.03 \pm 0.01 | 1 | 0.005 \pm 0.003 | 1 |
| Dicer-2/R2D2 | 2 \pm 1 | 0.3 | 0.09 \pm 0.06 | 3 | 0.03 \pm 0.02 | 1 | 0.02 \pm 0.02 | 4 |
| Dicer-2 + Loqs-PD | 0.4 \pm 0.1 | 0.07 | 0.06 \pm 0.02 | 2 | 0.03 \pm 0.01 | 1 | 0.08 \pm 0.03 | 10 |

Phosphate Inhibits Dicing of Pre-miRNA by Dicer-2

In flies, dicing of long dsRNA requires ATP (Nykanen et al., 2001). Typically, creatine kinase (CK) and creatine phosphate (CP) are included in dicing reactions to maintain high levels of ATP and constant levels of free Mg^{2+} . Relative to ATP alone, the inclusion of CK and CP modestly enhanced Dicer-2 processing of both a 106 bp dsRNA bearing a 5' monophosphorylated blunt end (Figure 2.2B) and a 316 bp dsRNA bearing a 5' triphosphorylated, 2 nt, 5' overhang (Figure 2.S1B), but did not enhance processing of a 104 bp dsRNA with a 5' monophosphorylated, 2 nt 3' overhang (Figure 2.2B). In contrast, standard "ATP" conditions (+ATP, +CP, +CK) inhibited pre-let-7 processing by Dicer-2. More detailed analyses revealed that CP sufficed to inhibit pre-let-7 dicing.

CP can be hydrolyzed in water to creatine and phosphate. We therefore tested whether inorganic phosphate inhibited pre-let-7 processing by Dicer-2. We measured the initial rate of processing (v_0) with increasing concentrations of potassium phosphate (KH_2PO_4/K_2HPO_4 , pH 7.4) for pre-let-7, a 106 bp blunt end dsRNA and a 104 bp dsRNA with 2 nt 3' overhanging ends. Physiological concentrations of phosphate (Burt et al., 1976; Erecinska et al., 1977; Auesukaree et al., 2004) inhibited processing of pre-let-7, but neither of the two dsRNAs (Figure 2.2C). We observed little or no inhibition of pre-miRNA processing by Dicer-2 with 25 mM acetate, chloride, or glutamate (Figure 2.2D). Moreover, none of the anions—including phosphate—had a significant effect on

the dicing of long dsRNA. Phosphate further inhibited the low level of pre-let-7 processing of Dicer-2/R2D2 (Figure 2.2C) and Dicer-2/R2D2 + Loqs-PD, and suppressed pre-let-7 processing by Dicer-2 + Loqs-PD (data not shown). In contrast, processing of pre-let-7 by Dicer-1 was unaffected by phosphate (Figure 2.2C). We conclude that under physiological conditions—2–25 mM phosphate and a majority of Dicer-2 complexed with R2D2—Dicer-2 is unlikely to use pre-miRNA as a substrate.

ATP Hydrolysis and siRNA Production by Dicer-2

The Dicer-2 G31R point mutation, which lies in the protein's DExDc motif, uncouples Argonaute2 loading from dsRNA dicing (Lee et al., 2004). The mutation is predicted to disrupt ATP binding. Consistent with the requirement for ATP in siRNA production by Dicer-2, dsRNA processing by purified recombinant Dicer-2^{G31R} was significantly less than wild-type Dicer-2 for both substrates with blunt and 3' overhanging ends (Figure 2.2D, p-value = 2.2×10^{-6} for blunt end substrate). While dsRNA processing by wild-type Dicer-2 was strongly stimulated by ATP, mutant Dicer-2^{G31R} was not (Figure 2.2D). Nonetheless, Dicer-2^{G31R} cleaved pre-let-7 as efficiently as the wild-type enzyme, consistent with the finding that ATP was not required for wild-type Dicer-2 to cleave pre-let-7 (Figure 2.2A). Moreover, processing of long dsRNA by wild-type Dicer-2 was inhibited by adenosine 5'-O-(3-thio)triphosphate (ATPgS), but processing of pre-let-7 was not (Figure 2.2D).

In agreement with ATP γ S inhibiting dsRNA processing, Dicer-2 hydrolyzed ATP to ADP (Figure 2.3A). Both ATP and ATP α S supported the production of siRNA from long dsRNA, but ATP γ S did not (Figure 2.3B). In the presence of 1 mM ATP, the rate of dicing long dsRNA declined exponentially with increasing ATP γ S, suggesting that ATP γ S competes with ATP for binding to Dicer-2 and that once bound, ATP γ S is not efficiently hydrolyzed, preventing the production of siRNAs from long dsRNA.

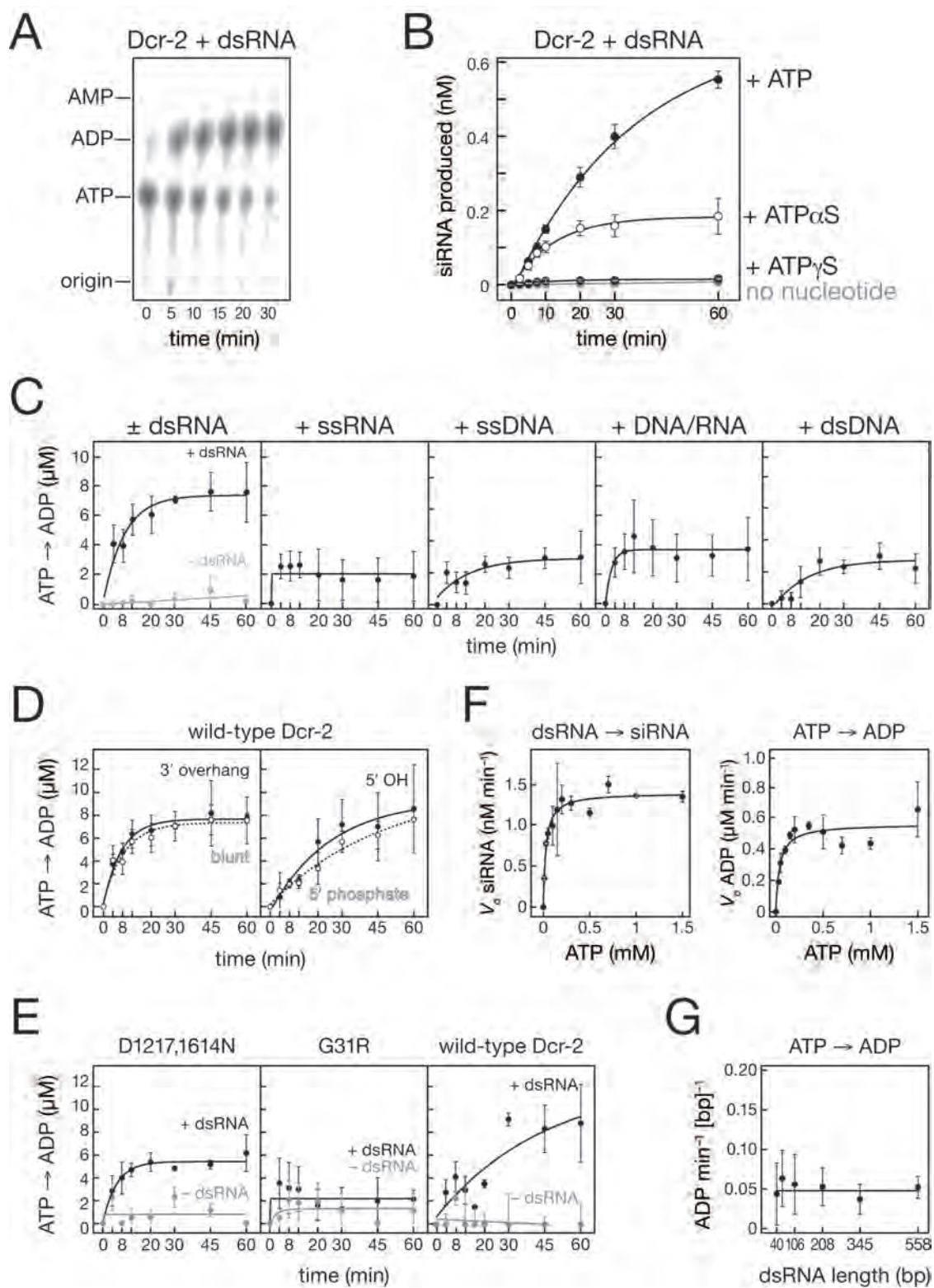


Figure 2.3. Dicer-2 Requires ATP Hydrolysis to Process Long dsRNA. (A)

Dicer-2 (10.8 nM) was incubated with 2.5 μM ^{-32}P ATP and 150 nM 515 bp dsRNA, and ATP hydrolysis monitored by thin-layer chromatography. (B) siRNA production by Dicer-2 (5.4 nM) from a 515 bp dsRNA (25 nM) was monitored with or without ATP or ATPgS (1 mM). (C) ATP hydrolysis by Dicer-2 (5.4 nM) was monitored for 120 nt long nucleic acid substrates (20 nM): dsRNA, single-stranded RNA, single-stranded DNA, RNA/DNA heteroduplex or double-stranded DNA in the presence of ATP (1 mM). (D) ATP hydrolysis by Dicer-2 (5.4 nM) in the presence of ATP (1 mM) was measured for 20 nM dsRNA bearing a 5 monophosphate and a 2 nt 3 overhang or blunt ends; the blunt ended dsRNA bearing either 5 monophosphate or 5 hydroxy termini. (E) ATP hydrolysis by mutant Dicer-2^{D1217,1614N} (3.3 nM), Dicer-2^{G31R} (2 nM) and wild-type Dicer-2 (2 nM) was measured in the absence or presence of 20 nM dsRNA bearing 5 monophosphate, blunt ends. (F) Left panel, initial rates for the conversion of substrate into siRNA by Dicer-2 (2.7 nM) for an internally ^{32}P -radiolabeled 515 bp dsRNA (150 nM) were measured at increasing concentrations of ATP. Right panel, initial rates for the hydrolysis of ATP to ADP by Dicer-2 in the presence of 150 nM, 515 bp dsRNA were measured for increasing ATP concentrations. The data were fit to the Michaelis-Menten equation. Table 2 reports the Michaelis-Menten parameters. (G) ATP consumption by Dicer-2 was measured in the presence of ATP (350 μM) and dsRNA substrates with blunt, 5 triphosphorylated termini for six different lengths: 40 bp (420 nM), 60 bp (280 nM), 106 bp (158

nM), 208 bp (80.5 nM), 345 bp (49 nM), 558 bp (30 nM). Substrate concentrations were selected to ensure an equal number of base pairs in each reaction. Values are mean \pm standard deviation for three independent experiments.

Table 2.2. Michaelis-Menten Analysis of siRNA Production and ATP Consumption by Dicer-2. Purified recombinant Dicer-2 was incubated with a 515 bp dsRNA and ATP. The initial rates of converting dsRNA into siRNA for increasing amounts of substrate were measured at saturating ATP (1 mM), and the initial rates of hydrolysis of ATP to ADP were measured at saturating dsRNA (150 nM) for increasing amounts of ATP. Different preparations of Dicer-2 were used here and in Table 1.

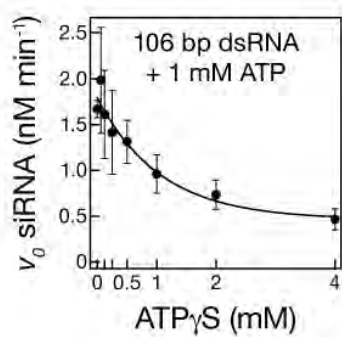
| | K_M (nM) | V_{max} (nM min ⁻¹) | [E _T] (nM) | k_{cat} (min ⁻¹) | k_{cat} / K_M (nM ⁻¹ min ⁻¹) |
|---------------------------|----------------|-----------------------------------|------------------------|--------------------------------|---|
| Substrate consumed | 6 ± 2 | 0.2 ± 0.1 | 1.2 | 0.2 ± 0.1 | 0.03 ± 0.02 |
| siRNA produced | 6 ± 2 | 5 ± 2 | 1.2 | 4 ± 1 | 0.7 ± 0.4 |
| ATP hydrolyzed | 14,000 ± 4,000 | 460 ± 70 | 5 | 93 ± 14 | 0.007 ± 0.002 |

Dicer-2 is a dsRNA-Stimulated ATPase

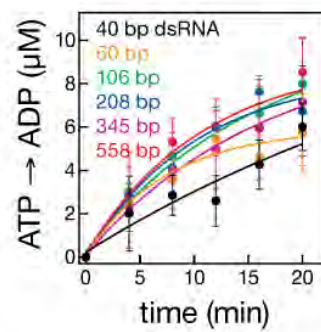
ATP hydrolysis by Dicer-2 increased when dsRNA was added (Figure 2.3C, p-value= 0.024, Wilcoxon rank sum test). The inclusion of 120 nt single-stranded RNA or DNA or 120 bp DNA/RNA heteroduplex or double-stranded DNA stimulated ATP hydrolysis less than the 120 bp dsRNA (all 20 nM; Figure 3C). A 25 nt single-stranded RNA or DNA, a 21 bp dsRNA or a 21 bp DNA/RNA heteroduplex, all failed to stimulate the Dicer-2 ATPase activity above the rate observed when no substrate was present (Figure 2.S3B). Neither the end structure of the dsRNA—blunt versus 3' overhang—nor the presence of a 5' monophosphate ($v_0^{5' \text{ PO}_4} = 0.4 \pm 0.1 \mu\text{M min}^{-1}$ versus $v_0^{5' \text{ OH}} = 0.3 \pm 0.1 \mu\text{M min}^{-1}$) significantly changed the ATP hydrolysis rate (Figure 2.3D).

ATP hydrolysis by Dicer-2 does not require dsRNA cleavage. A Dicer-2 mutant, D1217,1615N, in which a key aspartate in each of the two RNase III domain active sites was changed to asparagine, retained significant dsRNA-stimulated ATPase activity (p-value = 0.02; Figure 2.3E). Under multiple-turnover conditions, Dicer-2^{D1217,1615N} was essentially inactive for dicing, and siRNA production was detected only when [Dicer-2^{D1217,1615N}] >> [substrate] (data not shown). In contrast, Dicer-2^{G31R}, which also does not support multiple-turnover dicing of long dsRNA, did not hydrolyze ATP in the presence of dsRNA (Figure 2.3E). We conclude that the Dicer-2 helicase domain is responsible for the enzyme's dsRNA-stimulated ATPase activity.

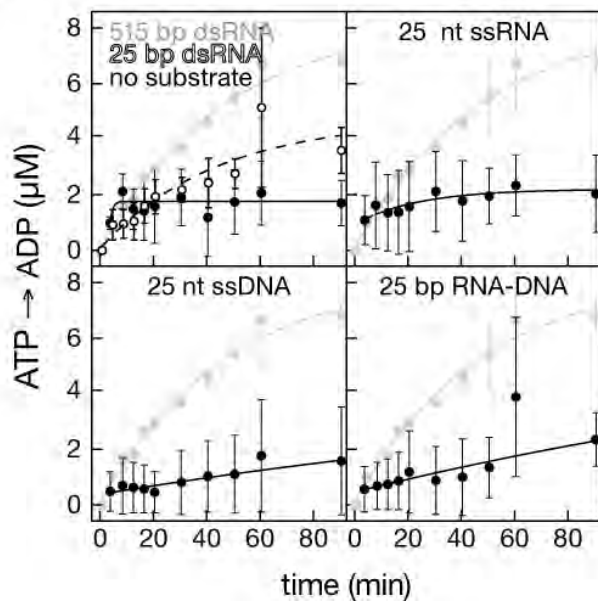
A



B



C



D

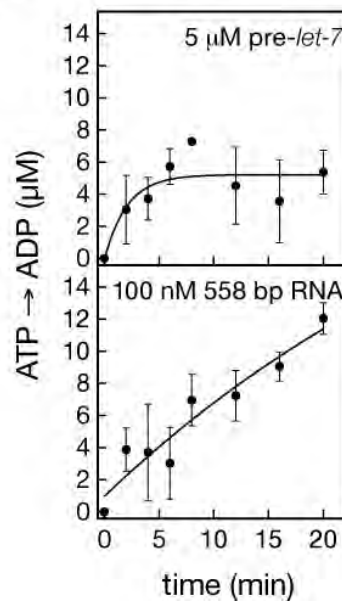


Figure 2.S3. ATP Hydrolysis is Required for Efficient Long dsRNA

Processing by Dicer-2. (A) siRNA formation from 5 monophosphorylated, internally ^{32}P -radiolabeled, blunt ended, 106 bp dsRNA was monitored in the presence of 1 mM ATP in the presence of increasing concentrations of ATPS. Initial rates of conversion of dsRNA to siRNA were plotted as a function of ATPS concentration. (B) ATP consumption was measured over time for a 40 bp (420 nM), 60 bp (280 nM), 106 bp (158 nM), 208 bp (81 nM), 345 bp (49 nM), and a 558 bp (30 nM) dsRNA substrate. Initial rates (v_0) were calculated from the fitted exponential curves and used in Figure 3F. (C) The hydrolysis of ATP to ADP was measured for Dicer-2 alone (5.4 nM) or with 25 nt single-stranded (ss) synthetic RNA, 25 nt synthetic ssDNA, 25 bp synthetic RNA or RNA-DNA duplex. ATP hydrolysis by Dicer-2 (5.4 nM) in the presence of a 515 bp dsRNA is shown in gray for reference. Substrate concentrations were chosen to achieve equivalent nt·mole·l⁻¹: ssRNA and ssDNA, 4.4 μM ; RNA and DNA-RNA duplex, 2.2 μM ; and 515 bp dsRNA, 100 nM. (D) The hydrolysis of ATP to ADP was measured for Dicer-2 (5 nM) with pre-let-7 (5 μM) or with 558 bp dsRNA (100 nM) in the presence of 1 mM ATP.

ATP Consumption and Dicing are Coupled

We measured the initial rates of siRNA production and ATP hydrolysis using a 515 bp dsRNA in the presence of increasing concentrations of ATP. The dependence on ATP concentration of both siRNA production and ATP hydrolysis fit well to the Michaelis-Menten kinetic scheme (Figure 2.3F). When both substrate and ATP were saturating (i.e., $\geq 10 \times K_M$), the k_{cat} for ATP hydrolysis was $93 \pm 14 \text{ min}^{-1}$, whereas the k_{cat} for siRNA production was $4 \pm 1 \text{ min}^{-1}$, as inferred from the k_{cat} for the complete conversion of a molecule of substrate into siRNA. These values predict that 23 ± 8 molecules of ATP are hydrolyzed for each 21 nt siRNA formed (Figure 2.3F, Table 2.2). Such a high rate of ATP hydrolysis for each siRNA produced might suggest that ATP energy and siRNA production are poorly coupled. Alternatively, ATP might be hydrolyzed to power translocation of Dicer-2 along the dsRNA, with approximately one ATP molecule consumed for each base pair traversed, a rate similar to DNA and RNA translocases containing ATPase/helicase domains (Bianco and Kowalczykowski, 2000; Patel and Donmez, 2006; Seidel et al., 2008). Consistent with this idea, the rate of ATP consumption was essentially unchanged for different substrate lengths when the molar concentration of base pairs was kept constant (Pearson correlation, $r = 0.98$; Figure 2.3G, 2.S3B).

Evidence that Dicer-2 is Processive

If ATP fuels the translocation of Dicer-2 along dsRNA, dicing should produce successive siRNAs along the substrate. In the presence of ATP, Dicer-2 would be predicted to produce the first siRNA—i.e., the terminal siRNA—at roughly the same rate as subsequent, internal siRNAs. In contrast, in the absence of ATP, the rate of production of siRNA should decline as its distance from 5' end increases. To test these predictions, we synthesized three, identical 120 nt dsRNA substrates, each bearing a single ^{32}P radiolabel at position 16, 35, or 104 from the 5' end of one strand (Figure 2.4A). All substrates contained two deoxynucleotides at one end, forcing Dicer-2 to initiate processing from the opposite end (Figure 2.S4A) (Rose et al., 2005). We used this “one-ended” substrate to examine the production of the first, second, or fourth siRNA (see Experimental Procedures) in the presence or absence of ATP (Figure 2.4A). With 1 mM ATP, the initial rates for the production of the first, second, and fourth siRNAs were essentially indistinguishable ($\sim 28 \text{ nM min}^{-1}$); with no ATP, the rate for the first siRNA ($1.4 \pm 0.3 \text{ nM min}^{-1}$) was greater than that of the second ($0.9 \pm 0.1 \text{ nM min}^{-1}$), which was greater than the rate for the fourth siRNA ($0.2 \pm 0.1 \text{ nM min}^{-1}$; Figure 2.4A).

We can envision two explanations consistent with these results and previous studies on Dicer enzymes: either ATP fuels processive dicing of long dsRNA or dicing of long dsRNA in the presence of ATP comprises a slow initial binding step to the end of the substrate, followed by rapid but ATP-dependent

production of subsequent siRNAs. Such a two-step process might occur if the rate of production of the first siRNA was slowed by the blunt structure of the substrate; the first dicing event would convert the substrate end to a 2 nt, 3' overhang bearing a 5' monophosphate, with all subsequent siRNAs produced rapidly. To test this idea, we prepared a substrate bearing one blocked end and a 2 nt, 3' overhang bearing a 5' monophosphate at the other end (Figure 2.S4B). The rates of first siRNA production from both blunt and 3' overhanging end substrates were indistinguishable when the reactions contained ATP—conditions where dicing was efficient. They were also similar when ATP was omitted—when dicing was slow (Figure 2.4B). The subsequent siRNAs were also produced at similar rates from the two different substrates (Figure 2.S4B). Thus, we favor the hypothesis that ATP converts Dicer-2 from an inefficient, distributive enzyme into a processive enzyme.

Surprisingly, ATP also enhanced the production of the terminal siRNA from long dsRNA under single-turnover conditions ($[Dicer-2] \gg [dsRNA]$); the enhancement by ATP γ S was considerably weaker (Figure 2.4C). This suggests that ATP hydrolysis is required for the production of even the first siRNA. ATP was required irrespective of the terminal structure of the dsRNA (blunt versus 3' overhang), excluding a role for ATP in the binding of Dicer-2 to a particular type of dsRNA end.

Moreover, when the $[enzyme] > [dsRNA]$, helicase mutant Dicer-2^{G31R} produced the first siRNA at similar rates in the presence or absence of either

ATP or ATP γ S (Figure 2.4D and 2.S4C). For Dicer-2^{G31R}, the rate of production of the terminal siRNA was faster than that of the fourth siRNA, consistent with the idea that ATP hydrolysis converts Dicer-2 from a distributive to a processive enzyme (Figure 2.4E).

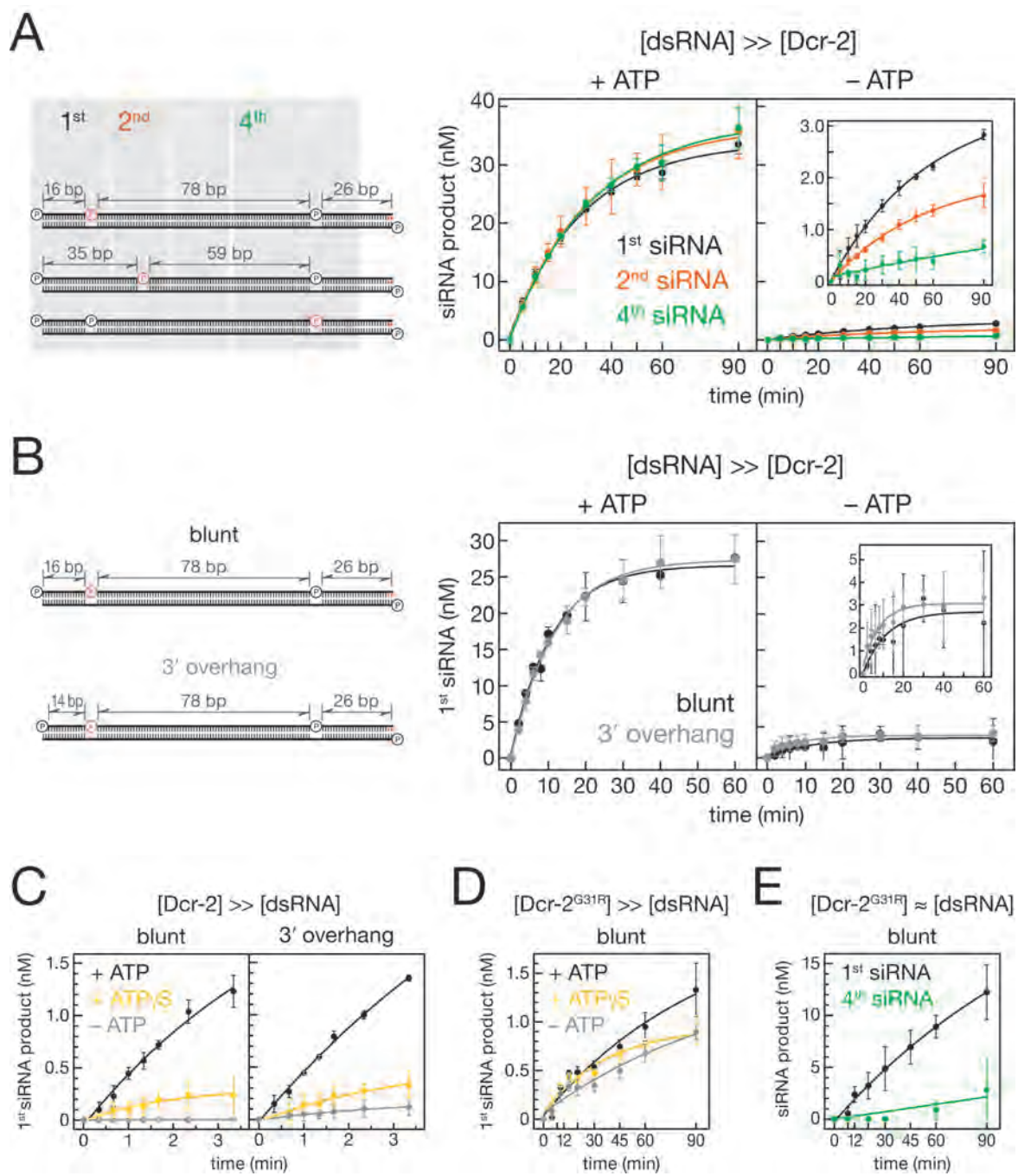


Figure 2.4. Evidence that Dicer-2 is Processive.

(A) Site-specifically ^{32}P -radiolabeled 120 bp dsRNAs were used to measure the rate of production of the first, second and fourth siRNAs from the all-RNA end of the substrate; the 3' end of the sense strand contained two deoxynucleotides (red) to block entry of Dicer-2 from the other end (Figure 2.S4A). Production of siRNA was monitored in the presence or absence of ATP using 100 nM dsRNA and 5.4 nM Dicer-2. Values correspond to the mean \pm standard deviation from three independent experiments. The data were fit to a single exponential function. In (A) and (B), pink denotes the position of the ^{32}P -radiolabel. (B) The rate of production of the initial siRNA was measured as in (A), but using a site-specifically ^{32}P -radiolabeled dsRNA with either a blunt (black) or a 2 nt 3 overhanging (red) end in the presence or absence of ATP. (C) The rate of production of the initial siRNA was monitored using 100 nM Dicer-2 and 10 nM dsRNA bearing either a blunt (left) or a 2 nt 3 overhanging (right) end in the presence or absence of ATP or ATPS. (D) The rate of production of the initial siRNA from a 120 bp blunt ended dsRNA (10 nM) was measured using 100 nM mutant Dicer-2^{G31R}. (E) The rate of production of the first and fourth siRNAs from a 120 bp blunt ended dsRNA (100 nM) was measured using 100 nM Dicer-2^{G31R}. Values are mean \pm standard deviation for three independent experiments.

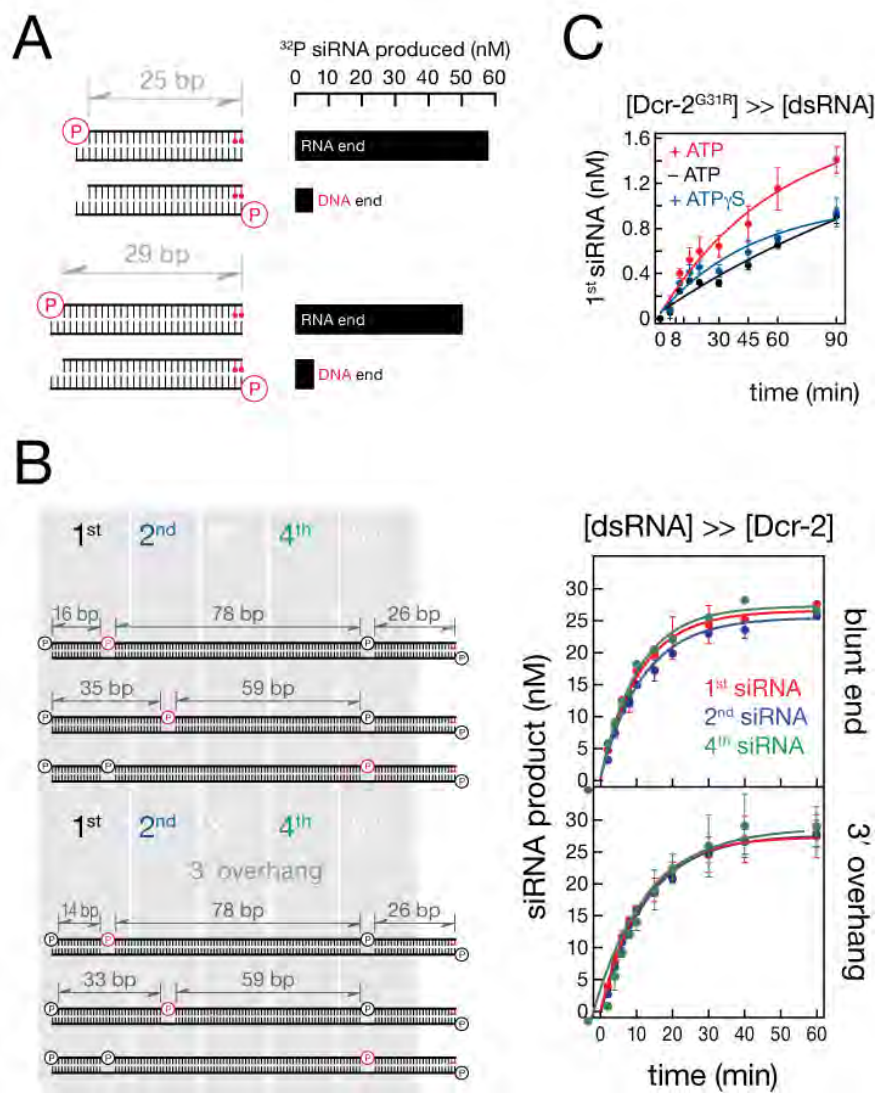


Figure 2.S4. The Rate of siRNA Production for Different dsRNA Termini:**Two 5 Deoxynucleotides Block Entry of Dicer-2, but Blunt and 3´****Overhanging Ends are Processed Similarly.** (A) Twenty-five and 29 bp dsRNA

substrates (73 nM) containing two 2' deoxycytidine nucleotides at the 3´ end

were incubated with 5.4 nM Dicer-2, and total siRNA production measured at 45

min. For each substrate, the 5´ ribonucleotide of the 5´ end or the end that

contained 3 deoxynucleotides on the complementary strand was ³²P-radiolabeled

to allow measurement of siRNA production from that end alone. (B) Site-

specifically ³²P-radiolabeled 120 bp dsRNAs (100 nM) were incubated with Dicer-

2 (5.4 nM) in the presence of 1 mM ATP to detect production of the first, second

and fourth siRNAs as diagrammed here and in Figure 4A. Right, the substrate

end where Dicer-2 enters was either blunt (top) or had a 2 nt 3´ overhang

(bottom). (C) Production of the initial siRNA (i.e., the first siRNA) by Dicer-2^{G31R}

(100 nM) from a 118 bp dsRNA (10 nM) bearing a 2 nt 3´ overhanging end was

measured in the presence or absence of ATP or in the presence of ATPγS (see

also Figures 4A and S4A).

Dicer-2 produces siRNAs without dissociating from the dsRNA

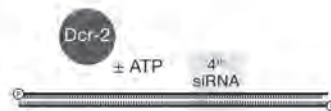
A processive enzyme can act multiple times on its substrate before dissociating. Thus, catalysis by processive enzymes resists dilution (Rivera and Blackburn, 2004). To test whether Dicer-2 remains physically associated with the long dsRNA after three subsequent dicing events, we used the 120 nt, site-specifically ^{32}P -radiolabeled dsRNA to monitor the rate of production of the fourth siRNA following dilution (Figure 2.4A). The dsRNA substrate (200 fmol) was first pre-incubated with Dicer-2 (54 fmol) for 2 min at 4°C to allow the enzyme to bind substrate. During this pre-incubation step, no siRNA was detected (data not shown). Next, the reaction was diluted 1,000-fold into buffer pre-warmed to 25°C. The diluted reaction was incubated at 25°C, and production of the fourth siRNA measured over time (Figure 2.5A). During the pre-incubation, the substrate concentration (20 nM) was > 3-fold greater than the K_M of Dicer-2 for long dsRNA; after dilution, the substrate concentration was ~300 times less than the K_M . For both the pre-incubation and the dilution steps, the dsRNA substrate was present at ~4-fold higher concentration than Dicer-2. When the pre-incubation was omitted, little fourth siRNA was produced (Figure 2.5B), demonstrating that the conditions largely prevented reassociation of Dicer-2 with dsRNA once it dissociated from the substrate.

When ATP was included in both the pre-incubation and the dilution buffer, 54 fmol of Dicer-2 produced 10 fmol of fourth siRNA in 1 h. About half as much fourth siRNA was produced when ATP was present in the dilution buffer but

omitted from the pre-incubation. When ATP was omitted from the dilution buffer, essentially no fourth siRNA was produced, regardless of whether ATP was present during the pre-incubation, suggesting that ATP dissociates rapidly from Dicer-2. More fourth siRNA was made when ATP was present in both the pre-incubation and dilution buffers than when pre-incubation was carried out in the absence of ATP (p -value = 0.015). We conclude that the initial binding of Dicer-2 to the end of long dsRNA is enhanced by ATP and that in the presence of ATP Dicer-2 remains associated with the dsRNA. We propose that the stably bound Dicer-2 then cleaves successive siRNAs along the dsRNA.

A

1. pre-incubate 2 min, 4°C



2. dilute 1000x, 25°C

3. assay 4th siRNA, 0–60 min

B

1. pre-incubation:

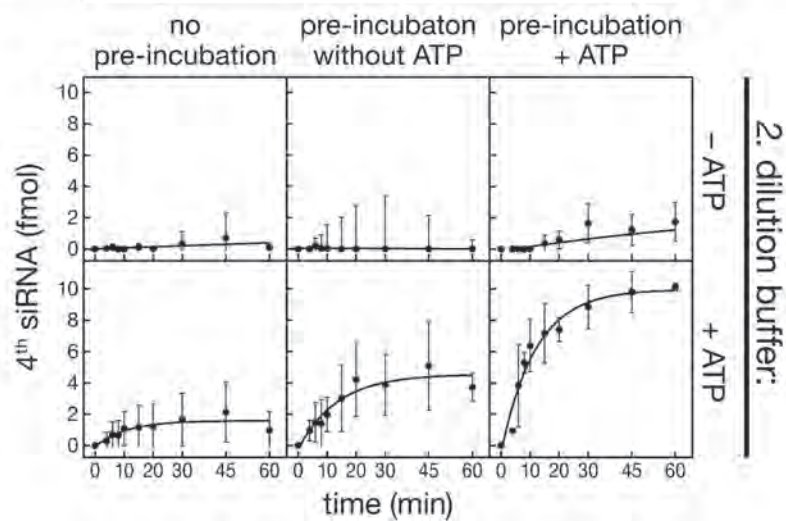


Figure 2.5. Dicer-2 Remains Associated with its Substrate in the Presence

of ATP. (A) Design of the experiment. The reaction contained 54 fmol Dicer-2 and 200 fmol dsRNA, corresponding to 5.4 nM enzyme and 20 nM dsRNA before dilution. (B) The rate of production of the fourth siRNA was measured for six different combinations of pre-incubation (no pre-incubation, without ATP, and with ATP) and dilution (without and with ATP). Values are mean \pm standard deviation for three independent experiments.

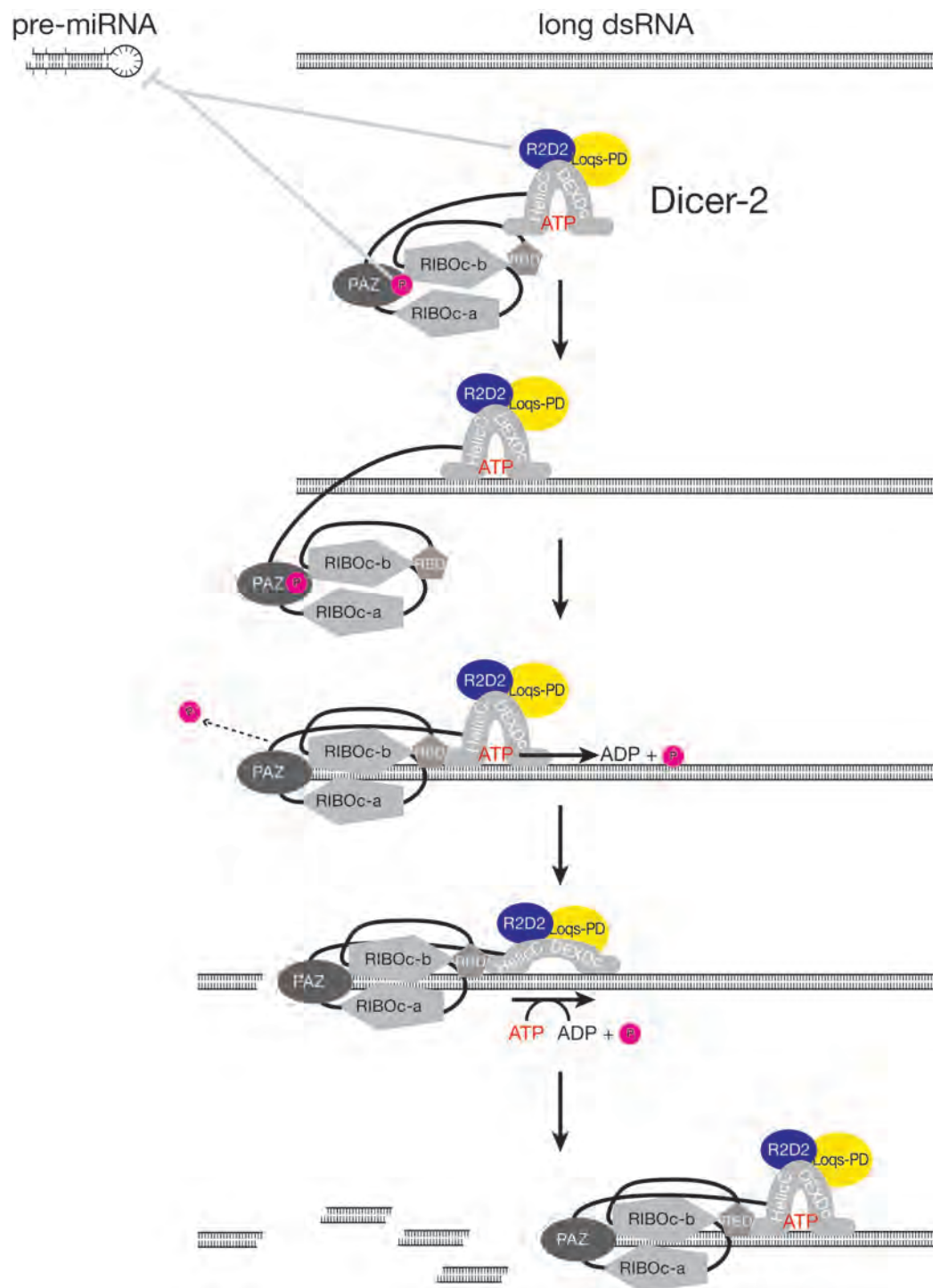


Figure 2.6. A Model for *Drosophila* Dicer-2.

