Title
Mechanisms of RNA Recognition and Cleavage in RNA Interference

Author
Kidwell, Mary Anne Rose

Publication Date
2014

Peer reviewed|Thesis/dissertation

Mechanisms of RNA Recognition and Cleavage in RNA Interference

By

Mary Anne Rose Kidwell

A thesis 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
Assistant Professor Lin He
Professor Jeremy Thorner
Professor David Wemmer

Fall 2014

ABSTRACT

Mechanisms of RNA Recognition and Cleavage in RNA Interference

By

Mary Anne Rose Kidwell

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Jennifer A. Doudna, Chair

Proteins that harbor double stranded RNA-binding domains (dsRBDs) play functional roles in processes as diverse as RNA localization, RNA splicing, RNA editing, nuclear export of RNA and translation, yet the mechanistic basis and functional significance of dsRBDs remain unclear. To unravel this enigma, we investigated transactivation response (TAR) RNA-binding protein (TRBP), comprising three dsRBDs, which has functions in HIV replication, PKR-mediated immune response and RNA silencing. In collaborative studies using single-molecule methods, I found that TRBP exhibits an ATP-independent diffusion activity exclusively on double-stranded RNA (dsRNA) in a length-dependent manner. The first two dsRBDs of TRBP are essential for diffusion, whereas the third dsRBD is dispensable. Two homologs of TRBP, PKR activator (PACT) and R3D1-L, displayed the same behavior, implying that the ability to diffuse along dsRNA is a universal property of this protein family. Furthermore, a Dicer-TRBP complex on dsRNA exhibited dynamic diffusion, which was correlated with Dicer's catalytic activity, suggesting that the dsRNA-specific diffusion activity of TRBP contributes to enhancing small interfering RNA (siRNA) and microRNA (miRNA) processing by Dicer.

The enzyme Dicer generates 21-25 nucleotide-long dsRNAs that target specific mRNAs for silencing during RNA interference and related pathways. Although the active site and RNA-binding regions of Dicer are functionally conserved within eukaryotes, the helicase domain has distinct activities in the context of different Dicer enzymes. To examine the evolutionary origins of Dicer helicase functions, we investigated two related Dicer enzymes from the thermophilic fungus *Sporotrichum thermophile*. RNA cleavage assays showed that *S. thermophile* Dicer-1 (StDicer-1) can process hairpin precursors to miRNAs, whereas StDicer-2 can only cleave linear dsRNAs. Furthermore, only StDicer-2 possesses robust ATP hydrolytic activity in the presence of dsRNA. Deletion of the StDicer-2 helicase domain increases both StDicer-2 cleavage activity and affinity for hairpin RNA. Notably, the full-length forms (but not truncated versions lacking their helicase domain) of both StDicer-1 and StDicer-2 could complement the phenotype of a mutant of a distantly related yeast *Schizosaccharomyces pombe* lacking its endogenous Dicer gene, underscoring the importance of the helicase domain for Dicer function. These results further suggest that an *in vivo* regulatory function for the helicase domain may be conserved from fungi to humans.

ACKNOWLEDGEMENTS

There are many people without whom this thesis would not have been possible. First off, I would like to thank Jennifer Doudna for her encouragement over the years. Jennifer has created a great lab environment and I am honored to have been a part of her group for the last five years. I am very grateful to all the members of the Doudna Lab with whom I have worked. This is an extremely talented group and it has been humbling to work in such a great environment. The first person I worked with was Cameron Noland, an extremely patient graduate student who taught me all the basics of biochemistry starting with my first ever Dicer prep. Enbo Ma was my baymate for many years, a co-author on my first paper in the Doudna lab, and always generous with all his wisdom and knowledge. Ross Wilson has been with me in lab from the very beginning and although I'll probably never publish in PLOS, we'll always have Lady Gaga. Ross and I have held strong on Team Dicer, but it would not have been complete without Ho Young Lee and Aaron Brewster. Ho Young's meticulous biochemistry was helpful for my own dicing experiments and Aaron provided me with tons of crystallography knowledge. Someday we will get the structure of Dicer. (Unless David Taylor's EM becomes so incredible that we don't need crystallography.)

Outside of Team Dicer there were so many amazing people in the Doudna lab over the years. Martin Jinek was always encouraging and unbelievably modest about his accomplishments. I am proud to be able to say that I sat next to the famous Martin for many years. However, Yun Bai Great is stiff competition for most supportive baymate – she has been a wonderful person to talk to about both science and life. Stephen Floor, Philip Kranzusch, and Yun Bai have all given great advice on my project and I'm looking forward seeing them become PIs in the future. Stephen especially helped me connect with his former lab at USCF, specifically David Paquette, to help with a seemingly simple experiment. For my fellow graduate student, Sam Sternberg, I am happy to have spent all my time in lab with you and I appreciate all your insightful questions. James Nuñez has been an excellent Filipino buddy in lab and Megan helps keep the humor alive. Kaihong Zhou is of course, the most important person in the lab, who has inadvertently helped me find my true path in life.

My scientific output would have not been possible without my talented undergraduate Jessica Chan who put hours into running gels over and over again. Spencer Knight was also a great roton and now fellow lab member. Sua Myong and Hye Ran Koh were essential for completing my PNAS paper. There were amazing collaborators and I'm looking forward to seeing new, exciting research from their lab. Stephanie Mortimer was a fun co-author on the review and I hope to someday achieve her perfect life.

My committee members have provided immense amounts of support and encouragement, especially when I thought it would not be possible to complete this monumental task. Jeremy Thorner, James Berger, Lin He, and Dave Wemmer have helped shape me into the scientist that I am today. In addition to their help, I would like to thank Ming Hammond for her invaluable tip on the review. The faculty at Berkeley has been a great resource with an incredibly collaborative environment.

The most important people in my life have been my family. It has been difficult for my family back in Virginia, but I will be joining them soon! To Jason Ji, my soon-to-be husband, it has been a great adventure coming out to California with you. No one

doubts how caring and loving you have been to me over the years. I love you and I'm looking forward to spending the rest of my life with you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Overview and implications

Diverse and important new roles for RNA have emerged over the last few decades (Cech & Steitz, 2014). On the basis of the functions of mRNA, rRNA and tRNA, it was previously thought that RNA species served only as transitional molecules between DNA and protein. We now understand that RNA is much more versatile and has many other roles in cells. One of the most exciting discoveries in the past few decades has been uncovering the regulatory role of small RNAs in the RNA interference (RNAi) pathways, whereby a single small RNA can control the expression of multiple genes. The biogenesis of these small RNAs has been thoroughly explored, but many questions remain unanswered. Here I address how these small RNAs are naturally created and the regulation behind each of the steps. With this knowledge, I hope I can assist the field to develop a deeper understanding of the biological roles of RNA and potentially direct further research and therapeutic development for the future (Castanotto & Rossi, 2009).

1.2 Types of small RNAs

During RNAi, 20- to 30-nucleotide small RNAs post-transcriptionally down-regulate target messenger RNAs (mRNAs) containing complementary sequences. There are three main classes of small RNAs: microRNAs (miRNAs), short interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs) (Jinek & Doudna, 2009). Both miRNAs and siRNAs originate from double-stranded RNA (dsRNA) precursors that depend on the RNase III enzyme Dicer (Bernstein *et al.*, 2001). After cleavage by Dicer, miRNAs and siRNAs are loaded into the Argonaute family of proteins to direct binding to complementary mRNAs and prevent protein translation (Hutvagner & Simard, 2008). Although piRNAs also interact with proteins belonging to the Argonaute family, their biogenesis and function is markedly different from that of miRNAs and siRNAs and is beyond the scope of my dissertation research (Brennecke *et al.*., 2007).

1.2.1 Discovery of RNAi

RNAi was first documented in plants when attempts were made to overexpress chalcone synthase in petunia petals (Napoli *et al.*, 1990; van der Krol *et al.*, 1990). Chalcone synthase catalyzes a key step in flavonoid biosynthesis, which is responsible for plant coloration. Instead of the expected increase in pigmentation due to the overexpression of chalcone synthase, introduction of this gene blocked transgenic and endogenous flavonoid biosynthesis. Consequently, approximately 40% of the transgenic plants exhibited either a decrease in coloration or completely white flowers. This phenomenon was referred to as posttranscriptional gene silencing (PTGS). A similar process, termed "quelling," was also documented in the filamentous fungus *Neurospora crassa* when analyzing ectopic expression of carotenoid genes (Romano & Macino, 1992).

Nearly contemporaneously, short RNA molecules were found to control developmental timing in *Caenorhabditis elegans* (Lee *et al.*, 1993; Wightman *et al.*, 1993). The pathway revolved around *lin-4* which acts early in *C. elegans* development to negatively regulate the expression of LIN-14 protein. Classic genetic techniques revealed that *lin-4* did not encode a protein, but instead two small transcripts that were 22 and 61 nucleotides (nt), respectively. These sequences were complementary to the 3'-untranslated region (UTR) of the *lin-14* mRNA, leading to the proposal that *lin-4*

RNAs base pair to *lin-14* mRNA to down-regulate protein expression in a process called RNA interference (RNAi).

The molecular events underlying PTGS, quelling, and RNAi were soon unified. In the seminal paper by Andrew Fire and Craig Mello they described that double-stranded RNA, instead of single-stranded, led to down-regulation of endogenous mRNA in *C. elegans* (Fire *et al.*, 1998). Small, antisense RNAs were also discovered to be crucial for PTGS indicating a common mechanism for all small RNA pathways (Hamilton & Baulcombe, 1999).

1.2.2 Sources of miRNAs

Collectively, thousands of miRNAs have now been discovered in humans, flies, plants, fungi, and viruses (Kozomara & Griffiths-Jones, 2014; Ambros *et al.*, 2003; German *et al.*, 2008). In humans, up to 5% of the genome encodes miRNAs which regulate at least 30% of protein-coding genes (Berezikov *et al.*, 2005; Macfarlane & Murphy, 2010). In addition to their roles in development and tissue formation, miRNAs play major roles in heterochromatin formation and cell proliferation and mutations to this pathway can lead to many diseases, including cancer, cardiovascular disease, and neurological disorders (Volpe *et al.*, 2002; He *et al.*, 2005; 2007; Chen *et al.*, 2008; Lu *et al.*, 2008).

After analyzing the human genome, it was found that approximately half of all miRNAs are located in the introns of coding genes, whereas the remaining are located in intergenic regions and are expressed as non-coding transcripts (Saini *et al.*, 2007). Most of these miRNAs are transcribed by RNA polymerase II as primary miRNA transcripts (pri-miRNAs) containing cap structures and poly(A) tails (Cai *et al.*, 2004; Smalheiser, 2003; Lee *et al.*, 2004a). The few exceptions are miRNAs located within Alu elements or other repetitive elements that are transcribed by RNA polymerase III (Borchert *et al.*, 2006). Regardless of their location, multiple miRNAs are often encoded in a single polycistronic transcript that can be thousands of nts in length.

In the canonical miRNA biogenesis pathway, the pri-miRNAs will form hairpin secondary structures with imperfectly base-paired stems that are recognized and cleaved within the base-pair stems by the Microprocessor, a protein complex containing the RNase III enzyme Drosha and its partner protein DGCR8/Pasha (Denli *et al.*, 2004). The approximately 60-nt liberated hairpin products have 2-nt 3'-overhangs and are called precursor miRNAs (pre-miRNAs) (Starega-Roslan *et al.*, 2011). The pre-miRNAs are exported from the nucleus into the cytoplasm by Exportin-5 (Lund *et al.*, 2004). Once in the cytoplasm, processing continues with Dicer and the miRNAs and siRNAs are loaded into Argonaute for mRNA silencing (Figure 1.1) (Wilson & Doudna, 2013). Exceptions to this pathway include miRNAs that originate from introns or are processed through non-canonical, Dicer-independent pathways (Kim & Kim, 2007; Cifuentes *et al.*, 2010; Cheloufi *et al.*, 2010). The miRNAs originating from introns can be processed by the Microprocessor complex or they can bypass this step and are processed by the splicing machinery.

While most miRNAs are endogenously derived, they can also be derived from external sources. Virus can produce miRNAs by hijacking the cellular machinery and miRNAs have been identified in all of the herpesviruses and other DNA viruses (Grundhoff *et al.*, 2006). Viruses that have an RNA genome seem to lack miRNAs, although this is controversial for HIV-1 (Ouellet *et al.*, 2008; Schopman *et al.*, 2012).

1.2.3 Sources of siRNAs

The precursors of siRNAs are long dsRNA that range in size from tens to thousands of base pairs in length. These long dsRNA bypass nuclear processing by the Microprocessor complex and are directly processed by Dicer into 20-30 nt products. Once the RNA molecules are processed by Dicer, the siRNA and miRNA pathway merge mechanistically. Mature siRNAs are also loaded into Argonaute proteins to induce gene silencing, primarily through mRNA cleavage (Figure 1.1) (Meister *et al.*, 2004).

Unlike miRNAs, most siRNAs are usually derived from an exogenous source. Perhaps silencing by siRNAs first evolved as an ancient antiviral mechanism to help protect against RNA viruses in plants and flies (Waterhouse *et al.*, 2001). However, this defense mechanism is believed to have been outpaced by the advent of the innate and adaptive immune systems in mammals, as siRNAs are not efficiently processed in mammalian somatic cells (Svoboda, 2014). There are, however, some notable examples where endogenous siRNAs have been found in mammals that target repetitive elements and regulate endogenous genes in early development (Babiarz *et al.*, 2008). Experiments with a transgenic mouse expressing a long dsRNA showed that robust RNAi was present in oocytes, but not in the somatic tissues due to poor processing of the precursor RNA (Nejepinska *et al.*, 2012). Although the role of siRNAs as an antiviral response in mammals is still under debate, use of synthetic dsRNAs for RNAi have had a huge impact as an experimental research tool and even in the pharmaceutical world as potential drugs.

