Title
Mechanistic Studies of Human RISC Loading Complex Function in RNA Interference

Author
Noland, Cameron Ladd

Publication Date
2012

Peer reviewed|Thesis/dissertation

Mechanistic Studies of Human RISC Loading Complex Function in RNA Interference

by

Cameron Ladd Noland

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jennifer A. Doudna, Chair
Professor Eva Nogales
Professor James M. Berger
Professor David E. Wemmer

Fall 2012

Abstract

Mechanistic Studies of Human RISC Loading Complex Function in RNA Interference

by

Cameron Ladd Noland

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Jennifer A. Doudna, Chair

RNA interference (RNAi) pathways are critical eukaryotic post-transcriptional gene regulatory systems that control the expression of at least one third of all human coding genes. In each of these pathways, short single-stranded RNAs (ssRNAs) bind to cognate messenger RNAs (mRNAs) and direct their endonucleolytic cleavage or translational repression by RNA-Induced Silencing Complexes (RISCs). In the cytoplasmic portion of these pathways, silencing is triggered by long double-stranded RNA (dsRNA) precursors (pre-siRNAs) or hairpin RNAs (pre-miRNAs), which are processed by the RNase III enzyme Dicer to yield short 21-23 nucleotide (nt) duplex RNAs termed short-interfering RNAs (siRNAs) and microRNAs (miRNAs), respectively. These short duplex RNAs are then loaded onto Argonaute2 (Ago2), the catalytic component of RISC. Ago2 retains one strand of the duplex (the "guide" strand) for subsequent gene targeting and discards the other. This process of asymmetrically loading siRNA or miRNA duplexes onto RISCs such that one strand is preferentially retained is fundamental to the target specificity of RNAi pathways.

Structural studies of the individual components of the RNAi machinery have proven immensely informative to our mechanistic understanding of the individual steps of RNAi pathways such as substrate cleavage by Dicer and subsequent mRNA targeting by Ago2. What has been missing is a clear picture of how different RNAi components functionally interact to transfer small RNAs from one protein to another in a way that ultimately allows for the selection of the proper guide strand by Ago2. In *Drosophila* and humans, protein complexes called RISC Loading Complexes (RLCs) have been implicated in facilitating this process. At their core, these complexes are composed of Ago2, Dicer, and a dsRNA binding protein (dsRBP). Two such complexes exist in humans and are defined by their associated dsRBP. One complex contains HIV-1 Trans-activation Response (TAR) RNA Binding Protein (TRBP) and the other contains Protein Activator of PKR (PACT). Here we have used electron microscopy (EM) and single particle reconstruction methods to show that human Dicer has an L-shaped architecture consisting of a long catalytic branch and a shorter DExH/D helicase-containing base branch. Further, we present a three-dimensional (3D) reconstruction of a highly dynamic human RLC containing Ago2, Dicer, and TRBP. TRBP interacts with the distal end of Dicer's helicase branch, whereas Ago2

interacts with Dicer's catalytic branch. The structural information garnered in this study, combined with previous biochemical data, led to a testable model for how human RLCs efficiently transfer Dicer products to Ago2 in an orientation that allows for accurate selection of the targeting strand.

The preferred guide strand of any given siRNA or miRNA can be predicted with high accuracy based on the thermodynamics of the duplex ends. The RNA strand that has its 5′ end at the less stable end of the duplex is preferentially loaded onto Ago2 as the guide strand. In flies, an RLC subcomplex containing Dicer-2 and the dsRBP R2D2 is involved in sensing the thermodynamics of siRNAs prior to their loading onto Ago2, although it is unclear which protein is the actual sensor of thermodynamic asymmetry. In humans the sensor of siRNA thermodynamics has been unclear, given that Dicer forms a complex with either TRBP or PACT. In plants, some Dicer/dsRBP heterodimers are involved in sensing siRNA thermodynamics while others are not. Using *in vitro* biochemical experiments that were guided by structural insights, we show that human Dicer products are released and then rebound at a novel site along Dicer's helicase domain in a TRBP- or PACT-dependent manner. This novel binding site allows either Dicer/dsRBP heterodimer to sense siRNA thermodynamic asymmetry such that Dicer binds to the less stable end of the duplex and the associated dsRBP binds to the more stable end. We further demonstrate that Dicer itself is the sensor of siRNA thermodynamics and that this functionality is activated upon association with either TRBP or PACT. These crucial insights into siRNA positioning by human Dicer/dsRBP complexes led to a revised model for how strand-selective RISC loading may be achieved by RLCs in human RNAi pathways.

Although structural and biochemical data supports the notion that human RLCs function similarly to the *Drosophila* Ago2/Dicer-2/R2D2 RLC in strand-selective RISC loading, a number of clear differences exist. For example, in humans it has been shown that Ago2 can bind siRNAs and miRNAs in the absence of Dicer, which is not the case in flies. Furthermore, in certain cases Ago2 alone binds to duplex RNAs strand-selectively. The extent to which human RLCs are actually important for Ago2 loading and strand selection has therefore been controversial. We have used a reconstituted system to determine the degree to which each of the core components of the human RNAi machinery contributes to RISC loading and guide strand selection. We show that Ago2 has intrinsic but substrate-dependent strand selection capabilities. This activity, however, is in many cases enhanced substantially when Ago2 is in complex with Dicer and TRBP or PACT as a functional RLC. Our findings suggest that rather than functioning exclusively in human strand selection, Ago2's binding preferences serve instead as a secondary RISC loading checkpoint that acts in concert with Dicer/dsRBP asymmetry sensors to ensure proper strand selection by RLCs. The specific roles of each component in this process are dictated and fine-tuned by specific duplex parameters such as thermodynamics, 5′ nucleotide identity, and duplex structure. Surprisingly, our results also show that strand selection for some miRNAs is enhanced by PACT-containing complexes but not by those containing TRBP. Furthermore, overall mRNA targeting by miRNAs is disfavored for complexes containing TRBP but not PACT. This key finding reveals the possibility that RLCs containing TRBP may be optimized for the siRNA pathway, whereas RLCs containing PACT are optimized for the miRNA pathway. Overall, the body of work presented herein represents a significant step forward in our understanding of how core components of human RNAi pathways functionally interact within RLCs to achieve proper target specificity through strand-selective RISC loading.

To
Mom & Dad

Acknowledgements

Without the considerable support and advice that I have received over the years from my mentors, labmates, friends, and family, this dissertation simply would not have come to fruition. First and foremost, I owe a debt of gratitude to Jennifer Doudna, whose mentorship, trust in my abilities, and contagious enthusiasm for scientific discovery has served as a driving force behind my work. I am grateful for the intellectual control over this research that Jennifer has allowed me, and also for the fact that her door has always been open for troubleshooting experiments or planning for the next steps.

Jennifer has also built up a lab of interesting and intelligent scientists and fostered an environment has been fun and exciting to be a part of. I am thankful to all of the past and present members of the Doudna Lab, both for their wisdom and their humor. In particular, I owe Srinivas Chakravarthy a debt of gratitude for putting up with me as a rotation student that had never purified a protein before and for somehow answering my many inane questions with a straight face. Kaihong Zhou has been a veritable treasure trove of experimental knowledge, and Enbo Ma has also been generous to collaborate and share his expertise. Dipa Sashital has been immensely generous with her time and knowledge; I am happy to consider her both a mentor and a friend. Ho Young Lee has also provided a great deal of insight into my experiments, as has Andy Mehle and Martin Jinek. Thanks also to the other members of the Dicer sub-group; Mary Anne Kidwell, Ross Wilson, Emine Kaya, and Aaron Brewster for valuable conversations and for sharing reagents. I was fortunate to work with a very hardworking and intelligent summer student, Ruty Miyazaki-Smith, who helped me optimize several tandem affinity purification protocols. Stephen Floor was a great resource to me for this project as well. In addition to providing intellectually stimulating conversations and experimental expertise, my coworkers have also made the lab a fun place to work over the years. Whether it was a contentious debate over the rules of the B-E-E-R Board, a particularly Wacky Wednesday, playing hooky to see an A's game, or a quick coffee at Yali's, the rapport between the members of the Doudna Lab has often added a much needed dose of sanity and perspective to my graduate school career.

I am also thankful to several members of the Nogales Lab, not only for their willingness to let me use their electron microscopy equipment, but also for teaching me the ins and outs of single particle reconstruction. Thank you especially to Hongwei Wang and Patricia Grob for teaching me nearly everything I know about electron microscopy. Thanks also to David Taylor, Karin Felderer, and Bunpote Siridechadilok for their part in our various collaborative electron microscopy studies.

I have been fortunate to work with a number of collaborators over the years. In particular, I would like to thank Ann Fischer for her incredibly dedicated assistance with tissue culture. Thank you as well to three mass spectrometry experts without whom a good portion of this work would not have been accomplished: David King, Lori Kohlstaedt, and Tony Iavarone. Further, I would like to thank Scott Gradia and the scientists in the Keck MacroLab for the use of their facilities. Dinshaw Patel was particularly generous in sharing unpublished structural information. Thank you also to John Hershey for providing eIF3 samples and to Lynne Maquatt's lab for providing purified UPF1 and for helpful discussions.

I would also like to thank my committee members, Eva Nogales, James Berger, and David Wemmer, for the advice and direction they have given me over the years as well as for

their thoughtful comments on this manuscript. Each member of my committee has played a crucial role in directing the scope of my projects and keeping me on track to complete them.

I am particularly grateful to the amazing group of caring, intelligent, and hilarious friends that have become my Bay Area family and supported me over the years. Thank you to Lacey Fairchild, Walker DePuy, Kate Engel, Sheena Degnan, Dan Pulver, Suzy Pulver, Johnny Oliver, Rich Price, Jaime Hernandez, Chelsie Moffatt, and Orlando Naranjo – both for lending an understanding ear and for indulging in the occasional impromptu dance party late into the night to help me get over any number of failed protein purifications and cracked gels.

Finally, thank you of course to my family. My brother, Chris, and my sister, Mandie, have been more helpful than they probably know. Whether lending an ear or a laugh, our conversations have always helped to keep things in perspective. I am also grateful to my aunt, Kirby, and uncle, Carl, whose selfless support and generosity I hope to someday share with someone as well. Most importantly, thank you to my parents, Lynn and Gary. Your love, unwavering support, and guidance over the years has meant the world to me. You taught me the meaning of hard work and integrity, and I would not be where I am today without your sacrifices. No words can ever thank you enough.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

Abbreviation	Description
17-AAG	17-(allylamino)-17-demethoxygeldanamycin
2D	two-dimensional
3D	three-dimensional
6xHis	hexahistidine
A	adenine or alanine
Å	Angstrom
ADP	adenosine diphosphate
Ago	Argonaute
ATP	adenosine 5′-triphosphate
bp	base pair
BSA	Bovine Serum Albumin
°C	degrees Celsius
C	cytosine
C. elegans	*Caenorhabditis elegans*
Ca	calcium
cal	calorie
CBP	Cap Binding Protein
Ci	curie
CCD	charge-coupled device
CHS	Chalcone Synthase
Cl	chloride
cryoEM	cryo-electron microscopy
D	aspartate
Da	Dalton
DGCR8	DiGeorge Critical Region 8
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
dsRBD	double-stranded RNA binding domain
dsRBP	double-stranded RNA binding protein
DTT	dithiothreitol
DUF	domain of unknown function
e^-	electron
E	glutamate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eIF	eukaryotic Initiation Factor
EJC	Exon Junction Complex
EM	electron microscopy
endo-siRNA	endogenous siRNA
ES	embryonic stem

FSC	Fourier shell correlation
g	grams
G	guanosine
GTP	guanosine 5´-triphosphate
H	histine
H_2O	water
HEPES	N-2-hydroxyethylpiperazine-N´-2-ethanesulfonic acid
HIV-1	Human Immunodeficiency Virus-1
IMAC	Immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactoside
k	kilo
K	potassium
K_D	dissociation constant
L	liter
LaB6	lanthanum hexaboride
LB	Luria-Bertani broth
Loqs	Loquacious
m	mili or meter
M	Molar / moles per liter
MAPK	Mitogen-Activated Protein Kinase
MBP	Maltose Binding Protein
Met	methionine
Mg^{2+}	magnesium
miRNA/miR	microRNA
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein
MRA	multi-reference alignment
MSA	multivariate statistical analysis
μ	micro
MudPIT	multidimensional protein identification technology
n	nano
Na	sodium
ncRNA	non-coding RNA
nm	nanometer
NMD	nonsense-mediated mRNA decay
NMR	nuclear magnetic resonance
nt	nucleotide
NTA	nitrilotriacetic acid
OD	optical density
p	pico
P	proline
PABP	Poly(A)-Binding Protein
PACT	Protein Activator of PKR
PAGE	polyacrylamide gel electrophoresis
PAZ	Piwi-Argonaute-Zwille
PBS	phosphate-buffered saline

PCR	polymerase chain reaction
PKR	Protein Kinase R
pre-mRNA	precursor mRNA
pre-miRNA	precursor miRNA
pre-siRNA	precursor siRNA
pri-miRNA	primary miRNA
PTC	premature termination codon
PTM	post-translational modification
PTGS	post-transcriptional gene silencing
PVDF	polyvinylidine fluoride
RCT	random conical tilt
RISC	RNA-induced silencing complex
RLC	RISC loading complex
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
rpm	revolutions per minute
S	serine
SDS	sodium dodecyl sulfate
SILAC	stable isotope labeling with amino acids in cell culture
siRNA	short interfering RNA
SPNR	Spermatid Perinuclear RNA Binding Protein
ssRNA	single-stranded RNA
T	thymine or threonine
TAP	tandem affinity purification
TAR	transactivation response
TBS	tris-buffered saline
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
TRBP	TAR RNA Binding Protein
tRNA	transfer RNA
U	uridine
UPF	up-frameshift
UTR	untranslated region
UV	ultraviolet
V	volt
WT	wild type
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside
Y	tyrosine

Chapter 1

Introduction

1.1 The biological importance of post-transcriptional gene regulation by RNAi

Post-transcriptional control of gene expression is essential to organismal survival in all kingdoms of life. This type of gene regulation operates by diverse means, fine-tuning gene expression in various cell types at different times. Gene regulation by these means allows for the exquisite control of biological processes such as developmental timing, defense against pathogens, and adaptation to novel environments. Additionally, post-transcriptional gene regulation provides essential safeguards against the expression of mRNAs harboring potentially deleterious mutations. RNA interference, or RNAi, is a well-conserved post-transcriptional gene regulation pathway in which genetically encoded or exogenous short, single-stranded RNAs (ssRNAs) target cognate messenger RNAs (mRNAs) for repression via RNA degradation and/or translational repression (Bernstein et al. 2001; Jinek and Doudna 2009; Hutvagner et al. 2001; Zamore et al. 2000). This process is reliant on long double-stranded RNA (dsRNA) precursors (pre-siRNAs) or stem-loop RNAs (pre-miRNAs), which are processed into shorter duplex RNAs – short-interfering RNAs (siRNAs) and microRNAs (miRNAs), respectively. These duplexes are then used as the triggers of gene silencing (Elbashir et al. 2001a; 2001b; Zamore et al. 2000; Lee and Ambros 2001; Lagos-Quintana et al. 2001; Lau et al. 2001). Since the discovery of RNAi in 1998, over 10,000 miRNA and endogenous siRNA genes have been annotated, with potential targets covering over a third of all human genes (Lewis et al. 2005; Kozomara and Griffiths-Jones 2011). RNAi pathways have been shown to be essential for a wide range of basic biological functions including the control of developmental timing (Pasquinelli et al. 2000; Reinhart et al. 2000; Wightman et al. 1993; Lee et al. 1993), brain morphogenesis (Giraldez et al. 2005), metabolism (Xu et al. 2003), apoptosis (Brennecke et al. 2003), and viral defense (Wilkins et al. 2005; Saleh et al. 2009; Wang et al. 2006). Additionally, this system has been found to play key roles in stem cell maintenance and division and the fine balance between those processes (Tay et al. 2008; Melton et al. 2010; Hatfield et al. 2005).

Given its regulation of myriad biological processes, it is not surprising that misregulation of the RNAi system can lead to disaster for an organism. This has been observed when genes involved in miRNA biogenesis such as the protein DiGeorge Critical Region 8 (DGCR8) are missing, which leads to DiGeorge syndrome (Gregory et al. 2004), or when downstream genes such as TRBP are mutated, as is the case in certain cancers (Melo et al. 2009). Furthermore, reductions in the levels of the key RNA processing enzyme Dicer have been shown to lead to geographic atrophy (Kaneko et al. 2011). Lack of Dicer altogether is fatal (Bernstein et al. 2003). Aside from the individual proteins involved in RNAi, certain miRNAs have been shown to be oncogenic, leading to the coinage of the term “oncomiRs” (Si et al. 2007; Lu et al. 2008).

RNAi systems have also garnered considerable interest within the biomedical community due to their potential for treating disease. For example, orally delivered siRNAs targeting a macrophage gene have been shown to reduce inflammation in mice, a finding that has implications for diseases such as rheumatoid arthritis and inflammatory bowel disease (Aouadi et al. 2009). In another key study, siRNAs targeting Herpes Simplex Virus 2 in mice either before or after exposure to the virus had a protective effect against viral infection (Palliser et al. 2006). RNAi-based therapeutics have also been aimed at the treatment of cancer. In one recent study, the administration of siRNAs targeting the nonsense-mediated decay machinery led to immune recognition and rejection of invading tumor cells in mice (Pastor et al. 2010).

In order to fully understand the interplay between the varied biological functions of RNAi systems and the deleterious effects that a malfunction may have on an organism, a detailed mechanistic understanding of those systems is necessary. This knowledge may well lead to the

development of more advanced RNAi-based treatments for disease. The body of work presented herein advances our understanding of these mechanisms, focusing on the human system and how a key protein complex called the RISC Loading Complex (RLC) processes RNAs to achieve proper target specificity. A sufficient understanding of the impact of this work requires knowledge of the current state of the field of RNAi and the crucial unanswered questions that drive the field. The following introduction serves as a primer on the mechanisms of RNAi as they are currently understood, providing a historical perspective as well as an outlook for the future.

1.2 The path to the discovery of RNAi

1.2.1 A novel mechanism for gene knockdown by dsRNAs

The field of RNAi was fortuitously born of the desire for an alternative to classical genetics for analyzing gene function. Although classical genetic techniques such as forward mutational screens had provided an indispensible means for assessing gene function in organisms such as yeast, *C. elegans*, and *Drosophila*, many higher organisms had proven intractable to classical genetic analysis, and alternative approaches for analyzing gene function were therefore necessary. A breakthrough came in 1984 when Izant and Weintraub expressed RNAs complementary to a known mRNA (antisense RNAs) as a means for sequence-specifically knocking down the expression of those genes approximately 4- to 5-fold (Izant and Weintraub 1984). This technique was quickly expanded for use in mammalian cells, *Xenopus* oocytes, *Drosophila* embryos, and *C. elegans* (Izant and Weintraub 1985; Harland and Weintraub 1985; Melton 1985; Rosenberg et al. 1985; Fire et al. 1991). Gene knockdown was thought to occur through a blockade of protein translation via RNA:RNA hybridization (Melton 1985). Unfortunately, this method was generally quite inefficient and variable from gene to gene, requiring anywhere from 10- to 100-fold excesses of antisense RNA (Melton 1985; Izant and Weintraub 1984; Harland and Weintraub 1985). Additionally, it soon became clear that the mechanism of gene knockdown by antisense RNA was not clear-cut when Fire and colleagues observed in 1991 that not only antisense, but also sense RNAs seemed to inhibit gene expression when injected into *C. elegans* oocytes (Fire et al. 1991). This phenomenon was again observed by Guo and Kemphues in 1995 when it was observed that both sense and antisense RNAs inhibited *par-1* gene expression in *C. elegans* (Guo and Kemphues 1995). These observations perplexed many and challenged the previously accepted notion that this form of gene knockdown was caused by antisense RNA hybridization to complementary sense mRNAs.

It was not until 1998 that Fire and colleagues solved this mystery and opened the door to the revolutionary new field of RNAi. The key to Fire's discovery lay in the minutia of the experimental procedures. In many cases, the antisense RNAs that had been used in the gene knockdown experiments of the late 80s and early 90s were synthesized *in vitro* using the bacteriophage SP6 RNA polymerase. This technique was highly efficient and specific, producing very few ectopic transcripts and only ~0.2% the transcribed RNAs were aberrant sense RNAs (Melton et al. 1984). For Fire and colleagues, however, that was not good enough. They hypothesized that this 0.2% could hybridize with antisense RNAs *in vivo*, and that the resulting dsRNAs were in fact responsible for knocking down gene expression. This group found that when either *unc-22* sense or antisense RNAs were injected into *C. elegans*, very high doses were required to knock down gene expression, which has the effect of producing twitching animals. Surprisingly, injecting a 1:1 mixture of sense and antisense RNAs produced strong twitching phenotypes even at concentrations as low as a few molecules per cell. When that RNA was then

extracted from the injected animals, they found it to be predominantly double stranded. When the extracted RNAs were further gel purified and re-injected, they caused a profound decrease in endogenous *unc-22* mRNAs. This effect was hypothesized to be post-transcriptional, given that dsRNAs corresponding to promoter and intronic regions had no effect on gene expression (Fire et al. 1998). Fire's discovery hinted at the existence of some unknown process of multiple-turnover gene knockdown, which pointed to the exciting possibility that a previously unknown endogenous system existed for post-transcriptional gene regulation by dsRNAs.

Around the same time that questions were arising regarding the mechanism of antisense interference, another mystery was taking shape in the field of plant genetics. In 1990, Jorgensen and colleagues – as well as a group led by Antoine Stuitje – had introduced a chalcone synthase (CHS) gene into petunias in an attempt to drive its overexpression and induce more brilliant coloration of the flowers. To their astonishment, this procedure had the opposite effect. Many of the transgenic flowers lacked pigment completely and others were patterned, which corresponded to a dramatic decrease in CHS expression (Napoli et al. 1990; van der Krol et al. 1990). This phenomenon was termed co-suppression or post-transcriptional gene silencing (PTGS), and although antisense interference was one possible explanation, initially no such antisense RNAs were found. The key here was that groups investigating PTGS in plants had not looked for small enough RNAs. By enriching for small RNAs, David Baulcombe's group was finally able to detect 25 nt dsRNAs (now referred to as siRNAs) following the transformation of transgenes into tomato plants (Hamilton and Baulcombe 1999). This finding hinted at a mechanism of action for PTGS that was similar to that of RNAi, and again implied that an unknown endogenous gene regulation system existed that used dsRNAs to target genes for repression.

1.2.2 An extensive endogenous gene regulatory system mediated by dsRNAs

While the mystery of gene knockdown by dsRNAs unraveled, other curious findings related to the control of developmental timing in *C. elegans* were being reported. These findings revolved around the discovery that the *lin-4* gene (a heterochronic gene in *C. elegans* known to negatively regulate the levels of the LIN-14 protein during development) in fact seemed to encode two small non-coding RNAs (ncRNAs). One of these RNAs was approximately 61 nt in length and was predicted to form a stem-loop structure, whereas the other RNA was 22 nt and derived from the stem of the larger RNA (Lee et al. 1993). Surprisingly, this RNA was complementary to a conserved repeat sequence found in the 3′ untranslated region (3′ UTR) of the *lin-14* mRNA, and a deletion mutant in which these regions were missing was found to be a strong gain of function allele (Lee et al. 1993; Wightman et al. 1993). These studies led to the conclusion that the larger *lin-4* RNA may be post-transcriptionally processed into the smaller one, and that LIN-14 depletion occurs via sense:antisense hybridization between the *lin-4* ncRNA and the *lin-14* mRNA. Years later, another 21 nt ncRNA called *let-7* was found to negatively regulate another heterochronic gene called *lin-41* (Reinhart et al. 2000). At the time, researchers were excited by the possibility that a conserved cascade of these ncRNAs may be responsible for regulating developmental timing across species (Pasquinelli et al. 2000), but in 2001 the simultaneous publication of three reports demonstrated that the implications of these early findings were actually even larger (Lee and Ambros 2001; Lau et al. 2001; Lagos-Quintana et al. 2001). These seminal studies each identified an extensive class of small regulatory RNAs (thereafter termed "microRNAs" (miRNAs)) that were derived from similar hairpin precursors and matched regions within the 3′ UTRs of other genes. Many of these miRNAs were conserved across a wide range

of species and were expressed tissue-specifically. Additionally, the miRNAs were derived from either the 5′- or the 3′ end of their precursor hairpin RNAs and had a strong preference for a 5′ uridine and against a 5′ guanosine. Notably, although the experimental protocols used would not have selected against natural siRNAs, not a single natural siRNA was identified in these studies, leading to the conclusion that miRNAs were the natural substrates for the RNAi machinery. This idea dovetailed nicely with the finding that the *let-7* RNA was processed from its hairpin precursor by one of the enzymes that had been found to be central to the siRNA pathway (see **Section 1.3.2**; (Hutvagner et al. 2001)). With these discoveries, the lines were blurred between the siRNA and miRNA pathways. Indeed, as will be discussed below, these pathways overlap at many points, particularly in humans.

1.3 miRNA and siRNA biogenesis

1.3.1 pri-miRNA processing and nuclear export of pre-miRNAs

With the exception of miRNA genes interspersed within Alu repeats, primary miRNA transcripts (pri-miRNAs) are produced in the nucleus by RNA Polymerase II and contain 5′ 7-methyl guanosine caps as well as poly(A) tails (Borchert et al. 2006; Cai et al. 2004; Lee et al. 2004a). These genes are often clustered and transcribed as a single transcript called a primary miRNA (pri-miRNA), although single, non-clustered pri-miRNAs are also transcribed (Lee et al. 2002). These pri-miRNAs are rapidly processed in the nucleus by the enzyme Drosha to create ~70 nt hairpin precursor miRNAs (pre-miRNAs) (**Figure 1.1**; (Lee et al. 2002; 2003)). Drosha is a member of a class of endoribonucleases with specificity for dsRNA called RNase III enzymes. These magnesium-dependent enzymes function as dimers of two RNase III domains in which a single cleavage event occurs on each strand of an RNA duplex. The geometry of the RNase III dimers creates a characteristic 2 nt 3′ overhang at the cleavage site. Drosha and the RNase III enzyme Dicer (see **Section 1.3.3**) represent special cases where each protein acts as a monomer but contains two RNase III domains that interact to form an intramolecular heterodimer (Han et al. 2004; Robertson et al. 1968; Blaszczyk et al. 2001; Gan et al. 2006; Macrae et al. 2006; Zhang et al. 2004). The pre-miRNA terminus created by Drosha cleavage remains unchanged in the mature miRNA and its 2 nt 3′ overhang is essential for later recognition, transport, and processing steps (see below). The presence of the pri-miRNA terminal loop structure is not important for cleavage by Drosha, although a larger terminal loop generally leads to more efficient cleavage. More important is the presence of at least one ssRNA segment flanking the pri-miRNA hairpin and a dsRNA stem of approximately 33 base pairs (bp) (Zeng et al. 2005; Han et al. 2006; Zeng and Cullen 2005).

Drosha itself is unable to effectively process miRNAs in the absence of another protein called DGCR8. Together, these two proteins form the Microprocessor complex (Denli et al. 2004; Gregory et al. 2004; Han et al. 2004; Landthaler et al. 2004). To accomplish specific cleavage, DGCR8 recognizes the base junction between the hairpin and flanking sequences, serving as a molecular anchor and allowing Drosha to cleave ~11 bp up the hairpin stem (**Figure 1.1**). How exactly DGCR8 recognizes the pri-miRNA junction is a matter of debate. DGCR8 is a heme-binding protein containing an N-terminal WW domain and two tandem dsRNA binding domains (dsRBDs). Binding to heme triggers dimerization and enhances pri-miRNA processing (Barr et al. 2011; Faller et al. 2007; Sohn et al. 2007; Senturia et al. 2010). Precisely how this dimerization enhances Drosha processing is still unclear. No structural data is available for the Microprocessor complex, although electron microscopy reconstructions and partial crystal structures of DGCR8 seem to suggest that this portion of the complex recognizes pri-miRNAs as

a trimer that allows for the transient association of Drosha with the pri-miRNA for cleavage (Faller et al. 2010; Sohn et al. 2007; Senturia et al. 2010; Han et al. 2006).

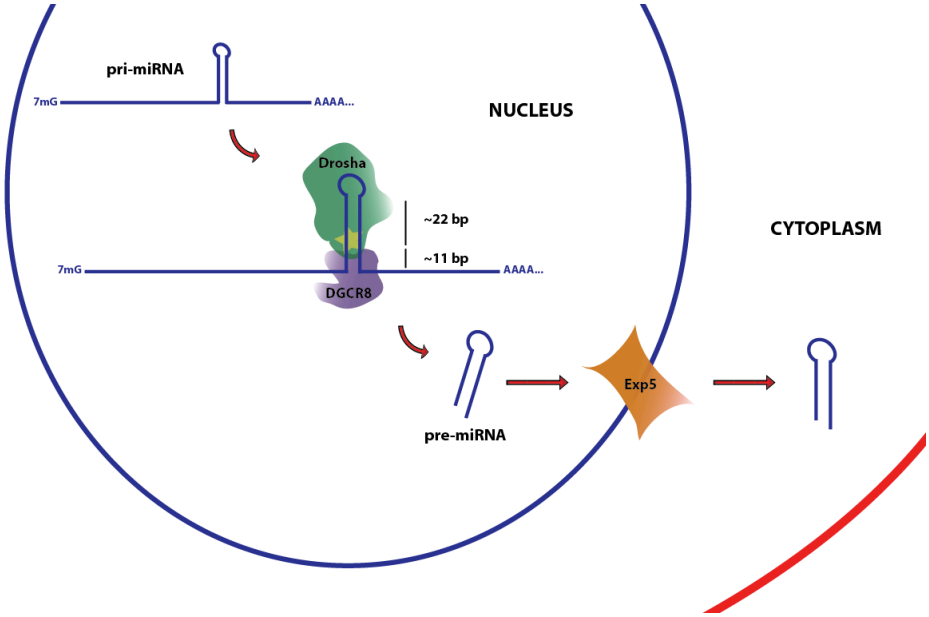

Figure 1.1: Canonical pri-miRNA processing in the nucleus.
RNA polymerase II-transcribed pri-miRNAs are processed by the Microprocessor (Drosha/DGCR8) complex in the nucleus to yield short hairpin RNAs called pre-miRNAs. These RNAs are then exported from the nucleus by Exportin-5 for further processing in the cytoplasm (see **Figure 1.2**).

Although most miRNAs arise from preliminary Drosha processing, some miRNAs derive from intronic pre-mRNA regions and are able to bypass this step. These non-canonical pri-miRNAs, termed "mirtrons," are instead generated from the canonical splicing machinery, forming a mirtron lariat intermediate that is resolved by Debranching Enzyme. Resolved mirtron lariats then fold into a more canonical pre-miRNA hairpin structure (Ruby et al. 2007; Okamura et al. 2007; Berezikov et al. 2007). The 5′- and 3′ ends of these mirtrons are therefore not defined by RNase III cleavage by Drosha, but rather by the splice sites themselves (Ruby et al. 2007).

Following cleavage by either the Microprocessor or Debranching Enzyme, the miRNA and mirtron pathways converge at the point of nuclear export (Okamura et al. 2007). Export from the nucleus is facilitated by the protein Exportin-5 in a RanGTP-dependent manner. Exportin-5 binds to the pre-miRNA stem through extensive weak interactions and specifically recognizes the Drosha-processed 2 nt 3′ overhang, binding sequence non-specifically with a positively-charged pocket. The pre-miRNA is anchored by a tight pi-stacking interaction between the guanidyl moiety of an arginine and the first base pair of the pre-miRNA stem (Okada et al. 2009). Due to its recognition of Drosha-processed ends, Exportin-5 can be thought of as a quality control checkpoint for pre-miRNAs, only allowing properly processed pre-miRNAs to pass through the nuclear envelope.

1.3.2 Biogenesis of endogenous siRNAs

Although it was initially thought that siRNAs are not encoded in the genome, the advent of deep sequencing technologies has allowed for the discovery of numerous endogenous siRNAs (endo-

siRNAs) in species ranging from *C. elegans* to mouse (Czech et al. 2008; Ruby et al. 2006; Ghildiyal et al. 2008; Tam et al. 2008; Watanabe et al. 2008). Endogenous pre-siRNAs derive from large inverted repeat structures, bidirectional transcription of overlapping mRNAs, and pseudogenes. Although endo-siRNAs are primarily involved in retrotransposon suppression, pseudogene-derived endogenous pre-siRNAs, which arise from base pairing between a protein-coding gene and an antisense-transcribed pseudogene, can regulate the expression of specific coding genes (Tam et al. 2008). Biogenesis of endo-siRNAs from their precursors is Drosha-independent, but Dicer-dependent (Okamura et al. 2008; Song et al. 2011; Ghildiyal et al. 2008; Watanabe et al. 2008; Tam et al. 2008).

1.3.3 pre-miRNA and pre-siRNA processing

Once they have reached the cytoplasm, pre-miRNAs and pre-siRNAs are processed by the RNase III enzyme Dicer into their mature miRNA and siRNA forms (**Figure 1.2** (Hutvagner et al. 2001)). Humans express only one Dicer enzyme, and this processing step thus marks a point of convergence for the miRNA and siRNA pathways (Bernstein et al. 2001). This is in contrast to *Drosophila*, which expresses two Dicer proteins (Dicer-1 and Dicer-2). In this case, Dicer-1 mediates pre-miRNA processing and Dicer-2 processes pre-siRNAs (Zamore et al. 2000; Lee et al. 2004b). The structural reasons for these differing substrate specificities is still under investigation, but preliminary studies have revealed that the Dicer-2's dsRNA binding protein (dsRBP; see **Section 1.4**) partner R2D2 participates in restricting its processing to dsRNAs (Cenik et al. 2011). Dicer-1 instead binds to a dsRBP called Loquacious-PB (Loqs-PB) and mounting evidence suggests that Dicer-1's helicase domain specifically recognizes pre-miRNA hairpin loops for exclusive production of miRNAs (Tsutsumi et al. 2011; Saito et al. 2005; Ye et al. 2007; Jiang et al. 2005).

Human Dicer is a ~220 kDa multi-domain enzyme consisting of an N-terminal DExH/D box RNA helicase domain, a domain of unknown function (DUF283), a PAZ domain, tandem RNase III domains (RNase IIIa and IIIb), and a C-terminal dsRBD. Structural data has been indispensible in uncovering the mechanism by which Dicer produces mature siRNAs and miRNAs. Human Dicer's PAZ domain recognizes duplex RNA ends at the Drosha-processed termini of pre-miRNAs (Park et al. 2011; Macrae et al. 2006; 2007). This domain, which is also present in Argonaute (Ago) proteins (see **Section 1.5**), anchors the 5′ phosphate in a conserved basic pocket, and the 2 nt 3′ overhang is bound in a largely hydrophobic pocket, stacking against a conserved phenylalanine and making extensive backbone contacts along a positively-charged groove (Park et al. 2011; Lingel et al. 2004; 2003; Ma et al. 2004; Yan et al. 2003). Following RNA binding, Dicer itself serves as a molecular ruler for ATP-independent cleavage of the duplex 21-23 nt away, yielding siRNAs and mature miRNAs with two sets of 2 nt 3′ overhangs (Takeshita et al. 2007; Macrae et al. 2007; Zhang et al. 2002; 2004; Macrae et al. 2006; Bernstein et al. 2001; Elbashir et al. 2001b; Zamore et al. 2000).

Dicer's helicase domain is required for processing endogenous pre-siRNAs into endo-siRNAs, and in *Drosophila* and *C. elegans* this processing requires that Dicer bind to a dsRNA binding protein (dsRBP) partner (Czech et al. 2008; Welker et al. 2010). In these cases Dicer's helicase domain likely recognizes blunt-ended long dsRNAs and triggers a conformational change in Dicer that allows it to process the entire dsRNA processively without dissociating (Welker et al. 2011). Interestingly, in humans Dicer's helicase domain actually has an autoinhibitory effect on processing of exogenous pre-siRNAs and may have a stimulatory effect on processing of thermodynamically unstable miRNAs, indicating that the function of this

domain across (and within) species is not clear-cut (Soifer et al. 2008; Tsutsumi et al. 2011; Welker et al. 2010; 2011; Ma et al. 2008).

The one known pre-miRNA that bypasses Dicer processing is pre-miR-451. This pre-miRNA is instead processed by the protein Ago2 (See **Section 1.5**). pre-miR-451 is cleaved after its 30th base and undergoes uridylation and trimming to generate the mature miR-451 of 23 nt (Liu et al. 2004; Rivas et al. 2005; Song et al. 2004; Cifuentes et al. 2010; Cheloufi et al. 2010).

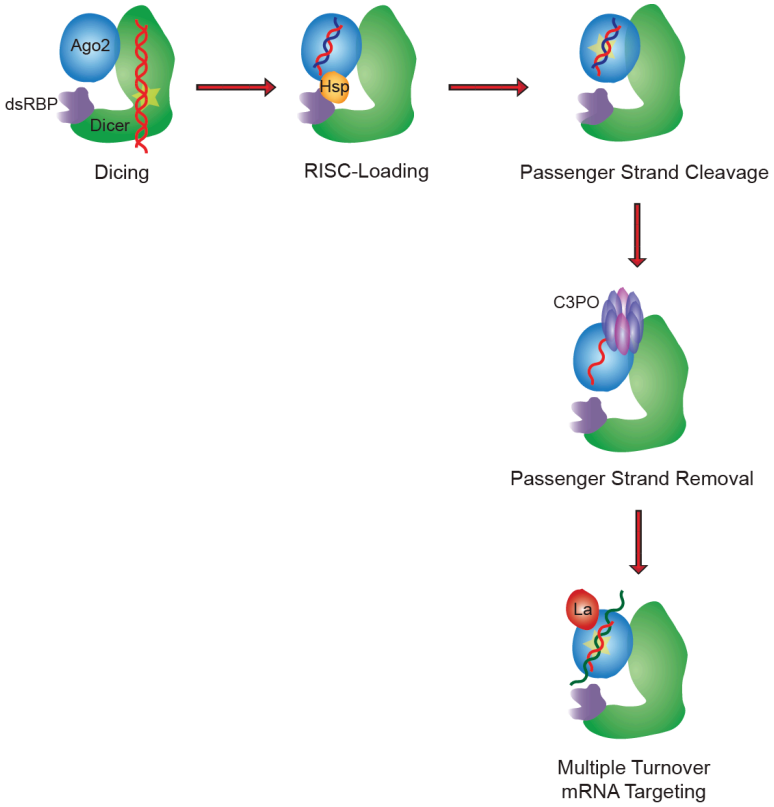

Figure 1.2: The canonical human cytoplasmic RNAi pathway.
pre-siRNAs and pre-miRNAs are engaged at their ends by Dicer's PAZ domain and are cleaved at both strands to yield a 21-23 nt siRNA or miRNA with 2 nt 3′ overhangs. Dicer products are then loaded onto Ago2 with the help of heat shock proteins and Dicer-associated dsRBPs. For siRNAs, the passenger strand is cleaved by Ago2 and removed with the aid of the endoribonucleases C3PO (for miRNAs, passenger strands are generally removed without cleavage). Finally, the guide strand targets cognate mRNAs and if sufficient complementarity exists, the mRNA is cleaved by Ago2 and removed with the help of Autoantigen La. If insufficient complementarity exists, as is the case for many natural miRNA targets, gene silencing is instead achieved by stimulating GW protein-dependent translational repression and mRNA decay (see **Section 1.6; Figure 1.4**).

1.4 dsRNA binding proteins in RNAi

1.4.1 Introduction to dsRBPs

Double-stranded RNA binding proteins are known to play various roles in RNAi pathways. These proteins generally contain a series of 2~3 conserved 65~75 amino acid dsRNA binding domains (dsRBDs) connected by disordered linkers (St Johnston et al. 1992). dsRBDs adopt an α-β-β-β-α fold in which the N- and C-terminal α-helices pack against one face of a three-stranded antiparallel β-sheet (Bycroft et al. 1995; Nanduri et al. 1998; Ryter and Schultz 1998; Ramos et al. 2000; Yang et al. 2010; Yamashita et al. 2011; Stefl et al. 2010; Qin et al. 2010; Kharrat et al. 1995). For each dsRBD, the protein-dsRNA interaction spans ~16 bp, including

two minor grooves and the intervening major groove of the helix. These interactions are in part mediated by hydrogen bonding between the N-terminal α-helix and 2′ hydroxyl groups in the minor groove. The following minor groove is similarly contacted by a loop between the first two β-strands, and the C-terminal α-helix binds across the intervening major groove, interacting with non-bridging phosphates (Ryter and Schultz 1998; Ramos et al. 2000). When multiple dsRBDs are present in a single protein, or when more than one of the same dsRBD bind to a single RNA, the binding register likely overlaps, as the domains are capable of binding multiple faces of the dsRNA helix (Ucci et al. 2007).

Not all dsRBDs bind to dsRNA. Instead, these non-canonical dsRBDs mediate protein-protein interactions. For example, the C-terminal dsRBDs of both TRBP and PACT are involved in binding to Dicer in humans (Haase et al. 2005; Laraki et al. 2008; Kok et al. 2007; Lee et al. 2006; Daniels et al. 2009). Additionally, in plants Dicer's DUF domain adopts a dsRBD fold and is involved in binding to different protein partners specific to that Dicer isoform (Dlakić 2006; Qin et al. 2010).

1.4.2 dsRBPs in RNAi

As mentioned above, in *Drosophila* two dsRBPs, R2D2 and Loqs, are known to be involved in RNAi. Each protein binds exclusively to a separate Dicer protein and assists with the processing of pre-siRNAs (R2D2) and pre-miRNAs (Loqs-PB) respectively (Liu et al. 2003; Saito et al. 2005; Jiang et al. 2005). R2D2 is not strictly required for dicing of pre-siRNAs, but instead inhibits the processing of pre-miRNAs and is then directly required for siRNA processing at later stages (see **Section 1.5** (Liu et al. 2003; 2006; Tomari et al. 2004b; Cenik et al. 2011; Marques et al. 2010)). Interestingly, for the processing of endogenous pre-siRNAs by Dicer-2, not R2D2, but an isoform of Loqs (Loqs-PD) is required (Marques et al. 2010).

Although in humans there is only one Dicer enzyme, two dsRBPs have been implicated in facilitating RNAi. The first of these dsRBPs is TRBP, which was originally shown to bind to the TAR RNA hairpin of the HIV-1 LTR, inducing expression of the LTR synergistically with the HIV-1 Tat protein (Gatignol et al. 1991). A second dsRBP, PACT, is also involved in RNAi. Both of these proteins bind to an α-helical insertion in Dicer's helicase domain through their C-terminal dsRBDs (Wilson et al. manuscript in preparation; (Kok et al. 2007; Lee et al. 2006; Haase et al. 2005; Laraki et al. 2008; Daniels et al. 2009). The delineation of specific roles for these proteins in human RNAi pathways has thus far been unclear, likely due to the field's tendency to focus exclusively on TRBP and due to a relative lack of sensitive *in vitro* systems to study each step of the human pathway. Knockdown of either TRBP or PACT leads to decreases in both pre-siRNA and pre-miRNA processing as well as downstream gene targeting, indicating that they may function at multiple and overlapping steps in RNAi pathways (Kok et al. 2007; Chendrimada et al. 2005; Haase et al. 2005; Lee et al. 2006).

TRBP is known to enhance Dicer's processing of pre-siRNAs and also facilitates product binding by Dicer (Chakravarthy et al. 2010; Chendrimada et al. 2005). The mechanism by which TRBP enhances pre-siRNA processing is unclear, although it may be involved in allosterically relieving the helicase domain's autoinhibition of this processing (Ma et al. 2008). Interestingly, both TRBP and Loqs-PB have been shown to induce the processing of different isomiRs by Dicer, biasing Dicer cleavage sites to produce miRNAs of slightly differing length. This has been shown to have downstream effects on Ago2 cleavage by altering guide strand selection as well as the seed sequence (see **Section 1.5.2**; (Lee and Doudna 2012; Fukunaga et al. 2012)). Although the mechanism of isomiR formation is unclear, increased production caused by dsRBPs may be

due to the slight underwinding of dsRNA caused by the binding of certain dsRBDs, by the straightening of bulged dsRNAs, or some combination of the two (Zheng and Bevilacqua 2000; Ryter and Schultz 1998).

1.5 RISC loading

Following dicing, siRNAs and miRNAs are loaded into the RNAi effector complexes, RISCs (Hammond et al. 2000). The catalytic core of these complexes is an Ago protein. The best characterized of these proteins in humans is Ago2, a magnesium-dependent, multiple-turnover endonuclease containing an N-terminal domain, a PAZ domain, a Mid domain, and a C-terminal PIWI domain that structurally resembles RNase H (Hutvagner and Zamore 2002; Martinez and Tuschl 2004; Schwarz et al. 2004; Song et al. 2004; Hammond et al. 2001; Liu et al. 2004). *Drosophila* expresses two Ago proteins (Ago1 and Ago2). As with the other *Drosophila* RNAi machinery, the siRNA and miRNA pathways can be delineated based on the propensities of each Ago protein to be loaded with either type of duplex. Ago2 is a functioning endonuclease, and is loaded predominantly with siRNAs, whereas Ago1 lacks robust endonuclease activity and is loaded primarily with miRNAs. The determinants of this loading preference are in this case the specific protein components that bind to each Ago (Förstemann et al. 2007; Tomari et al. 2007; Czech et al. 2009). In mammals, siRNAs and miRNAs do not appear to be differentially sorted into the four Ago proteins (Ago1-Ago4). Ago2, however, is the most highly expressed Ago protein and is the only human Ago that exhibits endonuclease activity (Meister et al. 2004; Liu et al. 2004; Dueck et al. 2012; Wang et al. 2012). Since Ago2 binds to both siRNAs and miRNAs, the delineation of these pathways at this level in humans has been unclear.