Table 2.S1. List of Synthetic DNA Oligonucleotides. m, 2'-O-methyl ribose.

| | |
|---|---|
| Forward primer for 515 bp product for sense RNA from pEGFP-N3 | 5'-GCG TAA TAC GAC TCA CTA TAG GAC TCA GAT CTC GAG CTC AAG-3' |
| Reverse primer for 515 bp product for sense RNA from pEGFP-N3 | 5'-GCT GTT GTA GTT GTA CTC CAG-3' |
| Forward primer for 515 bp product for antisense RNA from pEGFP-N3 | 5'-GCG TAA TAC GAC TCA CTA TAG GCT GTT GTA GTT GTA CTC CAG-3' |
| Reverse primer for 515 bp product for antisense RNA from pEGFP-N3 | 5'-GAC TCA GAT CTC GAG CTC AAG-3' |
| Forward primer for 316 bp product for sense RNA from pEGFP-N3 | 5'-GCGTAATACGACTCACTATAGGGCCACAAGTTCAGCGTGTCC-3' |
| Reverse primer for 316 bp product for sense RNA from pEGFP-N3 | 5'-TCGATGCCCTTCAGCTCG-3' |
| Forward primer for 316 bp product for antisense RNA from pEGFP-N3 | 5'-GCGTAATACGACTCACTATAGTCGATGCCCTTCAGCTCG-3' |
| Reverse primer for 316 bp product for antisense RNA from pEGFP-N3 | 5'-GCCACAAGTTCAGCGTGTCC-3' |
| <i>let-7</i> probe for Northern hybridization | 5'-CTA TAC AAC CTA CTA CCT CAA-3' |

| | |
|---|--|
| <i>let-7*</i> probe for Northern hybridization | 5'-AAA GCT AGC ACA TTG TAT AGT-3' |
| Forward primer for 558, 345, 208, 106 bp product for sense RNA from pEGFP-N3 (Figure 3G) | 5'-TAA TAC GAC TCA CTA TAG GCA AGC TGA CCC TGA AGT TC-3' |
| Reverse primer for 558 bp product for sense RNA from pEGFP-N3 | 5'-mGmGT CAC GAA CTC CAG CAG GAC-3' |
| Forward primer for 558 bp product for anti-sense RNA from pEGFP-N3 | 5'-TAA TAC GAC TCA CTA TAG GTC ACG AAC TCC AGC AGG AC-3' |
| Reverse primer for 558, 345, 208, 106 bp product for anti-sense RNA from pEGFP-N3 | 5'-mGmGC AAG CTG ACC CTG AAG TTC-3' |
| Reverse primer for 345 bp product for sense RNA from pEGFP-N3 | 5'-mGmGC CAT GAT ATA GAC GTT GTG-3' |
| Forward primer for 345 bp product for anti-sense RNA from pEGFP-N3 | 5'-TAA TAC GAC TCA CTA TAG GCC ATG ATA TAG ACG TTG TG-3' |
| Reverse primer for 208 bp product for sense RNA from pEGFP-N3 | 5'-mGmGG TCT TGT AGT TGC CGT CGT-3' |
| Forward primer for 208 bp product for anti-sense RNA from pEGFP-N3 | 5'-TAA TAC GAC TCA CTA TAG GGT CTT GTA GTT GCC GTC GT-3' |

| | |
|--|--|
| Reverse primer for 106 bp product for sense RNA from pEGFP-N3 | 5'-mGmGT AGC GGC TGA AGC ACT GCA-3' |
| Forward primer for 106 bp product for sense RNA from pEGFP-N3 (Figure 2) | 5'-GCG TAA TAC GAC TCA CTA TAG GGC CAC AAG TTC AGC GTG TCC-3' |
| Reverse primer for 106 bp product for sense RNA from pEGFP-N3 | 5'-mGmGG CCA GGG CAC GGG CAG CTT GCC G-3' |
| Forward primer for 106 bp product for anti-sense RNA (blunt end) from pEGFP-N3 (Figure 2) | 5'-mGmGG CCA CAA GTT CAG CGT GTC C-3' |
| Reverse primer for 106 bp product for anti-sense RNA (blunt end) from pEGFP-N3 (Figure 2) | 5'-GTA CTT AAT ACG ACT CAC TAT AGG GCC AGG GCA CGG GCA GCT TGC CG-3' |
| Forward primer for 104 bp product for anti-sense RNA (3'overhang) from pEGFP-N3 (Figure 2) | 5'-GTA CTT AAT ACG ACT CAC TAT AGC CAG GGC ACG GGC AGC TTG CCG-3' |
| Reverse primer for 104 bp product for sense RNA (3'overhang) from pEGFP-N3 (Figure 2) | 5'-mTmTG GGC CAC AAG TTC AGC GTG TCC-3' |

Table 2.S2. List of Synthetic RNA Oligonucleotides. d, deoxy ribose.

| | |
|---|--|
| 25 nt sense RNA for 25 bp dsRNA | 5'-ACC CUG AAG UUC AUC UGC ACC ACdC dG-3' |
| 27 nt anti-sense RNA for 25 bp dsRNA | 5'-CGG UGG UGC AGA UGA ACU UCA GGG UCA -3' |
| 29 nt sense RNA for 29 bp dsRNA | 5'-ACC CUG AAG UUC AUC UGC ACC GUC CAC dCdG-3' |
| 31 nt anti-sense RNA for 29 bp dsRNA | 5'-CGG UGG ACG GUG CAG AUG AAC UUC AGG GUC A-3' |
| pre- <i>let 7</i> | 5'-UGA GGU AGU AGG UUG UAU AGU AGU AAU UAC ACA UCA UAC UAU ACA AUG UGC UAG CUU UCU-3' |

Table 2.S3. DNA Templates for Transcription and DNA Splints for RNA Ligation, Used in Figures 2.3 and 2.4. m, 2'-O-methyl ribose.

| | |
|--|---|
| DNA template to transcribe antisense RNA strand in Figure 4 | 5'-ACT CCT CAA CAA ATC ATA AAC TAC AAT ATA CAT CAA TAC GAC ATT ACC CTC ACA ATC AAT CAT ACA ACC ATC CCT AAA GAC CAA CAG CAC CCC ACG ATC AAG AAT AAG AAC TAT AAT CCC TAT AGT GAG TCG TAT TAC CC 3' |
| T7 RNAP promoter sequence annealed to the DNA template for transcription | 5'-GCG TAA TAC GAC TCA CTA TAG-3' |
| DNA splint for RNA ligation to generate sense strand in Figure 4 | 5'-GGG ATT ATA GTT CTT ATT CTT GAT CGT GGG GTG CTG TTG GTC TTT AGG GAT GGT TGT ATG ATT GAT TGT GAG GGT AAT GTC GTA TTG ATG TAT ATT GTA GTT TAT GAT TTG TTG AGG AGT-3' |
| DNA template to transcribe 40 nt sense RNA in Figure 3G | 5'-mGGC TTC ATG TGG TCG GGG TAG CGG CTG AAG CAC TGC ACG CCT ATA GTG AGT CGT ATT ACG C-3' |
| DNA template to transcribe 40 nt anti-sense RNA in Figure 3G | 5'-mGGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CCT ATA GTG AGT CGT ATT ACG C-3' |
| DNA template to transcribe 60 nt sense RNA in Figure 3G | 5'-mGGT GAA CTT CAG GGT CAG CTT GCT TCA TGT GGT CGG GGT AGC GGC TGA AGC ACT GCA CGC CTA TAG TGA GTC GTA TTA CGC-3' |
| DNA template to transcribe 60 nt anti-sense RNA in Figure 3G | 5'-mGGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAA GCT GAC CCT GAA GTT CAC CTA TAG TGA GTC GTA TTA CGC-3' |

Table 2.S4. List of Synthetic RNA Oligonucleotides Assembled by Splinted Ligation into Site-Specifically Labeled Sense-Strand RNA and Used in Figure 2.4, 2.5 and S4.

| | |
|--------|--|
| RNA #1 | 5'-ACU CCU CAA CAA AUC A-3' |
| RNA #2 | 5'-UAA ACU ACA AUA UAC AUC AAU ACG ACA UUA CCC UCA CAA UCA AUC AUA CAA CCA UCC CUA AAG ACC AAC AG CAC CCC A-3' |
| RNA #3 | 5'-CGA UCA AGA AUA AGA ACU AUA AUC dCdC-3' |
| RNA #4 | 5'-ACU CCU CAA CAA AUC AUA AAC UAC AAU AUA CAU CA-3' |
| RNA #5 | 5'-AUA CGA CAU UAC CCU CAC AAU CAA UCA UAC AAC CAU CCC UAA AGA CCA ACA GCA CCC CA-3' |
| RNA #6 | 5'-UCC UCA ACA AAU CA-3' |
| RNA #7 | 5'-UCC UCA ACA AAU CAU AAA CUA CAA UAU ACA UCA-3' |

Discussion

Purified Dicer-1 and Dicer-2 both process pre-miRNAs, but generate different length products (22 versus 21 nt). Genetic analyses suggest that Dicer-1 and Dicer-2 are restricted to specific substrate classes in vivo (Lee et al. 2004). For example, Dicer-2 cannot replace Dicer-1 in the miRNA pathway. Similarly, dicer-2 mutants are defective for RNAi, even though they express normal levels of Dicer-1 (Lee et al., 2004). Despite structural similarities, Dicer-2 specifically processes esiRNA hairpins, while Dicer-1 cleaves pre-miRNAs (Lee et al., 2004; Förstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005; Miyoshi et al., 2010). This observation suggests that the length of a dsRNA is the primary determinant of substrate choice.

Our data argue that the combination of R2D2 and cellular phosphate restricts Dicer-2 to its biologically relevant substrates by inhibiting the processing of short substrates such as pre-miRNA. Thus, a protein, R2D2, and a small molecule, phosphate, convert a promiscuous dsRNA endonuclease into one specific for the long dsRNA substrates that trigger RNAi. It is tempting to speculate that inorganic phosphate interferes with recognition of the 5' monophosphate present on all pre-miRNAs, and that 5' phosphate recognition is unnecessary for longer substrates, because their greater length allows additional protein-RNA contacts—perhaps by the dsRNA-binding and the helicase domains—between Dicer-2 and long dsRNA. We note that human Dicer has

been reported to recognize a 5' monophosphate on single-stranded RNA (Kini and Walton, 2007).

In flies, the Dicer-2 partner proteins Loqs-PD and R2D2 likely enhance substrate specificity by increasing the affinity of the enzyme for long dsRNA. The k_{cat} for and Dicer-2 and Dicer-2/R2D2 were similar, but R2D2 decreased the K_M of Dicer-2 for long dsRNA (Table 2.1 and Figure 2.S2). We note that the specificity constant, k_{cat}/K_M , was ~4 fold higher for the Dicer-2/R2D2 heterodimer than for Dicer-2 alone. Similarly, Loqs-PD lowered the K_M of Dicer-2 for long dsRNA without reducing the catalytic rate, resulting in a ~10-fold higher k_{cat}/K_M .

Processing of pre-miRNA by Dicer-1 was unaffected by phosphate. We find that the intrinsic properties of Dicer-1, which cannot efficiently catalyze multiple-turnover processing of long dsRNA, restrict that enzyme to process pre-miRNA. We do not yet know whether the transition of substrates from Dicer-1 to Dicer-2 is gradual, such that some substrates are processed equally well by both enzymes. In theory, such intermediate substrates might be selected against in evolution, enforcing the distinction between Dicer-1 and Dicer-2 substrates.

The Dicer-2 helicase domain is similar to that of RIG-I, a sensor in the mammalian innate immune system. The RIG-like ATPase/helicase domain is conserved among plant and animal Dicercs. Yet, its function has remained unknown. Our data suggest that this domain of Dicer-2 is involved in ATP-dependent production of successive siRNAs from long dsRNA. Notably, two other members of this helicase family, DRH-3 and RIG-I are also bona fide

ATPases: DRH-3, a *C. elegans* protein required for RNA silencing and germ-line development (Nakamura et al., 2007), is a dsRNA-stimulated ATPase (Matranga and Pyle, 2010); and the mammalian protein RIG-I, which recognizes viral 5' triphosphorylated dsRNA and initiates an innate immune response, uses ATP to translocate along dsRNA (Myong et al., 2009). Our data are consistent with the idea that ATP hydrolysis fuels translocation of Dicer-2 along long dsRNA substrates. An alternative view—that the ATP-dependent binding of a molecule of Dicer at the end of the substrate promotes the complete and rapid oligomerization of Dicer-2 along the entire extent of the dsRNA—would require that the Dicer and RIG-I helicase domains share a conserved sequence but have highly divergent functions.

ATP was not required for Dicer-2 to process pre-miRNA, and a mutant Dicer-2 unable to hydrolyze ATP remained able to process pre-miRNA but not long dsRNA. These results help explain why in *C. elegans*, in which a single Dicer processes both long dsRNA and pre-miRNA, a mutation in the DCR-1 helicase domain disrupted endo-siRNA, but not miRNA, accumulation (Welker et al., 2010).

Four lines of evidence support a role for ATP hydrolysis in the production of successive siRNAs along the dsRNA by Dicer-2. First, Dicer-2 consumes a constant amount of ATP per base-pair. Second, ~23 molecules of ATP were consumed for each 21 nt siRNA produced. Third, the rate of production of the first, second, and fourth siRNAs from a long dsRNA substrate were

indistinguishable in the presence of ATP, but in the absence of ATP, the rate of siRNA production declined with increasing distance from the end of the dsRNA. Finally, the association of Dicer-2 with a long dsRNA was resistant to dilution provided ATP was present, suggesting that after binding the end of its substrate, Dicer-2 remains bound to the dsRNA and uses ATP energy to reposition itself to produce the next 21 bp siRNA. Translocation along the dsRNA seems a likely mechanism.

Although helicase mutant Dicer-2^{G31R} processed pre-miRNAs as efficiently as wild-type Dicer-2, the mutant was unable to produce even the terminal siRNA from a long RNA duplex under multiple turnover conditions. This suggests an inherent ability of the helicase domain of Dicer-2 to distinguish between long and short substrates. We note that the helicase domain of human Dicer auto-inhibits processing of an RNA duplex, and its dsRNA-binding protein partner TRBP, a homolog of R2D2 and Loqs, relieves this inhibition (Ma et al., 2008; Chakravarthy et al., 2010). We hypothesize that *Drosophila* Dicer-2 can occupy two distinct conformations. When inorganic phosphate is low, Dicer-2 assumes a conformation—perhaps similar to the auto-inhibited conformation of human Dicer—that can bind and load siRNA. This conformation is unaffected by ATP and, we presume, is involved in promiscuously processing pre-miRNA in vitro. When inorganic phosphate is higher and the enzyme's helicase and dsRNA-binding domains engage its substrate, Dicer-2 assumes a conformation that requires ATP for binding and hydrolysis to process dsRNA.

Materials and Methods

Protein Expression and Purification

Expression and purification of His₆-Dicer-2 or His₆-Dicer-2 and His₆-R2D2 in Sf21 cells was as described (Liu et al., 2003; Tomari et al., 2004). His₆-Dicer-2^{G31R}, His₆-Dicer-2^{D1217,1614N} and His₆-Dicer-1 were expressed in Sf9 insect cells using the BAC-to-BAC Baculovirus Expression System (Invitrogen, Carlsbad, CA) and purified from cell lysates by using Ni-NTA agarose (QIAGEN, Valencia, CA), HiTrap Q, HiTrap Heparin (GE Healthcare, Pittsburgh, PA), and Superdex 200 gel filtration. Loqs-PD was expressed in Escherichia coli Rosetta2(DE3), isolated using Ni-Sepharose (GE Healthcare), treated with HRV3C protease cleavage to remove the His-tag, and purified using HiTrap SP and HiTrap Heparin. Proteins were exchanged into 20 mM HEPES-KOH (pH 8.0), 100 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride. Protein concentrations were determined by quantitative amino acid analysis (Keck Biotechnology Resource Laboratory, New Haven, CT). Two different preparations of recombinant Dicer-2 were used in this study. Preparation 1 was used for Table 2.1 and Figures 2.2, 2.3E, 2.4C, 2.S2B, 2.S3A, 2.S3D, and 2.S2. Preparation 2 was used for Table 2.2 and Figures 2.1, 2.3A–D, 2.3F, 2.3G, 2.4A, 2.4B, 2.5, 2.S1, 2.S2A, 2.S3B, 2.S3C, and 2.S4.

RNA Substrates

dsRNAs were prepared as described (Haley et al., 2003). PCR templates for transcription of sense and antisense RNAs were generated from the EGFP sequence of pN3-eGFP using primers listed in Table 2.S1. Twenty-five and 29 bp dsRNAs (Figure 2.S2A, 2.S3B, 2.S4A, Table 2.S2) were as previously described (Rose et al., 2005). Synthetic RNAs and synthetic *Drosophila* pre-let-7 (Dharmacon, Lafayette, CO) were 5' ³²P-radiolabeled using γ -³²P ATP (6000 Ci/mmol; PerkinElmer, Waltham, MA) and T4 polynucleotide kinase (NEB, Ipswich, MA). After gel purification, RNA strands or pre-let-7 were incubated at 65°C for 5 min and then at 25°C for 30 min. Site-specifically radiolabeled 120 nt dsRNAs were prepared by DNA-splinted ligation (Table 2.S3 and 2.S4) (Moore and Sharp, 1993; Moore and Query, 2000). To monitor formation of the fourth siRNA in Figure 2.4B, we used a 120 nt substrate in which the 5th siRNA (the last siRNA generated by Dicer-2 from this substrate) was site-specifically ³²P-radiolabeled. The fourth siRNA produced corresponds to the sum of the two ³²P-radiolabeled cleavage products produced when the dsRNA was cleaved to generate the fourth and fifth siRNAs.

In Vitro RNA Processing

For Dcr-2 +Loqs-PD, Loqs-PD was first mixed with Dicer-2 or Dicer-2/R2D2 and incubated 10 min on ice followed by 5 min at room temperature. Dicing reactions contained 7.5 mM DTT, 3.3 mM magnesium acetate, 0.25% v/v glycerol, 100 mM potassium acetate, 18 mM HEPES-KOH (pH 7.4), 15 mM CP,

2.25 μg CK, and 1 mM ATP; –ATP reactions contained 1 mM EDTA but no CK or CP; ATP γ S reactions contained 1 mM ATP γ S only. In Figure 2.2A-D, 2.3, 2.4B-E, 2.5, 2.S3, and 2.S4, +ATP reactions contained no CP or CK; +CP and +CK reactions in Figure 2B and S2B contained 20 mM CP or 2.25 μg CK. Reactions were assembled on ice and pre-incubated at 25°C for 5 min before adding RNA. In Figure S4, dilution buffer contained 0.1% NP-40.

Aliquots (1 μl) of reactions with radiolabeled RNA substrate were quenched by the addition of 25 volumes of formamide loading buffer (98% v/v formamide, 0.1% w/v bromophenol blue and xylene cyanol, 10 mM EDTA), incubated 5 min at 95°C, and analyzed by electrophoresis through a denaturing polyacrylamide 7 M urea gel using 0.5 \times Tris-borate-EDTA buffer (National Diagnostics, Atlanta, GA). In Figure 5, 200 μL aliquots from dilution reactions were stopped with 300 mM sodium acetate and 25 mM EDTA, isopropanol precipitated, and dissolved in formamide loading buffer before gel analysis. Gels were exposed to image plates and analyzed with an FLA-5000 and ImageGauge 3.0 software (Fujifilm, Tokyo, Japan). In Figure 1, let-7 and let-7* strands were detected by Northern hybridization with 5' ^{32}P -radiolabeled DNA probes (Table S1).

ATP Hydrolysis

α - ^{32}P -ATP (250 nM, 3000 mmol/Ci; PerkinElmer) was used to monitor hydrolysis. Reactions were stopped with a 25 vol formamide loading dye and spotted onto 20 cm \times 20 cm cellulose plates (EMD, Darmstadt, Germany) and

chromatographed in 0.75 M KH_2PO_4 (adjusted to pH 3.3 with H_3PO_4) until the solvent reached the top of the plate. The plate was dried and analyzed by phosphorimager. For Figure 2.S3B, ATP hydrolysis was monitored using the ATP Bioluminescent Assay Kit (Sigma, St. Louis, MO). The reaction was stopped by diluting the sample 10 times in H_2O and immediately flash-freezing in liquid nitrogen. Samples were stored at -80°C until they were measured. A standard curve spanning at least 100-fold less than and greater than the experimental values was used to determine ATP concentrations.

Rate Analyses

Substrate converted to siRNA versus time was fit to $y = y_0 + A(1 - e^{-kt})$, where $dy/dt = Ake^{-kt}$. When $t = 0$, $dy/dt = Ak$ (Lu and Fei, 2003). Data were fit to the Michaelis-Menten scheme using Visual Enzymics 2008 (Softzynamics, Princeton, NJ) for Igor Pro 6.11 (WaveMetrics, Lake Oswego, OR). The rate of substrate consumption was fit for each replicate separately, and two-tailed, two sample equal variance t-test used to compare rates (Excel, Microsoft, Seattle, WA). R 2.6.0 software was used for other statistical analyses.

Acknowledgments

We thank Yukihide Tomari for help purifying recombinant Dicer-2, members of the Zamore and Hall laboratories and Can Cenik for advice and comments on the manuscript, and Sean Ryder for suggesting we test the effect of phosphate on dicing of pre-miRNA.

CHAPTER III: PROCESSING OF PRE-MIRNAS BY *DROSOPHILA***DICER-1**

Preface

This work was a collaborative effort between Dr. Phillip D. Zamore's and Dr. Traci M. Tanaka Hall's laboratories. Dr. Gang Lu cloned, recombinantly expressed and purified Dicer-1 and Dicer-1/LoqsPB complex. I carried out dicing and ATP consumption assays in Figures 3.1, 3.S2, Figure 3.2, and Table 3.1.

Abstract

Drosophila Dicer-1 is responsible for generating miRNAs from pre-miRNAs. Despite structural similarities to endo-siRNAs and other double stranded RNA, pre-miRNAs are processed in a very specific manner by Dicer-1 and not Dicer-2. What confers this specificity is unknown. Here we report that a partner protein of Dicer-1, Loqs-PB, and ATP enhance substrate specificity of Dicer-1. More specifically, ATP and Loqs-B likely facilitate efficient processing of the miRNA/miRNA* duplex by Dicer-1 by lowering the K_M . Our data highlight the role of ATP and regulatory dsRNA-binding partner proteins to achieve substrate specificity in *Drosophila* RNA silencing.

Introduction

Both siRNAs and miRNAs rely on members of the Dicer family of double-stranded (ds) RNA-specific RNase III endonucleases for their processing from longer RNA precursors (Bernstein et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). In *Drosophila melanogaster*, two separate pathways produce siRNAs and miRNAs. Primary miRNA precursors are transcribed by Pol II and contain one or a few miRNA precursors (pre-miRNAs) (Lee et al., 2002; Cai et al., 2004; Lee et al., 2004a). The RNase III enzyme Drosha, in collaboration with its dsRNA-binding domain partner protein, Pasha, cleaves primary miRNAs to release pre-miRNAs, ~70 nt long stem-loop structures that contain the mature miRNA within this structure (Lee et al., 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Han et al., 2006). Subsequently, pre-miRNAs are exported from the nucleus to the cytoplasm by Exportin 5 (Yi et al., 2003; Bohnsack et al., 2004; Lund et al., 2004). In the cytoplasm, the pre-miRNAs are cleaved by Dicer-1, which acts in concert with its dsRNA-binding protein partner, Loquacious-PB (Loqs-PB). This processing step liberates a duplex comprising the mature miRNA bound to its miRNA*, a partially complementary small RNA derived from the opposite arm of the pre-miRNA stem (Förstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005; Ye et al., 2007). Different mature miRNAs derive from either the 5' or 3' arm of the pre-miRNA stem.

Drosophila Dicer-1 contains two RNase III domains that form an

intramolecular heterodimer whose dimer interface creates two active sites (Zhang et al., 2004; Macrae et al., 2006; Ye et al., 2007). Each active site positions two Mg^{2+} ions that catalyze phosphodiester bond hydrolysis on one or the other strand of the dsRNA. Dicer-1 generates a dsRNA terminus bearing a two-nucleotide 3' overhang characteristic of all RNase III enzymes (Bass, 2000; Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006). Like other members of the Dicer family, Dicer-1 contains a C-terminal dsRBD and a central PAZ domain, which has an RNA-binding motif that can recognize the two-nucleotide, 3' single-stranded ends of Drosha and Dicer products (Cerutti et al., 2000; Lingel et al., 2003; Song et al., 2003; Yan et al., 2003; Lingel et al., 2004; Ma et al., 2004; Macrae et al., 2006; MacRae et al., 2007). The crystal structure of a simple Dicer protein from *Giardia intestinalis*, together with functional studies using human Dicer, suggest that the distance between the PAZ domain and the active sites of the RNase III domains establishes the length of the small RNA product (Zamore, 2001; Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006; MacRae et al., 2007; Takeshita et al., 2007).

Here, we present evidence that Dicer-1 requires ATP to efficiently process its substrates. ATP facilitates efficient processing of pre-miRNAs when in excess compared to Dicer-1, as would be expected in vivo. ATP modifies the properties of Dicer-1 by decreasing its K_M for pre-miRNA. Its partner protein Loqs-PB, like ATP, decreases the K_M of Dicer-1 for pre-miRNA. Thus, the characteristic substrate specificity of Dicer-1 arise both from its intrinsic properties, and its

interactions with a small molecule—ATP—and its cognate dsRNA-binding partner protein.

Results

ATP enhances processing of pre-let-7 by Dicer-1

Previous studies concluded that Dicer-1 requires no ATP for pre-miRNA processing (Jiang et al., 2005), yet it wasn't clear whether the substrate and enzyme concentrations were close to intracellular levels (pre-miRNA = 4×10^4 cpm, enzyme = not specified). Since the Helicase C domain of Dicer-1 contains an ATP-binding domain, we asked whether ATP binding or hydrolysis had any effect in pre-miRNA processing by Dicer-1 (Figure 3.S1B).

We analyzed pre-let-7 processing using purified recombinant Dicer-1 and Dicer-1/Loqs-PB in ATP presence or absence. Consistent with previous single-turnover analyses, ATP was dispensable for processing of pre-let-7 when the concentration of Dicer-1 or Dicer-1/Loqs-PB was greater than pre-let-7. However, when pre-let-7 substrate concentration was much greater than that of Dicer-1, ATP enhanced pre-let-7 processing for both Dicer-1 and Dicer-1/Loqs-PB (Figure 3.1A, right). This multiple-turnover reaction condition likely reflects the intracellular environment in which Dicer-1 functions. Surprisingly, ATP had no statistically significant effect on the rate of pre-miRNA processing by Dicer-1 at saturating substrate concentrations (Table 3.1), rather, it decreased the K_M (p-value = 0.02).

T. Thermophilus Hb8 UvrB Chain A
D. melanogaster Dicer-1
 HCV Ns3 helicase Chain A
 HCV Ns3 helicase Chain B
S. cerevisiae MAK5 ATP dependent helic
E. coli RhIE ATP dependent helicase
 Mouse MVH (Vasa homolog)
B. subtilis dbpA ATP dependent RNA helicase
S. pombe dbp10 ATP dependent RNA helicase
S. cerevisiae RRP3 ATP dependent rRNA helicase

473 . [5] .KRQA. [4] .LRLG. [2] .DCLVGLNLLRKGDDIP. [1] .VSLVAILD. [7] .RSERSLIQTIGRA. [1] .RNA 540
 545 . [5] .ROEF. [4] .FRMH. [2] .NYLIGTSVLEEGIDVF. [1] .CINLVVRMB. [2] .TTYRSYVQCKGBA .BAA 606
 229 . [2] .DVSV. [3] .PTSG. [1] .VVVVAEDALMTGFTGD FDSVIDCN. [23] .QDAVSRTOFRGRT. [1] .RGK 304
 408 . [2] .DVSV. [1] .PTIG. [1] .VVVVATDALTGFTGD FDSVIDCN. [23] .QDAVSRSGRGRRT. [1] .RGR 483
 496 . [5] .NRLK. [4] .FKOO. [18] .TVLIIASLSVAARGLDIP. [1] .VQHVHYH. [2] .KSTLIYIHRSGRT. [1] .RAG 574
 278 . [5] .ARTE. [4] .FKSG. [2] .RVLVATDIAARGLDIE. [1] .LPHVWVYE. [2] .NVPEDYVHRIGRT. [1] .RPR 340
 487 . [5] .REQ. [4] .FRCG. [2] .PVLVATSVAAARGLDIE. [1] .VQHVINFD. [2] .STIDYVHRIGRT. [1] .RCG 549
 273 . [5] .DRFD. [4] .FRRG. [2] .RYLVATDVAARGLDIE. [1] .ISLVINVD. [2] .LEKESYVHRIGRT. [1] .RAG 93
 379 . [5] .ARLN. [4] .FRUG. [2] .NLIWVVTIVASRGIDIP. [1] .LANVINVD. [2] .PQPRYFVHRVGRRT. [1] .RAG 441
 396 . [5] .QRMG. [4] .EKAG. [2] .SILVATDVAARGLDIP. [1] .VDIYVNVVD. [2] .VISRSYIHRVGRRT. [1] .RAG 458

Figure 3.S1. Alignment of the Predicted ATP-Binding Site of Dicer-1 with Known or Putative ATPase Proteins. The alignment is from the NCBI conserved domain database. An asterisk denotes residues predicted to be at the ATP-binding site of Dicer-1, based on the crystal structure of the UvrB protein bound to ATP (Theis et al., 1999; Marchler-Bauer et al., 2009)

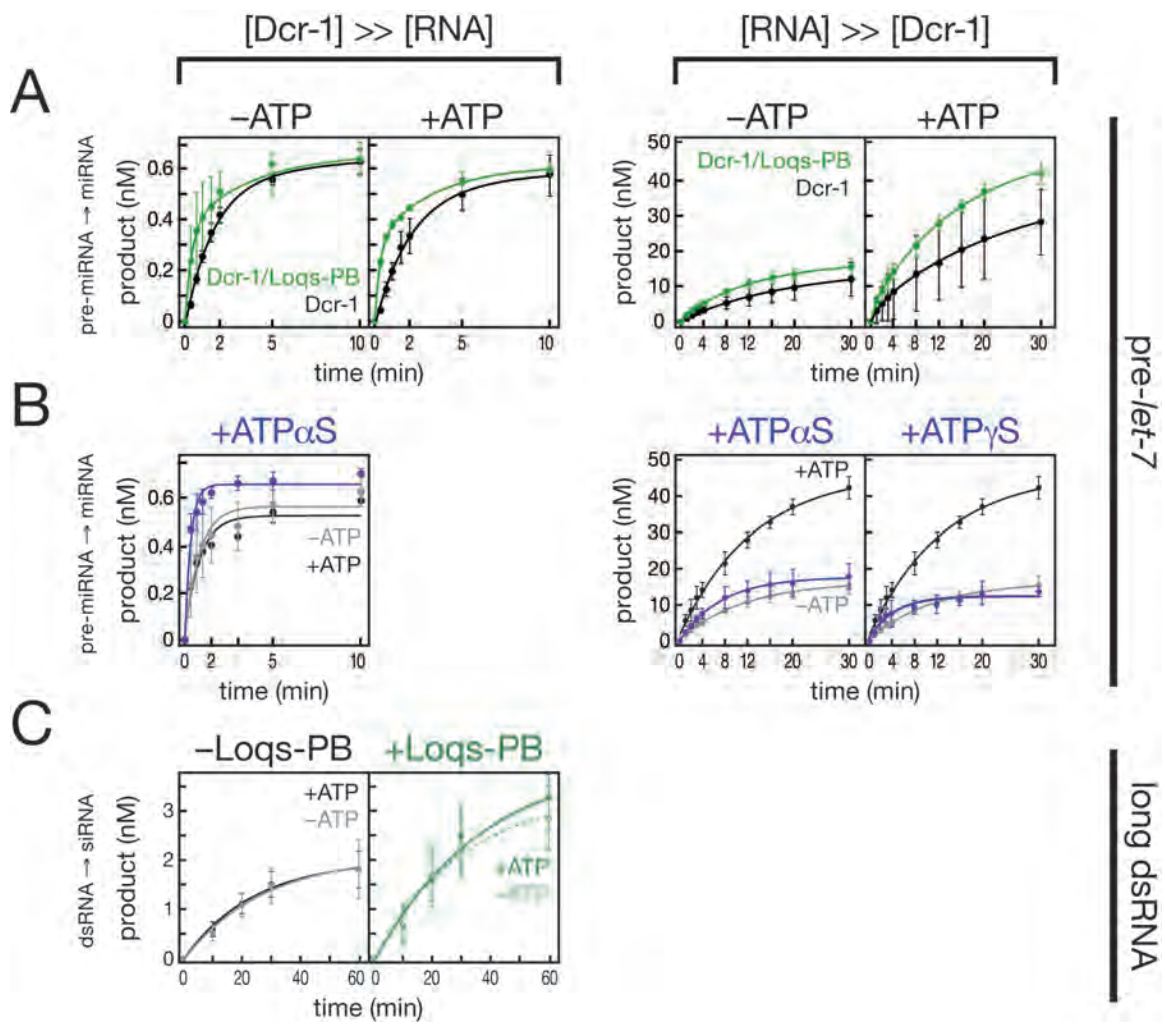


Figure 3.1. ATP and Loqs-PB Increase Multiple-Turnover Pre-miRNA

Processing by Dicer-1. (A) Production of *let-7* from a 5 ³²P-radiolabeled pre-*let-7* substrate by Dicer-1 (black) or Dicer-1/Loqs (green) was measured with or without 1 mM ATP under single- (left; 8 nM enzyme, 1 nM substrate) or multiple-turnover conditions (right; 2 nM enzyme, 150 nM substrate). (B) Pre-*let-7* processing by Dicer-1/Loqs-PB was monitored under single- (left) or multiple-turnover (right) conditions in the presence (black) or absence (gray) of ATP or in the presence of the ATP analogs (blue) ATP α S or ATP γ S. Values correspond to the mean \pm standard deviation from three independent experiments. The data were fit to a single exponential function. (C) siRNA production from a 517 bp dsRNA (25 nM) by Dicer-1 or Dicer-1/Loqs-PB (235 nM) was monitored in the presence or absence of ATP. Left, the amount of siRNA generated by Dicer-1 alone was measured in the presence (black) or absence (gray) of ATP. Right, the amount of siRNA generated by Dicer-1/Loqs-PB was measured in the presence (solid line) or absence (dashed line) of ATP.

Although ATP is required for efficient pre-miRNA processing in multi-turnover conditions, using the same thin-layer chromatography assay we were unable to detect ATP hydrolysis above background by Dicer-1 or Dicer-1/Loqs PB in pre-let-7 substrate presence or absence (Figure 3.S2A). ATP analogs modified at the β – γ bond did not affect multiple rounds of pre-miRNA processing by Dicer-1/Loqs PB (Figure 3.1B). Single-turnover cleavage of pre-let-7 by Dicer-1 required no ATP. Our results are suggestive of ATP promoting an alternative, higher affinity conformation for Dicer-1.

Dicer-1 Does Not Process Long dsRNA Efficiently

Loqs-PB has been proposed to prevent Dicer-1 from processing long dsRNA, thereby restricting siRNA production to Dicer-2 (Saito et al., 2005). However, we found that Dicer-1 alone did not efficiently process a 517 bp dsRNA, nor did Loqs-PB inhibit the low level of siRNA production by Dicer-1 (Figure 3.1C). In fact, we could only detect siRNA production from a 517 bp dsRNA when Dicer-1 was in ~10-fold excess compared to the dsRNA substrate (225 nM Dicer-1, 25 nM long dsRNA), (Figure 3.1C). Moreover, we were unable to detect siRNA production by Dicer-1 (data not shown) using the same multiple-turnover conditions (25 nM dsRNA and 2 nM Dicer) under which Dicer-2 efficiently generated siRNAs. Unlike Dicer-2, processing of long dsRNA by Dicer-1 or Dicer-1/ Loqs-PB was distributive, with intermediate products readily detected (Figure 3.2).

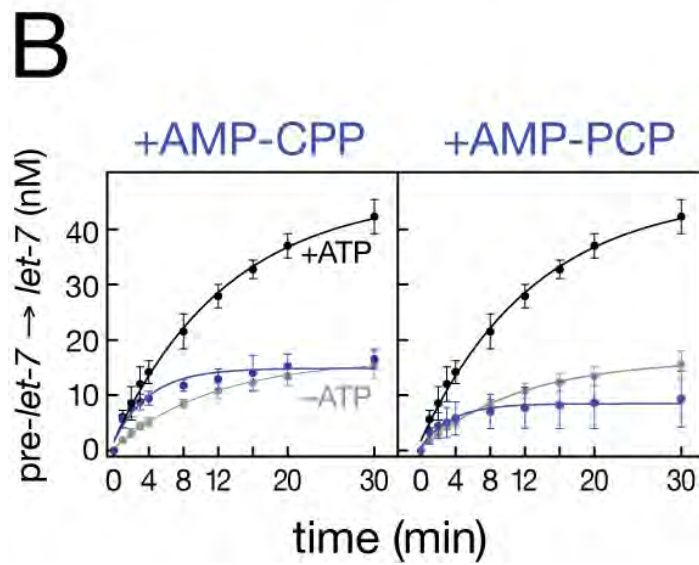
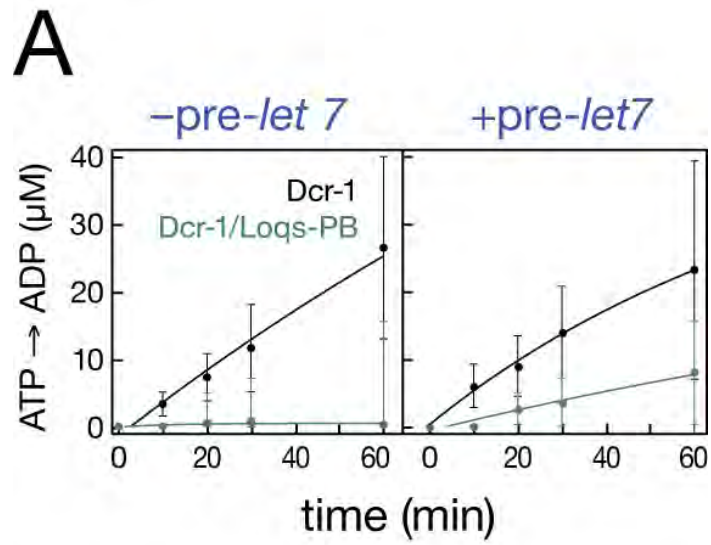


Figure 3.S2. Non-hydrolyzable β - γ ATP Analogs do not Replace ATP for Pre-miRNA Processing by Dicer-1. (A) ATP (1 mM) hydrolysis to ADP was measured for 450 nM Dicer-1 or Dicer-1/Loqs-PB in the presence and absence of 2 μ M pre-*let-7* substrate. (B) pre-*let-7* (150 nM) processing by Dicer-1/Loqs-PB (2 nM) was measured in the presence of the non-hydrolyzable ATP analogs, AMPPcP or AMPcPP (1mM) under conditions of substrate excess. For comparison, pre-*let-7* processing by Dicer-1/Loqs-PB in the absence (gray) or presence of ATP (black) is shown.