1.3 RNAi biogenesis

Molecular details of the enzymes involved in RNA processing have revealed great insights into how miRNAs and siRNAs are generated as well as the regulation behind their generation. Here I describe the enzymes involved in each step of the miRNA and siRNA biogenesis pathways.

1.3.1 Nuclear processing and export

After pri-miRNAs are transcribed,they must be processed by the Microprocessor complex prior to export into the cytoplasm. Many of the human intergenic miRNAs are encoded by transcripts that are 3-4 kb in length containing single or clustered hairpins that bear single-stranded 5’- and 3’-termini (Saini *et al.*, 2007). In addition to the hairpin secondary structure, pri-miRNAs must have specific sequence motifs for efficient processing in mammals, although the exact contributions of each motif are unknown (Auyeung *et al.*, 2013).

Pri-miRNAs transcripts are processed by the Microprocessor complex, which contains the RNase III enzyme Drosha and the RNA-binding partner protein DGCR8 (known as Pasha in invertebrates) (Denli *et al.*, 2004). Cleavage is performed by Drosha, which possesses a dimeric active site formed from tandem RNase III domains (Lee *et al.*, 2003). Cleavage occurs within the double-stranded region of the RNA, approximately 11bp from the ss/dsRNA junction, generating an approximately 60-nt pre-miRNA products with 2-nt 3’-overhangs (Han *et al.*, 2006). The tandem RNase III domains in Drosha are followed by a dsRNA-binding domain (dsRBD) that is required for processing *in vivo*, but not *in vitro* (Zeng & Cullen, 2005). However, Drosha is unable to accurately cleave pri-miRNAs in isolation and requires DGCR8 to correctly position itself on pri-miRNAs (Han *et al.*, 2006).

Part of DGCR8 has been crystallized in the absence of RNA and has given further insight into how the protein recognizes dsRNA (Sohn *et al.*, 2007). DGCR8 contains two dsRBD domains that are arranged with pseudo two-fold symmetry that could recognize a pri-miRNA in two possible orientations. In the first model, one molecule of DGCR8 spans an approximately 33 bp RNA, which is the length of a pri-miRNA stem. In the second model, one or two DGCR8 molecules cooperatively bind two pri-miRNAs, which has been supported by other studies suggesting that DGCR8 bound pri-miR30a as a trimer (Faller *et al.*, 2006). Regardless of the orientation, it is still unclear how the ssRNA is recognized to position DGCR8/Drosha 11bp from the ss/dsRNA junction in a pri-miRNA. The only ssRNA-binding activity is attributed to the serine/arginine-rich region of Drosha's N terminus and further investigation is needed to clearly define how the Microprocessor complex recognizes its substrates (Zeng & Cullen, 2005).

Once cleaved, the pre-miRNA is exported through the nuclear pore into the cytoplasm by exportin-5 (Xpo5) (Yi *et al.*, 2003; Lund *et al.*, 2004; Bohnsack *et al.*, 2004). Xpo5 is a member of the karyopherin family of proteins and is dedicated to the transport of pre-miRNAs. Transport by karyopherin proteins is driven by a gradient in the distribution of the GTP-bound state of the small Ras-related GTPase Ran. In the nucleus, Ran exists in its GTP-bound state, whereas in the cytoplasm it exists in its GDP-bound state (Cook & Conti, 2010). These two states are achieved because its guanine nucleotide exchange factor (Rcc1) is chromatin-bound and confined exclusively in the nucleus, whereas its GTPase-activating protein (Ran-GAP1) is bound on fibrils that extend in to the cytoplasm from the cytosolic face of the each nuclear pore complex. Biochemical experiments and crystal structures of the pre-miRNA export complex revealed that Xpo5, when bound to Ran-GTP, recognizes the 2-nt 3' overhang and the double-stranded stem of pre-miRNAs (Zeng & Cullen, 2004; Okada *et al.*, 2009). The pre-miRNA export complex interacts with 17 bp of the RNA duplex, helping to recognize structural features shared by different pre-miRNA hairpins. Once in the cytoplasm, the complex dissociates upon GTP hydrolysis, allowing cytoplasmic processing of the released pre-miRNA to begin.

Notable exceptions to this pathway are short, capped pre-miRNAs that are transcribed by RNA Pol II, but lack a poly(A) tail due to transcription termination (Xie *et al.*, 2013). These capped pre-miRNAs are exported by the sequential PHAX-exportin-1/Crm1/Xpo1pathway. Once in the cytoplasm, these substrates are also recognized and cleaved by Dicer.

1.3.2 Cytosolic processing by Dicer

Dicer is the main cytosolic component of the RNAi pathway that cleaves a variety of dsRNA precursors into mature, 20- to 30-nt miRNAs and siRNAs. The resulting RNA duplexes have 2 nt, 3'-overhangs that are characteristic of the cleavage products of RNase III family of enzymes, which includes Drosha. This signature, along with the specificity for dsRNA, allowed for biochemical discovery of Dicer (Bernstein *et al.*, 2001). Members of the three RNase III classes were epitope-tagged, purified, and tested for their ability to cleave RNA substrates into discrete products. Only one enzyme was able to perform this activity and was aptly named Dicer.

Subsequent biochemical and structural studies have revealed the mechanisms of RNA measurement and cleavage. Canonical Dicer consists of an N-terminal DExD/H

box RNA helicase, a DUF283 domain, a PAZ domain, tandem RNase III domains, and a C-terminal dsRBD (Figure 1.2). The DUF283, for "domain of unknown function," is now thought to contain a divergent dsRNA-binding element (Dlakić, 2006). The tandem RNase III domains make up the catalytic center of Dicer, which form an intramolecular dimer positioned to generate 2 nt, 3'-overhangs. These domains are related to its bacterial counterpart, a dsRNA-specific endonuclease that functions as a homodimer (Blaszczyk *et al.*, 2001). Mg^{2+} ions are used to catalyze RNA cleavage and the Mg^{2+} is coordinated by four conserved acidic residues in each active site.

In addition to the DUF283 domain, the RNase III domains are flanked by two other RNA-binding domains: the PAZ domain and the C-terminal dsRBD (Zhang *et al.*, 2004). The PAZ domain (named after the first three proteins in which it was recognized, Piwi, Argonaute, Zwille), which recognizes the 3'-overhanging nucleotides, helps to measure the distance from the cleavage site in conjunction with the RNase III domains. This hypothesis was confirmed by the crystal structure of a truncated Dicer from *G. intestinalis*, which lacks the N terminal helicase domain (MacRae *et al.*, 2006). The enzyme resembles an axe with tandem RNase III domains representing the blade and the PAZ domain representing the handle. The distance between the PAZ domains and the active site is 65 Å, which corresponds with the 25 nt products of *G. intestinalis* Dicer. Variations in the distances between these two sites could explain the variation in product sizes across species.

The helicase domain, which lies adjacent to the RNase III domains in three dimensional space, has emerged as an important domain for the cleavage of specific RNAs and may have developed specialized roles for different Dicer proteins (Figure 1.3). Humans and other mammals have only one Dicer and differences in the cleavage rates of pre-miRNA and pre-siRNA substrates are attributable to the helicase domain (Ma *et al.*, 2008; Chakravarthy *et al.*, 2010). Human Dicer cleaves pre-miRNAs much more efficiently than pre-siRNAs with the helicase domain inhibiting siRNA production. In *D. melanogaster*, where there are two Dicers, the processing of these two RNA substrates is segregated with Dicer-1 primarily responsible for cleaving pre-miRNAs and Dicer-2 cleaving siRNAs (Lee *et al.*, 2004b). Dicer-1 contains an inactive helicase domain, which facilitates binding to the loops of pre-miRNAs (Tsutsumi *et al.*, 2011), whereas Dicer-2 has an active helicase that allows for the processive cleavage of long dsRNAs (Cenik *et al.*, 2011; Welker *et al.*, 2010; Lau *et al.*, 2012). Finally, within *C. elegans* Dicer, the helicase domain appears to be important for the processing of siRNA for a dsRNA precursor, yet dispensable for the processing of pre-miRNAs into miRNAs. The helicase domain may have allowed Dicer to rapidly adapt to handle a diverse array of substrates and additional biochemical knowledge of the function of the helicase domain in the Dicer orthologs in other organisms could shed light on the evolution of this domain in the protein.

1.3.3 Partner proteins of Dicer

Similar to Drosha, Dicer does not function in isolation *in vivo*. Dicer partners with a family of dsRNA-binding proteins (dsRBPs) for both RNA processing and RNA loading into Argonaute. The family of dsRNA-binding proteins (dsRBPs) comprise one or more evolutionarily conserved dsRNA-binding domains (dsRBDs), each of 65-68 residues, found in eukaryotes, prokaryotes, and viral-encoded products (Ryter & Schultz, 1998). The dsRBPs interact exclusively with double-stranded RNA (dsRNA) in a non-sequence

specific manner (Ryter & Schultz, 1998; Manche *et al.*, 1992). The dsRBP family includes TAR RNA-binding protein (TRBP), protein activator of PKR (PACT), and in *Drosophila*, Loquacious and R2D2 (Doyle & Jantsch, 2003).

TRBP is one of the more studied RNAi dsRBPs. It was initially isolated from a HeLa cell expression library using TAR RNA as a binding probe (Gatignol *et al.*, 1991). After its role in activation of HIV-1 gene expression by binding TAR RNA was understood (Dorin *et al.*, 2003), TRBP was identified as an integral component of the RNA-induced silencing complex (RISC) (Chendrimada *et al.*, 2005). As an essential partner of Dicer, TRBP is required for optimal gene silencing induced by siRNA and miRNA (Gregory *et al.*, 2005). The diverse functions of TRBP are based on its ability to bind dsRNA, mediated by its dsRBDs. Therefore, it is crucial to understand how TRBP interacts with dsRNA as a first step toward deciphering its role in higher-order biological processes. Moreover, TRBP consists of three dsRBDs, where the first two N-terminal dsRBDs bind dsRNA and the third C-terminal dsRBD interacts with other partner proteins, but it remains unclear why two dsRBDs are required for interaction with dsRNA.

The dsRBPs also play crucial functions in iso-miR formation (Lee *et al.*, 2013). Iso-miR, a term coined by Morin et al. (2008), refers to the different miRNA variants that are generated *in vivo* from the same pre-miRNA precursor, as revealed by deep sequencing. Dicer cleavage can often lead to heterogeneous products with differences in the RNA length or 5'-or 3'-end identity (Morin *et al.*, 2008). These miRNA variants, or iso-miRs, can affect the effectiveness of RNAi by changing the mRNAs that they associate with (Cloonan *et al.*, 2011). Since the specificity of miRNA targeting is primarily determined by the seed region of the miRNA (nucleotides 2-8),small changes in the length or sequence can dramatically affect mRNA targeting (Lewis *et al.*, 2005).

Once processed by Dicer, the RNA duplexes are loaded into Argonaute by the RISC-loading complex, which consists of an Argonaute protein, Dicer and a dsRBD protein, primarily TRBP (MacRae *et al.*, 2008; Maniataki & Mourelatos, 2005). During this process, one strand will be selected as the guide stand and the other strand, the passenger strand, will be ejected or cleaved. The strand with its 5'-end at the less thermodynamically stable end of the duplex is selected as the guide (Schwarz *et al.*, 2003). TRBP and the helicase domain of Dicer are known to play key roles in choosing which strand becomes the guide. However, there are intrinsic preferences for Argonaute within the basic pocket of the MID domain (Noland *et al.*, 2011).

1.3.3 Argonaute

The guide RNA molecules in this pathway serve to recruit key protein components including the enzyme Argonaute (Ago), which functions as the catalytic engine of the RISC. Within the RISC, Argonaute accepts small RNA duplexes and selects one strand as the guide to mediate mRNA silencing by Ago-catalyzed target mRNA cleavage or translational silencing. Humans have four Argonautes (Ago1-4) associated with RNAi; however, only one, Ago2, is known to cleave target RNAs through its 'slicer' activity, whereas the other three silence mRNAs through translational repression followed by decay (Wilson & Doudna, 2013).

The Argonaute clade of proteins includes four structural domains: N-terminal (N), PAZ, MID, and PIWI. These domains form a bi-lobed architecture, a structure that helps explain how a guide RNA is recognized through two conserved sets of interactions

(Schirle & MacRae, 2012; Nakanishi *et al.*, 2012; Elkayam *et al.*, 2012). The 5'-phosphate and the first nucleotide of the guide strand are anchored by the MID domain, whereas the 3'-hydroxyl end is bound by the PAZ domain. The PIWI domain contains a catalytic tetrad of acidic residues that trigger the endonucleolytic cut in a target RNA (Nakanishi *et al.*, 2012). The human Ago1 and Ago4 proteins lack the intact catalytic tetrad, explaining why they lack RNA-cleaving activity. Surprisingly, however, both the catalytically-active Ago2 and the catalytically-inactive Ago3 possess these residues necessary for RNA cleavage, although only Ago2 is an active slicer enzyme. Using chimeric proteins, combining domains from human Ago2 and Ago3, it was found that the PIWI domain from Ago3, in the context of the remaining domains from Ago2, was found to be catalytically competent, confirming that the Ago3 PIWI domain does not lack intrinsic functionality(Hauptmann *et al.*, 2013; Schürmann *et al.*, 2013). This led to the hypothesis that Ago3 prevents RNA cleavage with additional inhibitory elements. Further chimeric proteins revealed two sequences in the N domain that inhibit the slicer activity of the Ago3 PIWI domain.

The same trend of inhibitory regions in the N domain is observed with human Ago1. Ago1 has a PIWI domain lacking the acidic residues, but once it is made catalytically active by the appropriate mutation, the N domain of Ago1 still prevents cleavage. The crystal structure of a bacterial Argonaute from *T. thermophilus* revealed that the N domain prevents proper base pairing with target RNA through steric clashes (Wang *et al.*, 2009). However, once this part of the N domain was deleted, corresponding to the first 106 residues of the protein, the *T. thermophilus* Ago was rendered completely catalytically inactive suggesting this domain could stabilize the active complex. In *D. melanogaster*, mutations to the N-terminal lobe also affected the activity of the PIWI domain (Hur *et al.*, 2013). Finally, another study using alanine scanning mutagenesis identified residues in the same region of Ago2 that could affect RISC assembly (Kwak & Tomari, 2012).