1.5.1 RISC-associated proteins

Although implicated as the catalytic endonuclease of RNAi, recombinant human Ago2 has been shown to be incapable of binding to dsRNA on its own (Rivas et al. 2005; Ye et al. 2011). Numerous studies have isolated active "RISCs" of vastly differing size ranging from as small as 160kD to a ribosome-associated, 80S complex (Mourelatos et al. 2002; Gregory et al. 2005; Caudy et al. 2002; Chendrimada et al. 2005; Pham et al. 2004). These complexes contain a host of expected and unexpected proteins, including Dicer (Maniataki and Mourelatos 2005; Gregory et al. 2005); Fragile X-Related Protein, VIG (Caudy et al. 2002); RNA helicases MOV10, Gemin3, and RNA Helicase A (Meister et al. 2005; Höck et al. 2007; Mourelatos et al. 2002; Robb and Rana 2007); the GW protein TNRC6 (Höck et al. 2007; Landthaler et al. 2008); and the dsRBPs TRBP and PACT (Haase et al. 2005; Chendrimada et al. 2005; Lee et al. 2006; Kok et al. 2007). For years little was known about the ways in which these proteins collaborate to facilitate target knockdown and at what steps, although more recently a clearer picture of the proteins involved in the RISC loading process has begun to emerge.

1.5.2 RISC loading

Across organisms, heat shock proteins facilitate optimal duplex RNA binding by Ago2 (Miyoshi et al. 2010; Iwasaki et al. 2010; Iki et al. 2010; Pare et al. 2009; Johnston et al. 2010). In *Drosophila,* RISC loading also requires a preformed complex called the RISC Loading Complex (RLC) containing Ago, Dicer, and a dsRBP. The existence of such a complex was first described in flies with the observation that R2D2, a dsRBP that binds to Dicer-2, functionally links dicing and RISC activity (Liu et al. 2003). Following cleavage, Dicer-2 releases its product siRNAs and is unable to rebind them in the absence of R2D2 (Liu et al. 2003; Preall et al. 2006). Dicer-2 and

R2D2 bind siRNA products coordinately, which is required for subsequent loading of siRNAs onto Ago2 (Liu et al. 2003; 2006). Interestingly, Dicer-2 and R2D2 serve as "gatekeepers" for Ago2 loading that promote siRNA loading, but select against miRNA loading. Conversely, miRNAs are preferentially loaded onto Ago1 by a poorly understood mechanism (Tomari et al. 2007).

Similar RISC Loading Complexes have been shown to exist in humans. In the minimal human RLCs, the Ago2 PIWI domain binds to Dicer's RNase IIIa domain, and the C-terminal dsRBD of TRBP (or PACT) binds to Dicer's DExH/D helicase domain (Tahbaz et al. 2004; Sasaki and Shimizu 2007; Haase et al. 2005; Kok et al. 2007). Beyond these biochemical studies, the molecular architecture of RLCs and how the structural arrangement of its component parts relates to RISC loading has not been investigated. Despite the clear existence of RLCs in humans, their strict requirement for Ago2 loading has been less apparent. When the human RNAi system is reconstituted in *Saccharomyces cerevisiae*, Dicer and TRBP are required for Ago2 loading (Suk et al. 2011). Furthermore, in certain respects TRBP has been shown to be analogous to R2D2, in that it is required for Dicer to bind siRNA products and potentially transfers those siRNAs to Ago2 (Chendrimada et al. 2005; Chakravarthy et al. 2010). On the other hand, mouse embryonic stem cells that are Dicer-deficient are still capable of siRNA-mediated gene knockdown, indicating that Ago2 is somehow loaded in these cells (Murchison et al. 2005; Kanellopoulou et al. 2005). Despite this observation, in a wild type setting Ago2/Dicer/dsRBP complexes have been shown to be the primary complexes involved in RISC loading in humans (Liu et al. 2012). However, whether Dicer and dsRBPs play important roles in RISC loading within these complexes remains an open question.

Numerous crystal structures of all or part of human Ago2 as well as various Ago homologs have demonstrated how loaded duplex RNAs are bound. This structural data indicates that the guide strand has its 3′ end bound by the PAZ domain and its 5′ end is anchored within a conserved basic pocket between the Mid and PIWI domains. Anchoring of the 5′ end of the guide RNA is critical for proper target cleavage (Yuan et al. 2005; Parker et al. 2004; Wang et al. 2008b; 2008c; 2009c; Ma et al. 2005; Nakanishi et al. 2012; Elkayam et al. 2012; Schirle and Macrae 2012). This pocket makes base-specific contacts with the 5′ nucleotide of the guide RNA, with a preference for a 5′ uridine or 5′ adenosine as opposed to cytidines or guanosines – a structural feature that was expected based on sequencing of Ago-bound small RNAs (Hu et al. 2009; Frank et al. 2010; Wang et al. 2009c; Elkayam et al. 2012). Once Ago2 has been loaded with a duplex RNA, the first two nucleotides of the guide are splayed apart from their complementary bases, which is supported by kinetic data indicating that 5′ terminal mismatches enhance Ago2 activity (Haley and Zamore 2004; Ma et al. 2005; Wang et al. 2008c; Parker et al. 2005; Elkayam et al. 2012). The guide strand directs the endonucleolytic cleavage of the passenger strand at the phosphodiester bond that lies directly across from the 10^{th} and 11^{th} nucleotides of the guide RNA, as measured from the 5′ end (Rand et al. 2005; Kim et al. 2007; Matranga et al. 2005; Leuschner et al. 2006). This processing site lies within the Ago2 PIWI domain and cleavage occurs via a catalytic tetrad consisting of Asp-Asp-Asp-Glu residues (Liu et al. 2004; Rivas et al. 2005; Wang et al. 2009c; Nakanishi et al. 2012). Based on these structural studies, it is important to note that the specific orientation of an siRNA or miRNA when bound to Ago2 is essential for the cleavage of the passenger strand as opposed to the guide strand. Loading of Ago2 in the wrong orientation would thus lead to the downstream targeting of an entirely different set of genes.

In the case of miRNAs, the passenger strand (also referred to as miRNA*) is not cleaved, but rather is removed from the complex intact (Kawamata et al. 2009). Argonaute itself makes few contacts with the passenger strand, priming it for release (Parker et al. 2005; Ma et al. 2005). Human and *Drosophila* Ago2, however, is thought to be incapable of releasing passenger strands on its own (Wang et al. 2009a). This process is facilitated by another magnesium-dependent endoribonuclease, C3PO, which has a strong stimulatory effect on RISC activity (Liu et al. 2009). C3PO is a multimeric complex composed of 6 copies of the protein Translin and two copies of TRAX. These proteins form an asymmetric barrel that places the complex's active site residues on the positively charged interior of the barrel. The mechanism by which this complex removes and degrades passenger strand RNA is still unclear (Ye et al. 2011; Tian et al. 2011).

Once the passenger strand has been removed from Ago2, the resulting single-stranded guide RNA-containing complex, termed "holo-RISC", is competent for the multiple-turnover targeting of mRNAs that are complementary to the guide strand (Kim et al. 2007; Martinez et al. 2002; Hutvagner and Zamore 2002). Structures of holo-RISC have shown that bases 2-6 of the guide strand RNA (called the "seed" sequence) remain solvent-exposed, which has important implications for target recognition. The complementarity of these bases with target mRNAs is absolutely essential for that process, and thus seed sequence pairing is the most accurate predictor of a miRNA target (Yuan et al. 2005; Haley and Zamore 2004; Wang et al. 2008b; Lewis et al. 2003; Selbach et al. 2008; Baek et al. 2008; Lim et al. 2005; Schirle and Macrae 2012; Elkayam et al. 2012; Nakanishi et al. 2012). As previous studies had alluded to, for effective target knockdown this complementarity generally must reside at either end of the 3′ UTR of a targeted gene (at least 15 nt from the termination codon). Additionally, miRNA-mediated gene silencing appears to be tunable based on the number of seed sequences present in close proximity within a given 3′ UTR, with multiple sites acting synergistically (Selbach et al. 2008; Doench et al. 2003; Grimson et al. 2007; Baek et al. 2008). These parameters (and the genomic studies that have validated them) reveal that each miRNA can directly regulate targets that number in the 100s, with over 30% of all human genes under the exquisite control of RNAi (Lewis et al. 2005; Lim et al. 2005; Selbach et al. 2008; Baek et al. 2008).

1.5.3 Strand selection parameters

Cloning of large populations of natural miRNAs led to the discovery that mature miRNA guide strands can derive from either end of their precursor hairpin duplex (Lee and Ambros 2001; Lagos-Quintana et al. 2001; Lau et al. 2001). Additionally, although each strand of an siRNA duplex could theoretically have an equal chance of being incorporated into RISC as the guide strand, this was often not the case (Schwarz et al. 2003; Khvorova et al. 2003). One of the most pressing mechanistic questions in the field of RNAi therefore involves determining how the guide strand is detected so that siRNAs and miRNAs are loaded onto Ago2 in the proper orientation to allow for retention of that strand. Ideally, a complete understanding of the parameters involved in strand selection will allow for the design of siRNAs that incorporate the desired RNA strand into RISC to the exclusion of the other.

One of the most salient strand selection parameters has to do with the thermodynamics of siRNA or miRNA duplex ends. The free energy of duplex unwinding can be calculated based on the nearest neighbor model, which takes into consideration the stabilities of each base pair within the context of its nearest neighbor (for example, two GC pairs separated by an AU pair are less stable then two neighboring GC pairs; see **Figure 1.3**; (Tinoco et al. 1971; 1973; Borer et al. 1974; Xia et al. 1998; O'Toole et al. 2005; Mathews et al. 1999; Freier et al. 1986)). This model

can be used to calculate the free energy differences between the duplex ends by determining the relative contributions that the 2 nt overhangs and last four base pairs at either end make to the overall duplex stability. Calculating the end stabilities in such a way allows for accurate guide strand prediction. When both duplex ends have similar stabilities, either strand is able to be loaded into RISC. However, when the siRNA ends are thermodynamically asymmetric, that is, when one end of the duplex is more stable than the other, the strand for which the 5′ end lies at the less stable end of the duplex will become the guide strand (Schwarz et al. 2003; Khvorova et al. 2003). The exact mechanism by which the RLC senses thermodynamic asymmetry of siRNA ends is unclear, but research in *Drosophila* has provided some clues. R2D2 and Dicer-2 bind to nascent siRNAs in a directional manner such that R2D2 consistently binds to the more stable siRNA end and Dicer-2 binds to the less stable end. At the time it was suggested that R2D2 is the actual sensor of thermodynamic asymmetry and recruits the less stable duplex end to Dicer-2. This conclusion was based on the observation that Dicer-2 cannot bind to siRNAs on its own (Tomari et al. 2004b). Additionally, although directional binding of siRNAs by human Dicer and TRBP or PACT has not been shown, TRBP has been suggested to sense the thermodynamic stabilities of siRNA (Gredell et al. 2010). The possibility that Dicer is in fact the sensor of thermodynamic asymmetry was never directly tested in either of these cases. The mechanism of strand selection in humans is unclear. Although Ago2/Dicer/dsRBP RLCs have been shown to be the primary complexes involved in RISC loading in humans (Liu et al. 2012), separate work has demonstrated that in Dicer knockout cells Ago2 is in some cases able to sense the proper guide strand (Betancur and Tomari 2012). The extent to which human Ago2/Dicer/dsRBP complexes are actually true RLCs that contribute to strand selection has therefore been a topic of vigorous debate.

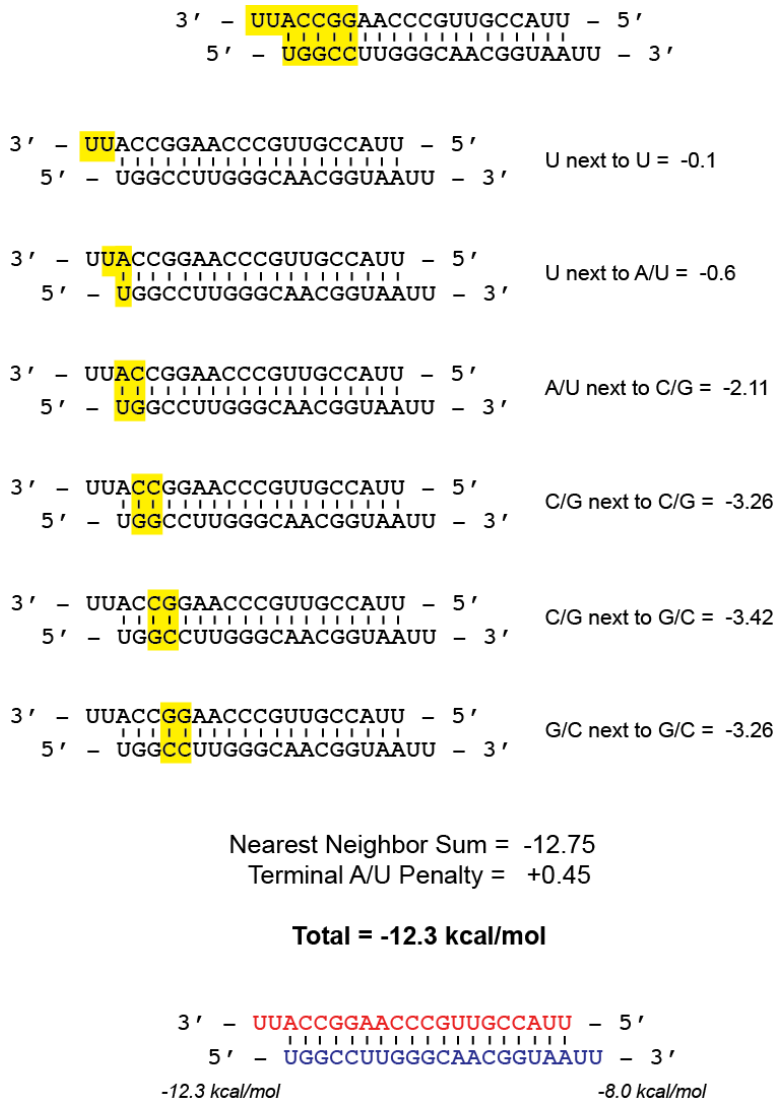

Figure 1.3: Nearest neighbor calculations of siRNA duplex end stabilities.
Free energy (kcal/mol) was calculated for the terminal 4 base pairs and the 2 nt 3′ overhang for each end of the duplex using experimentally-derived values (Xia et al. 1998; Mathews et al. 1999; Tinoco et al. 1973; Borer et al. 1974; O'Toole et al. 2005; Freier et al. 1986). Calculation steps for the left end of the duplex shown are given as an example. Based on these calculations, the top strand of the duplex is predicted to be the guide strand (shown in red).

1.6 Modes of gene regulation by RNAi

1.6.1 mRNA cleavage

Of the four human Ago proteins, both Ago1 and Ago2 are capable of cleaving the passenger strand, whereas only Ago2 is competent to cleave target mRNAs (Meister et al. 2004; Wang et al. 2009a). mRNA cleavage requires extensive complementarity between the guide RNA and its target. A number of factors are structurally important for coordinating this cleavage. 5′ complementarity is particularly important, whereas guide-target mismatches are tolerated better at the 3′ end of the guide strand. In fact, these 3′ mismatches promote product release and thus multiple turnover (Haley and Zamore 2004; Liu et al. 2011). Target binding is thought to occur in two steps. During nucleation of the seed sequence, the mRNA target begins to pair with the exposed nucleotides 2-6 of the guide RNA, and the guide strand is anchored at its 5′- and 3′ ends

by the Mid/PIWI and PAZ domains, respectively. As the duplex propagates, the 3′ end of the guide strand is then released due to topological restraints. This release likely also allows for an optimal geometry within the catalytic site (Wang et al. 2009c). Much like the passenger strand, target mRNAs are cleaved at the phosphodiester bond ten nucleotides from the 5′ end of the guide strand (Elbashir et al. 2001c; 2001b; Rivas et al. 2005). This cleavage is mediated by the catalytic tetrad, which is thought to coordinate two magnesium ions to activate a water for nucleophilic attack on the phosphate backbone of the target RNA and stabilize the leaving group (Wang et al. 2009c; Nakanishi et al. 2012). In humans, multiple-turnover mRNA targeting is facilitated by Autoantigen La, which removes cleaved mRNA products (Liu et al. 2011).

1.6.2 Mechanisms of miRNA-mediated translational repression

The mechanism of gene knockdown by miRNAs is far less clear-cut than for siRNAs. Although some miRNAs with extensive complementarity to their target genes are able to direct mRNA cleavage, this is generally not the case (Okamura et al. 2007; Yekta et al. 2004). Instead of directly cleaving target mRNAs, miRNAs are known to silence genes by translational repression and mRNA deadenylation and decay. These functions require the recruitment of Ago2 to cytoplasmic processing bodies, or P-bodies (Liu et al. 2005b; Sen and Blau 2005; Pillai et al. 2005). Given that translation and mRNA stability are intricately related, the relative contributions of each form of repression to miRNA-directed gene silencing has been the topic of lively debate for years (discussed in **Section 1.6.4**). This section will focus on the mechanisms of translational repression by miRNAs.

In humans, translational repression by miRNAs was first reported for the miRNA *let-7*, which exclusively represses cap-dependent translation of its target. At the time no change in mRNA levels was observed. Furthermore, repressed mRNAs were shifted to lighter fractions of polysomal density gradients, which indicated that repression occurred at the level of translation initiation (Pillai et al. 2005). miRNA-mediated translational repression requires the glycine-tryptophan repeat-containing protein GW182 (TNRC6 in humans), which binds directly to Ago2 (Eulalio et al. 2008). This interaction is thought to affect 5′ cap-dependent mRNA circularization during translation initiation. Normal cap-dependent translation requires the circularization of mRNA transcripts via Poly(A)-Binding Protein (PABP) interacting with eIF4G, which strengthens the cap-binding protein eIF4E's interaction with the 5′ cap. This causes mRNA transcripts to form a closed-loop that is essential for optimal translation efficiency. Translational repression by miRNAs can be partially explained by the finding that GW182 binds to PABP and interferes with the PABP-eIF4G interaction, thereby disrupting the closed-loop of the mRNA transcript and affecting cap binding by eIF4E (**Figure 1.4 A**; (Fabian et al. 2009; Zekri et al. 2009; Humphreys et al. 2005; Mathonnet et al. 2007; Thermann and Hentze 2007)). This likely has the synergistic effect of exposing the mRNA ends to deadenylase and decapping complexes (see **Section 1.6.3**). Beyond the effects of miRNAs on cap binding, RISC has also been shown to associate with eIF6. This factor prevents joining of the 40S and 60S ribosomal subunits, and its depletion causes the de-repression of miRNA targets (Chendrimada et al. 2007). As would be expected based on this finding, targeted mRNAs are known to be enriched for 40S, but not 60S ribosomal subunits (Wang et al. 2008a). These findings suggest dual modes of miRNA-mediated repression of translation initiation.

In addition to the evidence linking miRNAs to a decrease in translation initiation, some studies have pointed to an effect on the elongation stage of translation. This was based on the fact that some miRNAs were found to be associated with polysomes following sucrose gradient

sedimentation (Nottrott et al. 2006; Petersen et al. 2006; Maroney et al. 2006). A more recent study demonstrated that these populations, however, represent non-polysomal mRNPs that simply co-sediment with polysomes even under conditions where ribosomal subunit joining is blocked (Thermann and Hentze 2007). It therefore appears that the true effect of miRNAs on translation lies at the initiation stage (**Figure 1.4 B**).

1.6.3 Mechanisms of miRNA-mediated mRNA decay

In addition to having an effect on translation initiation, substantial evidence suggests that miRNAs destabilize their target mRNAs. miRNA-mediated mRNA deadenylation was first demonstrated in zebrafish, where it was shown that miR-430 expression early in embryogenesis promotes the deadenylation of hundreds of maternal mRNAs, leading to their degradation (Giraldez et al. 2006). miRNA-directed deadenylation has since been shown to be widespread in other organisms (Guo et al. 2010; Wu et al. 2006; Eulalio et al. 2009; Wu et al. 2010). As with translational repression, miRNA-mediated mRNA decay also requires GW proteins, which serve as a platform for recruiting mRNA degradation machinery to Ago2-bound transcripts (Eulalio et al. 2008; Liu et al. 2005a; Chen et al. 2009). GW proteins recruit both the CCR4-NOT and PAN2-PAN3 deadenylase complexes to Ago2-bound mRNAs (**Figure 1.4 C**; (Chen et al. 2009; Behm-Ansmant et al. 2006; Fabian et al. 2009; 2011)). GW182 binding to PABP is also essential for deadenylation, likely bringing the poly(A) tail in close proximity to the deadenylase complexes (Fabian et al. 2009; Zekri et al. 2009; Huntzinger et al. 2010; Jinek et al. 2010). As a result of translational inhibition and/or deadenylation, the 5′ cap is exposed for decapping by the DCP1-DCP2 complex, and subsequent 5′-3′ degradation by Xrn1 is triggered (**Figure 1.4 D**; (Chen et al. 2009; Behm-Ansmant et al. 2006; Rehwinkel et al. 2005)).

1.6.4 Ordered, biphasic gene silencing by miRNAs

Although evidence for both miRNA-mediated translational inhibition and mRNA decay is abundant, these processes are closely linked, and thus questions lingered for some time as to the relative importance of each type of inhibition on the overall miRNA-directed repression profile. It was unclear if translational inhibition occurs first, triggering mRNA decay, or if the deadenylation associated with mRNA decay was the true trigger for miRNA-mediated translational repression. Ultimately, the sequence of events during miRNA-mediated gene silencing has proven quite difficult to assess *in vivo*, and thus this subtle distinction has for years been a topic of debate.

The first studies to address this question on a global scale used stable isotope labeling with amino acids in cell culture (SILAC) and microarrays to simultaneously examine the effects of miRNAs on protein and mRNA levels (Baek et al. 2008; Selbach et al. 2008). Based on the fact that changes in protein levels as measured by SILAC and changes in mRNA levels as measured by microarray were similar, these groups concluded that mRNA destabilization is likely responsible for the bulk of the loss in protein synthesis observed in response to miRNAs. This observation was true for the majority of miRNA targets, although a small subset of genes emerged in each case that appeared to be regulated exclusively at the level of translation. Using ribosome profiling, which is less biased towards well-expressed proteins, one of these groups went on to show that for at least 84% of miRNA targets, the major contributor to decreased protein expression was mRNA decay (Guo et al. 2010). For the remaining 16% of targets inhibited through translational repression, the possibility existed that partial deadenylation and the resulting disruption of the mRNA closed-loop was the true cause of repression – a case where

the methods used would have artifactually indicated that translational repression was the major cause of repression.

A major caveat of the above studies is that they were carried out at steady state and were therefore insufficient to determine the initial steps of miRNA-mediated gene silencing. Even if mRNA decay was the major contributor to silencing at later time points, this destabilization could have been triggered initially by translational repression. Indeed, early kinetic analyses demonstrated that miRNA-induced translational inhibition occurs within 15 minutes in cell extracts, whereas deadenylation begins later (Mathonnet et al. 2007; Fabian et al. 2009). These studies, however, were conducted *in vitro* using cell lysates, leading to speculation that they may not accurately reflect the *in vivo* repression dynamics. It was not until recently that the kinetic resolution of these repression events was achieved *in vivo* in zebrafish, *Drosophila*, and human cells. Each of these reports independently demonstrated that translational repression precedes mRNA decay by assaying protein expression and mRNA levels at time points as early as one hour (Béthune et al. 2012; Bazzini et al. 2012; Djuranovic et al. 2012). Importantly, early translational repression was not accompanied by even minor deadenylation, ruling out the possibility that partial deadenylation would break the closed-loop of the mRNA, thereby causing translational repression. Furthermore, in one study ribosome density along repressed mRNAs was probed by ribosome profiling at early timepoints. miRNA repression led to an overall reduction in ribosome occupancy, but a uniform density along the transcript, bolstering the developing consensus that miRNAs direct translational repression at the initiation step rather than by affecting elongation or by causing ribosome drop-off (Bazzini et al. 2012). Taken together, these studies indicate that miRNA-mediated silencing is biphasic, beginning at the level of direct repression of translation initiation. This repression then triggers mRNA deadenylation and decay, which becomes the dominant effect at steady state (**Figure 1.4**).

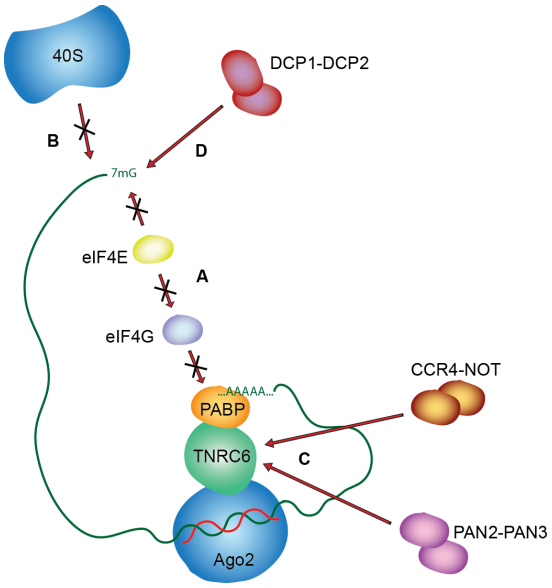

Figure 1.4: miRNA-mediated translational repression and mRNA decay.
(A) Following the targeting of a mRNA 3´ UTR by Ago2 and the GW protein TNRC6, PABP recruitment breaks the closed-loop of the mRNA by preventing eIF4G from interacting with PABP, which in turn reduces the affinity of eIF4E for the cap. **(B)** As a result, translation initiation is inhibited. **(C)** TNRC6 also recruits the CCR4-NOT and PAN2-PAN3 deadenylase complexes, placing them in the proximity of the poly-(A) tail with the help of PABP. **(D)** The transcript is then exposed to the decapping complex DCP1-DCP2.

1.7 Research rationale

Scientific interest in RNAi pathways has grown exponentially since their discovery in 1998. Despite this, ten years later when I joined the Doudna Lab numerous gaps still existed in our mechanistic understanding of these systems, particularly in humans. A paucity of structural data existed for the individual human components of the RNAi machinery and how they come together as a functional whole. Striking differences were also emerging between human RNAi systems and those systems in which RNAi was originally studied, in particular necessitating a detailed analysis of the RISC loading process in humans. Furthermore, the notion that a human RISC Loading Complex *per se* even existed had been called into question, which would require a thorough analysis of the individual contributions of each component part of the core RNAi machinery to overall gene silencing.

For the most part, structural information regarding the RNAi machinery has been piecemeal. A wealth of structural information now exists for both prokaryotic and eukaryotic Ago proteins – including human Ago2 – which has provided a detailed picture of how these proteins bind guide RNAs and direct the cleavage of cognate mRNAs (Elkayam et al. 2012; Schirle and Macrae 2012; Nakanishi et al. 2012; Wang et al. 2009c; 2008b; 2008c; Frank et al. 2010; Parker et al. 2005; Ma et al. 2005). In addition to Argonaute, the structure of a *Giardia* Dicer protein was pivotal to our understanding of how Dicer proteins recognize, measure, and cleave their dsRNA substrates into siRNAs and miRNAs of the expected size (Macrae et al. 2006). Despite all of this structural information, one glaring void has been structural data that tell us how the proteins that comprise the core RNAi machinery interact functionally. In other words, imagining dicing as Point A and siRNA/miRNA binding by Ago2 as Point B, no structural information that might indicate how a Dicer product gets from Point A to Point B existed. Based on biochemical data and analogy to the *Drosophila* system, it seemed clear that RLCs comprised of Ago2, Dicer, and TRBP or PACT might be involved. The interactions between each of these proteins had been mapped to specific domains of each protein, but knowledge of how these interactions result in a macromolecular protein complex capable of orienting Dicer products based on subtle duplex thermodynamics and transferring them to Ago2 was out of grasp. The aim of the work presented in **Chapter 2** was to break this barrier by determining the 3D molecular architecture of a core human RLC composed of Ago2, Dicer, and TRBP using electron microscopy and single particle reconstruction. This study yielded a view of a conformationally dynamic complex that gave insights into the overall RISC loading process.

The information garnered from this initial structural study led to a series of biochemically testable hypotheses as to how dsRNAs are processed by – and transferred through – this incredibly dynamic protein complex. Previous studies of the RISC loading process in *Drosophila* informed these hypotheses – a decision driven by an intuitive sense that these two systems simply were not as different as they seemed (Preall et al. 2006; Tomari et al. 2004b; Liu et al. 2003; 2006). The experiments presented in **Chapter 3** began to answer these questions biochemically, detailing how information gained for the human system makes mechanistic sense based on our EM reconstruction. A view of how siRNA orienting by Dicer/dsRBP complexes might lead to strand-specific Ago2 loading in humans emerged from this study.

Finally, a topic of lively debate in the RNAi field has been the extent to which human RLCs are even involved in the RISC loading process at all. Previous studies had shown that human Ago2 is capable of binding siRNAs and targeting genes for silencing in the absence of Dicer (Murchison et al. 2005; Kanellopoulou et al. 2005). Furthermore, it had been shown that – at least in some cases – Ago2 can select the proper guide strand of an siRNA or miRNA in the

absence of Dicer (Betancur and Tomari 2012). Despite this, certain evidence pointed to human RLCs being the main players in RISC loading in wild type cells (Liu et al. 2012). **Chapter 4** expands on these initial studies to show that original conclusions that the formation of RLCs in humans was inconsequential did not fully appreciate the true complexity of the RISC loading process. These expanded findings solidified the importance of human RLCs in RISC loading and had the unexpected result of unearthing specific roles for TRBP and PACT in both strand selection and target cleavage that are reminiscent of the functional specialization of dsRBPs in *Drosophila*.

Chapter 2

Electron Microscopic Analysis of the Human RISC Loading Complex*†

* A portion of the work presented in this chapter has previously been published as part of the following paper: Wang, H-W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., Nogales, E. (2009). Structural insights into RNA processing by the human RISC-loading complex. *Nature Structural & Molecular Biology* 16, 1148-1153.

† Dr. Hong-Wei Wang planned the experiments and performed EM analysis of various samples. Cameron Noland purified core proteins, reconstituted RLC complexes, and performed the preparation and analysis of the glutaraldehyde crosslinked complexes by EM. Dr. Bunpote Siridechadilok performed the preparation and analysis of the uncrosslinked RLC by EM. David Taylor performed the preparation and analysis of the GraFix crosslinked RLC by EM. Dr. Enbo Ma purified core proteins and reconstituted RLC complexes. Dr. Karin Felderer prepared and analyzed the human Dicer sample by EM.

2.1 Introduction

High-resolution structural information exists for a variety of single RNAi components and has been essential to informing our understanding of the key mechanistic steps of RNAi pathways. For example, the structures of human Ago2 bound to a guide RNA as well as homologous structures have provided a detailed understanding of how duplex binding and mRNA targeting occurs (Elkayam et al. 2012; Schirle and Macrae 2012; Nakanishi et al. 2012; Wang et al. 2009c). Additionally, the structure of a *Giardia intestinalis* Dicer protein has proven essential to our understanding of how this protein positions, measures, and cleaves pre-siRNAs to yield product siRNAs of a discrete size (Macrae et al. 2006). However, certain insights from this original Dicer structure do not translate to the human system. For example, subtle structural differences within domains that are conserved between *Giardia* and human Dicer proteins have dramatic effects on processes such as substrate recognition and product size measuring (Park et al. 2011). Furthermore, the *Giardia* Dicer protein is highly truncated, at only ~23% the size of human Dicer. Several domains found in human Dicer are not present in *Giardia*, namely the DExH/D helicase domain, the DUF domain, and the C-terminal dsRBD.

One of the key structural problems for the RNAi field regards determining how Dicer functionally interacts with other core RNAi proteins. Human Dicer is known to associate with Ago2 and either TRBP or PACT to form RLCs, and it is currently unclear how these complexes may achieve the hand-off of Dicer products to Ago2. In particular, it is unknown how nascent siRNAs might be oriented based on the thermodynamics of the duplex and loaded onto Ago2 in an orientation that allows for retention of the guide strand. Knowledge of the 3D structure of human RLCs would likely provide insights into this process. Unfortunately, due the lack of conservation between the *Giardia* and human Dicer proteins, the structure of *Giardia* Dicer is not informative in terms of the formation of these complexes. Biochemical studies have shown that a portion of the Ago2 PIWI domain binds to a region of Dicer's RNase IIIa domain that is not conserved in *Giardia* (Tahbaz et al. 2004; Sasaki and Shimizu 2007). Furthermore, the C-terminal dsRBPs of TRBP and PACT have both been shown to bind to a region within Dicer's helicase domain; a domain that *Giardia* lacks (Haase et al. 2005; Lee et al. 2006; Kok et al. 2007; Laraki et al. 2008; Daniels et al. 2009). This interaction in particular is essential for the RISC loading process in *Drosophila* and has been suggested to be important in humans as well (Liu et al. 2003; Tomari et al. 2004b; Chendrimada et al. 2005; Liu et al. 2006). A structural analysis of human Dicer in isolation and as a component of RLCs would therefore contribute to a more detailed understanding of the connections between the pre-siRNA processing and mRNA targeting steps of RNAi pathways.

In this study, we have used electron microscopy and single particle reconstruction to determine the low-resolution structure of human Dicer, revealing that it has an L-shaped architecture. Another EM structure of human Dicer supporting this architecture was independently solved by another group (Lau et al. 2009). We show that the short base branch of Dicer is composed of the DExH/D helicase domain, whereas the longer branch contains the catalytic domains. Single particle analysis was also used to determine the molecular architecture of an *in vitro* reconstituted human RLC containing Ago2, Dicer, and TRBP. This reconstruction revealed a conformationally dynamic complex in which TRBP binds to the distal end of Dicer's DExH/D helicase domain, and Ago2 binds to Dicer's catalytic branch. Docking of available Ago and Dicer structures into the EM density provided key insights into the transfer and orientation of nascent siRNAs during strand-specific RISC loading in humans.

2.2 Structure determination by electron microscopy

Electron microscopy can be used to determine the three-dimensional structures of macromolecules using a rapidly developing technique called single particle reconstruction. This technique holds many advantages over classical methods of structure determination such as x-ray crystallography and NMR. Single particle reconstruction is often referred to as "crystallography without crystals," the basis of which is rooted in the observation that the set of translationally repeating "images" in a crystal is similar to the image of a random field of isolated molecules (called single particles) that makes up an electron micrograph (Frank 1996). In fact, structure determination by single particle reconstruction holds many advantages over crystallography. One clear advantage is that the notorious "phase problem" of x-ray crystallography is avoided altogether, given that images – as opposed to diffraction – are the source of data in EM. Another advantage of EM is that it enables the study of conformationally dynamic proteins in conditions that mimic their native environment, allowing for the monitoring of biologically relevant protein movements in ways that crystallography simply does not allow for (Kostek et al. 2006; Hall et al. 2007; Leschziner and Nogales 2007). Single particle reconstruction by EM also requires the use of far less protein sample, which is advantageous in cases where protein yield is an issue. In general, an EM grid only requires ~4 μL of sample at ~25-50 nM concentration. Concentrations of at least 1 mg/mL are generally required for crystallography, and a typical sparse matrix screen would require ~100 μL of a sample at that concentration. Although crystallography techniques are continually being developed to allow the use of smaller sample volumes (Villaseñor et al. 2012; Chapman et al. 2011), electron microscopy is still generally the more viable option when considering a sample of low concentration and small volume.

Electron microscopy is not without its drawbacks. One major problem with EM arises when using the technique of negatively-staining samples with heavy metal salts such as uranyl formate, which increases contrast but only provides a picture of the solvent-accessible portion of the molecule, resulting in a significant loss of resolution and interior detail. Another significant problem that is unique to negatively-stained samples is that a high degree of flattening can occur due to air-drying of the sample (see **Section 2.3**). This flattening can in certain cases reach up to 60% of certain dimensions of a protein (Frank 1996). In a more advanced technique called cryo-electron microscopy (cryoEM) the sample is not stained but instead frozen in vitreous ice, circumventing these problems. Reconstructions using cryoEM reflect a macromolecule's native structure, including interior details. Rapid advancements in instrument design, data collection, and data processing techniques have all contributed to the attainment of sub-nanometer resolution structures by cryoEM, and the number of structures that approach atomic resolution continues to grow (Baker et al. 2010). Unfortunately, since protein assemblies scatter electrons only slightly more strongly than vitreous ice (see **Section 2.2.1**), this technique results in a very low signal:noise ratio that requires the collection of large data sets and precludes the study of low molecular weight proteins. Due to this size constraint, cryoEM is currently not amenable to the study of RISC Loading Complexes.

2.2.1 The basics of transmission electron microscopy

The maximal theoretical resolving power of any microscope is a function of the wavelength of the light source being used, and thus the resolution limit for a traditional light microscope is ~200 nm. For comparison, the length of a prokaryotic ribosome is ~20 nm and a peptide bond length is ~1.32 Å. In order to resolve the shorter distances required to image a macromolecular protein complex, a much shorter wavelength particle such as an electron must therefore be used

(Alberts et al. 2002; Berg et al. 2010). In principle, a transmission electron microscope shares many similarities to a traditional light microscope, where light from a light source is first focused by a condenser lens, passes through a specimen, and is then re-focused and magnified by an objective lens, allowing for imaging of the sample. In transmission electron microscopy (TEM), the "light source" is instead an electron source. These electrons travel in a vacuum through a set of electromagnetic condenser lenses, pass through the sample (which is held in place on a support grid), and are then magnified by an electromagnetic objective lens. The resulting electron beam can then be imaged either on a fluorescent screen or charge-coupled device (CCD).

The ultimate image viewed by TEM is dictated by the ways in which electrons interact with the sample. Most electrons from the incident beam will pass directly through a thin sample without interacting with it, while some will be scattered by the atoms in the sample. There are two forms of this scattering: "elastic" and "inelastic." Elastic scattering involves no transfer of energy to the sample, and the angle of such scattering contains high-resolution information about the sample itself. In inelastic scattering, energy is transferred to the sample and excites secondary electrons, which contribute to noise in the image and radiation damage of the sample. The ratio of elastic to inelastic scattering is tied to atomic number, where elastic scattering dominates at higher atomic number. This principle is problematic for the imaging of biological samples, which are composed of atoms with low atomic numbers. This leads to a situation where extremely poor contrast is achieved at electron doses that are low enough to limit radiation damage of the sample. As mentioned above, heavy metal staining of the sample is one way to partially alleviate this problem (Frank 1996).

2.2.2 Two-dimensional analysis by EM

Single particle reconstruction involves the task of deriving the 3D structure of a macromolecule from a set of theoretically random two-dimensional (2D) projection images of that molecule in every possible orientation. A prerequisite for 3D structure determination is therefore an analysis of these 2D particle projections and their in-plane angular relationships to one another. This process involves first computationally boxing/selecting each individual particle in an electron micrograph. Even using negative stain, the low signal:noise ratio for single particles poses a formidable problem in terms of 2D (and 3D) analysis. Large amounts of data with sufficient redundancy are required such that averaging and classification methods will boost the signal of each particle view. As a rule of thumb >10,000 particles are often used for negative stain and >40,000 particles for cryoEM data, although sample heterogeneity can easily push those limits to >100,000 particles.

Once a sufficient number of particles have been boxed, the goal is to class each particle according to its specific orientation on the EM grid and then average together particles that fall into the same classes to increase the signal:noise ratio. Classification is achieved computationally through iterative rounds of multivariate statistical analysis (MSA) and multi-reference alignment (MRA; (van Heel 1984)). This process typically begins with a reference-free alignment of the particles, which is achieved through a pixel-by-pixel cross-correlation analysis of the salient features that define the entire raw particle stack. Similar particles are grouped together as class averages. These classes can then be used as references for a new alignment of the raw particles in a process called MRA. Each raw particle in the stack is compared sequentially to each initial reference class by cross-correlation and placed into the class with the highest fit. In theory, after

several iterations this process results in a set of high-signal class averages, each of which represents a unique view of the macromolecule of interest (Frank 1996).

2.2.3 From 2D to 3D: random conical tilt, projection matching, and refinement

With a set of class averages in hand, the problem becomes one of relating each 2D view of the molecule to its relative orientation in 3D space. Solving this problem requires the determination of the Euler angles for each particle. This task is often initially achieved by the Random Conical Tilt (RCT) method, which involves taking two sets of images for each field of view (Radermacher et al. 1987). In one image, the sample is untilted, and in the other the sample is tilted by ~55°. The relative in-plane angular relationships between the particles are determined during 2D classification and averaging of the untilted particles, as discussed above. Using this angular information as well as the fixed tilt angle, the tilted particles can then be back projected to construct what is called a class volume. This class volume is missing a significant amount of data due to the specific orientation of data collection, so different class volumes must then be merged by cross-correlation analysis to arrive at an initial 3D reconstruction.

The initial model attained through RCT is often refined through a process called projection matching (Penczek et al. 1994). In this method, the RCT reconstruction is filtered to a low resolution and a set of reference projections are computed that encompass the entire Euler sphere (these projections can be thought of as an idealized set of class averages for each possible view of the macromolecule). Each raw particle is then classed with the best-fit reference projection and assigned its Euler angles accordingly. Particles are then back projected to arrive at a refined model. This process is repeated iteratively to attain a final, refined 3D reconstruction.

2.3 Methods

2.3.1 Expression and purification of recombinant TRBP

Human TRBP cDNA was cloned into the vector pHMGWA (Busso et al. 2005) from the entry vector pENTR-TEV-D-TOPO (Invitrogen), with the result that the expressed protein would have an N-terminal 6xHis-MBP-TEV tag. This vector was transformed into *E. coli* BL21 (DE3) cells and transformed cells were grown with shaking (180 rpm) at 37°C in 750 mL LB medium supplemented with 0.1% glucose and 100 μg/mL ampicillin until the cells reached an OD of 0.6. Expression was then induced by the addition of 0.4 mM IPTG and the cells were grown at 16°C overnight. Cells were pelleted by centrifugation at 4,000 rpm for 20 minutes at 4°C and resuspended in 30 mL Lysis Buffer (20 mM HEPES, pH 7.5; 500 mM NaCl; 10% glycerol; 10 mM imidazole; 0.5% Triton X-100; 1X EDTA-free protease inhibitor cocktail (Roche); and 1 mM TCEP). Cells were then lysed by two passes through a French press cell homogenizer. Lysates were cleared by centrifugation at 16,500 rpm for 30 minutes at 4°C. Supernatants were then applied to 5 mL nickel-NTA Superflow resin (Qiagen) that had been pre-equilibrated with Wash Buffer (20 mM HEPES, pH 7.5; 500 mM NaCl; 10% glycerol; 30 mM imidazole; and 1 mM TCEP) and rotated gently at 4°C for 30 minutes. The resin was then spun down at 2,400 rpm for 5 minutes and the supernatant removed prior to 4 washes in batch with 45 mL Wash Buffer. Bound proteins were eluted with 15 mL Elution Buffer (20 mM HEPES, pH 7.5; 500 mM NaCl; 10% glycerol; 300 mM imidazole; and 1 mM TCEP) by rotating for 20 minutes at 4°C. To remove the 6xHis-MBP tag, ~1 mg of 6xHis-tagged TEV protease was added and the sample was dialyzed against Dialysis Buffer (20 mM HEPES, pH 7.5; 500 mM NaCl; 10% glycerol; 10 mM imidazole; and 1 mM TCEP) overnight at 4°C.

Following dialysis, the sample was passed through a 5 mL HisTrap column (GE) to remove the 6xHis-tagged TEV protease, cleaved 6xHis-MBP tag, and any remaining uncleaved protein. The resultant flow through was then dialyzed against Buffer A (20 mM HEPES-KOH, pH 7.5; 150 mM NaCl; 5% glycerol; and 1 mM TCEP) for two hours, replacing the buffer with fresh Buffer A after 1 hour. To remove contaminating nucleic acids, the dialyzed sample was bound to a 5 mL HiTrap Heparin column (GE) that had been pre-equilibrated with Buffer A. TRBP was eluted from the column using a 100 mL linear gradient of NaCl (150 mM to 1 M). Fractions containing TRBP were then pooled and concentrated to ~1 mL using an Amicon Ultra concentrator (Millipore) before application to a HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE) that had been pre-equilibrated with Gel Filtration Buffer (50 mM HEPES-KOH, pH 8.0; 300 mM KCl; 10% glycerol; 1 mM TCEP). Fractions containing TRBP were pooled and concentrated to > 5 mg/mL and flash frozen in liquid nitrogen either as a large pool or in small aliquots depending on the downstream application. This purification strategy yielded a sample that was completely pure as visualized by SDS-PAGE (**Figure 2.1 B**) and contained an additional GSTF sequence at its N-terminus due to the original fusion construct.

2.3.2 Expression and purification of recombinant human Dicer

Human Dicer cDNA was cloned into a pFastBac vector (Invitrogen) using the restriction sites SfoI and XhoI. The purified vector was then transformed into DH10Bac *E. coli* cells (Invitrogen) and plated on LB plates containing 50 μg/mL kanamycin, 7 μg/mL gentamycin, 10 μg/mL tetracycline, 33 μg/mL X-Gal, and 0.1 mM IPTG. Colonies were grown in the dark at 37°C for 48-72 hours and a white colony was selected, which contained recombinant bacmid DNA. SF9 cells in a 6-well dish containing 2 mL cells/well at 10^6 cells/mL were then transfected with 1 μg purified bacmid DNA using the FuGENE HD (Roche) transfection reagent. Following transfection, cells were incubated at 27°C for 72 hours. A primary viral amplification was performed where 500 μL of the transfection supernatant was applied to 10 mL SF9 cells. After an additional 5 days at 27°C, a secondary amplification was performed where 1 mL of the primary amplification supernatant was applied to 100 mL SF9 cells and incubated at 27°C for 5 days. This process yielded high titer baculovirus expressing 6xHis-TEV-Dicer.

To induce large-scale Dicer expression, High5 cells (0.6 x 10^6 cells/mL) were infected with baculovirus (10 mL/L) for 48 hours. Cells were then harvested by centrifugation at 2,000 rpm for 20 minutes at 4°C and resuspended in 35 mL Lysis Buffer. Cells were gently lysed by 12 strokes using a B pestle in a 40 mL Dounce homogenizer. Lysates were then cleared by ultracentrifugation at 40,000 rpm for 1.5 hours at 4°C. Cleared lysates were applied to 5 mL nickel-NTA Superflow resin and a nickel purification, TEV protease cleavage, and an orthogonal nickel purification were each carried out in a manner that was identical to that performed for TRBP.