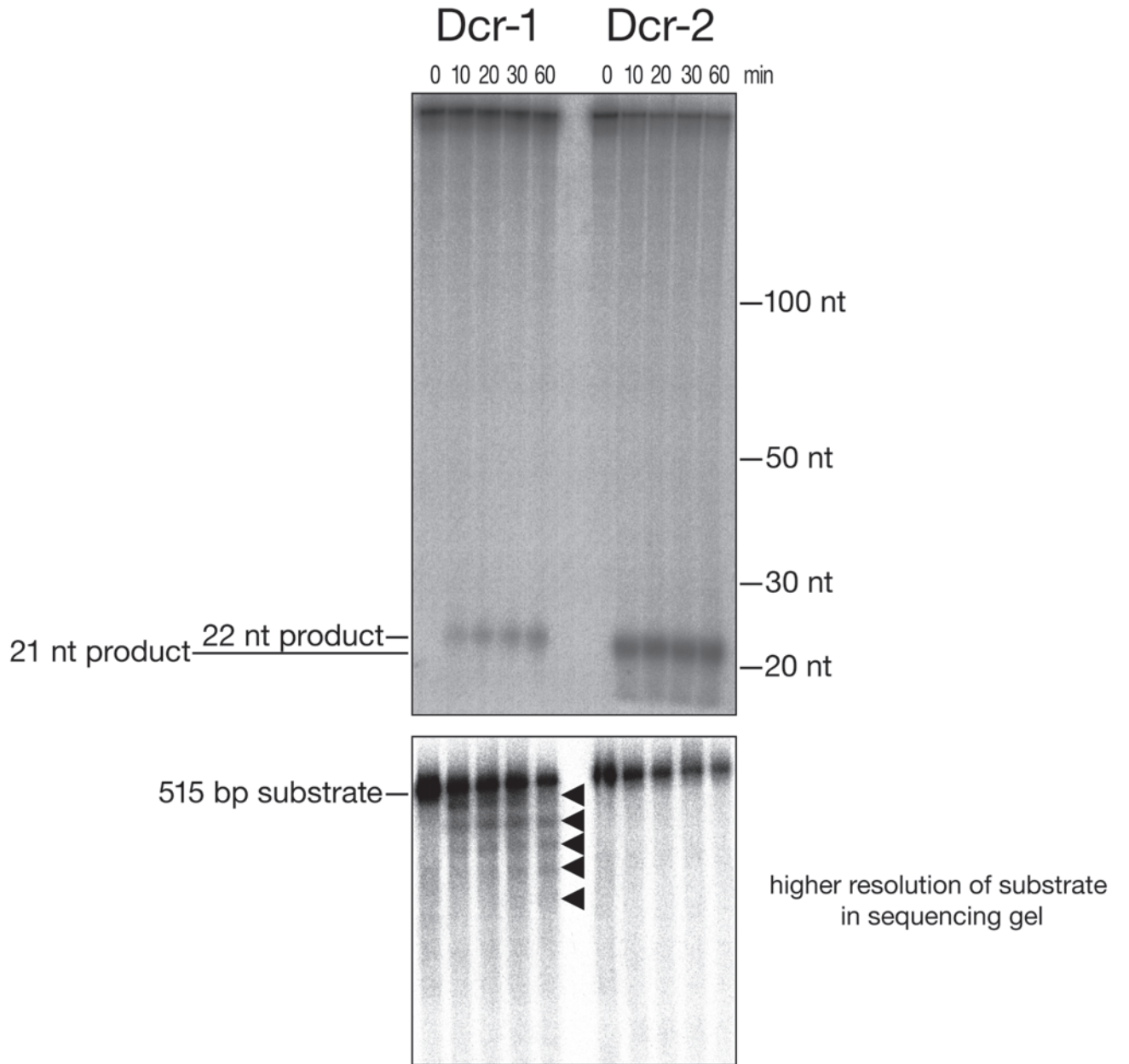


Figure 3.2. Dicer-1, but not Dicer-2, Generates Intermediates when

Processing Long dsRNA. 25 nM internally ^{32}P -radiolabeled 515 bp dsRNA was incubated with either 225 nM Dicer-1 or 5.4 nM Dicer-2. The top gel was run to visualize both the uncleaved substrate and the 21–22 bp siRNA products; the lower gel was run to resolve the intermediates (arrowheads) generated by Dicer-1.

Loqs-PB Decreases the K_M of Dicer-1 for Pre-let-7

Although Loqs-PB does not alter the affinity of Dicer-1 for long dsRNA, it might be expected to increase the affinity of Dicer-1 for its authentic substrate, pre-miRNA. Michaelis-Menten analyses of pre-let-7 processing by Dicer-1 and Dicer-1/Loqs-PB revealed a small increase in catalytic rate when Loqs-PB was present ($k_{cat} = 2.4 \pm 1.1 \text{ min}^{-1}$ for Dicer-1 versus $1.1 \pm 0.9 \text{ min}^{-1}$ for Dicer-1/Loqs-PB; p -value = 0.01) and a large decrease in K_M ($44.3 \pm 32.3 \text{ nM}$ for Dicer-1 versus $4.4 \pm 3.5 \text{ nM}$ for Dicer/Loqs-PB) in the presence of ATP (Table 3.1). The effect of ATP on the kinetics of pre-miRNA processing was similar for Dicer-1/Loqs-PB and Dicer-1 alone. For Dicer-1/Loqs-PB, ATP had no statistically significant effect on k_{cat} , but significantly decreased the K_M (Table 3.1, p -value = 0.008). Overall, the combination of ATP and Loqs-PB decreased the K_M for Dicer-1 by 25-fold and increased its specificity constant, k_{cat}/K_M by 75-fold.

Table 3.1. Michaelis-Menten Analysis of Dicer-1.

Purified recombinant Dicer-1 was incubated with pre-*let-7* and the production of *let-7* miRNA measured.

| Dicer-1 plus: | ATP | K_M (nM) | Change in K_M | k_{cat} (min^{-1}) | Change in k_{cat} | k_{cat} / K_M ($\text{nM}^{-1} \text{min}^{-1}$) | Change in k_{cat} / K_M |
|---------------|-----|------------------|-----------------|---------------------------------|---------------------|--|---------------------------|
| \emptyset | - | 116.5 ± 13.7 | 1 | 1.0 ± 0.5 | 1 | 0.008 ± 0.003 | 1 |
| | + | 44.3 ± 32.3 | 0.4 | 1.1 ± 0.9 | 1.1 | 0.04 ± 0.03 | 5 |
| Loqs-PB | - | 73.3 ± 12.8 | 0.6 | 3.1 ± 0.5 | 3.1 | 0.04 ± 0.009 | 5 |
| | + | 4.4 ± 3.5 | 0.04 | 2.4 ± 1.1 | 2.4 | 0.6 ± 0.2 | 75 |

Discussion

Our data suggest a model in which ATP and dsRNA binding partners channel Dicer-1 to its biologically appropriate substrate class. To the best of our knowledge, the use of a nucleotide cofactor to ensure biologically appropriate substrate specificity in an RNA silencing pathway is without precedent. Loqs-PB increases the efficiency of pre-miRNA processing by Dicer-1 most likely by increasing its affinity for pre-miRNA.

Genetic data suggested that Dicer-1 and Dicer-2 are restricted to specific substrate classes in vivo. Dicer-2 cannot replace Dicer-1 in the miRNA pathway, and dicer-2 mutants are defective for RNAi, even when Dicer-1 is expressed at normal levels. Despite structural similarity between endo siRNA hairpins, and pre-miRNAs, *Drosophila* Dicers can effectively discriminate these substrates suggesting that the length of a dsRNA is the primary determinant for substrate choice. In the presence of ATP and dsRNA-binding domain partner proteins, longer precursors are processed by Dicer-2, while shorter ones are processed more efficiently by Dicer-1. We do not yet know if the transition from Dicer-1 substrate to Dicer-2 substrate is gradual; with some substrates processed equally well by each enzyme. Nor do we know if biologically relevant precursor RNAs exist that are cleaved by both Dicer-1 and Dicer-2. Evolution might have selected against such dual substrates, ensuring the production of precise small RNA products.

While attractive, partner proteins specifying substrate choice for *Drosophila* Dicers is not entirely consistent with our data. Specifically, we did not observe a role for Loqs-PB in restricting Dicer-1 to the miRNA pathway. Instead, Loqs-PB appears to increase the affinity of Dicer-1 for pre-miRNA (i.e., it decreased the K_M).

The ATPase/helicase domain scaffold of Dicer is widely conserved among plants and animals, including in human Dicer. Yet, detailed functional analysis of the domain has yet to be carried out. The helicase domains of Dicers might potentially provide an interface for protein-protein interactions. Supporting this possibility, *Drosophila* Dicer-2' ATPase/helicase domain interacts with innate immunity pathway proteins, promoting an immune response upon entry of an RNA virus (Wang et al., 2006; Deddouche et al., 2008). Our data suggest that ATP binding or hydrolysis catalyzed by ATPase/helicase domains may also enhance the catalytic function of Dicers, by one or more of several possible mechanisms including improved substrate binding, enhanced catalytic efficiency, and promotion of processivity. In this regard, it is striking that the dsRNA-binding partner protein for human Dicer, TRBP, binds Dicer through its ATPase/helicase domain (Ma et al., 2008; Daniels et al., 2009).

Materials and Methods

Protein Expression and Purification

Dicer-1 protein was expressed from a recombinant baculovirus containing a Dicer-1 cDNA isolated from *Drosophila* S2 cells and inserted into the pFastBac1 vector (Invitrogen, Carlsbad, CA, USA). The sequence of this cDNA corresponds to the genomic sequence of S2 cells and some Oregon R fly stocks, but bears 9 polymorphic differences from the FlyBase reference genome. A His₆ tag was added to the amino-terminus of Dicer-1 to facilitate purification. Similarly, a cDNA encoding full-length *Drosophila* Loqs-PB (Förstemann et al., 2005) was inserted into pFastBac1 with an N-terminal His₆ tag. His-Dicer-1 alone or together with His-Loqs-PB was expressed in Sf9 insect cells using the BAC-to-BAC Baculovirus Expression System (Invitrogen) and purified from cell lysates by successive chromatography using Ni-NTA agarose (QIAGEN), HiTrap Q (GE Healthcare, Pittsburgh, PA, USA), HiTrap Heparin (GE Healthcare), and Superdex 200 gel filtration. The final purified proteins were exchanged into 20 mM HEPES-KOH (pH 8.0), 100 mM NaCl, and 1 mM tris(2-carboxyethyl)phosphine hydrochloride.

cDNAs encoding each Loqs isoform were generated by RT-PCR of fly total RNA and inserted in between the Sall and NotI restriction sites of a modified pCold I vector (Takara, Otsu, Shiga, Japan) containing the HRV3C protease recognition site. Amino-terminally His-tagged Loqs proteins were expressed in Rosetta2(DE3) *Escherichia coli* cells. The proteins were purified using Ni-

Sepharose (GE Healthcare), followed by HRV3C protease cleavage to remove the His-tag, and then further purified using HiTrap SP and HiTrap Heparin (GE Healthcare).

RNA Substrates

Synthetic pre-*let-7* (Dharmacon, Lafayette, CO, USA) (Sequence: 5'-UGA GGU AGU AGG UUG UAU AGU AGU AAU UAC ACA UCA UAC UAU ACA AUG UGC UAG CUU UCU-3') were 5' radiolabeled using γ -³²P ATP (6000 Ci/mmol; Perkin Elmer) and T4 polynucleotide kinase (NEB, Ipswich, MA, USA). Synthetic *Drosophila* pre-*let-7* was 5' ³²P-radiolabeled, gel purified, incubated at 65°C for 5 min and then at 25°C for 30 min.

In Vitro RNA Processing

For multiple-turnover dicing reactions, 150 nM substrate was incubated with 2 nM recombinant Dicer-1 and Dicer-1/Loqs-PB. For single-turnover reactions, 1 nM RNA substrate was incubated with 8 nM Dicer. Dicing reactions contained 7.5 mM DTT, 3.3 mM magnesium acetate, 0.25% v/v glycerol, 100 mM potassium acetate, 18 mM HEPES-KOH (pH 7.4) and 15 mM creatine phosphate, 2.25 μ g creatine kinase, and 1 mM ATP for +ATP reactions; 1 mM EDTA but no creatine kinase or creatine phosphate for –ATP and non-hydrolyzable ATP analog reactions. One mM ATP α S, ATP γ S, AMPcPP, or AMPPcP were used for ATP analog reactions. Reactions were assembled on ice and then pre-incubated at 25°C for 5 min before adding the RNA substrate (Haley et al., 2003).

Aliquots (1 μ l) of reactions with radiolabeled RNA substrate were quenched by the addition of 25 volumes of formamide loading buffer (98% v/v formamide, 0.1% w/v bromophenol blue, 0.1% w/v xylene cyanol, 10 mM EDTA), incubated 5 min at 95°C, and analyzed by electrophoresis through a denaturing polyacrylamide 7 M urea gel using 0.5 \times tris-borate-EDTA buffer (National Diagnostics, Atlanta, GA, USA). To resolve substrate from product a 15 % polyacrylamide/urea gel was used for pre-let-7. Gels were exposed to image plates and analyzed with an FLA-5000 instrument and ImageGauge 3.0 software (Fujifilm, Tokyo, Japan).

Measuring ATP Hydrolysis

To monitor ATP consumption by Dicer-1, α -³²P-ATP (250 nM, 3000 mmol/Ci; PerkinElmer, Waltham, MA, USA) was added. Reactions were quenched by adding a 25-fold excess of formamide loading dye, and samples were spotted onto 20 cm \times 20 cm cellulose thin-layer glass plates (EMD, Darmstadt, Germany) 1.5 cm above the bottom and chromatographed in 0.75 M KH_2PO_4 (adjusted to pH 3.3 with H_3PO_4) until the solvent reached the top of the plate, and then the plate was dried and analyzed by phosphorimager.

Rate Analyses

As described in the Material & Methods section of Chapter II.

Acknowledgements

We would like to thank Traci Tanaka Hall lab members, especially Gang Lu, Robert Dutcher and Yeming Wang for their help with Dicer-1 and Dicer-1/Loqs-PB protein expression and purification, their valuable inputs and helpful discussion; Zamore lab members and Ryuya Fukunaga for helpful discussions.

CHAPTER IV: PRESENCE OF PIRNAS IN *DROSOPHILA* HEADS

Preface

The work in this chapter was the result of collaboration. Dr. Jia Xu from Dr. Zhiping Weng's laboratory coded the pipeline that represented the results in 4.S1 and Figure 4.1. Dr. Can Cenik from Dr. Melissa Moore's laboratory and I analyzed the published small RNA datasets in Figure 4.1, 4.2, 4.3, 4.4, 4.5 Table 4.1 and 4.2, 4.3, 4.4, 4.5. Dr. Chengjian Li and Cindy Tipping prepared and maintained the fly stocks used in this study. Cindy Tipping prepared. I performed head dissections, carried out small RNA cloning and analyzed the results in Figure 4.6, and Figure 4.7 with the help of Dr. Can Cenik.

Abstract

Piwi-dependent small RNAs (piRNAs) are a distinct class of small RNAs differing from miRNAs and siRNAs in their biogenesis and function. piRNAs are discovered in animal germlines as a ~23-30nt small RNA family that function in repressing transposon expression. piRNA biogenesis and function are Dicer-independent, and instead require Piwi family proteins. Despite their critical role in the germline, the function and presence piRNA-like RNAs in the somatic tissues has not been clearly established. Here, we investigated whether reads that have piRNA-like properties generated from *Drosophila* head small RNA sequencing libraries come from the germline as a contaminant or are soma-specific. The distribution of piRNAs from our previously published head libraries revealed a very strong correlation to that of germline piRNAs. However, piRNAs from manually dissected heads are distinct from germline piRNAs, strongly suggesting the presence of a distinct set of somatic piRNAs. We are currently investigating whether these distinct head piRNAs are dependent on the Piwi family proteins in a manner similar to germline piRNAs.

Introduction

piRNAs guide PIWI proteins to silence transposons in the germ line of animals (Cox et al. 2000; Aravin et al. 2001; Aravin et al. 2003; Grivna et al. 2006; Brennecke et al. 2007; Batista et al. 2008; Das et al. 2008; Chamberyron et al. 2008; Desset et al. 2008). Unlike siRNAs and miRNAs, piRNA production does not require Dicer and likely involves only single-stranded precursor RNAs (Vagin et al. 2006; Girard et al. 2006; Saito et al. 2007; Kirino et al. 2007; Ohara et al. 2007; Brennecke et al. 2007; Batista et al. 2008). The biogenesis of piRNAs and their role in transposon silencing have been elucidated mainly from studies of *Drosophila* ovaries and testes (Aravin et al. 2001; Brennecke et al. 2007; Li et al. 2009). *Drosophila* gonads express three PIWI proteins: Piwi, Aubergine, and Ago3 (Ghildiyal et al. 2009; Cox et al. 1998; Cox et al. 2000; Harris et al. 2001; Vagin et al. 2006; Saito et al. 2006; Nishida et al. 2007; Yin et al. 2007; Klattenhoff et al. 2007; Desset et al. 2008; Malone et al. 2009). Piwi localizes to the nucleus in both the germline and the surrounding somatic follicle cells, and helps regulate differentiation and patterning of the germline nurse cells as well as the oocyte (Cox et al. 1998; Cox et al. 2000). Transposon derived piRNAs bound to Piwi silence transposon expression in the nucleus by a poorly understood mechanism.

In contrast, Ago3 and Aub reside in the germ cell cytoplasm (Kotelnikov et al. 2009; Li et al. 2009). High throughput sequencing of piRNAs bound to Piwi, Aubergine, and Ago3 suggested a model called the “Ping-Pong” mechanism for

the production and subsequent amplification of piRNAs in response to transcription of the transposons they target. The Ping-Pong model proposes that Aubergine and Ago3 collaborate both to increase the abundance of piRNAs and to bias piRNAs toward the antisense strand (Brennecke et al. 2007; Gunawarde et al. 2007). The detailed mechanism by which Aubergine and Piwi acquire primary piRNAs is unknown, but recent results suggest that they derive from long RNAs transcribed from “piRNA clusters”. These regions of the genome are specifically transcribed in the gonads and are characterized by transposon-richness, and a dearth of genes (Olivieri et al. 2010). The Ping-Pong model postulates that Aubergine, bound to an anti-sense primary piRNA pairs with the mRNA transcript of an active transposon, resulting in cleavage of the target and generation of a 3' cleavage product bearing a 5' monophosphate. The 5' end of the 3' cleavage fragment then becomes the 5' end of a sense piRNA bound to Ago3. This “secondary piRNA” can then direct cleavage of a primary piRNA transcript derived from a piRNA cluster. The next step reverses this process. The Ago3:sense piRNA complex cleaves the long transcript from a piRNA cluster, generating antisense RNA fragments that bind to Aub. These fragments are envisioned to be trimmed to mature piRNAs (Brennecke et al. 2007; Gunawarde et al. 2007; Li et al. 2009).

piRNAs are best known for their transposon repression function in the germline. Yet, they seem to be mapping to regions other than transposons and there is some preliminary evidence suggesting their presence in tissues other

than the germline. Of particular interest are piRNAs that were reported to map to coding genes (Robine et al. 2009; Saito, Inagaki et al. 2009). Furthermore, methylated small RNAs longer than 23nt were detected in *Drosophila* heads in mutants defective of siRNA production, suggesting the presence piRNAs (Ghildiyal et al. 2008). Longer species of small RNAs have also been reported in somatic tissues (Yan et al. 2011; Morazzani et al. 2012). Furthermore, the RNAs from the *Aplysia* central nervous system have been shown to be methylated and are found in complex with Piwi. One such piRNA species is antisense to the promoter region of CREB2 and its expression is regulated by serotonin. This piRNA species derives from a serotonin dependent methylation of a conserved CpG island in the promoter region of CREB2, and promotes long term changes in neurons upon stimulation (Rajasethupathy et al. 2012).

In *Drosophila* embryos, piRNA species have recently been proposed to bind 3' untranslated region (UTR) and promote deadenylation of the mRNA encoding Nanos, a protein required for anterior-posterior patterning of the developing embryo (Rouget et al. 2010).

In this work, we analyzed piRNA-like small RNAs reads derived from *Drosophila* heads. We analyzed the extent of contamination from the germline to understand whether they are soma specific or germline derived. Next, we optimized cloning of small RNAs from manually dissected *Drosophila* heads to minimize germline contamination. We observe a distinct population of methylated

piRNA-like species in manually dissected heads, and their distribution differs significantly from germline tissues.

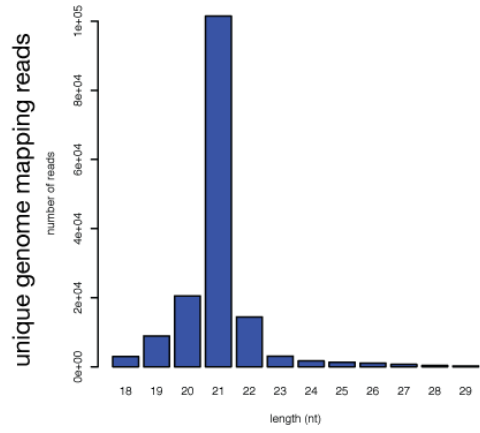
Results

piRNAs are observed in the absence of endo-siRNAs.

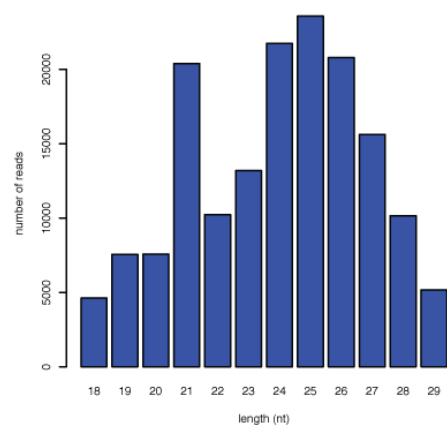
Endogenous methylated siRNA reads (endo-siRNAs) were discovered in *Drosophila* heads by oxidation enriched small RNA cloning (Ghildiyal et al. 2008). Endogenous siRNAs were dependent on siRNA pathway genes Dicer-2, R2D2 and Argonaute-2. Interestingly, in the absence of these genes, production of endo-siRNAs was abolished but longer transposon mapping reads were observed (Ghildiyal et al. 2008). In that study, *Drosophila* heads were extracted by rigorously shaking flies frozen in liquid nitrogen in a two layer sieve allowing the entrapment of bodies on the upper sieve while the heads were trapped in the lower sieve. Although this method is a very time efficient way to extract a large number of *Drosophila* heads, the chance of contamination of the head samples with other parts of the body is high. Therefore, one cannot confidently claim that the observed longer piRNA-like reads were from the head tissue as opposed to germline contamination from the body.

To explore this very important problem, we first undertook a bioinformatics approach. We used many small RNA sequencing datasets previously generated from *Drosophila* germline or heads. We specifically searched for the extent of similarity between the piRNA-like reads from the head and germline datasets (Figure 4.1, 4.1A, 4.1B).

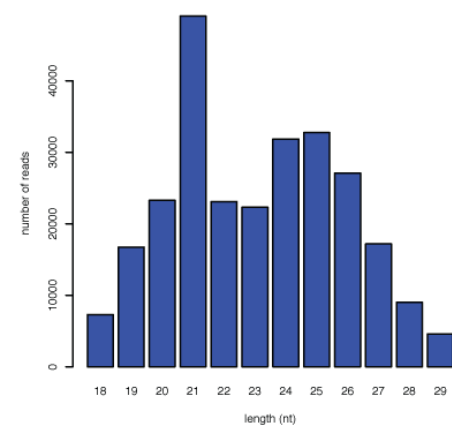
wild type oxidized heads



dcr2 oxidized heads



r2d2 oxidized heads



trasposon mapping reads

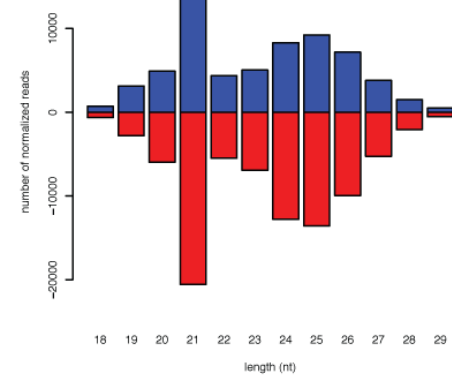
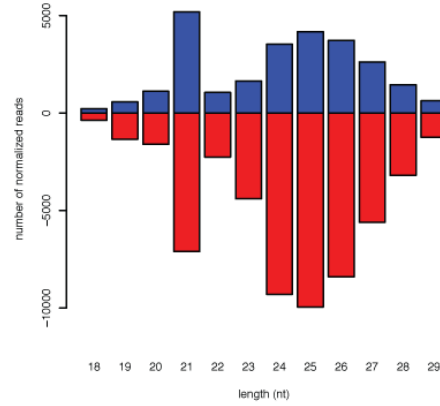
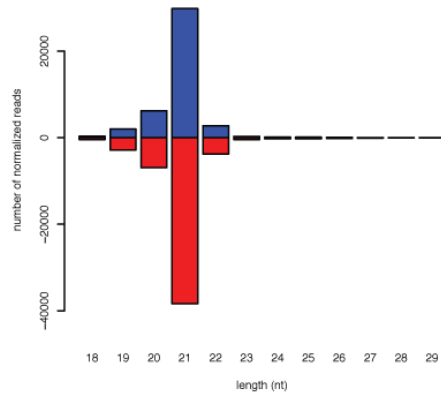
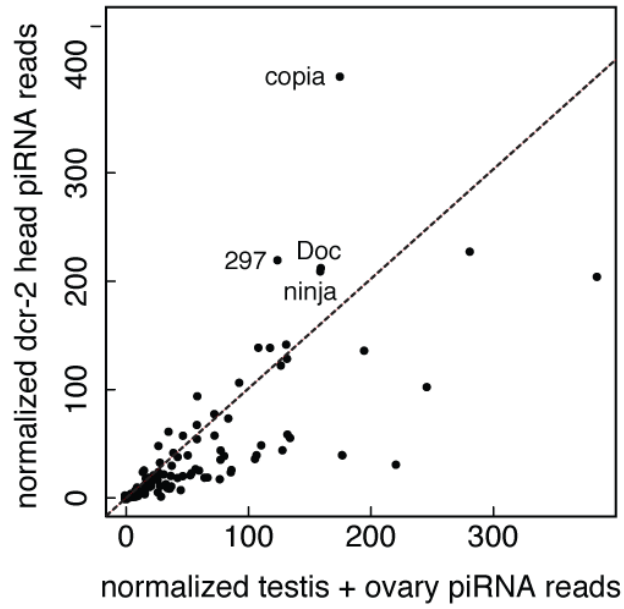


Figure 4.1. Summary of Sequencing Reads. (A) *dcr2* or *r2d2* homozygous mutations is sufficient to abolish the majority of 21nt small RNAs. Length distribution of genome mapped reads were shown as histograms. Three previously published small RNA head libraries from different genotype flies were used (left column: wildtype, center column: *dcr-2*, right column: *r2d2*; see Materials and Methods for details). First row depicts all unique genome mapping reads. The second row shows the length distribution for only transposon mapping reads. Blue and red bars annotate sense and anti-sense directions, respectively.

Transposon mapping piRNAs suggest a high probability of contamination.

piRNA reads from both head and germline datasets were mapped to the *Drosophila melanogaster* genome or transposons (see Methods for the full list of datasets and methods). The number of piRNA reads mapping to each transposon family was compared (Figure 4.2A). The spearman and pearson correlations are shown in Table 4.1. Given the extremely high correlation between two datasets, the most parsimonious explanation is a contamination from germline in the head datasets. The other but less likely possibility is that all transposons mapping piRNAs are expressed at the same level between *Drosophila* heads and germline.

A



B

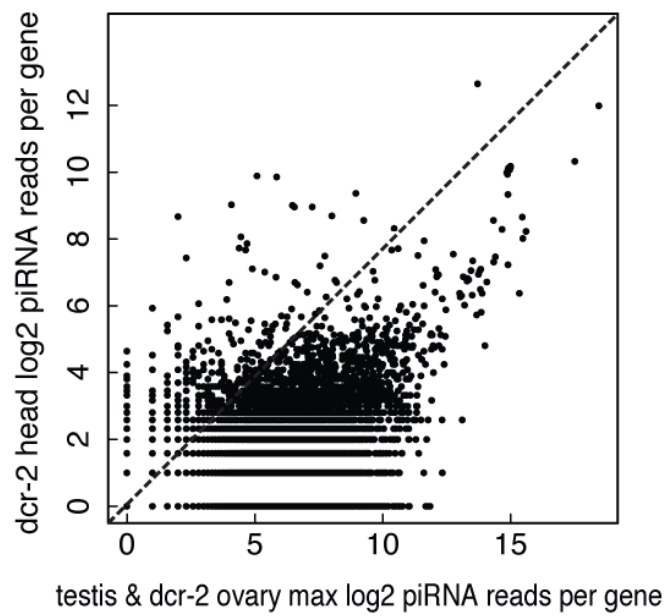


Figure 4.2. Transposon Mapping piRNAs Suggest a High Probability of Contamination. (A) piRNA reads that mapped to each *Drosophila* transposon family were compared in the *dcr2* mutant head and germline. For the germline piRNAs, the sum of reads from ovaries and testes for each transposon family was used. piRNA reads were normalized such that the total number of piRNA reads in the head and germline datasets are equal. (B) piRNA reads that mapped to *Drosophila* coding exons were grouped by transcripts. piRNAs from *dcr2* mutant heads were compared to germline (maximum number of piRNA reads from either ovaries or testes). Similar results were obtained from comparisons with *ago2* and *r2d2* mutant heads instead of *dcr-2* (data not shown).

Table 4.1. Pearson and Spearman Correlations between Transposon Mapping piRNA Reads from Several Published Datasets. Correlation values were calculated by using normalized number of piRNAs mapping to each transposon family.

| Comparison pair | Pearson's product-moment correlation | Spearman's rank correlation |
|---------------------------|--------------------------------------|-----------------------------|
| dcr2Het heads/wt testis | 0.81 | 0.94 |
| dcr2Het heads/wt ovary | 0.55 | 0.80 |
| dcr2 heads/wt testis | 0.86 | 0.95 |
| dcr2 heads/wt ovary | 0.48 | 0.77 |
| ago2 heads/wt testis | 0.44 | 0.55 |
| ago2 heads/wt ovary | 0.74 | 0.74 |
| r2d2 Het heads/wt testis | 0.27 | 0.80 |
| r2d2 Het heads t/wt ovary | 0.40 | 0.80 |
| r2d2 heads/wt testis | 0.71 | 0.88 |
| r2d2 heads/wt ovary | 0.76 | 0.86 |
| wt testis/ wt ovary | 0.20 | 0.72 |

Next, we mapped the piRNA-like reads from oxidized head libraries to the entire genome and plotted regions with mapped piRNAs as a histogram for each chromosome arm at a 1000 base resolution. Visual inspection of histograms for head versus ovary or testis datasets indicated a lower correlation for non-transposon mapping regions. A careful examination showed that a significant proportion of piRNAs mapped to coding genes (data not shown).

piRNAs map to coding genes

To more systematically analyze protein-coding gene mapping piRNAs, genomic coordinates of coding transcripts were extracted from Flybase. Each transcript was further divided into exonic, and intronic regions as well as into 5'UTR, CDS, and 3'UTR. We then matched the coordinates of genome mapping piRNA-like reads from each dataset to count the number of piRNA species mapping to 5', 3'UTR or open reading frame. We analyzed raw read counts as well as read density for each given region. The density was measured as reads per kilobase per million reads (RPKM). However, we should note that as we don't *a priori* expect the reads to tile the entire length of a transcript region, this length normalization might yield a biased estimator of piRNA abundance (Figure 4.3A). We found that more 3'UTR regions contained piRNAs on average in dcr2 mutant fly dataset compared to 5'UTR or open reading frame. piRNAs that map to the coding transcripts had a length distribution that is typical of piRNAs (23-30 nt). (Figure 4.3B). Furthermore, 70% of gene-mapping piRNAs had U as the 5'

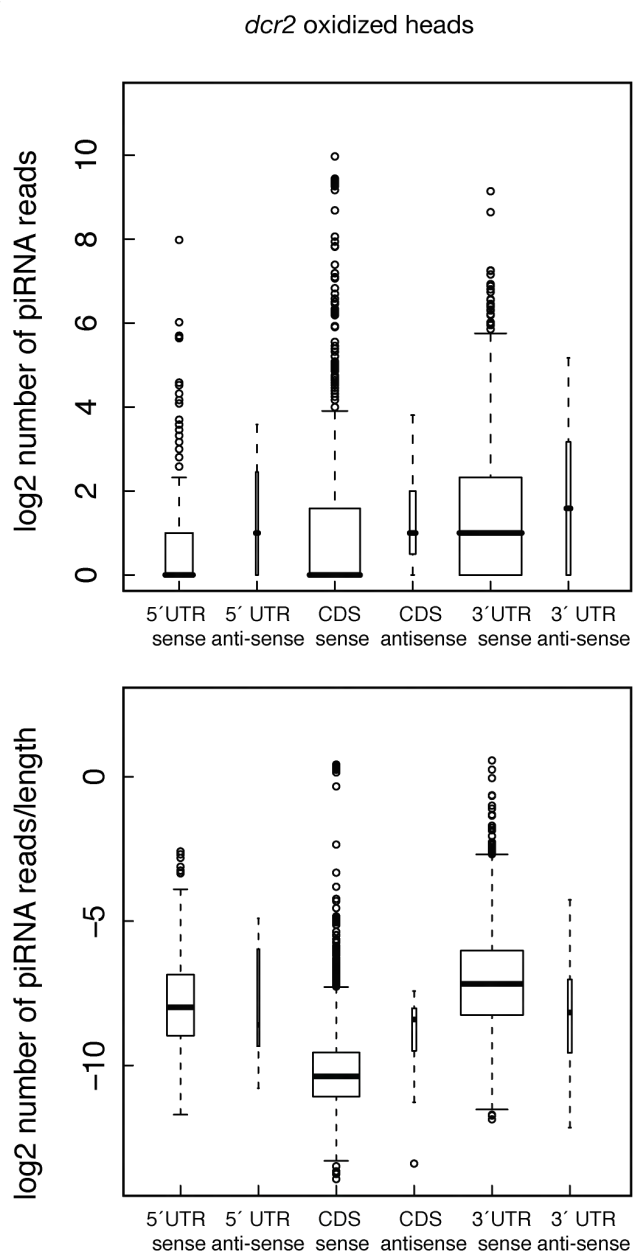
most nucleotide and were more likely to be oriented in the same direction as the transcript.

Gene mapping piRNAs abundance and gene expression correlate weakly.

Next, we wanted to address whether the observed piRNA-like reads are degradation products of mRNAs. Even though the characteristic U-bias and the enrichment of these reads in oxidation libraries argue against this bias, the predominantly sense orientation of these reads is suggestive of a degradation process. We predicted that if these piRNA-like reads are simply degradation products then the number of piRNA species for a given mRNA should be highly correlated to the corresponding mRNAs expression. To test this idea, we downloaded pre-normalized expression profiling data using Affymetrix microarrays from Flyatlas (Chintapalli VR et al. 2007, GEO accession number GSE7763). We extracted gene expression values from relevant tissues, namely ovary, testes and head.

The number of piRNA species mapping to a particular transcript and the corresponding mRNA expression in the relevant tissue (head, ovary or testis) were compared. The correlation values were as follows (Table 4.2.). Transcripts enriched in piRNA species are not necessarily from the most highly expressed genes. This result is suggestive that observed gene-mapping piRNAs are not artifactual products resulting from degradation (Figure 4.4).

A



B

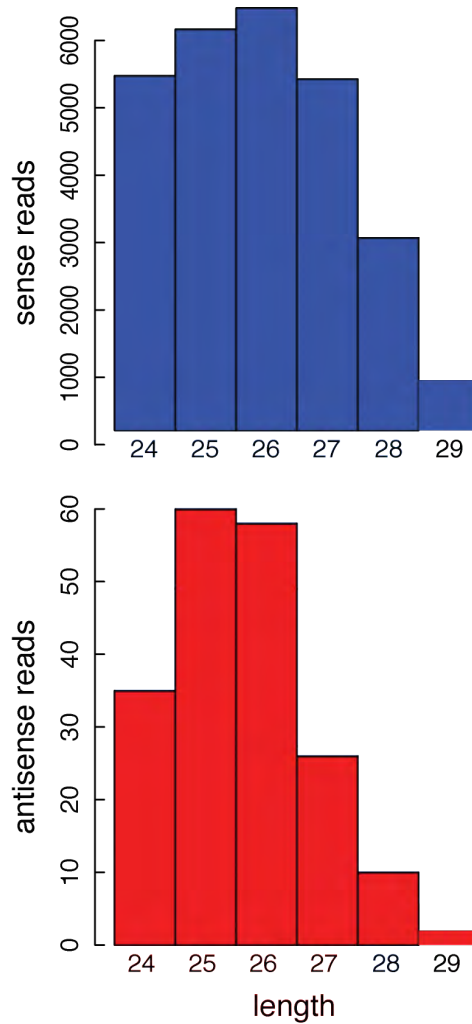


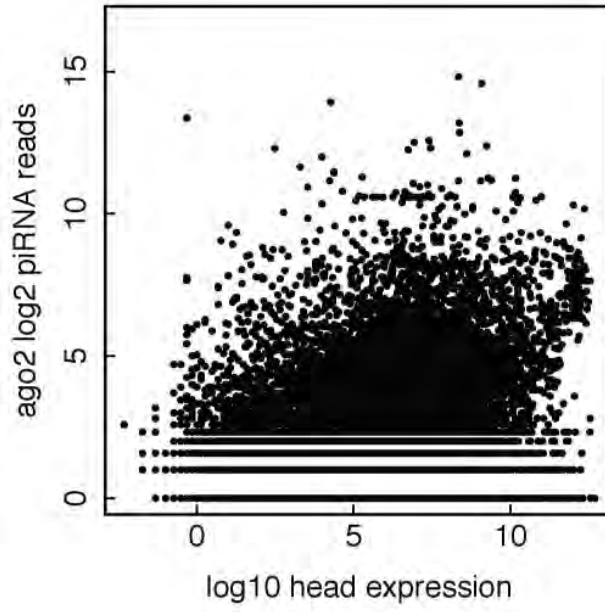
Figure 4.3 piRNAs are More Concentrated in the 3'UTR Region of Coding

Transcripts. (A) A boxplot was used to compare piRNA reads mapping to 5'UTR, coding region or 3'UTR in the sense or anti-sense direction. The width of the box is proportional to the number of datapoints shown with each box. Open circles denote data points that lie outside 1.5 times the interquartile range. The thick line in the box corresponds to the median of the distribution. The same boxplot was drawn using absolute read count (top) or read counts divided by the length of the corresponding region of the transcript (bottom) (B) Length distribution of gene mapping dcr-2 head piRNAs in the sense (blue) or anti-sense(red) orientation was shown as before.