Although it is now clear that mutations distal to the catalytic center can dramatically change the activity of Argonaute, it remains unclear why there are four distinct Argonaute proteins in humans. In considering the evolution of the Ago proteins, it is notable that Ago4 is the most divergent of the four proteins and is not expressed in most human cell lines (Petri *et al.*, 2011). Beyond Argonaute's intrinsic behavior, the interactions between Argonaute and other proteins also contribute to gene silencing efficiency. Dramatic conformational rearrangements mediated by the chaperone machinery are already known to help create a mature RISC(Iwasaki *et al.*, 2010; Iki *et al.*, 2010). Future studies should reveal the properties of different Argonaute complexes and how they contribute to small RNA biology.

1.3.5 RNA induced silencing complex and translational repression

Processing downstream of target binding involves many of the protein components from the RNA degradation pathway. GW182 proteins, which bind directly to Argonaute, are a key component. These proteins recruit additional factors to the RISC and assist in localizing the complex to processing (P) bodies.

Currently miRNAs and their targets are difficult to predict *de novo* due to lack of conservation in the sequence of small RNAs and competing structural elements in their targets. Degradome sequencing, which detects RNAs in the process of being degraded, has helped to facilitate the discovery of novel miRNA–target RNA pairs. Degradome

sequencing takes advantage of the fact that for ligation during the library preparation step of Illumina sequencing, RNAs with a free 5'-monophosphate are required, which occurs naturally through cellular pathways of RNA decay. RNA intermediates of mRNA turnover can therefore be ligated and sequenced to find novel miRNA–target pairs by working in reverse from the cleaved transcripts (Addo-Quaye *et al.*, 2008; German *et al.*, 2008; Gregory *et al.*, 2008; Willmann *et al.*, 2013). In *A. thaliana,* where the miRNA binding sites are completely complementary to the target sites, strong signals are observed for sites of small-RNA guided cleavage (Willmann *et al.*, 2013). A complementary method for organisms where small-RNA guided cleavage sites are more challenging to determine, is Argonaute HITS-CLIP (high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation) and its variants (Chi *et al.*, 2009; Zisoulis *et al.*, 2010; Yang *et al.*, 2011; Li *et al.*, 2014); this approach can simultaneously isolate miRNAs and mRNA target sites bound by Argonaute proteins. Linking both methods with transcriptome-wide RNA structure predictions may help accurately predict new small RNAs and their targets.

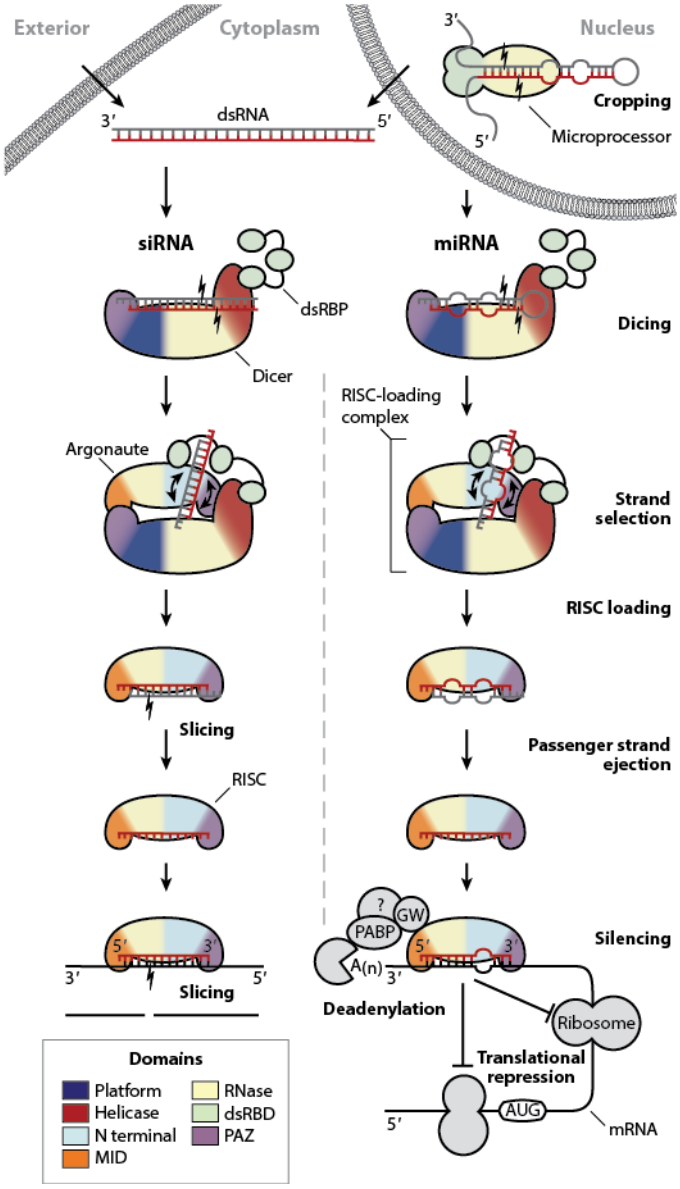

Figure 1.1 RNAi biogenesis pathways. The siRNA pathway is outlined on the left and the miRNA pathway is outlined on the right. The conserved domains of Drosha, Dicer, and Argonaute are colored according to the legend. Drosha, which contains tandem RNase III domains and a dsRBD, begins the miRNA pathways by cleaving pri-miRNAs into pre-miRNAs. The products are exported into the cytoplasm by Exportin-5 and are subsequently processed by Dicer with its partner dsRBPs. Dicer consists of an N-terminal DExD/H box RNA helicase, a DUF283 domain, a PAZ domain, tandem RNase III domains, and a C-terminal dsRBD. The mature products are loaded into Argonaute, which has a bilobed architecture and contains an N-terminal domain, PAZ domain, MID domain, and a PIWI domain. The siRNA precursors are often derived from exogenous sources and once they enter the cytoplasm, they enter a similar biogenesis pathway. Adapted from (Wilson & Doudna, 2013).

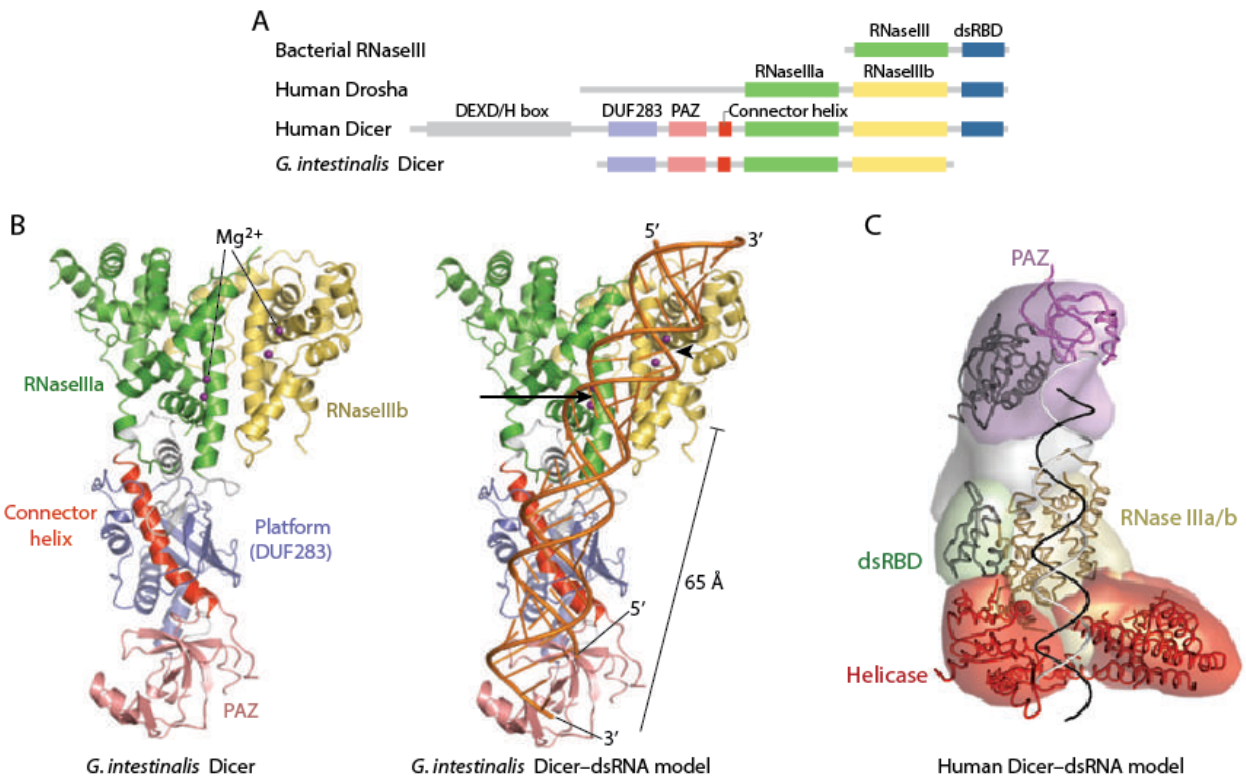

Figure 1.2 Domain architecture and structure of RNaseIII-family enzymes.(A) A schematic representation of the domain architecture of RNaseIII-family enzymes. (B) The crystal structure of *G. intestinalis* Dicer (left) and a model of a dsRNA substrate bound to *G. intestinalis* Dicer. In this model, the Mg^{2+} ions are shown in purple, indicating the sites of RNA cleavage. The distance between the 3' overhang in the PAZ domain and the active sites is 65 Å corresponding with the 25 nt products of Dicer cleavage. (C) Overall structure of Dicer. Homologous structures are docked onto a segmented electron microscopic map. The helicase domain is adjacent to the active sites and acts as a clamp to guide incoming RNA substrates. Adapted from (Jinek & Doudna, 2009) and (Wilson & Doudna, 2013).

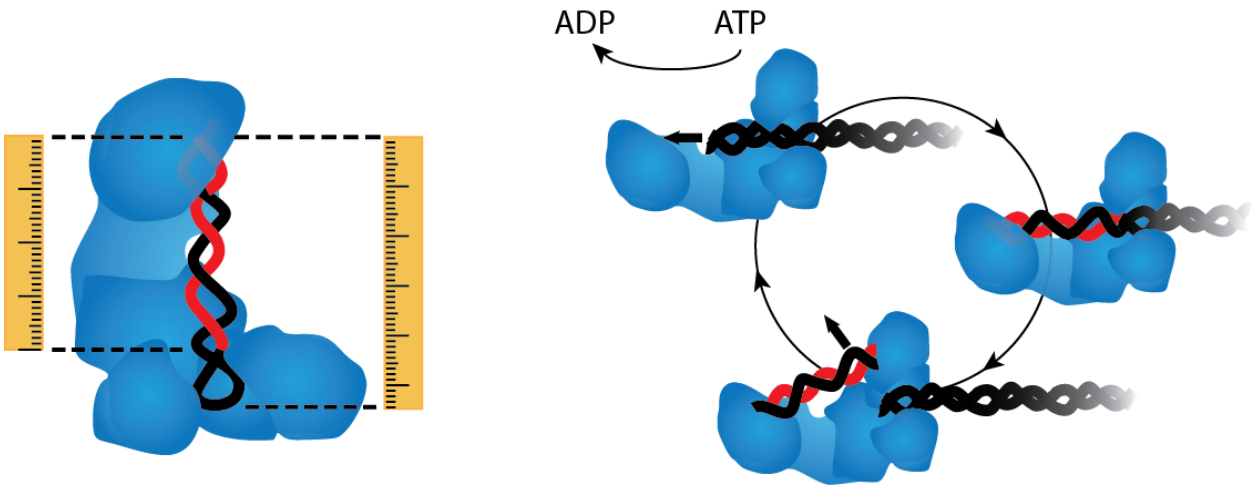

Figure 1.3 Two models for the function of the Dicer helicase domain. The helicase domain has emerged as an important domain for the cleavage of specific RNAs and may have developed specialized roles for different Dicer proteins. On the left is a model of Dicers with inactive helicase domains, such as Dicer-1 from *D. melanogaster*. The helicase domain interacts with the loop of pre-miRNAs and measures the distance from the 3' overhang to the single-stranded loop region. Once this requirement is satisfied, Dicer can cleave the RNA substrates by measuring the distance from the RNase III domains and the 3' overhang. On the right is a model of Dicers with active helicase domains, such as Dicer-2 from *D. melanogaster.* Dicer uses the energy from ATP hydrolysis to processively cleave an RNA substrate. The helicase domain acts as a clamp such that once the product is released, the remaining RNA is still bound by the protein.

Chapter 2

ATP-independent diffusion of double stranded RNA binding proteins

*A portion of the work presented in this chapter has been previously published as part of thefollowing paper:Koh HR, Kidwell MA, Ragunathan K, Doudna JA & Myong S (2013) ATP-independent diffusion of double-stranded RNA binding proteins. *Proceedings of the National Academy of Sciences***110(1):** 151-6.

*Hye Ran Koh, Mary Anne Kidwell, Jennifer Doudna, and Sua Myong designed research; Hye Ran Koh and Mary Anne Kidwell performed research; Kaushik Ragunathan contributed new reagents/analytic tools; Hye Ran Koh analyzed data; and Hye Ran Koh, Jennifer Doudna, and Sua Myong wrote the paper.