Flow through from the HisTrap column was dialyzed against HiTrap Q Buffer A (20 mM HEPES-KOH, pH 7.5; 150 mM KCl; 5% glycerol; and 1 mM TCEP) for two hours. Buffer was removed and replaced with fresh buffer after the first hour of dialysis. Following dialysis the sample was applied sequentially to 5 mL HiTrap S and HiTrap Q columns (GE) connected in tandem that had been pre-equilibrated with HiTrap Q Buffer A. Dicer flowed through the HiTrap S column and bound to the HiTrap Q column. Dicer was eluted from the column using a 100 mL linear gradient of KCl (150 mM to 1 M). Fractions containing Dicer were then pooled and concentrated to ~1 mL using an Amicon Ultra concentrator (Millipore) before application to a HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE) that had been pre-equilibrated

with Gel Filtration Buffer. Fractions containing Dicer were again pooled and concentrated to >2 mg/mL and flash frozen in liquid nitrogen either as a large pool or in small aliquots depending on the downstream application. This purification strategy yielded a sample that was completely pure as visualized by SDS-PAGE (**Figure 2.1 B**) and contained an additional GA sequence at its N-terminus due to the nature of the original fusion construct.

2.3.3 Expression and purification of recombinant human Ago2

Human Ago2 cDNA was cloned into the pFastBac vector and high-titer baculovirus was generating using a method identical to that used for Dicer. To induce large-scale Ago2 expression, ES-SF9 cells (10^6 cells/mL) were infected with baculovirus (10 mL/L) for 72 hours. Cells were then harvested by centrifugation at 2,000 rpm for 20 minutes at 4°C and resuspended in 35 mL Ago2 Lysis Buffer (50 mM Na_2HPO_4, pH 8; 300 mM NaCl; 5% glycerol; 10 mM imidazole; 1x EDTA-free protease inhibitor cocktail; 0.5% Triton X-100; and 1 mM TCEP). Cells were gently lysed by 12 strokes using a B pestle in a 40 mL Dounce homogenizer. Lysates were then cleared by ultracentrifugation at 40,000 rpm for 1.5 hours at 4°C. Cleared lysates were applied to 5 mL nickel-NTA Superflow resin that had been pre-equilibrated with High Salt Wash Buffer (50 mM Na_2HPO_4, pH 8; 1 M NaCl; 5% glycerol; 20 mM imidazole; 1 mM TCEP) and rotated gently at 4°C for 30 minutes. The resin was then spun down at 2,400 rpm for 5 minutes and the supernatant removed. Resin was resuspended in High Salt Wash Buffer and applied to a gravity column. The resin was then washed with 200 mL High Salt Wash Buffer followed by 200 mL Low Salt Wash Buffer (the same as High Salt Wash Buffer, but with 300 mM NaCl). Bound proteins were eluted from the column with 25 mL Ago2 Elution Buffer (Low Salt Wash Buffer supplemented with 300 mM imidazole). TEV protease cleavage was performed overnight at 4°C as the sample was dialyzed into Ago2 Dialysis Buffer (50 mM Na_2HPO_4, pH 8; 300 mM NaCl; 5% glycerol; 10 mM imidazole; 1 mM TCEP).

Following dialysis, an orthogonal nickel purification was performed as with TRBP and the HisTrap flow through was concentrated to 5 mL and dialyzed against Ago2 Buffer A (10 mM KH_2PO_4, pH 7.4; 150 mM KCl; 1 mM TCEP) for two hours at 4°C. Buffer A was removed and replaced with fresh buffer after the first hour. This sample was then applied to a 5 mL Bio-Scale CHT5 Ceramic Hydroxyapatite column (Type I; Bio-Rad) and Ago2 was eluted from the column using a 100 mL linear gradient of KH_2PO_4 (10 mM to 500 mM). Fractions containing Ago2 were then pooled and concentrated to ~1 mL using an Amicon Ultra concentrator (Millipore) before application to a HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE) that had been pre-equilibrated with Gel Filtration Buffer. Fractions containing Ago2 were again pooled and concentrated to >2 mg/mL and flash frozen in liquid nitrogen either as a large pool or in small aliquots depending on the downstream application. This purification strategy yielded a sample that was completely pure as visualized by SDS-PAGE (**Figure 2.1 B**) and contained an additional GA sequence at its N-terminus due to the nature of the original fusion construct.

2.3.4 *In vitro* reconstitution of the human RLC with or without TRBP

A RISC Loading Complex binding reaction was carried out in a final volume of 500 μL Gel Filtration Buffer containing 3 nmol Dicer, 6 nmol Ago2, and 11 nmol TRBP (assuming a dimer). In the case of the Dicer/Ago2 complex, TRBP was left out of the reaction. Reactions were incubated for 30 minutes at 4°C and then applied to a Superose 6 10/300 GL column (GE). A representative chromatogram is shown in **Figure 2.1 A**. For the RLC containing TRBP, fractions containing Dicer, Ago2, and TRBP were conservatively pooled to ensure that no free Ago2 or

TRBP was present, giving the purified sample shown in **Figure 2.1 B**. For Dicer/Ago2 complexes, fractions were pooled conservatively to ensure that no free Ago2 was present. Samples were concentrated to ~5 μM and flash frozen in small aliquots.

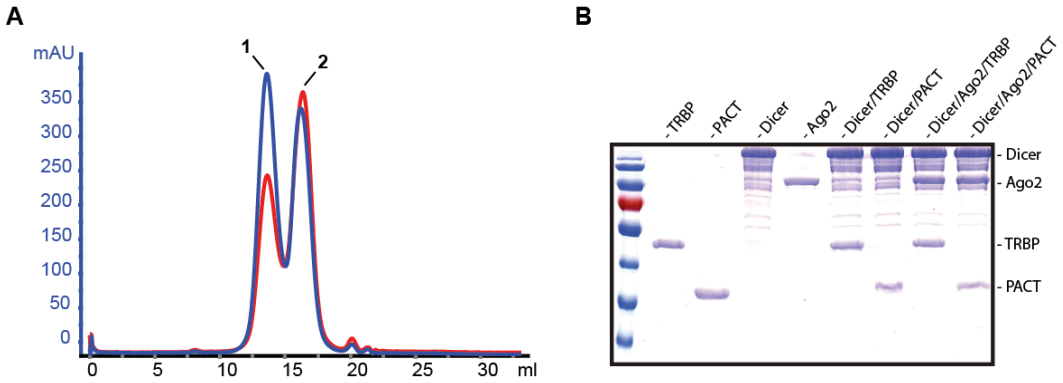

Figure 2.1: Purification of recombinant core human RNAi proteins and complexes.
(A) Superose 6 elution profile for the RISC Loading complex containing TRBP. Peak 1 contains Dicer, TRBP and Ago2 in complex, while Peak 2 contains free Ago2 and TRBP. Blue and red lines reflect the absorbance of the fractions at 280 and 260 nm wavelengths, respectively **(B)** SDS-PAGE analysis of purified core RLC proteins (TRBP, PACT, Dicer, and Ago2), Dicer/dsRBP heterodimers (Dicer/TRBP and Dicer/PACT), and RISC Loading Complexes (Ago2/Dicer/TRBP and Ago2/Dicer/PACT).

2.3.5 Glutaraldehyde crosslinking of protein complexes

A 25% EM grade glutaraldehyde stock (Sigma) was used to make a series of 1 mL secondary stocks containing 100 mM HEPES-KOH, pH 7.5 and 0-0.5% glutaraldehyde. For crosslinking reactions to test for the optimal glutaraldehyde crosslinking concentration, 5 μg RLC was combined with increasing amounts of glutaraldehyde (0-0.35%) and brought to a final volume of 14 μL with Gel Filtration Buffer. Reactions were incubated for 10 minutes at room temperature before quenching with 1 μL 1 M Tris-HCl, pH 7.5 followed by an additional 5-minute incubation at room temperature. Samples were then analyzed by SDS-PAGE to determine the optimal glutaraldehyde concentration for crosslinking (**Figure 2.2**).

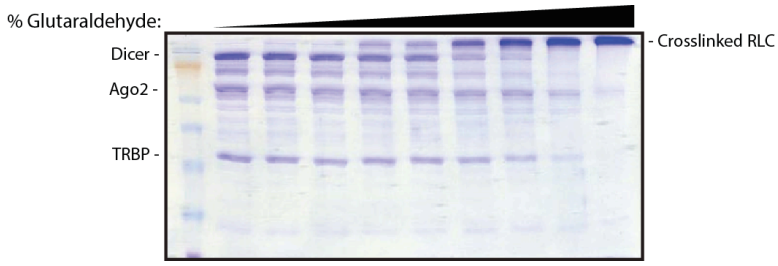

Figure 2.2: Glutaraldehyde crosslinking of the human RLC containing TRBP.
Coomassie Blue-stained SDS-PAGE gel of RLC complexes with increasing concentrations of glutaraldegyde. 0.02% glutaraldehyde completely shifts the Dicer band into the crosslinked RLC band.

Final 14 μL RLC crosslinking reactions for analysis by electron microscopy consisted of 1 μM RLC and 0.02% glutaraldehyde in Gel Filtration Buffer and were incubated for 10 minutes at room temperature. Reactions were then quenched by the addition of 1 μL 1 M Tris-HCl, pH 7.5 followed by a 5 minute incubation at room temperature.

2.3.6 GraFix crosslinking of protein complexes

The recently described GraFix sample preparation technique was used to enrich for crosslinked RLC (Kastner et al. 2008). 100 µl of RLC complex was applied to a 2 ml 5–25% (v/v) glycerol/0.02–0.1% (v/v) glutaraldehyde gradient in a 2.2 mL ultracentrifuge tube. After centrifugation at 45,000 g in a TLS-44 rotor (Beckman) for 6 h at 4°C, 200 µl fractions of the gradient were collected using a glass capillary with a peristaltic pump from the bottom of the tube. We then analyzed the fractions by SDS-PAGE and used the fraction corresponding to the ~350 kDa crosslinked RLC for single particle analysis (**Figure 2.9 A**).

2.3.7 Electron microscopy

Human Dicer, RLC or RLC subcomplex samples – either with or without crosslinking – were diluted to a concentration of 30-40 nM in EM Buffer (20 mM Tris-HCl, pH 7.5; 100 mM KCl; 3 mM $MgCl_2$; 1 mM TCEP) immediately before applying them to glow-discharged holey carbon grids with a thin layer of carbon over the holes. After adsorption for 1 min, we stained the sample consecutively with three 70 µL drops of 2% (w/v) uranyl formate solution, gently blotted off the residual stain and air-dried the sample in a hood. We examined the specimens under an FEI Tecnai-12 Bio-Twin electron microscope equipped with a LaB6 filament and operated at 120-kV acceleration voltage using a nominal magnification of 49,000x. For tilt-pair images, we recorded the same area of a specimen first at 55° and then at 0° tilt angles. We recorded images either on SO-163 film or on a Gatan 1k x 1k CCD camera using low-dose mode with an exposure dose of 20-30 $e^-/Å^2$. The defoci range used to collect the 0° tilt image was -0.8 µm to -1.1 µm, and that used for the 55° tilt image was -1.5 µm. We developed films using full-strength D19 for 12 min and digitized the micrographs without apparent drift and astigmatism using a Nikon Coolscan 8000 with a scanning step of 12.7 µm, resulting in a pixel size of 2.59 Å. The scanned images were first converted from their transmission values to optical densities using the PROC2D program of the EMAN image processing package (Ludtke et al. 1999). The CCD-recorded micrographs had a pixel size of 3.61 Å and were directly used in the image processing.

2.3.8 Image processing

For human Dicer, tilt pairs of particles were picked manually using WEB (Frank et al. 1996). Particle images from the micrographs were high-pass and low-pass filtered, normalized, boxed and reduced into boxes of 100 Å - 100 Å square pixels, with final pixel size of 5.18 Å using SPIDER (Frank et al. 1996). For the 3D reconstruction of human Dicer, the untilted images were imported into IMAGIC for reference-free 2D alignment and classification through several iterations of MSA analysis and MRA (van Heel et al. 1996). The tilted partners in each class were then used to generate 3D reconstructions by back projection using the RCT reconstruction method in SPIDER (Radermacher et al. 1987). The best six RCT class volumes were aligned, merged, and low-pass filtered to 50 Å resolution, then utilized as an initial model for projection matching refinement using the untilted particles in an iterative process with decreasing angular steps (Penczek et al. 1994). A final volume of human Dicer at a resolution of 34 Å was calculated from 3,433 untilted particles. The GraFix-prepared RLC was reconstructed using the same strategy. For this sample, a total of 2,600 tilted pairs were collected and used for RCT to generate an initial model. About 9,000 untilted single particle images of the GraFix-prepared RLC underwent heterogeneity analysis against the initial model by a maximum likelihood algorithm using Xmipp (Scheres et al. 2007) and 2,000 particles were used for the final reconstruction (see below). For the other specimens, the particle images underwent reference-

free 2D alignment and classification in IMAGIC to generate a final 100 to 200 class averages with 20~40 particles in each class. The class averages were compared to the 2D reprojections of either *ab initio* reconstructed or synthetic 3D models by MRA in SPIDER. The best-matched pairs were aligned and tiled together for clear illustration of their similarity and difference. The difference maps were calculated between class averages of the analyzed sample and the WT Dicer protein both matching the same projection view.

Supervised heterogeneity analysis of the RLC dataset was done following the same strategy as Gao and colleagues using two initial models (Gao et al. 2004). One initial model was the 3D reconstruction of human Dicer low-pass filtered to 50 Å resolution. The other was the 50 Å low-pass filtered model of the back-projected 3D reconstruction from all the crosslinked RLC class averages using their corresponding matched partners' Euler angles. Each particle in the RLC dataset was assigned two cross-correlation coefficients to the top matched reprojections of each of the two models. The difference between the two coefficients was used as a measure of the similarity of each particle to a certain model. The cross-correlation histogram of the entire dataset against the two models was then used to sort the particles into three groups, and to generate independent reconstructions from each group.

The uncrosslinked RLC, Dicer/Ago2 complex, and GraFix-prepared RLC were subjected to heterogeneity 3D analysis using the maximum likelihood reconstruction method (Scheres et al. 2007). Briefly, each dataset was first aligned to the Dicer 3D model (for the GraFix-prepared RLC, its own *ab initio* RCT initial model was used), and an arbitrary number of initial models from randomly defined subgroups were generated for further analysis (4 for the uncrosslinked RLC, 2 for the Dicer/Ago2, and 4 for the GraFix-prepared RLC datasets). For each dataset, the initial models only have statistical differences. We then used the Xmipp_ml_refine3d command in the Xmipp software package (Scheres et al. 2007), to reconstruct the four 3D models for the Dicer/Ago2/TRBP complex, the two models for the Dicer/Ago2 sample, and the four models for the GraFix-prepared RLC using an iterative maximum-likelihood reconstruction algorithm. The most abundant subpopulation of the GraFix-prepared RLC among the four had about 3,000 particles and was used for projection matching refinement to gain higher resolution than the maximum-likelihood model. The final reconstruction utilized ~2,000 particles after projection matching refinement. With overall similar shape to the above map, reconstructions from the other populations of the GraFix-prepared RLC particles have relatively low resolution and are noisy, probably due to the flexibility of the assemblies.

2.3.9 Docking of atomic models in the 3D EM density map

Docking of atomic models was done either using rigid-body docking programs within the Situs 2.0 software package (Wriggers and Birmanns 2001), or by visual inspection and manual docking using Chimera (Pettersen et al. 2004). More specifically, the atomic model of human DEAD-box RNA helicase DDX3X (PDB 2I4I; (Högbom et al. 2007)) was docked in the DExH/D domain using Colores within Situs 2.0l. The *Giardia intestinalis* Dicer crystal structure (PDB 2QVW; (Macrae et al. 2006)) was docked in the human Dicer platform density using the Fit-Model-In-Map function in Chimera guided by the biochemical evidence for the orientation, and then adjusted locally using Colacor of Situs 2.0. The guide strand-containing *Thermus thermophilus* Argonaute crystal structure (PDB 3DLH; (Wang et al. 2008c)), the *Giardia* Dicer crystal structure and the atomic model of human DEAD-box RNA helicase DDX3X were docked in the GraFix-prepared RLC 3D map by Fit-Model-in-Map of Chimera. All figures with three-dimensional models were generated using Chimera.

2.4 Results

2.4.1 Single particle reconstruction of human Dicer

We began by examining the structure of human Dicer by negative stain EM. The molecules appeared as well-dispersed, elongated particles with a length of ~20 nm (**Figure 2.3**). Using RCT methodology (Radermacher et al. 1987) followed by projection matching refinement (Penczek et al. 1994), we obtained a three-dimensional reconstruction of human Dicer at a resolution of ~34 Å (0.5 FSC criteria) from a final set of about 3,400 particle images (**Figure 2.3**). The final three-dimensional density map appears as an "L" shape containing a main vertical platform and a bilobal base branch that extends perpendicularly at the bottom. The curved 'cap' of the central platform protrudes in the same direction as the base branch. This arrangement defines an inner space or cavity delimited by an internal surface of the platform that is remarkably flat in the vertical direction and mostly concave in the horizontal direction (**Figure 2.3**).

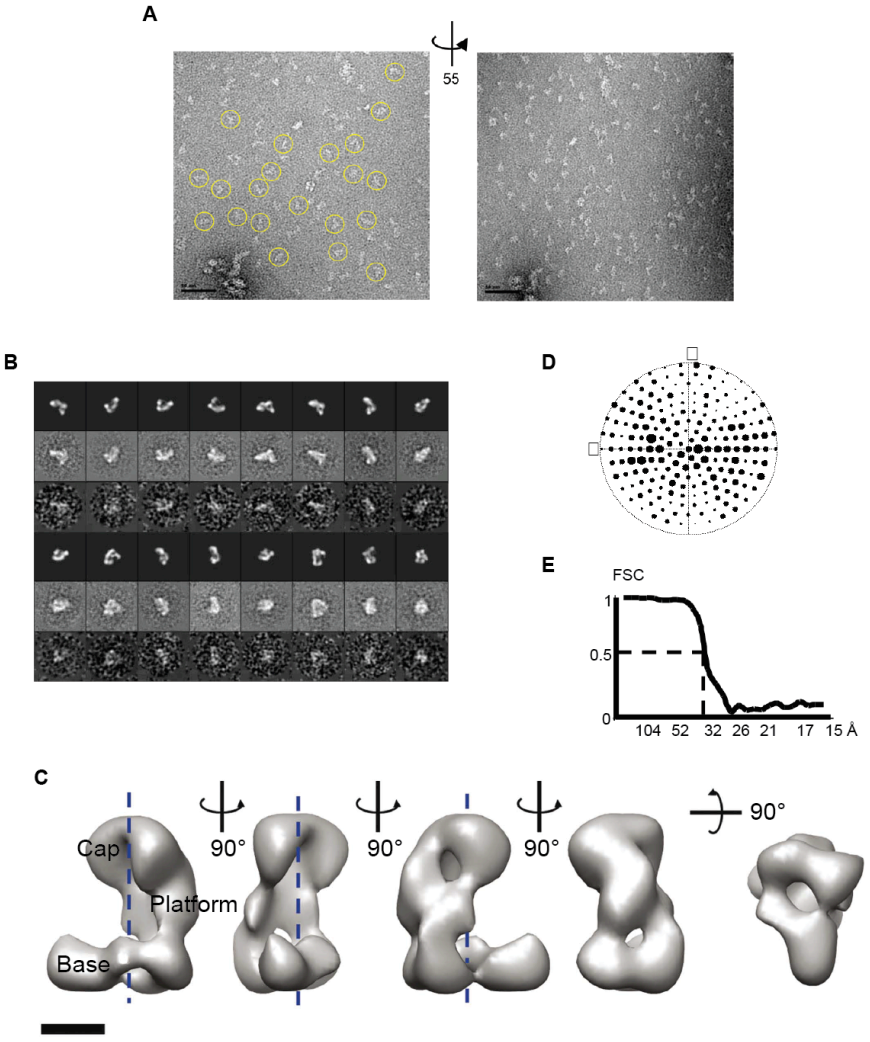

Figure 2.3: EM reconstruction of human Dicer.
(A) Representative tilted-pair electron micrographs of negatively-stained human Dicer particles at 0 and -55 degrees, respectively. Some individual particles are marked in yellow circles. Scale bar corresponds to 50 nm. **(B)**

Reprojections of the Dicer 3D reconstruction (first and fourth rows) and corresponding reference-free class averages (second and fifth rows). Box dimension corresponds to 41.8 nm. **(C)** 3D reconstruction of human Dicer shown in different orientations. The main structural elements are labeled. The channel in the middle of the volume is indicated by dashed lines. **(D)** Angular distribution map for the human Dicer reconstruction. The size of the spot relates to the number of particles falling in that particular view. **(E)** FSC curve for the human Dicer reconstruction showing the resolution of the map to be 34 Å using the 0.5 criterion.

With the overall protein shape defined, we next tried to dissect human Dicer's molecular architecture. Compared with the *Giardia intestinalis* Dicer, whose crystal structure is known (Macrae et al. 2006), human Dicer is substantially larger in size and includes an extra domain at its N-terminal end – the DExH/D helicase (**Figure 2.4 A**). Despite lacking the DExH/D domain, *Giardia* Dicer has full dicing activity. Similarly, a human Dicer mutant that lacks the DExH/D domain also retains dicing activity (Ma et al. 2008). These data suggest that the DExH/D domain is likely to be a discrete structural element within human Dicer. Although the catalytic core of Dicer, as approximately defined by the Giardia crystal structure, cannot fit within the base branch, it could fit within the reconstruction's platform, with an optimal cross-correlation coefficient of 0.55 in two possible orientations. The more biologically relevant orientation has the flat surface of the crystal structure, previously shown to be the site of RNA binding (Macrae et al. 2007), aligned along the flat vertical surface that faces the central cavity in our reconstruction (**Figure 2.4 B**). In contrast, the two-lobed volume of the base branch agrees well with the size of the DExH/D domain (68 kDa compared to the calculated 70 kDa from the volume at 3σ threshold). The crystal structure of its homolog, the human DEAD-box RNA helicase DDX3X (PDB 2I4I (Högbom et al. 2007)), can be docked within the base branch volume with a cross-correlation coefficient of 0.40 (**Figure 2.4 B**). These observations lead us to propose a working model in which the DExH/D domain of Dicer corresponds to the base branch region in our reconstruction.

To test this model, we examined a DExH/D domain-truncated version of human Dicer (ΔDExH/D) by EM. The truncated protein showed a strong tendency to dimerize, limiting the number of useful particles and precluding the attainment of a reliable three-dimensional reconstruction of this smaller protein (about 140 kDa). However, we were able to use reference-free two-dimensional alignment and classification to sort the ΔDExH/D particles and to obtain high-quality class averages. The class averages are more rod-like than those seen for the full-length enzyme and seem to lack the base branch, in agreement with our proposed assignment of the base branch to the DExH/D domain. To further test our hypothesis of the location of the DExH/D domain, we generated a 'synthetic' three-dimensional model of the ΔDExH/D protein by erasing the base branch from the three-dimensional reconstruction of the full-length Dicer (**Figure 2.4**). We then compared reprojections of this three-dimensional model to the experimental, reference-free two-dimensional class averages of the ΔDExH/D mutant. For each class average, we selected the best match out of the reprojections on the basis of the highest cross-correlation coefficient. This analysis retrieved more than 20 good matches to different views of the hypothesized three-dimensional model, covering most of the views. When compared with the same views of the full-length Dicer model and the experimental class averages, the density corresponding to the base branch is obviously missing in the ΔDExH/D mutant (**Figure 2.4 C**). This result, together with our docking of the DDX3X helicase, supports our initial proposal that the DExH/D domain of human Dicer is located at the base branch in the three-dimensional reconstruction.

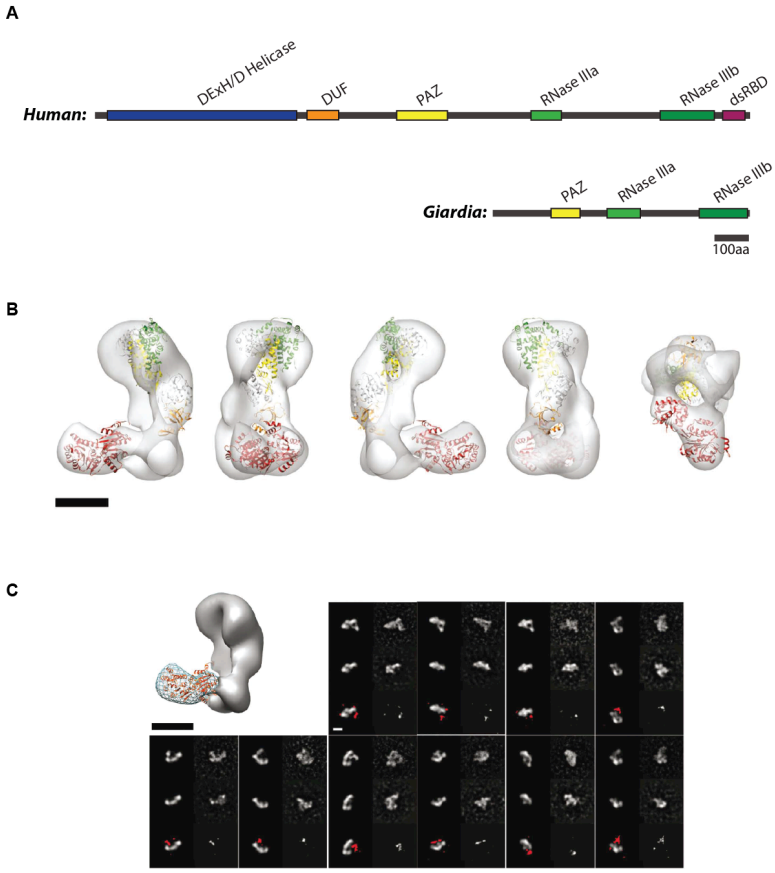

Figure 2.4: Architecture of human Dicer.
(A) Domain architecture schematics of human (upper) and *Giardia* (lower) Dicer. **(B)** Docking results of the atomic model of *Giardia* Dicer (Macrae et al. 2006) and the human DEAD-box RNA helicase DDX3X (Högbom et al. 2007) in red. **(C)** The synthetic three-dimensional model (grey solid surface) generated by removing the density corresponding to the base branch (blue wire) in the human Dicer reconstruction was used to match the reference-free 2D class averages (right) of the intact human Dicer. The second rows are reprojections from the synthetic Δbase branch model (left) and the corresponding reference-free class averages (right) of the ΔDExH/D mutant. The third rows are the difference maps of the 2D class averages between intact human Dicer and the mutant at 3σ threshold (right) and the superimposition of them (red) to the corresponding projections of the synthetic model (left). Scale bars in all figures indicate 5 nm.

2.4.2 The molecular architecture of the human RISC loading complex

Previous work has shown that the human RISC-loading complex can be reconstituted from individually purified Dicer, Ago2 and TRBP in a 1:1:1 stoichiometric ratio (Macrae et al. 2008). We used this complex to determine the molecular architecture of the human RLC by negative stain EM. Although this complex seems to be stable at the ~1 μM concentrations used during biochemical reconstitution, it tends to dissociate on EM grids either due to the lower concentration used (around 50 nM) or due to the stain. A 2D analysis revealed that in addition to the L-shaped particles corresponding to Dicer, a substantial number of globular particles with the shape and size of the dissociated Ago2 could be easily recognized in the micrographs of this sample (**Figure 2.5 A**, blue arrowheads).

To overcome the instability of the complex at low concentrations (or potentially in the negative stain solution) and achieve better definition of the RLC's architecture, we sought to obtain intact assemblies by crosslinking them with glutaraldehyde immediately after elution from a gel filtration column and before dilution for negative staining. Titration experiments showed

that 0.02% (v/v) glutaraldehyde crosslinks a substantial amount of the complex in its native 1:1:1 ratio but does not generate larger species (**Figure 2.2**). Under the electron microscope, the crosslinked specimen showed fewer dissociated components in the background than did the uncrosslinked sample (**Figure 2.5 B**). We performed iterative two-dimensional reference-free alignment and MSA classification (van Heel et al. 1996) of about 4,000 negatively-stained particles and sorted them into 200 classes. Some class averages emerged immediately with apparent additional densities when compared to the Dicer-only sample. We matched the 200 class averages with different projection views of Dicer's three-dimensional density using MRA and found that more than two-thirds of the class averages have significant additional densities (**Figure 2.5 C**). This indicated that although one-third of the particles still correspond to Dicer alone, the rest are likely to retain all or part of the RLC components.

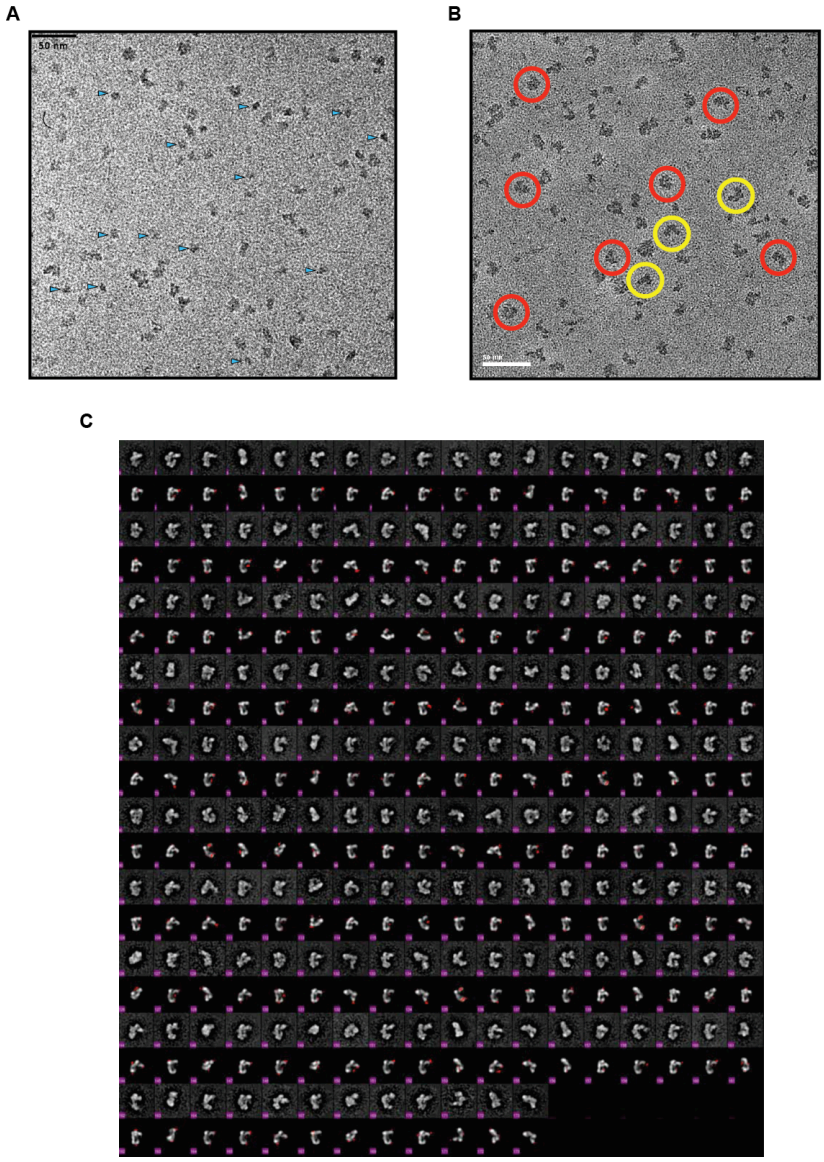

Figure 2.5: 2D EM analysis of glutaraldehyde crosslinked, negatively-stained RLC.
(A) Electron micrograph of negatively-stained uncrosslinked RLC. Small globular particles with dimensions and shape corresponding to Ago2 are indicated by blue arrowheads. **(B)** Electron micrograph of negatively-stained RLC complex that was crosslinked with 0.02% glutaraldehyde. Yellow circles indicate apo-Dicer particles. Red circles

indicate RLC particles or RLC subcomplexes. **(C)** Gallery comparing reference-free class averages of crosslinked RLC with their best-matched projection views of the 3D reconstruction of Dicer. The difference maps of the 2D class averages between the crosslinked RLC and the apo-Dicer at 3σ threshold (in red) are superimposed onto the projection views.

The class averages with additional densities were further sorted manually into three categories based on location and size of the extra densities. The first category includes particles with a small additional density attached to Dicer's base branch, the region of Dicer assigned to the DExH/D domain (**Figure 2.6 A**). The location of this density is variable around the tip of the DExH/D domain. The second category includes particles with a larger additional density located on the inner side of the L between the base branch and the cap at the top of the platform (**Figure 2.6 B**). This density also shows a range of positions and shapes but consistently remains close to the central cavity. The particles in the third category seem to have both the additional density connected to the DExH/D domain and the density facing the inner surface of the platform (**Figure 2.6 C**). We propose based on previous biochemical data that these additional densities correspond to TRBP and Ago2, respectively (Haase et al. 2005; Laraki et al. 2008; Kok et al. 2007; Sasaki and Shimizu 2007; Tahbaz et al. 2004).

To more firmly and objectively assign the extra densities, we examined a crosslinked complex composed of only Ago2 and Dicer, which was reconstituted in the same manner as the RLC. Similar image analysis revealed class averages clearly falling in the second category that show a large density occupying the inner space between Dicer's platform and the base branch (**Figure 2.6 D**). The larger size of Ago2 in these class averages compared to those in RLC suggests a more rigid configuration of Ago2 in the complex or a different orientation within the complex. These results further suggest that particles in Category 1 probably correspond to a Dicer/TRBP complex, that particles in Category 2 probably correspond to a Ago2/Dicer complex, and that Category 3 is formed by particles corresponding to the full RLC. Consistent with these assignments, our efforts to reconstitute Ago2 and TRBP with the ΔDExH/D Dicer mutant showed that Ago2 can still bind to Dicer but TRBP cannot (data not shown). This finding further supports a model in which Ago2 interacts with the C-terminal region of Dicer and TRBP interacts directly with the DExH/D domain.

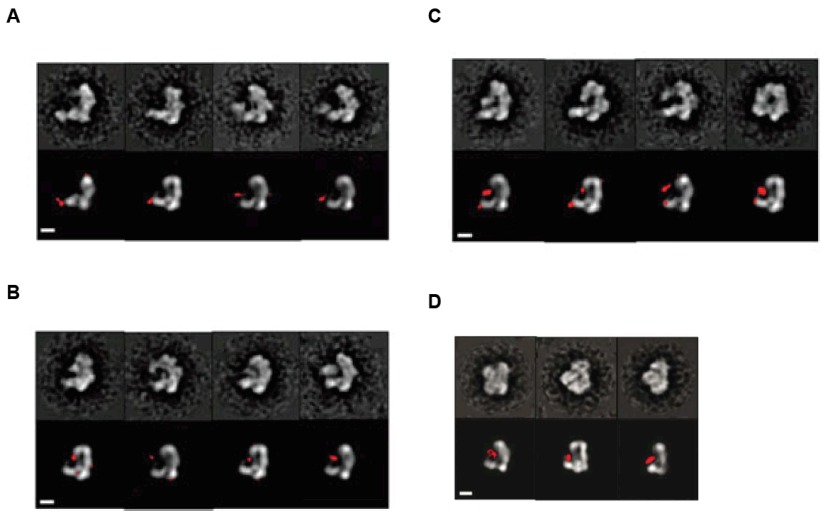

Figure 2.6: Molecular architecture of the human RLC containing TRBP.
(A) Comparison of Category 1 of the cross-linked RLC reference-free class averages (upper) with the corresponding projection views of the 3D map of human Dicer (lower). The difference maps of the 2D class averages between the

crosslinked RLC and the apo-Dicer at 3σ threshold (in red) are superimposed onto the projection views. **(B)** Comparison of Category 2 of the cross-linked RLC reference-free class averages (upper) with the corresponding projection views of the 3D map of human Dicer (lower). The difference maps of the 2D class averages between the crosslinked RLC and the apo-Dicer at 3σ threshold (in red) are superimposed onto the projection views. **(C)** Comparison of Category 3 of the cross-linked RLC reference-free class averages (upper) with the corresponding projection views of the 3D map of human Dicer (lower). The difference maps of the 2D class averages between the crosslinked RLC and the apo-Dicer at 3σ threshold (in red) are superimposed onto the projection views. **(D)** Comparison of the cross-linked Ago2/Dicer complex reference-free class averages (upper) with corresponding projection views of the three-dimensional map of human Dicer. The difference maps between the crosslinked Dicer/Ago2 complex and apo-Dicer class averages at 3σ threshold (in red) are superimposed onto the projection views. Shown here are the three clearest class averages out of a total of 100. Scale bars indicate 5 nm.

2.4.3 TRBP may stabilize the Ago2/Dicer interaction

Because we matched the 200 class averages of the crosslinked RLC sample to the corresponding projection views of the three-dimensional density of Dicer, we were able to reconstruct a three-dimensional map by back-projecting all the class averages with their assigned Euler angles. This 'all-in' reconstruction was similar to the Dicer reconstruction but had some additional density in the proposed location of Ago2 (**Figure 2.7 A**, gold structure as a low-pass filtered model at 40-Å resolution). We next performed supervised classification (Gao et al. 2004) on all the particles by matching them against two initial reference models, one being the low-pass filtered Dicer reconstruction (**Figure 2.7 A**, gray structure) and the other being the low-pass filtered all-in RLC reconstruction (**Figure 2.7 A**, gold structure). The histogram of cross-correlation values roughly agrees with our two-dimensional analysis that about one-third of particles are probably Dicer alone (**Figure 2.7 B**, model I). The one-third of particles that correlated better to the all-in model produced a final three-dimensional reconstruction with a prominent globular density narrowly attached to the cap and located in the inner space of the L between the platform and the base branch (**Figure 2.7 B**, model III). The remaining third of the particles in the center of the histogram produced a three-dimensional reconstruction with intermediate features, suggesting that they are a mixture of the two states (**Figure 2.7 B**, model II).

In none of the three models was there a clear density extending from the base branch corresponding to TRBP, probably owing to the flexibility of the TRBP molecule and/or its attachment to the DExH/D domain (based on sequence analysis, TRBP is composed of three small-sized RNA binding domains connected by flexible linkers. **See Figure 3.8 A**). Similarly, the density corresponding to Ago2 in the RLC reconstruction III could not account for the entire Ago2 atomic model (at the 3σ threshold, this density corresponds roughly to 50 kDa, about half of Ago2's molecular weight; **Figure 2.7 B**, **model III**). We propose on the basis of this result and on the initial reference-free 2D analysis in **Figure 2.6** that even when Ago2 is present in the complex it can take on different positions within the RLC. Therefore, it is likely that flexibility results in the reduced density seen in reconstruction III. The low occupancy of Ago2 and TRBP even in the crosslinked samples potentially reflects a relatively low affinity among the RLC components.

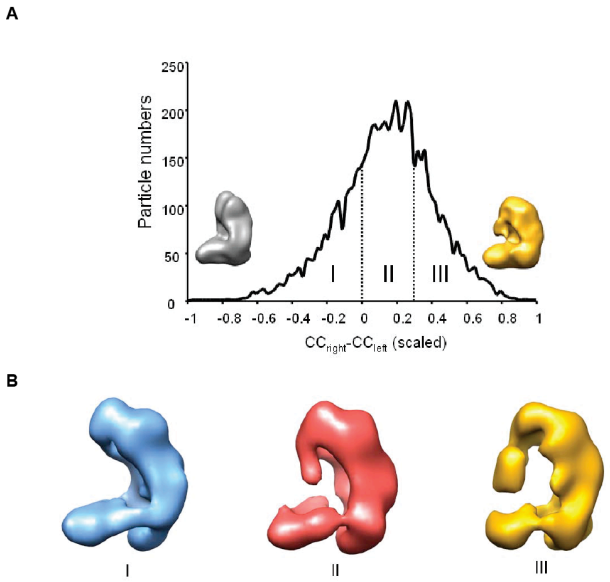

Figure 2.7: Supervised heterogeneity analysis of glutaraldehyde crosslinked RLC.
(A) Histogram of cross-correlation values obtained for the crosslinked RLC particles that compares two initial models: apo-Dicer (left, grey surface), and the "all-in" reconstruction from the crosslinked RLC sample (right, gold surface). The X-axis is the difference value between the cross-correlation coefficients of the raw particle with the right model (CC_{right}) and the left model (CC_{left}), respectively. The values were scaled to a range from -1 to 1. **(B)** The histogram in (A) was split into three groups of particles (I, II, and III) to be used in three separate reconstructions.

In addition to the data described above, we carried out an analysis of the heterogeneity present in the uncrosslinked RLC and Ago2/Dicer complexes using a maximum-likelihood strategy (Scheres et al. 2007). This study showed that in the absence of crosslinking, a prominent Ago2 density is observed in some particles of the sample that includes TRBP, in agreement with the three-dimensional reconstruction results (**Figure 2.8 A**). In the absence of TRBP, however, the Ago2/Dicer complex disassociates entirely under the conditions used for EM grid preparation, such that only the apo-Dicer structure is present in the particle data set (**Figure 2.8 B**). These results suggest that TRBP increases the affinity of Ago2 for Dicer and generally stabilizes the complex during dilution and grid preparation.

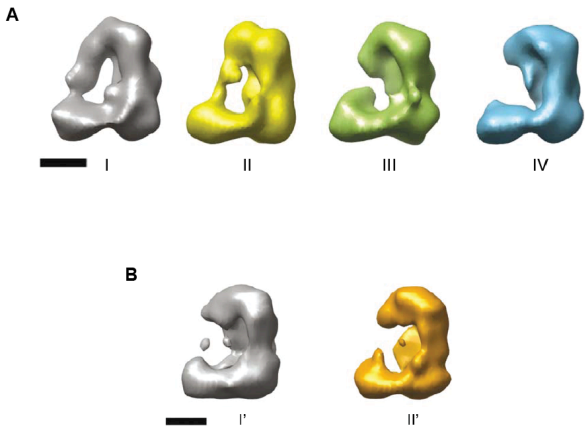

Figure 2.8: Maximum-likelihood analysis of uncrosslinked RLC and Dicer/Ago2 complex.
(A) From about 4,000 particles of uncrosslinked RLC, we generated four subclasses and calculated their 3D reconstructions. Among the four models, model I has an obvious additional density between the platform and the base branch, whereas model IV has no additional densities when compared with the apo-Dicer reconstruction. **(B)**

About 2,000 particle images of uncrosslinked Dicer/Ago2 complex underwent similar analysis to generate two subclasses and 3D reconstructions were created for each. Both models lack the additional density that exists in the RLC. Scale bars indicate 5 nm.

2.4.4 Ago2 may bind Dicer at its two termini

To further verify the RLC's architecture and to better characterize the interactions between its component parts, we used the GraFix technique to prepare RLC complexes for single particle analysis. This recently developed methodology has proven to be valuable for structural analysis of low-stability molecular assemblies by crosslinking the complexes gradually as they are separated on the basis of molecular weight and shape in a double gradient of glutaraldehyde and glycerol (Kastner et al. 2008; Herzog et al. 2009). The homogeneity of the RLC samples improved using this method, with the fraction containing the 360-kDa species appearing mono-disperse by EM observation (**Figure 2.9**).

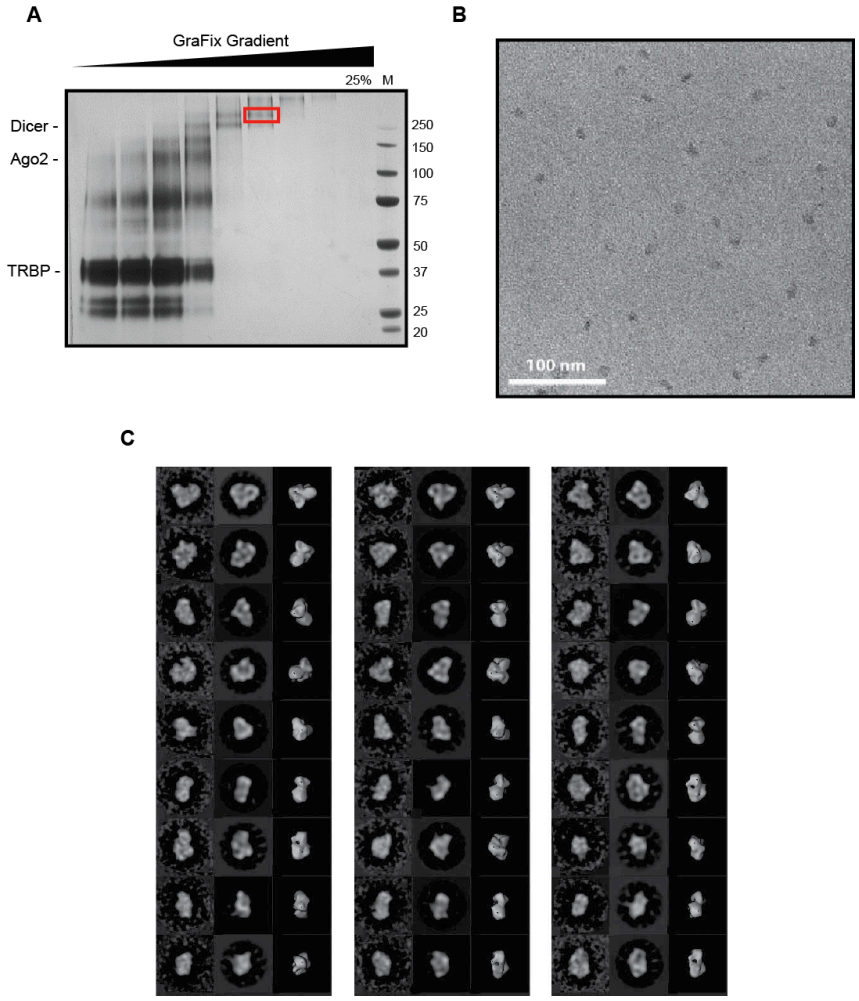

Figure 2.9: 2D EM analysis of GraFix-prepared human RLC.
(A) SDS-PAGE analysis of GraFix sedimentation fractions. The fraction shown in lane 6 was used for further single particle EM analysis. **(B)** Electron micrograph of negatively-stained particles of the GraFix-prepared RLC. **(C)** Comparison of the reference-free 2D class averages (Left column of each panel) matched with reprojections (Middle column of each panel) and the corresponding views (Right column of each panel) of the final 3D reconstruction of the GraFix-prepared RLC.