Table 4.2. Correlations between Transposon Mapping piRNA Reads from Several Published Datasets. Number of piRNA reads that map to coding transcripts and corresponding expression values in relevant tissues were compared.

| piRNA reads | Flyatlas Expression value | Pearson's Product-Moment's Correlation | Spearman's Rank correlation |
|-------------|---------------------------|--|-----------------------------|
| dcr2 ovary | wt ovary | 0.01 | 0.57 |
| wt testes | wt testes | 0.03 | 0.44 |
| dcr2 heads | wt heads | 0.03 | 0.58 |
| ago2 heads | wt heads | 0.04 | 0.47 |

ago2 head expression



dcr2 head expression

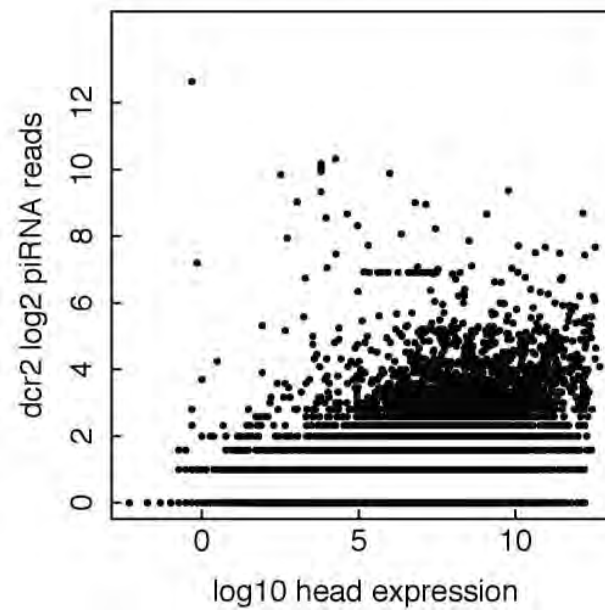


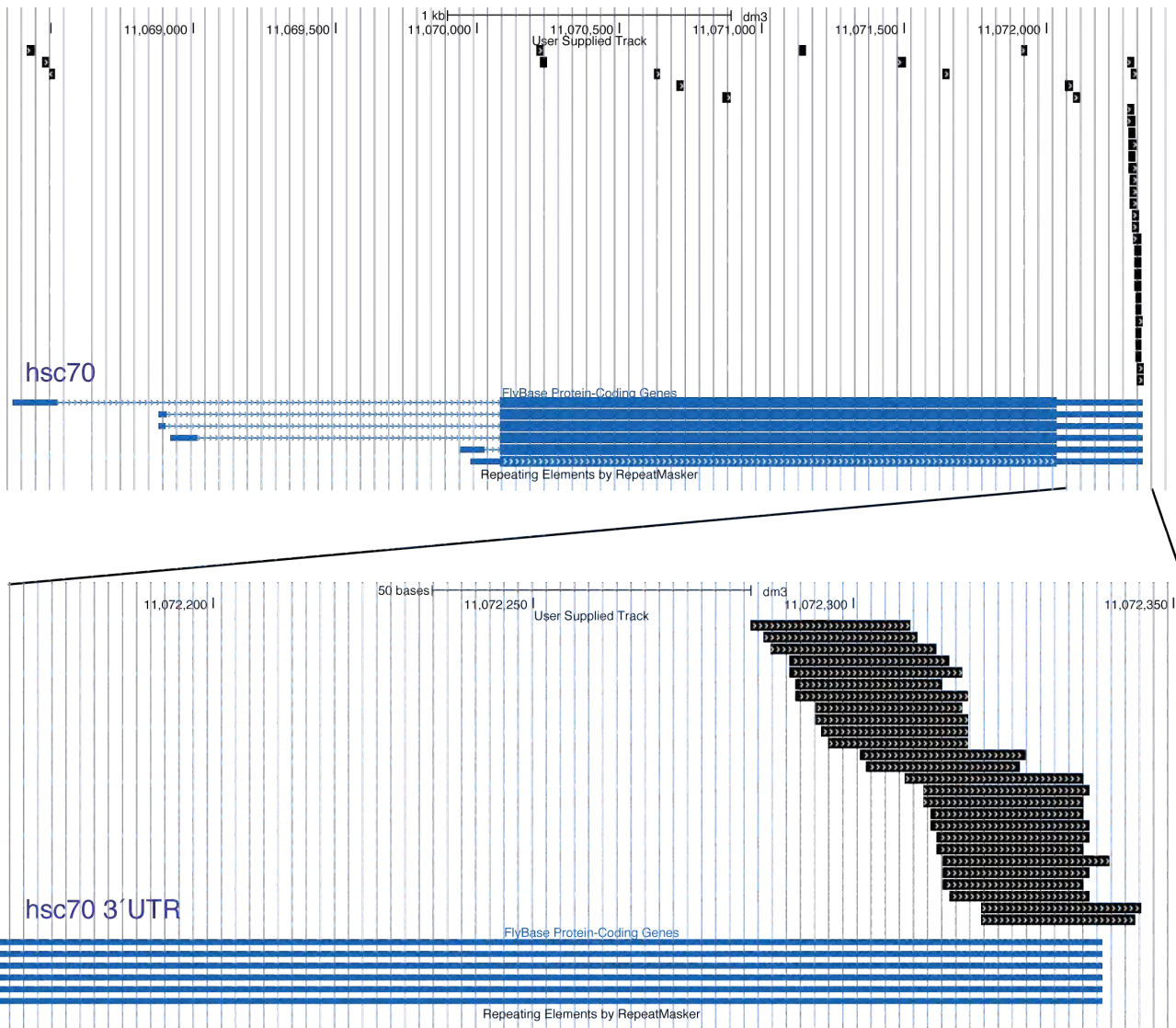
Figure 4.4. Gene Mapping piRNA Abundance and Gene Expression

Correlate Weakly. piRNAs (y axis) that mapped to coding transcripts from ago2 mutant (top panel) and dcr2 mutant (bottom panel) head small RNA libraries were compared with Drosophila head expression values of corresponding transcripts from Flyatlas (see Methods).

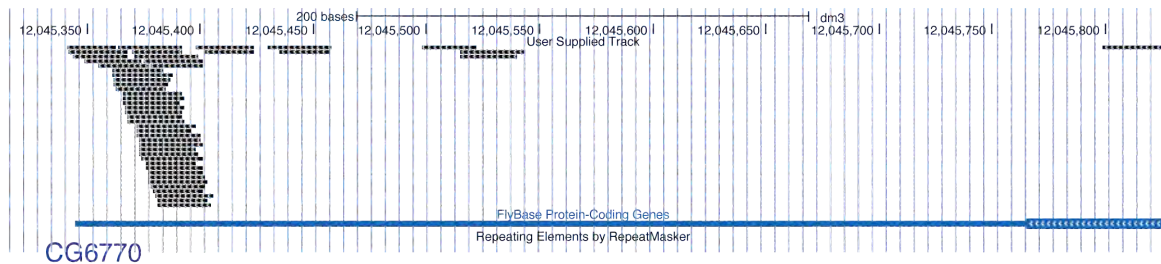
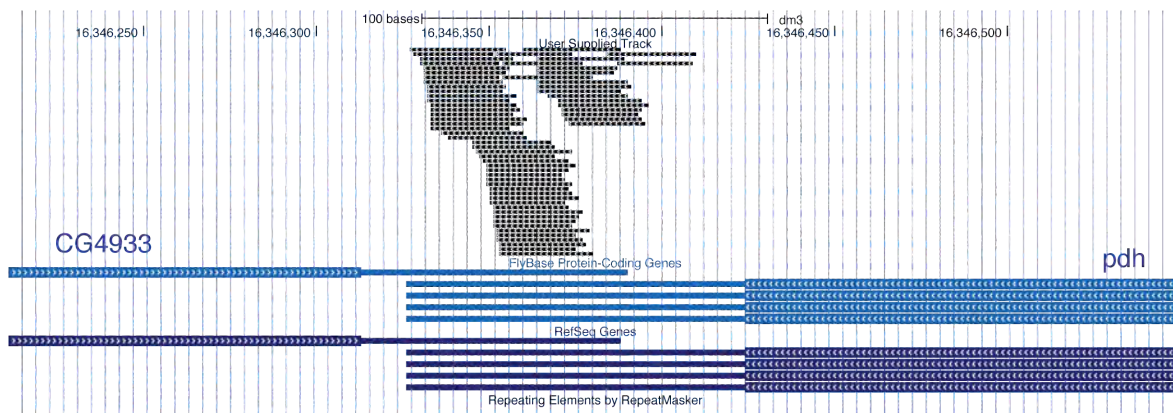
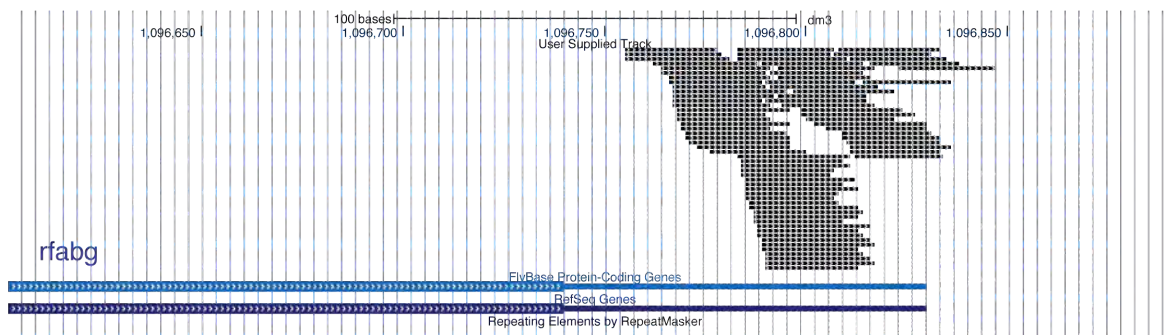
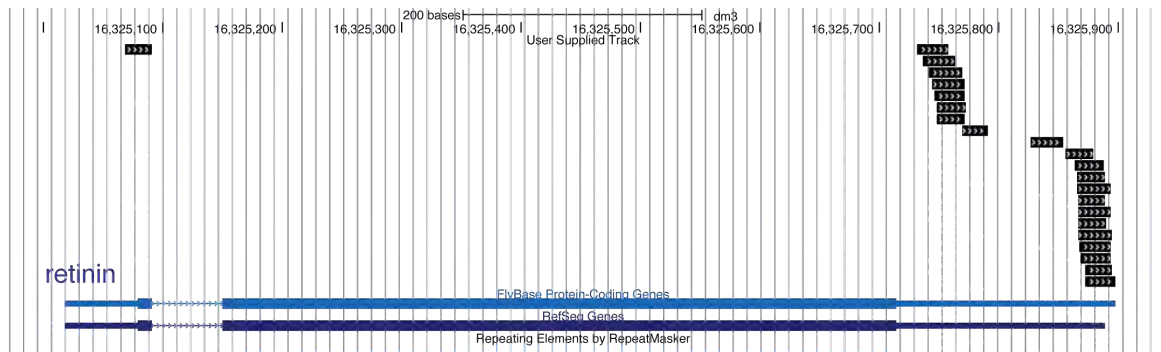
Gene mapping piRNAs predominantly map to a small set of coding genes.

The vast majority of genes do not have any transcripts with mapped piRNAs (Table 4.3). GO enrichment was done for the genes whose 3'UTR enriched for piRNA species. Among the enriched cellular processes for piRNA mapping genes are kinase activity, neural processes and phosphate metabolism (Table 4.4). Top 20 of transcripts with most 3'UTR exon mapped piRNA species, there are several with functions specific to nervous system (Table 4.5). These transcripts are likely to be important for neural functions. Images of examples of such transcripts with mapped piRNAs are shown in Table 4.3, 4.5 and Figure 4.5 (visualization was rendered using the UCSC Genome Browser).

A



B



C

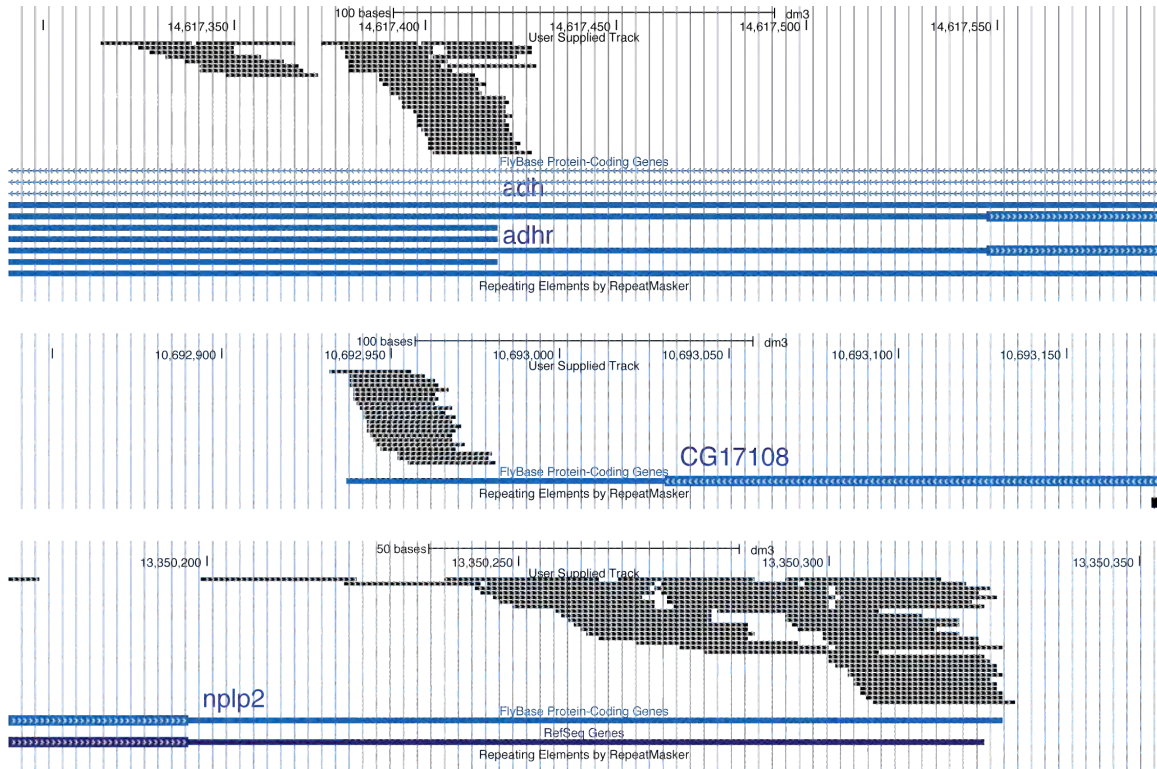


Figure 4.5. Representative coding genes with mapping piRNAs.

Representative transcripts were visualized using UCSC genome browser (Kent et al. 2002, Fujita et al. 2010). (A) The coding gene region of *hsc70* gene is shown in blue and piRNA reads from *dcr2* mutant heads (user track) are shown in black. 3'UTR region of *hsc70* is shown in more detail at the bottom revealing that several different species of reads were mapping to the 3'UTR. (B) Gene regions of *retinin*, *rfabg*, *CG4933*, *CG6770* genes are shown with mapping piRNA reads. (C) Gene regions of *adh*, *adhr*, *CG17108*, *nplp2* are shown with mapping piRNA reads.

Table 4.3. Gene mapping piRNA are Concentrated in a Small Subset of Coding Transcripts. Number of piRNAs mapping to transcripts in either the same (sense) or reverse (anti-sense) orientation was shown. Each transcript is further divided into 5', 3' UTR and coding sequence. Only the top 35 transcripts with the highest number of total species were shown.

| dcr2 Heads | 5' UTR | | | | CDS | | | | 3' UTR | | | | TOTAL |
|---------------|------------------|----------------|--------------------------|------------------------|------------------|----------------|--------------------------|------------------------|------------------|----------------|--------------------------|------------------------|---------|
| | Sense species | Sense reads | Anti sense species | Anti sense reads | Sense species | Sense reads | Anti sense species | Anti sense reads | Sense species | Sense reads | Anti sense species | Anti sense reads | Species |
| CG41580 | 0 | 0 | 0 | 0 | 1004 | 4051 | 4 | 15 | 0 | 0 | 0 | 0 | 1008 |
| CG33239 | 3 | 4 | 0 | 0 | 696 | 1103 | 2 | 2 | 36 | 50 | 0 | 0 | 737 |
| CG33241 | 3 | 4 | 0 | 0 | 696 | 1103 | 2 | 2 | 27 | 42 | 0 | 0 | 728 |
| CG33237 | 3 | 4 | 0 | 0 | 681 | 1084 | 2 | 2 | 37 | 52 | 0 | 0 | 723 |
| CG33245 | 3 | 4 | 0 | 0 | 674 | 1077 | 2 | 2 | 37 | 52 | 0 | 0 | 716 |
| CG33244 | 3 | 4 | 0 | 0 | 674 | 1077 | 2 | 2 | 37 | 52 | 0 | 0 | 716 |
| CG33236 | 0 | 0 | 0 | 0 | 674 | 1077 | 2 | 2 | 37 | 52 | 0 | 0 | 713 |
| CG33242 | 3 | 4 | 0 | 0 | 667 | 1061 | 2 | 2 | 37 | 52 | 0 | 0 | 709 |
| CG33247 | 3 | 4 | 0 | 0 | 660 | 1056 | 2 | 2 | 37 | 52 | 0 | 0 | 702 |
| CG33240 | 3 | 4 | 0 | 0 | 639 | 1022 | 1 | 1 | 37 | 52 | 0 | 0 | 680 |
| CG32616 | 0 | 0 | 0 | 0 | 677 | 1082 | 2 | 2 | 0 | 0 | 0 | 0 | 679 |
| CG33238 | 0 | 0 | 0 | 0 | 630 | 1013 | 2 | 2 | 17 | 21 | 0 | 0 | 649 |
| CG33243 | 3 | 4 | 0 | 0 | 614 | 990 | 1 | 1 | 17 | 21 | 0 | 0 | 635 |
| CG33246 | 3 | 4 | 0 | 0 | 576 | 923 | 2 | 2 | 37 | 52 | 0 | 0 | 618 |
| CG17461 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 565 | 1270 | 9 | 12 | 574 |
| CG42398 | 5 | 7 | 1 | 2 | 412 | 627 | 0 | 0 | 7 | 9 | 0 | 0 | 425 |
| CG32320 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 400 | 6416 | 2 | 2 | 402 |

| | | | | | | | | | | | | | |
|---------|-----|-----|---|---|-----|-----|---|---|-----|-----|----|-----|-----|
| CG11064 | 2 | 2 | 0 | 0 | 168 | 189 | 0 | 0 | 107 | 222 | 0 | 0 | 277 |
| CG40715 | 0 | 0 | 0 | 0 | 267 | 376 | 0 | 0 | 0 | 0 | 0 | 0 | 267 |
| CG34421 | 253 | 389 | 1 | 3 | 0 | 0 | 0 | 0 | 11 | 11 | 0 | 0 | 265 |
| CG32377 | 0 | 0 | 0 | 0 | 246 | 246 | 0 | 0 | 0 | 0 | 0 | 0 | 246 |
| CG40609 | 0 | 0 | 0 | 0 | 226 | 250 | 8 | 8 | 0 | 0 | 0 | 0 | 234 |
| CG41529 | 0 | 0 | 0 | 0 | 228 | 312 | 0 | 0 | 0 | 0 | 0 | 0 | 228 |
| CG5119 | 4 | 4 | 0 | 0 | 4 | 4 | 0 | 0 | 153 | 173 | 0 | 0 | 161 |
| CG12717 | 6 | 6 | 0 | 0 | 146 | 161 | 2 | 2 | 6 | 7 | 0 | 0 | 160 |
| CG4027 | 2 | 3 | 0 | 0 | 8 | 8 | 0 | 0 | 143 | 192 | 0 | 0 | 153 |
| CG17870 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 152 | 209 | 0 | 0 | 152 |
| CG41423 | 0 | 0 | 0 | 0 | 137 | 156 | 2 | 2 | 0 | 0 | 0 | 0 | 139 |
| CG33715 | 0 | 0 | 0 | 0 | 134 | 134 | 0 | 0 | 1 | 1 | 0 | 0 | 135 |
| CG33653 | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 118 | 291 | 3 | 5 | 124 |
| CG34339 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 122 | 162 | 0 | 0 | 122 |
| CG3481 | 3 | 3 | 0 | 0 | 5 | 5 | 0 | 0 | 112 | 171 | 0 | 0 | 120 |
| CG32120 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 94 | 573 | 25 | 354 | 119 |
| CG9674 | 3 | 3 | 0 | 0 | 7 | 7 | 0 | 0 | 86 | 431 | 21 | 219 | 117 |
| CG40748 | 0 | 0 | 0 | 0 | 114 | 134 | 2 | 2 | 0 | 0 | 0 | 0 | 116 |

Table 4.4. Gene Ontology (GO) Enrichment of Transcripts with the Most Mapped piRNAs in dcr-2 Heads. GO enrichment was performed using FuncAssociate. Transcripts with at least 20 piRNA species in their 3'UTR exons was provided as a query set (Berriz et al. 2009).

| GO annotation | Odds ratio | adjusted P value |
|---|------------|------------------|
| metarhodopsin inactivation | 316.5 | 0.009 |
| protein kinase CK2 complex | 56.1 | <0.001 |
| sex differentiation | 40.3 | 0.001 |
| kinase regulator activity | 13.8 | <0.001 |
| protein kinase regulator activity | 13.5 | <0.001 |
| bristle morphogenesis | 12.7 | 0.019 |
| cell morphogenesis involved in differentiation | 10.6 | 0.02 |
| neuron differentiation | 10.4 | 0.019 |
| regulation of kinase activity | 10.2 | <0.001 |
| regulation of protein kinase activity | 10.2 | <0.001 |
| regulation of transferase activity | 10.2 | <0.001 |
| ATP metabolic process | 8.9 | 0.019 |
| regulation of protein phosphorylation | 8.7 | <0.001 |
| ATPase activity, coupled to transmembr. mov. of ions, phosphorylative mechanism | 8.7 | 0.001 |
| ATPase activity, coupled to transmembrane movement of ions | 7.8 | 0.002 |
| regulation of phosphorylation | 7.7 | <0.001 |
| regulation of phosphate metabolic process | 7.0 | <0.001 |
| regulation of phosphorus metabolic process | 7.0 | <0.001 |
| lipid particle | 6.6 | <0.001 |
| regulation of protein modification process | 6.4 | <0.001 |
| growth | 6.1 | <0.001 |
| learning or memory | 5.9 | 0.012 |
| cognition | 5.9 | 0.012 |

Table 4.5. Many transcripts with 3'UTR Mapping piRNAs have Neurological Function. Top 20 transcripts with the most 3'UTR mapped piRNA species are shown along with their functions. 7 out of 20 genes are likely to be involved in neurobiological processes (shown in red).

| Gene name | GO annotation |
|-----------|--|
| senseless | axon guidance |
| CG9674 | glutamate biosynthetic process |
| pathetic | amino acid transmembrane transport, gro |
| CG18854 | inositol-1,4,5-trisphosphate 3-kinase activity |
| dcaps | neurotransmitter secretion |
| CG9941 | zinc ion binding |
| dHCF | histone H3 acetylation |
| 14-3-3ζ | learning or memory |
| CG8765 | lateral inhibition |
| CG14216 | mRNA processing |
| RfaBp | smoothened signaling pathway |
| raspberry | axon guidance |
| CG32850 | zinc ion binding |
| CG40228 | cellular component |
| unkempt | compound eye development |
| Atf6 | regulation of transcription, DNA-dependent |
| Tequila | long-term memory |
| Hr4 | intracellular steroid hormone receptor signaling pathway |
| Pdh | oxidation-reduction process |
| desat1 | male mating behavior |

Distribution of piRNA species from manually dissected male heads does not correlate with that of testis piRNA species.

Given the high contamination risk from germline in previous head piRNA datasets, we decided to generate small RNA sequencing libraries from *Drosophila* heads using a different method. Instead of using the double sieve shaking technique described above, we first froze the flies in liquid nitrogen and manually dissected their heads on dry ice.

Using 1000-2000 fly heads, we were able to isolate very minute amounts of methylated small RNAs. In fact, we were unable to see any visible RNA footprint after 15 cycles of PCR forcing us to increase the number of cycles to an undesirably high number of 25. The lack of sufficient starting material also manifested itself in the low complexity of the sequenced reads. Using stringent criteria for 5' and 3' adapter matching, we obtained a total of ~2.3M reads. However, the library complexity was very poor. We observed only 123K different species and ~33K out of these mapped to the genome. This poor complexity and minute amount of starting material suggests that sample preparation needs to be scaled up at least an order of magnitude (see Table 4.S1 for number of reads and species).

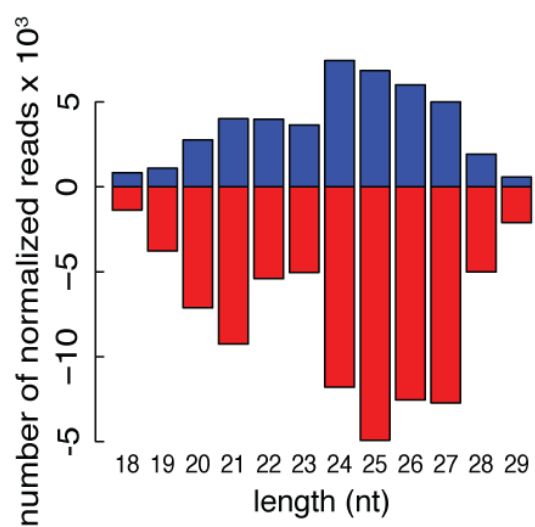
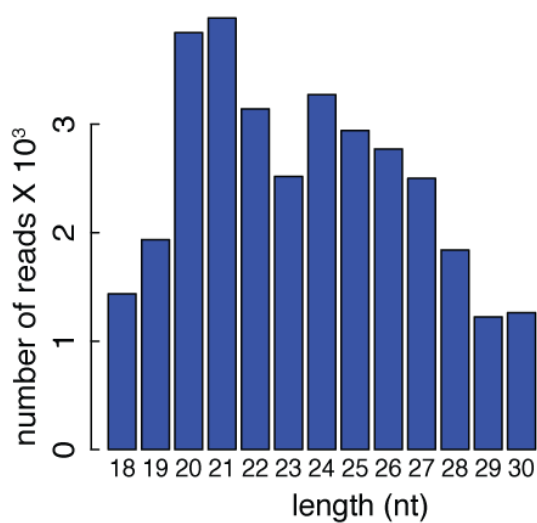
Despite the limited size of this sequencing library and the issues with sample complexity, we next tried to address the issue of contamination. As an additional way to test for contamination, we utilized only male heads in this study. Nearly 50% of testis piRNAs map to a single locus called suppressor of stellate

(su(ste)), and these piRNAs are required for suppressing stellate elements in the germline. By contrast, in the manually dissected head dataset we observed that su(ste) mapping piRNAs are not abundant (Figure 4.6). The choice of male heads, thus, allowed us to estimate an upper bound on the level of contamination using the assumption of no su(ste) expression in heads. Using this assumption, the overall germline piRNA contamination in the manually dissected head dataset was shown to be <1%. Using this upper bound for contamination, we cannot explain the presence of other piRNAs that map to somatic transposons such as gypsy that are more enriched in the heads compared to the germline.

Although gene mapping piRNAs were still detected in this manually dissected small datasets, due to small number of species the conclusions should be drawn with a larger scale manually dissected head library (Figure 4.7).

These results strongly suggest that manually dissected heads is a good alternative way to avoid germline contamination. However, only a minute amount of material can be extracted from even 1000 heads. Therefore, an order of magnitude scale-up is needed for analysis.

A



B

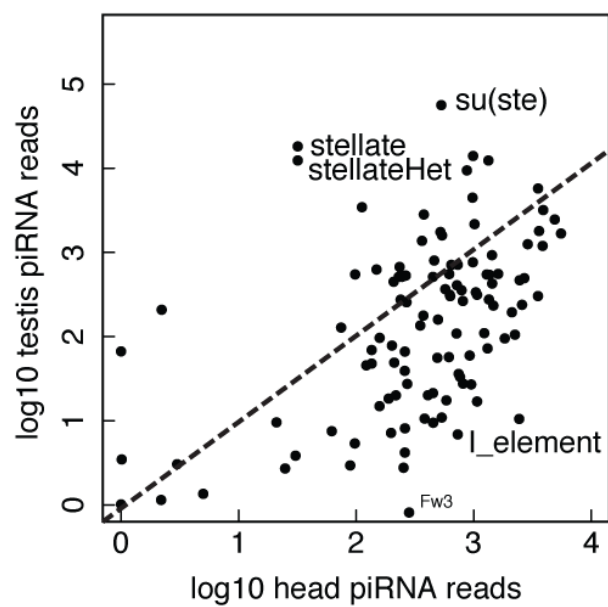
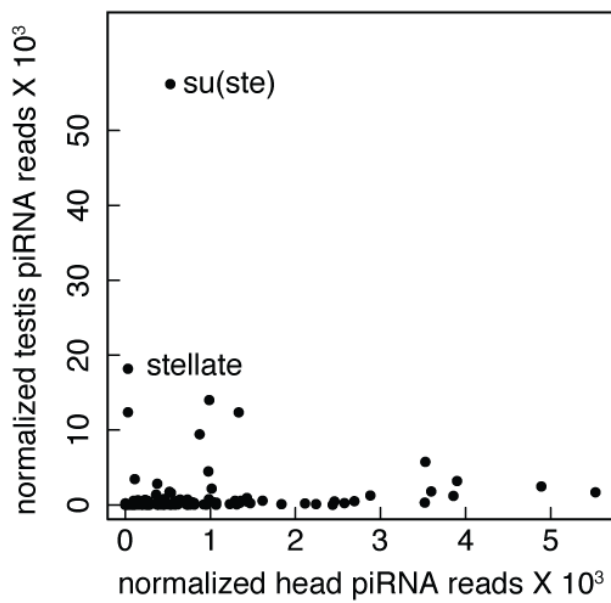


Figure 4.6. piRNAs Isolated from Manually Dissected Male Heads do not Originate from Germline Contamination. (A) Length distribution of all small RNA reads (left), and that of transposon mapping small RNAs divided into sense (blue) and antisense (red) orientation from manually dissected male heads (right) were plotted as before. (B) piRNA reads from from manually dissected male heads were mapped to the *Drosophila melanogaster* transposons. The number of normalized piRNA reads mapping to each transposon family was shown as a scatter plot. Linear (left) and logarithmic (right) scale graphs were plotted.

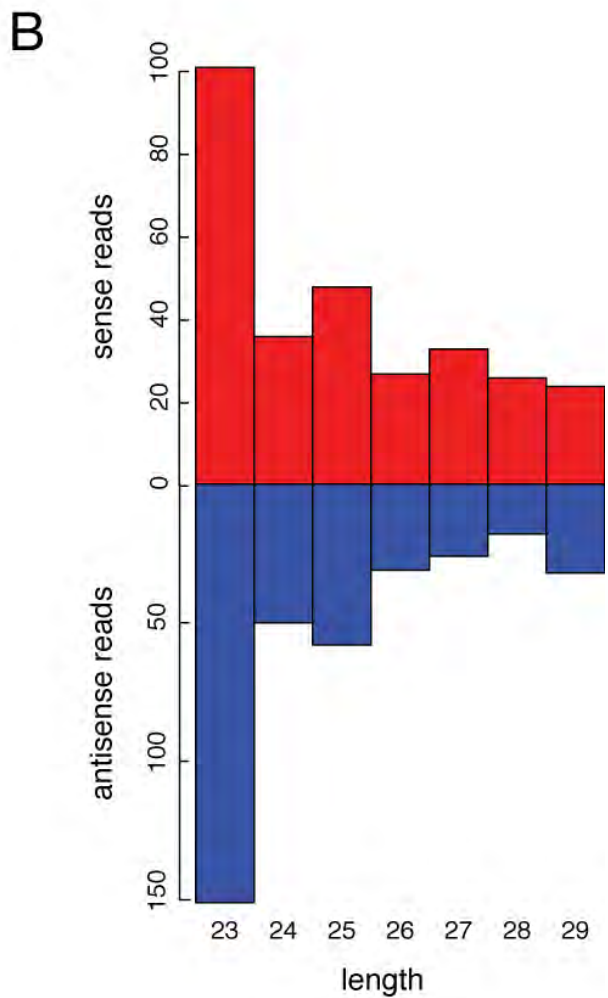
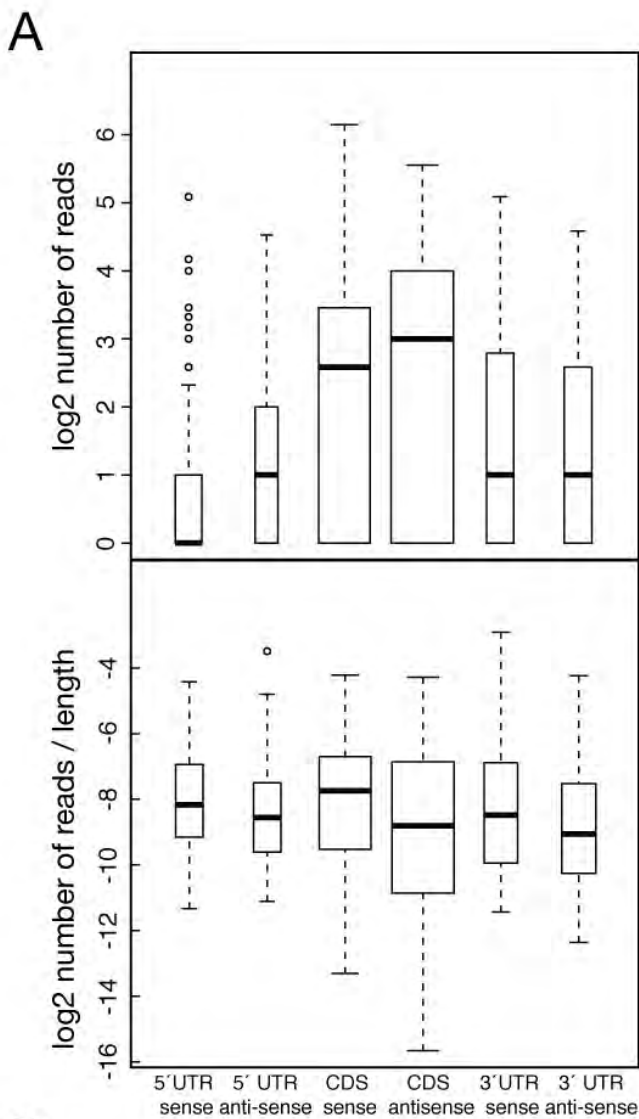


Figure 4.7. Coding Region Mapping piRNAs are Found in Manually Dissected Head Small RNA Library. (A) Distribution of head piRNAs from manually dissected flies that map to each 5'UTR, coding region or 3'UTR in the sense or anti-sense direction were plotted as a boxplot. The width of the box is proportional to the number of datapoints shown with each box. Open circles denote data points that lie outside 1.5 times the interquartile range. The thick line in the box corresponds to the median of the distribution. (B) Length distribution of gene mapping head piRNAs in the sense (blue) or anti-sense (red) direction of the coding gene. (C) Nucleotide frequency plot of the gene mapping piRNA species. Nucleotides G and U are represented more frequently compared with C and A on the 5' most nucleotide.

Table 4.S1. Summary of Manually Dissected piwi Mutant, ago2 Heterozygous Male Head Small RNA Library Reads.

| | Number of reads | Number of species |
|--------------------------------------|------------------------|--------------------------|
| Inserts | 2285582 | 123344 |
| Genome mapping | 624897 | 32661 |
| miRNAs | 47100 | 158 |
| piRNAs | 262009 | 16756 |
| Transposon mapping piRNAs | 98935 | 5825 |
| siRNAs | 61637 | 3875 |
| Transposon mapping siRNAs | 13479 | 692 |

Discussion

piRNA-like small RNAs have previously been reported in somatic tissues of several organisms. Yet, the specific biochemical and genetic properties of these RNAs have not been fully explored. For example, it is unclear whether these longer species are methylated or depend on piwi family proteins for their biogenesis. To overcome these limitations we utilize two strategies. First, we use *Drosophila* lines that have homozygous null mutations in piwi family proteins to test whether the longer species observed in the head are dependent on the piwi family proteins. Second, we use an oxidation procedure that specifically enriches for methylated small RNA species during small RNA sequencing library preparation. This approach is well suited for resolving the question of piRNA presence in *Drosophila* heads.

With this work, we attempted to characterize the sequence landscape of somatic piRNAs in detail. In addition to previously well characterized, transposon mapping piRNAs, we observed a clear signature of piRNAs that map to intergenic and coding regions of the genome. An important question is whether gene mapping piRNAs result merely from unspecific degradation of coding mRNAs? Given that these reads are enriched in oxidized libraries, their 3' end are very likely methylated. This modification argues against these products to be unspecific degradation product. Yet, it would be important to detecting these piRNAs on northern blot after oxidation followed by β elimination as a more direct way to test their methylation.

Another key question to address is how gene mapping piRNAs are made. These piRNAs are likely to be produced by RNA polymerase II, which also is responsible for synthesizing protein-coding mRNAs. Then, what differentiates whether a transcript will be a mature mRNA or get processed into piRNAs? To answer this question the first step would be to test whether piRNAs are processed from the protein-coding mRNAs or they are independently transcribed. One speculative idea is the presence of chromatin associated factors that mark certain transcripts for processing by the piRNA machinery.

Finally, the function of somatic piRNAs especially those that map to protein-coding mRNAs need to be explored. Are these piRNAs function in a similar silencing role as the germline specific transposons mapping piRNAs? To address this question, genome-wide expression profiling of piwi mutant *Drosophila* heads could be used. However, it should be noted that isolation of pure samples from *Drosophila* without any contamination is very time consuming. Therefore, other model organisms that are amenable to genetic alternations might be preferred for studying this problem.

Materials and methods

Small RNA cloning and sequencing

Small RNA cloning was performed as described before with the same set of 5' and 3' adapters (Li et al. 2009). Briefly, 2-4 day flies were frozen with liquid nitrogen and their heads were manually dissected. Between 1000-2000 fly heads were used for generating one sequencing library. In addition to the previously described methods, we depleted 2S rRNA using biotinylated antisense oligos that were subsequently removed using 2.5mg of Dynabeads MyOne streptavidin C1 beads (Invitrogen, Carlsbad, CA, USA). Beads were washed 3 times in B&W buffer (5 mM Tris-HCl(pH 7.5), 0.5 mM EDTA, 1M NaCl) and 2 times in Solution A (0.1 M NaOH, 50 mM NaCl), and resuspended in 600 μ l 0.5X SSC buffer before incubation with the RNA sample. A DNA oligo complementary to Drosophila 2S rRNA was added to a final concentration of 125 nM. After incubation of the reaction for 30 min, heat denatured RNA sample (80 °C for 5 min) was added to the reaction. The reaction was incubated at 50 °C for 1 hr. Supernatant and magnetic beads were separated with a magnet and the RNA from the supernatant was ethanol precipitated. For oxidation of 2S rRNA depleted small RNAs, RNA was incubated with 25 mM NaIO₄ in 60 mM borax, 60 mM boric acid, pH 8.6, for 30 min at 25°C. RNA was similarly extracted by ethanol precipitation.

Genome mapping of small RNA reads

Perfect matching 5' and 3' linkers were selected and removed to identify insert sequences. Extracted raw inserts were mapped to the *Drosophila* genome using pre-compiled Bowtie index for *D. melanogaster*, Flybase, r5.22 (Langmead et al. 2009; McQuilton, et al. 2012). For the mapping, one mismatch was allowed and only the best match was reported using Bowtie parameters <a --best --strata -v 1>. Only unique mapping reads were reported using bowtie parameter <k 1>. To identify 5'UTR, coding region and 3'UTR mapping reads, exon sequences were extracted from Flybase using Galaxy website (Goecks et al. 2010; Blankenberg et al. 2010; Giardine et al. 2005). Separate Bowtie indexes were prepared from 5'UTR, coding region and 3'UTR exonic regions using bowtie-build function. In case of multiple isoforms for a given Flybase gene id, only the longest isoform was selected to simplify quantification of mapped piRNA reads. To identify and annotate transposon mapping reads, *Drosophila melanogaster* transposon sequences were downloaded from Berkeley *Drosophila* genome project and a bowtie index was built as described above. The alignment parameters were the same with the following exception. In addition to unique mapping reads, piRNAs mapping to multiple location were used after dividing the number of reads to the number of mapped locations.

Published datasets used for the analyses

dcr-2, ago-2 and r2d2 mutant fly head small RNA datasets (Ghildiyal et al. 2008), wild type testis and ovary small RNA datasets (Li et al. 2009), dcr-2

mutant ovary small RNA datasets (Czech et al. 2008) were used for the analyses presented throughout the paper.

Graphs and statistical analysis

R software were used for all statistical tests, graphs and correlation analyses.

Frequency analysis

Frequency plots were generated using randomly selected 10 000 sequences by using a sequence logo generator, weblogo (Crooks et al. 2004; Schneider et al. 1990).

Gene expression analysis

For gene expression analysis tab de-limited pre-normalized data from Affymetrix arrays was used (Chintapalli VR et al. 2007, GEO accession number GSE7763).

GO enrichment analysis

Genes with more than 20 mapped piRNA species were used for GO enrichment analysis. Genes were ranked by the number of species and analyzed using the “ordered mode” in FuncAssociate software (Berriz et al. 2009).

Acknowledgements

We would like to thank Can Cenik, Chenjian Li and Jia Xu for their valuable input and discussions, Chenjian Li for providing mutant fly lines for manual head dissection and members of Zamore lab for helpful discussions.

CHAPTER V: FINAL SUMMARY AND CONCLUSIONS

The first part of my thesis focused on the function of *Drosophila* Dicers; mechanism of substrate specificity and the function of the helicase domain.