2.1 Introduction

The diverse functions of TRBP are based on its ability to bind dsRNA, mediated by dsRBDs. Therefore, it is crucial to understand how TRBP interacts with dsRNA as a first step toward deciphering its role in higher-order biological processes. Moreover, TRBP consists of three dsRBDs, where first two N-terminal dsRBDs bind dsRNA and the third C-terminal dsRBD interacts with other partner proteins, but it remains unclear why two dsRBDs are required for interaction with dsRNA. Using single molecule fluorescence detection, we sought to determine the fundamental action of TRBP upon dsRNA binding and how it may contribute to the biological function of its partner protein, Dicer when in a complex. Here, we report an unanticipated dsRNA diffusion behavior of TRBP which requires two dsRBDs, and its correlation with the role of TRBP in promoting Dicer-induced RNA cleavage. Furthermore, the diffusion activity is conserved in orthologous members of tandem dsRBDs-containing proteins indicating that diffusion may be a general mechanism by which dsRBPs mediate diverse biological processes.

Such functional diversity is not surprising in light of the majority of cellular and viral RNA species which likely exist in highly dsRNAs, primarily by forming secondary RNA structures (Saunders & Barber, 2003). However, it remains puzzling why most dsRBPs require multiple dsRBDs as truncation or mutation of one dsRBD greatly diminishes or completely abolishes the protein's ability to bind dsRNA (Schmedt *et al.*, 1995; Krovat & Jantsch, 1996) and their biological function (Micklem *et al.*, 2000). Here, we focused our study on the mechanism of the simplest multiple dsRBDs-containing proteins, TRBP and its homologs, PACT and Loquacious (Loqs)-PB (also known as R3D1-L).

2.2 Materials and Methods

2.2.1 Protein Purification

N-terminal His_6-Dicer was purified as previously described (MacRae *et al.*, 2008)with several modifications. After the protein was eluted from the Ni-nitrilotriacetic acid (NTA) resin (Qiagen), the N-terminal His_6-tag protein was not cleaved, and instead the protein was dialyzed overnight into a low salt buffer [100 mM KCl, 10% (vol/vol) glycerol, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), and 20 mM HEPES, pH 7.5]. The concentrated protein was applied to HiTrap 5 mL Q HP column (GE Healthcare) with a 0–800 mM KCl gradient. The peak material, which eluted at about 400 mM KCl, was concentrated to <1 mL and applied to a Superdex 200 16/60 column (Amersham Pharmacia) equilibrated in gel filtration buffer [150 mM KCl, 10% (vol/vol) glycerol, 1 mM TCEP, and 20 mM HEPES, pH 7.5]. Fractions containing nonaggregated protein were pooled, concentrated to ~3 mg/mL, and used in subsequent reconstitution experiments. All purification steps were carried out at 4 °C. All protein concentrations were determined by the Nanodrop (Bio-Rad).

TRBP, PACT, and their truncations were cloned as cleavable N-terminal His6-MBP fusions. Each was purified separately from bacterial overexpression using methods previously described (MacRae *et al.*, 2008). One truncation, TRBP dsRBD2 and -3, would stick to the Ni-NTA resin. Therefore, this protein was applied to the Ni-NTA column with a 100–800 mM Imidazole gradient, with peak material eluting at about 500 mM Imidazole.

R3D1 was cloned as a cleavable N-terminal His_6-GST fusion and overexpressed in bacteria. The protein was purified as previously described (MacRae *et al.*, 2008)with the NTA column being replaced by a 1 mL GSTrap Column (GE Healthcare).

2.2.2 Dicer-TRBP Reconstitution

705 µg (3 nmol) of Dicer and 550 µg (11 nmol) of TRBP were mixed in 250 µl (final volume) of gel filtration buffer. The protein solution was incubated on ice for 1 hour and then was applied to a Superose 6 10/30 column (Amersham Pharmacia) equilibrated in gel filtration buffer. Fractions were analyzed by SDS/PAGE, and those fractions containing the complex were pooled and concentrated to 1.5 mg/ml.

2.2.3 RNA Constructs Preparation, Labeling, and Annealing

The sequences of all RNA constructs were displayed in Table S1. *25RNA*, *40RNA* and *55RNA* were purchased from Dharmacon, and 3'-biotin or 3'-DY547 was incorporated in the process of each RNA synthesis. RNA constructs containing 5' triphosphate were *in vitro* transcribed and biotinylated at 3' end. All other RNA constructs were synthesized and HPLC-purified from IDT with proper chemical modifications of each RNA such as an internal C6 amine modifier dT, biotin or/and Cy3 incorporation at 5' or/and 3' end.

Cy3/Cy7 NHS ester and Alexa 647 NHS ester were purchased from GE Healthcare and Invitrogen, respectively. Excess dyes were mixed together with RNA containing amino modifier C6 dT in 100 mM NaHCO3, pH 8.5, and incubated overnight for the RNA labeling. Residual dyes were removed by Biorad P-6 column or ethanol precipitation. Labeling efficiency was higher than 90% for all labeling reactions. For generating dsRNA from each labeled or nonlabeled ssRNA, each strand of dsRNA was mixed with its complementary one in an annealing buffer (10 mM Tris, pH 8, and 100 mM NaCl), heated at 90°C for 2 min, and slowly cooled down to room temperature. For single-molecule FRET assay, we labeled RNAs and proteins with a fluorophore.

With *pre-siRNA*, we showed the diffusion activity by TRBP and its truncation mutants by 2-color FRET as well as 3-color FRET. *siRNA*, *25RNA*, *presiRNA*, *40RNA*, and *55RNA* were for investigating the length dependence of TRBP's diffusion activity, and *DNA:RNA*, *50ssRNA*, *DsrA*, *rpoS1*, *rpoS2*, and *prelet-7_n* for the substrate structure dependence. For the cleavage assay by Dicer-TRBP complex, non-biotinylated *prelet-7_n* was examined.

2.2.4 Protein Labeling

Protein and Cy3 NHS ester or Alexa 647 NHS ester were mixed together in 100 mM $NaHCO_3$, pH 8.5, buffer and incubated for 30 min at room temperature. The ratio between protein and dye was adjusted to achieve a 1:1 ratio of dye–protein by trial and error. N-terminal–specific labeling was done by an enzymatic reaction of Acp Synthase (New England Biolabs) using A1- tagged TRBP and CoA-modified fluorophores. Biorad P-6 column was used to remove the excess dyes, and then the labeling efficiency was calculated by measuring UV-VIS absorbance.

2.2.5 Immobilization of RNA and Protein

For the single-molecule FRET assays, RNAs and proteins were immobilized onto a poly ethylene glycol (PEG)-coated quartz surface through neutravidin-biotin interaction. 0.05 mg/ml of neutravidin was incubated for 5 mins and then 30-50 pM biotinylated RNA in T50-BSA buffer (10mM Tris, pH8, 50mM NaCl, and 0.1mg/ml BSA) was added for RNA immobilization. 10-100 nM of protein was added onto the immobilized RNA for the diffusion assay. For the RNA cleavage assay, N-terminal His_6-tag protein was immobilized onto a neutravidin-treated PEG surface through biotinylated anti-Penta-His antibody (Qiagen). 1-2 ug/ml of the antibody was incubated for 5 mins, following the incubation of 0.05 mg/ml neutravidin for 5 mins on a PEG surface. Then, 10 nM of His_6-tag protein was incubated for another 5mins, and 1-2 nM of RNA was added afterward for the single-molecule assays.

2.2.6 Single-molecule FRET Assays

Single-molecule FRET detection was achieved by a homemade wide-field TIRF microscopy. Single fluorophores immobilized on a PEG-coated quartz surface emit light by an excitation of a solid-state 532 nm laser (Spectra physics). The fluorescent emission was collected through a water immersion Olympus objective (60×, Numerical Aperture = 1.2), passed through 555 nm long pass filter for rejection of laser's Rayleigh scattering, separated by two, a green and red emission, at a dichroic mirror (cutoff, 630 nm), and detected by an EMCCD camera (Andor). For three-color FRET measurement, two sets of dichroic mirrors (cutoff, 630 nm and 725 nm) were used to separate the emission by three, the one from Cy3, Cy5, and Cy7, respectively. The exposure time was 30 ms, and the detected fluorescent signals were analyzed with a custom-written IDL and MATLAB program.

An oxygen scavenger system [0.5% (wt/vol) glucose, 100 mg/mL glucose oxidase (Sigma), and 8.8 kU/mL catalase (Calbiochem)] was used together with 5–10 mM trolox (Sigma) to stabilize fluorophore during a single-molecule data acquisition. TRBP sliding assay was performed in 20 mM Tris, pH 7.5, 25 mM NaCl, 1 mM DTT, and 0.1 mg/mL BSA in combination with the oxygen scavenger system. A total of 2 mM $MgCl_2$ was added to initiate an RNA cleavage by Dicer–TRBP.

2.2.7 Data Analysis: Auto-correlation and Cross-correlation

The auto-correlation function was analyzed by MATLAB program, and the equation of auto-correlation used is like below:

$$G(\tau) = \int \mathrm{E(t)} \cdot \mathrm{E(t\text{-}\tau)dt}$$

E(t) represents FRET and the FRET auto-correlation, G(τ), gives us an average diffusion rate because FRET changes in our system reflect the diffusion of a protein. More than 30 molecules were analyzed for each auto-correlation curve, and it was fitted with a single exponential decay to obtain an average diffusion rate of the molecules. In 3-color FRET experiment, the cross-correlation between the fluorescence intensity of Cy5 and the one of Cy7 was analyzed using the following equation:

$$XC(\tau) = \int \mathrm{I_{Cy5}(t)} \cdot \mathrm{I_{Cy7}(t\text{-}\tau)dt}$$

2.3 Results

2.3.1 Observation of TRBP Diffusion on dsRNA

TRBP possesses three dsRBDs, the N-terminal two of which (dsRBD1 & 2) bind dsRNA tightly (Yamashita *et al.*, 2011), whereas the third dsRBD (dsRBD3) participate in higher-order complex assembly with proteins such as PKR and Dicer (Haase *et al.*, 2005; Parker *et al.*, 2008). To examine the interaction between TRBP and dsRNA by single molecule fluorescence resonance energy transfer (smFRET) (Myong *et al.*, 2005), we prepared a donor (Cy3, green)-labeled dsRNA and an acceptor (Alexa 647, red)-labeled TRBP. We immobilized the RNA via biotin-neutravidin linkage and added the labeled TRBP to single molecule imaging surface (Figure 2.1A). Unexpectedly, we observed rapid FRET fluctuations caused by distance changes between TRBP and one end of the dsRNA (Figure 2.1B). The FRET fluctuation doesn't come from repetitive TRBP binding and dissociation because the FRET fluctuates between 0.3 and 0.8, rather than going down to 0 as expected if dissociation is occurringas shown in Figure 2.2. This repetitive distance change could result either from a diffusion of TRBP on dsRNA or a conformational change of TRBP's subdomain. To differentiate between the two possibilities, we labeled TRBP site-specifically at its N- terminus which is expected to contact dsRNA tightly. This protein also yielded FRET fluctuations highly analogous to the non-specifically labeled TRBP shown in Figure 2.1B (Figure 2.3). The similarity in FRET fluctuation regardless of Cy3 labeling position implies that the FRET fluctuations are likely due to the movement of the whole TRBP, rather than its subdomain's motion relative to the static dsRNA-bound domain. Furthermore, the initial FRET values are heterogeneous due to the binding of TRBP in a non-sequence specific manner (Figure 2.3).

To test further whether the observed FRET fluctuation could arise from diffusion of TRBP along dsRNA, we performed one color protein induced fluorescence enhancement (PIFE) (Hwang *et al.*, 2011) (Figure 2.4) and three color FRET (Hohng *et al.*, 2004) assays (Figure 2.5). In the PIFE assay, the intensity fluctuation of Cy3 in the absence of Cy5 indicates that TRBP comes in contact with the fluorophore repeatedly. In three color FRET assay, the anti-correlated changes between the two acceptors (Cy5 and Cy7) located at both ends of dsRNA indicates that the donor (Cy3)-labeled TRBP is moving across the dsRNA from one end to the other in succession. Again, we rule out the possibility of TRBP association and dissociation because the anti-correlated change between the two acceptor dyes (Cy5 and Cy7) can only result from the continuous and periodic movement of the donor labeled TRBP from one end to the other end of dsRNA axis. Both assays independently support the conclusion that TRBP diffuses on dsRNA. It is noteworthy that this activity was ATP independent, indicating that the diffusion of TRBP does not require an external energy source. Furthermore, this activity likely arises from a monomer of TRBP rather than dimer based on the observation that the same activity persists even after removal of excess protein by buffer wash. In addition, the majority of the initial binding and dissociation of TRBP (95% labeled) yield single step fluorescence increase and decrease corresponding to TRBP monomer In addition, the electophoretic mobility shift assay displayed a clear band which corresponds to a monomer TRBP-RNA complex in our reaction condition of 10nM TRBP. TRBP dimer-RNA complex is only seen when TRBP concentration reaches 200 nM. The majority of TRBP (>90%) in 10-50nM range binds to dsRNA as a monomer, suggesting that the diffusion behavior we detected at 10 nM TRBP concentration is likely due to TRBP monomer (Figure 2.6A). Also, we observed that the fluorophore labeling of TRBP didn't

affect its binding affinity to dsRNA (Figure 2.6B). Here, we use the term “diffusion” to ascribe 1-D diffusion of TRBP on dsRNA (Blainey *et al.*, 2006).

2.3.2 dsRNA Length Dependence of TRBP Diffusion

We tested whether the length of the dsRNA could modulate the diffusion activity of TRBP. We prepared dsRNAs of 19, 25, 38 and 55 base pairs (bp) each labeled with a donor at one end. TRBP diffusion on 38 bp (Figure 2.7A) results in larger amplitude of FRET change (Figure 2.7B) compared to that observed using 19 bp (Figure 2.7 C and D), suggesting that TRBP diffuses along the entire length of the dsRNA. To demonstrate this hypothesis, we performed an autocorrelation analysis of the FRET fluctuations obtained for all four dsRNAs of different lengths (Figure 2.7E). As a way of signal processing, the autocorrelation gives the magnitude of signal change as its initial value, and the rate of signal change as its decay time. The initial values of the autocorrelation curve, 0.07, 0.12, 017 and 0.20, were obtained for 19, 25, 38 and 55 bp dsRNA, respectively, indicating that a longer dsRNA displayed larger distance changes as expected. We also confirmed the result by comparing the average FRET histograms obtained from 19 and 55 bp dsRNAs which showed a shift towards the low FRET state in the case of longer dsRNA (Figure 2.8). Furthermore, the rates calculated from an exponential fitting of each autocorrelation curve, 0.106, 0.117, 0.143 and 0.190 sec of diffusion time for 19, 25, 38 and 55 bp dsRNA, respectively, indicates that it takes longer time to diffuse on the longer dsRNA (Figure 2.7F).