We performed *de novo* single particle reconstruction of the dataset collected from this sample using the RCT method, maximum-likelihood analysis and projection matching refinement and obtained a reconstruction at a resolution of ~33 Å (**Figure 2.10 A**). When compared with the reconstruction of Dicer alone, the reconstruction of this GraFix-prepared RLC has a prominent density in the middle of the molecule that connects the top of the platform and the tip of the base branch. The location of the density agrees well with the two-dimensional class averages in Categories 2 and 3 (**Figure 2.6 B, C**) and the maximum-likelihood reconstructed models I and II of the uncrosslinked RLC (**Figure 2.8 A**), suggesting the density to be Ago2. In agreement with this observation, the atomic model of a *Thermus thermophilus* Argonaute can be docked unambiguously in the difference map between RLC and Dicer reconstructions (**Figure 2.10 B**; (Wang et al. 2008c)). The docking suggests that Ago2 connects the top of the Dicer platform and the tip of Dicer base branch, giving rise to a triangular architecture for the complex. On one end, Ago2's PIWI domain interacts with Dicer's platform; on the other end, Ago2's N-terminal domain interacts with Dicer's base branch density. Although there are additional densities in this reconstruction beyond those corresponding to Ago2 and Dicer that could represent TRBP (especially at the base branch), it is difficult to interpret the location and shape of TRBP because of its possibly elongated shape and flexibility. Nonetheless, these data, together with those derived from the conventionally crosslinked sample, suggest that Ago2 may transiently interact with TRBP to form a closed complex with Dicer. We also note the substantial difference density observed in the region assigned to the DExH/D domain of Dicer in the GraFix sample when compared to Dicer alone. This implies a possible structural rearrangement or change in relative conformational dynamics of the helicase region upon association with Ago2 and/or TRBP.

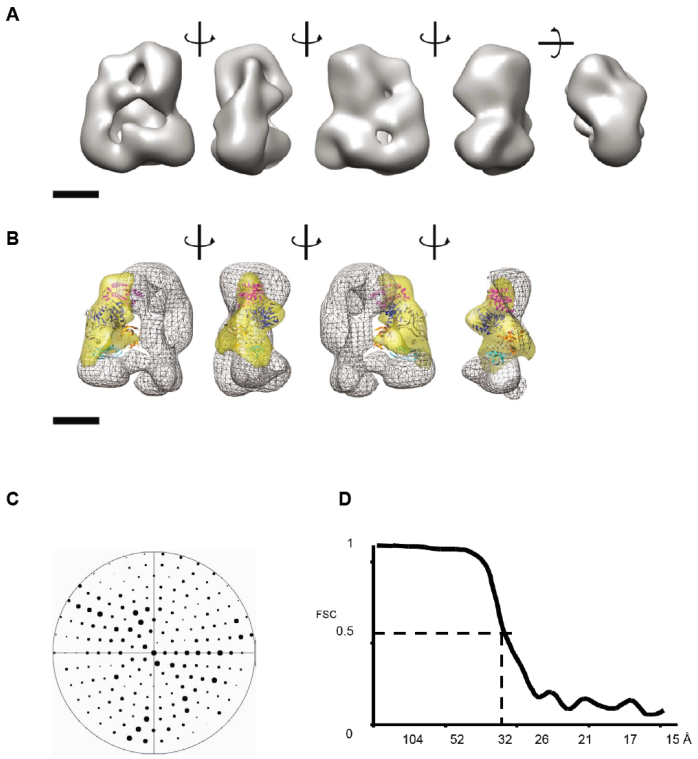

Figure 2.10: 3D EM reconstruction of GraFix-prepared human RLC.
(A) 3D reconstruction of the RLC shown in different orientations. **(B)** Docking of the atomic model of *Thermus thermophilus* Argonaute (Wang et al. 2008c) in the major part of the difference map (yellow transparent map)

calculated between the 3D reconstructions of the RLC shown in (A) and that of Dicer alone (shown as the wire map). Argonaute domain color code: cyan, N-terminal domain; orange, PAZ domain; pink, Mid domain; blue, PIWI domain. The docking is shown in four different views along the vertical axis. The front part of the RLC in the fourth view was removed to show the Argonaute docking more clearly. Scale bars represent 5 nm. **(C)** Angular distribution map for the GraFix-prepared RLC reconstruction. The size of the spot relates to the number of particles falling in that particular view. **(D)** FSC curve for the GraFix-prepared reconstruction showing the resolution of the map to be ~33 Å using the 0.5 criterion.

2.5 Discussion

The model presented here represents the first view of the molecular architecture of a human RLC and provides a structural framework for predicting and testing how siRNA duplexes may be passed from Dicer to Ago2 during the RISC loading process (**Figure 2.11**). Ago2 binds to Dicer's catalytic (platform) branch near the cap and shows substantial conformational flexibility in its attachment to Dicer. Previous biochemical studies suggested that this interface corresponds to the Ago2 PIWI domain in contact with a conserved insertion in Dicer's RNase IIIa domain (Sasaki and Shimizu 2007; Tahbaz et al. 2004). This information suggested the overall orientation of Dicer's catalytic core structure in the reconstruction map, with the C-terminal RNase III domains located near the top of the platform density. Beginning with this approximate orientation, we carried out quantitative docking of the Giardia Dicer crystal structure into the platform region of the three-dimensional volume of human Dicer using an automatic software algorithm (Fit-Model-In-Map in Chimera followed by COLACOR in Situs 2.0; (**Figure 2.4** (Pettersen et al. 2004; Wriggers and Birmanns 2001)). The most biologically relevant of the top two docking results positions the flat surface of the crystal structure, which has been shown to accommodate double-stranded RNA, facing the flat surface of the reconstruction (Macrae et al. 2007). Such a model implies that this is the initial binding site of substrate pre-siRNA. Interestingly, a later report that used EM to image internally epitope-tagged Dicer proteins indicated that the catalytic domain of Dicer is actually flipped 180° within the EM density of the catalytic branch compared to our model (Lau et al. 2012). This places the PAZ domain at the cap region of Dicer's catalytic branch and the RNase III domains near the helicase branch. Additionally, based on a recently solved structure of the RIG-I helicase, which is more closely related to Dicer's DExH/D domain than the DDX3X helicase, it appears that the helicase domain likely extends the full length of the base of human Dicer (Kowalinski et al. 2011).

Our RLC reconstruction shows considerable extra density at the proposed binding interface between Ago2 and Dicer's cap region. This connecting density may account for the extra 127-amino-acid insertion in human Dicer's RNAse IIIa domain that is not present in the *Giardia* Dicer enzyme (Sasaki and Shimizu 2007). This connection thus likely requires the extension of that region up from the RNase IIIa domain, but this extension is certainly possible given the insertion's size. An internal epitope tagging approach to verify the location of this insertion in the structure would likely be fruitful in refining our model. The extra density in the RLC reconstruction presented here can accommodate the atomic model of the *Thermus thermophilus* Argonaute containing a guide strand RNA adjacent to the long axis of the Dicer density (Wang et al. 2008c). Modeling a duplex RNA onto Dicer shows that the position of Ago2 leaves ample room for pre-siRNA binding (**Figure 2.11 A**). These Dicer substrates may potentially be fed to Dicer though a channel that is present in the helicase base branch of our apo-Dicer reconstruction (**Figure 2.3 C**; (Lau et al. 2009; 2012)).

Notably, our docking of the *Thermus thermophilus* Ago positions the Ago PAZ domain above the Dicer DExH/D domain in such a manner that the distance along Dicer's helicase

branch between the Ago2 PAZ domain and Dicer's catalytic branch can accommodate a 21-23 nt siRNA. Thus, it would in theory be possible for a product RNA to be bound along this axis in the step directly preceding complete RISC loading (**Figure 2.11 B**).

TRBP is known to be important for Dicer binding to its siRNA product and potentially for recruiting siRNAs to Ago2 (Chendrimada et al. 2005; Chakravarthy et al. 2010). Our 2D analysis indicates that TRBP is flexibly bound to the Dicer DExH/D domain, in agreement with existing biochemical data (**Figure 2.6**; (Haase et al. 2005; Kok et al. 2007)). The position of TRBP at the distal end of the Dicer DExH/D domain indicates that it may bind to Dicer products coordinately with Dicer along the helicase domain length. Furthermore, TRBP's high level of flexibility indicates that it may be able to bridge the Dicing and mRNA targeting steps by handing off nascent siRNAs to Ago2 directly. Such a process would be reminiscent of that shown biochemically in *Drosophila* with an RLC containing Ago2, Dicer-2, and the dsRBP R2D2 (Liu et al. 2003; 2006).

This model also has potential implications for guide strand selection during RISC loading. In general, the strand of a thermodynamically asymmetric siRNA duplex whose 5′ end lies at the less stable end of the duplex will become the guide strand when loaded into Ago2 (Schwarz et al. 2003; Khvorova et al. 2003). In *Drosophila*, Dicer-2/R2D2 RLC subcomplexes facilitate this process by binding thermodynamically asymmetric siRNAs in a directional manner such that Dicer-2 binds the less stable end and R2D2 binds the more stable end prior to loading onto Ago2 (Tomari et al. 2004b). Although it is currently unknown whether human Dicer and TRBP (or PACT) orient siRNAs in a similar manner (We address this question in **Chapter 3**), it is tempting to propose such an siRNA-orienting mechanism in light of our structure. An analogous thermodynamics sensor in which TRBP binds the more stable end of siRNA duplexes would position the 3′ end of the guide strand at the distal end of Dicer's helicase domain, near Ago2's PAZ domain. TRBP could then conformationally shift to directly pass this end of the guide strand to the Ago2 PAZ domain (**Figure 2.11 B**). Based on known structures of guide RNA-bound Ago proteins, this would prime Ago2 for loading of the siRNA in the proper orientation to allow for cleavage of the passenger strand and retention of the guide strand (Rand et al. 2005; Wang et al. 2009c; Matranga et al. 2005; Leuschner et al. 2006; Schirle and Macrae 2012; Elkayam et al. 2012; Nakanishi et al. 2012).

In the event than an siRNA is produced by dicing of a pre-siRNA that is positioned in the less favored direction with respect to RISC loading (for example, if the end of the duplex that is bound by the PAZ domain of Dicer is more stable than the end to be produced by RNase III cleavage), the siRNA could potentially be reoriented by release into the bulk solvent followed by rebinding in the correct orientation, as has been observed in *Drosophila* extracts (Preall et al. 2006). In this respect, TRBP would act as a checkpoint for strand-selectivity that ensures the proper orientation of the siRNA prior to its handoff to Ago2. Given that human Dicer forms two separate RLCs that are defined by the presence of either TRBP or PACT, it is unclear if both Dicer/TRBP and Dicer/PACT subcomplexes would have similar roles in strand-selective RISC loading. A study of strand selection in plant RNAi pathways has shown that certain Dicer/dsRBP heterodimers are able to orient duplex RNAs based on end stabilities while others cannot (Eamens et al. 2009). Further biochemical studies will therefore be necessary to determine the specific roles that TRBP- and PACT-containing RLCs may play in the process of strand-selective RISC loading in humans.

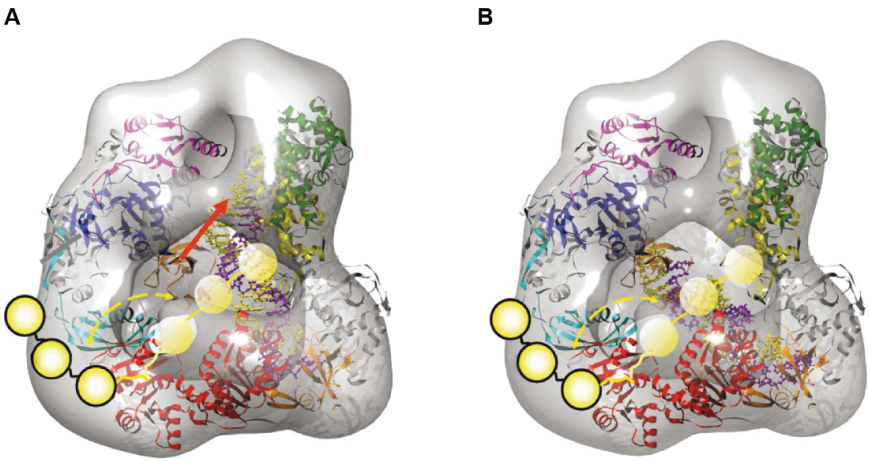

Figure 2.11: Proposed working model for the human RLC.
The 3D density map of the RLC is shown as a semitransparent surface. The atomic model of the DExH/D helicase domain (red ribbon), the *Giardia* Dicer atomic model (grey-yellow-green-orange ribbon, colored as in **Figure 2.4 B**) and the *Thermus thermophilus* Argonaute (grey-cyan-orange-pink-blue, colored as in **Figure 2.10 B**) are docked in the density map. TRBP is shown as a string of three yellow spheres with a flexible linker connecting it to the DE×H/D domain. Its potential motion range is marked by dashed yellow arrows. **(A)** Proposed RNA orientation during and directly following Dicer. An atomic model of an siRNA (guide strand in purple and passenger strand in yellow) is aligned vertically adjacent to Dicer's catalytic branch. **(B)** Proposed transfer of nascent siRNA onto Ago2's PAZ domain. The distance between the PAZ domain of Ago2 and the "elbow" of Dicer allows the accommodation for a 21 nt siRNA. TRBP would likely facilitate this handoff.

Chapter 3

Human RLCs Pre-Orient siRNAs for Strand-Selective RISC Loading*†

* A portion of the work presented in this chapter has previously been published as part of the following paper: Noland, C.L., Ma, E., Doudna, J.A. (2011). siRNA repositioning for guide strand selection by human Dicer complexes. *Molecular Cell* 43, 110-121.

† Dr. Enbo Ma designed and performed the activity assay with the bipartite Dicer enzyme. Cameron Noland performed all other published experiments and unpublished work described.

3.1 Introduction

RISC association with duplex siRNAs and miRNAs involves the selection of a guide strand to be used for downstream targeting events. Based on crystallographic studies of human Ago2, it is clear that Dicer-generated duplex RNAs bind to Ago2 such that the 3′ end of the guide strand binds to the PAZ domain, while the 5′ end is anchored by the Mid domain (Schirle and Macrae 2012; Elkayam et al. 2012). The specific loading of the guide strand in this way is essential to the overall efficiency of siRNAs an miRNAs, as accidental flipping of this orientation would lead to passenger strand retention and off-target gene silencing. Determining both the mechanism by which the RNAi machinery properly orients siRNAs and miRNAs for loading to Ago2, as well as the specific duplex parameters required for designing siRNAs that maximize the loading of the guide strand have thus been the subject of significant scientific inquiry (Czech et al. 2009; Khvorova et al. 2003; Krol et al. 2004; Okamura et al. 2009; Sano et al. 2008; Schwarz et al. 2003). Currently, little is known about the transit of nascent siRNAs to Ago2 in humans, although TRBP has been implicated in facilitating this process (Chendrimada et al. 2005).

The most accurate determinant of guide strand selection has to do with the thermodynamic properties of siRNA and miRNA duplexes. In thermodynamically asymmetric duplexes, the strand that has its 5′ end at the less stable end of the helix will be preferentially loaded onto Ago2 as the guide strand (Czech et al. 2009; Schwarz et al. 2003; Khvorova et al. 2003). In *Drosophila*, this selection process is assisted by the Dicer-2/R2D2 RLC subcomplex, which binds siRNAs in a directional manner such that the dsRBP R2D2 interacts with the more thermodynamically stable siRNA end and Dicer-2 interacts with the less stable end (Tomari et al. 2004b; Liu et al. 2006). It remains unclear which RLC component protein represents the actual sensor of siRNA thermodynamic asymmetry. dsRBPs have been proposed to sense siRNA thermodynamics, but these propositions were based on indirect evidence that did not simultaneously test for the ability of Dicer proteins to sense duplex thermodynamics (Tomari et al. 2004b; Gredell et al. 2010; Eamens et al. 2009).

How human Dicer and its associated dsRBPs might pre-orient duplex RNAs for proper guide strand selection upon Ago2 loading has been unclear, but our structural study of a human RLC containing Ago2, Dicer, and TRBP has indicated that a system analogous to that noted in *Drosophila* would make structural sense (see **Chapter 2**). Specifically, it would allow TRBP to bind the 3′ end of the guide strand and potentially transfer that end directly to the Ago2 PAZ domain (Wang et al. 2009b). However, given that Dicer binds to both TRBP and PACT to form two separate RLCs in humans, it is difficult to draw a direct analogy to the *Drosophila* thermodynamics sensor. Indeed, it has been shown that certain plant Dicer/dsRBP heterodimers are capable of sensing duplex thermodynamics, while others are not (Eamens et al. 2009).

In the present study, we use a structurally informed approach to determine how human Dicer/dsRBP heterodimers orient nascent siRNA duplexes for strand-selective RISC loading. Our EM data have shown human Dicer has an L-shaped architecture. The long arm of Dicer contains the catalytic center and binds to Ago2, and the short arm contains the N-terminal helicase domain and binds to TRBP and PACT (Kok et al. 2007; Lau et al. 2009; Tahbaz et al. 2004; Wang et al. 2009b). Using reconstituted human Dicer/dsRBP/siRNA complexes, we show here that nascent siRNAs are released and repositioned along Dicer's helicase domain following cleavage in order to enable sensing of siRNA thermodynamic asymmetry. We demonstrate that Dicer itself is capable of sensing this asymmetry and that this functionality is activated upon association with either TRBP or PACT. Taken together, these findings support a model for human siRNA processing in which two distinct RNA binding sites on Dicer enable the

repositioning of product RNAs, allowing the for sensing of siRNA thermodynamic asymmetry by Dicer/dsRBP RLC subcomplexes to guarantee the proper orientation of the duplex prior to strand-selective RISC loading.

3.2 Methods

3.2.1 Expression and purification of recombinant PACT

Human PACT cDNA was cloned into the vector pHMGWA (Busso et al. 2005) from the entry vector pENTR-TEV-D-TOPO (Invitrogen), with the result that the expressed protein would have an N-terminal 6xHis-MBP-TEV tag. The initial steps of this purification up through the TEV protease cleavage step were identical to that of TRBP (see **Section 2.3.1**).

Following overnight TEV protease cleavage, the sample was dialyzed against HiTrap S Buffer A (20 mM HEPES-KOH, pH 7.5; 150 mM KCl; 5% glycerol; and 1 mM TCEP) for two hours, replacing the buffer with fresh Buffer A after 1 hour. The dialyzed sample was then bound to a 5 mL HiTrap S column that had been pre-equilibrated with Buffer A. PACT was eluted from the column using a 100 mL linear gradient of KCl (150 mM to 1 M). Fractions containing PACT were then pooled and concentrated to ~1 mL using an Amicon Ultra concentrator (Millipore) before application to a HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE) that had been pre-equilibrated with Gel Filtration Buffer (50 mM HEPES-KOH, pH 8.0; 300 mM KCl; 10% glycerol; 1 mM TCEP). Fractions containing PACT were pooled and concentrated to >5 mg/mL and flash frozen in liquid nitrogen either as a large pool or in small aliquots depending on the downstream application. This purification strategy yielded a sample that was completely pure as visualized by SDS-PAGE (**Figure 2.1**) and contained an additional GSTF sequence at its N-terminus due to the original fusion construct.

3.2.2 Expression and purification of a recombinant TRBP truncation mutant

The first two dsRBDs of TRBP were cloned into the vector pHMGWA (Busso et al. 2005) from the entry vector pENTR-TEV-D-TOPO (Invitrogen), with the result that the expressed protein would have an N-terminal 6xHis-MBP-TEV tag. The expression and purification of this construct was identical to that of wild type TRBP (see **Section 2.3.1**).

3.2.3 Expression and purification of recombinant wild type SPNR

Human SPNR cDNA was cloned into the vector pHMGWA (Busso et al. 2005) from the entry vector pENTR-TEV-D-TOPO (Invitrogen), with the result that the expressed protein would have an N-terminal 6xHis-MBP-TEV tag. The expression and purification of this construct was identical to that of wild type TRBP (see **Section 2.3.1**).

3.2.4 Expression and purification of a recombinant SPNR-TRBP fusion protein

The two dsRBDs of SPNR including the intervening linker were fused to the C-terminal dsRBD of TRBP and its N-terminal linker by PCR (see **Figure 3.10 B**) and cloned into the vector pHMGWA (Busso et al. 2005) from the entry vector pENTR-TEV-D-TOPO (Invitrogen), with the result that the expressed protein would have an N-terminal 6xHis-MBP-TEV tag. The expression and purification of this construct was identical to that of wild type TRBP (see **Section 2.3.1**).

3.2.5 Expression and purification of a recombinant SPNR-Dicer protein

The two dsRBDs of SPNR including the intervening linker were ligated at their C-terminus to a second copy of the linker via an engineered BamHI restriction site. This cassette was then fused to the N-terminus of human Dicer via PCR using primers that would introduce a TEV protease cleavage site at Dicer's N-terminus (see **Figure 3.11 A**). This construct was then cloned into the pFastBac vector. Baculovirus was amplified and protein was expressed via a method identical to that of wild-type Dicer (see **Section 2.3.2**) with the exception that following the initial nickel purification, the sample was immediately loaded onto a HiLoad 16/60 Superdex 200 prep grade gel filtration column.

3.2.6 *In vitro* reconstitution of Dicer/dsRBP heterodimers

Heterodimer binding reactions were carried out in a final volume of 500 µL Gel Filtration Buffer containing 3 nmol Dicer and 11 nmol dsRBP (assuming a dimer). Reactions were incubated for 30 minutes at 4°C and then applied to a Superose 6 10/300 GL column (GE). Fractions were pooled conservatively to ensure that no free dsRBP was present. Samples were concentrated to ~1 mg/mL and flash frozen in small aliquots.

3.2.7 Preparation of dsRNAs

All ssRNAs studied were chemically synthesized by Integrated DNA Technologies or Dharmacon. RNAs were gel purified on a 15% denaturing PAGE gel and washed thoroughly to remove contaminating urea. dsRNAs were prepared by first incubating equimolar amounts of each ssRNA together in Annealing Buffer (100 mM HEPES, pH 7.5; 30 mM KCl; 3 mM $MgCl_2$) at 65°C for 10 minutes. Reactions were then slow-cooled by removing the heat block and placing it on the benchtop to equilibrate to room temperature.

3.2.8 UV photocrosslinking assays

10 µL photocrosslinking reactions composed of 100 nM protein, 10,000 cpm 5′ ^{32}P-labeled RNA (~50 nM), and 1x Dicing Buffer (20 mM HEPES, pH 7.5; 25 mM KCl; 1.5 mM $MgCl_2$; 1% glycerol; 0.01% Igepal 630-CA; 0.1 mg/mL BSA; 0.1 mg/mL tRNA; 2.5 mM TCEP) were pre-incubated at room temperature under foil for one hour. Reactions were then moved to a 96-well plate and exposed to 302 nm light for 15 minutes by placing a handheld UV lamp over the plate. In the case of the chimeric Dicer protein, reactions were then subjected to TEV-cleavage for 30 minutes under foil at room temperature. Reactions were run on a 4-20% SDS-PAGE gel. Gels were dried for 2 hours and observed by phosphorimaging following an overnight exposure. Data was quantified using ImageQuant TL software.

3.2.9 Competition assays

18 µL competition assays containing 2,000 cpm ^{32}P-labeled dsRNA (~15 nM), 10 nM protein, and 1x Dicing Buffer were preincubated for one hour on ice to allow for substrate binding but not cleavage. Reactions were then transferred to pre-warmed tubes containing 2 µL unlabeled siRNA of the appropriate concentration (from 0-fold to 100-fold over the concentration of substrate dsRNA) in 1x Dicing Buffer and incubated at 37°C for an additional 60 minutes. Reactions were split and 10 µL aliquots were run on either a 15% denaturing PAGE gel to analyze dicing efficiency or a 6% native PAGE gel to analyze product binding. Gels were dried overnight and exposed for ~8 hours prior to analysis by phosphorimaging. Data quantification was carried out using ImageQuant TL.

3.2.10 Biotin affinity purifications

18 μL reactions containing 15 nM dsRNA 40a/b, 10 nM protein, and 1x Dicing Buffer were pre-incubated for one hour on ice to allow for substrate binding but not cleavage. Reactions testing for siRNA pulldown were then transferred to tubes containing 2 μL siRNA 12a/b of the appropriate concentration (from 0-fold to 100-fold over the concentration of substrate dsRNA) and containing 1,000 cpm ^{32}P-labeled siRNA and incubated at 4°C for an additional 60 minutes. Reactions testing for protein pulldown were left at 4°C without adding siRNA. Samples were applied to 10 μL Avidin agarose resin and incubated for one hour at 4°C. Each sample was washed 5 times and before boiling in 20 μL 2x SDS Sample Buffer (125 mM Tris-HCl, pH 6.8; 20% glycerol; 4% SDS; 0.04% (w/v) bromophenol blue; 200 mM DTT) and run on either a 15% denaturing PAGE gel in the case of the siRNA pulldown assay or a 4-20% gradient SDS-PAGE gel in the case of the protein pulldown assay. Denaturing PAGE gels were dried overnight and exposed for ~8 hours prior to analysis by phosphorimaging. SDS-PAGE gels were silver stained.

3.2.11 Nickel affinity purifications

Dicer was incubated at a final concentration of 1 μM in a 50 μL reaction containing Wash Buffer (20 mM HEPES, pH 7.5; 150 mM KCl; 30 mM imidazole; 10% glycerol; 1 mM TCEP) either with or without 1 μM 6xHis-TRBP or 6xHis-SPNR for one hour on ice. Reactions were then applied to 20 μL nickel-NTA agarose and incubated for an additional hour on ice. Samples were washed 3 times with 200 μL Wash Buffer and then boiled in 20 μL 2x SDS Sample Buffer and run on a 4-20% gradient gel. Proteins were visualized by Coomassie Blue staining.

3.2.12 Western blotting

Mouse embryonic stem cells (conditional and Dicer null) were cultured as described (Murchison et al., 2005) and lysed in Lysis Buffer (50 mM Tris-HCl, pH 8.0; 170 mM NaCl; 5 mM EDTA; 20 mM NaF; 0.5% Nonpidet-40; 1 mM DTT) supplemented with protease inhibitors (Roche). Lysates were cleared by centrifugation at 14,000 rpm for 15 minutes at 4°C. Supernatants were applied to centrifugal filter units and spun for an additional 5 minutes at 14,000 rpm at 4°C. Samples were then run on a 10% SDS-PAGE gel. 10 ng of recombinant Dicer and bipartite Dicer were run as controls. Following transfer to a PVDF membrane, proteins were probed with antibody 1414 (1:750), which targets Dicer's N-terminal region (Kanellopoulou et al., 2005).

3.2.13 Immunoprecipitations for mass spectrometry

1.4 mL lysate from Dicer null mouse embryonic stem cells was incubated with antibody 1414 (1:75) for one hour at 4°C with rocking. This lysate was then transferred to 30 μL Protein A beads and incubated for 2 hours with rocking at 4°C. The supernatant was removed and beads were washed 5 times with 1 mL Immunoprecipitation Buffer (50 mM Tris-HCl, pH 7.4; 150 mM NaCl; 10% glycerol; 0.5% Nonpidet-40) supplemented with protease inhibitors (Roche). Beads were then boiled in SDS Sample Buffer, passed through a centrifugal filter unit and run on a 10% SDS-PAGE gel.

3.2.14 Native gel shifts

20 μL binding reactions were carried out on ice for 45 minutes and consisted of 0.5 pM siRNA 12a/b (see **Table 3.1**) and either 0, 0.5, or 1 pM protein in Binding Buffer (20 mM HEPES, pH 7.5; 300 mM KCl; 5% glycerol; 1 mM TCEP). Following this incubation, 6 μL 25% sucrose was

added to each tube and binding reactions were resolved by 6% native PAGE at 4°C and RNAs were stained for 10 minutes with SYBR Gold (Invitrogen) before imaging.

3.3 Results

3.3.1 Either TRBP or PACT is required for product binding by Dicer

We were interested to know the extent to which the roles of TRBP and PACT overlap in RNAi. TRBP had previously been shown to be necessary for product binding by Dicer (Chendrimada et al. 2005; Chakravarthy et al. 2010), and although PACT had been shown to bind to Dicer and had been generally demonstrated to be involved in RNAi, a more specific role had yet to be defined (Lee et al. 2006; Kok et al. 2007). In order to determine whether or not PACT is sufficient to impart product-binding activity on Dicer, we incubated constant concentrations of a 21nt siRNA (siRNA 12a/b, **see Table 3.1**) with increasing concentrations of either Dicer alone or the Dicer/TRBP or Dicer/PACT heterodimers. Following this initial incubation, binding reactions were run on a 6% native gel at 4°C to resolve free RNA from bound RNA. As expected, Dicer alone did not bind to the product RNA, whereas the Dicer/TRBP heterodimer bound efficiently to the same siRNA (**Figure 3.1 A**). Interestingly, the same was true for a Dicer/PACT heterodimer, although the specific affinities of each complex for siRNAs may differ, as evidenced by the presence of unbound RNA even at the highest concentration of Dicer/PACT (**Figure 3.1 B**).

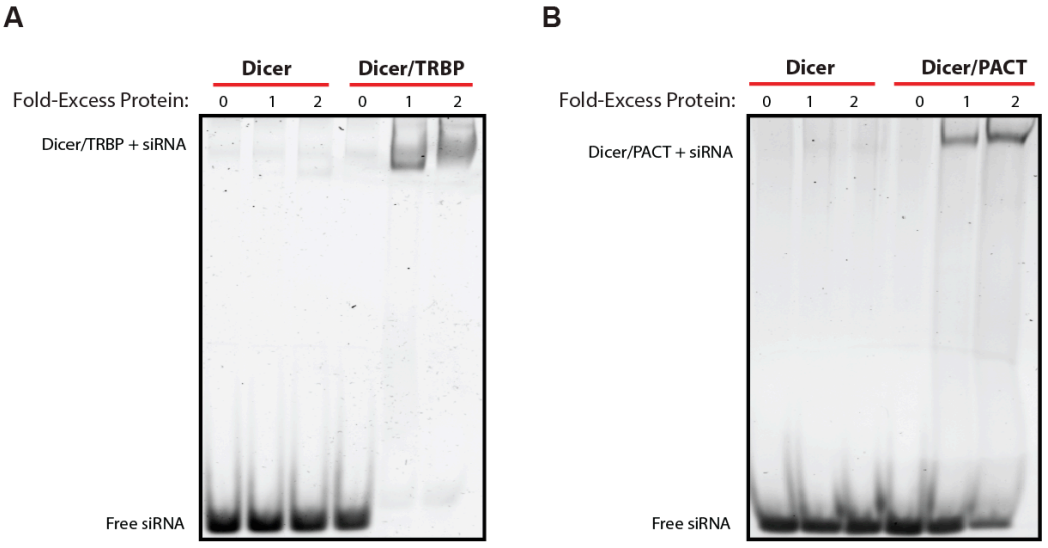

Figure 3.1: TRBP and PACT are sufficient to impart product binding activity on Dicer.
(A) Increasing concentrations of Dicer or Dicer/TRBP heterodimer were incubated with a constant amount of siRNA 12a/b on ice for one hour prior to being applied to a 6% native gel. **(B)** Increasing concentrations of Dicer or Dicer/PACT heterodimer were incubated with a constant amount of siRNA 12a/b on ice for one hour prior to being applied to a 6% native gel.

3.3.2 The Dicer/TRBP heterodimer binds differently to substrate and product RNAs

Negative stain electron microscopy indicates that Dicer has an L-shaped structure, and that TRBP binds to the distal end of Dicer's N-terminal helicase base branch, consistent with biochemical data (Lau et al. 2009; Lee et al. 2006; Wang et al. 2009b). Additionally, TRBP is required for high affinity binding of product, but not substrate dsRNAs (see **Section 3.3.1** (Chakravarthy et al. 2010; Chendrimada et al. 2005)). These studies suggest the possibility that

substrate and product dsRNAs might bind to different regions of Dicer. For pre-siRNA cleavage to occur, a substrate must be bound along the catalytic arm of Dicer. However, the similarity between the length of the helicase domain and the length of an siRNA duplex in our 3D EM reconstructions hinted that following dicing, products might re-localize to Dicer's helicase domain.

To test this hypothesis, we reconstituted the human Dicer enzyme by co-expression of an N-terminal "Helicase" fragment containing the N-terminus through the PAZ domain and a separate C-terminal "Catalytic" fragment containing the RNaseIII domains and C-terminal dsRBD (**Figure 3.2 A**). This bipartite Dicer lacks a small region of polypeptide corresponding to a flexible loop and co-elutes with full length Dicer on a sizing column (data not shown), thus its structure is likely to be identical to that of the intact protein. Dicing assays confirmed that this two-piece Dicer retains catalytic activity that is highly similar to that of the wild type protein (**Figure 3.2 B**). Following purification, this bipartite Dicer was combined with an equimolar amount of TRBP to reconstitute a Dicer/TRBP complex.

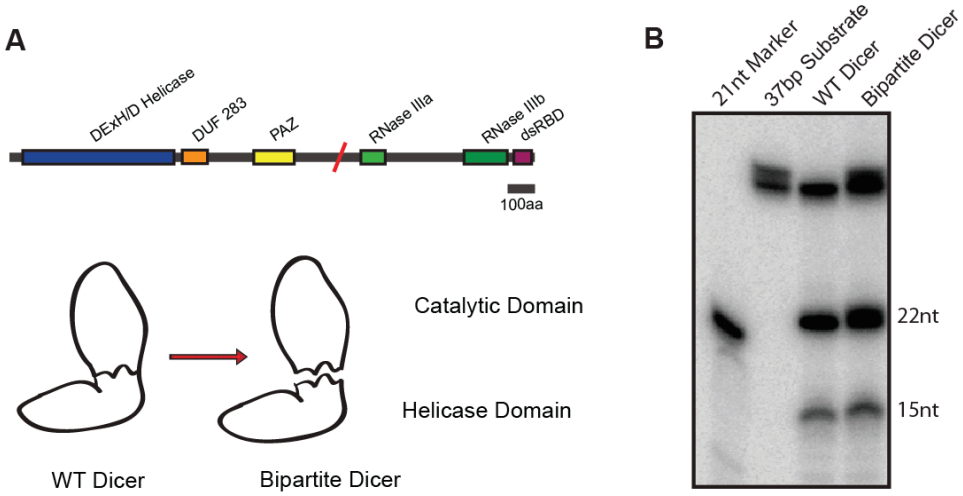

Figure 3.2: A bipartite Dicer is reconstituted by co-expressing two Dicer fragments.
(A) *Upper:* Diagram of human Dicer's domain structure. Red slash indicates the approximate location of the interface between the two parts of the bipartite Dicer. *Lower:* Cartoon of recombinant bipartite Dicer. **(B)** Recombinantly expressed wild type (WT) or bipartite Dicer was incubated for one hour at 37°C with a ^{32}P-labeled, 37 nt dsRNA substrate containing 2 nt 3′ overhangs. Samples were then run on a 15% denaturing polyacrylamide gel. Separate product bands represent the products formed when Dicer cleaves from either end of the dsRNA substrate.

To determine how the Dicer/TRBP complex positions Dicer substrates relative to products, we used a set of 5-iodo-uracil-modified substrate and product dsRNAs that could be photocrosslinked to nearby tryptophan, tyrosine, and histidine residues of the protein complex by exposing binding reactions to 302 nm light (**Figure 3.3 A, Table 3.1** (Tomari et al. 2004b; 2004a; Dietz and Koch 1987; Willis et al. 1993; Norris et al. 1996)). Substrate mimics consisted of a 37 bp chimeric dsRNA with one blunt end and a 2 nt 3′ overhang at the other end to promote unidirectional binding of the overhang by Dicer's PAZ domain (Rose et al. 2005). Deoxyribonucleotides present at the cleavage sites in each strand prevented the dicing reaction from occurring (data not shown). A single 5-iodo-uracil modification was incorporated at different locations along each dsRNA. Given that the 3′ overhangs of Dicer substrates are proposed to bind the PAZ domain (present in the helicase portion of the bipartite Dicer), care was taken to place substrate modifications at sites distant from this overhang. Product mimics

consisted of 21 nt thermodynamically symmetric siRNA duplexes with a single 5-iodo-uracil modification. Modified strands were radiolabeled at their 5′ ends so that photocrosslinked proteins could be identified by mobility using SDS-PAGE.

These experiments showed that regardless of sequence, Dicer substrates photocrosslink to both the C-terminal catalytic Dicer fragment and the N-terminal helicase fragment, indicating that pre-siRNAs may be bound in both catalytic- and noncatalytic conformations. Interestingly, Dicer products preferentially crosslink to the helicase fragment (**Figure 3.3 B, C**). These differential photocrosslinking propensities for substrates versus products are consistent with siRNA repositioning within the Dicer/TRBP complex following substrate cleavage.

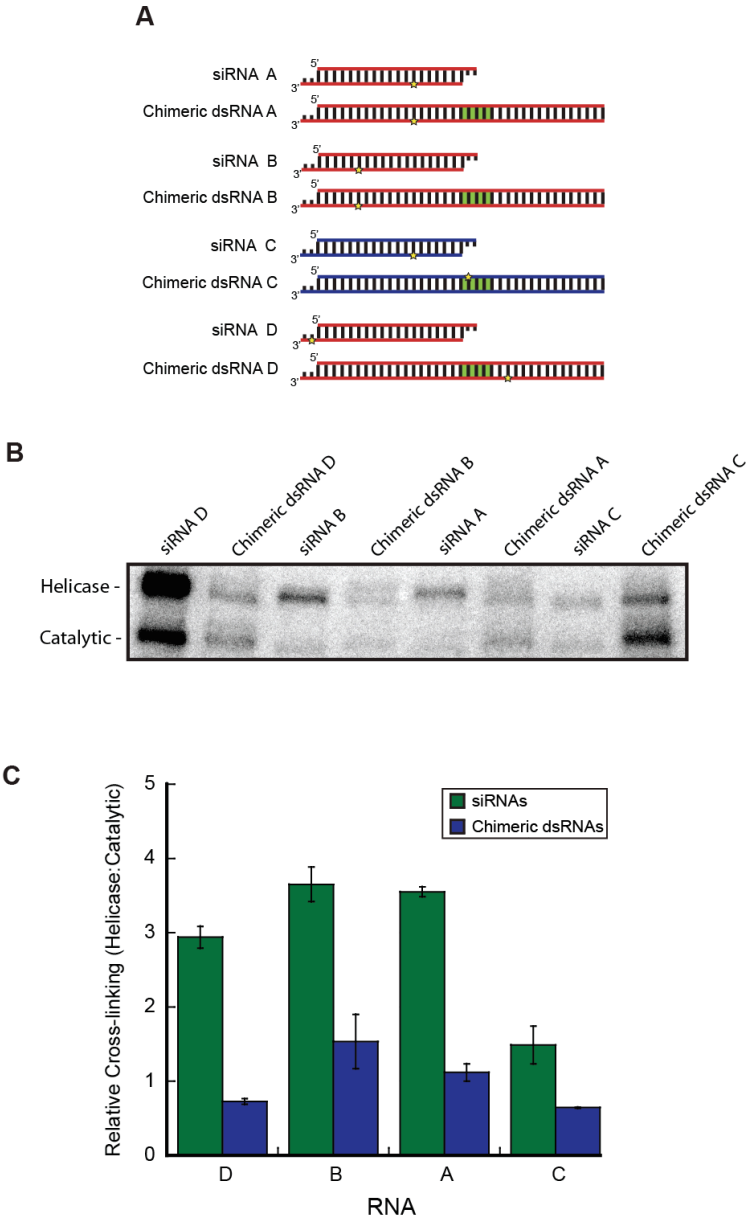

Figure 3.3: Human Dicer binds pre-siRNA substrates and siRNA products at distinct sites.
(A) Diagram of synthetic, modified RNAs used for photocrosslinking to the bipartite Dicer enzyme. Stars indicate the location of 5-iodo-uracil modifications. Green areas indicate the presence of deoxyribonucleotides at the cleavage site. Red and blue strands indicate RNAs designed based on the same starting sequence. Overhangs are 2 nt

3′ overhangs. siRNAs are thermodynamically symmetric. Strands containing 5-iodo-uracil modifications were 5′ radiolabeled and all unmodified strands had a 5′ phosphate. See also **Table 3.1**. **(B)** Photocrosslinking assay using bipartite Dicer enzyme. Substrate mimics or siRNAs were incubated with the bipartite Dicer/TRBP complex at room temperature followed by irradiation with 302 nm light. Photocrosslinked proteins were detected based on size by SDS-PAGE. **(C)** Quantification of three experimental replicates (mean +/- SD) showing the ratio of photocrosslinking to the bipartite Dicer's helicase domain versus the catalytic domain.

Table 3.1: dsRNA sequences. Green lettering indicates the presence of a deoxyribonucleotide. Red lettering denotes a 5-iodo-uracil base. Yellow highlights mark mismatched bases.

Chimeric dsRNA A	5′– GUCACAUUGCCCAAGUCUCUTTTGAAAGUUCACGAUU –3′ 3′– UUCAGUGUAACGGGUUCAGAGAAAACUUUCAAGUGCUAA –5′
Chimeric dsRNA B	5′– GUCACAUUGCCCAAGUCUCUTTTGAAAGUUCACGAUU –3′ 3′– UUCAGUGUAACGGGUUCAGAGAAAACUUUCAAGUGCUAA –5′
Chimeric dsRNA C	5′– GAUUUCUUCCCUUCUGAUGUAAACGCCCUAAGCGGUU –3′ 3′– AACUAAAGAAGGGAAGACUACTTUUGCGGGAUUCGCCAA –5′
Chimeric dsRNA D	5′– GUCACAUUGCCCAAGUCUCUTTTGAAAGUUCACGAUU –3′ 3′– UUCAGUGUAACGGGUUCAGAGAAACUUUCAAGUGCUAA –5′
siRNA A	5′– GUCACAUUGCCCAAGUCUCTT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′
siRNA B	5′– GUCACAUUGCCCAAGUCUCTT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′
siRNA C	5′– CAUCAGAAGGGAAGAAAUCAA –3′ 3′– AUGUAGUCUUCCCUUCUUUAG –5′
siRNA D/(1)	5′– GUCACAUUGCCCAAGUCUCTT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′
dsRNA 40a/b	5′– UGAGGUAGUAGGUUGUAUAGUUUGAAAGUUCACGAUUAUU –3′ 3′– UUACUCCAUCAUCCAACAUAUCAAACUUUCAAGUGCUAAUAA-biotin –5′
siRNA 12a/b	5′– GAUUUCUUCCCUUCUGAUGUU –3′ 3′– AACUAAAGAAGGGAAGACUAC –5′
siRNA 2	5′– UUCACAUUGCCCAAGUCUCTT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′
siRNA 3	5′– UUCACAUUGCCCAAGUCUCUT –3′ 3′– TTCAGUGUAACGGGUUCAGAG –5′
siRNA 4	5′– UUCACAUUGCCCAAGUCUCUT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′
siRNA 5	5′– GUCACAUUGCCCAAGUCUATT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′
siRNA 6	5′– GUCACAUUGCCCAAGUCUAUT –3′ 3′– TTCAGUGUAACGGGUUCAGAG –5′
siRNA 7	5′– GUCACAUUGCCCAAGUCUAUT –3′ 3′– TUCAGUGUAACGGGUUCAGAG –5′

3.3.3 The Dicer/TRBP heterodimer releases nascent siRNAs for re-binding

The product repositioning that we detected could occur by two potential mechanisms. Either siRNAs are released from the Dicer/TRBP complex and rebound in a new orientation, or they are reorganized without ever dissociating from the complex (**Figure 3.4 A**). Studies using

Drosophila embryo lysates or recombinant human Dicer and human cell extracts have led to conflicting views of siRNA fate immediately following dicing. In *Drosophila*, competition experiments indicated that siRNA products are released from Dicer prior to RISC loading (Preall et al. 2006). Rose and colleagues, on the other hand, presented a model in which siRNAs remain bound to human Dicer and enter directly into RISC based on the observation that in some cases Dicer's processing polarity is predictive of silencing efficacy (Rose et al. 2005).

To clarify the immediate fate of nascent siRNAs within the Dicer/TRBP complex before Ago2 loading, we developed a native PAGE product-binding assay using a reconstituted human dicing system. A 40 bp pre-siRNA substrate was used (dsRNA 40a/b; see **Table 3.1**) that had one blunt end and a 2 nt 3′ overhang at the other end. The strand containing the 2 nt 3′ overhang also had a 5′ biotin modification to further promote unidirectional dicing. As anticipated, this dsRNA yields a single cleavage product upon incubation with Dicer (data not shown). The amount of product bound could be readily tracked by mobility shift on a 6% native gel (**Figure 3.4 B**). Briefly, substrates were pre-incubated with the Dicer/TRBP heterodimer on ice to allow for binding but not cleavage (**Figure 3.4 B, Lane 3**). Reactions were then transferred to tubes containing increasing concentrations of unlabeled siRNA 12a/b competitor (**Table 3.1**) at 37°C to initiate the dicing reaction. If products remain bound to the Dicer/TRBP complex as they are moved to the helicase branch, increasing amounts of cold siRNA would compete with product binding to a similar extent as when the complex was instead pre-incubated with a labeled siRNA product. However, if products were first released and then re-bound by the helicase domain, one would expect a higher level of competition than in the case of a pre-bound product.