Genetic studies suggested that Dicer-1 and Dicer-2 are restricted to specific substrate classes in vivo. Dicer-2 mutants are defective for RNAi, even though they express normal levels of Dicer-1. However, Dicer-2 cannot replace Dicer-1 in the miRNA pathway (Lee et al. 2004). It was unclear how Dicers specifically recognize their structurally similar substrates. Given that Dicer-2 processes esiRNA hairpins RNAs while Dicer-1 cleaves pre-miRNAs, suggests that the length of a dsRNA is the primary determinant of substrate choice by the two *Drosophila* Dicer enzymes.

We found that R2D2 and phosphate are two key factors that confer Dicer-2 specificity to longer dsRNA substrates. An intracellular ion, phosphate, inhibits Dicer-2 from processing only short substrates but not longer substrates. In the presence of R2D2 and inorganic phosphate Dicer-2 became specific to longer siRNA precursors. Furthermore, in the presence of ATP, R2D2 and Loqs-PD, Dicer-2 becomes more specific to its long dsRNA substrate with a K_M of ~2 nM compared to ~8 nM when Dicer-2 is alone.

Similarly, ATP and dsRNA binding partners channel Dicer-1 to its biologically appropriate substrate class. To our knowledge, the use of a nucleotide cofactor to ensure biologically appropriate substrate specificity in an RNA silencing pathway is without precedent. Loqs-PB increases pre-miRNA processing efficiency by Dicer-1, potentially by increasing its affinity for substrate.

To understand the role of ATP in Dicer-1 function, the first experiment was to test whether Dicer-1 binds ATP and measure the affinity of ATP binding. We could not detect any ATP hydrolysis by Dicer-1, but we did not test whether Dicer-1 binds ATP. Future work could test whether ATP binding affects substrate binding. A Dicer-1 mutant that can bind to its substrate but cannot process them can be utilized. Another question is whether other molecules such as ADP, AMP or other ATP analogs have effects similar to ATP. The efficiency of dicing and effects on Dicer-1's K_M for pre-miRNAs should be identified.

We do not yet know if the transition of substrates from Dicer-1 to Dicer-2 is gradual such that some substrates are processed equally well by each enzyme. It is also unclear whether there are any biologically relevant precursor RNAs that are cleaved by both Dicer-1 and Dicer-2. One speculative possibility is that evolution has selected against such dual substrates, ensuring the production of precise small RNA products.

Future work could explore processing of substrates as a function of length. Is there a certain length of RNA where Dicer-1 processing is inhibited or is there a gradual effect of length on Dicer-1's affinity? Similarly, our results suggested that ATP dependency of Dicer-2 is length dependent, and there is likely a cutoff length at which ATP becomes necessary for Dicer-2 activity. Future work needs to clearly identify this length by testing varying lengths of double stranded RNA substrates. Another related question is the role of R2D2 and phosphate as a function of substrate length. We know that short hairpins such as pre-miRNAs

are not processed by Dicer-2 in the presence of phosphate and R2D2. Yet, it is unclear whether Dicer-1 or Dicer-2 or both process intermediate length hairpin substrates such as long pre-miRNA hairpins or short endogenous siRNA precursor hairpins.

My thesis also explored the ATPase/helicase domain scaffold of Dicers. This domain is widely conserved among plants and animals, including in humans. Yet, the detailed analysis of its function remains unexplored. One possibility is that the helicase domains provide Dicers with the ability to interact with proteins in other pathways. In *Drosophila*, the Dicer-2 ATPase/helicase domain interacts with components of the innate immunity pathway, promoting an immune response upon entry of an RNA virus into a cell (Wang et al., 2006; Deddouche et al., 2008). Our data suggest that ATP binding or hydrolysis catalyzed by ATPase/helicase domains may also enhance the catalytic function of Dicers.

Two different conformations of Dicer-2 might be formed. One for dicing of long dsRNA and one for loading siRNAs into RNA induced silencing complex. We observed that in the absence of phosphate ion and R2D2, Dicer-2 cleaved short dsRNAs without the requiring ATP. In the presence of phosphate R2D2 and ATP, Dicer-2 might be in a specific conformation where it specifically binds to long dsRNA. It would be interesting to test whether Dicer-2 changes conformation upon ATP binding/hydrolysis and binding of R2D2 and phosphate. Different conformations of Dicer-2 might affect the RNA substrate binding and

this could be tested by footprinting Dicer-2–dsRNA complexes in the presence of ATP, phosphate and R2D2.

We don't know the exact mechanism of phosphate inhibition of pre-miRNA processing by Dicer-2. Does the phosphate ion directly bind Dicer-2? Is it the same domain –PAZ– that binds to the 5' and 3' end of the RNA substrate? From preliminary studies, phosphate seems to mainly increase K_M . Therefore, the inhibition may be competitive. Phosphate might be binding to the same site as the substrate and inhibit its binding. This could be tested by using an RNase III mutant Dicer, which can bind to dsRNA substrate but cannot process it. Specifically, gel shift dsRNA binding assays could be performed and dissociation constant (K_D) of dsRNA binding by RNase III mutant Dicer-2 could be calculated at different phosphate concentrations.

Understanding the function of Dicer's helicase domain is not only important for the miRNA/siRNA field, but also for the helicase field. Our work is one of the few mechanistic studies on RIG-I like helicases and translocation along double-stranded RNAs. A previous study utilized single-molecule translocation assays and detected directional movement of human RIG-I along dsRNA (REF). Our study was focused on the coupling of ATPase function to the translocation of Dicer-2 along dsRNA.

We discovered that *Drosophila* Dicer-2 is a double-stranded RNA induced ATPase. ATP hydrolysis was required for Dicer-2's processivity. We found that Dicer-2 hydrolyzed ~1 molecule of ATP was per base-pair of processed long

dsRNA. ATP seems to be required for Dicer-2's processivity suggesting that Dicer-2 does not release its substrate or any intermediate before fully processing the substrate into siRNAs in ATP's presence. For *Drosophila*, this might be advantageous during viral infections, as the viral nucleic acids will be efficiently processed into siRNAs. Given that neither viral dsRNA precursors nor endogenous dsRNA precursors are abundant, energetic cost of ATP hydrolysis could be offset by the advantage of increased efficiency and specificity. The helicase domain doesn't make Dicer-2 specific to longer substrates, yet ATP dependent mode of dicing is only active in the presence of long substrates. Interestingly, Dicer-2 did not hydrolyze as much ATP when processing dsRNAs shorter than 40 base-pairs compared to longer substrates.

We propose an inchworm model of translocation along dsRNA by Dicer-2 (Pyle 2008). To test whether Dicer-2 moves along dsRNA, traps along the double stranded RNA could be introduced such as stretches of deoxyribonucleic acid, mismatches or hairpins on one of the strands. Effects of these traps to siRNA production by Dicer-2 could be measured. Alternatively, single molecule assays could be performed to measure Dicer-2 translocation along double stranded RNA.

Our dilution experiment suggested that ATP is not only required for translocation along long dsRNA but also enhances binding to the substrate. Directly measuring the effect of ATP on dsRNA substrate binding of Dicer-2 should be measured by binding assays with RNase III mutant Dicer-2.

ATP dependent processing of siRNAs might not be conserved in mammals since published work established that human Dicer function is ATP independent (Ma et al., 2008). Alternatively, ATP dependent function or hydrolysis might be dependent on the presence of additional factors such as TRBP or PACT, two RNA binding partner proteins of human Dicer. It is crucial to test the effect of ATP, phosphate or partner proteins on human Dicer.

While the first part of my thesis elucidated how *Drosophila* Dicers process siRNA and miRNA precursors, the second part focused on piRNAs. It is unknown what marks primary piRNA precursors for processing into mature piRNAs. We know that primary piRNAs are loaded to Piwi, and the primary piRNA precursors are processed into smaller fragments by a nuclease.

Similarly, piRNAs function is still poorly understood. Role of piRNAs in DNA methylation and interaction of Piwi with histone proteins such as HP1a were reported (Brower-Toland et al., 2007; Watanabe et al., 2011) yet we still lack a mechanistic understanding of piRNA function. Repression of transposon activity by piRNAs is well known, but their mechanism of action is not clear. They might destabilize transposon transcripts or they might alter DNA methylation or chromatin structure.

In addition to the well-established role of piRNAs in transposon repression of transposons in the germline (Cox et al. 2000; Aravin et al. 2001; Aravin et al. 2003; Grivna et al. 2006; Brennecke et al. 2007; Batista et al. 2008; Das et al. 2008; Chamberyron et al. 2008; Desset et al. 2008), recent studies

elucidated germline independent functions of piRNAs in *Drosophila* embryos and *Aplysia* central nervous system (Rouget et al. 2010, Rajasethupathy et al. 2012). Although the presence of piRNAs in somatic tissues in *Drosophila* and mammals has been reported, the functions of these somatic piRNAs remain unknown (Yan et al. 2011; Morazzani et al. 2012).

Our study elucidated the presence of piRNAs mapping to the coding genes, especially to their 3' untranslated regions (3'UTRs). A few of these 3'UTR mapping piRNAs overlap with polyadenylation cleavage signal sequences. It is tempting to speculate that these to be involved in polyadenylation site selection. Future studies should test whether mutations in piRNA genes alter mRNA profiles of somatic tissues.

BIBLIOGRAPHY

- Ameres, S. L., Martinez, J., & Schroeder, R. (2007). Molecular basis for target RNA recognition and cleavage by human RISC. *Cell*, 130(1), 101-112.
- Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B. et al. (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev Cell*, 5(2), 337-350.
- Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., & Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol*, 11, 1017–1027.
- Aukerman, M. J., & Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, 15(11), 2730-2741.
- Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209), 64-71.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), 281-297.
- Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136, 215–233.
- Bass, B. L. (2000). Double-stranded RNA as a template for gene silencing. *Cell*, 101(3), 235-238.
- Batista, P. J., Ruby, J. G., Claycomb, J. M., Chiang, R., Fahlgren, N., Kasschau, K. D. et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell*, 31(1), 67-78.
- Bennasser, Y., & Jeang, K. T. (2006). HIV-1 Tat interaction with Dicer: requirement for RNA. *Retrovirology*, 3, 95.
- Beran, R. K., Bruno, M. M., Bowers, H. A., Jankowsky, E., & Pyle, A. M. (2006). Robust translocation along a molecular monorail: the NS3 helicase from hepatitis C virus traverses unusually large disruptions in its track. *J Mol Biol*, 358(4), 974-982.

- Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, *409*(6818), 363-366.
- Berriz GF, Beaver JE, Cenik C, Tasan M, Roth FP. (2009). Next generation software for functional trend analysis. *Bioinformatics*. *25*(22):3043-4.
- Bianco, P. R., & Kowalczykowski, S. C. (2000). Translocation step size and mechanism of the RecBC DNA helicase. *Nature*, *405*(6784), 368-372.
- Billy, E., Brondani, V., Zhang, H., Muller, U., & Filipowicz, W. (2001). Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc Natl Acad Sci U S A*, *98*(25), 14428-14433.
- Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., & Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J*, *17*(1), 170-180.
- Bohnsack, M. T., Czapinski, K., & Gorlich, D. (2004). Exportin 5 is a Ran GTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, *10*(2), 185-191.
- Boutla, A., Delidakis, C., Livadaras, I., Tsagris, M., & Tabler, M. (2001). Short 5'-phosphorylated double-stranded RNAs induce RNA interference in *Drosophila*. *Current Biology*, *11*, 1776-1780.
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. et al. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089-1103.
- Brennecke, J., & Cohen, S. M. (2003). Towards a complete description of the microRNA complement of animal genomes. *Genome Biol*, *4*(9), 228.
- Brennecke, J., Stark, A., Russell, R. B., & Cohen, S. M. (2005). Principles of microRNA-target recognition. *PLoS Biol*, *3*(3), e85.
- Brodersen, P., & Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol*, *10*(2), 141-148.

Brower-Toland, B., Findley, S. D., Jiang, L., Liu, L., Yin, H., Dus, M. et al. (2007). *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev*, 21(18), 2300-2311.

Brower-Toland, B., Riddle, N. C., Jiang, H., Huisinga, K. L., & Elgin, S. C. (2009). Multiple SET methyltransferases are required to maintain normal heterochromatin domains in the genome of *Drosophila melanogaster*. *Genetics*, 181(4), 1303-1319.

Cai, X., Hagedorn, C. H., & Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10(12), 1957-1966.

Catalanotto, C., Azzalin, G., Macino, G., & Cogoni, C. (2002). Involvement of small RNAs and role of the qde genes in the gene silencing pathway in *Neurospora*. *Genes Dev*, 16(7), 790-795.

Cerutti, L., Mian, N., & Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends in Biochemical Sciences*, 481-482.

Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A., & Doudna, J. A. (2010). Substrate-Specific Kinetics of Dicer-Catalyzed RNA Processing. *J Mol Biol*.

Chambeyron S, Popkova A, Payen-Groschêne G, Brun C, Laouini D, Pelisson A, Bucheton A.(2008). piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci U S A*. 105(39):14964-9.

Chen, C. Z., Li, L., Lodish, H. F., & Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science*, 303(5654), 83-86.

Chintapalli, V. R., Wang, J., & Dow, J. A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet*, 39(6), 715-720.

Chiu, Y. L., & Rana, T. M. (2002). RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol Cell*, 10(3), 549-561.

Chung, W. J., Okamura, K., Martin, R., & Lai, E. C. (2008). Endogenous RNA Interference Provides a Somatic Defense against *Drosophila* Transposons. *Curr Biol*, 18(11), 795-802.

Colmenares, S. U., Buker, S. M., Buhler, M., Dlakic, M., & Moazed, D. (2007). Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell*, 27(3), 449-461.

Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., & Lin, H. (1998). A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev*, 12, 3715–3727.

Cox, D. N., Chao, A., & Lin, H. (2000). *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development*, 127(3), 503-514.

Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. (2004) WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190.

Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M. et al. (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature*, 453(7196), 798-802.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C. et al. (2009). Hierarchical rules for Argonaute loading in *Drosophila*. *Mol Cell*, 36(3), 445-456.

Daniels, S. M., Melendez-Pena, C. E., Scarborough, R. J., Daher, A., Christensen, H. S., El Far, M. et al. (2009). Characterization of the TRBP domain required for dicer interaction and function in RNA interference. *BMC Mol Biol*, 10, 38.

Das, P. P., Bagijn, M. P., Goldstein, L. D., Woolford, J. R., Lehrbach, N. J., Sapetschnig, A. et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol Cell*, 31(1), 79-90.

Deddouche, S., Matt, N., Budd, A., Mueller, S., Kemp, C., Galiana-Arnoux, D. et al. (2008). The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *Drosophila*. *Nat Immunol*.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., & Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014), 231-235.

- Desset, S., Buchon, N., Meignin, C., Coiffet, M., & Vaury, C. (2008). In *Drosophila melanogaster* the COM Locus Directs the Somatic Silencing of Two Retrotransposons through both Piwi-Dependent and -Independent Pathways. *PLoS ONE*, 3(2), e1526.
- Dumont, S., Cheng, W., Serebrov, V., Beran, R. K., Tinoco, I. J., Pyle, A. M. et al. (2006). RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature*, 439(7072), 105-108.
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836), 494-498.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806-811.
- Flynt, A., Liu, N., Martin, R., & Lai, E. C. (2009). Dicing of viral replication intermediates during silencing of latent *Drosophila* viruses. *Proc Natl Acad Sci U S A*.
- Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y., & Zamore, P. D. (2007). *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell*, 130(2), 287-297.
- Förstemann, K., Tomari, Y., Du, T., Vagin, V. V., Denli, A. M., Bratu, D. P. et al. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol*, 3(7), e236.
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Gardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res*. 2010 Oct 18.
- Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A., & Imler, J. L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nat Immunol*, 7(6), 590-597.

Gan, J., Tropea, J. E., Austin, B. P., Court, D. L., Waugh, D. S., & Ji, X. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, *124*(2), 355-366.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S. et al. (2008). Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells. *Science*, *320*(5879), 1077-1081.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., & Zamore, P. D. (2009). Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA*.

Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet*, *10*(2), 94-108.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*. *15*(10):1451-5.

Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S. et al. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, *308*(5723), 833-838.

Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy. (2010). A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. *11*(8):R86.

Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199-202.

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. et al. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature*, *432*(7014), 235-240.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N. et al. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, *455*(7217), 1193-1197.

- Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I. et al. (2001). Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control *C. elegans* Developmental Timing. *Cell*, 106(1), 23-34.
- Grivna, S. T., Beyret, E., Wang, Z., & Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev*, 20(13), 1709-1714.
- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T. et al. (2007). A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in *Drosophila*. *Science*, 315(5818), 1587-1590.
- Haley, B., Tang, G., & Zamore, P. D. (2003). In vitro analysis of RNA interference in *Drosophila melanogaster*. *Methods*, 30(4), 330-336.
- Haley, B., & Zamore, P. D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol*, 11(7), 599-606.
- Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E., & Lambowitz, A. M. (2007). Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol*, 365(3), 835-855.
- Hamilton, A. J., & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441), 950-952.
- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., & Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, 293, 1146-1150.
- Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., & Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev*, 18(24), 3016-3027.
- Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K. et al. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5), 887-901.

Harris, A. N., & Macdonald, P. M. (2001). *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development*, 128(14), 2823-2832.

Hartig, J. V., Esslinger, S., Bottcher, R., Saito, K., & Forstemann, K. (2009). Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J*, 28(19), 2932-2944.

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P. et al. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol*, 17(14), 1265-1272.

Houseley, J., & Tollervey, D. (2009). The many pathways of RNA degradation. *Cell*, 136(4), 763-776.

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H. et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, 129(1), 69-82.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, É., Tuschl, T., & 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(5531), 834-838.

Hutvagner, G., & Zamore, P. D. (2002). A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science*, 297, 2056-2060.

Iwasaki, S., Kawamata, T., & Tomari, Y. (2009). *Drosophila* argonaute1 and argonaute2 employ distinct mechanisms for translational repression. *Mol Cell*, 34(1), 58-67.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., & Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev*, 19(14), 1674-1679.

Johnston, R. J., & Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature*, 426(6968), 845-849.

Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T. et al. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev*, 19(4), 489-501.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y. et al. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet*, 16(19), 2272-2280.

Kawamata, T., Seitz, H., & Tomari, Y. (2009). Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol*, 16(9), 953-960.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T. et al. (2008). *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature*, 453(7196), 793-797.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002 Jun;12(6):996-1006.

Khurana JS, Theurkauf W.(2010). piRNAs, transposon silencing, and *Drosophila* germline development. *J Cell Biol*.191(5):905-13.

Kennerdell, J. R., & Carthew, R. W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell*, 95(7), 1017-1026.

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., & 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(20), 2654-2659.

Kim, K., Lee, Y. S., & Carthew, R. W. (2006). Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *RNA*, 13, 22-29.

Kirino, Y., & Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol*, 14(4), 347-348.

Klattenhoff, C., Bratu, D. P., McGinnis-Schultz, N., Koppetsch, B. S., Cook, H. A., & Theurkauf, W. E. (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell*, 12(1), 45-55.

Klattenhoff C, Theurkauf W. (2008). Biogenesis and germline functions of piRNAs. *Development*. 135(1):3-9.

- Kotelnikov, R. N., Klenov, M. S., Rozovsky, Y. M., Olenina, L. V., Kibanov, M. V., & Gvozdev, V. A. (2009). Peculiarities of piRNA-mediated post-transcriptional silencing of Stellate repeats in testes of *Drosophila melanogaster*. *Nucleic Acids Res*.
- Krutzfeldt, J., Kuwajima, S., Braich, R., Rajeev, K. G., Pena, J., Tuschl, T. et al. (2007). Specificity, duplex degradation and subcellular localization of antagomirs. *Nucleic Acids Res*, 35(9), 2885-2892.
- Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M. et al. (2005). Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, 438(7068), 685-689.
- Kurth, H. M., & Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA*, 15(4), 675-685.
- Lai, E. C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet*, 30(4), 363-364.
- Langmead B, Trapnell C, Pop M, Salzberg SL. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*.10(3):R25.
- Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294(5543), 858-862.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J. et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956), 415-419.
- Lee, Y., Jeon, K., Lee, J. T., Kim, S., & Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J*, 21(17), 4663-4670.
- Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H. et al. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, 23(20), 4051-4060.
- Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J. et al. (2004). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, 117(1), 69-81.

- Lees-Murdock, D. J., De Felici, M., & Walsh, C. P. (2003). Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics*, *82*(2), 230-237.
- Leuschner, P. J., Ameres, S. L., Kueng, S., & Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep*, *7*(3), 314-320.
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, *115*(7), 787-798.
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H. et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell*, *137*(3), 509-521.
- Li, J., Yang, Z., Yu, B., Liu, J., & Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol*, *15*(16), 1501-1507.
- Lingel, A., Simon, B., Izaurralde, E., & Sattler, M. (2003). Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature*, *426*(6965), 465-469.
- Lingel, A., Simon, B., Izaurralde, E., & Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol*, *11*(6), 576-577.
- Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J. et al. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, *305*(5689), 1437-1441.
- Liu X, Park JK, Jiang F, Liu Y, McKearin D, Liu Q. (2007). RNA. Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* Dec;13(12):2324-9.
- Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H. E., Smith, D. P. et al. (2003). R2D2, a Bridge Between the Initiation and Effector Steps of the *Drosophila* RNAi Pathway. *Science*, *301*(5641), 1921-1925.
- Liu, X., Jiang, F., Kalidas, S., Smith, D., & Liu, Q. (2006). Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA*, *12*(8), 1514-1520.

- Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., & Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, *303*(5654), 95-98.
- Ma, E., MacRae, I. J., Kirsch, J. F., & Doudna, J. A. (2008). Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol*, *380*(1), 237-243.
- Ma, J. B., Ye, K., & Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, *429*(6989), 318-322.
- Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., & Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, *434*(7033), 666-670.
- MacRae, I. J., Zhou, K., & Doudna, J. A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol*, *14*(10), 934-940.
- Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z. et al. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science*, *311*(5758), 195-198.
- Mallory, A. C., Dugas, D. V., Bartel, D. P., & Bartel, B. (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr Biol*, *14*(12), 1035-1046.
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R. et al. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. *Cell*.
- Martinez, J., Patkaniowska, A., H, H. U., Lührmann, R., & Tuschl, T. (2002). Single stranded antisense siRNA guide target RNA cleavage in RNAi. *Cell*, *110*, 563-574.
- Martinez, J., & Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev*, *18*, 975-980.
- Matranga, C., & Pyle, A. M. (2010). Double-stranded RNA-dependent ATPase DRH-3: insight into its role in RNAsilencing in *Caenorhabditis elegans*. *J Biol Chem*, *285*(33), 25363-25371.

Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., & Zamore, P. D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, 123(4), 607-620.

McQuilton P., St. Pierre S.E., Thurmond J., and the FlyBase Consortium (2012). FlyBase 101 – the basics of navigating FlyBase. *Nucleic Acids Res.* 40(Database issue):D706-14.

Miyoshi, K., Miyoshi, T., Hartig, J. V., Siomi, H., & Siomi, M. C. (2010). Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA*, 16(3), 506-515.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., & Siomi, M. C. (2005). Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev*, 19(23), 2837-2848.

Montgomery, M. K., & Fire, A. (1998). Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet*, 14(7), 255-258.

Moore, M. J., & Sharp, P. A. (1992). Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*, 256(5059), 992-997.

Moore MJ, Sharp PA. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*. 365(6444):364-8.

Moore MJ, Query CC. (2000). Joining of RNAs by splinted ligation. *Methods Enzymol*. 317:109-23.

Moshkovich, N., & Lei, E. P. (2010). HP1 recruitment in the absence of argonaute proteins in *Drosophila*. *PLoS Genet*, 6(3), e1000880.

Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. (2012). Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog*.(1):e1002470.

Mullen, T. E., & Marzluff, W. F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev*, 22(1), 50-65.

- Myong, S., Cui, S., Cornish, P. V., Kirchhofer, A., Gack, M. U., Jung, J. U. et al. (2009). Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science*, 323(5917), 1070-1074.
- Nishida, K. M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S. et al. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA*, 13(11), 1911-1922.
- Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S. et al. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet*, 36(11), 1174-1180.
- Nykanen, A., Haley, B., & Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, 107(3), 309-321.
- Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., & Suzuki, T. (2007). The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol*, 14(4), 349-350.
- Okamura, K., Balla, S., Martin, R., Liu, N., & Lai, E. C. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol*, 15(6), 581-590.
- Okamura, K., Chung, W. J., Ruby, J. G., Guo, H., Bartel, D. P., & Lai, E. C. (2008). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature*, 453(7196), 803-806.
- Okamura, K., Ishizuka, A., Siomi, H., & Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev*, 18(14), 1655-1666.
- Okamura, K., Liu, N., & Lai, E. C. (2009). Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Mol Cell*, 36(3), 431-444.
- Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K., & Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J*, 29(19), 3301-3317.
- Papp, I., Mette, M. F., Aufsatz, W., Daxinger, L., Schauer, S. E., Ray, A. et al. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol*, 132(3), 1382-1390.

Park, W., Li, J., Song, R., Messing, J., & Chen, X. (2002). CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in *Arabidopsis thaliana*. *Current Biology*, 12, 1484–1495.

Parker, J. S., Roe, S. M., & Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J*, 23(24), 4727-4737.

Parker, J. S., Roe, S. M., & Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, 434(7033), 663-666.

Pelisson, A., Sarot, E., Payen-Groschene, G., & Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J Virol*, 81(4), 1951-1960.

Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W., & Sontheimer, E. J. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell*, 117(1), 83-94.

Pham, J. W., & Sontheimer, E. J. (2005). Molecular requirements for RNA-induced silencing complex assembly in the *Drosophila* RNA interference pathway. *J Biol Chem*, 280, 39278-39283.

Pyle, A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys*, 37, 317-336.

Rand, T. A., Ginalski, K., Grishin, N. V., & Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. *Proc Natl Acad Sci U S A*, 101(40), 14385-14389.

Rand, T. A., Petersen, S., Du, F., & Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, 123(4), 621-629.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E. et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403(6772), 901-96.

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., & Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell*, 110(4), 513-520.

- Rissland, O. S., Mikulasova, A., & Norbury, C. J. (2007). Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Mol Cell Biol*, 27(10), 3612-3624.
- Rissland, O. S., & Norbury, C. J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat Struct Mol Biol*, 16(6), 616-623.
- Robine, N., Lau, N. C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S. et al. (2009). A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol*, 19(24), 2066-2076.
- Rouget, C., Papin, C., Boureux, A., Meunier, A. C., Franco, B., Robine, N. et al. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature*, 467(7319), 1128-1132.
- Patel SS, Donmez I. (2006). Mechanisms of helicases. *J Biol Chem*. 281(27):18265-8.
- Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E. et al. (2009). A regulatory circuit for piwi by the large Maf gene *traffic jam* in *Drosophila*. *Nature*, 461(7268), 1296-1299.
- Saito, K., Ishizuka, A., Siomi, H., & Siomi, M. C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol*, 3(7), e235.
- Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T. et al. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev*, 20(16), 2214-2222.
- Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., & Siomi, M. C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev*, 21(13), 1603-1608.
- Saleh, M. C., Tassetto, M., van Rij, R. P., Goic, B., Gausson, V., Berry, B. et al. (2009). Antiviral immunity in *Drosophila* requires systemic RNA interference spread. *Nature*, 458(7236), 346-350.

- Schupbach, T., & Wieschaus, E. (1991). Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics*, 129(4), 1119-1136.
- Schneider TD, Stephens RM. (1990). Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* 18:6097-6100.
- Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., & Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115(2), 199-208.
- Schwarz, D. S., Tomari, Y., & Zamore, P. D. (2004). The RNA-Induced Silencing Complex Is a Mg(2+)-Dependent Endonuclease. *Curr Biol*, 14(9), 787-791.
- Seidel, R., Bloom, J. G., Dekker, C., & Szczelkun, M. D. (2008). Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J.*
- Seitz, H., Ghildiyal, M., & Zamore, P. D. (2008). Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. *Curr Biol*, 18(2), 147-151.
- Selbach, M., Schwanhauser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209), 58-63.
- Sempere, L. F., Dubrovsky, E. B., Dubrovskaya, V. A., Berger, E. M., & Ambros, V. (2002). The Expression of the *let-7* Small Regulatory RNA Is Controlled by Ecdysone during Metamorphosis in *Drosophila melanogaster*. *Developmental Biology*, 244, 170-179.
- Shen, B., & Goodman, H. M. (2004). Uridine addition after microRNA-directed cleavage. *Science*, 306(5698), 997.
- Soifer, H. S., Sano, M., Sakurai, K., Chomchan, P., Saetrom, P., Sherman, M. A. et al. (2008). A role for the Dicer helicase domain in the processing of thermodynamically unstable hairpin RNAs. *Nucleic Acids Res*, 36(20), 6511-6522.

Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A. et al. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol*, 10(12), 1026-1032.

Song, J. J., Smith, S. K., Hannon, G. J., & Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, 305(5689), 1434-1437.

Stark, A., Brennecke, J., Russell, R. B., & Cohen, S. M. (2003). Identification of *Drosophila* MicroRNA targets. *PLoS Biol*, 1(3), E60.

Takehita, D., Zenno, S., Lee, W. C., Nagata, K., Saigo, K., & Tanokura, M. (2007). Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human dicer. *J Mol Biol*, 374(1), 106-120.

Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S. et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, 453(7194), 534-538.

Tomari, Y., Du, T., & Zamore, P. D. (2007). Sorting of *Drosophila* small silencing RNAs. *Cell*, 130(2), 299-308.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., & Zamore, P. D. (2004). A protein sensor for siRNA asymmetry. *Science*, 306(5700), 1377-1380.

Tomari, Y., & Zamore, P. D. (2005). Perspective: machines for RNAi. *Genes Dev*, 19(5), 517-529.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785), 320-324.

van Rij, R. P., & Berezikov, E. (2009). Small RNAs and the control of transposons and viruses in *Drosophila*. *Trends Microbiol*, 17(4), 163-171.

van Wolfswinkel, J. C., Claycomb, J. M., Batista, P. J., Mello, C. C., Berezikov, E., & Ketting, R. F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell*, 139(1), 135-148.

- Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, Saleh MC. (2012). Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. *PLoS One*. 2012;7(1):e30861.
- Wang, H. W., Noland, C., Siridechadilok, B., Taylor, D. W., Ma, E., Felderer, K. et al. (2009). Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol*, 16(11), 1148-1153.
- Wang, X. H., Aliyari, R., Li, W. X., Li, H. W., Kim, K., Carthew, R. et al. (2006). RNA Interference Directs Innate Immunity Against Viruses in Adult *Drosophila*. *Science*, 312(5772), 452-454.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T. et al. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, 461(7265), 754-761.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y. et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194), 539-543.
- Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S. et al. (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science*, 332(6031), 848-852.
- Weinberg, D. E., Nakanishi, K., Patel, D. J., & Bartel, D. P. (2011). The inside-out mechanism of dicers from budding yeasts. *Cell*, 146(2), 262-276.
- Welker NC, Maity TS, Ye X, Aruscavage PJ, Krauchuk AA, Liu Q, Bass BL. (2011). Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol Cell*. 41(5):589-99.
- Wickens, M., & Kwak, J. E. (2008). Molecular biology. A tail tale for U. *Science*, 319(5868), 1344-1345.
- Wienholds, E., Koudijs, M. J., van Eeden, F. J., Cuppen, E., & Plasterk, R. H. (2003). The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat Genet*, 35(3), 217-218.

Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., & Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, *426*(6965), 468-474.

Yan, Z., Hu, H. Y., Jiang, X., Maierhofer, V., Neb, E., He, L. et al. (2011). Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res*, *39*(15), 6596-6607.

Yang, N., & Kazazian, H. H. J. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*, *13*(9), 763-771.

Yang, Z., Ebright, Y. W., Yu, B., & Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res*, *34*(2), 667-675.

Ye, X., Paroo, Z., & Liu, Q. (2007). Functional anatomy of the *Drosophila* microRNA-generating enzyme. *J Biol Chem*.

Yi, R., Qin, Y., Macara, I. G., & Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*, *17*(24), 3011-3016.

Yin, H., & Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature*, *450*(7167), 304-308.

Yoneyama, M., & Fujita, T. (2007). Function of RIG-I-like receptors in antiviral innate immunity. *J Biol Chem*, *282*(21), 15315-15318.

Yuan, Y. R., Pei, Y., Chen, H. Y., Tuschl, T., & Patel, D. J. (2006). A potential protein-RNA recognition event along the RISC-loading pathway from the structure of *A. aeolicus* Argonaute with externally bound siRNA. *Structure*, *14*(10), 1557-1565.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W. et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science*, *307*(5711), 932-935.

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

Zhang, H., Kolb, F. A., Brondani, V., Billy, E., & Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J*, 21(21), 5875-5885.

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., & Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118(1), 57-68.

Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J. A., Perrimon, N. et al. (2009). Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform. *RNA*, 15(10), 1886-1895.

Zou, J., Chang, M., Nie, P., & Secombes, C. J. (2009). Origin and evolution of the RIG-I like RNA helicase gene family. *BMC Evol Biol*, 9(1), 85.

APPENDICES

Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease

Elif Sarinay Cenik,^{1,3} Ryuya Fukunaga,^{1,3} Gang Lu,² Robert Dutcher,² Yeming Wang,² Traci M. Tanaka Hall,² and Phillip D. Zamore^{1,*}

¹Department of Biochemistry and Molecular Pharmacology and Howard Hughes Medical Institute, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA

²Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA

³These authors contributed equally to this work

*Correspondence: phillip.zamore@umassmed.edu

DOI 10.1016/j.molcel.2011.03.002

SUMMARY

Drosophila Dicer-2 generates small interfering RNAs (siRNAs) from long double-stranded RNA (dsRNA), whereas Dicer-1 produces microRNAs (miRNAs) from pre-miRNA. What makes the two Dicers specific for their biological substrates? We find that purified Dicer-2 can efficiently cleave pre-miRNA, but that inorganic phosphate and the Dicer-2 partner protein R2D2 inhibit pre-miRNA cleavage. Dicer-2 contains C-terminal RNase III domains that mediate RNA cleavage and an N-terminal helicase motif, whose function is unclear. We show that Dicer-2 is a dsRNA-stimulated ATPase that hydrolyzes ATP to ADP; ATP hydrolysis is required for Dicer-2 to process long dsRNA, but not pre-miRNA. Wild-type Dicer-2, but not a mutant defective in ATP hydrolysis, can generate siRNAs faster than it can dissociate from a long dsRNA substrate. We propose that the Dicer-2 helicase domain uses ATP to generate many siRNAs from a single molecule of dsRNA before dissociating from its substrate.

INTRODUCTION

In *Drosophila melanogaster*, distinct pathways produce 21 nt small interfering RNAs (siRNAs) and ~22 nt microRNAs (miRNAs). The RNase III enzyme Drosha, aided by its partner protein, Pasha, cleaves primary miRNAs to release pre-miRNAs, ~70 nt long stem-loop structures that contain a mature miRNA within their stems (Lee et al., 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2004, 2006). The pre-miRNA is then cleaved by Dicer-1, acting with its double-stranded RNA (dsRNA)-binding domain (dsRBD) protein partner, Loquacious-PB (Loqs-PB), to liberate a duplex comprising the mature miRNA bound to its miRNA*, a partially complementary small RNA derived from the opposite arm of the pre-miRNA stem (Förstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005; Ye et al., 2007). Mature miRNAs can derive from either the 5' or 3' arm of the pre-miRNA stem.

In contrast to miRNAs, *Drosophila* siRNAs are generated by Dicer-2 (Lee et al., 2004), which forms a stable complex with the dsRNA-binding protein R2D2 (Liu et al., 2003). In vitro, Dicer-2 can produce siRNAs in the absence of R2D2, but both Dicer-2 and R2D2 are required to load siRNAs into Ago2 (Liu et al., 2003, 2006; Tomari et al., 2004, 2007; Pham and Sontheimer, 2005). Exogenous siRNAs derive from long dsRNA molecules that are generated experimentally, from viral RNA genomes or intermediates of replication, whereas endo-siRNAs derive from convergent transcription of mRNAs or from RNA from mobile genetic elements (Yang and Kazazian, 2006; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a, 2008b; Tam et al., 2008; Watanabe et al., 2008). A special class of endogenous siRNAs, hp-esiRNAs, derive from partially self-complementary hairpin transcripts. Production of esiRNAs by Dicer-2 requires an alternative partner protein, Loqs-PD. This Loqs isoform contains only two of the three dsRBDs found in the Dicer-1 partner protein, Loqs-PB (Okamura et al., 2008b; Hartig et al., 2009; Zhou et al., 2009; Miyoshi et al., 2010; Hartig and Förstemann, 2011).

Dicer-1 and Dicer-2 each contain two RNase III domains, which form an intramolecular heterodimer whose dimer interface creates two active sites (Zhang et al., 2004; Macrae et al., 2006; Ye et al., 2007). Like other members of the Dicer family, Dicer-1 and Dicer-2 each contain a C-terminal dsRBD and a central PAZ domain, an RNA-binding motif specialized to recognize the two-nucleotide, 3' single-stranded tails of Drosha and Dicer products (Bass, 2000; Cerutti et al., 2000; Lingel et al., 2003, 2004; Song et al., 2003; Yan et al., 2003; Ma et al., 2004; Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006; MacRae et al., 2007). The structure of *Giardia intestinalis* Dicer and functional studies using human Dicer suggest that the distance between the PAZ domain and the active sites of the RNase III domains establishes the length of the small RNA product (Zhang et al., 2004; Gan et al., 2006; Macrae et al., 2006; MacRae et al., 2007; Takeshita et al., 2007).

Differences in the domain architecture of Dicer-1 and Dicer-2 (Figure S1A) are unlikely to explain their distinct substrate specificities. *Drosophila* Dicer-2 shares its domain architecture—including N-terminal DEXDc, Helicase C, PAZ, RNase IIIa, RNase IIIb, and dsRBD domains—with human Dicer, which produces both siRNAs and miRNAs. In contrast, Dicer-1 lacks a DEXDc

domain, yet has a Helicase C domain. DExDc/H and DEAD box domains are found in a wide range of RNA “helicases,” proteins that couple ATP hydrolysis to RNA binding or unwinding (Pyle, 2008). In addition to unwinding nucleic acids, helicase domains can couple ATP hydrolysis to translocation along nucleic acid molecules, rearrange RNA:protein or protein:protein interactions, or act as RNA chaperones (Bianco and Kowalczykowski, 2000; Beran et al., 2006; Bowers et al., 2006; Dumont et al., 2006; Halls et al., 2007; Pyle, 2008; Franks et al., 2010).

In *Drosophila*, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe*, siRNA production from long dsRNAs requires ATP (Zamore et al., 2000; Bernstein et al., 2001; Ketting et al., 2001; Nykänen et al., 2001; Colmenares et al., 2007). Moreover, a point mutation (G31R) in the *Drosophila* Dicer-2 helicase domain blocks siRNA production in vivo, although the protein retains the ability to collaborate with R2D2 to load synthetic siRNAs (Lee et al., 2004; Pham et al., 2004) and highly paired miRNA/miRNA* duplexes (Fürsteman et al., 2007) into Ago2. In contrast, a mutation predicted to inhibit nucleotide binding by the human Dicer helicase domain does not affect dicing (Zhang et al., 2002).

What restricts a given Dicer to a specific dsRNA substrate? We find that purified, recombinant Dicer-2 can cleave pre-miRNA, but that R2D2 inhibits processing of pre-miRNA by Dicer-2, while promoting use of its biologically relevant substrate by reducing the K_M of Dicer-2 for long dsRNA. Moreover, physiological concentrations of inorganic phosphate block pre-miRNA processing by Dicer-2, but do not inhibit processing of long dsRNA by Dicer-2 or pre-miRNA processing by Dicer-1. Thus, the characteristic specificity of Dicer-2 for long dsRNA is not intrinsic to the enzyme, but rather emerges in the presence of inorganic phosphate and R2D2. We also find that Dicer-2 is a dsRNA-stimulated ATPase that hydrolyzes ATP to ADP; ATP hydrolysis is required for Dicer-2 to process long dsRNA, but not pre-miRNA. Wild-type Dicer-2, but not a mutant defective in ATP hydrolysis, can generate siRNAs faster than it dissociates from its long dsRNA substrate. We envision that the Dicer-2 helicase domain uses ATP to drive the movement of Dicer-2 along dsRNA, enabling it to generate many siRNAs from a single molecule of substrate before dissociating from the dsRNA.