Taken together, our data demonstrate that the observed FRET fluctuation is due to TRBP’s diffusion movement along the entire RNA rather than a conformational change of TRBP’s subdomain.

2.3.3 TRBP Diffuses Exclusively on dsRNA

We varied the composition of RNA substrates to investigate the substrate specificity of TRBP diffusion (Figure 2.9A-F). While TRBP showed approximately 60% binding to 38-bp dsRNA, it did not show any binding or movement on a RNA:DNA hybrid or a single-stranded RNA, indicating that both strands of dsRNA are required for TRBP binding (Figure 2.9G and H). It also indicates that A-form structure of nucleic acids is not sufficient for TRBP’s binding because TRBP didn’t bind to RNA:DNA heteroduplex exhibiting A-form duplex structure. Consistent with this possibility, TRBP diffusion behavior was observed only when bound to dsRNA in which most or all nucleotides are predicted to form canonical Watson-Crick base pairs, and not to dsRNAs with more complex secondary structure such as bulges and loops (Figure 2.9H).

2.3.4 Two dsRBDs are Responsible for Diffusion Activity

We generated truncation mutants of TRBP to investigate which of three dsRBDs give rise to the diffusion activity. Based on the dissociation constants of individual dsRBD1 (220nM) and dsRBD2 (113nM) to dsRNA which are three orders of magnitude higher than that of dsRBD1+dsRBD2 (0.25nM) (Yamashita *et al.*, 2011), it is clear that both dsRBD1 and dsRBD2 are required for an efficient binding to dsRNA. In addition, both dsRBD1 and 2 play functionally important role with respect to RNA processing (Chendrimada *et al.*, 2005; Daviet *et al.*, 2000) whereas dsRBD3 serves as a connector

to Dicer (Haase *et al.*, 2005; Parker *et al.*, 2008; Gatignol *et al.*, 1993). Consistent with these previous findings, we find that the deletion of dsRBD3 did not interfere with the observed diffusion behavior, whereas the deletion of dsRBD1 or dsRBD2 completely abrogated the diffusion activity (Figure 2.10A-E). Therefore, we conclude that the intact dsRBD1 and dsRBD2 are directly responsible for the diffusion movement of TRBP and that dsRBD3 is dispensable. This result also implies that the diffusion activity requires intact tandem dsRBD1 and 2, and one dsRBD alone is insufficient. In addition, it supports that the diffusion activity is a characteristic of TRBP monomer because dsRBD3-truncated mutant (dsRBD1+dsRBD2) wouldn't form a dimer due to the lack of dsRBD3, the protein interaction domain. We hypothesize that one dsRBD's binding affinity is not strong enough to show diffusion whereas the affinity of two dsRBDs is sufficient to support this activity. However, the three orders of magnitude difference in the binding affinity between one dsRBD and two dsRBDs (dsRBD1+2) suggests that it is not due to a simple summation of two independent binding bodies but a new conformational state of two dsRBDs which contributes to non-additive binding affinity to dsRNA as well as the unique diffusion activity along dsRNA.

2.3.5 Conservation of Diffusion Activity in PACT and R3D1-L

We tested if this diffusion activity is conserved in orthologous proteins with tandem dsRBDs. For this, we purified PACT, another partner protein of human Dicer (Lee *et al.*, 2006) as well as an activator of PKR, and R3D1-L-L, a cofactor of Dicer-1 found in *Drosophila* (Jiang *et al.*, 2005; Saito *et al.*, 2005). Both of these proteins possess three dsRBDs analogous to those in TRBP. In smFRET experiments similar to those described for TRBP previously, both PACT and R3D1-L exhibited behavior similar to that of TRBP and consistent with diffusion on dsRNA substrates (Figure 2.11). The finding that two additional proteins possess dsRNA diffusion activity strongly suggests that this behavior may be an intrinsic feature of these types of RNA-binding proteins.

2.3.6 TRBP-driven diffusion of Dicer-TRBP complex

We asked if TRBP diffusion contributes to the function of its partner protein, Dicer. To test whether TRBP diffusion occurs in the context of a Dicer-TRBP complex, we analyzed a reconstituted sample of Dicer-TRBP in smFRET assays in which Dicer-TRBP and dsRNA substrates were labeled as before (Figure 2.12A). Upon association with dsRNA, which is detected by the appearance of FRET, some Dicer-TRBP binding events produced a constant FRET signal whereas others produced a fluctuating FRET signal, consistent with two populations of complexes exhibiting either static binding or diffusion on dsRNA, respectively (Figure 2.12*B*). The molecules exhibiting static binding are not likely due to Dicer alone because Dicer's binding affinity to dsRNA is at least ten times lower than TRBP as quantified by fluorescence spot numbers. In these assays, we found that about 60 % of the Dicer-TRBP complex also exhibits smFRET fluctuation pattern which is consistent with diffusion on dsRNA, although almost none (< 5%) of Dicer alone and most of (>85%) of TRBP alone does (Figure 2.12*C*), suggesting that TRBP induces the diffusion of Dicer-TRBP complex.

2.3.7 TRBP-mediated Diffusion Correlates with Dicer-induced RNA Cleavage

Next, we tested dicing activity of dsRNA by Dicer-TRBP complex. To minimize possible surface effects that may interfere with the access of the Dicer-TRBP complex to dsRNA, we performed a complementary experiment in which the protein complex, rather than the dsRNA, was immobilized (Figure 2.12*D*). The protein complex labeled with red fluorescent dye (Alexa 647) nonspecifically was surface-immobilized using an antibody against the Histidine$_6$-tag (Jain *et al.*, 2011) and green dye (Cy3)-labeled RNA substrate was added, generating Dicer-TRBP-RNA ternary complex. Labeling position in Dicer-TRBP complex can be either in Dicer or TRBP, but either case will lead to FRET fluctuation regardless of the labeling position if the protein complex moves along dsRNA. All immobilized molecules should be Dicer-TRBP complex rather than TRBP or Dicer alone. First, they cannot be TRBP alone because the protein complex was immobilized via histidine tag on Dicer. Second, they cannot be Dicer alone because the binding affinity of Dicer to the RNA substrate is too low to pull down dsRNA. The K_d for Dicer is approximately 1.8 nM, whereas, K_d for Dicer-TRBP is less than 50 pM (Chakravarthy *et al.*, 2010).

The RNA cleavage activity was monitored by the loss of fluorescent spots on the surface over time after the addition of Mg^{2+}. The tested cleavable dsRNA contains the sequence of the well-characterized pre-miRNA, pre-let7a, but with a nick in the middle of its hairpin loop, which is predicted by mFold to retain the same structure (Zuker, 2003). The cleavage rate of Dicer-TRBP on this nicked pre-miRNA was consistent with that observed in a bulk cleavage assay (Chakravarthy *et al.*, 2010) with $t_{1/2}$ of ~ 3 minutes (Figure 2.12*E*). Further analysis of individual single molecule traces revealed both static binding and dynamic diffusion population of molecules. We counted these two classes of molecules over the time interval corresponding to active RNA cleavage after Mg^{2+} addition.

Surprisingly, our data indicate that the diffusing molecules, but not the static molecules, are selectively lost (Figure 2.12*F*). The calculated ratio of the diffusion molecules over both static and diffusion molecules exhibited the disappearance rate following the same kinetic rate of dsRNA cleavage (Figure 2.12*G*). There are two possible explanations for the two different populations of Dicer-TRBP complex. Our RNA substrate is asymmetric because only one end of RNA possesses the 2-nt overhang, which PAZ domain of Dicer needs to latch on to cleave the RNA substrate at 21-23 bp away from the end. If Dicer binds in the opposite orientation, Dicer cannot cleave the dsRNA because PAZ domain cannot function as a ruler to induce the RNase's cleavage. The wrong orientation effect may explain the static population that didn't get cleaved over time. Another possibility is that there are two types of associations between Dicer and TRBP, one that inhibits diffusion of the complex and the other that permits it. In this model, the inhibitory association will not allow cleavage unless the binding occurs in the correct position where PAZ domain engages with the 2-nt overhang. In contrast, the diffusing molecules can slide along dsRNA to search for the cleavage site. Both scenarios support that only the diffusing Dicer-TRBP complex cleaves RNA substrate, suggesting that TRBP-driven diffusion may play a role in enhancing the cleavage rate of Dicer.

2.4 Discussion

Herein, we present the first case of a protein family which diffuses exclusively on dsRNA. Such movement can only be detected by real-time monitoring of individual proteins. It is interesting to note that TRBP, PACT and R3D1-L which contain tandem dsRBDs possess the intrinsic ability to diffuse on dsRNA, whereas Dicer with only one dsRBD shows transient static binding without diffusion on dsRNA. The diffusion activity seen in three dsRBPs here may represent a general mode of multiple dsRBDs. Diffusion activity tested on TRBP truncation mutants also confirms that one dsRBD is insufficient for its diffusion. The substrate specificity shown in Figure 2.9 in which TRBP diffusion was observed in dsRNA with internal mismatches but not with more complex secondary structures may serve as a basis for its RNA scanning function. It is noteworthy that miRNAs also contain internal mismatches which TRBP can diffuse on.

Reduced dimension in diffusion has been suggested to increase the efficiency of protein-nucleic acid interaction. The association and dissociation kinetics of *Lac* repressor protein showed that it searches for its target not just by a three-dimensional random collision but a 1-D diffusion (Riggs *et al.*, 1970; Winter *et al.*, 1981), which was supported by theoretical approaches (Berg & Ehrenberg, 1982; Berg *et al.*, 1981) Recently, single molecule studies have identified proteins that diffuse on nucleic acids. Rad51 diffuses laterally on double stranded DNA (dsDNA) (Graneli *et al.*, 2006), DNA glycosylase 1 (hOgg1) diffuses on dsDNA by one-dimensional diffusion (Blainey *et al.*, 2006), reverse transcriptase of HIV shuttles back and forth on RNA:DNA hybrid (Liu *et al.*, 2008), single stranded DNA binding protein diffuses on single stranded DNA (Roy *et al.*, 2009) and DNA repair protein, Msh2-Msh6 also diffuses on dsDNA (Gorman *et al.*, 2010). Herein, we present the first case of a protein family which diffuses exclusively on dsRNA. This motion is ATP-independent unlike RIG-I which translocates on dsRNA fueled by ATP (Myong *et al.*, 2009; Kowalinski *et al.*, 2011; Luo *et al.*, 2011; Jiang *et al.*, 2011). It is interesting and intriguing to note that TRBP, PACT and R3D1-L, which possess the intrinsic ability to diffuse on dsRNA, are multiple dsRBDs-containing proteins as well as cofactors of Dicer in different organisms.

This diffusion activity arising from multiple tandem dsRBDs, could at least in part, unravel the unsolved biological questions regarding the role of dsRBPs. *In vitro*, TRBP facilitates Dicer's cleavage of pre-siRNA or pre-miRNA (Haase *et al.*, 2005; Chakravarthy *et al.*, 2010), but how it contributes to the function of Dicer has remained unknown. Based on the correlation between TRBP-driven diffusion of Dicer-TRBP complex and its cleavage activity, we propose that TRBP-induced diffusion could aid in positioning Dicer at the proper cleavage site (Figure 2.12F), resulting in an enhanced cleavage rate of Dicer-TRBP compared to Dicer alone (Haase *et al.*, 2005; Chakravarthy *et al.*, 2010). Dicer by itself will find the cleavage site through multiple trials of random binding and dissociation while Dicer-TRBP complex can find it more efficiently through diffusion mechanism (Figure 2.12F). In other words, Dicer-TRBP diffusion can serve to scan pre-miRNA and pre-siRNA substrates and to facilitate Dicer's catalytic activity by locating the complex at the RNA cleavage site with an improved efficiency and precision (MacRae *et al.*, 2006; MacRae *et al.*, 2007). It is also likely that TRBP's diffusion plays another role in RNA-mediated gene silencing, for instance, a recruitment of Dicer-TRBP complex to Ago2 or guide strand selection by dsRNA scanning (Noland *et al.*, 2011; Gredell *et al.*, 2010).

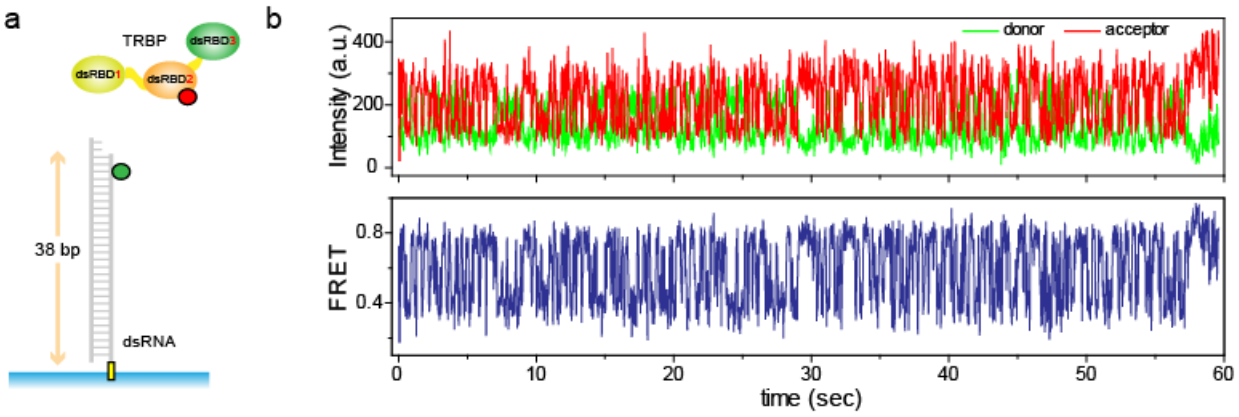

Figure 2.1 TRBP's interaction with dsRNA at the single-molecule level. (A) Alexa 647(red)-labeled TRBP was added to an immobilized Cy3(green)-labeled dsRNA, and their interaction was visualized by TIRF microscopy. (B) Repetitive FRET fluctuation was observed at the single molecule level without TRBP dissociation from dsRNA, reflecting a repetitive distance change between TRBP and the end of dsRNA.