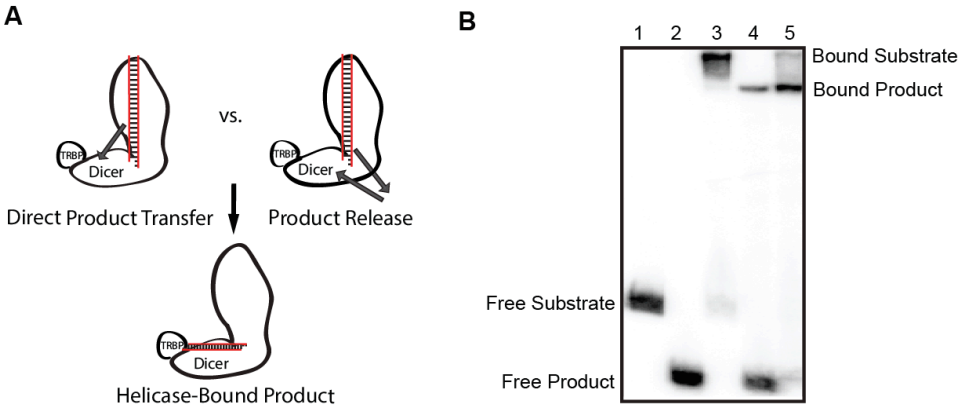

Figure 3.4: A Dicer product binding native PAGE assay.
(A) Schematic of potential post-processing mechanisms of siRNA repositioning. Products are either reorganized within the complex (*upper left*) or released and re-bound (*upper right*) prior to repositioning along Dicer's helicase domain (*lower*). **(B)** Dicer product binding native PAGE assay. ^{32}P-labeled pre-siRNA 40a/b (Lanes 1, 3 and 5) or siRNA 12a/b (Lanes 2 and 4) were pre-incubated either alone (Lanes 1 and 2) or with the Dicer/TRBP heterodimer on ice for one hour and either left on ice for another hour (Lanes 1-4) or transferred to 37°C to initiate dicing for one hour (Lane 5). 3 µL 25% sucrose was added to each sample prior to running on a 6% native polyacrylamide gel.

Product binding and dicing assays showed that excess unlabeled siRNA effectively competed with the 40a/b dicing product for binding to the Dicer/TRBP heterodimer without significantly affecting dicing efficiency (**Figure 3.5 A**). We reproducibly observed a ~67% decrease in product binding when 100-fold excess competitor was present. By contrast, binding decreased by only ~11% when Dicer was pre-incubated with a 21 nt siRNA instead of the 40a/b substrate, indicating that we are seeing *bona fide* product release (**Figure 3.5 B**). To control for non-specific effects, we repeated these competition experiments with an unrelated, unlabeled 116

nt structured human Alu RNA and observed only a ~16% decrease in product binding (**Figure 3.5 B**). None of these decreases were accompanied by a commensurate decrease in overall dicing activity (**Figure 3.5 C**). To address the possibility that these data simply reflect an ability of Dicer to bind both substrates and products simultaneously, we performed biotin affinity purification assays using the biotinylated 40a/b pre-siRNA substrate. Dicer/TRBP was pre-bound to this substrate on ice, and increasing concentrations of radiolabeled siRNA 12a/b were then added to the reactions. Despite the fact that the Dicer/TRBP complex was purified by this method, radiolabeled product RNA was not precipitated at any of the concentrations tested (**Figure 3.5 D, E**). Our data are consistent with a model in which the Dicer/TRBP heterodimer releases nascent siRNAs into the bulk solvent and then rebinds them along Dicer's helicase domain rather than reorienting them within the confines of the complex.

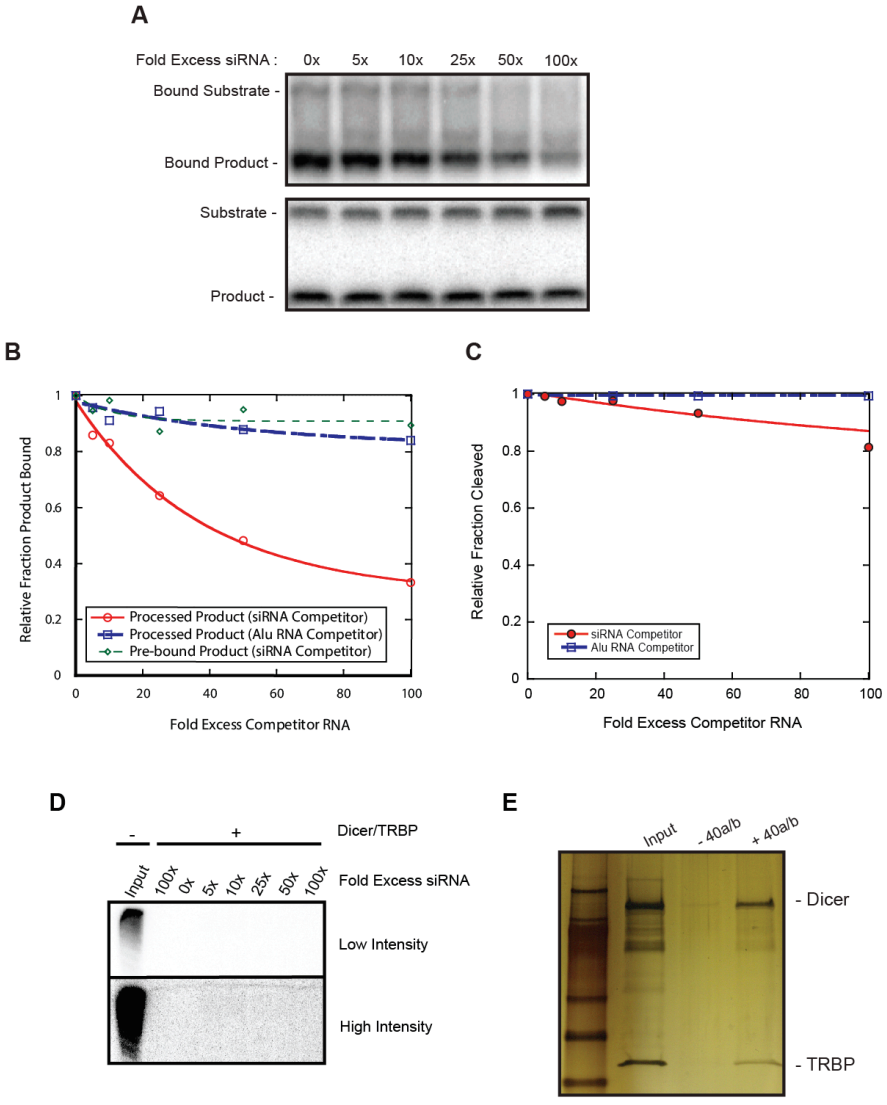

Figure 3.5: Human Dicer/TRBP releases nascent siRNA products prior to repositioning.
(A) Product binding competition assay. Dicer/TRBP complex was pre-incubated with radiolabeled 40a/b substrate at 4°C. To initiate cleavage, reactions were transferred to tubes at 37°C containing increasing concentrations of unlabeled competitor siRNA 12a/b. *Upper*: 6% native PAGE analysis shows a decrease in bound product as the concentration of competitor siRNA is increased. *Lower*: 15% denaturing PAGE analysis demonstrates that

competitor siRNA 12a/b has a negligible effect on dicing activity. **(B)** Quantification of product bound by Dicer/TRBP (Bound siRNA/Total siRNA) using data from three experimental replicates (means). Red circles: As the competitor concentration is increased, the fraction of bound product decreases substantially. Blue squares: Increasing concentrations of a structured Alu RNA did not compete away bound product. Green diamonds: When siRNA 12a/b is pre-bound, increasing concentrations of unlabeled siRNA 12a/b do not lead to a significant decrease in the fraction of bound product. Values for each condition were normalized to the fraction of product bound when no competitor was present. **(C)** Quantification of product formation for dicing reactions (Fraction cleaved/Total) using data from three experimental replicates (means). Red circles: Increasing concentrations of cold siRNA 12a/b have a limited effect on dicing activity. Blue squares: Unlabled Alu RNA also fails to compete with dicing activity. Values for each condition were normalized to the fraction substrate cleaved when no competitor was present. **(D)** Dicer/TRBP cannot bind both substrates and products simultaneously. Dicer/TRBP complex was pre-incubated with dsRNA 40a/b on ice followed by a secondary one hour incubation on ice with increasing concentrations of siRNA 12a/b containing 1,000 cpm ^{32}P-labeled siRNA 12a/b. Reactions were then applied to Avidin agarose resin and incubated on ice for one hour. Reactions were washed 5 times and run on a 15% denaturing PAGE gel. **(E)** Biotinylated substrate RNAs can precipitate the Dicer/TRBP heterodimer. Dicer/TRBP was pre-incubated in the presence and absence of dsRNA 40a/b on ice for one hour. Reactions were then applied to Avidin agarose resin and incubated on ice for one hour. Reactions were washed 5 times run on a 4-20% gradient gel. Proteins were visualized by silver staining.

3.3.4 The Dicer/TRBP heterodimer senses siRNA thermodynamic asymmetry

A previous report demonstrated that the Dicer-2/R2D2 heterodimer functions as a sensor for siRNA thermodynamic asymmetry to direct strand specific RISC loading in *Drosophila* (Tomari et al. 2004b). We sought to determine the extent to which the Dicer/TRBP heterodimer functions analogously in humans. To test for sensing of siRNA asymmetry by human complexes, we used a set of synthetic siRNAs identical to that used previously (**Figure 3.6 A; Table 3.1** (Tomari et al. 2004b)). Briefly, these RNAs were each based on the thermodynamically symmetric sequence of siRNA 1 (this siRNA is identical to siRNA D in **Figure 3.3 A**, see **Table 3.1**), in which one strand contains a 5-iodo-uracil modification at the 20th base counting from the 5′ end. siRNAs 2-4 contained a single mismatch at one end to introduce thermodynamic asymmetry. The 5-iodo-uracil modification lies on either the 3′ overhang of the presumptive guide strand, passenger strand, or both. siRNAs 5-7 instead had a mismatch introduced at the opposite end of the duplex, flipping the asymmetry and thus the predicted guide strands of the duplexes.

These modified siRNAs were incubated with the Dicer/TRBP complex, followed by irradiation with 302 nm light and separation on an SDS-PAGE gel. If the Dicer/TRBP heterodimer behaves analogously to the Dicer-2/R2D2 heterodimer, Dicer would photocrosslink preferentially to siRNAs that have a 5-iodo-uracil modification located at the less stable end of the duplex. TRBP, on the other hand, would be expected to photocrosslink preferentially to siRNAs containing the photocrosslinkable base at the more stable end. We found that the Dicer/TRBP heterodimer did exhibit this crosslinking pattern (**Figure 3.6 B, C**), indicating that this human Dicer/dsRBP heterodimer functions as a sensor for siRNA asymmetry similar to that reported in flies. Based on additional crosslinking data, this sensor is also functional in the larger context of a full RLC containing Ago2 (**Figure 3.6 B, C**).

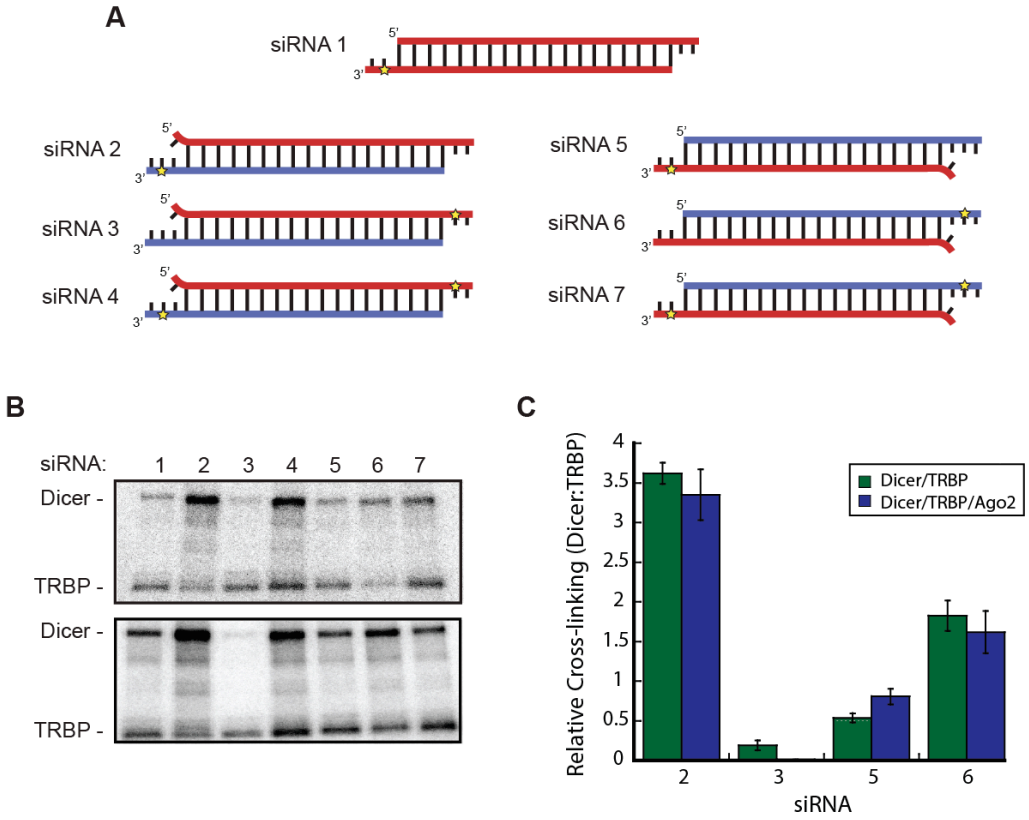

Figure 3.6: Human Dicer/TRBP heterodimers sense siRNA thermodynamic asymmetry.
(A) Diagram of synthetic, modified siRNAs used for asymmetry sensing assays. Stars indicate the location of 5-iodo-uracil modifications. Strands containing 5-iodo-uracil modifications were 5′ end radiolabeled and all unmodified strands were 5′ phosphorylated. Overhangs are 2 nt 3′ overhangs. Curved ends indicate the presence of a mismatched base pair. Predicted guide strands are colored red and passenger strands are colored blue. See text for further details. **(B)** Photocrosslinking assay using reconstituted wild type Dicer/TRBP (upper panel) and Dicer/TRBP/Ago2 (lower panel) complexes. Modified siRNAs were incubated with Dicer/TRBP complex at room temperature followed by irradiation with 302 nm light. Photocrosslinked proteins were detected based on size by SDS-PAGE. **(C)** Quantification of the relative photocrosslinking of siRNAs 2, 3, 5, and 6 to Dicer versus TRBP in asymmetry sensing assays using data from three experimental replicates (mean +/- SD). In the context of both the Dicer/TRBP heterodimer and the core RLC containing TRBP, Dicer was preferentially photocrosslinked when the 5-iodo-uracil modification was introduced at the less stable end of the duplex, whereas TRBP preferentially photocrosslinked to the more stable end.

3.3.5 Heterodimer formation is required for asymmetry sensing in humans

To test the extent to which Dicer or TRBP alone are capable of sensing siRNA thermodynamic asymmetry, we subjected each protein separately to photocrosslinking assays. Although Dicer photocrosslinked to siRNAs in control reactions using the Dicer/TRBP heterodimer, Dicer alone was unable to photocrosslink to any of the siRNAs tested (**Figure 3.7 A, B**). This observation is consistent with the low affinity that this protein has for siRNAs in the absence of a dsRBP (see **Figure 3.1**; (Chakravarthy et al. 2010)). We also found that although TRBP photocrosslinked to each siRNA, it did so with no pattern related to siRNA thermodynamic asymmetry (**Figure 3.7 A, B**).

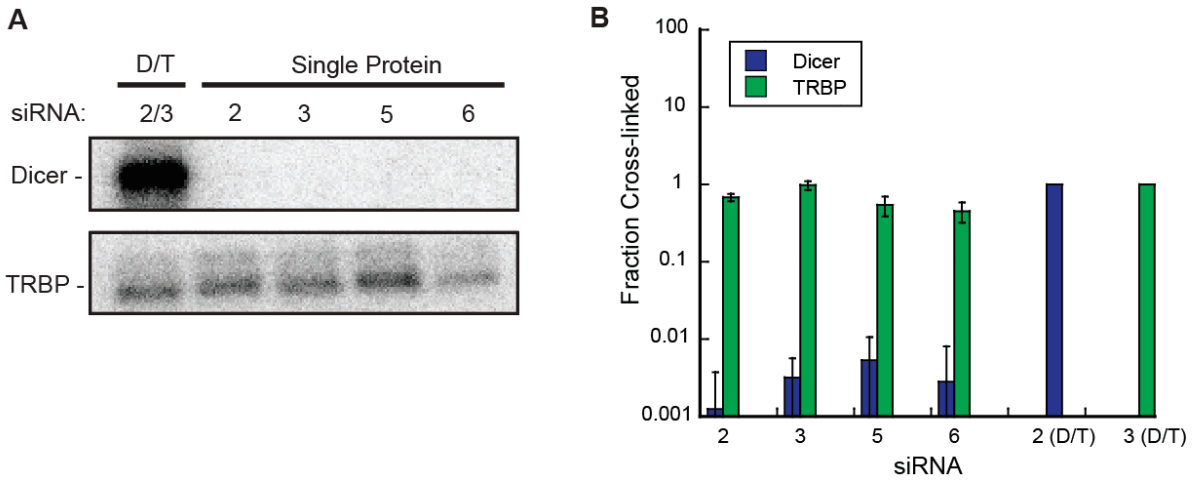

Figure 3.7: Neither Dicer nor TRBP alone sense siRNA thermodynamic asymmetry.
(A) Photocrosslinking assay using free Dicer and TRBP. Dicer alone or TRBP alone was incubated with siRNA 2, 3, 5, or 6 prior to irradiation with 302 nm light. Photocrosslinked proteins were detected based on size by SDS-PAGE. Upper: Dicer alone does not detectably photocrosslink to any of the modified siRNAs used in this study. For comparison, the left lane shows the level of Dicer that photocrosslinks to siRNA 2 in the context of the Dicer/TRBP heterodimer (D/T). Lower: TRBP alone photocrosslinked to each siRNA, but not with any pattern related to the thermodynamic asymmetry of siRNA ends. For comparison, the left lane shows the level of TRBP that photocrosslinks to siRNA 3 in the context of the Dicer/TRBP heterodimer (D/T). **(B)** Quantification of the relative photocrosslinking of siRNAs 2, 3, 5, and 6 to Dicer or TRBP alone (Fraction crosslinked/Total) in asymmetry sensing assays using data from three experimental replicates (means +/- SD). Data were normalized to the levels of Dicer or TRBP that photocrosslinked to siRNA 2 or 3, respectively, in the context of the Dicer/TRBP heterodimer. Note that the y-axis is plotted on a logarithmic scale.

In light of the fact that neither Dicer nor TRBP alone could recapitulate asymmetric photocrosslinking to siRNA ends, we wondered whether heterodimer formation is necessary for the human asymmetry sensor to function. Wild type TRBP consists of three dsRBDs connected by flexible linker regions (**Figure 3.8 A**). The two N-terminal dsRBDs have a high affinity for dsRNA (Parker et al. 2008), whereas the C-terminal dsRBD is responsible for binding to Dicer (Kok et al. 2007). We therefore expressed a TRBP truncation mutant that lacks the C-terminal, Dicer-binding dsRBD. When combined with purified Dicer, no evidence of complex formation was detected by size exclusion chromatography (**Figure 3.8 B, C**). This truncated TRBP was pre-incubated with equimolar amounts of human Dicer and subjected to photocrosslinking assays. We found that under these conditions, Dicer again did not detectably photocrosslink to any of the siRNAs. By contrast, the truncated TRBP did photocrosslink, but again with no pattern related to the thermodynamic asymmetry of the siRNA ends (**Figure 3.8 D**). These findings indicate that heterodimer formation is required for sensing of siRNA thermodynamic asymmetry, and that neither Dicer nor TRBP is competent for this function in their free states.

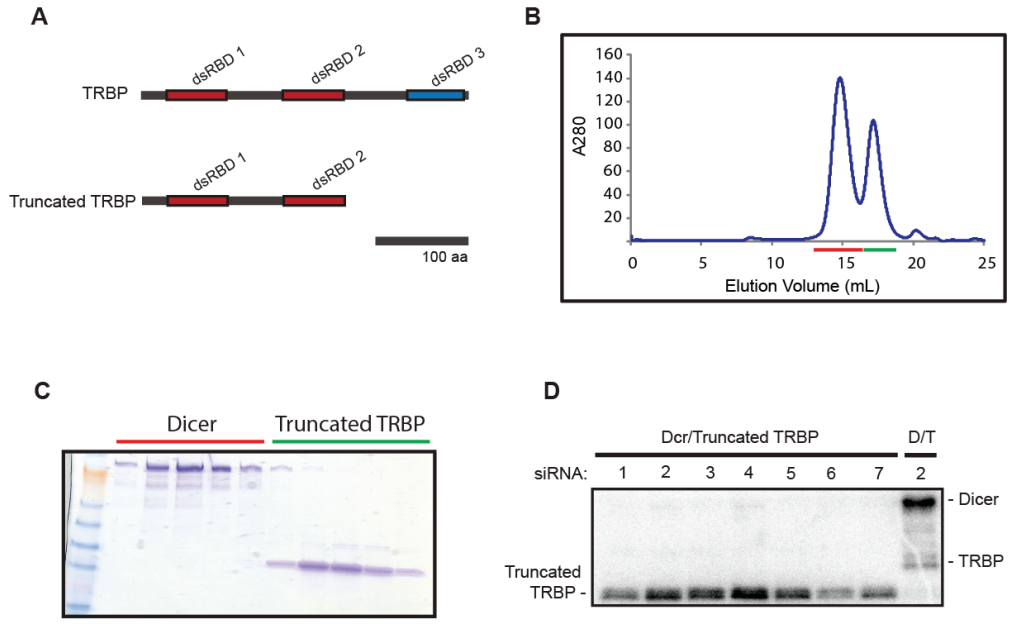

Figure 3.8: Heterodimer formation is required to sense siRNA thermodynamic asymmetry.
(A) Diagram of wild type and truncated TRBP domain structures. The dsRBD responsible for binding to Dicer is shown in blue. **(B)** Superose 6 elution profile for Dicer and truncated TRBP. The peak underlined in red contains Dicer and the peak underlined in green contains unincorporated truncated TRBP. **(C)** SDS-PAGE analysis of elution fractions from B. Lanes are labeled as in B. **(D)** Dicer/TRBP heterodimer formation is necessary for sensing siRNA thermodynamic asymmetry. Dicer and truncated TRBP were pre-incubated on ice prior to incubation with siRNAs 1-7 at room temperature. Reactions were then irradiated with 302 nm light and analyzed by SDS-PAGE. siRNAs were only photocrosslinked to the truncated TRBP protein, and with no pattern related to siRNA thermodynamic asymmetry. For reference, the far right reaction contained wild type Dicer/TRBP complex photocrosslinked to siRNA 2.

3.3.6 dsRBPs activate Dicer's asymmetry sensing functionality

In light of the finding that Dicer cannot sense siRNA asymmetry outside of the context of the Dicer/TRBP complex, we asked whether or not siRNA asymmetry sensing could be rescued by substituting PACT for TRBP, since PACT also binds to Dicer's helicase domain (Lee et al. 2006) and imparts product-binding activity on Dicer (**Figure 3.1**). We subjected the Dicer/PACT heterodimer to the same photocrosslinking assays as for the Dicer/TRBP heterodimer. Likely due to the specific geometry of the photocrosslinking reaction, we saw no crosslinking to PACT in this experiment (**Figure 3.9 A**). Interestingly, however, we did observe photocrosslinking to Dicer in the context of this heterodimer. When we compared the total fraction of Dicer photocrosslinked for each siRNA, we saw a similar pattern of crosslinking to that of Dicer in the context of the Dicer/TRBP heterodimer (**Figure 3.9 A, B**). Dicer reproducibly photocrosslinked more readily when the 5-iodo-uracil modification was located at the less stable end of the duplex. This pattern indicates that the Dicer/PACT heterodimer is also capable of sensing the thermodynamic asymmetry of siRNA ends.

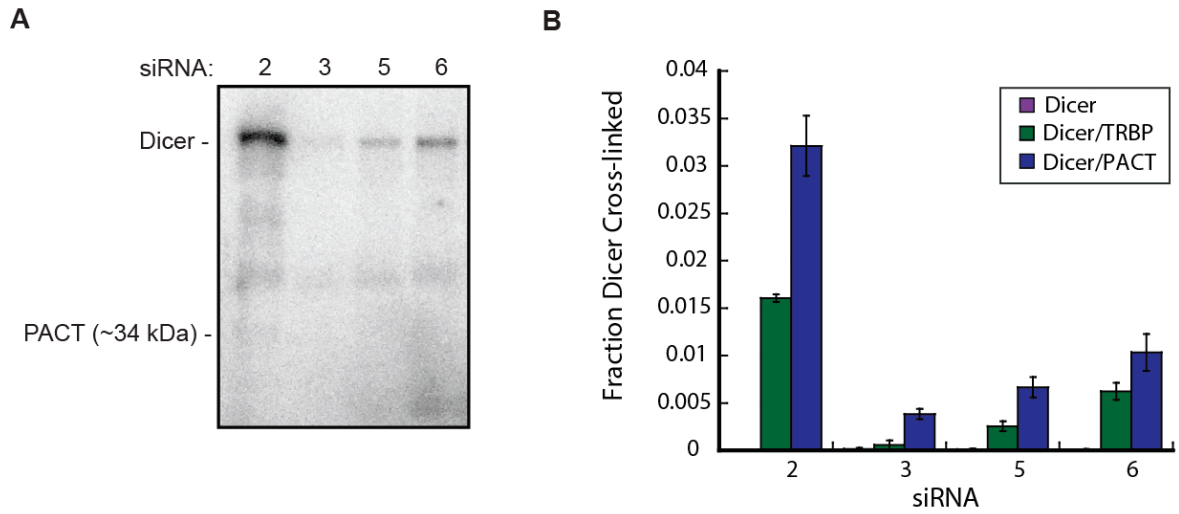

Figure 3.9: Human Dicer/PACT heterodimers sense siRNA thermodynamic asymmetry.
(A) The Dicer/PACT heterodimer senses siRNA thermodynamic asymmetry. siRNAs 2, 3, 5, and 6 were incubated with Dicer/PACT complex at room temperature followed by irradiation with 302 nm light. Photocrosslinked proteins were detected based on size by SDS-PAGE. No detectable crosslinking to PACT was observed. **(B)** Quantification of the fraction Dicer photocrosslinked (Fraction crosslinked/Total) to siRNAs 2, 3, 5, and 6 in asymmetry sensing assays using data from 3 experimental replicates (mean +/- SD) using either Dicer, the Dicer/PACT heterodimer, or the Dicer/TRBP heterodimer.

Based on this finding, we hypothesized that asymmetry sensing requires Dicer to be in complex with a dsRBP, but not necessarily TRBP or PACT. To investigate this prediction, we created a chimeric dsRBP in which we fused the two dsRBDs from the unrelated spermatid perinuclear RNA binding protein (SPNR) to the C-terminal Dicer binding domain of TRBP (**Figure 3.10 B**; (Schumacher et al. 1995; 1998)). Wild type SPNR does not bind to Dicer (**Figure 3.10 A**). We reconstituted a heterodimer consisting of this chimeric dsRBP and Dicer (**Figure 3.10 C, D**) and subjected the complex to photocrosslinking assays to test for a functional asymmetry sensor. This tethering of SPNR dsRBDs, which are not implicated in RNAi, to TRBP's Dicer binding dsRBD was sufficient to impart asymmetry sensing capability to the heterodimer (**Figure 3.10 E, F**).

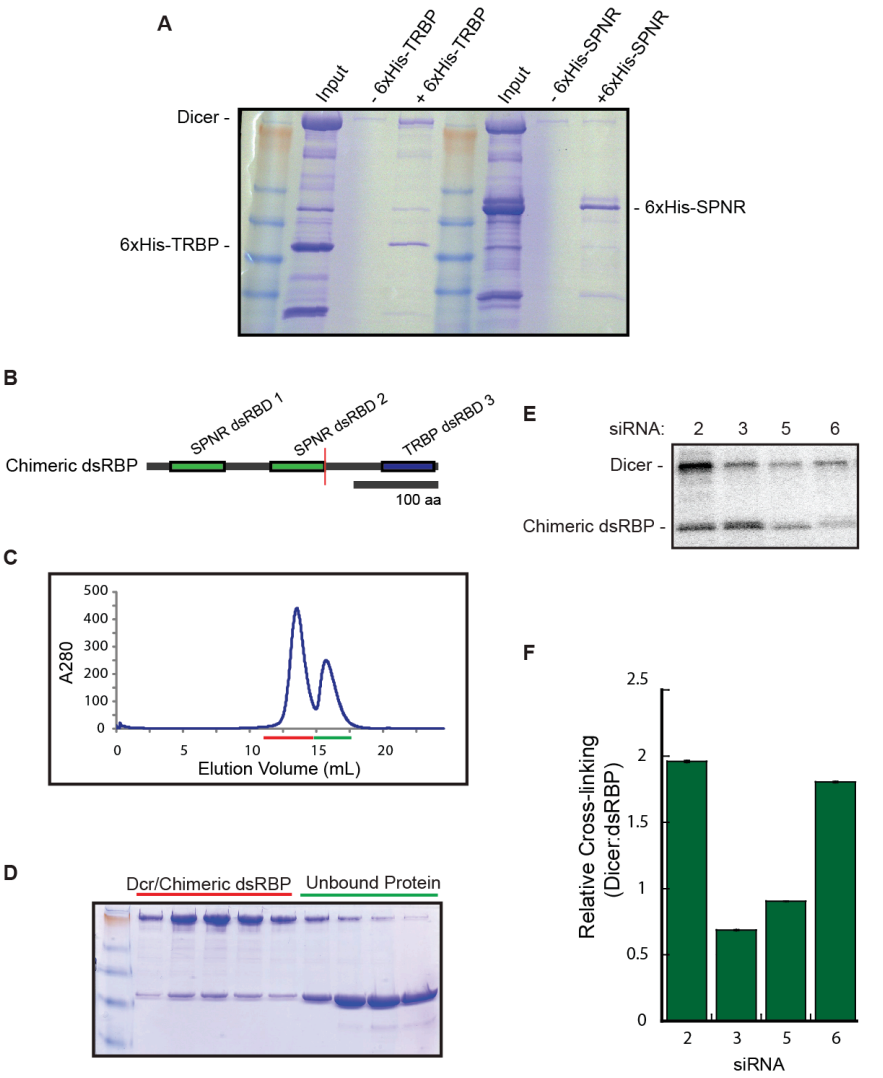

Figure 3.10: Dicer/Chimeric dsRBP heterodimers sense siRNA thermodynamic asymmetry.
(A) Wild type SPNR does not bind to Dicer. Dicer was pre-incubated with or without 6xHis-TRBP or 6xHis-SPNR for one hour. Binding reactions were then applied to nickel-NTA agarose resin and incubated at 4°C for one hour. The resin was washed 3 times and samples were run on a 4-20% gradient gel. Proteins were visualized by Coomassie Blue staining. **(B)** Diagram of chimeric dsRBP domain structure. Sequence N-terminal to the vertical red line derives from SPNR. Green boxes indicate dsRBDs from SPNR. Sequence C-terminal of the vertical red line derives from TRBP. Blue box indicates TRBP's Dicer binding dsRBD. **(C)** Superose 6 elution profile for the Dicer/Chimeric dsRBP heterodimer. The peak underlined in red contains the heterodimer and the peak underlined in green contains unincorporated protein. **(D)** SDS-PAGE analysis of elution fractions from C. Lanes are labeled as in C. **(E)** Dicer/Chimeric dsRBP complex was incubated with siRNAs 2, 3, 5, or 6 at room temperature prior to irradiation with 302 nm light. Photocrosslinked proteins were detected based on size by SDS-PAGE. **(F)** Quantification of the relative photocrosslinking of siRNAs 2, 3, 5, and 6 to Dicer versus the dsRBP in asymmetry sensing assays using data from 3 experimental replicates (mean +/- SD). Dicer was preferentially photocrosslinked when the 5-iodo-uracil modification was introduced at the less stable end of the duplex, whereas the dsRBP preferentially photocrosslinked to the more stable end.

The above results strongly suggest that Dicer is the protein sensor for siRNA thermodynamic asymmetry in humans. One clear question, however, was whether or not the third dsRBD of TRBP may itself retain some residual asymmetry sensing capabilities in a complex with Dicer. To rule out this possibility, we constructed a second fusion protein in which

the SPNR dsRBDs were fused directly to the N-terminus of Dicer via an SPNR-derived linker containing an engineered TEV cleavage site (**Figure 3.11 A, B**), giving a physically linked heterodimer containing no sequence from TRBP or PACT. We performed photocrosslinking assays with this protein with an added TEV cleavage step following exposure to 302 nm light. Although the background of SPNR crosslinking was higher in this case (likely due to residual crosslinking that occurred following TEV cleavage), Dicer clearly retained its preference for crosslinking to the less stable end of siRNA duplexes in the absence of TRBP (**Figure 3.11 C, D**). Taken together, these data demonstrate that Dicer itself is the functional sensor of siRNA thermodynamic asymmetry in humans, and that Dicer-associated dsRBPs play an essential positioning role to allow this sensing to occur.

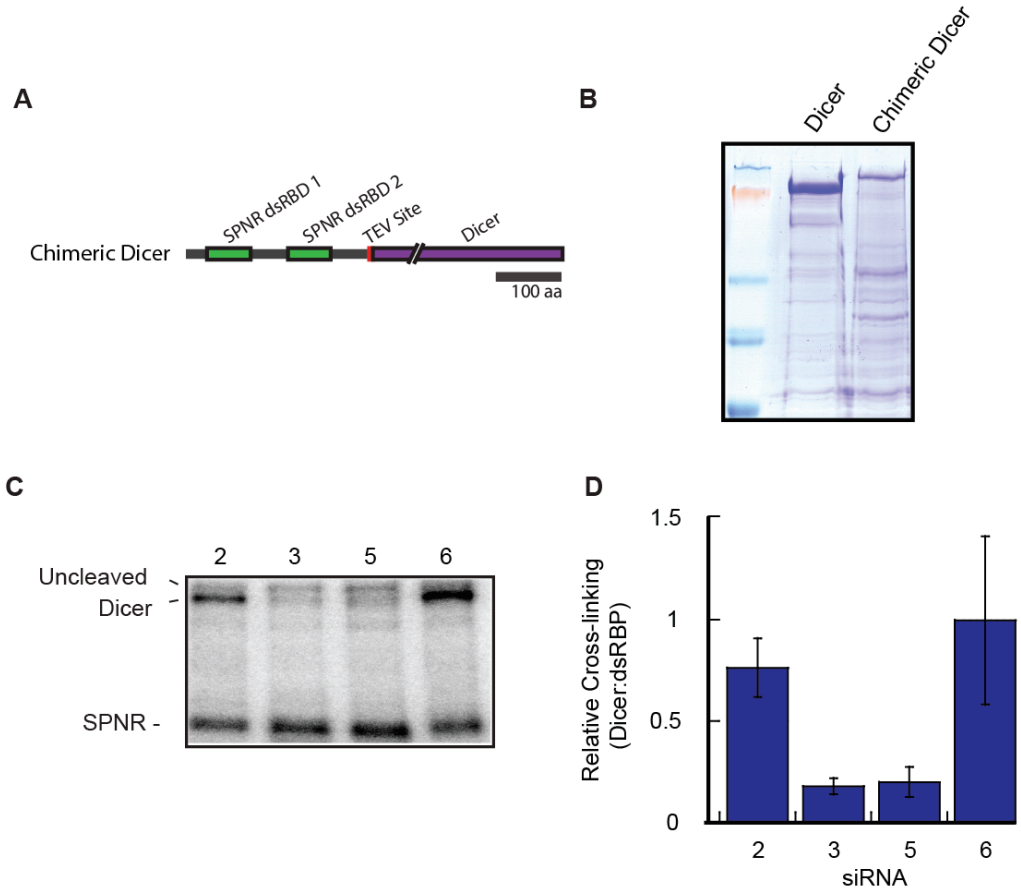

Figure 3.11: Human Dicer is a protein sensor of siRNA thermodynamic asymmetry.
(A) Diagram of chimeric Dicer domain structure. Green boxes indicate SPNR dsRBDs. Small red box indicates TEV cleavage site. Purple box represents full length Dicer. Diagonal break indicates a break from scale. **(B)** SDS-PAGE analysis of wild type recombinant Dicer versus chimeric Dicer. **(C)** Chimeric Dicer was incubated with siRNAs 2, 3, 5, or 6 at room temperature prior to irradiation with 302 nm light. Following TEV cleavage, photocrosslinked proteins were detected based on size by SDS-PAGE. **(D)** Quantification of the relative photocrosslinking of siRNAs 2, 3, 5, and 6 to Dicer versus the dsRBP in asymmetry sensing assays using data from 3 experimental replicates (mean +/- SD). Care was taken to ensure that the small residual signal corresponding to uncleaved fusion protein was not entered into the analysis. Dicer was preferentially photocrosslinked when the 5-iodo-uracil modification was introduced at the less stable end of the duplex, whereas the dsRBP preferentially photocrosslinked to the more stable end.

Previous characterizations of mouse embryonic stem cells in which portions of Dicer's RNaseIII domains are excised have demonstrated that although these cells cannot process Dicer substrates, exogenous siRNAs are still able to knock down gene expression (although strand selection was not specifically examined in either case (Murchison et al. 2005; Kanellopoulou et al. 2005). One interpretation of these results is that Dicer is not necessary for RISC loading in mammals. However, in both cases it was possible that a truncated form of Dicer containing the DExH/D helicase domain was still expressed. Such a Dicer fragment could retain its TRBP- and siRNA-binding capacity and may therefore sense siRNA thermodynamic asymmetry through the interactions proposed here. To investigate the possibility that such a Dicer fragment is expressed in these knockout cells, we performed a western blot analysis of cell lysates using an antibody that recognizes Dicer's N-terminus (Murchison et al. 2005; Kanellopoulou et al. 2005). Interestingly, in the mouse knockout cells, but not the conditional (un-floxed) control cells, we detected a band corresponding to ~100 kDa (**Figure 3.12**). To rule out the possibility that this signal represented a non-specific antibody interaction with a protein expressed at higher levels in the mouse knockout cells than in the control cells, we used the same antibody to immunoprecipitate this fragment from cell lysates and ran the eluate on an SDS-PAGE gel. We then cut a band from the gel corresponding to the size of the fragment and submitted the sample for analysis by mass spectrometry. We were able to detect three tryptic fragments of mouse Dicer each corresponding to the N-terminal portion of the protein using a cross-correlation cutoff that yielded results with greater than 95% confidence (**Table 3.2**). A large, N-terminal fragment of Dicer is therefore expressed in these mouse Dicer "knockout" cells, indicating that although Dicer's main catalytic function is absent in these cells, it is possible that other Dicer functions such as asymmetry sensing are still carried out.

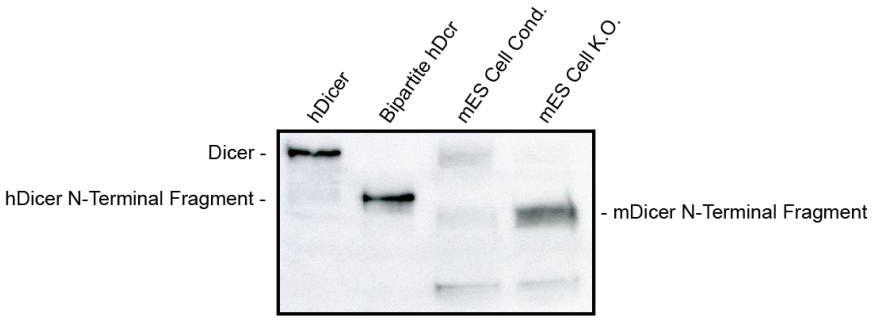

Figure 3.12: Mouse Dicer null embryonic stem cells express an N-terminal Dicer fragment.
Whole cell lysates were run on a 10% SDS-PAGE gel, transferred to a PVDF membrane, and probed with antibody 1414 (Kanellopoulou et al. 2005). Recombinant wild type and bipartite Dicer proteins were run as controls.

Table 3.2: Mouse Dicer peptides present in Dicer null embryonic stem cells.

dicer1 [Mus musculus]			
Ion%	**#**	**Sequence**	**Amino Acids**
81.2%	1	K.TFIAVLLTK.E	70-80
77.8%	1	K.FLLFTDTLLR.K	343-354
72.7%	1	R.AQTASDAGVGVR.S	1071-1084

3.4 Discussion

Considerable efforts have shed light on the mechanisms of both dicing and target mRNA cleavage in RNAi, but the process by which an siRNA is strand-selectively transferred to Ago2

from Dicer in humans has been poorly understood. In this study, we used biochemical approaches to investigate the post-processing fates of siRNAs within the Dicer/TRBP and Dicer/PACT RLC subcomplexes as they are positioned and pre-oriented for strand-selective RISC loading. The findings presented here allow us to propose a model for siRNA positioning and RISC loading by RLCs that begins to establish the molecular basis of this process in humans (**Figure 3.13**).

Our data provide evidence that the Dicer/TRBP heterodimer binds Dicer substrates and products by distinct mechanisms. Dicer substrates are engaged by the catalytic region of the enzyme, as would be necessary for productive dicing. TRBP seems to stabilize this interaction, as human Dicer did not photocrosslink to substrates in the absence of TRBP in our assays (data not shown). Following cleavage, dicing products are released into the bulk solvent, similar to the *Drosophila* system and in contrast to previous models of the human enzyme (Preall et al. 2006; Rose et al. 2005). Rose and colleagues approached this question by using modified pairs of synthetic pre-siRNAs that restricted Dicer binding to one or the other end of the duplex, but that would nonetheless each produce the same siRNA products. They reported that the processing polarity of Dicer affects the extent of gene silencing in human cells, and concluded based on this finding that human Dicer does not release its products before RISC loading. This report, however, did not examine the potential effects that differing dicing efficiencies across different substrates would have on overall silencing. Additionally, the effects of heterogeneous product formation (21- versus 22 nt siRNAs) on silencing were not examined for all cases in which it occurred. While it is possible that the product release observed here is not obligate or that additional factors prevent siRNA dissociation *in vivo*, our *in vitro* data strongly favor a model of product release prior to repositioning.

We show that the Dicer/TRBP heterodimer rebinds its siRNA products (or exogenously administered siRNAs) at Dicer's helicase domain. Given PACT's similar binding site on Dicer and the fact that this dsRBP also imparts product binding on Dicer, the same is likely true for the Dicer/PACT heterodimer. Based on recently published EM structures, this repositioning represents an almost 90° shift in binding orientation, although it is possible that RNA binding to the helicase domain induces a conformational shift in Dicer that alters this angle (Lau et al. 2009; Wang et al. 2009b). This staging step allows an siRNA to be bound coordinately between Dicer and TRBP (which binds at the distal end of Dicer's helicase branch, see **Figures 2.6 and 3.13**). Bound in this way, an siRNA can be oriented by the heterodimer according to the thermodynamic asymmetry of its ends in a way that would not be possible if the products were bound along Dicer's catalytic domain, far from TRBP or PACT. Interestingly, a recent EM study showed that Dicer's PAZ domain is located at the distal end of the catalytic branch (Lau et al. 2012). This finding suggests that the end of the siRNA that is bound to Dicer might not be engaged by the PAZ domain.

Our findings have demonstrated that the human core RLCs are capable of sensing the thermodynamic asymmetry of siRNA ends *in vitro*. These complexes bind siRNAs directionally such that Dicer interacts with the less stable end of the duplex and its associated dsRBP interacts with the more stable end. This finding alone, however, does not directly address the question of which component represents the functional thermodynamic asymmetry sensing machinery. A previous report concluded that TRBP alone preferentially photocrosslinks to the more stable end of siRNA duplexes, which led to the conclusion that TRBP functions to sense siRNA thermodynamic asymmetry in humans, whereas Dicer acts as more of a passive element in the asymmetry sensing process (Gredell et al. 2010). These findings, however, were inconsistent

across all siRNAs tested. Perhaps more importantly, given that in the same report the very siRNA strands that were found to photocrosslink to TRBP preferentially in the context of an asymmetric duplex were also crosslinked preferentially to TRBP when incubated as separate ssRNAs, one cannot rule out the possibility that these findings reflect a photocrosslinking artifact related to the tendencies of individual ssRNAs to crosslink preferentially rather than to duplex stability.

In our study, the fact that TRBP alone photocrosslinks with no pattern related to thermodynamic asymmetry argues that its function is not to bind the more stable end of an siRNA and then directionally recruit that RNA to Dicer, as was previously proposed. Rather, our experiments using various heterodimers and chimeric proteins demonstrate that asymmetry sensing is only activated upon heterodimer formation, and that Dicer itself is the *bona fide* sensor of siRNA thermodynamic asymmetry. In this light, Dicer-associated dsRBPs likely act as positioning elements that hold siRNA products in place while Dicer senses the thermodynamic stabilities of siRNA ends and binds preferentially to the less stable end.

Given that our data show siRNAs to be bound by Dicer's DExH/D helicase domain, it is tempting to implicate this domain in Dicer's asymmetry sensing functionality. Although in our assays asymmetry sensing does not require ATP, recent reports have shown that certain helicases are capable of unwinding short stretches of RNA in an ATP-independent manner (Liu et al. 2008). Partial unwinding of an unstable siRNA end could potentially allow Dicer to bind that siRNA end more favorably, imparting the observed directionality of product binding by Dicer/dsRBP heterodimers. Unfortunately, point mutations in Dicer's helicase domain render the protein highly unstable, precluding a direct analysis of this hypothesis. Clearly, further structural and biochemical analyses of the precise nature of siRNA binding by Dicer/dsRBP complexes are necessary to firmly establish the mechanistic basis of this asymmetry sensor.

In terms of strand-selective RISC loading by the core human RLCs, it is of mechanistic significance that asymmetry sensing by the heterodimers investigated here leads TRBP or PACT to preferentially bind the 3′ end of the guide strand (which lies at the more stable end of the duplex). In our previously reported EM structure of a reconstituted RLC containing TRBP, we modeled the crystal structure of a guide strand-containing *Thermus thermophilus* Argonaute into the EM density (Wang et al. 2009b). Intriguingly, this docking placed the Ago PAZ domain near TRBP and the outer edge of Dicer's helicase domain. In binding the 3′ end of the guide strand, TRBP or PACT thus likely holds that end of the siRNA in close proximity to the Ago2 PAZ domain and in the proper orientation for guide-strand binding by the apposed domain. Given TRBP's apparently high level of flexibility, it is possible that this protein or PACT could directly hand off the 3′ end of the guide strand to Ago2 in a RISC-priming step that leads to a previously proposed state in which an siRNA is transiently coordinately bound between Dicer and the Ago2 PAZ domain (Wang et al. 2009b). Further studies will be needed to definitively establish whether or not this RISC loading intermediate exists.