RESULTS

Dicer-2 Processes Pre-miRNA Inaccurately

In vivo, Dicer-1 but not Dicer-2 is required to produce miRNAs from the stems of pre-miRNA. Surprisingly, purified, recombinant Dicer-2 cleaved pre-miRNA (Figure 1). However, the size of the miRNA and miRNA* products generated by Dicer-2 differed from those produced by Dicer-1: the predominant Dicer-2 product was one nucleotide shorter than that produced by Dicer-1. For miRNAs residing on the 5' arm of their pre-miRNA, such a difference in size would not alter the miRNA seed sequence, but could promote their inappropriate loading into Argonaute2, which favors 21 nt RNAs, rather than Argonaute1, which prefers 22-mers (Ameres et al., 2011). For the ~60% of *D. melanogaster* miRNAs derived from the 3' arm of their pre-miRNA, the seed sequence of the Dicer-2 product would differ from the authentic miRNA and would therefore regulate a repertoire of mRNAs different from that controlled by the

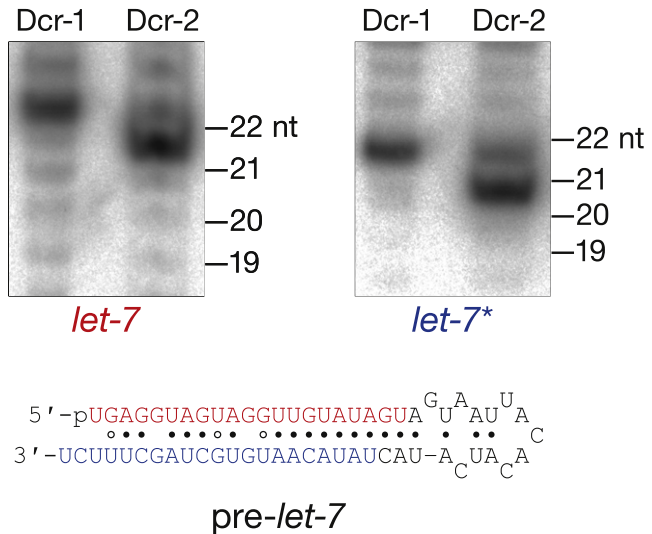


Figure 1. Dicer 1 and Dicer 2 Produce Different Products from Pre *let-7*

Synthetic, 5' monophosphorylated pre *let-7* (1.0 μ M) was incubated with Dicer 1 (6.0 nM) or Dicer 2 (16.2 nM) for 1 hr. Products were resolved by electrophoresis. *let-7* and *let-7** were detected by northern hybridization.

authentic miRNA. The biological consequences of such misregulation are predicted to be dramatic, suggesting that processing of pre-miRNA by Dicer-2 is suppressed in vivo.

Dicer-1 Does Not Efficiently Process Long dsRNA

In the absence of Dicer-2, flies do not accumulate siRNAs (Lee et al., 2004; Ghildiyal et al., 2008; Okamura et al., 2008a, 2008b; Czech et al., 2008). Why does Dicer-1 not make siRNAs in the absence of Dicer-2, especially since immunopurified Dicer-1 has been reported to dice long dsRNA (Saito et al., 2005)? Loqs-PB has been proposed to prevent Dicer-1 from processing long dsRNA, restricting it to the miRNA pathway (Saito et al., 2005). However, we find that Dicer-1 is unable to catalyze multiple-turnover cleavage of long dsRNA (data not shown). In fact, 225 nM Dicer-1 was approximately as active at siRNA production from 25 nM long dsRNA as 5.4 nM Dicer-2. Unlike Dicer-2, Dicer-1 generated intermediates when processing long dsRNA (Figure S1B). Our data argue against Loqs-PB restricting Dicer-1 to the miRNA pathway. Instead, they suggest that the fundamental block to Dicer-1 processing long dsRNA is its inherent inefficiency in using this substrate.

R2D2 Inhibits Dicing of Pre-miRNA by Dicer-2

We compared the rate of pre-*let-7* processing by Dicer-2 alone to the rate of purified Dicer-2/R2D2 heterodimer, Dicer-2 supplemented with equimolar Loqs-PD, and Dicer-2/R2D2 heterodimer supplemented with Loqs-PD. R2D2 significantly inhibited pre-*let-7* processing by Dicer-2 when either enzyme (data not shown) or substrate was in excess (Figure 2A) (p value for excess substrate = 0.0009). Similar inhibition of Dcr-2 processing by R2D2 was observed for a 25 bp RNA duplex (Figure S2A), suggesting that R2D2 suppresses processing of short,

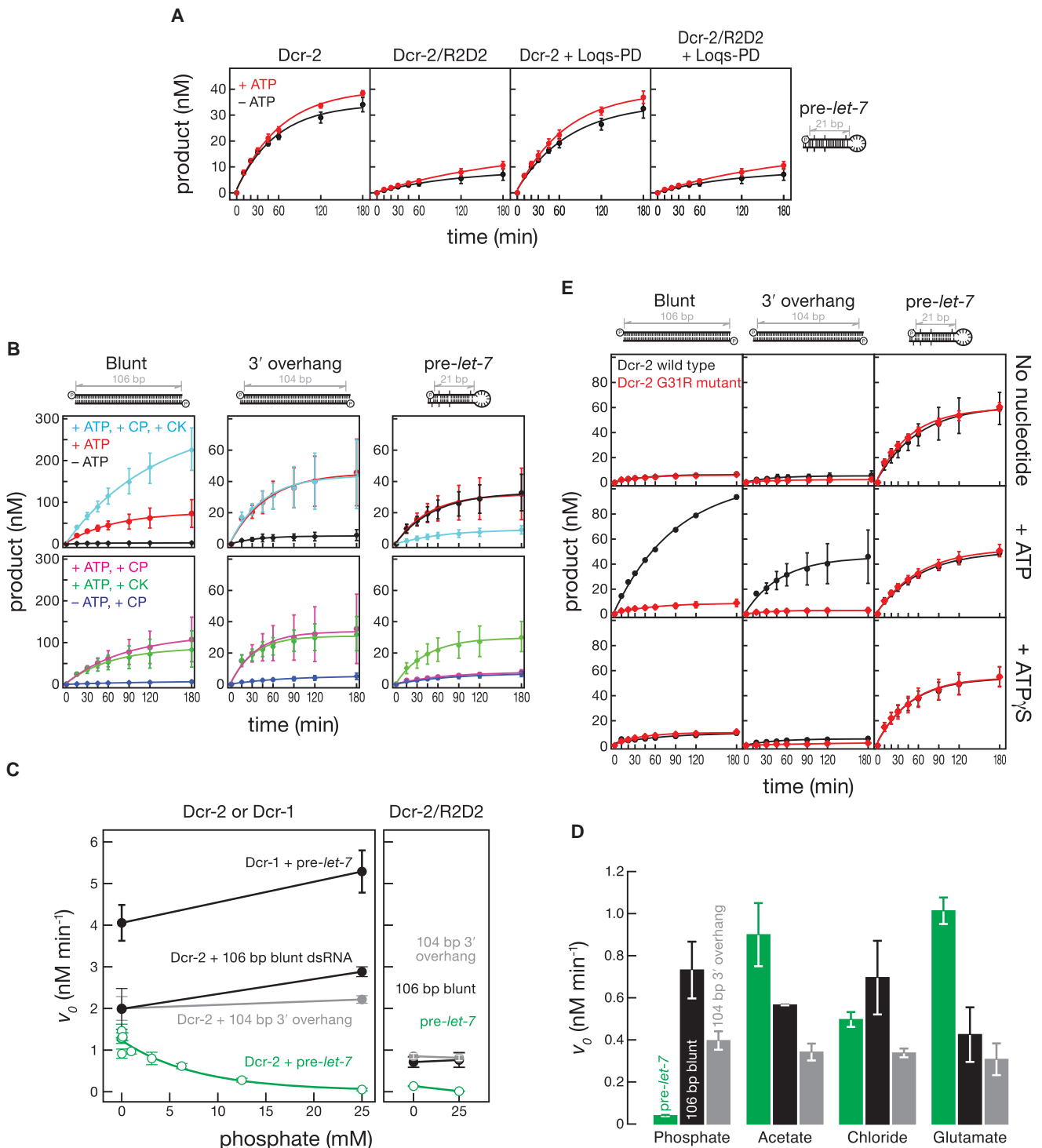


Figure 2. R2D2 and Phosphate Inhibit Dicer 2 Processing of Short Substrates

(A) *let 7* production was monitored using 5' ³²P radiolabeled pre *let 7* (100 nM) with or without ATP for Dicer 2 alone (8 nM), Dicer 2/R2D2 (8 nM), Dicer 2 + Loqs PD (8 nM + 8 nM), or Dicer 2/R2D2 + Loqs PD (8 nM + 8 nM).

(B) Processing of internally ³²P radiolabeled long dsRNA or 5' ³²P radiolabeled pre *let 7* (100 nM) by Dicer 2 (8 nM). CK, creatine kinase; CP, creatine phosphate.

(C) Initial velocities for the processing of 100 nM pre *let 7* or long dsRNA (106 bp blunt ended or 104 bp with 2 nt, 3' overhanging ends) by 8 nM Dicer 2 or Dicer 2/R2D2 in the presence of increasing concentrations of potassium phosphate. Total potassium in the reaction was kept constant.

(D) Initial velocities for Dicer 2 processing pre *let 7*, 106 bp blunt ended dsRNA or 104 bp dsRNA with 2 nt, 3' overhanging ends in the presence of 25 mM potassium phosphate, acetate, chloride, or glutamate.

Table 1. Michaelis-Menten Analysis of Dicer-2

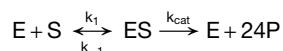
| | K_M (nM) | Change in K_M | V_{max} (nM min ⁻¹) | $[E_T]$ (nM) | k_{cat} (min ⁻¹) | Change in k_{cat} | k_{cat}/K_M (nM ⁻¹ min ⁻¹) | Change in k_{cat}/K_M |
|-------------------|------------|--------------------|-----------------------------------|--------------|--------------------------------|------------------------|---|----------------------------|
| Dicer 2 | 6 ± 2 | 1 | 0.07 ± 0.01 | 2 | 0.03 ± 0.01 | 1 | 0.005 ± 0.003 | 1 |
| Dicer 2/R2D2 | 2 ± 1 | 0.3 | 0.09 ± 0.06 | 3 | 0.03 ± 0.02 | 1 | 0.02 ± 0.02 | 4 |
| Dicer 2 + Loqs PD | 0.4 ± 0.1 | 0.07 | 0.06 ± 0.02 | 2 | 0.03 ± 0.01 | 1 | 0.08 ± 0.03 | 10 |

Dicer 2, Dicer 2/R2D2, or Dicer 2 supplemented with equimolar Loqs PD was incubated with a 515 bp dsRNA and saturating ATP (1 mM) ATP. The initial rates of converting dsRNA into siRNA for increasing amounts of substrate were measured and fit to the Michaelis-Menten equation (Figure S5). The table reports mean ± standard deviation for three trials.

double-stranded substrates irrespective of the extent of complementarity or the presence of a loop. In contrast, we did not detect any inhibition of pre-*let-7* processing when Loqs-PD was added to Dicer-2, even though the same preparation of Loqs-PD lowered the K_M of Dicer-2 for long dsRNA 10-fold (Table 1). These data suggest that, R2D2, but not Loqs-PD, helps suppress pre-miRNA processing by Dicer-2 in vivo.

R2D2 and Loqs-PD Decrease the K_M of Dicer-2 for Long dsRNA

Dicer-2 forms a stable heterodimer with R2D2 (Liu et al., 2003, 2006; Tomari et al., 2004), but it is unknown whether R2D2 modulates dicing rate. We measured the initial rate of dicing by Dicer-2 alone or by the Dicer-2/R2D2 heterodimer using increasing concentrations of long dsRNA substrate and saturating ATP (1 mM). For both Dicer-2 and Dicer-2/R2D2, the data fit well to the Michaelis-Menten kinetic scheme (Figure S5)



where k_{cat} is the rate of complete conversion of substrate into siRNAs at saturating dsRNA concentration. The k_{cat} of Dicer-2/R2D2 processing a 515 bp dsRNA ($0.03 \pm 0.02 \text{ min}^{-1}$) was indistinguishable from that of Dicer-2 alone ($0.03 \pm 0.01 \text{ min}^{-1}$) (Table 1). In contrast, the K_M for Dicer-2/R2D2 ($2 \pm 1 \text{ nM}$) was less than that of Dicer-2 alone ($6 \pm 2 \text{ nM}$, p value = 0.04), suggesting that R2D2 increases the affinity of Dicer-2 for long dsRNA (Table 1). Similarly, supplementing Dicer-2 (2 nM) with purified recombinant Loqs-PD (2 nM) did not alter the k_{cat} but did decrease K_M ($0.4 \pm 0.1 \text{ nM}$, p value = 0.02), suggesting that Loqs-PD also increases the affinity of Dicer-2 for long dsRNA (Table 1).

Phosphate Inhibits Dicing of Pre-miRNA by Dicer-2

In flies, dicing of long dsRNA requires ATP (Nykänen et al., 2001). Typically, creatine kinase (CK) and creatine phosphate (CP) are included in dicing reactions to maintain high levels of ATP and constant levels of free Mg^{2+} . Relative to ATP alone, the inclusion of CK and CP modestly enhanced Dicer-2 processing of both a 106 bp dsRNA bearing a 5' monophosphorylated blunt end (Figure 2B) and a 316 bp dsRNA bearing a 5' triphosphorylated, 2 nt, 5' overhang (Figure S2B), but did not enhance processing

of a 104 bp dsRNA with a 5' monophosphorylated, 2 nt, 3' overhang (Figure 2B). In contrast, standard "ATP" conditions (+ATP, +CP, +CK) inhibited pre-*let-7* processing by Dicer-2. More detailed analyses revealed that CP sufficed to inhibit pre-*let-7* dicing.

CP can be hydrolyzed in water to creatine and phosphate. We therefore tested whether inorganic phosphate inhibited pre-*let-7* processing by Dicer-2. We measured the initial rate of processing (v_0) with increasing concentrations of potassium phosphate (KH_2PO_4/K_2HPO_4 , pH 7.4) for pre-*let-7*, a 106 bp blunt end dsRNA, and a 104 bp dsRNA with 2 nt, 3' overhanging ends. Physiological concentrations of phosphate (Burt et al., 1976; Ercińska et al., 1977; Auesukaree et al., 2004) inhibited processing of pre-*let-7* but neither of the two dsRNAs (Figure 2C). We observed little or no inhibition of pre-miRNA processing by Dicer-2 with 25 mM acetate, chloride, or glutamate (Figure 2D). Moreover, none of the anions—including phosphate—had a significant effect on the dicing of long dsRNA. Phosphate further inhibited the low level of pre-*let-7* processing of Dicer-2/R2D2 (Figure 2C) and Dicer-2/R2D2 + Loqs-PD and suppressed pre-*let-7* processing by Dicer-2 + Loqs-PD (data not shown). In contrast, processing of pre-*let-7* by Dicer-1 was unaffected by phosphate (Figure 2C). We conclude that under physiological conditions—2–25 mM phosphate and a majority of Dicer-2 complexed with R2D2—Dicer-2 is unlikely to use pre-miRNA as a substrate.

ATP Hydrolysis and siRNA Production by Dicer-2

The Dicer-2 G31R point mutation, which lies in the protein's DExDc motif, uncouples Argonaute2 loading from dsRNA dicing (Lee et al., 2004). The mutation is predicted to disrupt ATP binding. Consistent with the requirement for ATP in siRNA production by Dicer-2, dsRNA processing by purified recombinant Dicer-2^{G31R} was significantly less than wild-type Dicer-2 for substrates with either blunt or 3' overhanging ends (Figure 2D) (p value = 2.2×10^{-6} for blunt end substrate). While dsRNA processing by wild-type Dicer-2 was strongly stimulated by ATP, mutant Dicer-2^{G31R} was not (Figure 2D). Nonetheless, Dicer-2^{G31R} cleaved pre-*let-7* as efficiently as the wild-type enzyme, consistent with the finding that ATP was not required for wild-type Dicer-2 to cleave pre-*let-7* (Figure 2A). Moreover, processing of long dsRNA by wild-type Dicer-2 was inhibited by

(E) Processing of internally ³²P radiolabeled long dsRNA or 5' ³²P radiolabeled pre-*let-7* substrate (100 nM) by wild type or G31R mutant Dicer 2 (8 nM) with or without ATP or with ATP_γS (1 mM). Values are mean ± standard deviation for three independent experiments.

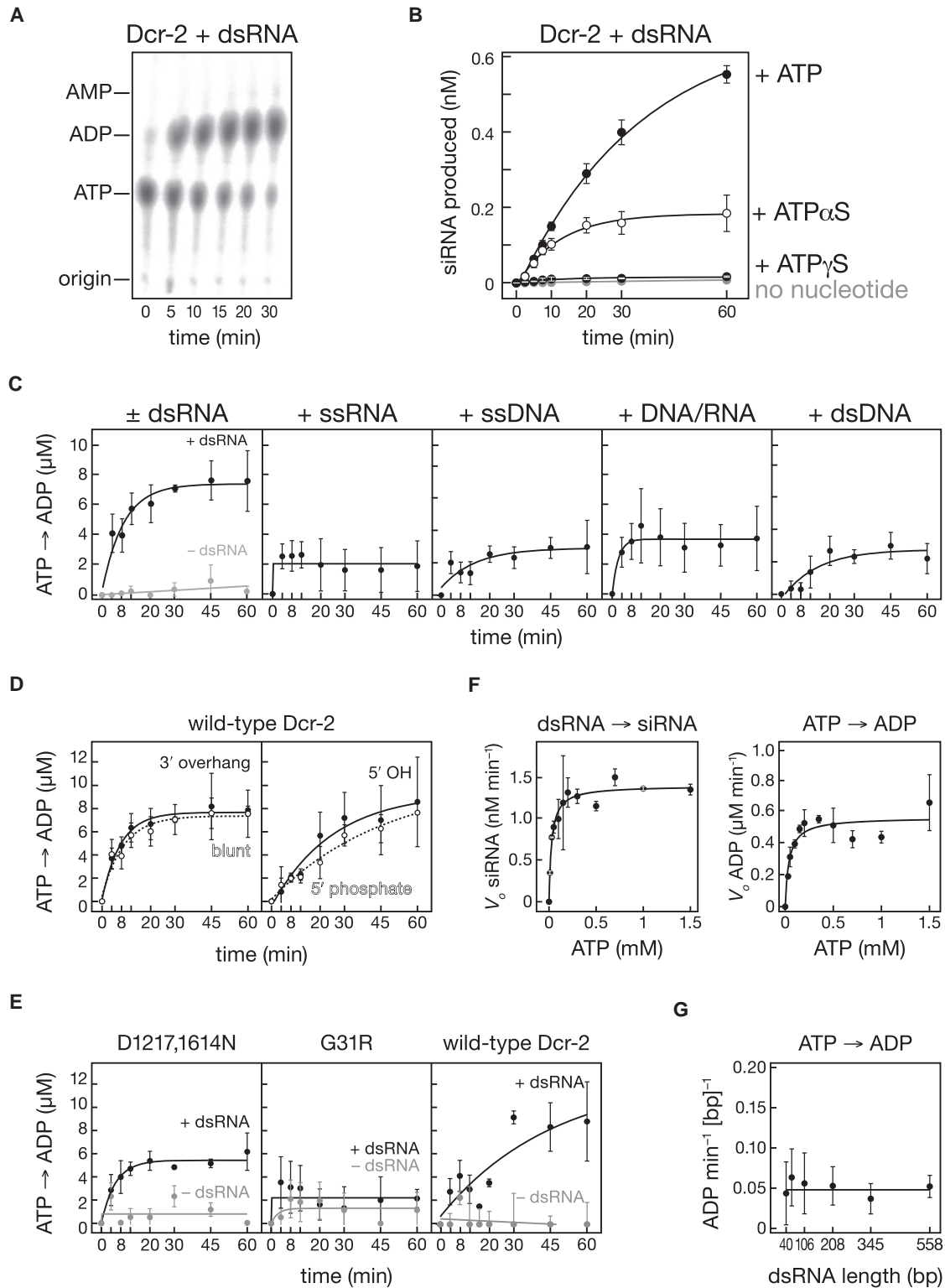


Figure 3. Dicer 2 Requires ATP Hydrolysis to Process Long dsRNA

(A) Dicer 2 (10.8 nM) was incubated with 2.5 μM [α - ^{32}P]ATP and 150 nM 515 bp dsRNA, and ATP hydrolysis was monitored by thin layer chromatography.

(B) siRNA production by Dicer 2 (5.4 nM) from a 515 bp dsRNA (25 nM) was monitored with or without ATP or ATP γ S (1 mM).

(C) ATP hydrolysis by Dicer 2 (5.4 nM) was monitored for 120 nt long nucleic acid substrates (20 nM): dsRNA, single stranded RNA, single stranded DNA, RNA/DNA heteroduplex, or double stranded DNA in the presence of ATP (1 mM).

Table 2. Michaelis-Menten Analysis of Dicer-2 Incubated with a 515 bp dsRNA and ATP

| | K_M (nM) | V_{max} (nM min ⁻¹) | $[E_T]$ (nM) | k_{cat} (min ⁻¹) | k_{cat}/K_M (nM ⁻¹ min ⁻¹) |
|--------------------|----------------|-----------------------------------|--------------|--------------------------------|---|
| Substrate consumed | 6 ± 2 | 0.2 ± 0.1 | 1.2 | 0.2 ± 0.1 | 0.03 ± 0.02 |
| siRNA produced | 6 ± 2 | 5 ± 2 | 1.2 | 4 ± 1 | 0.7 ± 0.4 |
| ATP hydrolyzed | 14,000 ± 4,000 | 460 ± 70 | 5 | 93 ± 14 | 0.007 ± 0.002 |

The initial rates of converting dsRNA into siRNA for increasing amounts of substrate were measured at saturating ATP (1 mM), and the initial rates of hydrolysis of ATP to ADP were measured at saturating dsRNA (150 nM) for increasing amounts of ATP. The table reports mean ± standard deviation for four trials. Different preparations of Dicer 2 were used here and in Table 1.

adenosine 5'-O-(3-thio)triphosphate (ATP γ S), but processing of pre-*let-7* was not (Figure 2D).

In agreement with ATP γ S inhibiting dsRNA processing, Dicer-2 hydrolyzed ATP to ADP (Figure 3A). Both ATP and ATP α S supported the production of siRNA from long dsRNA, but ATP γ S did not (Figure 3B). In the presence of 1 mM ATP, the rate of dicing long dsRNA declined exponentially with increasing ATP γ S, suggesting that ATP γ S competes with ATP for binding to Dicer-2 and that, once bound, ATP γ S is not efficiently hydrolyzed, preventing the production of siRNAs from long dsRNA.

Dicer-2 Is a dsRNA-Stimulated ATPase

ATP hydrolysis by Dicer-2 increased when dsRNA was added (Figure 3C) (p value = 0.024, Wilcoxon rank-sum test). The inclusion of 120 nt single-stranded RNA or DNA or 120 bp DNA/RNA heteroduplex or double-stranded DNA stimulated ATP hydrolysis less than the 120 bp dsRNA (all 20 nM) (Figure 3C). A 25 nt single-stranded RNA or DNA, a 21 bp dsRNA, and a 21 bp DNA/RNA heteroduplex all failed to stimulate the Dicer-2 ATPase activity above the rate observed when no substrate was present (Figure S3B). Neither the end structure of the dsRNA—blunt versus 3' overhang—nor the presence of a 5' monophosphate ($v_0^{5' PO_4} = 0.4 \pm 0.1 \mu\text{M min}^{-1}$ versus $v_0^{5' OH} = 0.3 \pm 0.1 \mu\text{M min}^{-1}$) significantly changed the ATP hydrolysis rate (Figure 3D).

ATP hydrolysis by Dicer-2 does not require dsRNA cleavage. A Dicer-2 mutant, D1217,1615N, in which a key aspartate in each of the two RNase III domain active sites was changed to asparagine, retained significant dsRNA-stimulated ATPase activity (p value = 0.02) (Figure 3E). Under multiple-turnover conditions, Dicer-2^{D1217,1615N} was essentially inactive for dicing, and siRNA production was detected only when [Dicer-2^{D1217,1615N}] \gg [substrate] (data not shown). In contrast, Dicer-2^{G31R}, which also does not support multiple-turnover dicing of long dsRNA, did not hydrolyze ATP in the presence of dsRNA (Figure 3E). We conclude that the Dicer-2 helicase domain is responsible for the enzyme's dsRNA-stimulated ATPase activity.

ATP Consumption and Dicing Are Coupled

We measured the initial rates of siRNA production and ATP hydrolysis using a 515 bp dsRNA in the presence of increasing concentrations of ATP. The dependence on ATP concentration of both siRNA production and ATP hydrolysis fit well to the Michaelis-Menten kinetic scheme (Figure 3F). When both substrate and ATP were saturating (i.e., $\geq 10 \times K_M$), the k_{cat} for ATP hydrolysis was $93 \pm 14 \text{ min}^{-1}$, whereas the k_{cat} for siRNA production was $4 \pm 1 \text{ min}^{-1}$, as inferred from the k_{cat} for the complete conversion of a molecule of substrate into siRNA. These values predict that 23 ± 8 molecules of ATP are hydrolyzed for each 21 nt siRNA formed (Figure 3F and Table 2). Such a high rate of ATP hydrolysis for each siRNA produced might suggest that ATP energy and siRNA production are poorly coupled. Alternatively, ATP might be hydrolyzed to power translocation of Dicer-2 along the dsRNA, with approximately one ATP molecule consumed for each base pair traversed, a rate similar to DNA and RNA translocases containing ATPase/helicase domains (Bianco and Kowalczykowski, 2000; Patel and Donmez, 2006; Seidel et al., 2008). Consistent with this idea, the rate of ATP consumption was essentially unchanged for different substrate lengths when the molar concentration of base pairs was kept constant (Pearson correlation, $r = 0.98$) (Figures 3G and S3B).

Evidence that Dicer-2 Is Processive

If ATP fuels the translocation of Dicer-2 along dsRNA, dicing should produce successive siRNAs along the substrate. In the presence of ATP, Dicer-2 would be predicted to produce the first siRNA—i.e., the terminal siRNA—at roughly the same rate as subsequent, internal siRNAs. In contrast, in the absence of ATP, the rate of production of siRNA should decline as its distance from the 5' end increases. To test these predictions, we synthesized three identical 120 nt dsRNA substrates, each bearing a single ³²P radiolabel at position 16, 35, or 104 from the 5' end of one strand (Figure 4A). All substrates contained two deoxynucleotides at one end, forcing Dicer-2 to initiate processing from the opposite end (Figure S4A) (Rose et al., 2005). We used this “one-ended” substrate to examine the

(D) ATP hydrolysis by Dicer 2 (5.4 nM) in the presence of ATP (1 mM) was measured for 20 nM dsRNA bearing a 5' monophosphate and a 2 nt, 3' overhang or blunt ends (left) or a blunt ended dsRNA bearing either 5' monophosphate or 5' hydroxy termini (right).

(E) ATP hydrolysis by mutant Dicer 2^{D1217,1614N} (3.3 nM), Dicer 2^{G31R} (2 nM), and wild type Dicer 2 (2 nM) was measured in the absence or presence of 20 nM dsRNA bearing 5' monophosphate, blunt ends.

(F) Left panel: Initial rates for the conversion of substrate into siRNA by Dicer 2 (2.7 nM) for an internally ³²P radiolabeled 515 bp dsRNA (150 nM) were measured at increasing concentrations of ATP. Right panel: Initial rates for the hydrolysis of ATP to ADP by Dicer 2 in the presence of 150 nM 515 bp dsRNA were measured for increasing ATP concentrations. The data were fit to the Michaelis Menten equation. Table 2 reports the Michaelis Menten parameters.

(G) ATP consumption by Dicer 2 was measured in the presence of ATP (350 μM) and dsRNA substrates with blunt, 5' triphosphorylated termini for six different lengths: 40 bp (420 nM), 60 bp (280 nM), 106 bp (158 nM), 208 bp (80.5 nM), 345 bp (49 nM), and 558 bp (30 nM). Substrate concentrations were selected to ensure an equal number of base pairs in each reaction. Values are mean ± standard deviation for three independent experiments.

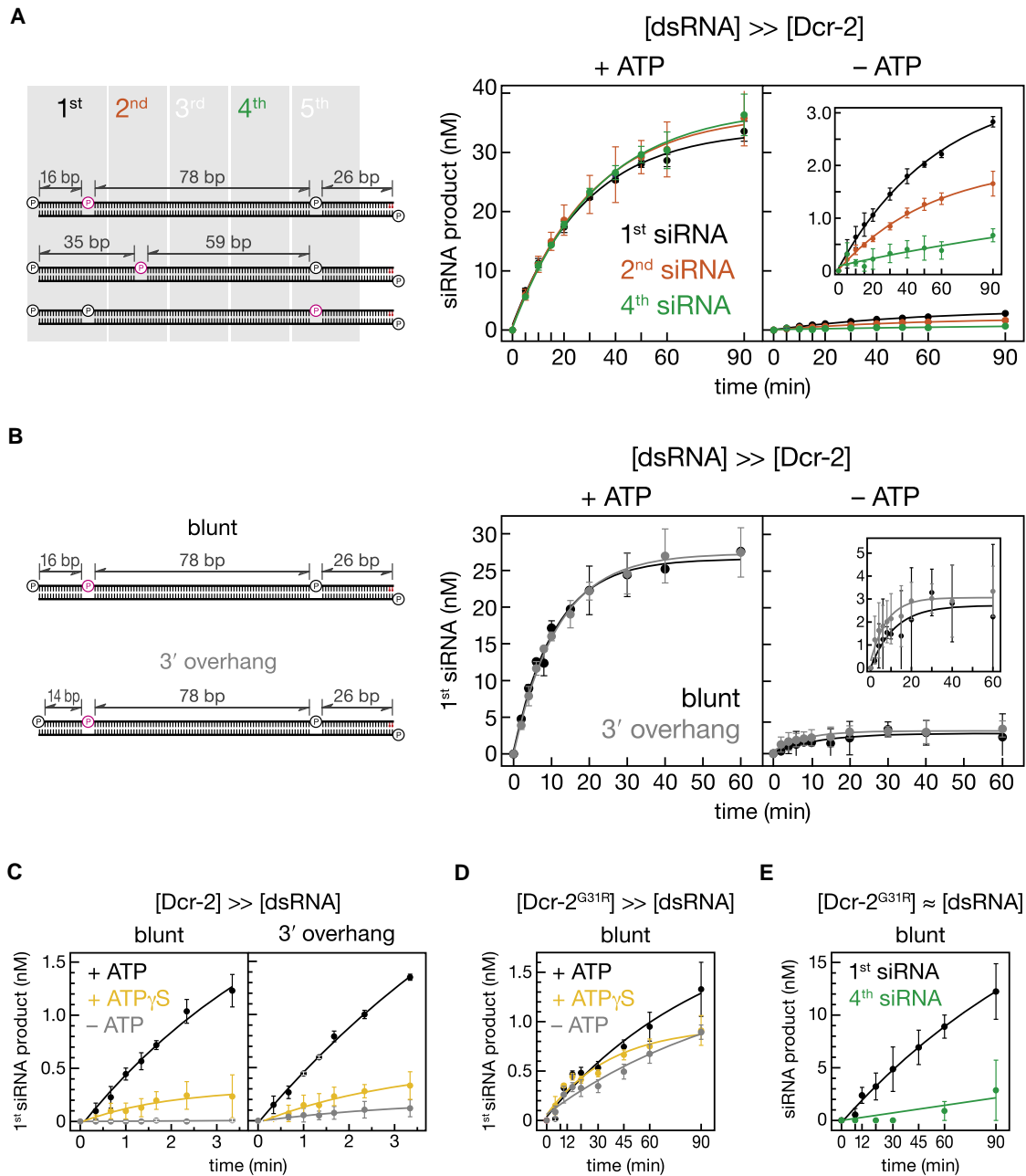


Figure 4. Evidence that Dicer 2 Is Processive

(A) Site specifically ^{32}P radiolabeled 120 bp dsRNAs were used to measure the rate of production of the first, second, and fourth siRNAs from the all RNA end of the substrate; the 3' end of the sense strand contained two deoxynucleotides (red) to block entry of Dicer 2 from the other end (Figure S4A). Production of siRNA was monitored in the presence or absence of ATP using 100 nM dsRNA and 5.4 nM Dicer 2. Values correspond to the mean \pm standard deviation from three independent experiments. The data were fit to a single exponential function. In (A) and (B), pink denotes the position of the ^{32}P radiolabel.

(B) The rate of production of the initial siRNA was measured as in (A) but using a site specifically ^{32}P radiolabeled dsRNA with either a blunt (black) or a 2 nt, 3' overhanging (red) end in the presence or absence of ATP.

(C) The rate of production of the initial siRNA was monitored using 100 nM Dicer 2 and 10 nM dsRNA bearing either a blunt (left) or a 2 nt 3' overhanging (right) end in the presence or absence of ATP or ATP γ S.

(D) The rate of production of the initial siRNA from a 120 bp blunt ended dsRNA (10 nM) was measured using 100 nM mutant Dicer 2^{G31R}.

(E) The rate of production of the first and fourth siRNAs from a 120 bp blunt ended dsRNA (100 nM) was measured using 100 nM Dicer 2^{G31R}. Values are mean \pm standard deviation for three independent experiments.

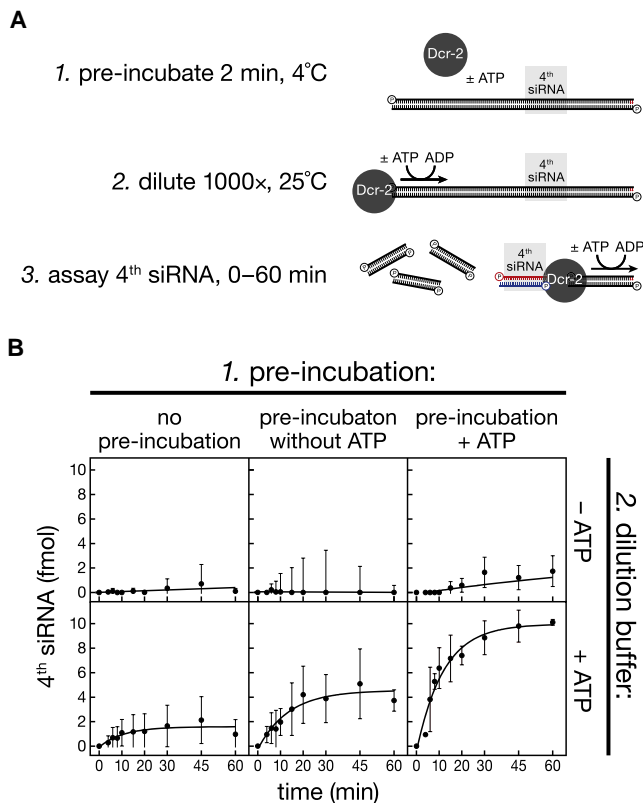


Figure 5. Dicer 2 Remains Associated with Its Substrate in the Presence of ATP

(A) Design of the experiment. The reaction contained 54 fmol Dicer 2 and 200 fmol dsRNA, corresponding to 5.4 nM enzyme and 20 nM dsRNA before dilution.

(B) The rate of production of the fourth siRNA was measured for six different combinations of preincubation (no preincubation, without ATP, and with ATP) and dilution (without and with ATP). Values are mean \pm standard deviation for three independent experiments.

production of the first, second, or fourth siRNA (see [Experimental Procedures](#)) in the presence or absence of ATP ([Figure 4A](#)). With 1 mM ATP, the initial rates for the production of the first, second, and fourth siRNAs were essentially indistinguishable (~ 28 nM min⁻¹); with no ATP, the rate for the first siRNA (1.4 ± 0.3 nM min⁻¹) was greater than that of the second (0.9 ± 0.1 nM min⁻¹), which was greater than the rate for the fourth siRNA (0.2 ± 0.1 nM min⁻¹) ([Figure 4A](#)).

We can envision two explanations consistent with these results and previous studies on Dicer enzymes: either ATP fuels processive dicing of long dsRNA, or dicing of long dsRNA in the presence of ATP comprises a slow initial binding step to the end of the substrate, followed by rapid but ATP-dependent production of subsequent siRNAs. Such a two-step process might occur if the rate of production of the first siRNA was slowed by the blunt structure of the substrate; the first dicing event would convert the substrate end to a 2 nt, 3' overhang bearing a 5' monophosphate, with all subsequent siRNAs produced rapidly. To test this idea, we prepared a substrate bearing one blocked end and a 2 nt, 3' overhang bearing a 5' monophosphate at the

other end ([Figure S4B](#)). The rates of first siRNA production from both blunt and 3'-overhanging end substrates were indistinguishable when the reactions contained ATP—conditions where dicing was efficient. They were also similar when ATP was omitted—when dicing was slow ([Figure 4B](#)). The subsequent siRNAs were also produced at similar rates from the two different substrates ([Figure S4B](#)). Thus, we favor the hypothesis that ATP converts Dicer-2 from an inefficient, distributive enzyme into a processive enzyme.

Surprisingly, ATP also enhanced the production of the terminal siRNA from long dsRNA under single-turnover conditions ($[Dicer-2] \gg [dsRNA]$); the enhancement by ATP γ S was considerably weaker ([Figure 4C](#)). This suggests that ATP hydrolysis is required for the production of even the first siRNA. ATP was required irrespective of the terminal structure of the dsRNA (blunt versus 3' overhang), excluding a role for ATP in the binding of Dicer-2 to a particular type of dsRNA end.

Moreover, when the $[enzyme] > [dsRNA]$, helicase mutant Dicer-2^{G31R} produced the first siRNA at similar rates in the presence or absence of either ATP or ATP γ S ([Figures 4D](#) and [S4C](#)). For Dicer-2^{G31R}, the rate of production of the terminal siRNA was faster than that of the fourth siRNA, consistent with the idea that ATP hydrolysis converts Dicer-2 from a distributive to a processive enzyme ([Figure 4E](#)).

Dicer-2 Produces siRNAs without Dissociating from the dsRNA

A processive enzyme can act multiple times on its substrate before dissociating. Thus, catalysis by processive enzymes resists dilution ([Rivera and Blackburn, 2004](#)). To test whether Dicer-2 remains physically associated with the long dsRNA after three subsequent dicing events, we used the 120 nt, site-specifically ³²P-radiolabeled dsRNA to monitor the rate of production of the fourth siRNA following dilution ([Figure 4A](#)). The dsRNA substrate (200 fmol) was first preincubated with Dicer-2 (54 fmol) for 2 min at 4°C to allow the enzyme to bind substrate. During this preincubation step, no siRNA was detected (data not shown). Next, the reaction was diluted 1000-fold into buffer prewarmed to 25°C. The diluted reaction was incubated at 25°C, and production of the fourth siRNA was measured over time ([Figure 5A](#)). During the preincubation, the substrate concentration (20 nM) was >3 -fold greater than the K_M of Dicer-2 for long dsRNA; after dilution, the substrate concentration was ~ 300 times less than the K_M . For both the preincubation and the dilution steps, the dsRNA substrate was present at ~ 4 -fold higher concentration than Dicer-2. When the preincubation was omitted, little fourth siRNA was produced ([Figure 5B](#)), demonstrating that the conditions largely prevented reassociation of Dicer-2 with dsRNA once it dissociated from the substrate.