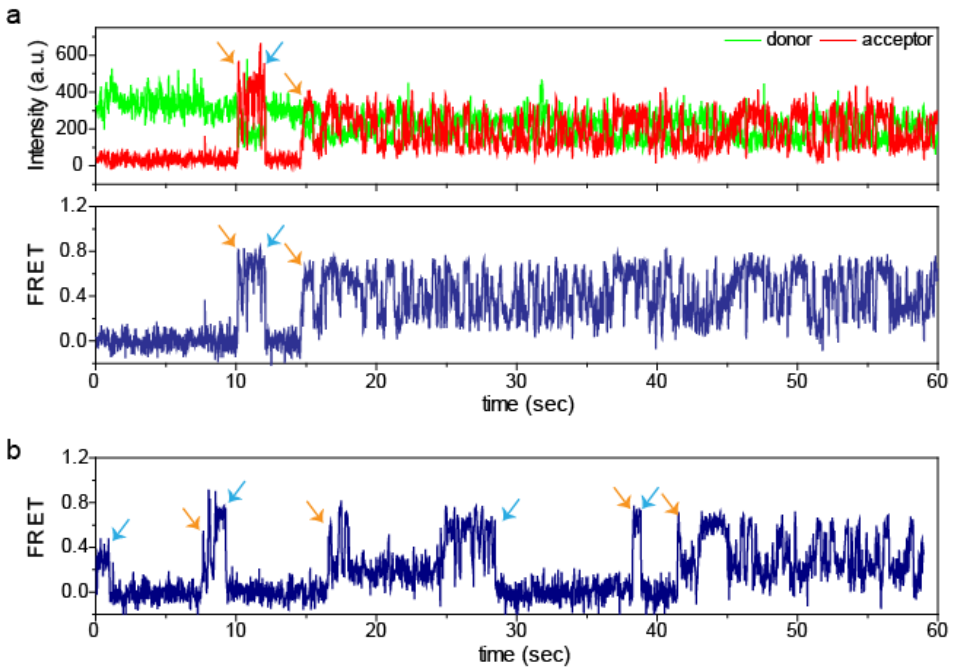

Figure 2.2 Multiple binding and dissociation of TRBP while sliding on dsRNA. (A and B) Single-molecule time traces show TRBP's binding (displayed by an orange arrow) and dissociation (blue arrow) as well as its repetitive FRET change at its dsRNA-bound state, demonstrating that the observed FRET changes aren't caused by rapid binding and dissociation of TRBP.

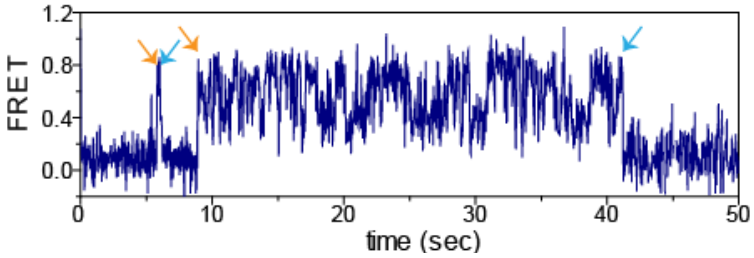

Figure 2.3 Interaction of N-terminal–labeled TRBP with dsRNA. N-terminal–labeled TRBP using A1-tag–engineered TRBP, CoA-647, and ACP synthase showed a similar FRET trace with the randomly labeled TRBP as shown in Fig. 1. The similar FRET change regardless of a dye position suggests that it is likely due to the movement of an entire TRBP along dsRNA rather than a specific subdomain's conformational change.

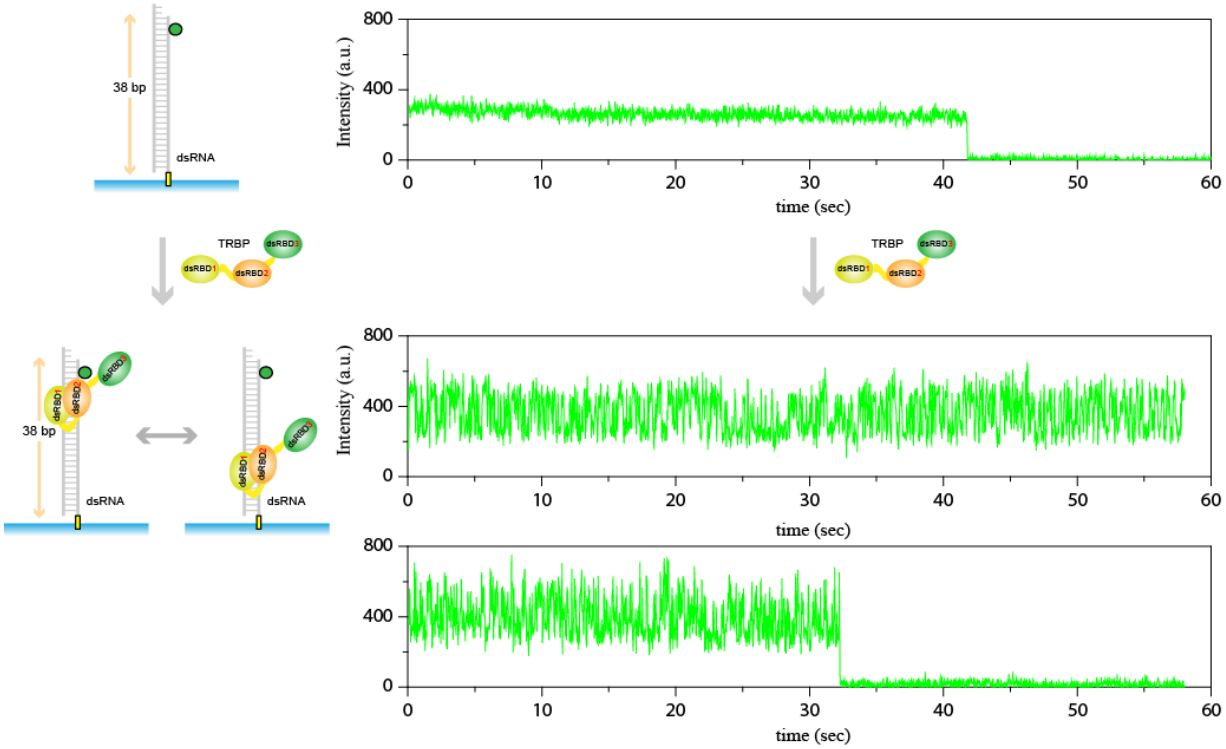

Figure 2.4 PIFE (one-color) visualization of TRBP's diffusion on dsRNA. Repetitive Cy3 intensity fluctuations were obtained by adding unlabeled TRBP to Cy3-labeled dsRNA. The intensity of Cy3 can increase only by a proximal protein contact, so the rapid intensity changes are consistent with the TRBP's repetitive sliding along dsRNA as seen in Figure 2.1.

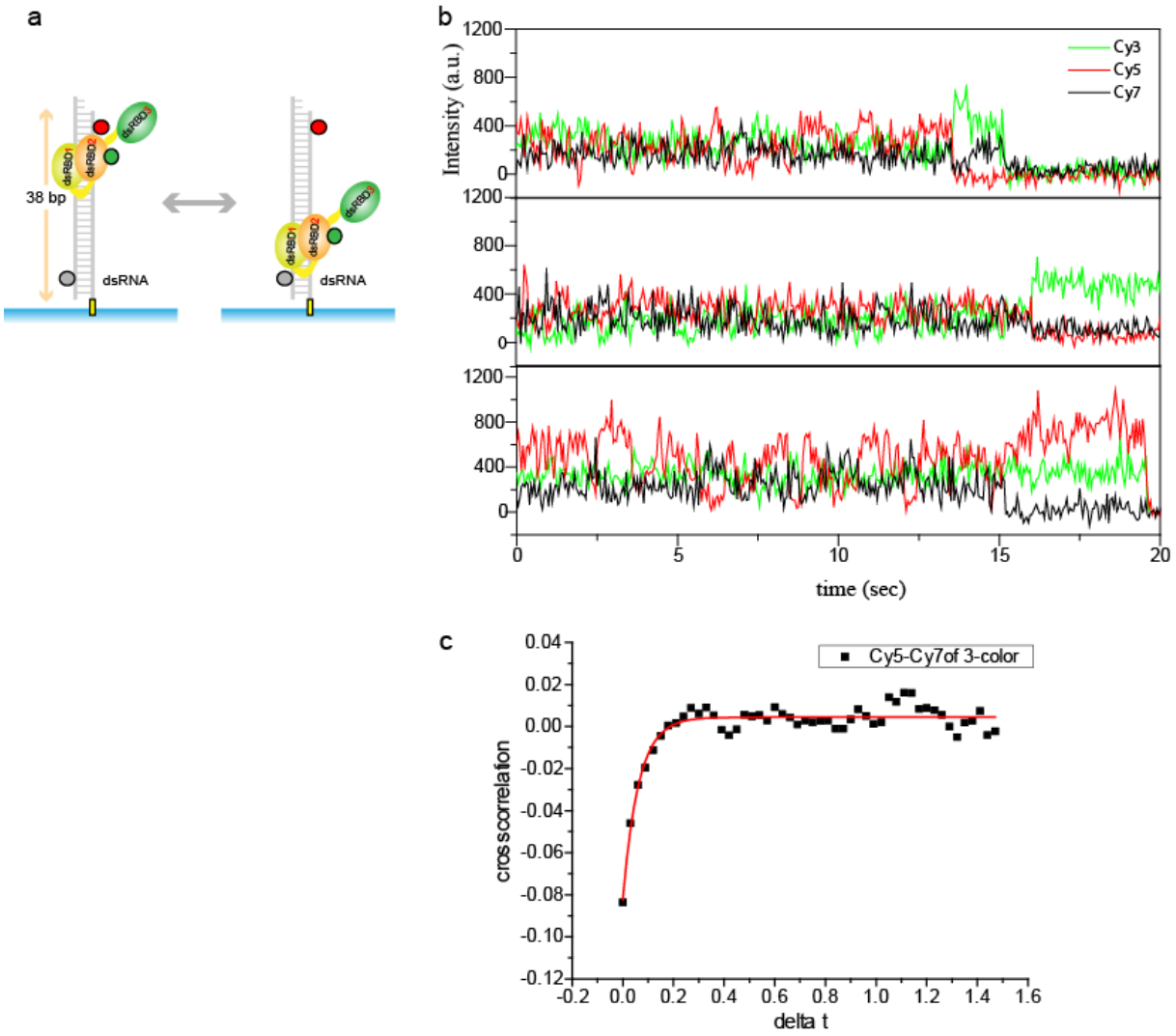

Figure 2.5 Three-color visualization of TRBP's diffusion on dsRNA. (A) Cy3-labeled TRBP was added to dsRNA labeled with Cy5 and Cy7 at each end for the three-color FRET assay. (B) Anticorrelation between Cy5 and Cy7 intensity was observed, supporting TRBP's sliding along dsRNA. An enlarged view of Cy5 and Cy7 intensity was displayed to show the anticorrelation clearly. (C) Cross-correlation analysis on Cy5 and Cy7 intensity generated a curve fitted to an exponential rise, further demonstrating the anticorrelation between Cy5 and Cy7 intensity.

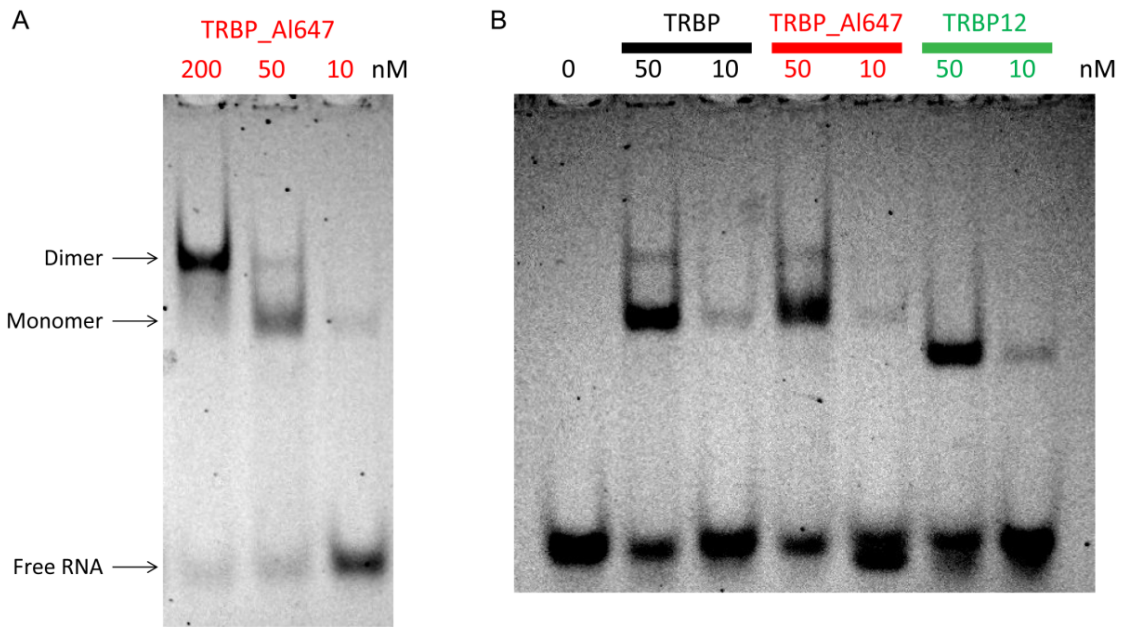

Figure 2.6 TRBP binds to dsRNA by a single event at 10 nM TRBP concentration. (A) An electrophoretic mobility shift assay using Alexa 647–labeled TRBP (named as TRBP_Al647) and 10 nM Cy3-labeled 19 bp dsRNA exhibited one TRBP–RNA complex at 10 nM TRBP concentration and two TRBP–RNA complexes at higher TRBP concentration. Even at 50 nM TRBP concentration, ~90% of TRBP binds to dsRNA as a single complex. (B) We compared nonlabeled TRBP, TRBP_Al647, and dsRBD3-truncated TRBP mutant (named as TRBP12). Alexa 647 labeling of TRBP didn't affect its binding affinity to dsRNA, and dsRBD-truncated TRBP mutant showed a single TRBP–RNA complex.