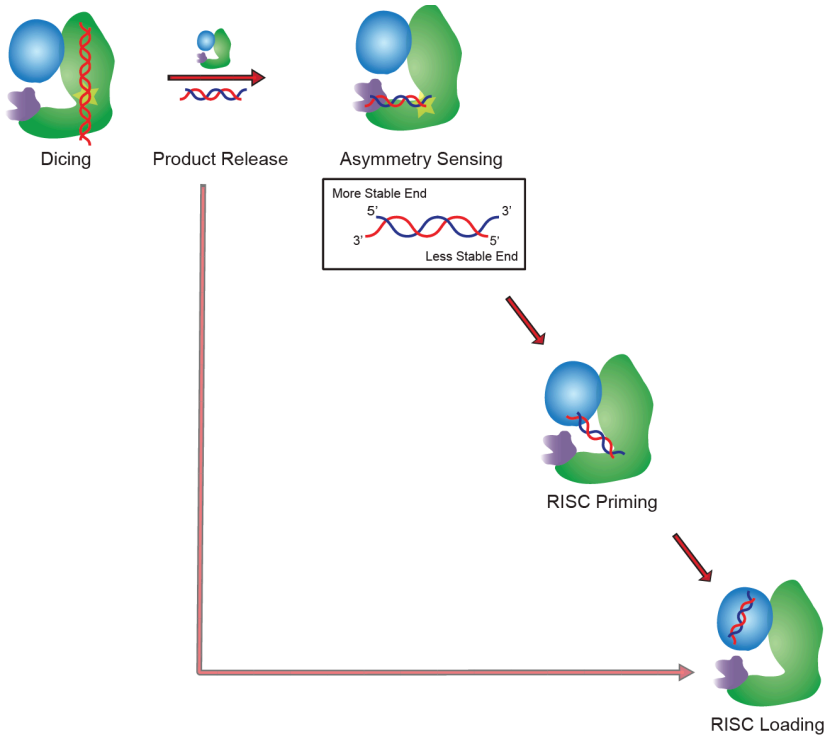

Figure 3.13: Model for dsRNA positioning by human RISC Loading Complexes.
Dicer engages pre-siRNA substrates along its catalytic domain, positioning them in the active site for cleavage. Products are then released and may potentially be bound directly by Ago2 or rebound along Dicer's helicase branch. In this conformation, Dicer products are bound directionally based on the thermodynamics of the duplex such that Dicer actively senses and binds to the less stable end, leaving the dsRBP to bind the more stable end. In the following hypothetical RISC priming step, the dsRBP passes the 3′ end of the guide strand to the Ago2 PAZ domain, securing the orientation of the siRNA to ensure a properly loaded RISC with an enhanced level of strand-selectivity.

Our findings do not exclude the possibility that the asymmetry sensing observed here represents only one part of a more complex system of semi-redundant guide strand selection checkpoints *in vivo*. For example, prokaryotic Argonaute structures indicate that the two 5′ bases of the guide strand splay apart and do not base pair with the passenger strand (Wang et al. 2009c), indicating that Ago2 itself may also bind siRNAs in a preferential orientation that serves as a secondary proofreading mechanism for proper guide strand loading. Additionally, human Ago2 loading may be affected by internal miRNA structural elements, as is the case in *Drosophila* (Okamura et al. 2009). Nonetheless, the results of this study demonstrate that human Dicer/dsRBP RLC subcomplexes are able to position siRNAs distinctly from pre-siRNA substrates and orient them with respect to the relative thermodynamic stabilities of duplex ends, strongly suggesting a role for these heterodimers in guide-strand selection during the RISC loading process in humans.

Chapter 4

Multiple Sensors Ensure Accurate Guide Strand Selection in Human RNAi

4.1 Introduction

In certain organisms, the siRNA and miRNA pathways are clearly delineated. For example, in *Drosophila*, which express two separate Dicer enzymes, Dicer-2 generates siRNAs whereas Dicer-1 generates miRNAs (Lee et al. 2004b). This distinction is less clear in mammals, which have only one Dicer enzyme that is capable of processing both pre-siRNAs and pre-miRNAs (Provost et al. 2002; Zhang et al. 2002). Human Dicer has two known dsRBP partners, TRBP and PACT. Both proteins form heterodimers with Dicer and are important for the processing of pre-siRNAs and pre-miRNAs, although a more specific delineation of the roles of these two dsRBPs has not yet been achieved. A general consensus has been that there is some degree of functional overlap between the two in terms of their involvement in the siRNA and miRNA pathways, but a quantitative comparison of the relative contributions that each protein makes to these pathways has not been carried out (Chendrimada et al. 2005; Haase et al. 2005; Lee et al. 2006; Kok et al. 2007).

In *Drosophila*, the delineation of the siRNA and miRNA pathways continues at the RISC loading step, although a fair amount of overlap exists. Loading of Dicer product RNAs onto Ago proteins in *Drosophila* requires an RLC comprised of an Ago protein, a Dicer protein, and a dsRBP. This process is significantly enhanced by the Hsc70/Hsp90 chaperone machinery (Pare et al. 2009; Johnston et al. 2010; Iwasaki et al. 2010; Miyoshi et al. 2010; Iki et al. 2010; Liu et al. 2003). In terms of pathway delineation, the *Drosophila* RLC composed of Ago2, Dicer-2, and R2D2 has been shown to promote the loading of siRNAs but selects against miRNAs. Conversely, miRNAs are preferentially loaded onto Ago1 by a poorly understood mechanism (Tomari et al. 2007). In mammals, siRNAs and miRNAs do not appear to be differentially sorted into the four Ago proteins (Ago1-Ago4), although Ago2 appears to be the most highly expressed and is the only human Ago that exhibits endonuclease activity (Meister et al. 2004; Liu et al. 2004; Dueck et al. 2012; Wang et al. 2012). Since Ago2 binds to both siRNAs and miRNAs, the delineation of these pathways at this level in humans remains unclear.

Guide strand selection in *Drosophila* is based primarily on the detection of RNA duplex thermodynamics by the RLC (Schwarz et al. 2003; Khvorova et al. 2003). Thermodynamically asymmetric siRNAs are pre-oriented by Dicer-2 and R2D2 such that Dicer-2 binds to the less stable end and R2D2 binds to the more stable end prior to loading onto Ago2 (Tomari et al. 2004b). The fidelity of this process is fundamental to the target specificity of RNAi pathways. We have shown that analogous human complexes containing either TRBP or PACT function as siRNA thermodynamics sensors as well, but the ultimate importance of these complexes for strand-selective RISC loading has yet to be fully demonstrated (see **Chapter 3**; (Noland et al. 2011)). Although Ago2/Dicer/dsRBP complexes have been shown to be the primary complexes involved in RISC loading in humans (Liu et al. 2012), for some siRNAs and miRNAs human Ago2 is sufficient for loading and accurate guide strand selection in the absence of Dicer (Betancur and Tomari 2012; Kanellopoulou et al. 2005; Murchison et al. 2005). The extent to which human complexes containing Dicer and either of its associated dsRBPs are true RLCs and contribute to strand selection has therefore been a topic of debate. Furthermore, it is unknown how competing parameters such as duplex thermodynamics, 5′ nucleotide identity, and structure may affect the relative contributions of these different RNAi components to overall guide strand selection efficiency in humans.

Here we have used a reconstituted system to determine the extent to which the core components of the human RNAi machinery contribute to RNA guide strand selection. We show that recombinant Ago2 alone can utilize duplex siRNAs and miRNAs to target and cleave RNAs

in vitro without the aid of chaperones or C3PO. Furthermore, we demonstrate that Ago2 has an intrinsic but substrate-dependent strand selection capability. Dicer and TRBP or PACT are often essential for achieving maximal levels of strand selection, and in certain cases the precise level of strand selection depends on the particular dsRBP present in the complex. Interestingly, complexes containing TRBP are considerably less efficient at RNA targeting using miRNAs than complexes containing PACT. These findings demonstrate that specific RNA duplex features dictate the relative contributions of human RNAi components to overall guide strand selection and suggest that TRBP and PACT may play distinct roles in this process.

4.2 Methods

4.2.1 *In vitro* reconstitution of Ago2/Dicer/dsRBP complexes

Binding reactions were carried out in a final volume of 500 μL Gel Filtration Buffer containing a 1:2:4 molar ratio of Dicer:Ago2:dsRBP (assuming an initial dsRBP dimer). Reactions were incubated at 4°C for 30 minutes and then applied to a Superose 6 10/300 GL column (GE). Fractions containing the desired complex were pooled conservatively to ensure that no free Ago2 or dsRBPs were present. All complexes were concentrated to 5 μM and aliquots stored at -80°C.

4.2.2 RNA preparation

All ssRNAs were chemically synthesized by Integrated DNA Technologies. ssRNAs to be used for siRNA and miRNAs were 5′ phosphorylated. All target RNAs were perfectly complimentary to their respective guide strands except for mismatches with the last 4 bases at the 3′ end of the guide strand to promote turnover (Haley and Zamore 2004; Liu et al. 2011). All ssRNAs were gel purified on a 15% denaturing PAGE gel. Eluted RNAs were washed thoroughly to remove contaminating urea. dsRNAs were prepared by incubating equimolar amounts of each ssRNA together in Annealing Buffer (100 mM HEPES, pH 7.5; 30 mM KCl; 3 mM $MgCl_2$) at 65°C for 10 minutes. Annealing reactions were then slow-cooled by removing the heat block and placing it on the benchtop to equilibrate to room temperature. Annealed dsRNAs were further gel purified on a 15% native gel to remove any remaining ssRNAs.

4.2.3 Nearest neighbor RNA duplex thermodynamics analysis

Thermodynamic stabilities of the RNA duplex ends were calculated using nearest neighbor methods for the terminal four base pairs and 2 nt 3′ overhangs at each end of the duplex (see **Figure 1.3**; (O'Toole et al. 2005; Freier et al. 1986; Xia et al. 1998; Mathews et al. 1999; Borer et al. 1974; Tinoco et al. 1973)).

4.2.4 Strand-selective RNA cleavage assays

For target RNA cleavage assays using siRNAs, 50 nM Ago2, or either complex was pre-incubated on ice for 30 minutes with 50 nM siRNA or miRNA in 18 μL Reaction Buffer (20 mM Tris-HCl, pH 6.5; 75 mM KCl; 1 mM $MgCl_2$; 2.5% glycerol; 0.1 mg/mL BSA; 2.5 mM TCEP). Following an incubation on ice for 30 minutes, these RISC loading reactions were then divided into two 9 μL slicing reactions and 1 μL (1,000 cpm, ~0.5 nM) 5′ ^{32}P-radiolabeled target RNA corresponding to the guide strand was added to one reaction and 1 μL (1,000 cpm, ~0.5 nM) 5′ ^{32}P-radiolabeled target RNA corresponding to the passenger strand was added to the other. Target cleavage reactions were carried out for 30 minutes at 37°C and were stopped by phenol:chloroform extraction and ethanol precipitation of the RNAs. RNA pellets were resuspended in 2x Formamide RNA Loading Dye (95% Formamide, 18 mM EDTA, 0.025%

SDS, 0.1% xylene cyanol, 0.1% bromophenol blue). Product RNAs were resolved by 15% denaturing PAGE and visualized by phosphorimaging.

Target RNA cleavage assays using miRNAs were identical to those with siRNAs with the exception that two separate 9 μL reactions were set up for each protein. One of these reactions contained a 5′ biotinylated DNA capture oligonucleotide with complementarity to the passenger strand and the other contained a capture oligonucleotide complementary to the guide strand (this is to prevent re-loading of unwound strands, since Ago2 binds efficiently to ssRNAs). Following an initial 30 minute incubation on ice, 1 μL (1,000 cpm, ~0.5 nM) 5′ ^{32}P-radiolabeled target RNA corresponding to the guide strand was added to the reaction containing the passenger strand capture oligonucleotide. A radiolabeled target corresponding to the passenger strand was added to the reaction containing the guide strand capture oligonucleotide. Target cleavage reactions were then carried out as above.

4.3 Results

4.3.1 Ago2 cleaves target RNAs *in vitro* using duplex siRNAs and miRNAs

We wanted to establish a minimal *in vitro* system for assessing the individual contributions of different RNAi proteins to strand-selective target cleavage. As a starting point, we attempted to load recombinantly purified human Ago2 with a thermodynamically asymmetric duplex siRNA or miRNA to test for conditions that would allow for slicing activity (**Figure 4.1 A, B**). Importantly, these duplex RNAs were gel purified by 15% native PAGE to ensure that no single stranded RNAs (ssRNAs) were present. Surprisingly, although recombinant human Ago2 alone had previously been shown to be unable to bind duplex siRNAs (Rivas et al. 2005; Ye et al. 2011), we found that Ago2 was able to cleave target RNAs using a duplex siRNA or miRNA without the need for either ATP or accessory proteins (**Figure 4.1 C**). A native gel analysis performed at 37°C showed that the duplex RNAs remained double-stranded after 30 minutes, ruling out the possibility that strand dissociation and subsequent loading of ssRNAs accounted for the observed slicing activity (**Figure 4.1 D**).

Heat shock proteins have been shown to enhance Ago2 loading in both *Drosophila* and humans (Miyoshi et al. 2010; Iwasaki et al. 2010), and it was therefore possible that these results were simply due to trace amounts of insect Hsp90 that may have co-purified with Ago2. To determine if this was the case, we performed RNA cleavage assays in the presence of 1 mM of the Hsp90 inhibitor 17-(allylamino)-17-demethoxygeldanamycin (17-AAG). In the presence of this inhibitor, we saw no decrease in cleavage activity for either the siRNA or miRNA, whereas EDTA strongly inhibited cleavage activity (**Figure 4.1 C**). These findings support the conclusion that Ago2 alone can bind to duplex RNAs and select one strand as a guide for target recognition.

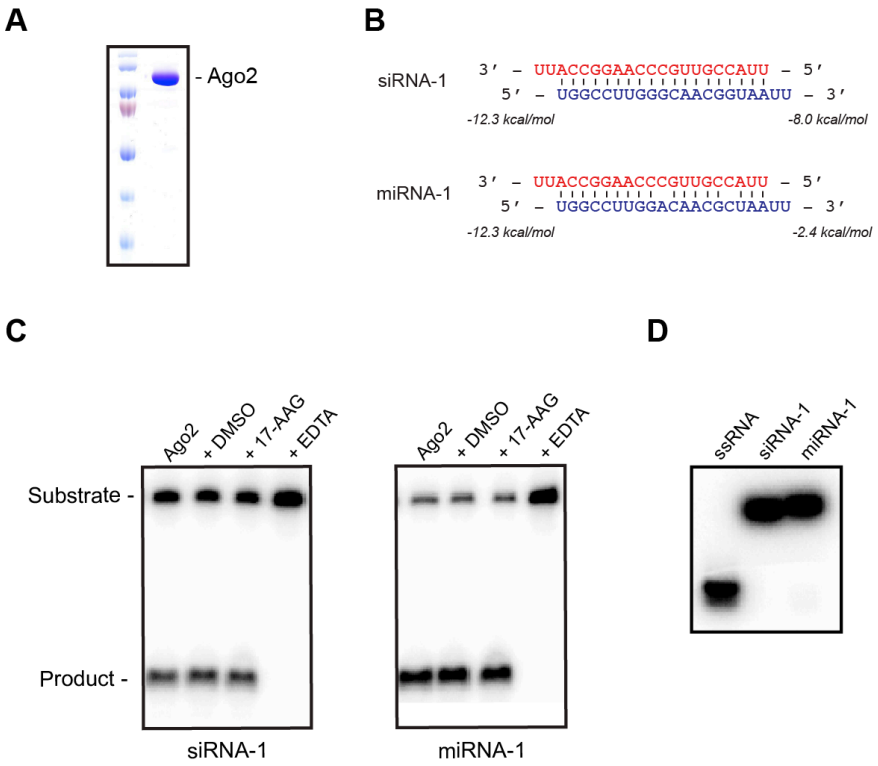

Figure 4.1: Human Ago2 cleaves target RNAs *in vitro* using duplex siRNAs and miRNAs.
(A) Coomassie Blue-stained SDS-PAGE gel of recombinantly purified Ago2. **(B)** Sequences and thermodynamic analysis of siRNA-1 and miRNA-1. Thermodynamics were calculated for the 4 terminal base pairs and 2 nt 3′ overhangs at each end of the duplex. The presumptive guide and passenger strands are shown in red and blue, respectively. **(C)** Recombinant human Ago2 is active for target cleavage using either a duplex siRNA (siRNA-1, left panel) or duplex miRNA (miRNA-1 right panel) as a guide. Neither DMSO nor the Hsp90 inhibitor 17-AAG inhibit the activity of human Ago2, whereas EDTA does inhibit activity. **(D)** siRNA-1 and miRNA-1 remain double stranded at 37°C, as demonstrated by 15% native PAGE analysis.

4.3.2 Dicer and dsRBPs are required for efficient strand selection with some siRNAs

We next tested the extent to which Dicer and TRBP or PACT contribute to strand-selective RISC loading in humans. For this experiment we used siRNA-1, which is thermodynamically asymmetric and has a 5′ uridine on both duplex strands (**Figure 4.1 B**, siRNA-1). We designed a pair of 41 nucleotide target RNAs that were perfectly complementary to one or the other strand of the duplex (except for mismatches corresponding the last four nucleotides at the 3′ end of either strand) and carried out target cleavage assays to determine the extent to which each duplex strand is used for target RNA cleavage.

Ago2 alone or in complex with Dicer and either TRBP or PACT was pre-incubated with siRNA-1 on ice for 30 minutes. Following this initial RISC loading reaction, radiolabeled target RNAs were added and incubated for 30 minutes at 37°C. We found that under these conditions, Ago2 used both the predicted guide strand and passenger strands to nearly equal levels. Both Dicer/dsRBP-containing complexes, however, preferentially used the predicted guide strand for targeting (**Figure 4.2 A, B**). This finding demonstrates that for an siRNA that has the same 5′ nucleotide on both strands, pre-formed Ago2/Dicer/dsRBP complexes are essential for achieving proper strand-selective RISC loading and thus target specificity. In this respect, human Dicer/dsRBP-containing complexes have the ability to function similarly to the *Drosophila* RLC.

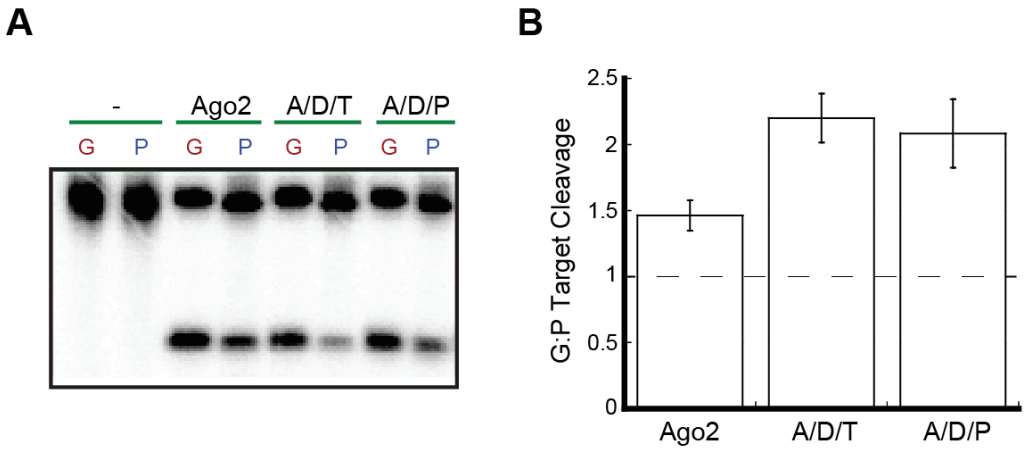

Figure 4.2: Dicer and dsRBPs are required for strand selection with some siRNAs.
(A) Ago2 alone exhibits minimal levels of strand selection when loaded with siRNA-1. Complexes containing Dicer and either TRBP or PACT enhance strand selection. Lanes labeled "G" contain a target corresponding to the guide strand and lanes labeled "P" contain a target that is complementary to the passenger strand. **(B)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates.

4.3.3 Ago2/Dicer/dsRBP complexes are sensitive to changes in duplex stability

We were interested to know if a wider window of siRNA thermodynamic asymmetry would allow for strand selection by Ago2 alone or result in further enhancement of strand selection by Dicer/dsRBP-containing complexes. To this end, we altered siRNA-1 by changing the terminal A:U base pair at the 5′ end of the guide strand to a G:U wobble pair (**Figure 4.3 A**, siRNA-2). With this RNA we again found that Ago2 exhibited only low levels of strand-selective target cleavage. As expected, strand selection was enhanced by both Ago2/Dicer/dsRBP complexes to a higher degree than was observed for siRNA-1 (**Figure 4.3 B, C**). This finding indicates that human Ago2/Dicer/dsRBP complexes are more sensitive to siRNA thermodynamics than Ago2 alone and are capable of translating these small changes in duplex stability to large changes in strand-selective RISC loading.

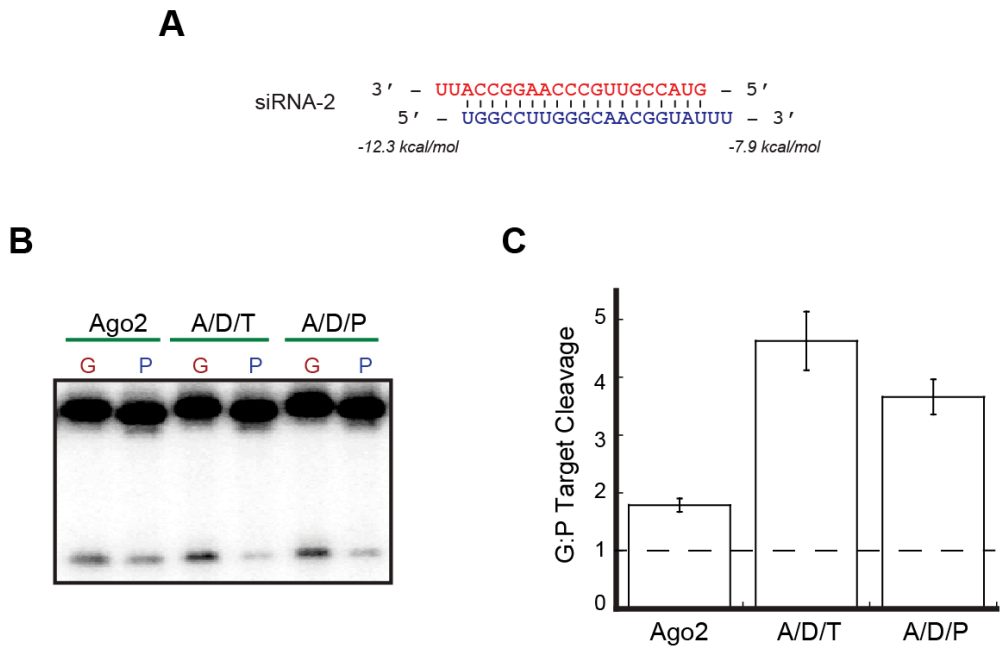

Figure 4.3: Ago2/Dicer/dsRBP complexes are sensitive to changes in duplex stability.
(A) Sequence and thermodynamic analysis of siRNA-2. Thermodynamics were calculated for the 4 terminal base pairs and 2 nt 3′ overhangs at each end of the duplex. The presumptive guide and passenger strands are shown in red and blue, respectively. **(B)** Ago2 alone exhibits minimal levels of strand selection when loaded with siRNA-2. Complexes containing Dicer and either TRBP or PACT enhance strand selection. Lanes labeled "G" contain a target corresponding to the guide strand and lanes labeled "P" contain a target that is complementary to the passenger strand. **(C)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates.

4.3.4 5′ nucleotide identity affects strand selection by human Ago2

In addition to purely thermodynamic considerations, Ago2 is known to have a binding preference for small RNAs with 5′ uridine and adenosine nucleotides, while 5′ cytidines and guanosines are selected against during RISC loading. This is due in part to a conserved selectivity loop in the Ago2 Mid domain (Frank et al. 2010; Hu et al. 2009). In light of this 5′ nucleotide specificity, we were intrigued by the finding that Ago2/Dicer/dsRBP complexes achieved stronger strand-selective target cleavage with siRNA-2 despite the fact that it contained a guanosine nucleotide at the 5′ end of the guide strand. This result demonstrates that the thermodynamics of the duplex are more important than 5′ nucleotides in strand selection by human Ago2/Dicer/dsRBP complexes, but does not rule out the possibility that 5′ nucleotides might play some role. In particular, if an unfavorable nucleotide were present at the 5′ end of the passenger strand, the combined thermodynamic and 5′ nucleotide effects may allow for Ago2 alone to achieve strand-selective target cleavage. In a previous study using an siRNA that had a 5′ C on the passenger strand of an siRNA duplex, target RNAs were cleaved strand-specifically by Ago2 in the absence of Dicer (Betancur and Tomari 2012). We used the same siRNA – which has a small window of thermodynamic asymmetry – in our strand-selective RNA cleavage assay to further test the effect that 5′ nucleotide preference may have on strand selection (**Figure 4.4 A**, siRNA-3). In support of the previous results (Betancur and Tomari 2012), Ago2 was able to cleave targets strand-selectively on its own using this duplex (**Figure 4.4 B, C**). Interestingly, in our hands Ago2 alone exhibited a higher degree of strand selection than either Dicer/dsRBP-containing complex in this case, indicating that these complexes may predetermine the guide

strand based on thermodynamics such that unfavorable 5′ nucleotides on the passenger strand do not have a stimulatory effect. Nonetheless, these findings indicate that an unfavorable 5′ nucleotide on the passenger strand facilitates strand-selectivity for human Ago2 in the absence of Dicer or dsRBPs.

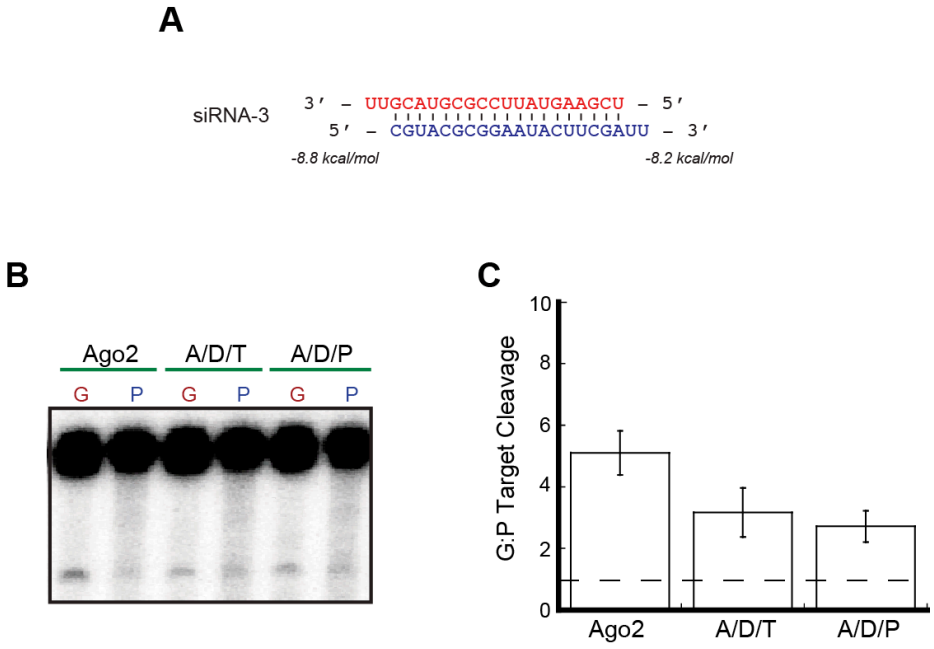

Figure 4.4: 5′ nucleotide identity affects strand selection by Ago2.
(A) Sequence and thermodynamic analysis of siRNA-3. Thermodynamics were calculated for the 4 terminal base pairs and 2 nt 3′ overhangs at each end of the duplex. The presumptive guide and passenger strands are shown in red and blue, respectively. **(B)** Ago2 alone cleaves target RNAs strand-selectively using siRNA-3. This level of strand selection is slightly attenuated by complexes containing Dicer and either TRBP or PACT. Lanes labeled "G" contain a target corresponding to the guide strand and lanes labeled "P" contain a target that is complementary to the passenger strand. **(C)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates.

4.3.5 RNA duplex terminal base pairing affects strand selection by Ago2

Prokaryotic Ago proteins bind to duplex nucleic acids such that the 5′ nucleotide of the guide strand is splayed apart from the corresponding base of the target – a property that appears to be conserved in human Ago2 (Wang et al. 2008b; Elkayam et al. 2012). We were interested to know if this binding geometry might allow Ago2 to preferentially select the guide strand in the absence of Dicer. To this end, we modified siRNA-1 by introducing a mismatch at the 5′ end of the guide strand (**Figure 4.5 A**, siRNA-4). With a similar RNA it was recently shown that endogenous mammalian Ago2 alone can be asymmetrically loaded as efficiently as when it is supplemented with recombinant Dicer (Betancur and Tomari 2012). In support of those data, we found that with this siRNA, Ago2 was again able to cleave its target in a strand-selective manner with no further enhancement in the presence of Dicer and TRBP or PACT (**Figure 4.5 B, C**). Thus, in addition to the specific identity of the 5′ nucleotide of the passenger strand, certain structural features of duplex ends likely provide a preferred binding geometry that allows Ago2 to selectively cleave the guide strand target in the absence of Dicer.

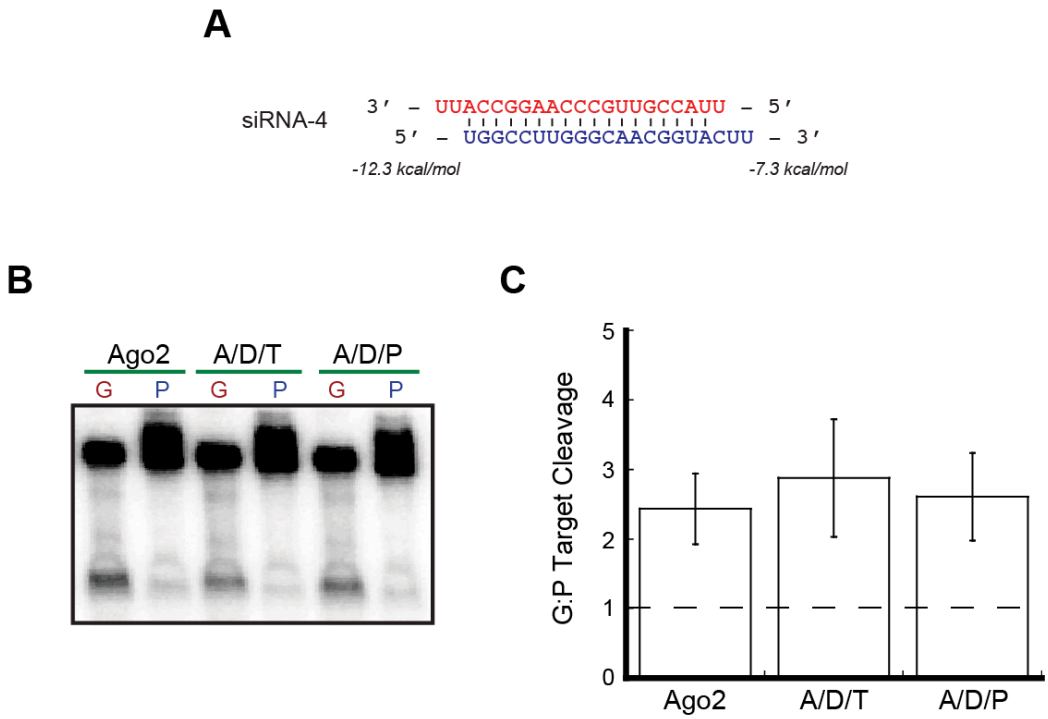

Figure 4.5: RNA duplex terminal base pairing affects strand selection by Ago2.
(A) Sequence and thermodynamic analysis of siRNA-4. Thermodynamics were calculated for the 4 terminal base pairs and 2 nt 3´ overhangs at each end of the duplex. The presumptive guide and passenger strands are shown in red and blue, respectively. **(B)** Ago2 alone cleaves targets strand-selectively using siRNA-4. Complexes containing Dicer and either TRBP or PACT do not enhance this strand selection. Lanes labeled "G" contain a target corresponding to the guide strand and lanes labeled "P" contain a target that is complementary to the passenger strand. **(C)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates.

4.3.6 Ago2 alone cleaves target RNAs strand-selectively using an miRNA

We were curious to know how internal duplex structural features such as those present in natural miRNAs might affect the ability of Ago2 alone to achieve strand-selective target cleavage. Such structural features are known to affect asymmetric RISC loading in *Drosophila* (Okamura et al. 2009; Czech et al. 2009). Additionally, miRNAs often have much wider windows of thermodynamic asymmetry compared with siRNAs, and it is possible that such a large thermodynamic difference between the duplex ends would be beneficial for strand-selective Ago2 loading in a manner similar to that of a terminal mismatch. We therefore created a model miRNA by introducing a central mismatch at position 10 of the guide strand of siRNA-1 as well as a seed mismatch at position 4 of the guide strand (**Figure 4.6 A**, miRNA-1). Target RNAs were again complementary to either strand to allow for their endonucleolytic cleavage. Ago2 alone exhibited strong strand selection with this miRNA (**Figure 4.6 B, C**), indicating that certain combinations of internal duplex structure and thermodynamics greatly enhance strand selection by human Ago2.

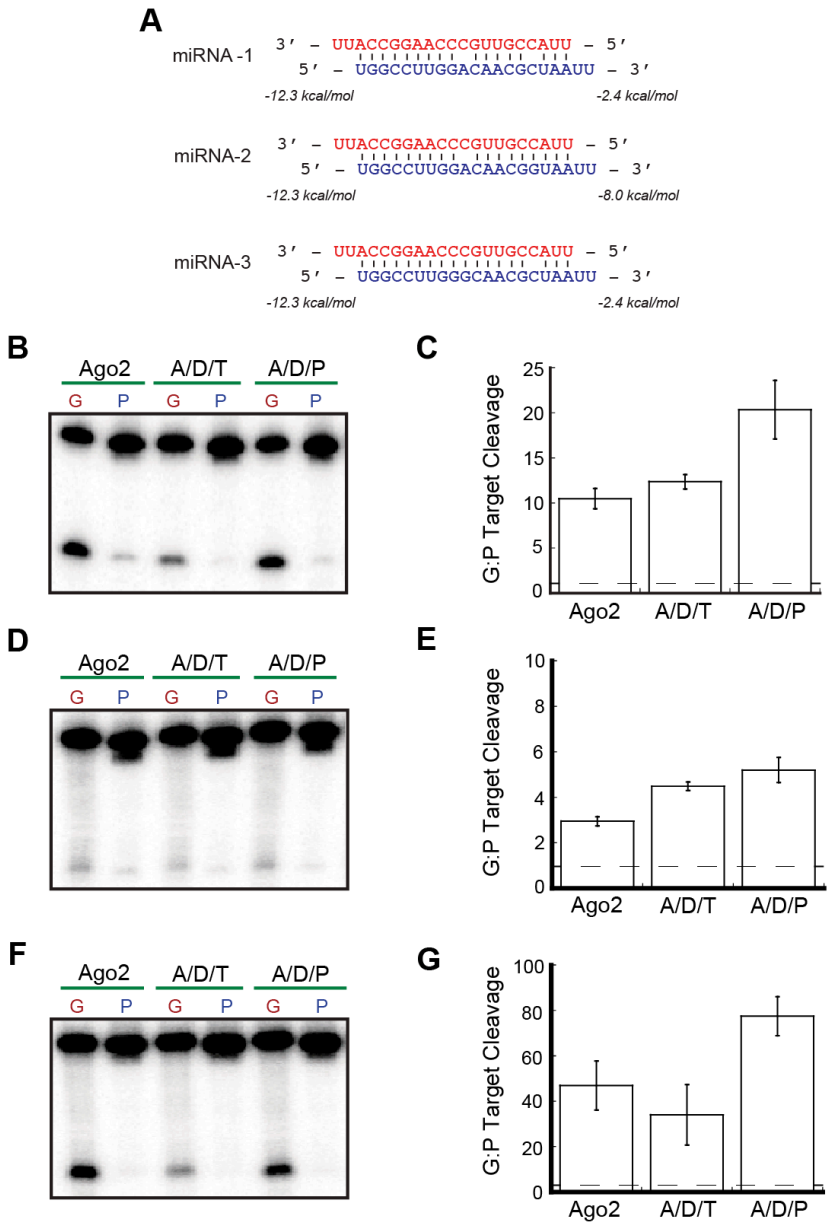

Figure 4.6: Complexes containing PACT but not TRBP enhance miRNA strand selection.
(A) Sequence and thermodynamic analysis of miRNA-1, miRNA-2, and miRNA-3. Thermodynamics were calculated for the 4 terminal base pairs and 2 nt 3′ overhangs at each end of the duplex. The presumptive guide and passenger strands are shown in red and blue, respectively. **(B)** Ago2 alone cleaves target RNAs strand-selectively using miRNA-1. This strand selection is not enhanced by the presence of Dicer and TRBP, but is strongly enhanced by the presence of Dicer and PACT. Lanes labeled "G" contain a target corresponding to the guide strand and lanes labeled "P" contain a target that is complementary to the passenger strand. **(C)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates. **(D)** Ago2 alone exhibits low levels of strand-selectivity using miRNA-2. Complexes containing Dicer and either TRBP or PACT enhance this strand-selectivity. **(E)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates. **(F)** Ago2 alone exhibits strong strand-selectivity using miRNA-3. This strand selection is not enhanced by the presence of Dicer and TRBP, but is strongly enhanced by the presence of Dicer and PACT. **(G)** Quantification of the ratio of target cleavage mediated by the guide strand compared to that mediated by the passenger strand. Data shown are means (+/- SD) from three separate experimental replicates.

4.3.7 miRNA strand selection is affected to differing degrees by TRBP and PACT

With the siRNA duplexes that we tested, Ago2/Dicer/dsRBP complexes generally exhibited similar levels of strand selection enhancement. We were surprised to find that with miRNA-1, strand selection by Ago2/Dicer/TRBP complexes was not enhanced above that observed for Ago2 alone, whereas a complex containing PACT enhanced strand-selective target cleavage (**Figure 4.6 B, C**). Additionally, the overall cleavage efficiency was substantially lower for the complex containing TRBP, indicating that such complexes may disfavor RNAs with internal duplex structures such as those found in most natural miRNAs (**Figure 4.6 B**). Together, these findings point to the possibility of delineating specific roles for TRBP and PACT in the fine-tuning of target specificity in RNAi.

To determine the specific elements of miRNA-1 that are important for this distinction (and for the ability of Ago2 alone to select the guide strand), we designed two miRNAs with a single mismatch either at position 10 (**Figure 4.6 A**, miRNA-2) or position 4 (**Figure 4.6 A**, miRNA-3) of the guide strand. Targets were designed to match each miRNA strand. The overall target cleavage activity for miRNA-2 was much lower than for miRNA-1 (**Figure 4.6 D**), which can be explained by previous reports that central miRNA mismatches promote Ago loading but not unwinding in both humans and *Drosophila* (Kawamata et al. 2009; Yoda et al. 2010). Ago2 alone was again able to cleave target RNAs strand-specifically in this case, indicating that internal duplex structure likely plays a role in dictating Ago2's ability to select the guide strand, as in *Drosophila* (Okamura et al. 2009). Interestingly, both Ago2/Dicer/dsRBP complexes enhanced Ago2's basal strand selection to similar degrees with this miRNA, indicating that the mismatch at position 10 is not responsible for the disproportionate enhancement of strand selection that was noted for the Ago2/Dicer/PACT complex with miRNA-1 (**Figure 4.6 D, E**).

The overall activities of each protein when programmed with miRNA-3 were similar to those observed with miRNA-1 (**Figure 4.6 B, F**). Again in this case, Ago2 exhibited strong strand selection on its own, indicating that seed mismatches and/or high levels of duplex thermodynamic asymmetry influence strand selection in the absence of Dicer (**Figure 4.6 F, G**). The abilities of Ago2/Dicer/dsRBP complexes to enhance this strand selection was also similar for miRNA-3 and miRNA-1, such that TRBP did not enhance Ago2's strand-selectivity, whereas PACT strongly enhanced strand selection (**Figure 4.6 F, G**). Thus, complexes containing PACT are more sensitive to duplex structure in the seed region than those containing TRBP when it comes to strand selection. Furthermore, complexes containing TRBP are generally less equipped to tolerate internal mismatches such as those present in many natural miRNAs.

4.4 Discussion

This study has revealed the nuanced roles of the human RNAi machinery in strand-selective target cleavage and lays the groundwork for the further delineation of those roles. Our results indicate that recombinant human Ago2 expressed in insect cells is fully competent to bind duplex siRNAs and miRNAs and use them as guides for target cleavage. This finding is contrary to previous reports that recombinant Ago2 alone is only active when programmed with ssRNAs (Rivas et al. 2005; Ye et al. 2011). A separate study showed that while bacterially expressed human Ago2 binds to duplex siRNAs and miRNAs, with siRNAs it could not release cleaved passenger strands (Wang et al. 2009a). It is unclear why such discrepancies in the field have arisen. One possibility is that purification of N-terminal GST-fusion proteins from bacterial sources may lead to a less active protein than a wild type protein from insect cells, given that

human Ago2's N-terminal domain has been implicated in duplex unwinding (Kwak and Tomari 2012). Similarly, we have found that a number of purification strategies for Ago2 lead to less active or inactive protein (unpublished observations). Another possible reason for these discrepancies is that some groups anneal duplex RNAs by adding an excess of passenger strand, which may compete with duplex RNAs for Ago2 binding, leading to the conclusion that the recombinant protein is not active with dsRNAs (Ye et al. 2011). Here we have gel purified each duplex RNA to ensure that no ssRNAs are present in our reactions (**Figure 4.1 D**).

Heat shock proteins are known to facilitate Ago2 loading (Pare et al. 2009; Johnston et al. 2010; Miyoshi et al. 2010; Iwasaki et al. 2010; Iki et al. 2010), but the fact that our reactions do not contain ATP and that the Hsp90 inhibitor 17-AAG does not inhibit Ago2 activity indicates that the activity we observe is not likely the result of chaperone activity. With respect to passenger strand removal, we cannot rule out the possibility that insect C3PO has co-purified with Ago2, although the protein appears to be pure by SDS-PAGE (**Figure 4.1 A**). It is likely that the addition of either C3PO or heat shock proteins to our assays would further enhance the activity of Ago2 as well as each Ago2/Dicer/dsRBP complex.

The results of our strand-selective target cleavage assays support and expand upon previous findings that in certain cases human Ago2 alone is able to cleave targets strand-specifically (Betancur and Tomari 2012). It was previously concluded that Dicer is dispensable for asymmetric RISC loading, but our expanded set of duplex RNAs has demonstrated that this conclusion is true only for certain types of duplex RNAs such as those with unfavorable 5′ nucleotides on the passenger strand or with terminal mismatches. Furthermore, even when Ago2 alone is capable of robust strand selection, Dicer and TRBP or PACT in some cases enhance that selection, as is the case with the miRNAs tested here. It should also be noted that despite the ability of Ago2 alone to bind certain duplex RNAs asymmetrically, Ago2/Dicer/dsRBP complexes were recently shown to be the primary machinery responsible for loading miRNAs onto Ago2 in mammals (Liu et al. 2012). In light of this finding and the data presented here, it seems clear that rather than Ago2 functioning exclusively in human strand selection, its binding preferences serve instead as a secondary RISC loading checkpoint that acts in concert with Dicer/dsRBP asymmetry sensors within RLCs to ensure proper strand selection.

The data presented here indicate the presence of a hierarchy of parameters that govern the necessity of human Dicer and its associated dsRBPs for strand selection enhancement that considers thermodynamic, 5′ nucleotide, and structural contributions (**Figure 4.7**). For perfectly duplexed siRNAs with a 5′ uridine on both strands, Dicer/dsRBP-containing RLCs are required for optimal strand selection. This remains true when unfavorable 5′ nucleotides are present on the guide strand. When present on the passenger strand, however, unfavorable 5′ nucleotides are able to drive strand selection by Ago2 alone. Similarly, terminal mismatches, which increase the thermodynamic asymmetry of the duplex and likely provide a favorable binding geometry for Ago2, allow for strand selection by that protein alone. Central and seed mismatches also allow robust strand-selective target cleavage by Ago2. In the case of miRNAs containing seed mismatches, Ago's strand-selectivity is significantly enhanced in the presence of Dicer and PACT but not TRBP, demonstrating that these dsRBPs play differing roles in strand selection that are likely governed by internal duplex structural elements. Many more natural miRNAs will need to be tested to fully define the duplex features that dictate these roles.

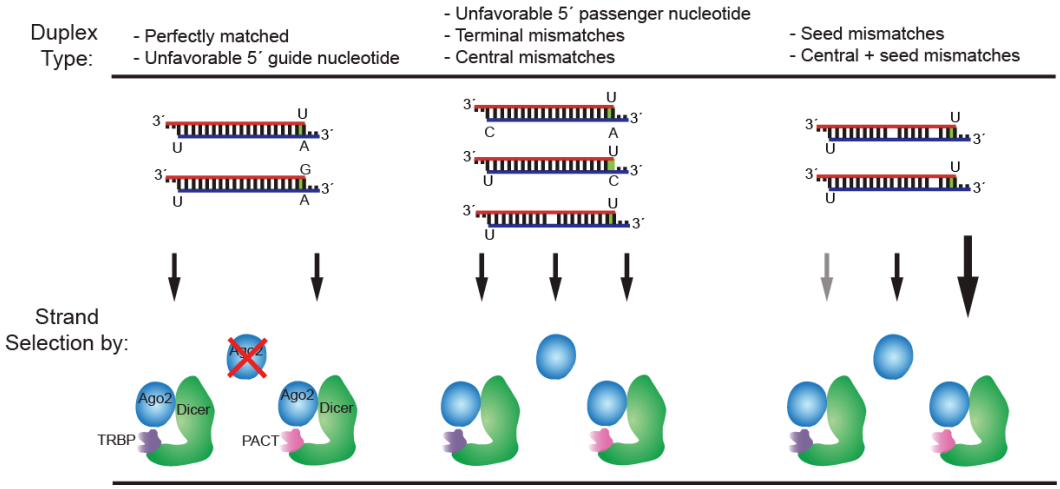

Figure 4.7: Parameters dictating the contributions of RLC components to strand selection.
Left: Only Ago2/Dicer/dsRBP complexes are able to efficiently direct strand selection with perfectly matched siRNAs containing identical 5´ nucleotides on both strands or a disfavored 5´ nucleotide on the guide strand. *Center:* Ago2 alone as well as Dicer/dsRBP-containing complexes are capable of strand selection with duplex RNAs harboring a disfavored 5´ nucleotide on the passenger strand, terminal mismatches, or central mismatches. *Right:* Ago2 and Ago2/Dicer/TRBP complexes exhibit robust strand selection with duplex RNAs containing seed mismatches or combined seed and central mismatches, although overall activity is diminished with a TRBP-containing complex (denoted by grey arrow). Strand selection with these RNAs is enhanced by Ago2/Dicer/PACT complexes (denoted by large arrow).