When ATP was included in both the preincubation and the dilution buffer, 54 fmol of Dicer-2 produced 10 fmol of fourth siRNA in 1 hr. About half as much fourth siRNA was produced when ATP was present in the dilution buffer but omitted from the preincubation. When ATP was omitted from the dilution buffer, essentially no fourth siRNA was produced, regardless of whether ATP was present during the preincubation, suggesting that ATP dissociates rapidly from Dicer-2. More fourth siRNA was made when ATP was present in both the preincubation

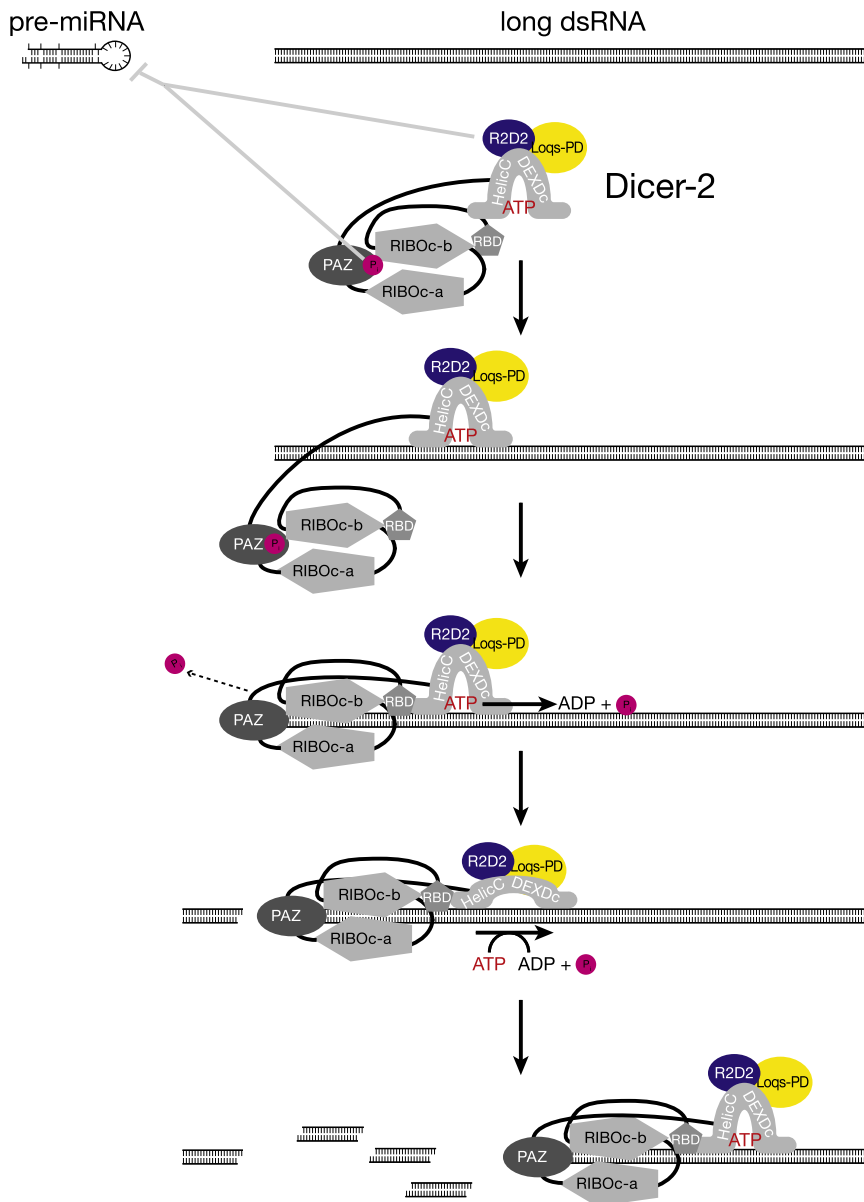


Figure 6. A Model for *Drosophila* Dicer 2

dicer-2 mutants are defective for RNAi, even though they express normal levels of Dicer-1 (Lee et al., 2004). Despite structural similarities, Dicer-2 specifically processes esiRNA hairpins, while Dicer-1 cleaves pre-miRNAs (Lee et al., 2004; Förstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005; Miyoshi et al., 2010). This observation suggests that the length of a dsRNA is the primary determinant of substrate choice.

Our data argue that the combination of R2D2 and cellular phosphate restricts Dicer-2 to its biologically relevant substrates by inhibiting the processing of short substrates such as pre-miRNA. Thus, a protein, R2D2, and a small molecule, phosphate, convert a promiscuous dsRNA endonuclease into one specific for the long dsRNA substrates that trigger RNAi (Figure 6). It is tempting to speculate that inorganic phosphate interferes with recognition of the 5' monophosphate present on all pre-miRNAs and that 5' phosphate recognition is unnecessary for longer substrates, because their greater length allows additional protein-RNA contacts—perhaps by the dsRNA-binding and the helicase domains—between Dicer-2 and long dsRNA. We note that human Dicer has been reported to recognize a 5' monophosphate on single-stranded RNA (Kini and Walton, 2007).

In flies, the Dicer-2 partner proteins Loqs-PD and R2D2 likely enhance substrate specificity by increasing the affinity of the enzyme for long dsRNA. The k_{cat} for Dicer-2 and Dicer-2/R2D2 were similar, but R2D2 decreased the

and dilution buffers than when preincubation was carried out in the absence of ATP (p value = 0.015). We conclude that the initial binding of Dicer-2 to the end of long dsRNA is enhanced by ATP and that in the presence of ATP, Dicer-2 remains associated with the dsRNA. We propose that the stably bound Dicer-2 then cleaves successive siRNAs along the dsRNA (Figure 6).

DISCUSSION

Purified Dicer-1 and Dicer-2 both process pre-miRNAs, but generate products of different length (22 versus 21 nt). Genetic analyses suggest that Dicer-1 and Dicer-2 are restricted to specific substrate classes in vivo (Lee et al., 2004). For example, Dicer-2 cannot replace Dicer-1 in the miRNA pathway. Similarly,

K_M of Dicer-2 for long dsRNA (Table 1 and Figure S5). We note that the specificity constant, k_{cat}/K_M , was ~ 4 fold higher for the Dicer-2/R2D2 heterodimer than for Dicer-2 alone. Similarly, Loqs-PD lowered the K_M of Dicer-2 for long dsRNA without reducing the catalytic rate, resulting in an ~ 10 -fold higher k_{cat}/K_M .

Processing of pre-miRNA by Dicer-1 was unaffected by phosphate. We find that the intrinsic properties of Dicer-1, which cannot efficiently catalyze multiple-turnover processing of long dsRNA, restrict that enzyme to process pre-miRNA. We do not yet know whether the transition of substrates from Dicer-1 to Dicer-2 is gradual, such that some substrates are processed equally well by both enzymes. In theory, such intermediate substrates might be selected against in evolution, enforcing the distinction between Dicer-1 and Dicer-2 substrates.

The Dicer-2 helicase domain is similar to that of RIG-I, a sensor in the mammalian innate immune system. The RIG-like ATPase/helicase domain is conserved among plant and animal Dicers. Yet its function has remained unknown. Our data suggest that this domain of Dicer-2 is involved in ATP-dependent production of successive siRNAs from long dsRNA. Notably, two other members of this helicase family, DRH-3 and RIG-I, are also bona fide ATPases: DRH-3, a *C. elegans* protein required for RNA silencing and germline development (Nakamura et al., 2007), is a dsRNA-stimulated ATPase (Matranga and Pyle, 2010), and the mammalian protein RIG-I, which recognizes viral 5' triphosphorylated dsRNA and initiates an innate immune response, uses ATP to translocate along dsRNA (Myong et al., 2009). Our data are consistent with the idea that ATP hydrolysis fuels translocation of Dicer-2 along long dsRNA substrates. An alternative view—that the ATP-dependent binding of a molecule of Dicer at the end of the substrate promotes the complete and rapid oligomerization of Dicer-2 along the entire extent of the dsRNA—would require that the Dicer and RIG-I helicase domains share a conserved sequence but have highly divergent functions.

ATP was not required for Dicer-2 to process pre-miRNA, and a mutant Dicer-2 unable to hydrolyze ATP remained able to process pre-miRNA but not long dsRNA. These results help explain why in *C. elegans*, in which a single Dicer processes both long dsRNA and pre-miRNA, a mutation in the DCR-1 helicase domain disrupted endo-siRNA, but not miRNA, accumulation (Welker et al., 2010).

Four lines of evidence support a role for ATP hydrolysis in the production of successive siRNAs along the dsRNA by Dicer-2. First, Dicer-2 consumes a constant amount of ATP per base pair. Second, ~23 molecules of ATP were consumed for each 21 nt siRNA produced. Third, the rate of production of the first, second, and fourth siRNAs from a long dsRNA substrate were indistinguishable in the presence of ATP, but in the absence of ATP, the rate of siRNA production declined with increasing distance from the end of the dsRNA. Finally, the association of Dicer-2 with a long dsRNA was resistant to dilution provided ATP was present, suggesting that after binding the end of its substrate, Dicer-2 remains bound to the dsRNA and uses ATP energy to reposition itself to produce the next 21 bp siRNA. Translocation along the dsRNA seems a likely mechanism.

Although helicase mutant Dicer-2^{G31R} processed pre-miRNAs as efficiently as wild-type Dicer-2, the mutant was unable to produce even the terminal siRNA from a long RNA duplex under multiple turnover conditions. This suggests an inherent ability of the helicase domain of Dicer-2 to distinguish between long and short substrates. We note that the helicase domain of human Dicer autoinhibits processing of an RNA duplex, and its dsRNA-binding protein partner TRBP, a homolog of R2D2 and Loqs, relieves this inhibition (Ma et al., 2008; Chakravarthy et al., 2010). We hypothesize that *Drosophila* Dicer-2 can occupy two distinct conformations. When inorganic phosphate is low, Dicer-2 assumes a conformation—perhaps similar to the autoinhibited conformation of human Dicer—that can bind and load siRNA. This conformation is unaffected by ATP and, we presume, is involved in promiscuously processing pre-miRNA in vitro. When inorganic phosphate is higher and the enzyme's helicase and

dsRBDs engage its substrate, Dicer-2 assumes a conformation that requires ATP for binding and hydrolysis to process dsRNA.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

Expression and purification of His₆ Dicer 2 or His₆ Dicer 2 and His₆ R2D2 in Sf21 cells was as described (Liu et al., 2003; Tomari et al., 2004). His₆ Dicer 2^{G31R}, His₆ Dicer 2^{D1217,1614N}, and His₆ Dicer 1 were expressed in Sf9 insect cells using the BAC to BAC Baculovirus Expression System (Invitrogen, Carlsbad, CA) and purified from cell lysates by using Ni NTA agarose (QIAGEN, Valencia, CA), HiTrap Q, HiTrap Heparin (GE Healthcare, Pittsburgh), and Superdex 200 gel filtration. Loqs PD was expressed in *Escherichia coli* Rosetta2(DE3), isolated using Ni Sepharose (GE Healthcare), treated with HRV3C protease cleavage to remove the His tag, and purified using HiTrap SP and HiTrap Heparin. Proteins were exchanged into 20 mM HEPES KOH (pH 8.0), 100 mM NaCl, 1 mM tris(2 carboxyethyl)phosphine hydrochloride. Protein concentrations were determined by quantitative amino acid analysis (Keck Biotechnology Resource Laboratory, New Haven, CT). Two different preparations of recombinant Dicer 2 were used in this study. Preparation 1 was used for Table 1 and Figures 2, 3E, 4C, S2B, S3A, S3D, and S5. Preparation 2 was used for Table 2 and Figures 1, 3A–3D, 3F, 3G, 4A, 4B, 5, S1, S2A, S3B, S3C, and S4.

RNA Substrates

DsRNAs were prepared as described (Haley et al., 2003). PCR templates for transcription of sense and antisense RNAs were generated from the EGFP sequence of pN3 eGFP using primers listed in Table S1. Twenty five and 29 bp dsRNAs (Figures S2A, S3B, and S4A and Table S2) were as previously described (Rose et al., 2005). Synthetic RNAs and synthetic *Drosophila* pre *let 7* (Dharmacon, Lafayette, CO) were 5' ³²P radiolabeled using [γ -³²P]ATP (6000 Ci/mmol) (PerkinElmer, Waltham, MA) and T4 polynucleotide kinase (NEB, Ipswich, MA). After gel purification, RNA strands or pre *let 7* were incubated at 65°C for 5 min and then at 25°C for 30 min. Site specifically radiolabeled 120 nt dsRNAs were prepared by DNA splinted ligation (Tables S3 and S4) (Moore and Sharp, 1993; Moore and Query, 2000). To monitor formation of the fourth siRNA in Figure 4B, we used a 120 nt substrate in which the fifth siRNA (the last siRNA generated by Dicer 2 from this substrate) was site specifically ³²P radiolabeled. The fourth siRNA produced corresponds to the sum of the two ³²P radiolabeled cleavage products produced when the dsRNA was cleaved to generate the fourth and fifth siRNAs.

In Vitro RNA Processing

For Dcr 2 +Loqs PD, Loqs PD was first mixed with Dicer 2 or Dicer 2/R2D2 and incubated for 10 min on ice, followed by 5 min at room temperature. Dicing reactions contained 7.5 mM DTT, 3.3 mM magnesium acetate, 0.25% v/v glycerol, 100 mM potassium acetate, 18 mM HEPES KOH (pH 7.4), 15 mM CP, 2.25 μ g CK, and 1 mM ATP; ATP reactions contained 1 mM EDTA but no CK or CP; ATP γ S reactions contained 1 mM ATP γ S only. In Figures 2A–2D, 3, 4B–4E, 5, S3, and S4, +ATP reactions contained no CP or CK; +CP and +CK reactions in Figures 2B and S2B contained 20 mM CP or 2.25 μ g CK. Reactions were assembled on ice and preincubated at 25°C for 5 min before adding RNA. In Figure S4, dilution buffer contained 0.1% NP 40.

Aliquots (1 μ l) of reactions with radiolabeled RNA substrate were quenched by the addition of 25 volumes of formamide loading buffer (98% v/v formamide, 0.1% w/v bromophenol blue and xylene cyanol, 10 mM EDTA), incubated for 5 min at 95°C and analyzed by electrophoresis through a denaturing polyacrylamide 7 M urea gel using 0.5 \times Tris borate EDTA buffer (National Diagnostics, Atlanta). In Figure 5, 200 μ l aliquots from dilution reactions were stopped with 300 mM sodium acetate and 25 mM EDTA, isopropanol precipitated, and dissolved in formamide loading buffer before gel analysis. Gels were exposed to image plates and analyzed with an FLA 5000 and Image Gauge 3.0 software (Fujifilm, Tokyo). In Figure 1, *let 7* and *let 7** strands were detected by northern hybridization with 5' ³²P radiolabeled DNA probes (Table S1).

ATP Hydrolysis

[α - 32 P]ATP (250 nM, 3000 mmol/Ci) (PerkinElmer) was used to monitor hydrolysis. Reactions were stopped with a 25 vol formamide loading dye, spotted onto 20 × 20 cm cellulose plates (EMD, Darmstadt, Germany), and chromatographed in 0.75 M KH₂PO₄ (adjusted to pH 3.3 with H₃PO₄) until the solvent reached the top of the plate. The plate was dried and analyzed by phosphorimager. For Figure S3B, ATP hydrolysis was monitored using the ATP Bioluminescent Assay Kit (Sigma, St. Louis). The reaction was stopped by diluting the sample ten times in H₂O and immediately flash freezing in liquid nitrogen. Samples were stored at -80°C until they were measured. A standard curve spanning at least 100 fold less than and greater than the experimental values was used to determine ATP concentrations.

Rate Analyses

Substrate converted to siRNA versus time was fit to $y = y_0 + A(1 - e^{-kt})$, where $dy/dt = Ake^{-kt}$. When $t = 0$, $dy/dt = Ak$ (Lu and Fei, 2003). Data were fit to the Michaelis-Menten scheme using Visual Enzymics 2008 (Softzyms, Princeton, NJ) for Igor Pro 6.11 (WaveMetrics, Lake Oswego, OR). The rate of substrate consumption was fit for each replicate separately, and two tailed, two sample equal variance t test was used to compare rates (Excel, Microsoft, Seattle). R 2.6.0 software was used for other statistical analyses.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, four tables, and Supplemental References and can be found with this article online at doi:10.1016/j.molcel.2011.03.002.

ACKNOWLEDGMENTS

We thank Yukihide Tomari for help purifying recombinant Dicer 2, members of the Zamore and Hall laboratories and Can Cenik for advice and comments on the manuscript, and Sean Ryder for suggesting we test the effect of phosphate on dicing of pre-miRNA. This work was supported in part by grants from the National Institutes of Health to P.D.Z. (GM62862 and GM65236), the Intramural Research Program of the National Institute of Environmental Health Sciences to T.M.T.H., and a JSPS Research Fellowship for Research Abroad and a Charles A. King Trust Postdoctoral Fellowship to R.F.

Received: May 10, 2010

Revised: January 26, 2011

Accepted: March 3, 2011

Published online: March 17, 2011

REFERENCES

Ameres, S.L., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. (2011). Target RNA directed tailing and trimming purifies the sorting of endo siRNAs between the two *Drosophila* Argonaute proteins. *RNA* 17, 54–63.

Auesukaree, C., Homma, T., Tochio, H., Shirakawa, M., Kaneko, Y., and Harashima, S. (2004). Intracellular phosphate serves as a signal for the regulation of the PHO pathway in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 279, 17289–17294.

Bass, B.L. (2000). Double stranded RNA as a template for gene silencing. *Cell* 101, 235–238.

Beran, R.K., Bruno, M.M., Bowers, H.A., Jankowsky, E., and Pyle, A.M. (2006). Robust translocation along a molecular monorail: the NS3 helicase from hepatitis C virus traverses unusually large disruptions in its track. *J. Mol. Biol.* 358, 974–982.

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, 363–366.

Bianco, P.R., and Kowalczykowski, S.C. (2000). Translocation step size and mechanism of the RecBC DNA helicase. *Nature* 405, 368–372.

Bowers, H.A., Maroney, P.A., Fairman, M.E., Kastner, B., Lüthmann, R., Nilsen, T.W., and Jankowsky, E. (2006). Discriminatory RNP remodeling by the DEAD box protein DED1. *RNA* 12, 903–912.

Burt, C.T., Glonek, T., and Bárány, M. (1976). Analysis of phosphate metabolites, the intracellular pH, and the state of adenosine triphosphate in intact muscle by phosphorus nuclear magnetic resonance. *J. Biol. Chem.* 251, 2584–2591.

Cerutti, L., Mian, N., and Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem. Sci.* 25, 481–482.

Chakravarty, S., Sternberg, S.H., Kellenberger, C.A., and Doudna, J.A. (2010). Substrate specific kinetics of Dicer catalyzed RNA processing. *J. Mol. Biol.* 404, 392–402.

Colmenares, S.U., Buker, S.M., Buhler, M., Diakić, M., and Moazed, D. (2007). Coupling of double stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol. Cell* 27, 449–461.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* 432, 231–235.

Dumont, S., Cheng, W., Serebrov, V., Beran, R.K., Tinoco, I.J., Jr., Pyle, A.M., and Bustamante, C. (2006). RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature* 439, 105–108.

Ercińska, M., Stubbs, M., Miyata, Y., and Ditre, C.M. (1977). Regulation of cellular metabolism by intracellular phosphate. *Biochim. Biophys. Acta* 462, 20–35.

Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer 1. *Cell* 130, 287–297.

Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ line stem cell maintenance requires Loquacious, a double stranded RNA binding domain protein. *PLoS Biol.* 3, e236.

Franks, T.M., Singh, G., and Lykke Andersen, J. (2010). Upf1 ATPase dependent mRNP disassembly is required for completion of nonsense mediated mRNA decay. *Cell* 143, 938–950.

Gan, J., Tropea, J.E., Austin, B.P., Court, D.L., Waugh, D.S., and Ji, X. (2006). Structural insight into the mechanism of double stranded RNA processing by ribonuclease III. *Cell* 124, 355–366.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., and Zamore, P.D. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077–1081.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240.

Haley, B., Tang, G., and Zamore, P.D. (2003). In vitro analysis of RNA interference in *Drosophila melanogaster*. *Methods* 30, 330–336.

Halls, C., Mohr, S., Del Campo, M., Yang, Q., Jankowsky, E., and Lambowitz, A.M. (2007). Involvement of DEAD box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis dependent and independent mechanisms, and general RNA chaperone activity. *J. Mol. Biol.* 365, 835–855.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha DGCR8 complex in primary microRNA processing. *Genes Dev.* 18, 3016–3027.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha DGCR8 complex. *Cell* 125, 887–901.

- Hartig, J.V., Esslinger, S., Böttcher, R., Saito, K., and Förstemann, K. (2009). Endo siRNAs depend on a new isoform of *loquacious* and target artificially introduced, high copy sequences. *EMBO J.* *28*, 2932–2944.
- Hartig, J.V., and Förstemann, K. (2011). Loqs PD and R2D2 define independent pathways for RISC generation in *Drosophila*. *Nucleic Acids Res.* Published online January 17, 2011.
- Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. (2005). Dicer 1 and R3D1 L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* *19*, 1674–1679.
- Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* *453*, 793–797.
- 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*, 2654–2659.
- Kini, H.K., and Walton, S.P. (2007). In vitro binding of single stranded RNA by human Dicer. *FEBS Lett.* *581*, 5611–5616.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* *425*, 415–419.
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for *Drosophila* Dicer 1 and Dicer 2 in the siRNA/miRNA silencing pathways. *Cell* *117*, 69–81.
- Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* *426*, 465–469.
- Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3' end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* *11*, 576–577.
- Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* *301*, 1921–1925.
- Liu, X., Jiang, F., Kalidas, S., Smith, D., and Liu, Q. (2006). Dicer 2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA* *12*, 1514–1520.
- Lu, W.P., and Fei, L. (2003). A logarithmic approximation to initial rates of enzyme reactions. *Anal. Biochem.* *316*, 58–65.
- Ma, E., MacRae, I.J., Kirsch, J.F., and Doudna, J.A. (2008). Autoinhibition of human dicer by its internal helicase domain. *J. Mol. Biol.* *380*, 237–243.
- Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang specific small interfering RNA recognition by the PAZ domain. *Nature* *429*, 318–322.
- MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.* *14*, 934–940.
- Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double stranded RNA processing by Dicer. *Science* *311*, 195–198.
- Matranga, C., and Pyle, A.M. (2010). The double stranded RNA dependent ATPase DRH 3: insight into its role in RNA silencing in *Caenorhabditis elegans*. *J. Biol. Chem.* *285*, 25363–25371.
- Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H., and Siomi, M.C. (2010). Molecular mechanisms that funnel RNA precursors into endogenous small interfering RNA and microRNA biogenesis pathways in *Drosophila*. *RNA* *16*, 506–515.
- Moore, M.J., and Query, C.C. (2000). Joining of RNAs by splinted ligation. *Methods Enzymol.* *317*, 109–123.
- Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre mRNA splicing. *Nature* *365*, 364–368.
- Myong, S., Cui, S., Cornish, P.V., Kirchofer, A., Gack, M.U., Jung, J.U., Hopfner, K.P., and Ha, T. (2009). Cytosolic viral sensor RIG I is a 5' triphosphate dependent translocase on double stranded RNA. *Science* *323*, 1070–1074.
- Nakamura, M., Ando, R., Nakazawa, T., Yudazono, T., Tsutsumi, N., Hatanaka, N., Ohgake, T., Hanaoka, F., and Eki, T. (2007). Dicer related *drh 3* gene functions in germ line development by maintenance of chromosomal integrity in *Caenorhabditis elegans*. *Genes Cells* *12*, 997–1010.
- Nykänen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* *107*, 309–321.
- Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008a). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat. Struct. Mol. Biol.* *15*, 581–590.
- Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008b). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* *453*, 803–806.
- Patel, S.S., and Donmez, I. (2006). Mechanisms of helicases. *J. Biol. Chem.* *281*, 18265–18268.
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., and Sontheimer, E.J. (2004). A Dicer 2 dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* *117*, 83–94.
- Pham, J.W., and Sontheimer, E.J. (2005). Molecular requirements for RNA induced silencing complex assembly in the *Drosophila* RNA interference pathway. *J. Biol. Chem.* *280*, 39278–39283.
- Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu. Rev. Biophys.* *37*, 317–336.
- Rivera, M.A., and Blackburn, E.H. (2004). Processive utilization of the human telomerase template: lack of a requirement for template switching. *J. Biol. Chem.* *279*, 53770–53781.
- Rose, S.D., Kim, D.H., Amarzguioui, M., Heidel, J.D., Collingwood, M.A., Davis, M.E., Rossi, J.J., and Behlke, M.A. (2005). Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res.* *33*, 4140–4156.
- Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2005). Processing of pre microRNAs by the Dicer 1 Loquacious complex in *Drosophila* cells. *PLoS Biol.* *3*, e235.
- Seidel, R., Bloom, J.G., Dekker, C., and Szelczkun, M.D. (2008). Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J.* *27*, 1388–1398.
- Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Mol. Biol.* *10*, 1026–1032.
- Takeshita, D., Zenno, S., Lee, W.C., Nagata, K., Saigo, K., and Tanokura, M. (2007). Homodimeric structure and double stranded RNA cleavage activity of the C terminal RNase III domain of human dicer. *J. Mol. Biol.* *374*, 106–120.
- Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. (2008). Pseudogene derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* *453*, 534–538.
- Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of *Drosophila* small silencing RNAs. *Cell* *130*, 299–308.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004). A protein sensor for siRNA asymmetry. *Science* *306*, 1377–1380.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* *453*, 539–543.
- Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S., and Bass, B.L. (2010). Dicer's helicase domain is required for accumulation of some, but not all, *C. elegans* endogenous siRNAs. *RNA* *16*, 893–903.
- Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature* *426*, 468–474.

- Yang, N., and Kazazian, H.H.J., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* *13*, 763–771.
- Ye, X., Paroo, Z., and Liu, Q. (2007). Functional anatomy of the *Drosophila* microRNA generating enzyme. *J. Biol. Chem.* *282*, 28373–28378.
- 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.
- Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* *21*, 5875–5885.
- Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* *118*, 57–68.
- Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of *Drosophila* endo siRNAs depends on a specific Loquacious isoform. *RNA* *15*, 1886–1895.

Supplemental Information

Molecular Cell, Volume 42

Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease

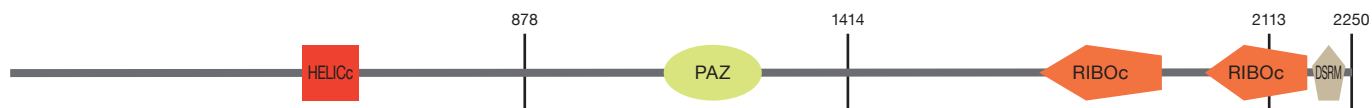
Elif Sarinay Cenik, Ryuya Fukunaga, Gang Lu, Robert Dutcher, Yeming Wang, Traci M. Tanaka Hall, and Phillip D. Zamore

Figure S1, related to Figure 1. Dicer-1, but not Dicer-2, Generates Intermediates when Processing Long dsRNA

(A) Domain Structures of *Drosophila* Dicer-1 and Dicer-2 and human Dicer. Domain structures were adapted from SMART protein domain database (Schultz et al., 1998; Ponting et al., 1999). The DEXDc motif in these proteins has the sequence DECH.

(B) 25 nM internally ³²P-radiolabeled 515 bp dsRNA was incubated with either 225 nM Dicer-1 or 5.4 nM Dicer-2. The top gel was run to visualize both the uncleaved substrate and the 21–22 bp siRNA products; the lower gel was run to resolve the intermediates (arrowheads) generated by Dicer-1.

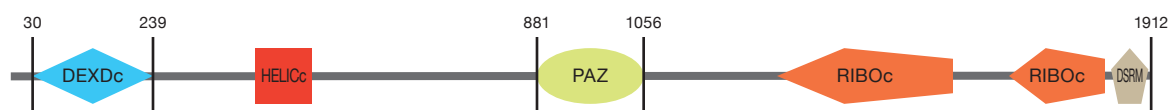
Drosophila Dicer-1




Drosophila Dicer-2



Human Dicer



 DEXDc DEAD-like helicase superfamily

 HELICc Helicase superfamily C-terminal domain

 PAZ PAZ domain

 RIBOc Ribonuclease III family

 DSRM Double-stranded RNA binding motif

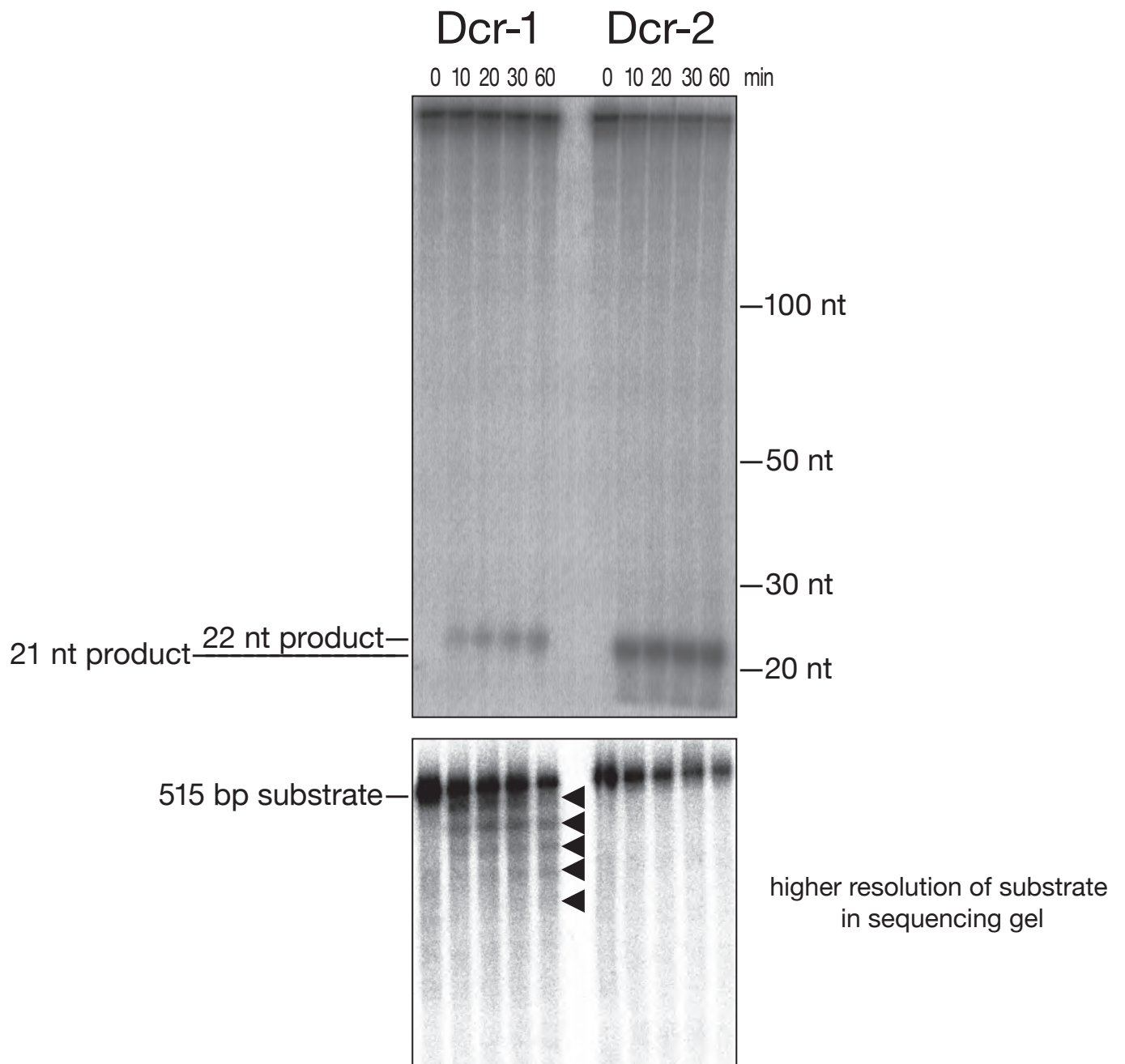
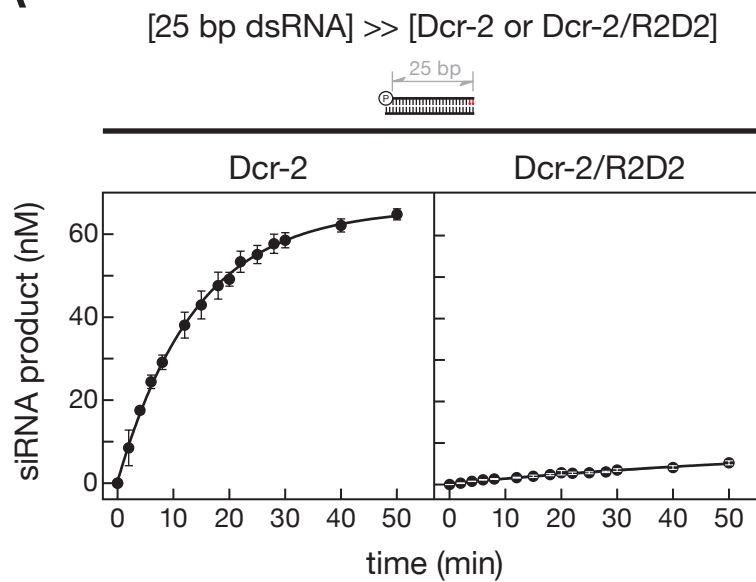


Figure S2, related to Figure 2. R2D2 Inhibits Dicer-2 Processing of a 25 bp dsRNA Substrate and Creatine Phosphate Inhibits Dicer-2 Processing of a Long dsRNA Bearing 5' Triphosphate termini

(A) siRNA production by Dcr-2 (5.4 nM) or Dcr-2/R2D2 heterodimer (12 nM) was measured for a 25 bp dsRNA (100 nM). Red indicates deoxynucleotides.

(B) A 5' triphosphorylated 316 bp dsRNA (100 nM) bearing 2 nt 5' overhanging ends was incubated with Dicer-2 (8 nM) in the presence or absence of ATP, creating kinase (CK), or creatine phosphate (CP). Values are mean \pm standard deviation for three independent experiments.

A



B

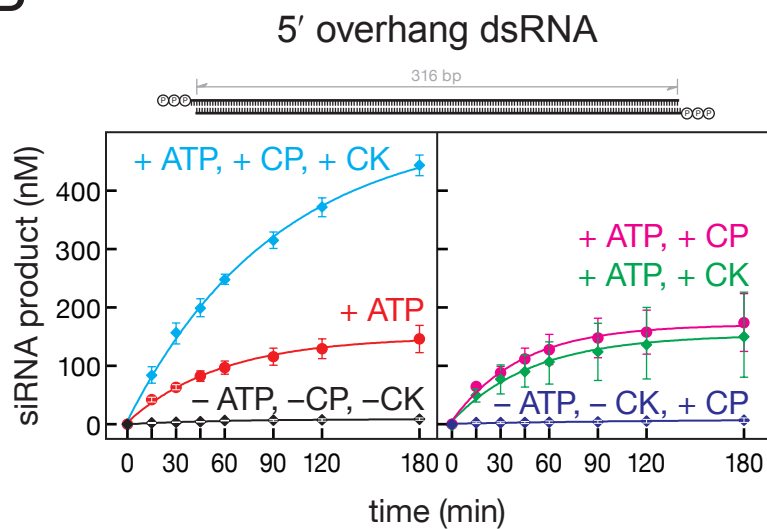


Figure S3, related to Figure 3. ATP Hydrolysis is Required for Efficient Long dsRNA Processing by Dicer-2

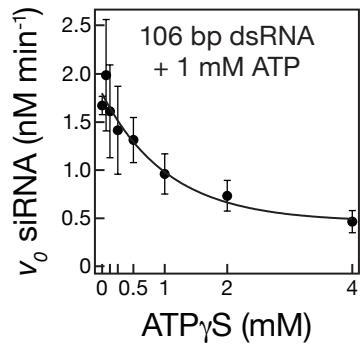
(A) siRNA formation from 5' monophosphorylated, internally ^{32}P -radiolabeled, blunt ended, 106 bp dsRNA was monitored in the presence of 1 mM ATP in the presence of increasing concentrations of ATP γ S. Initial rates of conversion of dsRNA to siRNA were plotted as a function of ATP γ S concentration.

(B) ATP consumption was measured over time for a 40 bp (420 nM), 60 bp (280 nM), 106 bp (158 nM), 208 bp (81 nM), 345 bp (49 nM), and a 558 bp (30 nM) dsRNA substrate. Initial rates (v_0) were calculated from the fitted exponential curves and used in Figure 3F.

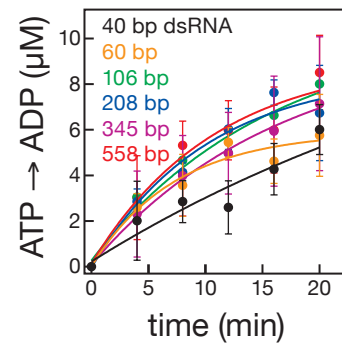
(C) The hydrolysis of ATP to ADP was measured for Dicer-2 alone (5.4 nM) or with 25 nt single-stranded (ss) synthetic RNA, 25 nt synthetic ssDNA, 25 bp synthetic RNA or RNA-DNA duplex. ATP hydrolysis by Dicer-2 (5.4 nM) in the presence of a 515 bp dsRNA is shown in gray for reference. Substrate concentrations were chosen to achieve equivalent nt·mole·l $^{-1}$: ssRNA and ssDNA, 4.4 μM ; RNA and DNA-RNA duplex, 2.2 μM ; and 515 bp dsRNA, 100 nM.

(D) The hydrolysis of ATP to ADP was measured for Dicer-2 (5 nM) with prelet-7 (5 μM) or with 558 bp dsRNA (100 nM) in the presence of 1 mM ATP. Values are mean \pm standard deviation for three independent experiments.

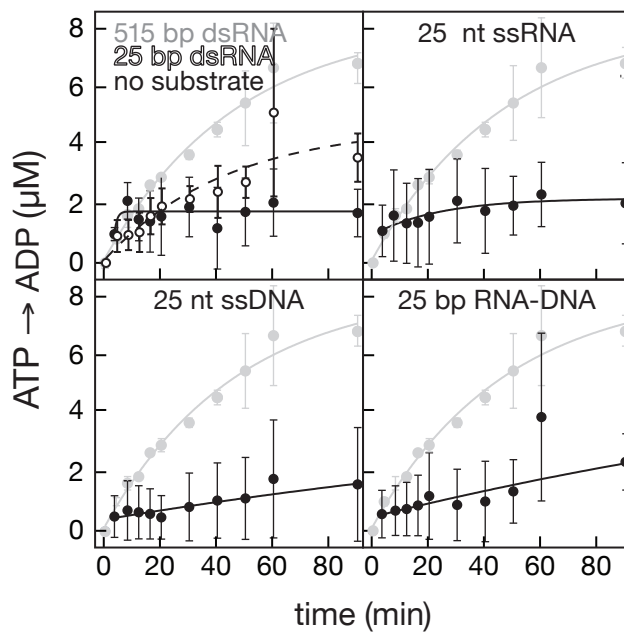
A



B



C



D

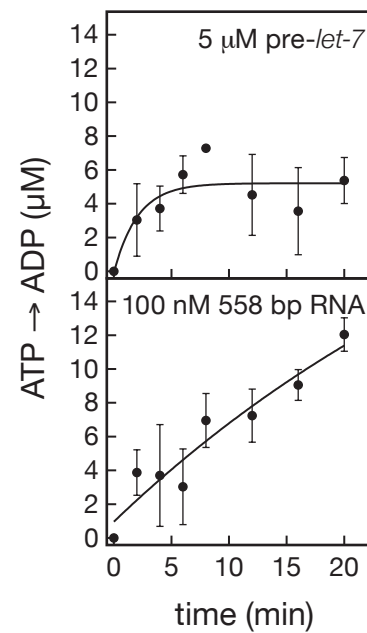


Figure S4, related to Figure 4. The Rate of siRNA Production for Different dsRNA Termini: Two 5' Deoxynucleotides Block Entry of Dicer-2, but Blunt and 3' Overhanging Ends are Processed Similarly

(A) Twenty-five and 29 bp dsRNA substrates (73 nM) containing two 2' deoxycytidine nucleotides at the 3' end were incubated with 5.4 nM Dicer-2, and total siRNA production measured at 45 min. For each substrate, the 5' ribonucleotide of the 5' end or the end that contained 3' deoxynucleotides on the complementary strand was ³²P-radiolabeled to allow measurement of siRNA production from that end alone.

(B) Site-specifically ³²P-radiolabeled 120 bp dsRNAs (100 nM) were incubated with Dicer-2 (5.4 nM) in the presence of 1 mM ATP to detect production of the first, second and fourth siRNAs as diagrammed here and in Figure 4A. Right, the substrate end where Dicer-2 enters was either blunt (top) or had a 2 nt 3' overhang (bottom).