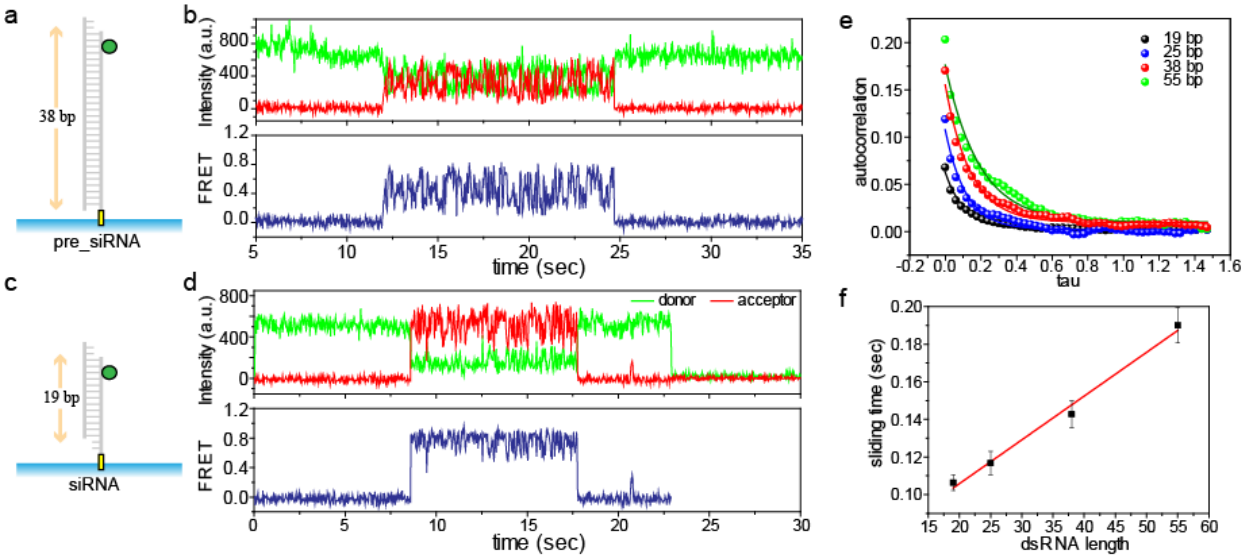

Figure 2.7 TRBP diffuses on the entire length of dsRNA. (A–F) TRBP diffuses on dsRNA in a length-dependent manner. (A–D) Diagrams of RNA substrates tested (38 bp dsRNA and 19 bp dsRNA with 2 nt 3′ overhang) and their representative FRET traces. Larger FRET changes were observed with a longer dsRNA. (E) Autocorrelation analysis on FRET signal using dsRNAs with four different lengths: 19, 25, 38, and 55 bp. The initial value of the autocorrelation curve (0.07, 0.12, 0.17, and 0.20 for 19, 25, 38, and 55 bp dsRNA, respectively) indicates a larger distance change with a longer dsRNA. More than 100 TRBP diffusion events were analyzed for each dsRNA. (F) Diffusion time (±SD) calculated by an exponential fit to an autocorrelation curve as shown in Fig. 2E displays a longer diffusion time for a longer dsRNA.

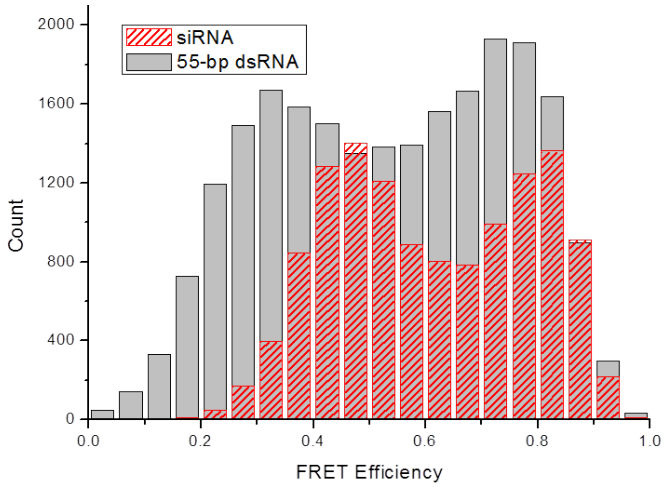

Figure 2.8 dsRNA length dependence of TRBP diffusion. FRET histogram of siRNA (19 bp) and 55 bp dsRNA shows a larger FRET dynamic range with longer dsRNA. Between 60–80 molecules were collected to generate the FRET histograms. The y-axis displays counts per 30 ms for selected FRET traces among the collected molecules.

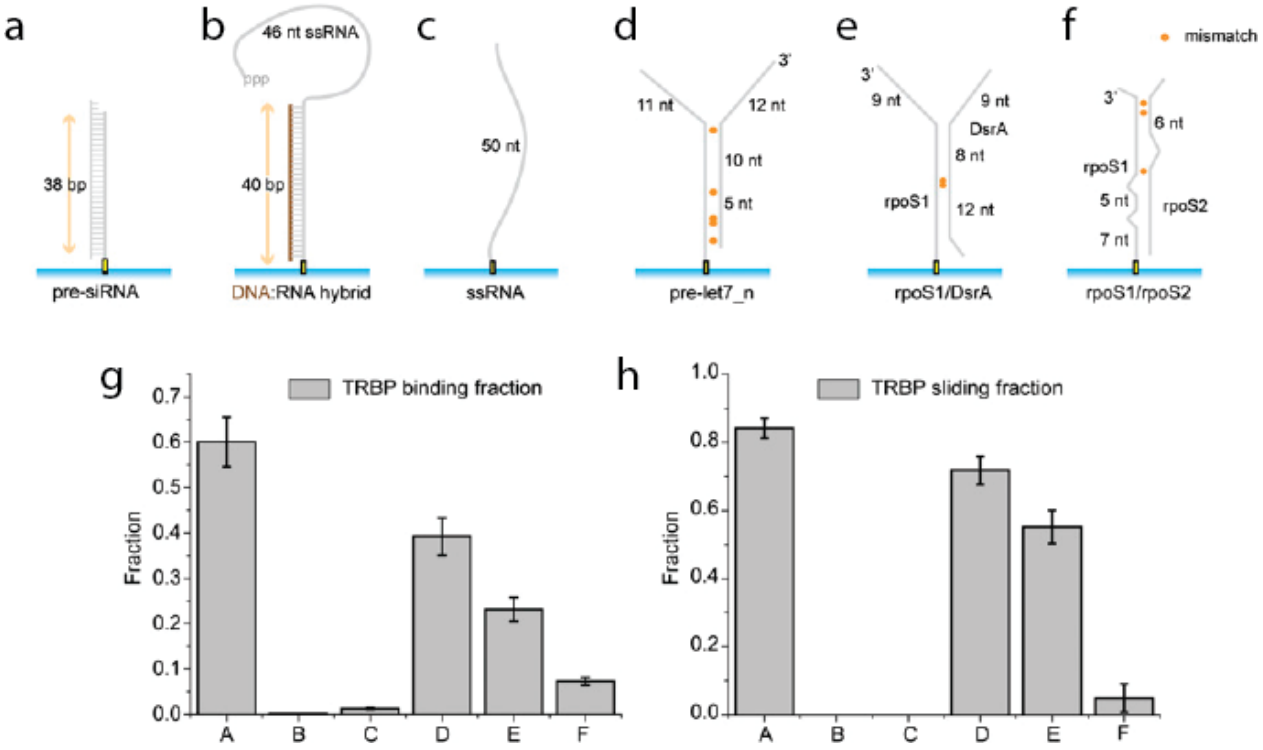

Figure 2.9 Substrate specificity of TRBP binding and its diffusion along dsRNA. (A–F) Pre-siRNA, DNA–RNA heteroduplex, ssRNA, and three dsRNAs with different secondary structure were prepared for testing the substrate specificity. (G) The fraction (±SEM) of the substrate (A–F) bound to TRBP was calculated from ~5,000 substrate molecules at 10 nM TRBP concentration. TRBP does not bind to DNA–RNA heteroduplex and ssRNA, but binds to all dsRNAs to varying degrees at 10 nM TRBP concentration. (H) The fraction (±SEM) of diffusion among TRBP-bound substrates shows that TRBP diffuses on dsRNAs with several single mismatches but not on dsRNA with more complex secondary structures.

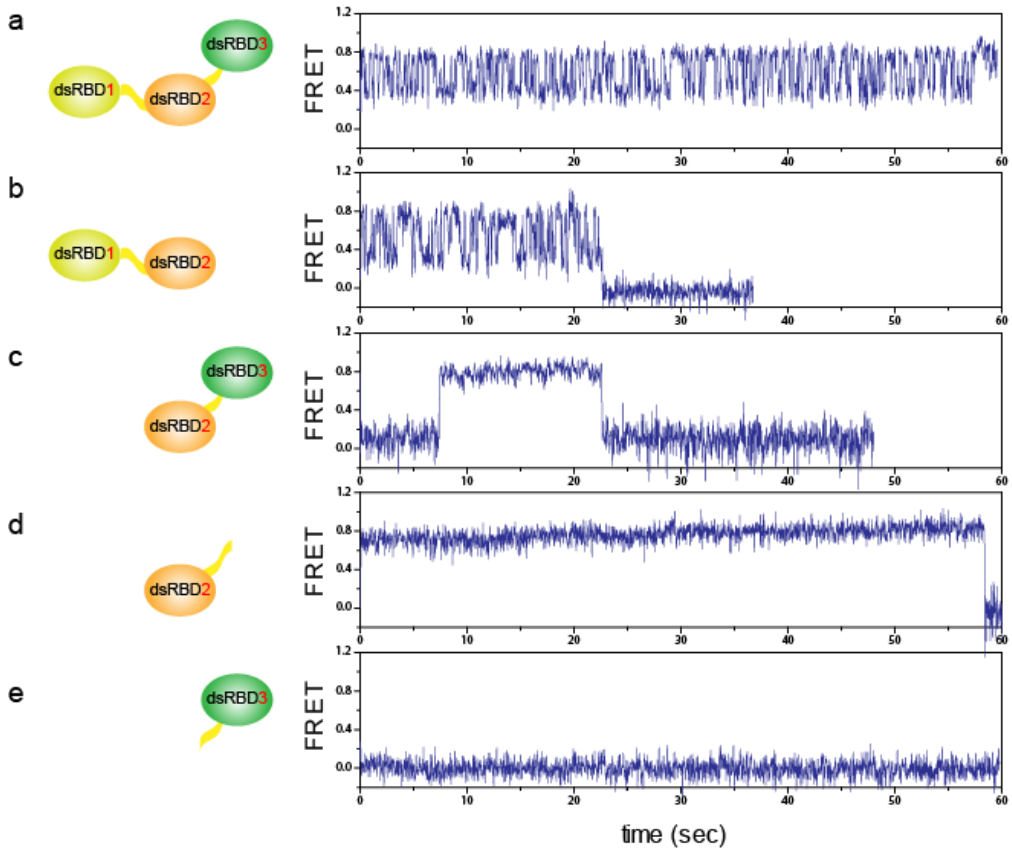

Figure 2.10 dsRBD1 and -2 as indispensable components for diffusion. (A) Representative single-molecule trace of wild-type TRBP diffusion on 38 bp dsRNA. (B–E) Several truncation mutants of TRBP show that both dsRBD1 and -2 are required for diffusion. (B) dsRBD1/2 shows a similar FRET trace as those from wild-type TRBP. (C, D) Short-lived binding but no diffusion was observed with dsRBD2/3 and dsRBD2 only. (D) DsRBD3 does not bind to dsRNA.

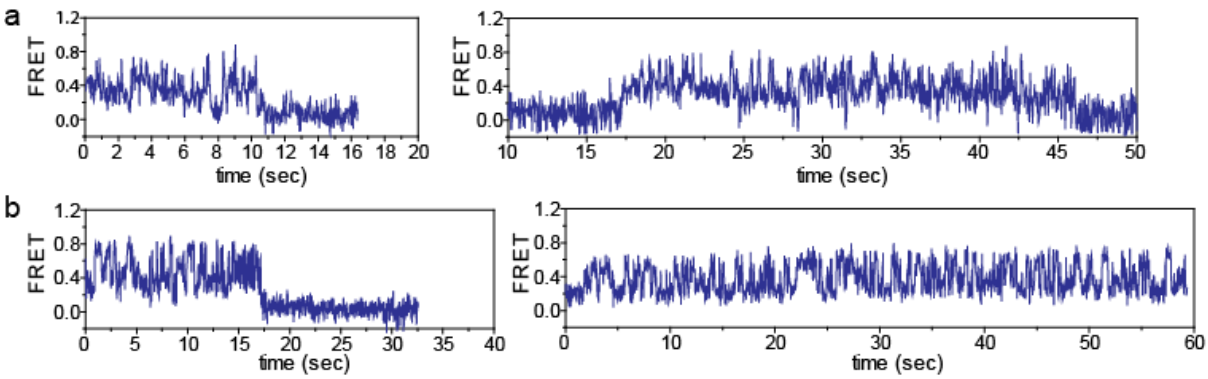

Figure 2.11 Diffusion of PACT and R3D1-L (A) Representative FRET traces of Alexa 647–labeled PACT with Cy3-labeled 38 bp dsRNA showing its repetitive diffusion motion. (B) R3D1-L also showed a similar diffusion motion on 38 bp dsRNA as seen in the wild-type TRBP.