We also found that the presence of TRBP substantially attenuates overall RISC activity when programmed with a miRNA containing a seed sequence mismatch. This finding is reminiscent of small RNA sorting by the *Drosophila* RLC, where Dicer-2/R2D2 heterodimers disfavor miRNA loading onto Ago2, and a separate mechanism supports preferential loading of miRNAs onto Ago1 (Tomari et al. 2007). Our study indicates that in humans the Dicer/TRBP RLC subcomplex may be analogous to the Dicer-2/R2D2 subcomplex in its relative selection against miRNAs. Interestingly, while we found that TRBP-containing RLCs disfavored miRNA processing at the level of mRNA targeting, separate studies have found that PACT-containing RLCs disfavor siRNA processing at the level of dicing and that TRBP enhances pre-siRNA processing by Dicer (Lee et al. manuscript in preparation; (Chakravarthy et al. 2010)). These studies combined with our data provide a picture of human siRNA and miRNA pathway delineation wherein Ago2/Dicer/TRBP RLCs are optimized for the siRNA pathway, whereas Ago2/Dicer/PACT RLCs are optimized for the miRNA pathway.

Chapter 5

Summary

The principal goal of this work has been to elucidate the mechanisms of RISC loading in humans. In particular, a focus has been placed on the means by which the core human RNAi machinery achieves strand-selective RISC loading, as the precision of this portion of RNAi pathways is absolutely essential to the overall specificity of mRNA targeting. We have shown that human RISC Loading Complexes are central to this multi-checkpoint process. As a starting point, we determined the molecular architecture of a human RLC containing Ago2, Dicer, and TRBP. This structural information then informed our efforts to elucidate the roles of Dicer/dsRBP RLC subcomplexes in pre-orienting nascent siRNAs and miRNAs based on the thermodynamic properties of the duplex prior to RISC loading. Finally, we have shown that although it is possible that in some cases Ago2 may function in isolation, Dicer/dsRBP thermodynamic sensors are often essential for fine-tuning the process of strand-selective RISC loading in humans. Ago2 therefore likely acts as a secondary checkpoint within the context of each RLC to ensure optimal strand selection. Furthermore, we have opened the door to the possibility of delineating specific roles for different human RLCs with respect to the siRNA and miRNA pathways that are dependent on the identity of the RLC-associated dsRBP.

When this research commenced, little was known of the mechanisms of RISC loading in humans. TRBP had been suggested to play a role in this process, although the significance of that role was under debate. The role that PACT plays in RNAi pathways was still quite vague. One barrier to a clear understanding of the human RISC loading process was a lack of structural information regarding how the individual components of the RNAi machinery functionally interact in a way that allows for pre-orienting and loading nascent siRNAs and miRNAs onto Ago2. To this end, we determined the molecular architecture of an RLC containing Ago2, Dicer, and TRBP. This study yielded a view of a highly dynamic complex, and intriguingly, the dimensions of the complex seemed optimized for the transfer of a 21-23 nt duplex RNA between each component protein during RISC loading. Based on this structure and analogy to a biochemically characterized *Drosophila* complex, we formed a series of testable hypotheses regarding the mechanistic steps bridging substrate cleavage by Dicer and siRNA or miRNA binding by Ago2.

The lack of mechanistic information regarding the human RISC loading process also reflected the absence of a robust *in vitro* system for studying the individual contributions of the human RNAi components to this process. Using individually expressed proteins and reconstituted RISC Loading Complexes and subcomplexes, we developed a series of assays to determine how human RLCs orient Dicer product RNAs prior to RISC loading. As suggested by our EM reconstruction, these experiments revealed a novel RNA binding site at Dicer's helicase domain for product RNAs. We further showed that Dicer likely releases nascent siRNAs prior to rebinding them at this new site. Importantly, we were able to demonstrate that human Dicer/dsRBP heterodimers pre-orient thermodynamically asymmetric siRNAs such that Dicer binds to the less stable end of the duplex and TRBP or PACT binds to the more stable end. On their own, dsRBPs bound to either end of the duplexes indiscriminately, whereas Dicer did not bind at all. However, by tethering a dsRBP that is not involved in RNAi to Dicer's helicase domain, we were able to rescue thermodynamic sensing. This indicated that Dicer itself is likely the active sensor of siRNA thermodynamic asymmetry, whereas dsRBPs serve more of a passive role, holding the RNA in place and binding the duplex end that Dicer does not prefer.

At the time that this research was being conducted, one of many significant topics of debate in the RNAi field regarded the extent to which human RLCs are in fact *bona fide* RLCs. This question arose from the observation that human Ago2 is in some cases capable of binding siRNAs strand-specifically in the absence of Dicer. From these observations it was concluded that human Dicer (and thus the human RLC) is not involved in RISC loading. These conclusions, however, underestimated the complexity of the human RISC loading process. It was soon shown by another group that in wild type human cells, RLCs are the primary complexes involved in RISC loading. Despite this clear step forward, a question that remained was that since Dicer products are released prior to asymmetry sensing, it was possible that siRNAs might bypass the canonical Dicer/dsRBP asymmetry sensing machinery in humans and bind directly to Ago2. We felt that the available data were more consistent with the hypothesis that Ago2 itself serves as a secondary checkpoint within RLCs to ensure proper strand selection. To test this, we developed an assay to test for strand-selective target cleavage and found that in certain cases where 5′ uridines were present on both duplex strands or when an unfavorable 5′ nucleotide was present on the guide strand, Ago2 alone was insufficient for accurate guide strand selection. RLCs were essential to achieving proper target specificity in these cases. In other cases where an unfavorable 5′ nucleotide was present on the passenger strand or the duplex contained a terminal mismatch, Ago2 alone was as capable of strand selection as either RLC. This was also true in the case of miRNAs, but RLCs containing PACT achieved much higher levels of strand-selectivity than did Ago2 alone. Our findings provide evidence for a complex, multi-checkpoint strand selection system within each RLC wherein the individual contributions of each protein are dictated by specific RNA duplex characteristics such as end thermodynamics, 5′ nucleotide identity, and duplex structure. Interestingly, in the process of studying strand-selective target cleavage, we also found that mRNA targeting by miRNAs is strongly disfavored by RLCs containing TRBP as opposed to PACT. This finding opens the door to further delineation of the specific roles of TRBP and PACT in the siRNA and miRNA pathways. Overall, the body of work presented here represents a significant advance in our mechanistic understanding of the RISC loading process in human RNA interference.

Bibliography

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. *Molecular Biology of the Cell, Fourth Edition*. 4(null) ed. Garland Science.

Aouadi M, Tesz GJ, Nicoloro SM, Wang M, Chouinard M, Soto E, Ostroff GR, Czech MP. 2009. Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation. *Nature* **458**: 1180–1184.

Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* **455**: 64–71.

Baker ML, Zhang J, Ludtke SJ, Chiu W. 2010. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat Protoc* **5**: 1697–1708.

Barr I, Smith AT, Senturia R, Chen Y, Scheidemantle BD, Burstyn JN, Guo F. 2011. DiGeorge critical region 8 (DGCR8) is a double-cysteine-ligated heme protein. *J Biol Chem* **286**: 16716–16725.

Bazzini AA, Lee MT, Giraldez AJ. 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**: 233–237.

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Gene Dev* **20**: 1885–1898.

Beltrao P, Albanèse V, Kenner LR, Swaney DL, Burlingame A, Villén J, Lim WA, Fraser JS, Frydman J, Krogan NJ. 2012. Systematic functional prioritization of protein posttranslational modifications. *Cell* **150**: 413–425.

Bennett RL, Blalock WL, May WS. 2004. Serine 18 phosphorylation of RAX, the PKR activator, is required for PKR activation and consequent translation inhibition. *J Biol Chem* **279**: 42687–42693.

Berezikov E, Chung W-J, Willis J, Cuppen E, Lai EC. 2007. Mammalian mirtron genes. *Mol Cell* **28**: 328–336.

Berg JM, Tymoczko JL, Stryer L. 2010. *Biochemistry*. Seventh Edition. W. H. Freeman.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366.

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. 2003. Dicer is essential for mouse development. *Nat Genet* **35**: 215–217.

Betancur JG, Tomari Y. 2012. Dicer is dispensable for asymmetric RISC loading in mammals. *Rna* **18**: 24–30.

Béthune J, Artus-Revel CG, Filipowicz W. 2012. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *Embo Rep* **13**: 716–723.

Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**: 1111–1124.

Blaszczyk J, Tropea JE, Bubunenko M, Routzahn KM, Waugh DS, Court DL, Ji X. 2001. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure* **9**: 1225–1236.

Borchert GM, Lanier W, Davidson BL. 2006. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**: 1097–1101.

Borer PN, Dengler B, Tinoco I, Uhlenbeck OC. 1974. Stability of ribonucleic acid double-stranded helices. *J Mol Biol* **86**: 843–853.

Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**: 25–36.

Busso D, Delagoutte-Busso B, Moras D. 2005. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli. *Anal Biochem* **343**: 313–321.

Bycroft M, Grünert S, Murzin AG, Proctor M, St Johnston D. 1995. NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *Embo J* **14**: 3563–3571.

Cai X, Hagedorn CH, Cullen BR. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**: 1957–1966.

Caudy AA, Myers M, Hannon GJ, Hammond SM. 2002. Fragile X-related protein and VIG associate with the RNA interference machinery. *Gene Dev* **16**: 2491–2496.

Cenik ES, Fukunaga R, Lu G, Dutcher R, Wang Y, Tanaka Hall TM, Zamore PD. 2011. Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol Cell* **42**: 172–184.

Chakravarthy S, Sternberg SH, Kellenberger CA, Doudna JA. 2010. Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J Mol Biol* **404**: 392–402.

Chang YF, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* **76**: 51–74.

Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, Hunter MS, Schulz J, DePonte DP, Weierstall U, et al. 2011. Femtosecond X-ray protein nanocrystallography. *Nature* **470**: 73–77.

Cheloufi S, Santos Dos CO, Chong MMW, Hannon GJ. 2010. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**: 584–589.

Chen C-YA, Zheng D, Xia Z, Shyu A-B. 2009. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* **16**: 1160–1166.

Chendrimada TP, Finn KJ, Ji X, Baillat D, Gregory RI, Liebhaber SA, Pasquinelli AE, Shiekhattar R. 2007. MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447**: 823–828.

Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**: 740–744.

Cheng Z, Muhlrad D, Lim MK, Parker R, Song H. 2007. Structural and functional insights into the human Upf1 helicase core. *Embo J* **26**: 253–264.

Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND, et al. 2010. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328**: 1694–1698.

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

Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ. 2009. Hierarchical rules for Argonaute loading in Drosophila. *Mol Cell* **36**: 445–456.

Daniels SM, Melendez-Peña CE, Scarborough RJ, Daher A, Christensen HS, Far El M, Purcell DFJ, Lainé S, Gatignol A. 2009. Characterization of the TRBP domain required for dicer interaction and function in RNA interference. *BMC Mol Biol* **10**: 38.

Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231–235.

Dephoure N, Zhou C, Villén J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP. 2008. A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences* **105**: 10762–10767.

Dietz TM, Koch TH. 1987. Photochemical coupling of 5-bromouracil to tryptophan, tyrosine and histidine, peptide-like derivatives in aqueous fluid solution. *Photochem Photobiol* **46**: 971–978.

Djuranovic S, Nahvi A, Green R. 2012. miRNA-mediated gene silencing by translational

repression followed by mRNA deadenylation and decay. *Science* **336**: 237–240.

Dlakić M. 2006. DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics* **22**: 2711–2714.

Doench JG, Petersen CP, Sharp PA. 2003. siRNAs can function as miRNAs. *Gene Dev* **17**: 438–442.

Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G. 2012. microRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res* **40**: 9850–9862.

Eamens AL, Smith NA, Curtin SJ, Wang M-B, Waterhouse PM. 2009. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *Rna* **15**: 2219–2235.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001a. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494–498.

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

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. 2001c. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J* **20**: 6877–6888.

Elkayam E, Kuhn C-D, Tocilj A, Haase AD, Greene EM, Hannon GJ, Joshua-Tor L. 2012. The structure of human argonaute-2 in complex with miR-20a. *Cell* **150**: 100–110.

Eulalio A, Huntzinger E, Izaurralde E. 2008. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* **15**: 346–353.

Eulalio A, Huntzinger E, Nishihara T, Rehwinkel J, Fauser M, Izaurralde E. 2009. Deadenylation is a widespread effect of miRNA regulation. *Rna* **15**: 21–32.

Fabian MR, Cieplak MK, Frank F, Morita M, Green J, Srikumar T, Nagar B, Yamamoto T, Raught B, Duchaine TF, et al. 2011. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat Struct Mol Biol* **18**: 1211–1217.

Fabian MR, Mathonnet G, Sundermeier T, Mathys H, Zipprich JT, Svitkin YV, Rivas F, Jinek M, Wohlschlegel J, Doudna JA, et al. 2009. Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* **35**: 868–880.

Faller M, Matsunaga M, Yin S, Loo JA, Guo F. 2007. Heme is involved in microRNA processing. *Nat Struct Mol Biol* **14**: 23–29.

Faller M, Toso D, Matsunaga M, Atanasov I, Senturia R, Chen Y, Zhou ZH, Guo F. 2010. DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *Rna* **16**: 1570–1583.

Fernandez-Valverde SL, Taft RJ, Mattick JS. 2010. Dynamic isomiR regulation in Drosophila development. *Rna* **16**: 1881–1888.

Fire A, Albertson D, Harrison SW, Moerman DG. 1991. Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. *Development* **113**: 503–514.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806–811.

Förstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. 2007. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**: 287–297.

Frank F, Sonenberg N, Nagar B. 2010. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**: 818–822.

Frank J. 1996. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. 1st ed. Academic Press.

Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**: 190–199.

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *P Natl Acad Sci USA* **83**: 9373–9377.

Fukunaga R, Han BW, Hung J-H, Xu J, Weng Z, Zamore PD. 2012. Dicer Partner Proteins Tune the Length of Mature miRNAs in Flies and Mammals. *Cell* **151**: 533–546.

Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X. 2006. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* **124**: 355–366.

Gao H, Valle M, Ehrenberg M, Frank J. 2004. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J Struct Biol* **147**: 283–290.

Gatignol A, Buckler-White A, Berkhout B, Jeang KT. 1991. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science* **251**: 1597–1600.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW, Zapp ML, Weng Z, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**: 1077–1081.

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**: 833–838.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**: 75–79.

Gredell JA, Dittmer MJ, Wu M, Chan C, Walton SP. 2010. Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry-Us* **49**: 3148–3155.

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**: 631–640.

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

Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91–105.

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835–840.

Guo S, Kemphues KJ. 1995. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**: 611–620.

Haase AD, Jaskiewicz L, Zhang H, Lainé S, Sack R, Gatignol A, Filipowicz W. 2005. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *Embo Rep* **6**: 961–967.

Haley B, Zamore PD. 2004. Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol* **11**: 599–606.

Hall RJ, Siridechadilok B, Nogales E. 2007. Cross-correlation of common lines: a novel approach for single-particle reconstruction of a structure containing a flexible domain. *J Struct Biol* **159**: 474–482.

Hamilton AJ, Baulcombe DC. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950–952.

Hammond SM, Bernstein E, Beach D, Hannon GJ. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**: 293–296.

Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. 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 VN. 2004. The Drosha-DGCR8 complex in

primary microRNA processing. *Gene Dev* **18**: 3016–3027.

Han J, Lee Y, Yeom K-H, Nam J-W, Heo I, Rhee J-K, Sohn SY, Cho Y, Zhang B-T, Kim VN. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**: 887–901.

Harland R, Weintraub H. 1985. Translation of mRNA injected into Xenopus oocytes is specifically inhibited by antisense RNA. *J Cell Biol* **101**: 1094–1099.

Hatfield SD, Shcherbata HR, Fischer KA, Nakahara K, Carthew RW, Ruohola-Baker H. 2005. Stem cell division is regulated by the microRNA pathway. *Nature* **435**: 974–978.

Herzog F, Primorac I, Dube P, Lenart P, Sander B, Mechtler K, Stark H, Peters JM. 2009. Structure of the anaphase-promoting complex/cyclosome interacting with a mitotic checkpoint complex. *Science Signalling* **323**: 1477.

Höck J, Weinmann L, Ender C, Rüdel S, Kremmer E, Raabe M, Urlaub H, Meister G. 2007. Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. *Embo Rep* **8**: 1052–1060.

Högbom M, Collins R, van den Berg S, Jenvert R-M, Karlberg T, Kotenyova T, Flores A, Karlsson Hedestam GB, Schiavone LH. 2007. Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. *J Mol Biol* **372**: 150–159.

Hu HY, Yan Z, Xu Y, Hu H, Menzel C, Zhou YH, Chen W, Khaitovich P. 2009. Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* **10**: 413.

Humphreys DT, Westman BJ, Martin DIK, Preiss T. 2005. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *P Natl Acad Sci USA* **102**: 16961–16966.

Huntzinger E, Braun JE, Heimstädt S, Zekri L, Izaurralde E. 2010. Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. *Embo J* **29**: 4146–4160.

Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.

Hutvagner G, Zamore PD. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056–2060.

Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M. 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* **39**: 282–291.

Ishigaki Y, Li X, Serin G, Maquat LE. 2001. Evidence for a pioneer round of mRNA translation:

mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**: 607–617.

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, Maquat LE. 2008. Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* **133**: 314–327.

Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y. 2010. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39**: 292–299.

Izant JG, Weintraub H. 1985. Constitutive and conditional suppression of exogenous and endogenous genes by anti-sense RNA. *Science* **229**: 345–352.

Izant JG, Weintraub H. 1984. Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. *Cell* **36**: 1007.

Jackson RJ, Hellen CUT, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Bio* **11**: 113–127.

Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, D'Orso I, Fernandes J, Jang G, Frankel AD, et al. 2011. Purification and characterization of HIV-human protein complexes. *Methods* **53**: 13–19.

Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. 2005. Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Gene Dev* **19**: 1674–1679.

Jinek M, Doudna JA. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**: 405–412.

Jinek M, Fabian MR, Coyle SM, Sonenberg N, Doudna JA. 2010. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nat Struct Mol Biol* **17**: 238–240.

Johnston M, Geoffroy M-C, Sobala A, Hay R, Hutvagner G. 2010. HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Mol Biol Cell* **21**: 1462–1469.

Kadlec J, Guilligay D, Ravelli RB, Cusack S. 2006. Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *Rna* **12**: 1817–1824.

Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, et al. 2011. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* **471**: 325–330.

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Gene Dev* **19**: 489–501.

Kastner B, Fischer N, Golas MM, Sander B, Dube P, Boehringer D, Hartmuth K, Deckert J, Hauer F, Wolf E, et al. 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* **5**: 53–55.

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

Kharrat A, Macias MJ, Gibson TJ, Nilges M, Pastore A. 1995. Structure of the dsRNA binding domain of E. coli RNase III. *Embo J* **14**: 3572–3584.

Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209–216.

Kim K, Lee YS, Carthew RW. 2007. Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *Rna* **13**: 22–29.

Kok KH, Ng M-HJ, Ching Y-P, Jin D-Y. 2007. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J Biol Chem* **282**: 17649–17657.

Kostek SA, Grob P, De Carlo S, Lipscomb JS, Garczarek F, Nogales E. 2006. Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* **14**: 1691–1700.

Kowalinski E, Lunardi T, McCarthy AA, Louber J, Brunel J, Grigorov B, Gerlier D, Cusack S. 2011. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**: 423–435.

Kozak M. 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**: 1109–1123.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152–7.

Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ. 2004. Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J Biol Chem* **279**: 42230–42239.

Kwak PB, Tomari Y. 2012. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat Struct Mol Biol* **19**: 145–151.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–858.

Landthaler M, Gaidatzis D, Rothballer A, Chen PY, Soll SJ, Dinic L, Ojo T, Hafner M, Zavolan M, Tuschl T. 2008. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *Rna* **14**: 2580–2596.

Landthaler M, Yalcin A, Tuschl T. 2004. The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* **14**: 2162–2167.

Laraki G, Clerzius G, Daher A, Melendez-Peña C, Daniels S, Gatignol A. 2008. Interactions between the double-stranded RNA-binding proteins TRBP and PACT define the Medipal domain that mediates protein-protein interactions. *Rna Biol* **5**: 92–103.

Lau NC, Lim, L. P., Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858–862.

Lau P-W, Guiley KZ, De N, Potter CS, Carragher B, Macrae IJ. 2012. The molecular architecture of human Dicer. *Nat Struct Mol Biol* **19**: 436–440.

Lau P-W, Potter CS, Carragher B, Macrae IJ. 2009. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure* **17**: 1326–1332.

Lee HY, Doudna JA. 2012. TRBP alters human precursor microRNA processing in vitro. *Rna* **18**: 2012–2019.

Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, Wang K. 2010. Complexity of the microRNA repertoire revealed by next-generation sequencing. *Rna* **16**: 2170–2180.

Lee RC, Ambros V. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**: 862–864.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–854.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415–419.

Lee Y, Hur I, Park S-Y, Kim Y-K, Suh MR, Kim VN. 2006. The role of PACT in the RNA silencing pathway. *Embo J* **25**: 522–532.

Lee Y, Jeon K, Lee J-T, Kim S, Kim VN. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* **21**: 4663–4670.

Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN. 2004a. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**: 4051–4060.

Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004b. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**: 69–82.

Leschziner AE, Nogales E. 2007. Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions. *Annu Rev Biophys Biomol Struct* **36**: 43–62.

Leung AKL, Vyas S, Rood JE, Bhutkar A, Sharp PA, Chang P. 2011. Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. *Mol Cell* **42**: 489–499.

Leuschner PJF, Ameres SL, Kueng S, Martinez J. 2006. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *Embo Rep* **7**: 314–320.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20.

Lewis BP, Shih I-H, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769–773.

Lingel A, Simon B, Izaurralde E, Sattler M. 2004. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11**: 576–577.

Lingel A, Simon B, Izaurralde E, Sattler M. 2003. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**: 465–469.

Liu F, Putnam A, Jankowsky E. 2008. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proceedings of the National Academy of Sciences* **105**: 20209–20214.

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song J-J, Hammond SM, Joshua-Tor L, Hannon GJ. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**: 1437–1441.

Liu J, Rivas FV, Wohlschlegel J, Yates JR, Parker R, Hannon GJ. 2005a. A role for the P-body component GW182 in microRNA function. *Nature cell biology* **7**: 1261–1266.

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R. 2005b. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nature cell biology* **7**: 719–723.

Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, 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, Liu Q. 2006. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *Rna* **12**: 1514–1520.

Liu X, Jin D-Y, McManus MT, Mourelatos Z. 2012. Precursor microRNA-programmed silencing complex assembly pathways in mammals. *Mol Cell* **46**: 507–517.

Liu Y, Tan H, Tian H, Liang C, Chen S, Liu Q. 2011. Autoantigen La promotes efficient RNAi, antiviral response, and transposon silencing by facilitating multiple-turnover RISC catalysis. *Mol Cell* **44**: 502–508.

Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q. 2009. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* **325**: 750–753.

Lu Z, Liu M, Stribinskis V, Klinge CM, Ramos KS, Colburn NH, Li Y. 2008. MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene. *Oncogene* **27**: 4373–4379.

Ludtke SJ, Baldwin PR, Chiu W. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**: 82–97.

Lykke-Andersen J, Shu MD, Steitz JA. 2000. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**: 1121–1131.

Ma E, Macrae IJ, Kirsch JF, Doudna JA. 2008. Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol* **380**: 237–243.

Ma J-B, Ye K, Patel DJ. 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**: 318–322.

Ma J-B, Yuan Y-R, Meister G, Pei Y, Tuschl T, Patel DJ. 2005. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**: 666–670.

Macrae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. 2008. In vitro reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences* **105**: 512–517.

Macrae IJ, Zhou K, Doudna JA. 2007. Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol* **14**: 934–940.

Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. 2006. Structural basis for double-stranded RNA processing by Dicer. *Science* **311**: 195–198.

Maniataki E, Mourelatos Z. 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Gene Dev* **19**: 2979–2990.

Mann M, Jensen ON. 2003. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**: 255–261.

Maroney PA, Yu Y, Fisher J, Nilsen TW. 2006. Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol* **13**: 1102–1107.

Marques JT, Kim K, Wu P-H, Alleyne TM, Jafari N, Carthew RW. 2010. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. *Nat Struct Mol Biol* **17**: 24–30.

Martinez J, Patkaniowska A, Urlaub H, Lührmann R, Tuschl T. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**: 563–574.

Martinez J, Tuschl T. 2004. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Gene Dev* **18**: 975–980.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911–940.

Mathonnet G, Fabian MR, Svitkin YV, Parsyan A, Huck L, Murata T, Biffo S, Merrick WC, Darzynkiewicz E, Pillai RS, et al. 2007. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* **317**: 1764–1767.

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**: 607–620.

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**: 185–197.

Meister G, Landthaler M, Peters L, Chen PY, Urlaub H, Lührmann R, Tuschl T. 2005. Identification of novel argonaute-associated proteins. *Curr Biol* **15**: 2149–2155.

Melo SA, Ropero S, Moutinho C, Aaltonen LA, Yamamoto H, Calin GA, Rossi S, Fernandez AF, Carneiro F, Oliveira C, et al. 2009. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet* **41**: 365–370.

Melton C, Judson RL, Blelloch R. 2010. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**: 621–626.

Melton DA. 1985. Injected anti-sense RNAs specifically block messenger RNA translation in vivo. *P Natl Acad Sci USA* **82**: 144–148.

Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* **12**: 7035–7056.

Miyoshi T, Takeuchi A, Siomi H, Siomi MC. 2010. A direct role for Hsp90 in pre-RISC formation in Drosophila. *Nat Struct Mol Biol* **17**: 1024–1026.

Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Gene Dev* **16**: 720–728.

Murchison EP, Partridge JF, Tam OH, Cheloufi S, Hannon GJ. 2005. Characterization of Dicer-deficient murine embryonic stem cells. *P Natl Acad Sci USA* **102**: 12135–12140.

Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. 2012. Structure of yeast Argonaute with guide RNA. *Nature* **486**: 368–374.

Nanduri S, Carpick BW, Yang Y, Williams BR, Qin J. 1998. Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *Embo J* **17**: 5458–5465.

Napoli C, Lemieux C, Jorgensen R. 1990. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* **2**: 279–289.

Noland CL, Ma E, Doudna JA. 2011. siRNA repositioning for guide strand selection by human Dicer complexes. *Mol Cell* **43**: 110–121.

Norris CL, Meisenheimer PL, Koch TH. 1996. Mechanistic studies of the 5-iodouracil chromophore relevant to its use in nucleoprotein photo-cross-linking. *J Am Chem Soc* **118**: 5796–5803.

Nottrott S, Simard MJ, Richter JD. 2006. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* **13**: 1108–1114.

O'Toole AS, Miller S, Serra MJ. 2005. Stability of 3“ double nucleotide overhangs that model the 3” ends of siRNA. *Rna* **11**: 512–516.

Okada C, Yamashita E, Lee SJ, Shibata S, Katahira J, Nakagawa A, Yoneda Y, Tsukihara T. 2009. A high-resolution structure of the pre-microRNA nuclear export machinery. *Science* **326**: 1275–1279.

Okamura K, Balla S, Martin R, Liu N, Lai EC. 2008. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. *Nat Struct Mol Biol* **15**: 581–590.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* **130**: 89–100.

Okamura K, Liu N, Lai EC. 2009. Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. *Mol Cell* **36**: 431–444.

Palliser D, Chowdhury D, Wang Q-Y, Lee SJ, Bronson RT, Knipe DM, Lieberman J. 2006. An siRNA-based microbicide protects mice from lethal herpes simplex virus 2 infection. *Nature* **439**: 89–94.

Pare JM, Tahbaz N, López-Orozco J, LaPointe P, Lasko P, Hobman TC. 2009. Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies. *Mol Biol Cell* **20**: 3273–3284.

Park H, Davies MV, Langland JO, Chang HW, Nam YS, Tartaglia J, Paoletti E, Jacobs BL, Kaufman RJ, Venkatesan S. 1994. TAR RNA-binding protein is an inhibitor of the interferon-induced protein kinase PKR. *P Natl Acad Sci USA* **91**: 4713–4717.

Park J-E, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. 2011. Dicer

recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**: 201–205.

Parker GS, Maity TS, Bass BL. 2008. dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *J Mol Biol* **384**: 967–979.

Parker JS, Roe SM, Barford D. 2004. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *Embo J* **23**: 4727–4737.

Parker JS, Roe SM, Barford D. 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**: 663–666.

Paroo Z, Ye X, Chen S, Liu Q. 2009. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell* **139**: 112–122.

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P, et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**: 86–89.

Pastor F, Kolonias D, Giangrande PH, Gilboa E. 2010. Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. *Nature* **465**: 227–230.

Patel RC, Sen GC. 1998. PACT, a protein activator of the interferon-induced protein kinase, PKR. *Embo J* **17**: 4379–4390.

Penczek PA, Grassucci RA, Frank J. 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* **53**: 251.

Peters GA, Li S, Sen GC. 2006. Phosphorylation of specific serine residues in the PKR activation domain of PACT is essential for its ability to mediate apoptosis. *J Biol Chem* **281**: 35129–35136.

Petersen CP, Bordeleau M-E, Pelletier J, Sharp PA. 2006. Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* **21**: 533–542.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612.

Pham JW, Pellino JL, Lee YS, Carthew RW, Sontheimer EJ. 2004. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117**: 83–94.

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**: 1573–1576.

Preall JB, He Z, Gorra JM, Sontheimer EJ. 2006. Short interfering RNA strand selection is independent of dsRNA processing polarity during RNAi in Drosophila. *Curr Biol* **16**: 530–

535.

Provost P, Dishart D, Doucet J, Frendewey D, Samuelsson B, Rådmark O. 2002. Ribonuclease activity and RNA binding of recombinant human Dicer. *Embo J* **21**: 5864–5874.

Qi HH, Ongusaha PP, Myllyharju J, Cheng D, Pakkanen O, Shi Y, Lee SW, Peng J, Shi Y. 2008. Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature* **455**: 421–424.

Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA. 2010. Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *Rna* **16**: 474–481.

Radermacher M, Wagenknecht T, Verschoor A, Frank J. 1987. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc* **146**: 113–136.

Ramos A, Grünert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G. 2000. RNA recognition by a Staufen double-stranded RNA-binding domain. *Embo J* **19**: 997–1009.

Rand TA, Petersen S, Du F, Wang X. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**: 621–629.

Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *Rna* **11**: 1640–1647.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901–906.

Rivas FV, Tolia NH, Song J-J, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L. 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**: 340–349.

Robb GB, Rana TM. 2007. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol Cell* **26**: 523–537.

Robertson HD, Webster RE, Zinder ND. 1968. Purification and properties of ribonuclease III from Escherichia coli. *J Biol Chem* **243**: 82–91.

Rose SD, Kim D-H, Amarzguioui M, Heidel JD, Collingwood MA, Davis ME, Rossi JJ, Behlke MA. 2005. Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res* **33**: 4140–4156.

Rosenberg UB, Preiss A, Seifert E, Jäckle H, Knipple DC. 1985. Production of phenocopies by Krüppel antisense RNA injection into Drosophila embryos. *Nature* **313**: 703–706.

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127**: 1193–1207.

Ruby JG, Jan CH, Bartel DP. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**: 83–86.

Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, Zhang H, Zha X-M, Polakiewicz RD, Comb MJ. 2005. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* **23**: 94–101.

Rüdel S, Wang Y, Lenobel R, Körner R, Hsiao H-H, Urlaub H, Patel D, Meister G. 2011. Phosphorylation of human Argonaute proteins affects small RNA binding. *Nucleic Acids Res* **39**: 2330–2343.

Rybak A, Fuchs H, Hadian K, Smirnova L, Wulczyn EA, Michel G, Nitsch R, Krappmann D, Wulczyn FG. 2009. The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. *Nature cell biology* **11**: 1411–1420.

Ryter JM, Schultz SC. 1998. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *Embo J* **17**: 7505–7513.

Saito K, Ishizuka A, Siomi H, Siomi MC. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *Plos Biol* **3**: e235.

Saleh M-C, Tassetto M, van Rij RP, Goic B, Gausson V, Berry B, Jacquier C, Antoniewski C, Andino R. 2009. Antiviral immunity in Drosophila requires systemic RNA interference spread. *Nature* **458**: 346–350.

Sano M, Sierant M, Miyagishi M, Nakanishi M, Takagi Y, Sutou S. 2008. Effect of asymmetric terminal structures of short RNA duplexes on the RNA interference activity and strand selection. *Nucleic Acids Res* **36**: 5812–5821.

Sasaki T, Shimizu N. 2007. Evolutionary conservation of a unique amino acid sequence in human DICER protein essential for binding to Argonaute family proteins. *Gene* **396**: 312–320.

Scheres SHW, Gao H, Valle M, Herman GT, Eggermont PPB, Frank J, Carazo J-M. 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat Methods* **4**: 27–29.

Schirle NT, Macrae IJ. 2012. The crystal structure of human Argonaute2. *Science* **336**: 1037–1040.

Schumacher JM, Artzt K, Braun RE. 1998. Spermatid perinuclear ribonucleic acid-binding protein binds microtubules in vitro and associates with abnormal manchettes in vivo in mice. *Biol Reprod* **59**: 69–76.

Schumacher JM, Lee K, Edelhoff S, Braun RE. 1995. Spnr, a murine RNA-binding protein that is localized to cytoplasmic microtubules. *J Cell Biol* **129**: 1023–1032.

Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199–208.

Schwarz DS, Tomari Y, Zamore PD. 2004. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Curr Biol* **14**: 787–791.

Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**: 58–63.

Sen GL, Blau HM. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature cell biology* **7**: 633–636.

Senturia R, Faller M, Yin S, Loo JA, Cascio D, Sawaya MR, Hwang D, Clubb RT, Guo F. 2010. Structure of the dimerization domain of DiGeorge critical region 8. *Protein Sci* **19**: 1354–1365.

Seo J, Lee K-J. 2004. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J Biochem Mol Biol* **37**: 35–44.

Si M-L, Zhu S, Wu H, Lu Z, Wu F, Mo Y-Y. 2007. miR-21-mediated tumor growth. *Oncogene* **26**: 2799–2803.

Singh M, Castillo D, Patel CV, Patel RC. 2011. Stress-induced phosphorylation of PACT reduces its interaction with TRBP and leads to PKR activation. *Biochemistry-Us* **50**: 4550–4560.

Siridechadilok B, Fraser CS, Hall RJ, Doudna JA, Nogales E. 2005. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310**: 1513–1515.

Sohn SY, Bae WJ, Kim JJ, Yeom K-H, Kim VN, Cho Y. 2007. Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* **14**: 847–853.

Soifer HS, Sano M, Sakurai K, Chomchan P, Saetrom P, Sherman MA, Collingwood MA, Behlke MA, Rossi JJ. 2008. A role for the Dicer helicase domain in the processing of thermodynamically unstable hairpin RNAs. *Nucleic Acids Res* **36**: 6511–6522.

Song J-J, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434–1437.

Song R, Hennig GW, Wu Q, Jose C, Zheng H, Yan W. 2011. Male germ cells express abundant endogenous siRNAs. *Proceedings of the National Academy of Sciences* **108**: 13159–13164.

St Johnston D, Brown NH, Gall JG, Jantsch M. 1992. A conserved double-stranded RNA-binding domain. *P Natl Acad Sci USA* **89**: 10979–10983.

Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB, et al. 2010. The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**: 225–237.

Suk K, Choi J, Suzuki Y, Ozturk SB, Mellor JC, Wong KH, MacKay JL, Gregory RI, Roth FP. 2011. Reconstitution of human RNA interference in budding yeast. *Nucleic Acids Res* **39**: e43.

Sun C, Todorovic A, Querol-Audí J, Bai Y, Villa N, Snyder M, Ashchyan J, Lewis CS, Hartland A, Gradia S, et al. 2011. Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proceedings of the National Academy of Sciences* **108**: 20473–20478.

Tahbaz N, Kolb FA, Zhang H, Jaronczyk K, Filipowicz W, Hobman TC. 2004. Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. *Embo Rep* **5**: 189–194.

Takeshita D, Zenno S, Lee WC, 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**: 106–120.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534–538.

Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**: 1124–1128.

Thermann R, Hentze MW. 2007. Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature* **447**: 875–878.

Tian Y, Simanshu DK, Ascano M, Diaz-Avalos R, Park AY, Juranek SA, Rice WJ, Yin Q, Robinson CV, Tuschl T, et al. 2011. Multimeric assembly and biochemical characterization of the Trax-translin endonuclease complex. *Nat Struct Mol Biol* **18**: 658–664.

Tinoco I, Borer PN, Dengler B, Levin MD, Uhlenbeck OC, Crothers DM, Bralla J. 1973. Improved estimation of secondary structure in ribonucleic acids. *Nature New Biol* **246**: 40–41.

Tinoco I, Uhlenbeck OC, Levine MD. 1971. Estimation of secondary structure in ribonucleic acids. *Nature* **230**: 362–367.

Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, Cook HA, Koppetsch BS, Theurkauf WE, Zamore PD. 2004a. RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell* **116**: 831–841.

Tomari Y, Du T, Zamore PD. 2007. Sorting of Drosophila small silencing RNAs. *Cell* **130**: 299–308.

Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. 2004b. A protein sensor for siRNA asymmetry. *Science* **306**: 1377–1380.

Tsutsumi A, Kawamata T, Izumi N, Seitz H, Tomari Y. 2011. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat Struct Mol Biol* **18**: 1153–1158.

Tyers M, Jorgensen P. 2000. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr Opin Genet Dev* **10**: 54–64.

Ucci JW, Kobayashi Y, Choi G, Alexandrescu AT, Cole JL. 2007. Mechanism of interaction of the double-stranded RNA (dsRNA) binding domain of protein kinase R with short dsRNA sequences. *Biochemistry-Us* **46**: 55–65.

van der Krol AR, Mur LA, Beld M, Mol JN, Stuitje AR. 1990. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**: 291–299.

van Heel M. 1984. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* **13**: 165–183.

van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. 1996. A new generation of the IMAGIC image processing system. *J Struct Biol* **116**: 17–24.

Villaseñor AG, Wong A, Shao A, Garg A, Donohue TJ, Kuglstatter A, Harris SF. 2012. Nanolitre-scale crystallization using acoustic liquid-transfer technology. *Acta Crystallogr D Biol Crystallogr* **68**: 893–900.

Walsh CT, Garneau-Tsodikova S, Gatto GJ. 2005. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**: 7342–7372.

Wang B, Li S, Qi HH, Chowdhury D, Shi Y, Novina CD. 2009a. Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. *Nat Struct Mol Biol* **16**: 1259–1266.

Wang B, Yanez A, Novina CD. 2008a. MicroRNA-repressed mRNAs contain 40S but not 60S components. *Proceedings of the National Academy of Sciences* **105**: 5343–5348.

Wang D, Zhang Z, O'Loughlin E, Lee T, Houel S, O'Carroll D, Tarakhovsky A, Ahn NG, Yi R. 2012. Quantitative functions of Argonaute proteins in mammalian development. *Gene Dev* **26**: 693–704.

Wang H-W, Noland C, Siridechadilok B, Taylor DW, Ma E, Felderer K, Doudna JA, Nogales E. 2009b. Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol* **16**: 1148–1153.

Wang X-H, Aliyari R, Li W-X, Li H-W, Kim K, Carthew R, Atkinson P, Ding S-W. 2006. RNA interference directs innate immunity against viruses in adult Drosophila. *Science* **312**: 452–454.

Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. 2008b. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**: 921–926.

Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ. 2009c. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**: 754–761.

Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. 2008c. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**: 209–213.

Washburn MP, Wolters D, Yates JR. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242–247.

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 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**: 589–599.

Welker NC, Pavelec DM, Nix DA, Duchaine TF, Kennedy S, Bass BL. 2010. Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. *Rna* **16**: 893–903.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855–862.

Wilkins C, Dishongh R, Moore SC, Whitt MA, Chow M, Machaca K. 2005. RNA interference is an antiviral defence mechanism in Caenorhabditis elegans. *Nature* **436**: 1044–1047.

Williams BR. 1999. PKR; a sentinel kinase for cellular stress. *Oncogene* **18**: 6112–6120.

Willis MC, Hicke BJ, Uhlenbeck OC, Cech TR, Koch TH. 1993. Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science* **262**: 1255–1257.

Witze ES, Old WM, Resing KA, Ahn NG. 2007. Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* **4**: 798–806.

Wriggers W, Birmanns S. 2001. Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J Struct Biol* **133**: 193–202.

Wu E, Thivierge C, Flamand M, Mathonnet G, Vashisht AA, Wohlschlegel J, Fabian MR, Sonenberg N, Duchaine TF. 2010. Pervasive and cooperative deadenylation of 3'UTRs by embryonic microRNA families. *Mol Cell* **40**: 558–570.

Wu L, Fan J, Belasco JG. 2006. MicroRNAs direct rapid deadenylation of mRNA. *P Natl Acad Sci USA* **103**: 4034–4039.

Xia T, SantaLucia J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry-Us* **37**: 14719–14735.

Xu P, Vernooy SY, Guo M, Hay BA. 2003. The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol* **13**: 790–795.

Yamashita S, Nagata T, Kawazoe M, Takemoto C, Kigawa T, Güntert P, Kobayashi N, Terada T, Shirouzu M, Wakiyama M, et al. 2011. Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci* **20**: 118–130.

Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou M-M. 2003. Structure and conserved RNA binding of the PAZ domain. *Nature* **426**: 468–474.

Yang SW, Chen H-Y, Yang J, Machida S, Chua N-H, Yuan YA. 2010. Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* **18**: 594–605.

Ye X, Huang N, Liu Y, Paroo Z, Huerta C, Li P, Chen S, Liu Q, Zhang H. 2011. Structure of C3PO and mechanism of human RISC activation. *Nat Struct Mol Biol* **18**: 650–657.

Ye X, Paroo Z, Liu Q. 2007. Functional anatomy of the Drosophila microRNA-generating enzyme. *J Biol Chem* **282**: 28373–28378.

Yekta S, Shih I-H, Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**: 594–596.

Yi R, O'Carroll D, Pasolli HA, Zhang Z, Dietrich FS, Tarakhovsky A, Fuchs E. 2006. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat Genet* **38**: 356–362.

Yoda M, Kawamata T, Paroo Z, Ye X, Iwasaki S, Liu Q, Tomari Y. 2010. ATP-dependent human RISC assembly pathways. *Nat Struct Mol Biol* **17**: 17–23.

Yuan Y-R, Pei Y, Ma J-B, Kuryavyi V, Zhadina M, Meister G, Chen H-Y, Dauter Z, Tuschl T, Patel DJ. 2005. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* **19**: 405–419.

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

Zekri L, Huntzinger E, Heimstädt S, Izaurralde E. 2009. The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol Cell Biol* **29**: 6220–6231.

Zeng Y, Cullen BR. 2005. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**: 27595–27603.

Zeng Y, Sankala H, Zhang X, Graves PR. 2008. Phosphorylation of Argonaute 2 at serine-387 facilitates its localization to processing bodies. *Biochem J* **413**: 429–436.

Zeng Y, Yi R, Cullen BR. 2005. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *Embo J* **24**: 138–148.

Zhang H, Kolb FA, Brondani V, Billy E, Filipowicz W. 2002. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *Embo J* **21**: 5875–5885.

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

Zheng X, Bevilacqua PC. 2000. Straightening of bulged RNA by the double-stranded RNA-binding domain from the protein kinase PKR. *P Natl Acad Sci USA* **97**: 14162–14167.

Appendix A

Post-Translational Modifications in RNAi

A.1 Introduction

A.1.1 Post-translational modifications and the control of biological pathways

Post-translational modifications (PTMs) are covalent protein modifications that occur at a number of amino acids and exert a range of effects on protein structure, stability, localization, and interactions. These modifications include phosphorylation, ubiquitylation, acetylation, glycosylation, methylation, hydroxylation, and many others (Walsh et al. 2005; Seo and Lee 2004; Mann and Jensen 2003). Such modifications are known to add an exquisite layer of control to essential biological systems such as the cell cycle, during which ubiquitylation of cyclins at defined time points marks cycle progression (Tyers and Jorgensen 2000). Another example of the control exerted by PTMs regards the autophosphorylation of PKR, which converts that protein – essential for apoptotic pathways that are triggered upon viral infection – from an inactive to an active state. This activation causes PKR to inhibit translation through its phosphorylation of another protein, eIF2α (Williams 1999).

The rapid development of mass spectrometry techniques for proteome-wide identification of post-translational modifications has led to a massive expansion of our ability to identify and characterize novel modifications, such that over 200,000 eukaryotic PTMs have been identified to date (Beltrao et al. 2012; Witze et al. 2007; Rush et al. 2005; Dephoure et al. 2008). Until recently, it was uncertain as to whether or not post-translational modifications might play a role in the regulation of the RNAi pathway. Recently, however, multiple studies have begun to show the extensive network of post-translational modifications that control different aspects of RNAi.