(C) Production of the initial siRNA (i.e., the first siRNA) by Dicer-2^{G31R} (100 nM) from a 118 bp dsRNA (10 nM) bearing a 2 nt 3' overhanging end was measured in the presence or absence of ATP or in the presence of ATP γ S (see also Figures 4A and S4A).

Panels (B) and (C) report mean \pm standard deviation for three independent experiments.

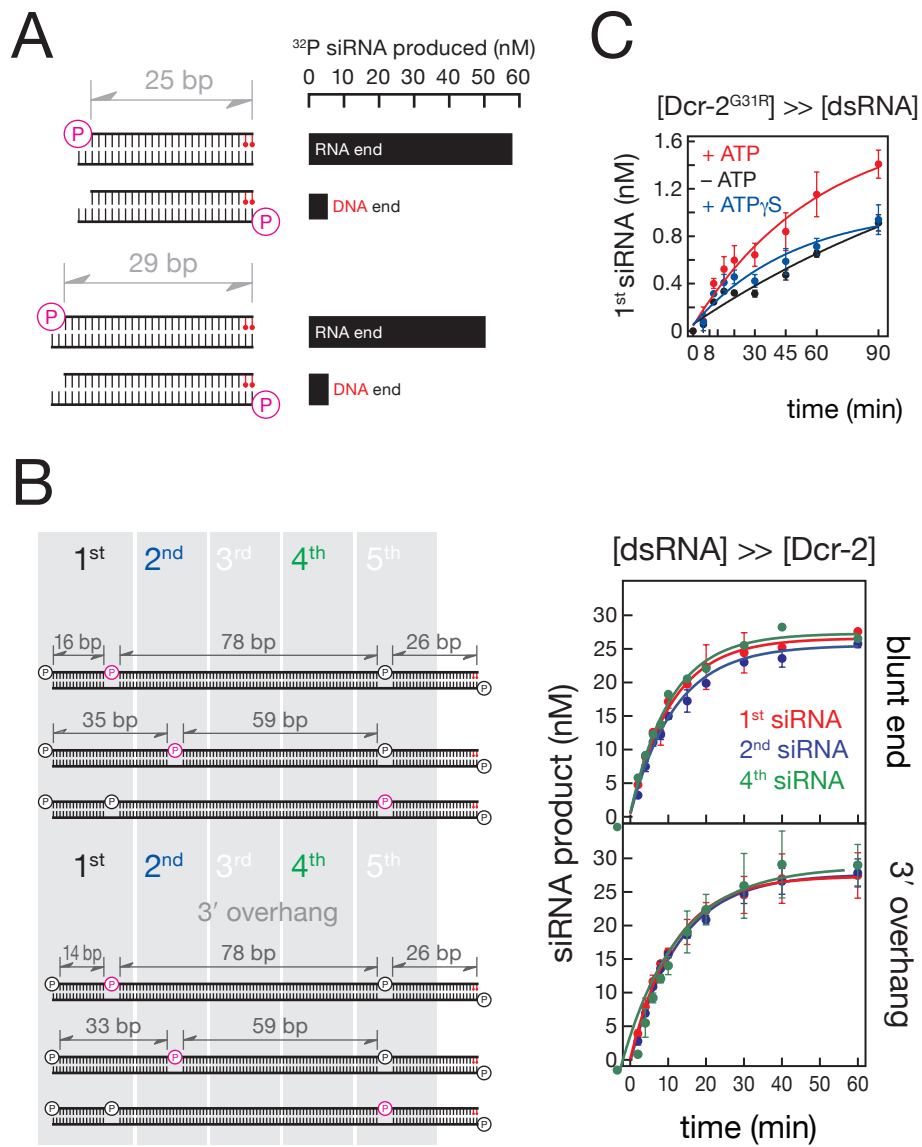


Figure S5, related to Table 1. Michaelis-Menten analysis of Dicer-2, Dicer-2 + Loqs-PD, and Dicer-2/R2D2 using a 515 bp dsRNA substrate

The initial rate (velocity) of substrate processing by 2 nM Dicer-2 , 3 nM Dicer-2/R2D2, and 2 nM Dicer-2 + 2 nM Loqs-PD was measured for increasing concentrations of a 515 bp long dsRNA in the presence of 1 mM ATP. The data were fit to the Michaelis-Menten equation. V_{max} corresponds to the rate of complete conversion of a molecule of substrate into 24 siRNAs at saturating dsRNA concentration. Values are mean \pm standard deviation for three independent experiments.

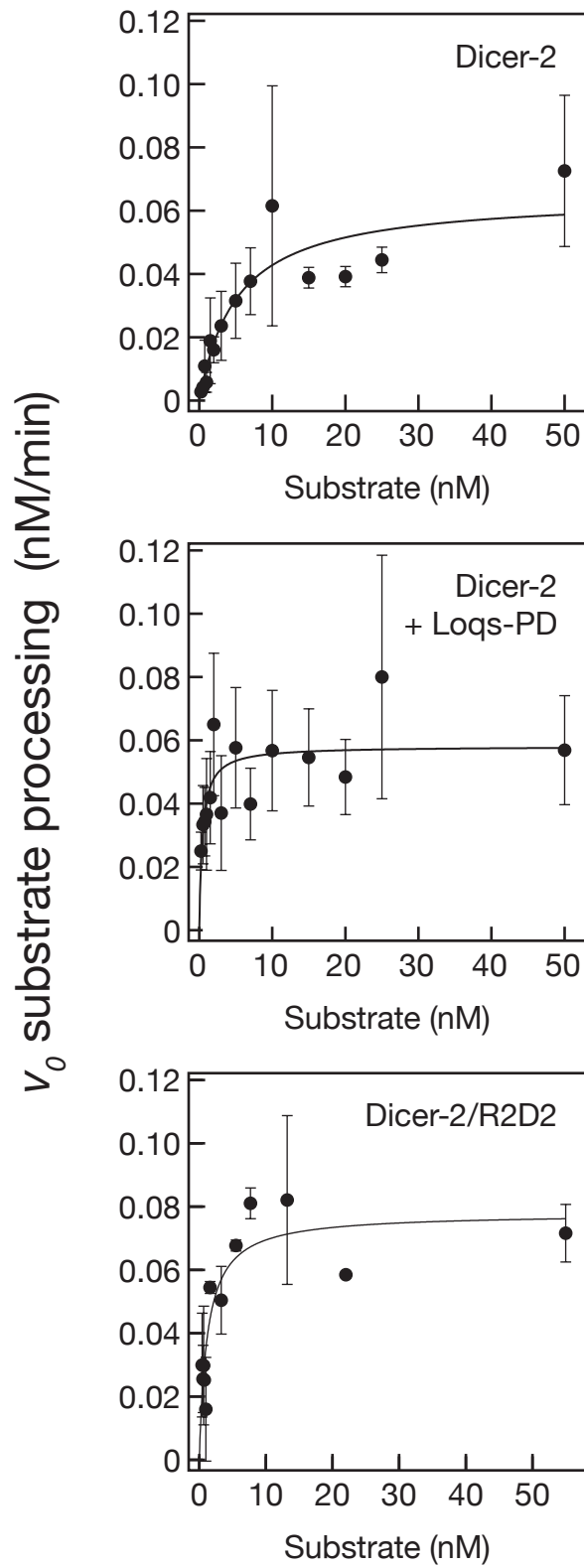


Table S1. Synthetic DNA Oligonucleotides. m, 2'-O-methyl ribose.

| | |
|---|---|
| Forward primer for 515 bp product for sense RNA from pEGFP-N3 | 5'-GCG TAA TAC GAC TCA CTA TAG GAC TCA GAT CTC GAG CTC AAG-3' |
| Reverse primer for 515 bp product for sense RNA from pEGFP-N3 | 5'-GCT GTT GTA GTT GTA CTC CAG-3' |
| Forward primer for 515 bp product for antisense RNA from pEGFP-N3 | 5'-GCG TAA TAC GAC TCA CTA TAG GCT GTT GTA GTT GTA CTC CAG-3' |
| Reverse primer for 515 bp product for antisense RNA from pEGFP-N3 | 5'-GAC TCA GAT CTC GAG CTC AAG-3' |
| Forward primer for 316 bp product for sense RNA from pEGFP-N3 | 5'-GCGTAATACGACTCACTATAGGGCCACAAGTTCAGCGTGTCC-3' |
| Reverse primer for 316 bp product for sense RNA from pEGFP-N3 | 5'-TCGATGCCCTTCAGCTCG-3' |
| Forward primer for 316 bp product for antisense RNA from pEGFP-N3 | 5'-GCGTAATACGACTCACTATAGTCGATGCCCTTCAGCTCG-3' |
| Reverse primer for 316 bp product for antisense RNA from pEGFP-N3 | 5'-GCCACAAGTTCAGCGTGTCC-3' |

| | |
|---|--|
| <i>let-7</i> probe for Northern hybridization | 5'-CTA TAC AAC CTA CTA CCT CAA-3' |
| <i>let-7*</i> probe for Northern hybridization | 5'-AAA GCT AGC ACA TTG TAT AGT-3' |
| Forward primer for 558, 345, 208, 106 bp product for sense RNA from pEGFP-N3 (Figure 3G) | 5'-TAA TAC GAC TCA CTA TAG GCA AGC TGA CCC TGA AGT TC-3' |
| Reverse primer for 558 bp product for sense RNA from pEGFP-N3 | 5'-mGmGT CAC GAA CTC CAG CAG GAC-3' |
| Forward primer for 558 bp product for anti-sense RNA from pEGFP-N3 | 5'-TAA TAC GAC TCA CTA TAG GTC ACG AAC TCC AGC AGG AC-3' |
| Reverse primer for 558, 345, 208, 106 bp product for anti-sense RNA from pEGFP-N3 | 5'-mGmGC AAG CTG ACC CTG AAG TTC-3' |
| Reverse primer for 345 bp product for sense RNA from pEGFP-N3 | 5'-mGmGC CAT GAT ATA GAC GTT GTG-3' |
| Forward primer for 345 bp product for anti-sense RNA from pEGFP-N3 | 5'-TAA TAC GAC TCA CTA TAG GCC ATG ATA TAG ACG TTG TG-3' |
| Reverse primer for 208 bp product for sense RNA from pEGFP-N3 | 5'-mGmGG TCT TGT AGT TGC CGT CGT-3' |

| | |
|--|--|
| Forward primer for 208 bp product for anti-sense RNA from pEGFP-N3 | 5'-TAA TAC GAC TCA CTA TAG GGT CTT GTA GTT GCC GTC GT-3' |
| Reverse primer for 106 bp product for sense RNA from pEGFP-N3 | 5'-mGmGT AGC GGC TGA AGC ACT GCA-3' |
| Forward primer for 106 bp product for sense RNA from pEGFP-N3 (Figure 2) | 5'-GCG TAA TAC GAC TCA CTA TAG GGC CAC AAG TTC AGC GTG TCC-3' |
| Reverse primer for 106 bp product for sense RNA from pEGFP-N3 | 5'-mGmGG CCA GGG CAC GGG CAG CTT GCC G-3' |
| Forward primer for 106 bp product for anti-sense RNA (blunt end) from pEGFP-N3 (Figure 2) | 5'-mGmGG CCA CAA GTT CAG CGT GTC C-3' |
| Reverse primer for 106 bp product for anti-sense RNA (blunt end) from pEGFP-N3 (Figure 2) | 5'-GTA CTT AAT ACG ACT CAC TAT AGG GCC AGG GCA CGG GCA GCT TGC CG-3' |
| Forward primer for 104 bp product for anti-sense RNA (3'overhang) from pEGFP-N3 (Figure 2) | 5'-GTA CTT AAT ACG ACT CAC TAT AGC CAG GGC ACG GGC AGC TTG CCG-3' |
| Reverse primer for 104 bp product for sense RNA (3'overhang) from pEGFP-N3 (Figure 2) | 5'-mTmTG GGC CAC AAG TTC AGC GTG TCC-3' |

Table S2. Synthetic RNA Oligonucleotides. d, deoxy ribose.

| | |
|---|--|
| 25 nt sense RNA for 25 bp dsRNA | 5'-ACC CUG AAG UUC AUC UGC ACC ACdC dG-3' |
| 27 nt anti-sense RNA for 25 bp dsRNA | 5'-CGG UGG UGC AGA UGA ACU UCA GGG UCA -3' |
| 29 nt sense RNA for 29 bp dsRNA | 5'-ACC CUG AAG UUC AUC UGC ACC GUC CAC dCdG-3' |
| 31 nt anti-sense RNA for 29 bp dsRNA | 5'-CGG UGG ACG GUG CAG AUG AAC UUC AGG GUC A-3' |
| <i>pre-let 7</i> | 5'-UGA GGU AGU AGG UUG UAU AGU AGU AAU UAC ACA UCA UAC UAU ACA AUG UGC UAG CUU UCU-3' |

Table S3. DNA Templates for Transcription and DNA Splints for RNA Ligation, used in Figures 3 and 4. m, 2'-O-methyl ribose.

| | |
|--|--|
| DNA template to transcribe antisense RNA strand in Figure 4 | 5'-ACT CCT CAA CAA ATC ATA AAC TAC AAT ATA CAT CAA TAC GAC ATT ACC CTC ACA ATC AAT CAT ACA ACC ATC CCT AAA GAC CAA CAG CAC CCC ACG ATC AAG AAT AAG AAC TAT AAT CCC TAT AGT GAG TCG TAT TAC GC-3' |
| T7 RNAP promoter sequence annealed to the DNA template for transcription | 5'-GCG TAA TAC GAC TCA CTA TAG-3' |
| DNA splint for RNA ligation to generate sense strand in Figure 4 | 5'-GGG ATT ATA GTT CTT ATT CTT GAT CGT GGG GTG CTG TTG GTC TTT AGG GAT GGT TGT ATG ATT GAT TGT GAG GGT AAT GTC GTA TTG ATG TAT ATT GTA GTT TAT GAT TTG TTG AGG AGT-3' |
| DNA template to transcribe 40 nt sense RNA in Figure 3G | 5'-mGGC TTC ATG TGG TCG GGG TAG CGG CTG AAG CAC TGC ACG CCT ATA GTG AGT CGT ATT ACG C-3' |
| DNA template to transcribe 40 nt anti-sense RNA in Figure 3G | 5'-mGGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CCT ATA GTG AGT CGT ATT ACG C-3' |
| DNA template to transcribe 60 nt sense RNA in Figure 3G | 5'-mGGT GAA CTT CAG GGT CAG CTT GCT TCA TGT GGT CGG GGT AGC GGC TGA AGC ACT GCA CGC CTA TAG TGA GTC GTA TTA CGC-3' |
| DNA template to transcribe 60 nt anti-sense RNA in Figure 3G | 5'-mGGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAA GCT GAC CCT GAA GTT CAC CTA TAG TGA GTC GTA TTA CGC-3' |

Table S4. Synthetic RNA Oligonucleotides Assembled by Splinted Ligation into Site-Specifically Labeled Sense-Strand RNA and Used in Figure 4 and S5.

| | |
|--------|--|
| RNA #1 | 5'-ACU CCU CAA CAA AUC A-3' |
| RNA #2 | 5'-UAA ACU ACA AUA UAC AUC AAU ACG ACA UUA CCC UCA CAA UCA AUC AUA CAA CCA UCC CUA AAG ACC AAC AG CAC CCC A-3' |
| RNA #3 | 5'-CGA UCA AGA AUA AGA ACU AUA AUC dCdC-3' |
| RNA #4 | 5'-ACU CCU CAA CAA AUC AUA AAC UAC AAU AUA CAU CA-3' |
| RNA #5 | 5'-AUA CGA CAU UAC CCU CAC AAU CAA UCA UAC AAC CAU CCC UAA AGA CCA ACA GCA CCC CA-3' |
| RNA #6 | 5'-UCC UCA ACA AAU CA-3' |
| RNA #7 | 5'-UCC UCA ACA AAU CAU AAA CUA CAA UAU ACA UCA-3' |

SUPPLEMENTAL REFERENCES

- Marchler-Bauer, A. et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37, D205-10.
- Ponting, C. P., Schultz, J., Milpetz, F., and Bork, P. (1999). SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res* 27, 229-232.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95, 5857-5864.
- Theis, K., Chen, P. J., Skovvaga, M., Van Houten, B., and Kisker, C. (1999). Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *EMBO J* 18, 6899-6907.

Why did you move back to China?

First of all, I feel extremely privileged to be educated and trained at the University of Minnesota. There are many superb scientists studying vision and brain imaging there. My collaboration with them led to some intriguing findings on visual adaptation, unconscious visual processing, and contextual modulation in early visual cortical areas. From them I learned not just experimental skills, but also various distinct perspectives on these same scientific questions.

Career-wise, working in China is very attractive to me. Government research funds in China have been growing at an annual rate of more than 20%. Ample funding allows me to explore and carry out much larger and more risky projects. At Peking University, I have been enjoying working with the country's most intelligent and hardworking students. In addition, I am a big ping-pong and soccer fan, and living in China gives me a lot more opportunity to enjoy these sports.

Tell us something about neuroscience in China?

Neuroscience in China has a tradition of excellence. I would like to mention the founders of modern Chinese neuroscience — Robert Kho-Seng Lim, Te-Pei Feng and Hsiang-Tung Chang. Lim and Feng were members of the US National Academy of Sciences. Lim carried out pioneering work on the physiology of neuromuscular junction and synaptic plasticity. Interested readers might want to read a chapter published in the *Annual Review of Neuroscience* in 1988 (11, 1–12) about Lim's career development and the early history of neuroscience in China. Chang was one of the pioneers of studying dendritic potentials and among the first to recognize the functional significance of dendrites in the central nervous system.

Neuroscience in China has grown steadily since the 1920s, and started to flourish in the 1990s. In 1995, the Chinese Neuroscience Society was founded and it now has more than 2500 members. Major neuroscience research programs are located in the Chinese Academy of Sciences, Peking University, Fudan University, Beijing Normal University, University of Science

and Technology of China, many medical universities and institutes, and many more places. Research areas include molecular, cellular and developmental neurobiology, systems and computational neuroscience, as well as cognitive and behavioral neuroscience. Chinese neuroscientists are making their contribution to the development of this field *on a par* with their peers in the international arena, as demonstrated by their frequent publications in almost all prestigious journals (including *Current Biology*).

And what about psychology in China? Psychology, on the other hand, took a slightly different turn. In 1917, the first psychology laboratory in China was set up at Peking University, under the guidance of the university president Yuen-Pei Tsai. Tsai studied psychology with Wilhelm Wundt when he was in Germany. Unfortunately, the development of psychology was suppressed for a long time, even halted during the Cultural Revolution from 1966 to 1976. This is because psychology was criticized as a pseudo-science. In 1981, only four universities had a psychology department. Interestingly, the turning point for the development of psychology was also in the 1990s, almost in parallel with the time when neuroscience started to thrive. Up to now, there are more than two hundred psychology departments/institutes in China. Founded in 1921, the Chinese Psychological Society now has about 8000 members. Psychological research in China covers almost all basic and applied fields. Brain and cognitive science has been identified as one of the eight research frontiers by the central government in 2006 and two national key laboratories have been set up targeting fundamental issues in this area. The rapid development of psychology (and neuroscience) in China is partly due to the nation's economic boom and thus a rapid growth in research funds. I feel honored to live in this era and to experience the dramatic (positive) changes of science and research in China.

Department of Psychology, Peking University, Beijing 100871, P.R. China.
E-mail: ffang@pku.edu.cn

Quick guides

Argonaute proteins

Elif Sarinay Cenik¹
and Phillip D. Zamore^{1,2}

What are Argonaute proteins?

Argonaute proteins form an evolutionarily conserved family whose members silence gene expression in pathways such as RNA interference (RNAi). Argonaute family proteins can be divided into AGO and PIWI proteins (Figure 1). Both types of Argonaute proteins bind 21–35 nt long small RNA guides whose sequence identifies the genes to be silenced. Argonaute–small-RNA complexes can repress the transcription of genes, target mRNAs for site-specific cleavage or general degradation, or block mRNA translation into protein. AGO proteins bind ~21 nt small interfering RNAs (siRNAs) and 21–23 nt microRNAs (miRNAs). Both siRNAs and miRNAs are cut from double-stranded RNA precursors by RNase III enzymes such as Dicer. AGO proteins are essential for development and differentiation, and in most plants and animals, defend cells against viral infection. In contrast, PIWI proteins bind 23–30 nt PIWI-interacting RNAs (piRNAs), whose production does not appear to involve double-stranded RNA or Dicer. piRNAs are unique to animals, where they repress transposon expression and ensure the successful production of sperm and eggs.

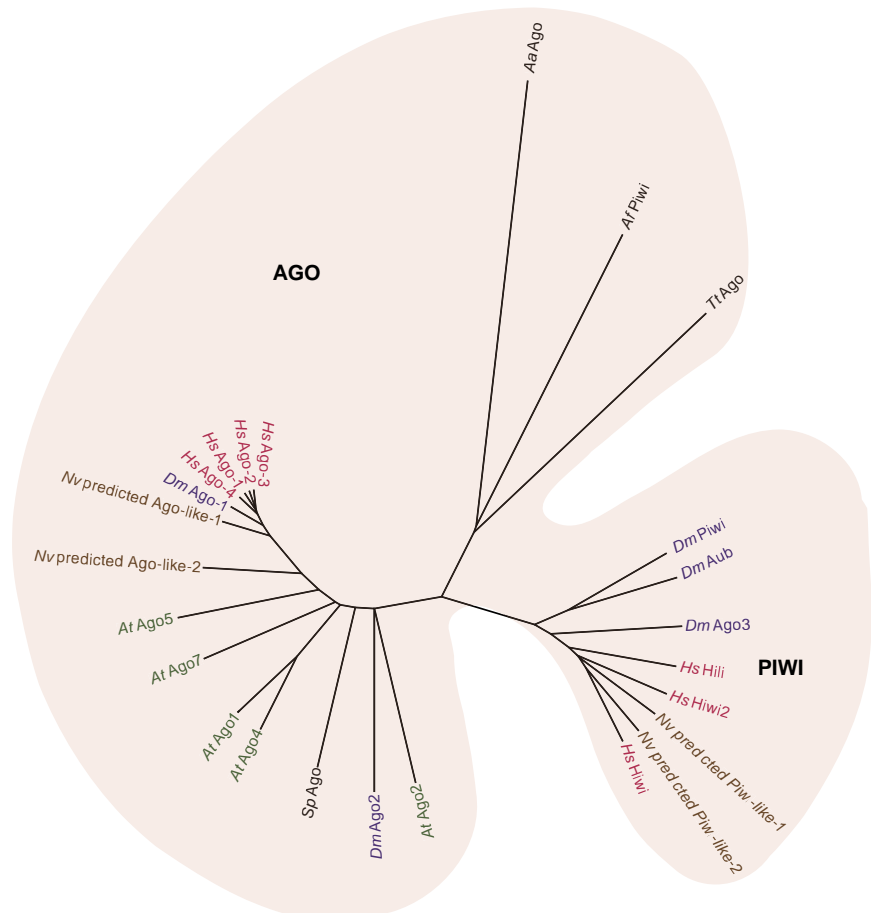
How do Argonautes function? An Argonaute protein plus its small RNA guide compose the RNA-induced silencing complex (RISC). RISC complexes can also contain additional proteins thought to extend the functions of Argonautes or to direct RISC to specific sub-cellular locations. The simplest, and likely ancestral, Argonaute function is endonucleolytic cleavage of its RNA target at a single phosphodiester bond. The structure of Argonaute ensures that the bond cleaved always lies between the target nucleotides paired to the tenth and eleventh nucleotides of the guide RNA. Increasingly, Argonaute aficionados refer to these nucleotides as g10 and g11 for the small RNA and t10 and t11 for the target, viewing both the guide (g) and the target (t) from

the 5'-to-3' perspective of the small silencing RNA.

What does Argonaute look like?

No three-dimensional structure is available for an entire eukaryotic Argonaute protein, but a series of structures of eubacterial and archaeal Argonautes, as well as structures of individual domains of eukaryotic Argonaute proteins, together reveal broad principles that hold true in fungi, plants, and animals. All Argonaute/Piwi proteins comprise three key domains: PIWI, MID, and PAZ (Figure 2). The Argonaute endonuclease, which requires Mg²⁺ to slice a target RNA into products bearing 3' hydroxyl and 5' phosphate groups, resides in the carboxy-terminal PIWI domain (Figure 2A). The PIWI domain resembles another nuclease, RNase H, a DNA-guided ribonuclease. Like RNase H, the PIWI domain contains three negatively charged, evolutionarily conserved amino acids — typically aspartate-aspartate-glutamate (DDE) — that form a Mg²⁺-binding catalytic triad. Unlike RNase H, the small RNA guide remains stably bound to the Argonaute protein through many rounds of target cleavage.

The amino-terminal PAZ domain uses its oligonucleotide-binding (OB) fold to secure the 3' end of the small RNA guide strand to Argonaute (Figure 2B). A conserved hydrophobic cavity within the PAZ domain recognizes the characteristic two-nucleotide, 3' overhanging end of the guide-passenger siRNA or miRNA/miRNA* duplex generated by Dicer. The PAZ domain can also detect chemical modification of the 3' end of the small RNA. In plants, all small RNAs bear a 2'-O-methyl modification at their final ribose sugar. In animals, piRNAs, but not typically miRNAs or siRNAs, are 2'-O-methyl modified, although in insects siRNAs bound to Argonaute2 are also 2'-O-methyl modified. The PAZ domains of human PIWI proteins (Hiwi1, Hiwi2, Hili) bind more tightly to the two-nucleotide, 3' overhanging end of a dsRNA when it bears a 2'-O-methyl than when the end is 2' hydroxyl. Conversely, the PAZ domain of the human AGO protein Ago1 prefers a 2' hydroxyl. Comparison of the structures of single-stranded RNA bound to the PAZ domain from mouse Miwi with that of the PAZ domain from human Ago1 suggests that the wider RNA-binding cleft of the Miwi PAZ domain better



Current Biology

Figure 1. Sequence relationships among AGO and PIWI sub-families of Argonaute proteins. Protein sequences from the thermophilic bacterium *Aquifex aeolicus* (Aa), the sulfur-reducing archaea *Archaeoglobus fulgidus* (Af), the eubacteria *Thermus thermophilus* (Tt), the yeast *Schizosaccharomyces pombe* (Sp), the plant *Arabidopsis thaliana* (At), and the animals *Drosophila melanogaster* (Dm), *Homo sapiens* (Hs), and *Nematostella vectensis* (Nv; sea anemone) were aligned using MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>) and displayed using Archaeopteryx 0.957beta (<http://www.phylosoft.org/archaeopteryx/>).

accommodates a terminal 2'-O-methyl group than does that of Ago1. However, *Drosophila* Ago1 — whose guide RNAs *in vivo* are thought always to end with a 2' OH — can accept 2'-O-methyl guides, at least *in vitro*.

Finally, the MID domain anchors the 5' monophosphate of a small silencing RNA to the Argonaute protein, securing the guide through multiple cycles of target cleavage (Figure 2A). *In vitro* studies suggest that 5' phosphate binding helps align the small RNA on the surface of Argonaute, ensuring that the correct bond of the target is positioned in the endonuclease active site. The MID domain also participates in sorting small RNAs among various Argonaute paralogs according to the identity of the first nucleotide of the RNA guide. For example, in *Arabidopsis* Ago1 prefers small RNAs that begin

with uridine, Ago2 and Ago4 prefer an initial adenosine, and Ago5 prefers cytosine. Swapping the MID domain of one *Arabidopsis* Argonaute for that of another exchanges the initial nucleotide that is favored. Such nucleotide preferences likely arise from subtle differences in the amino acid side chains near the 5' phosphate-binding pocket: structures of the human Ago2 MID domain bound to nucleoside monophosphates suggest that Ago2 favors small RNAs that begin with uridine or adenosine because a rigid loop in the domain inhibits binding to cytidine or guanosine.

How do Argonautes obtain their small RNA guides? On their own, AGO proteins do not accept an siRNA or miRNA/miRNA* duplex, yet all available evidence suggests that AGO proteins

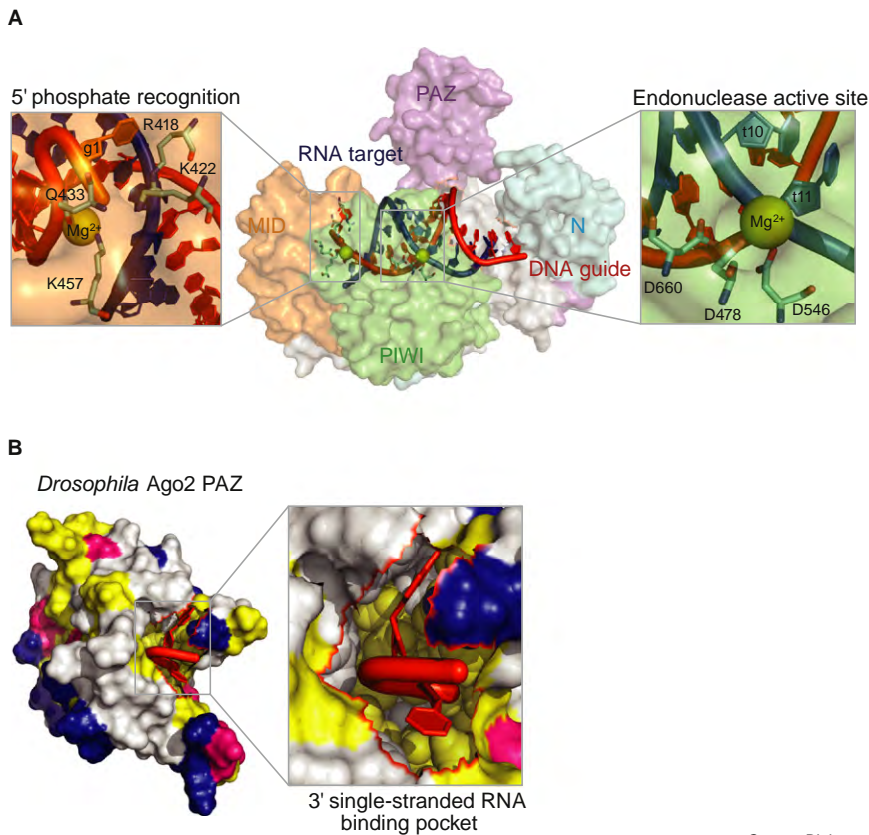


Figure 2. The domains of Argonaute proteins. (A) The structure of *Thermus thermophilus* Argonaute (PDB ID 3HM9) shows the three key functional domains common to all Argonaute proteins: the single-stranded RNA-binding PAZ domain (purple), the PIWI endonuclease domain (green) with its characteristic catalytic triad (D478, D546 and D660 in *T. thermophilus*) that cleaves the target RNA (dark blue), and the MID domain, which binds the 5' phosphate and first nucleotide of the nucleic acid guide (red; here, a DNA guide, but typically an RNA in eukaryotes). Adapted from Wang *et al.* (2009). (B) The structure of the *Drosophila melanogaster* Ago2 PAZ domain (PDB ID 1T2R) reveals how this domain binds the 3' single-stranded tail of the guide RNA. Amino acids are colored according to their chemical properties: hydrophobic, yellow; acidic, pink; and basic, blue. Adapted from Lingel *et al.* (2004).

are initially loaded with double-stranded small RNAs, which then mature to single-stranded RNA-containing, functional RISC. Loading of a small RNA duplex into an AGO protein requires both ATP and the chaperones Hsc70 and Hsp90. Hsc70 and Hsp90 are thought to use ATP energy to open AGO proteins to permit them to bind double-stranded small RNAs, but the details of this process remain unknown.

In *Drosophila*, mammals, and likely other higher plants and animals, RISC complexes load a small RNA duplex in a defined orientation: the thermodynamically less stable 5' end of the double-stranded small RNA ends up in the phosphate-binding pocket of the MID domain, establishing that strand as the future small RNA guide. The other RNA strand becomes the passenger. Such thermodynamic

asymmetry, together with first nucleotide identity, determines which strand of a small RNA duplex becomes the guide for an AGO protein. Most small RNA duplexes preferentially produce a readily predictable siRNA guide or miRNA strand, although some are bi-functional, with both strands loaded into AGO proteins. Even for these 'symmetric' small RNAs, a single molecule of duplex can only load one of its two strands; the other strand is ultimately destroyed.

An AGO protein bound to an siRNA or miRNA/miRNA* duplex is called pre-RISC. Once the siRNA passenger or miRNA* strand of the duplex small RNA is evicted from pre-RISC, the complex becomes mature RISC. For catalytically active AGO proteins, the siRNA passenger strand is thought to be cleaved as if it were a target mRNA.

The heterodimeric protein C3PO has been proposed to facilitate release of the cleaved passenger strand. For catalytically inactive AGO proteins, mismatches within the duplex — in the seed or between guide positions g12 to g15 — promote maturation of pre-RISC to RISC.

Many eukaryotes produce multiple Argonaute proteins, which are often functionally distinct. Flies and plants, for example, devote different AGO proteins to the RNAi and miRNA pathways. *Caenorhabditis elegans* produces 27 Argonautes (including two PIWI proteins). Mice and humans make four AGO proteins, only one of which, Ago2, retains the ability to cleave its RNA targets. Whether the functions of Ago1, Ago3, and Ago4 differ is not known.

How does Argonaute recognize and repress its mRNA targets? The 'seed sequence' of a small silencing RNA guide — nucleotides 2 to 7 or 2 to 8 — provides nearly all of the specificity for target binding. Argonaute proteins pre-organize the seed sequence into a one-stranded helix whose conformation makes it ready to pair with a target without loss of entropy. Argonaute proteins accomplish this by binding the negatively charged phosphodiester backbone of seed sequence nucleotides, displaying the edges of bases g2 to g8 so that they are ready to base-pair with t2 to t8 of a target mRNA.

Why the rest of the nucleotides of the small RNA guide contribute so little to target binding remains unexplained. The *two-state* model envisions that guide strand nucleotides 3' to the seed sequence alternate between two isoenergetic conformations: bound to Argonaute, with the 3' terminus anchored in the PAZ domain, and paired to the target mRNA, with the 3' end free in solution. The *fixed-end* model proposes that the 3' end of the guide strand remains anchored in the PAZ domain at all times, irrespective of the presence of a target RNA. Both models assume that the 5' end of the small RNA guide always resides in the MID domain phosphate-binding pocket.

The structures of *T. thermophilus* Argonaute bound to a DNA guide and to a DNA guide together with an RNA target support the two-state model: the 3' end of the guide strand binds the PAZ domain when only the DNA guide is present, but not when a target RNA

is present. Interestingly, the side chains of the catalytic amino acid triad move closer to the phosphodiester bond to be cut when the target RNA pairs with the guide. Much additional work will be needed to determine which, if either, of the two models best explains how Argonautes bind their targets.

Many Argonaute proteins lack a functional DDE catalytic triad and, thus, repress their targets through mechanisms other than endonucleolytic cleavage. For example, miRNA-guided Argonaute proteins such as fly or human Ago1 reduce the stability of their mRNA targets and can also block mRNA translation. The miRNA-binding Argonaute proteins, miRNAs, and their mRNA targets all localize to P bodies, cytoplasmic loci where general mRNA decay has been proposed to occur. In *Drosophila* and mammals, the Argonaute-binding protein GW182 is required for miRNA-directed mRNA repression and the recruitment of Argonaute proteins to P bodies, but is dispensable for RNAi.

What do we know about how piRNAs are made and how they function?

piRNAs guide PIWI proteins to silence transposons in the germ line of animals. We know little about the early steps in the biogenesis of piRNAs except that unlike siRNAs and miRNAs, piRNA production does not require Dicer and likely involves only single-stranded precursor RNAs.

Hints about the biogenesis of piRNAs and their role in transposon silencing come mainly from studies of *Drosophila* ovaries and testes. *Drosophila* gonads express three PIWI proteins: Piwi, Aubergine, and Ago3. Piwi localizes to the nucleus in both the germline and the surrounding somatic follicle cells, which help regulate the differentiation and patterning of the germline nurse cells and oocytes. Transposon-derived piRNAs bound to Piwi silence transposon expression in the nucleus by poorly understood mechanisms. In contrast, Ago3 and Aub reside in the germ cell cytoplasm.

What is the 'Ping-Pong' mechanism of amplification of piRNAs?

High throughput sequencing of piRNAs bound to Piwi, Aubergine, and Ago3 suggest a model, the 'Ping-Pong' mechanism, for how this class of small silencing RNAs is produced and subsequently amplified in response to transcription of the transposons they

target. The Ping-Pong model proposes that Aubergine and Ago3 collaborate both to increase the abundance of piRNAs and to bias piRNAs toward the antisense strand. The detailed mechanism by which Aubergine and Piwi acquire primary piRNAs is unknown, but current evidence suggests they derive from long RNAs transcribed from 'piRNA clusters': transposon-rich, gene-poor regions of the genome which are specifically transcribed in the gonads. The Ping-Pong model postulates that Aubergine, bound to an anti-sense primary piRNA, pairs with the mRNA transcript of an active transposon, cleaving it in two and generating a 3' cleavage product bearing a 5' monophosphate. The 5' end of the 3' cleavage fragment then becomes the 5' end of a sense piRNA bound to Ago3. This 'secondary piRNA' can then direct cleavage of a primary piRNA transcript derived from a piRNA cluster. The next step reverses the process: the Ago3-sense-piRNA complex cleaves the long transcript of a piRNA cluster, generating antisense RNA fragments that bind to Aub and are envisioned to be trimmed to piRNA length by an unidentified 3'-to-5' exonuclease.

In addition to the transposon-targeting piRNAs produced by the Ping-Pong cycle early in spermatogenesis ('pre-pachytene piRNAs'), mammals generate piRNAs from non-repetitive but gene-poor loci as the developing spermatocytes enter the pachytene stage of meiosis. Mutations that block production of these pachytene piRNAs arrest spermatogenesis at the round spermatid stage. We do not know why pachytene piRNAs are required for mammalian sperm maturation or what genes or DNA structures pachytene piRNAs regulate.

What is on the horizon? piRNAs protect the germline from invading transposons, but do they actually silence expression? We know that Piwi proteins cleave transposons to generate more piRNAs, which then cleave more transposons, but it remains to be established that this amplification cycle is the primary mechanism for transposon silencing by piRNAs. In many animals, piRNAs repress transcription, but we do not know how. Do Piwi proteins bind DNA or nascent transcripts? Do they recruit factors that alter chromatin structure or histone modifications?

Moreover, piRNA biogenesis, especially primary piRNA production, remains obscure. Are the piRNA precursors transcribed from large clusters initially fragmented to generate the 5' monophosphorylated ends of piRNAs, then trimmed by exonucleases to generate discrete 3' ends? What is the function of piRNAs derived from non-repetitive sequences? In *Drosophila* embryos, piRNAs have recently been proposed to bind the 3' untranslated region (UTR) and promote deadenylation of the mRNA encoding Nanos, a protein required for anterior-posterior patterning of the developing embryo. Are piRNAs that map to 3' UTRs generally involved in the temporal or spatial control of gene expression?

The sorting of small RNAs among different Argonaute proteins according to the structure and sequence of an siRNA or miRNA duplex suggests that each Argonaute protein might have a unique or specific regulatory function or that individual Argonautes might be specifically retained or degraded during cell differentiation. Yet no functional difference has been discovered distinguishing human Ago1, Ago3, and Ago4. Perhaps these three human AGOs bind different regulatory proteins. Defining the repertoire of proteins bound by each human Argonaute protein may help reveal their specific functions.

Where can I learn more?

- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.* 12, 19–31.
- Filipowicz, W. (2005). RNAi: the nuts and bolts of the RISC machine. *Cell* 122, 17–20.
- Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.
- Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* 11, 576–577.
- Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M., and Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* 33, 204–214.
- Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. *Genes Dev.* 19, 517–529.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* 461, 754–761.

¹Department of Biochemistry and Molecular Pharmacology and ²Howard Hughes Medical Institute and RNA Therapeutics Institute, University of Massachusetts Medical School, LRB 822, 364 Plantation Street, Worcester, MA 01605, USA.
E-mail: phillip.zamore@umassmed.edu