A.1.2 Post-translational modifications in RNAi

Mass spectrometry has demonstrated that Ago2 is phosphorylated at several residues (S253, T303, T307, S387, Y393, Y529, and S798). Interestingly, phosphorylation of Y529, which is located in the Mid domain, was shown to reduce small RNA binding by Ago2, indicating that the cell may be able to regulate complete unloading of an Ago2 protein or shut off RNAi altogether under certain conditions (Rüdel et al. 2011). The timed unloading of Ago2 would likely have further consequences on turnover, given that free Ago2 is relatively unstable (Elkayam et al. 2012; Johnston et al. 2010). Rather than being involved in Ago2 turnover, The MAPK-dependent phosphorylation of S387 is important for P-body localization, particularly under conditions of cellular stress (Zeng et al. 2008). To date, the effects of Ago2 phosphorylation at the other sites mentioned above – as well as the specific kinases responsible for those modifications – have not been investigated.

In addition to phosphorylation, Ago2 goes through several other post-translational modifications that are important to regulating its function. These modifications alter both Ago2 stability and localization, with resulting effects on overall silencing *in vivo*. One such modification is the poly(ADP-ribose)ylation of Ago2's PIWI domain. This modification, which occurs under conditions of cellular stress, both drives the localization of Ago2 to stress granules and also decreases miRNA-directed translational silencing and siRNA-directed mRNA cleavage (Leung et al. 2011). Another modification that leads to a decrease in miRNA activity is the ubiquitylation of Ago2 by the E3 ubiquitin ligase Lin-41 (a *let-7* target). This modification leads to Ago2 turnover (Rybak et al. 2009). A final known modification that stabilizes Ago2 is the prolyl-4-hydroxylation of the protein at P700 by the type I collagen prolyl-4-hydroxylase. Lack of this hydroxylation leads to higher Ago2 turnover, likely due to a folding defect, with the result of decreasing miRNA-directed translational repression (Qi et al. 2008).

Beyond Ago2, it is now understood that the dsRBPs involved in RNAi are also targets of phosphorylation. Specifically, TRBP is phosphorylated at two sites on each of its flexible linkers. In the first linker, S142 and S152 are modified, and in the second, S283 and S286 are modified. These phosphorylation events are regulated by the MAPK/Erk pathway and are essential to the stability of the Dicer/TRBP heterodimer. Due to this stabilizing effect, phosphorylation at these sites leads to an increase in miRNA-directed silencing (Paroo et al. 2009).

Given the importance of TRBP phosphorylation to the stability of the Dicer/TRBP complex, it is possible that PACT may be similarly phosphorylated in a way that would stabilize the analogous Dicer/PACT complex. If this were the case, it would be interesting to know if PACT's phosphorylation is regulated in the same manner as TRBP's. Although currently it is unknown if PACT's linker regions are phosphorylated, the protein has been shown to be phosphorylated at two sites in its third dsRBD. These phosphorylation sites, however, have only been investigated as they relate to the PKR pathway. PACT was originally identified as a protein activator of PKR under conditions of cellular stress, whereas TRBP inhibits PKR activity (Patel and Sen 1998; Park et al. 1994). PACT's activation of PKR is phosphorylation- and cellular stress-dependent. Within PACT's C-terminal dsRBD, S246 is known to be constitutively phosphorylated, and upon cellular stress S287 becomes phosphorylated, leading to the activation of PKR (Peters et al. 2006). The mouse homolog of PACT, RAX is also known to be phosphorylated at S18 following cellular stress, although this modification and its importance for activating PKR in humans has yet to be demonstrated (Bennett et al. 2004). Interestingly, it has recently been demonstrated that TRBP may inhibit PKR by forming heterodimers with PACT in unstressed cells, thereby sequestering it from PKR. Upon stress-induced phosphorylation of S287 of PACT this heterodimer dissociates, freeing PACT to activate PKR (Singh et al. 2011). It stands to reason that phosphorylation events on PACT's third dsRBD may affect its binding to Dicer as well, leading to an increase or decrease in the cellular levels of this complex relative to the Dicer/TRBP complex. Regardless, the dissociation of TRBP and PACT heterodimers may itself alter the ratio of Dicer/TRBP and Dicer/PACT heterodimers, which would have implications for gene silencing efficiency in RNAi (See **Chapter 4**).

Currently, the ability of post-translational modifications to directly alter Dicer activity is unknown, despite the fact that three phosphorylation sites on Dicer have been identified by mass spectrometry. One proteome-wide search for mitotically regulated phosphorylation sites identified S1016, located in Dicer's PAZ domain, and S1255, located in the linker between Dicer's PAZ and RNase IIIa domains (**Figure A.3**; (Dephoure et al. 2008)). A second proteome-wide study identified Y664, which is located in Dicer's DUF domain, as a phosphotyrosine (Rush et al. 2005). None of these phosphorylation sites, however, have been further studied to determine the extent to which they may modulate Dicer's function.

A.1.3 Research rationale

Due to the paucity of information regarding the effects of phosphorylation on the function of Dicer and PACT, we undertook to identify additional phosphorylation sites on each protein and determine their effects on pre-miRNA processing and protein-protein or protein-RNA interactions. We used a tandem affinity purification approach to purify proteins from human 239T cells for a mass spectrometry analysis of phosphorylation sites. We were only able to identify one phosphorylation site on PACT by this method. However, we have also shown that Dicer is phosphorylated at multiple sites in human cells. Although preliminary, our results point to the possibility that PAZ domain phosphorylation leads to subtle but reproducible changes in

Dicer's processing efficiency for a subset of pre-miRNAs. On the other hand, the production of isomiRs was not affected for the specific pre-miRNAs investigated. These preliminary results warrant future study, which may unveil the first evidence of the direct cellular regulation of Dicer's activity achieved via post-translational modification.

A.2 Methods

A.2.1 Tandem affinity purification of PACT and Dicer

Human PACT and Dicer cDNAs were cloned into a vector containing N-terminal 3xFLAG- and 2xStrep-tags that was derived from the vector pcDNA4/TO (Invitrogen; (Jäger et al. 2011)). For each protein, ten 10 cm dishes of 293T cells at ~70% confluence were transfected with 24 μg plasmid DNA using Lipofectamine 2000 (Invitrogen). Following a 48-hour incubation at 37°C, each plate of cells was washed twice with 7 mL PBS and harvested with 500 μL TAP Lysis Buffer (50 mM HEPES-KOH, pH 7.5; 170 mM NaCl; 0.5% Nonpidet-40; 5 mM EDTA; 1 mM TCEP; 20 mM NaF; 1x EDTA-free protease inhibitor (Roche); and 1x phosphatase inhibitor (Roche)). Cells were lysed by sonication and lysates were cleared by centrifugation at 16,000 g for 15 minutes at 4°C. Cleared lysates were applied to Spin-X centrifugal filters (Corning) and spun again for 5 minutes at 4°C at 16,000 g.

Cleared lysates for either protein were applied to 1 mL Strep-Tactin Superflow resin (Qiagen) that had been pre-equilibrated with TAP Lysis Buffer in a 15 mL Falcon tube and incubated overnight with gentle rotation at 4°C. Samples were then loaded onto a gravity column. Strep-Tactin resin was washed 3x with 5 mL TAP Wash Buffer (1x phosphate inhibitors and 0.5% Nonpidet-40 in TBS). The columns were then capped, bound proteins were resuspended with 4 mL Desthiobiotin Elution Buffer (2.5 mM desthiobiotin in TBS), moved to a fresh 15 mL Falcon tube, and rotated 30 minutes at 4°C. Following elution, samples and resin were returned to gravity columns to collect the eluates. The Strep-Tactin resin was then washed with 1 mL Desthiobiotin Elution Buffer and this fraction was pooled with the eluates.

Strep-Tactin eluates were next applied to 500 μL Anti-FLAG M2 resin (Sigma) that had been pre-equilibrated with TBS in a 15 mL Falcon tube and incubated at 4°C with gentle rotating for 4 hours. Samples were loaded onto gravity columns and the FLAG resin was washed 5x with TBS. The columns were then capped and bound proteins were resuspended in 1 mL FLAG Elution Buffer (0.2 μg/μL 3xFLAG peptide in TBS) and transferred to a 1.5mL tube. Samples were incubated with FLAG Elution Buffer for 30 minutes at 4°C with end-over-end rotation. Tubes were then centrifuged at 8,200 g for 30 seconds at 4°C. Supernatants were removed, leaving ~200 μL in the tubes. Resin was resuspended in the remaining eluate and applied to a Spin-X filter and spun at 8,200 g for 30 seconds at 4°C to collect final eluates. For each protein, the final eluates were pooled and concentrated to 150 μL in a SpeedVac. For Dicer, this method gave a highly purified protein, whereas for PACT there was still a significant number of contaminating proteins present in the elution fractions (**Figure A.1**).

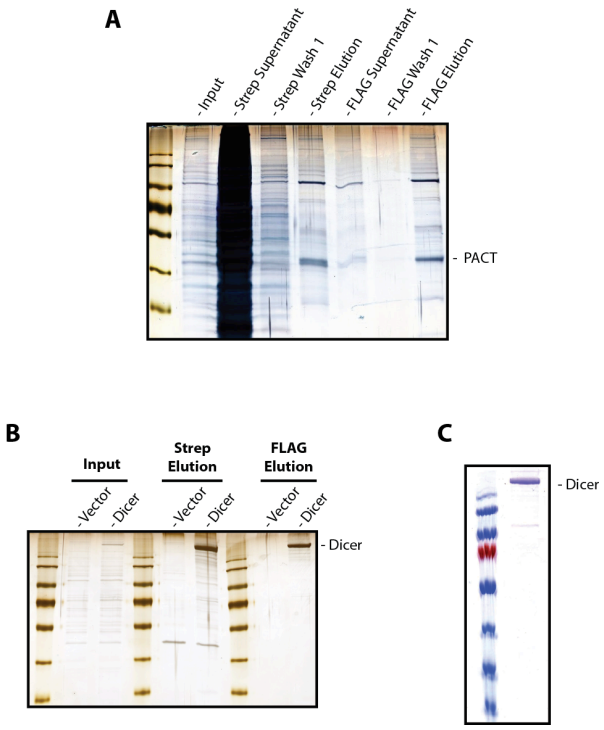

Figure A.1: Tandem affinity purification of PACT and Dicer.
(A) Silver stained SDS-PAGE gel of tandem affinity purification of 3xFLAG-2xStrep-PACT. **(B)** Silver stained SDS-PAGE gel of tandem affinity purification of 3xFLAG-2xStrep-Dicer. **(C)** Coomassie Blue-stained SDS-PAGE gel of FLAG eluate from 3xFLAG-2xStrep-Dicer tandem affinity purification shows that Dicer has been purified to near homogeneity.

A.2.2 Trypsin digestion and phosphopeptide enrichment for mass spectrometry

To chloroform/methanol precipitate FLAG eluates, 600 µL methanol was added to each sample and vortexed for 5 seconds. Samples were then centrifuged for 20 seconds at 9,000 g. 150 µL chloroform was then added and the samples were again vortexed and centrifuged for 20 seconds at 9,000 g. Next, 450 µL H_2O was added to each sample, vortexed, and centrifuged for 1 minute at 9,000 g. The aqueous (upper) layer was then removed and discarded and 450 µL of methanol was added. The sample was vortexed and spun for 2 minutes at 16,000 g. The supernatants were then discarded and the samples were air-dried. Protein pellets were stored at -80 °C until needed.

For Trypsin digestion, protein pellets were resuspended in 80 µL Resuspension Buffer (100 mM Tris-HCl, pH 8.5 and 8 M Urea) and 2.5 µL of 100 mM TCEP was added. Samples were then incubated at room temperature for 20 minutes. Next, 1.7 µL of fresh 500 mM iodoacetamide was added and the samples were incubated in the dark at room temperature for 15 minutes. Samples were then diluted 4-fold by the addition of 282.6 µL Tris-HCl, pH 8.5. Next, 3.4 µL 100 mM $CaCl_2$ was added, followed by 500 ng sequencing-grade Trypsin (Promega). Samples were incubated overnight at 37°C in the dark. Following overnight incubation, samples were spun down quickly and then dried in a SpeedVac.

Tryptic peptides were enriched for phosphopeptides using an iron-NTA phosphopeptide enrichment kit (Pierce) and desalted using graphite spin columns (Pierce). Samples were then dried in a SpeedVac and resuspended in a solution of 4% acetonitrile and 1% formic acid for analysis by tandem mass spectrometry or mudPIT mass spectrometry.

A.2.3 Expression and purification of Dicer phosphomimics and phosphomutants

Dicer phosphomimic (S1015D/S1016D) and phosphomutant (S1015A/S1016A) proteins were cloned by site-directed mutagenesis and expressed in a method identical to wild type Dicer (see **Section 2.3.2**).

A.2.4 Single-turnover dicing timecourse assays

Single-turnover dicing assays were set up in pooled 50 µL reactions containing 50 nM wild type, phosphomimic, or phosphomutant Dicer, 1x Dicing Buffer, and 500 cpm/µL 5′ ^{32}P-labeled substrate RNA (<0.5 nM). Reactions were incubated at 37°C and 5 µL aliquots were removed at the indicated timepoints and stopped by the addition of 6 µL 2x Formamide RNA Loading Dye. Reactions were run on a 15% denaturing PAGE gel. Gels were then dried under a vacuum and exposed to a phosphor-screen overnight before scanning on a STORM phosphorimager and analyzing with the ImageQuant software suite.

A.3 Preliminary results and discussion

A.3.1 Minimal phosphorylation identified for PACT

In order to verify the phosphorylation state of PACT in human cells, we cloned PACT cDNA into a vector derived from the pcDNA4/TO that would place 3xFLAG-and 2xStrep-tags at the N-terminus of PACT (Jäger et al. 2011). This vector was then transfected into human 293T cells. Lysates were subjected to tandem affinity purification, which would allow for purification of the PACT construct, while minimizing non-specific proteins. Surprisingly, by this method our final purified protein still contained many contaminants (**Figure A1 A**). Despite this, we trypsin digested the sample and enriched for phosphopeptides via iron-IMAC. After desalting the sample, these peptides were subjected to tandem mass spectrometry for identification of PACT phosphopeptides. Likely due to the high number contaminating peptides, only one phosphorylation site was identified for PACT (**Table A1, Figure A2**).

Table A.1: PACT phosphorylation site determined by mass spectrometry.

Peptide	Amino Acid	Domain
EDSGTFSLGK	18	N-terminus

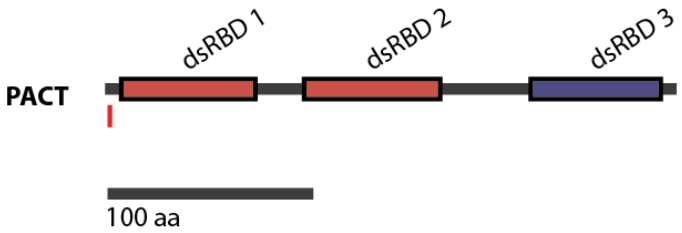

Figure A.2: Diagram of PACT's domain structure.
Phosphorylation site identified by mass spectrometry is denoted with a red line.

Interestingly, we found that S18 of human PACT is indeed phosphorylated, as it is in RAX, the mouse homolog of PACT (Bennett et al. 2004). It is likely that this phosphorylation site is important for the activation of PKR, as in mice, but we were interested to know how such a modification might affect PACT's function in RNAi. Thus far, we have not been able to

conduct any conclusive functional assays to determine the effects of this modification or the S246 or S287 phosphorylation sites on PACT. It is possible that this and other phosphorylation sites may alter the stability of PACT and/or PACT-containing complexes. Future studies investigating the stabilities of either wild type or phosphomimic PACT proteins *in vivo* by ^{35}S-labeling and pull-downs of Dicer using either wild type of phosphomimic PACT would shed light on this. Such alterations of stability may result in changes in small RNA pools, as is the case with TRBP phosphorylation (Paroo et al. 2009).

Although it is unlikely that phosphorylation at S18 would affect PACT's binding to Dicer, there is a strong possibility that phosphorylation at either S246, S287 or both would have an effect on formation of this heterodimer. If this were the case, these modifications would have quite interesting implications for the ratio of Dicer/TRBP to Dicer/PACT present in the cell. Since TRBP has been shown to alter Dicer processing activity (Chakravarthy et al. 2010; Lee and Doudna 2012; Fukunaga et al. 2012) and TRBP and PACT have been shown to have differing effects on Ago2's target specificity (see **Chapter 4**), alterations in the ratios of complexes containing these proteins could result in dramatic changes in the repression of certain genes. One tempting example would be the de-repression of certain miR-122 targets upon cellular stress. Although this de-repression was found to be HuR-dependent, it is possible that multiple factors are involved (Bhattacharyya et al. 2006). Since S287 of PACT is phosphorylated in a cell stress-dependent manner, it would be interesting to know if this phosphorylation similarly de-represses these and other miRNA targets.

RNA binding is unlikely to be affected by modification at the phosphorylation sites on PACT's third dsRBD. However, phosphorylation at S18 may have an effect on RNA binding, as it is located near dsRBD1. Filter binding assays comparing the affinities of wild type and phosphomimic proteins for both siRNAs and miRNAs may reveal such an effect. This and other possibilities presented above would have important implications for RNA processing in RNA interference. It is possible that subtle changes in the ratios of TRBP-containing and PACT-containing protein complexes in the cell would translate to large changes in small RNA pools and thus the specific genes targeted for repression. These potential changes have as yet remained unexplored and may provide expanded insights into the cellular regulation of RNAi pathways in humans.

A.3.2 Human Dicer is phosphorylated at multiple sites

We took a similar approach to identifying phosphorylation sites with human Dicer. 3xFLAG-2xStrep-Dicer was transfected into 293T cells and Dicer was purified by tandem affinity purification. Unlike PACT, this method yielded a very pure Dicer protein (**Figure A1 B, C**). MudPIT mass spectrometry (Washburn et al. 2001) revealed 12 phosphorylation sites on human Dicer, 10 of which are novel (**Table A.2, Figure A.3**). One previously known phosphorylation site – Y664 – was not identified by this method.

Table A.2: Dicer phosphorylation sites determined by mass spectrometry.

Peptide	Amino Acid	Domain
IK**S**FEEDLK	571	Helicase
ALPL**S**SAEKR	1015	PAZ
ALPLS**S**AEK	1016	PAZ
TLLSE**S**PGK	1160	PAZ - RNase IIIa Linker
STSDGSPVMAVMPGTTDTIQVLK	1250	PAZ - RNase IIIa Linker
S**T**SDGSPVMAVMPGTTDTIQVLK	1251	PAZ - RNase IIIa Linker
ST**S**DGSPVMAVMPGTTDTIQVLK	1252	PAZ - RNase IIIa Linker
STSDG**S**PVMAVMPGTTDTIQVLK	1255	PAZ - RNase IIIa Linker
MDSEQ**S**PSIGYSSR	1280	PAZ - RNase IIIa Linker
AAQLFLC**S**LGLK	1585	RNase IIIa - RNase IIIb Linker
FSANVPR**S**PVR	1852	dsRBD
ANQPQVPN**S**	1922	C-terminus

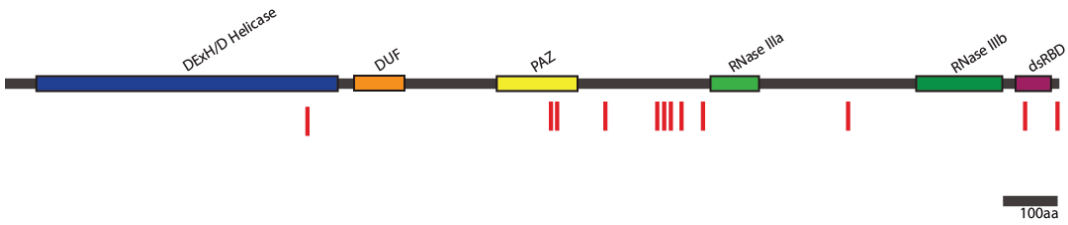

Figure A.3: Diagram of Dicer's domain structure.
Phosphorylation sites identified by mass spectrometry are denoted with red lines.

The phosphorylation sites identified by this method are all highly conserved among vertebrates, and two sets occurred in clusters, including a tandem pair of serines in Dicer's PAZ domain and a set of three serines and one threonine clustered near the end of the linker N-terminal to Dicer's RNase IIIa domain. Although both clusters were of particular interest, we chose to focus on the tandem serines in the PAZ domain, S1015 and S1016. The reasoning behind this decision was based on unpublished structural information generously shared by Dinshaw Patel's lab, which seemed to indicate that modifications at these sites might interfere with substrate RNA binding.

A.3.3 PAZ domain phosphorylation subtly affects Dicer's processing efficiency

We expressed both the PAZ double-phosphomimic (S1015D, S1016D) and double-phosphomutant (S1015A, S1016A) forms of Dicer from insect cells using a baculoviral expression system. To determine the effects of PAZ domain phosphorylation, we determined the substrate cleavage rates for a set of six RNAs: pre-let7-a, pre-miR-34c, pre-miR-21, pre-miR-200a, pre-miR-29a, and a 38bp perfect duplex pre-siRNA (**Figure A.4**).

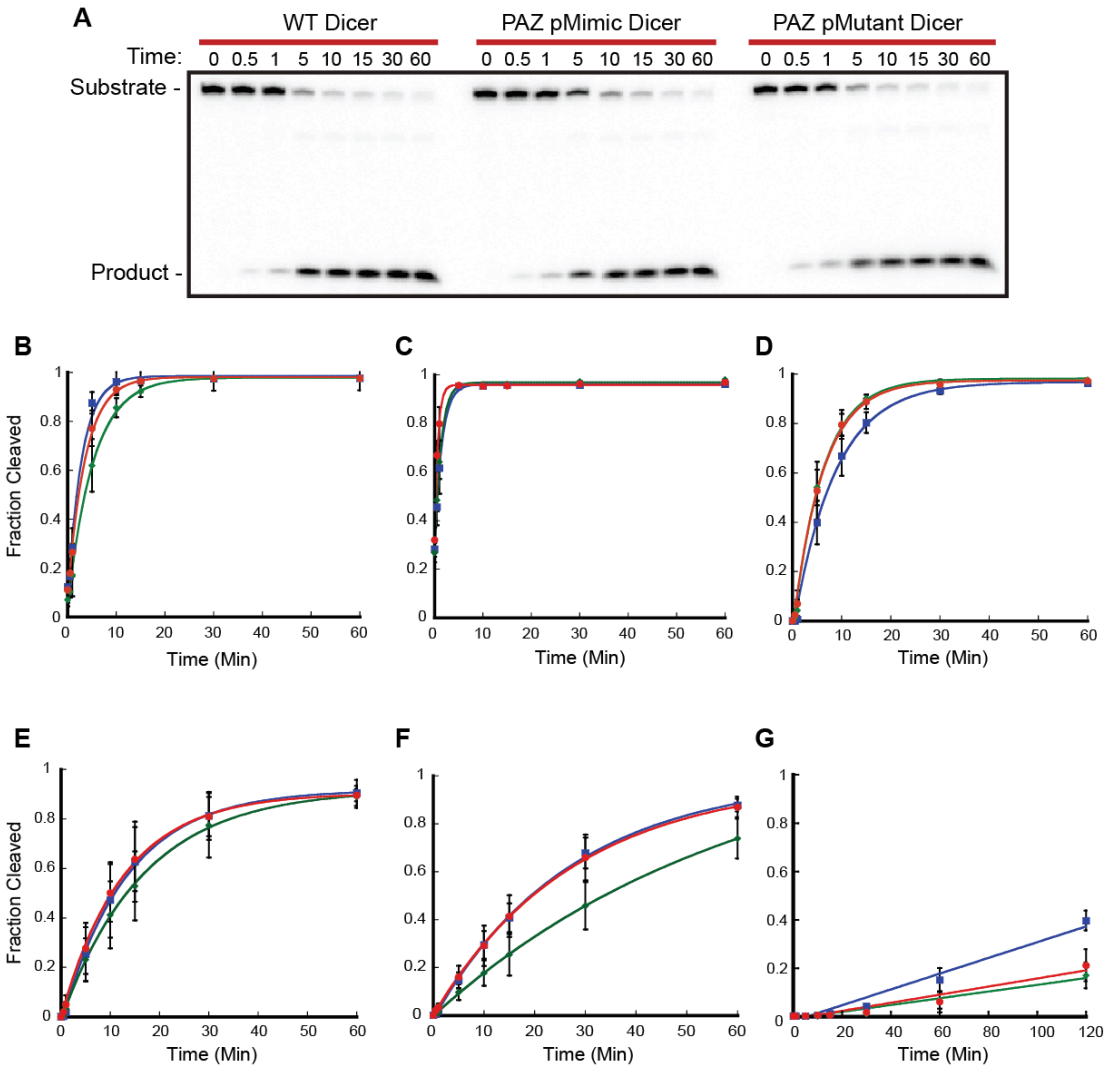

Figure A.4: PAZ phosphorylation subtly affects the processing of some substrates.
(A) Phosphorylation of Dicer's PAZ domain does not affect pre-let-7a processing rates. Dicer proteins were incubated with 5′ radiolabeled pre-let-7a at 37°C for the indicated time points and products were resolved by 15% denaturing PAGE. **(B)** Phosphorylation of Dicer's PAZ domain does not affect pre-let-7a processing rates. **(C)** Phosphorylation of Dicer's PAZ domain does not affect pre-miR-34c processing rates. **(D)** Phosphorylation of Dicer's PAZ domain does not affect pre-miR-21 processing rates. **(E)** Phosphorylation of Dicer's PAZ domain does not affect pre-miR-200a processing rates. **(F)** Phosphorylation of Dicer's PAZ domain results in a subtle decrease in pre-miR 29a processing rates. **(G)** Phosphorylation of Dicer's PAZ domain subtly affects the processing rates for a perfectly matched dsRNA. Data represent means (+/- SD) from three experimental replicates (Blue: WT; Green: phosphomimic; Red: phosphomutant).

For most of the substrate RNAs tested, we found no significant change in cleavage rate (**Figure A.4 A-E**). Importantly, since dsRBPs are known to affect Dicer processing, processing of these RNAs will need to be tested with phosphomimic and phosphomutant Dicer/dsRBP heterodimers in order to fully conclude that the processing rates of these RNAs are not affected by PAZ domain phosphorylation. Additionally, the conditions tested here represent single-turnover conditions, where the concentration of RNA is limiting. Multiple-turnover conditions should also be tested to further investigate the effects of PAZ domain phosphorylation. Phosphorylation did have a subtle effect on pre-miR-29a processing, as well as with the

processing of a perfectly matched long dsRNA (**Figure A.4 F-G**). Thus, as least for a subset of substrates, PAZ domain phosphorylation likely has a small effect on single-turnover Dicer processing efficiency. Further study of these and other RNAs may reveal the structural elements of particular Dicer substrates that are affected by PAZ domain phosphorylation or the extent to which the processing of these substrates is affected *in vivo*.

A.3.4 PAZ domain phosphorylation does not affect isomiR production by Dicer

In addition to canonical miRNAs, with particular pre-miRNAs Dicer produces two to three isomiRs, or products that differ in length by one nucleotide (Fernandez-Valverde et al. 2010; Lee et al. 2010; Lee and Doudna 2012; Fukunaga et al. 2012). We reasoned that relatively small changes in how a substrate RNA is bound by the PAZ domain might lead to changes in isomiR formation. It was therefore possible that PAZ domain phosphorylation would lead to an altered isomiR profile. Dicer is known to generate isomiRs from both pre-miR-29a and pre-miR-34c. Similarly, while Dicer alone does not generate isomiRs from pre-miR-200a, the Dicer/TRBP heterodimer does (Lee and Doudna 2012). We tested each of these substrates for isomiR formation with wild type Dicer, PAZ phosphomimic Dicer, or PAZ phosphomutant Dicer, and found no clear difference in isomiR formation (**Figure A.5**). However, this finding does not rule out the possibility that PAZ phosphorylation might lead to isomiR generation with some substrates, particularly given that we have not yet extended these experiments to Dicer/TRBP or Dicer/PACT heterodimers.

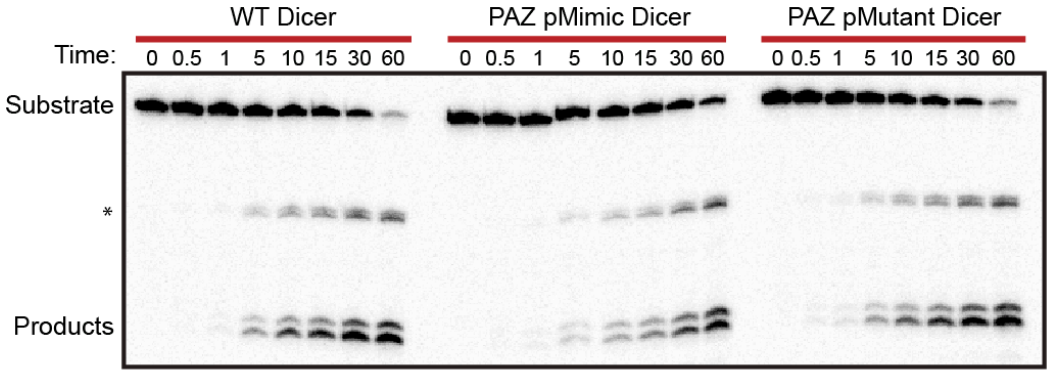

Figure A.5: PAZ phosphorylation does not affect isomiR production with pre-miR-29a.
Dicer proteins were incubated with 5′ radiolabeled pre-let-7a at 37°C for the indicated time points and products were resolved by 15% denaturing PAGE. Asterisk denotes a band resulting from cleavage at only the 3′ cleavage site.

The number of post-translational modifications on Dicer identified here – particularly when considering the vast number of potential substrate RNAs, protein complexes, and steps of the RNAi pathway that could be affected by those modifications – presents a challenging practical problem in terms of identifying the effects of specific post-translational modifications in a practical manner. Perhaps the most logical approach to this problem would involve using deep-sequencing technologies. For this approach, a Dicer knockout cell line could be used (Yi et al. 2006; Murchison et al. 2005; Kanellopoulou et al. 2005), and transfected with either wild type or phosphomimic Dicer. Small RNA populations could then be analyzed on a global scale by deep sequencing to assess changes in the levels of particular miRNAs or differences in isomiR formation. Promising hits could then be followed up on with more specific binding and kinetics assays *in vitro*. A number of Dicer phosphomimics could be tested in this manner and the

samples multiplexed. One approach would be to individually create each of the 12 phosphomimics for phosphorylation sites identified here. Perhaps a more fruitful approach, however, would be to test combinations of phosphorylation sites such as the two clusters we identified. Additionally, it may be informative to analyze a phosphomutant that combines the phosphorylation sites present in different dsRNA binding domains such as the PAZ and dsRBD domains, identified here (**Table A.2**), as well as the DUF domain, identified by Rush and colleagues (Rush et al. 2005). In addition to possibly indicating a change in overall Dicer activity, global changes in RNA processing efficiency could indicate merely a change in Dicer stability or even localization. This caveat aside, the proposed approach may in the future uncover the first evidence of the direct cellular regulation of Dicer's processing activity via post-translational modification.

Appendix B

EM Analysis of the Human phospho-UPF1-eIF3 Interaction†

† This project was collaborative effort carried out by Cameron Noland and Dr. Karin Felderer. Sample crosslinking, grid preparation, and preliminary EM imaging was carried out by Dr. Karin Felderer. Cameron Noland performed sample imaging and the 2D EM data analysis.

B.1 Introduction

B.1.1 The nonsense-mediated mRNA decay pathway

Nonsense-mediated mRNA decay (NMD) is a highly conserved post-transcriptional quality control pathway that exists to prevent the translation of mRNAs harboring premature termination (nonsense) codons (PTCs; (Chang et al. 2007)). NMD occurs exclusively with CBP80/20-bound transcripts during the pioneer round of translation (see below), preventing subsequent rounds of translation from occurring on the same transcript by coordinating translational repression and mRNA decay (Ishigaki et al. 2001; Chang et al. 2007). The process of NMD requires that PTCs can be distinguished from *bona fide* stop codons, which is achieved by large protein complexes called exon junction complexes (EJCs) that are deposited at exon-exon junctions following splicing. During translation of a wild type transcript, ribosomes displace these complexes as they move along the mRNA. However, when a PTC is present, the ribosome will often stop upstream of an EJC, which signals for translational repression and mRNA decay (Chang et al. 2007; Lykke-Andersen et al. 2000).

B.1.2 A brief primer on eukaryotic translational initiation

A full understanding of the mechanisms of translational repression during NMD requires some knowledge of the mechanisms of translational initiation. In eukaryotes, this process is one of elegant complexity, involving at least 9 protein complexes called eukaryotic initiation factors (eIFs). The first round of translation of any given transcript is known as the "pioneer" round. In this round of translation, the 5′ 7-methyl-guanosine cap of an mRNA is bound by the cap binding protein (CBP) heterodimer CBP80/20 (Ishigaki et al. 2001). This complex recruits what is called the 43S "pre-initiation complex", composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF2-GTP-Met-tRNA$_i^{Met}$, eIF3, and eIF5. eIF3 is an essential component of this complex, and binds directly to the 40S subunit. The 43S complex scans along the 5′ UTR in the 5′-3′ direction until it recognizes an AUG start codon by codon-anticodon base pairing. This recognition event triggers a subunit rearrangement to form the 48S complex, committing the ribosome for translation. Finally, interface factors are released from the 40S subunit and joining of this subunit with the 60S ribosomal subunit occurs, resulting in an 80S ribosome bound to the AUG start codon at the P-site via an initiator tRNA (Jackson et al. 2010; Kozak 1978).

B.1.3 UPF1 and the mechanisms of NMD

Up-frameshift (UPF) proteins are essential to NMD. UPF1 is a helicase that binds mRNAs upon PTC recognition by the ribosome during the pioneer round of translation and recruits UPF2 and UPF3, which triggers rapid mRNA decay. The phosphorylation status of UPF1 is critical to NMD and is regulated by SMG proteins. SMG-1 is responsible for phosphorylating UPF1, whereas SMG-5, SMG-6, and SMG-7 promote dephosphorylation of the protein. UPF1 is known to exist as a part of a complex that contains SMG-1, called the SURF complex, that interacts directly with ribosomes stalled at PTCs. This complex forms a molecular bridge by binding to the protein UPF2, which is present on a downstream EJC, joining the ribosome and the EJC. The recognition of prematurely stopped ribosomes by the formation of this SURF/EJC bridge is thought to drive the phosphorylation of UPF1.

Phosphorylation of UPF1 triggers the recruitment of cellular RNA degradation machinery such as the 5′-3′ exonuclease Xrn1 as well as the exosome, which degrades RNA in the 3′-5′ direction. In conjunction with this mRNA decay, another important aspect of NMD is the repression of further translation. Translational repression is crucial to mRNA decay, as it

precludes additional ribosomes from binding to the NMD-targeted mRNA, ensuring that it is accessible to RNA degradation activities. Phosphorylated Upf1 mediates translational repression by binding directly to eIF3 during NMD and inhibiting the joining of the 40S and 60S ribosomal subunits (Isken et al. 2008). The specific mechanism by which phospho-UPF1 binding to eIF3 leads to the inhibition of ribosomal subunit joining and the resulting inhibition of translation is currently unknown. In an effort to shed light on this process, we have conducted a 2D EM analysis of negatively-stained phospho-UPF1 bound to eIF3.

B.2 Methods

B.2.1 Sample and grid preparation

Purified human eIF3 was a generous gift from John Hershey, and hyperphosphorylated FLAG-UPF1 was purified by Lynne Maquat's lab from SF2 cells that were pre-treated with 75 nM okadaic acid. A binding reaction was performed using a 1:3 molar ratio of eIF3 to FLAG-UPF1 in binding buffer containing 20 mM HEPES, pH 7.5; 50 mM NaCl; 1 mM EDTA; and 1 mM DTT. The resulting complex was then crosslinked by adding 0.01% glutaraldehyde for 15 minutes at room temperature. The crosslinking reaction was quenched by the addition of 1 M Tris, pH 7.5 for an additional 3 minutes. Crosslinked samples were then diluted to a final concentration of 25 nM before application of 4 μL to a 6-second glow-discharged, continuous carbon grid. Sample was adsorbed for 1 minute before staining with three consecutive drops of 70 μL 2% uranyl formate for 10 seconds each. Excess stain was removed by gentle blotting and grids were then air-dried.

B.2.2 Electron microscopy

Dried grids were analyzed using an FEI Technai-12 Bio Twin electron microscope at 120 kV using a magnification of 49,000. Images were recorded in low dose mode using a Gatan 1k x 1k CCD camera and a defocus of -0.8 μm.

B.2.3 Data processing

CCD images were binned to give a pixel size of 5.18 Å. Individual particles were picked using the program Boxer with a box size of 120 x 120 square pixels. Each particle was then band-pass filtered and normalized using the program IMAGIC. Images were subjected to reference-free 2D alignment and classification through iterated MSA and MRA using IMAGIC. Twenty classes were originally obtained from 5939 particles, with 200-400 particles per class. Particles from classes showing extra density when compared with classes of eIF3 alone (Siridechadilok et al. 2005) were separately extracted and subjected to another round of MSA. For each class, this yielded 5 subclasses with 40-80 particles in each subclass. Difference maps were created by subtracting subclasses containing eIF3 alone from subclasses containing UPF1-bound eIF3.

B.3 Preliminary results and discussion

B.3.1 Phosphorylated UPF1 binds to the eIF3a and eIF3c subunits of eIF3

We were interested in structurally investigating the interaction between phospho-UPF1 and eIF3. Such an understanding would hopefully reveal the mechanism of translational inhibition by phospho-UPF1 during NMD. It had been shown that when expressed from SF2 cells that are pre-treated with okadaic acid, UPF1 is hyperphosphorylated (Isken et al. 2008). This hyperphosphorylated UPF1 was bound to native eIF3 and analyzed by negative stain EM. 2D alignment and classification of the particles yielded 20 class averages containing 300-400

particles in each class. Several of these classes showed a blurred extra density at the eIF3 'left arm' and 'head' regions, indicating that this is the binding region for Upf1 (**Figure B1 A, D**, red arrowheads). For each of these classes, the raw particles were extracted and subjected to another round of MSA, yielding 5 subclasses that contained 40-80 particles per subclass. Each subclass within a given original class consisted of the same view of the complex, but with some subclasses showing unoccupied eIF3 and some showing a clear extra density corresponding to bound UPF1 (**Figure B1 B**, red arrowheads). This analysis revealed a high degree of conformational flexibility in the UPF1 binding interaction, with UPF1 located closer to the left arm of eIF3 in some cases (**Figure B1 B, Class I**), and closer to the head in others (**Figure B1, Class II**). For computational confirmation of the added density, difference maps were created by subtracting each unoccupied subclass from the matching occupied subclass. The difference densities were similar in size in each case, but differed in their overall location (**Figure B1 C**).

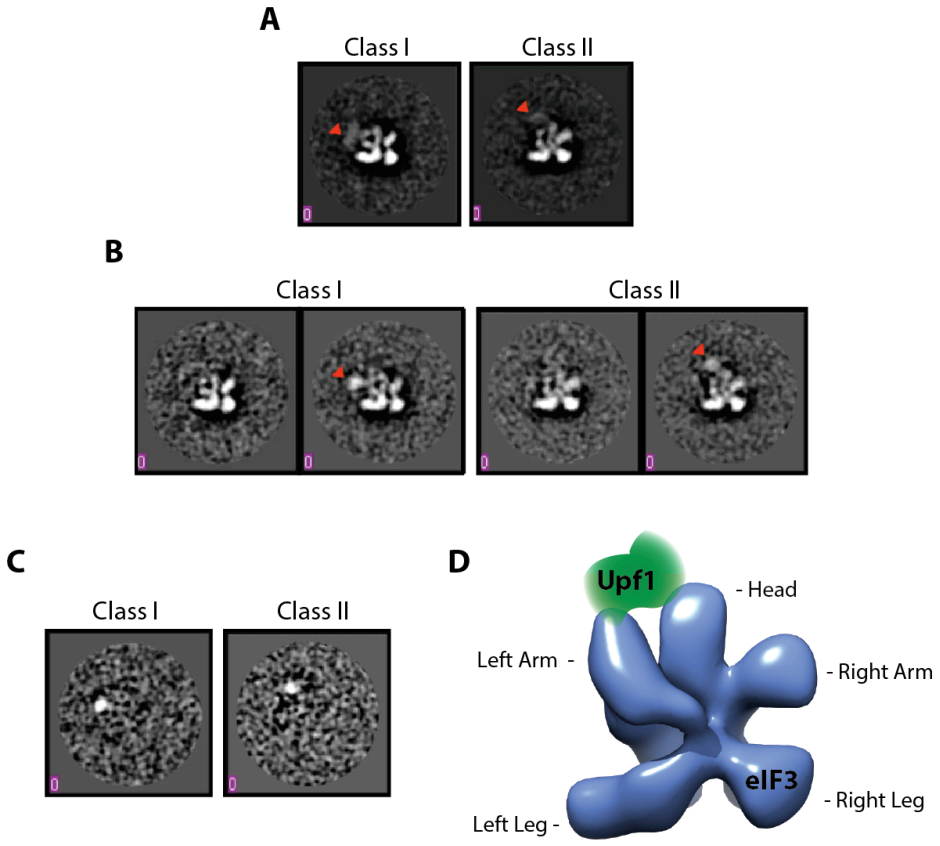

Figure B.1: Phospho-UPF1 binds to eIF3's left arm and head regions.
(A) Two representative classes containing extra density at the eIF3 left arm and head regions (red arrowheads). The left class consisted of 290 particles and the right class contained 318 particles. **(B)** Subclasses of the original classes in (A) that show either unoccupied eIF3 (left panel for each class) or extra density showing eIF3 interacting with UPF1 (red arrowheads; right panel for each class). **(C)** Difference maps showing the extra density in the UPF1-bound subclasses. Subclasses containing eIF3 alone were subtracted from subclasses containing the UPF1-eIF3 complex in (B). **(D)** Cartoon modeling the general binding site for UPF1 on the 3D reconstruction of eIF3 solved by cryoEM (Siridechadilok et al. 2005). UPF1 binds to the left arm and head regions of UPF1.

Recent subunit mapping of the eight eIF3 core proteins by EM has demonstrated that the left arm of eIF3 corresponds to the eIF3a subunit, whereas the head corresponds to the eIF3c

subunit. Furthermore, conserved motifs within these subunits (particularly for eIF3a) have been shown to be essential for eIF3 binding to the 40S ribosomal subunit, and ultimately for subunit joining to form the 80S ribosome complex (Sun et al. unpublished data). In light of these data, our 2D analysis of the phospho-UPF1-eIF3 interaction hints at a likely mechanism of translational repression by phospho-UPF1 during NMD. In binding to eIF3a and eIF3c, phospho-UPF1 likely competitively inhibits eIF3 binding to the 40S subunit, preventing formation of the 43S complex and therefore translational initiation.

With the advent of techniques for the recombinant expression and reconstitution of the human eIF3 core subunits from *E. coli* (Sun et al. 2011), many possibilities now exist for the biochemical validation of the interactions we have observed by EM. Specifically, mutations of the residues within eIF3a and eIF3c known to be important for binding to the 40S subunit could be used to determine the extent to which they affect binding to phospho-UPF1. This would hopefully allow for a more specific mapping of the phospho-UPF1-eIF3 binding interaction. Additionally, the identification of specific phosphorylation sites on UPF1 by means similar to those outlined in **Appendix A** will allow for the expression of specific phosphomimic and phosphomutant forms of UPF1. Such experiments will expand on our initial characterization of the interaction between phospho-UPF1 and eIF3 and potentially reveal a more specific mechanism of translational inhibition by phospho-UPF1 during NMD.

B.3.2 Phospho-UPF1 triggers a conformational change in the eIF3e subunit of eIF3

In addition to a potential mechanism of translation inhibition during NMD, our analysis of the phospho-UPF1-eIF3 class averages has given a preliminary indication that that the binding of UPF1 in some orientations may lead to a conformational change in the right arm of eIF3 (**Figure B2 Row 2**, yellow arrowhead).

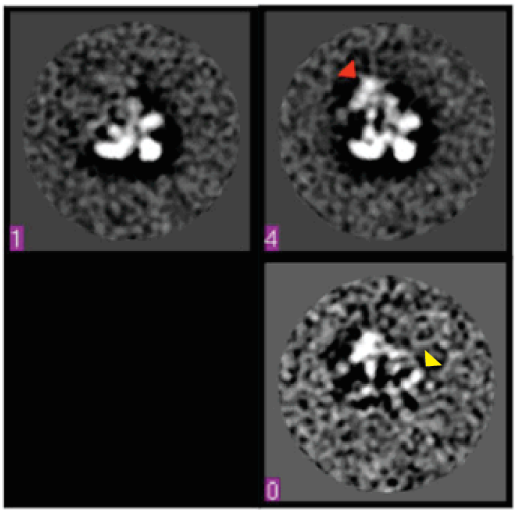

Figure B.2: Phospho-UPF1 binding induces conformational changes in eIF3.
Row 1: Subclasses showing unoccupied (left) and occupied (right, red arrowhead) eIF3. *Row 2:* Difference map obtained by subtracting the unoccupied eIF3 subclass from the UPF1-bound eIF3 subclass. Difference density indicating a conformational shift in the right arm of eIF3 is denoted by a yellow arrowhead.

The right arm of eIF3 corresponds to eIF3e, which is structurally similar to the proteosomal subunit Rpn6 (Sun et al. unpublished data). It is currently unclear if there are functional consequences to this conformational change that contribute to translational repression

by phospho-UPF1 or if this is merely a secondary effect of binding. A 3D cryoEM reconstruction of phospho-UPF1 bound to native eIF3 or to the bacterially expressed eIF3 core would likely reveal a more detailed view of this conformational change. Additionally, such a structure would help to reveal in greater detail the specific interactions that mediate translational repression by phospho-UPF1. This information would likely inform and complement biochemical characterizations of this interaction. Specifically, previously solved crystal structures of portions of UPF1 could be docked into the EM density to determine the precise domains of UPF1 that interact with eIF3 (Kadlec et al. 2006; Cheng et al. 2007). These future experiments will hopefully map the interaction between phospho-UPF1 and eIF3 in much finer detail, providing a clearer picture of the mechanisms of translational repression during nonsense-mediated mRNA decay.