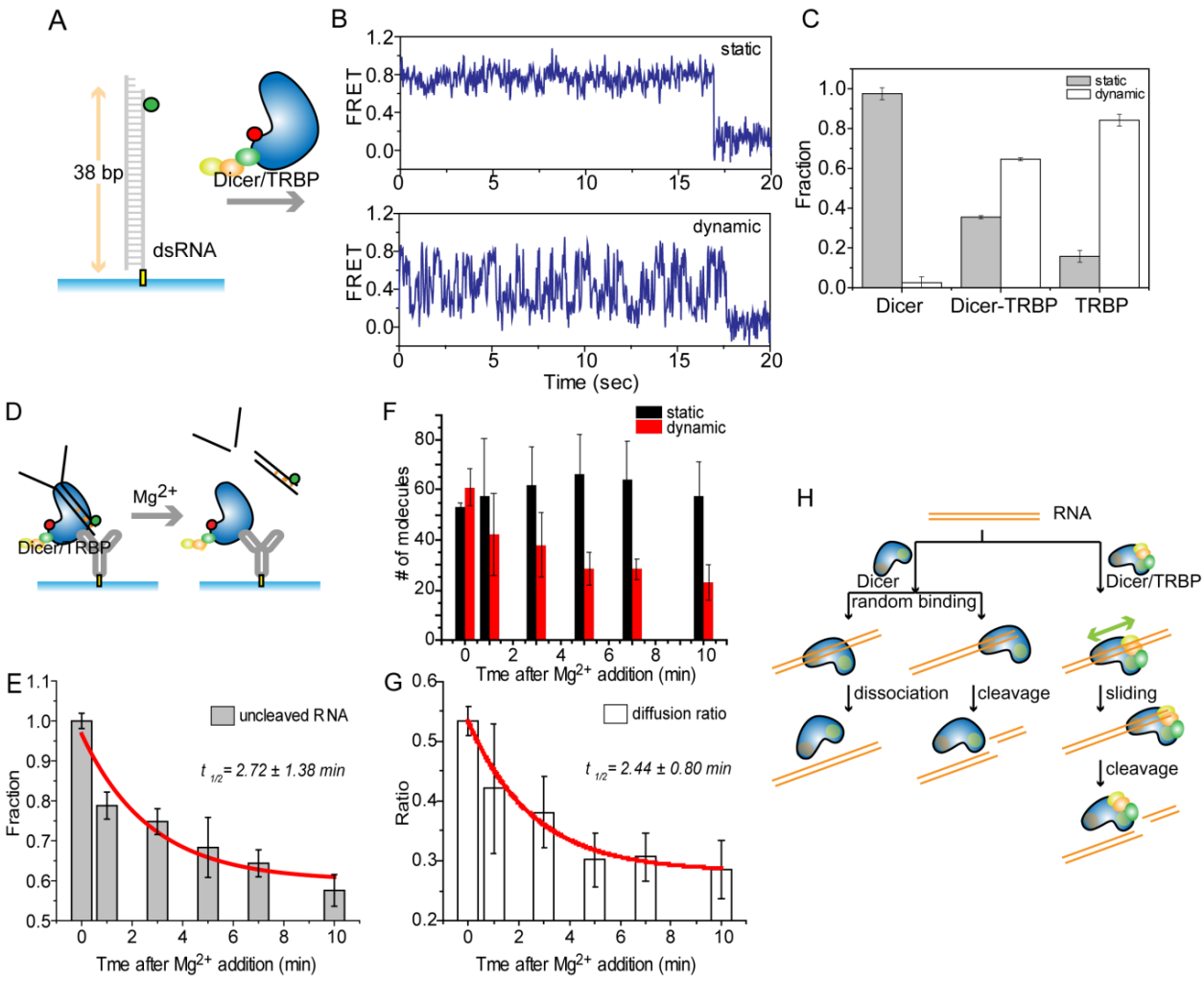

Figure 2.12 Dicer–TRBP diffuses on dsRNA and thereby promotes dsRNA cleavage activity. (A and B) Alexa 647–labeled Dicer–TRBP complex exhibits two populations-static vs. dynamic FRET. Dynamic FRET changes similar to those observed for TRBP only suggest that Dicer–TRBP also diffuses on dsRNA. (C) Fraction (±SEM) of static and dynamic population for Dicer, Dicer–TRBP, and TRBP. Most Dicer showed a static binding, whereas about 60% of Dicer–TRBP and more than 85% of TRBP showed diffusion. (D) Diagram of the cleavage assay by pull-down of Dicer–TRBP using anti-His antibody. The cleavage reaction was initiated by adding Mg^{2+} on the immobilized Dicer–TRBP–RNA complex. (E) The normalized number of fluorescence spots (±SEM) decreases upon Mg^{2+} addition, which triggers the cleavage of dsRNA. More than 3,000 molecules were investigated. (F) The number of diffusion molecules (red bar) (±SEM) of Dicer–TRBP decreases over time upon Mg^{2+} addition, whereas the number of static molecules (black bar) (±SEM) doesn't change. (G) The calculated ratio of diffusion molecules over both static and diffusion molecules exhibited the similar rate with the RNA cleavage rate by Dicer-TRBP, suggesting the correlation between the RNA cleavage and the diffusion by Dicer–TRBP. (H) The suggested model explains how Dicer–TRBP's diffusion may help the dsRNA cleavage by positioning Dicer at the correct cleavage site and thereby enhancing the cleavage rate.

Chapter 3

Evolutionarily Conserved Roles of the Dicer Helicase Domain in Regulating RNAi Processing

*A portion of the work presented in this chapter has been previously published as part of the following paper: Kidwell MA, Chan JM & Doudna JA (2014) Evolutionarily Conserved Roles of the Dicer Helicase Domain in Regulating RNAi Processing. *Journal of Biological Chemistry***289**(41):28352-62.

* Mary Anne Kidwell and Jennifer Doudna designed research; Mary Anne Kidwell and Jessica Chan performed research; Mary Anne Kidwell analyzed data; and Mary Anne Kidwell and Jennifer Doudna wrote the paper.

3.1 Introduction

Dicer is the enzyme responsible for cleaving double-stranded RNA (dsRNA) precursors into microRNAs (miRNAs) and small interfering RNAs (siRNAs) during RNA interference in eukaryotes (Wilson & Doudna, 2013). miRNAs are typically derived from hairpin precursors containing imperfectly base-paired stems, whereas siRNAs are formed from two separate RNA strands that comprise perfectly base-paired duplexes (Ambros *et al.*, 2003). Once processed by Dicer, both miRNAs and siRNAs function as sequence-specific guides to recruit Argonaute proteins and associated factors to complementary mRNAs for post-transcriptional silencing of gene expression (Bazzini *et al.*, 2012; Djuranovic *et al.*, 2012). Although RNAi is conserved among many eukaryotes, multiple duplications of its protein components have expanded the diversity and complexity of the pathway (Cerutti & Casas-Mollano, 2006; Moran *et al.*, 2013). Many species express multiple Dicer proteins, and it is often unclear if these copies have functionally distinct or overlapping roles in RNAi. Here we have investigated two fungal Dicers from *Sporotrichum thermophile* and show that their different functions are likely due to changes in their helicase domains.

Canonical Dicer consists of an N-terminal DExD/H box RNA helicase, a DUF283 domain, a PAZ domain, tandem RNase III domains, and a C-terminal double-stranded RNA-binding domain (dsRBD) (Figure 3.1A). Previous biochemical and structural studies show that the size of product RNAs is determined by the spatial arrangement of the PAZ and RNase III domains, as well as the binding pocket for the RNA 5′ end (Zhang *et al.*, 2004; MacRae *et al.*, 2006). The RIG-I-like helicase domain, which is located adjacent to the RNase III domains, has emerged as a substrate specificity determinant and may have developed specialized roles for different Dicer proteins.

Mammals and nematodes have only one Dicer enzyme, and differences in the cleavage rates of pre-miRNA and pre-siRNA substrates are attributable to the helicase domain (Ma *et al.*, 2008; Chakravarthy *et al.*, 2010). Human Dicer cleaves pre-miRNAs much more efficiently than pre-siRNAs, with the helicase domain inhibiting siRNA production. In *D. melanogaster*, which contains two Dicers, processing of these two types of RNA substrates is segregated such that Dicer-1 is responsible for generating miRNAs and Dicer-2 generates siRNAs (Lee *et al.*, 2004b). Dicer-1 contains an inactive helicase domain that is unable to hydrolyze ATP, which facilitates binding to the loops of pre-miRNAs (Tsutsumi *et al.*, 2011), whereas Dicer-2 has an active helicase that enables processive cleavage of long dsRNAs (Cenik *et al.*, 2011; Welker *et al.*, 2010; Lau *et al.*, 2012). The helicase domain may have allowed Dicer to adapt rapidly to a diverse array of substrates. Insight into the function of Dicer in other organisms could shed light on the evolution of this important enzyme.

An obstacle to studying Dicers biochemically has been their large sizes and difficultly in purification, so we turned to the thermophilic fungus *S. thermophile* as a system well-suited to purification of its two stable Dicers: StDicer-1 and StDicer-2. We discovered different biochemical activities for the isolated helicase domain of each of these proteins that could in turn influence the function of the full-length enzyme. The activities of these two Dicers are reminiscent of the RNAi pathway in *D. melanogaster*,highlighting how the helicase domain evolved as a conserved regulator of small RNA processing in eukaryotes.

3.2 Materials and Methods

3.2.1 Protein expression and purification

The genes encoding StDicer-1 (Q2H0G2.2) and StDicer-2 (XP_001228335.1) were amplified from the genomic DNA of *Sporotrichum thermophile* and the introns were removed using site-directed mutagenesis PCR.The full-length coding sequences as well as sequences encoding truncated proteins lacking the helicase domain (ΔHel) were cloned into a customized pFastBac expression vector (4C, Addgene #30116) using ligation-independent cloning, resulting in protein expression constructs that were fused downstream of a hexahistidine-maltose-binding protein (6×His-MBP) tag and a tobacco etch virus (TEV) protease cleavage site. The resulting plasmids were used in the Bac-to-Bac Baculovirus Expression System (Invitrogen) to produce StDicer expressing baculovirus that was used to transfect Sf-9 cells. After 72 hours, transfected Sf-9 cells were harvested and lysed by sonication in purification buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10% glycerol, 1 mM Tris (2-carboxyethyl)phosphine (TCEP)) supplemented with 10 mM imidazole, 0.5% Triton X-100, and protease inhibitors (Roche). The cleared lysate was incubated with Ni-NTA affinity resin (Qiagen) in purification buffer with 30 mM imidazole and was eluted with 300 mM imidazole. Eluted proteins were then dialyzed against purification buffer, followed by a second Ni-NTA step to remove the 6×His-MBP tag and TEV protease. Proteins were further purified by size-exclusion chromatography using a Superdex 200 (16/60) column (GE Life Sciences) in purification buffer.

The helicase domains from StDicer-1 and StDicer-2 were cloned into a customized pET expression vector (1M, Addgene #29656) using ligation-independent cloning, resulting in protein expression constructs that were fused downstream of a 6×His-MBP tag and a TEV protease cleavage site. The plasmids were transformed into BL21(DE3) cells and the proteins were purified using the protocol described above, except that the final size-exclusion chromatography was performed using a Superdex 75 (16/60) column (GE Life Sciences). All purification steps were carried out at 4°C. All protein concentrations were determined using a Nanodrop (Thermo).

3.2.2 DNA and RNA substrates

All of the DNA and RNA substrates used, with the exception of the hairpin pre-miRNA substrate, were synthesized by Integrated DNA Technologies (IDT). The sequence of the hairpin RNA substrate was derived from human pre-let-7a-1, and it was synthesized by *in vitro* transcription using T7 RNA polymerase. The DNA template contained a double ribozyme system to ensure homogeneous 5′ and 3′ ends, with a hammerhead ribozyme at the 5′ end and a HDV ribozyme at the 3′ end (Ferré-D'Amaré & Doudna, 1996). All RNA and DNA substrates were gel purified using 12.5% urea-PAGE prior to use. The RNA oligo 37a RNA can form a perfectly matched duplex containing a 2-nt 3′ overhang with either 37b RNA (35 bp dsRNA) or 37b DNA (35 bp DNA/RNA heteroduplex). The RNA oligo 27a can hybridize with 27b (25 bp dsRNA) and 24a can hybridize with 24b (22 bp dsRNA), both with 2-nt 3′ overhangs. The DNA oligos 53a DNA and 53b DNA form a completely complementary duplex (53 bp dsDNA). The sequences of all of substrates used in this study are in Table 3.1.

For the dicing assays and electrophoretic mobility shift assays (EMSAs), the substrates were 5′-end labeled with using T4 polynucleotide kinase (New England

Biolabs, Inc. Beverly, MA) and γ-^{32}P-ATP (Perkin Elmer), gel-purified, and annealed before use. Annealing was performed in 100 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$ and 30 mM NaCl by heating at 95º C for 5 minutes and either slow cooling (perfect duplex RNAs) or flash cooling (hairpin RNAs).

3.2.3 Dicing assays

Labeled and annealed RNA substrates were incubated with 1 μM StDicer at 37℃ for the specified time in a 10 μl reaction volume containing 20 mM Tris-HCl (pH 6.5), 5 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT and 1% glycerol. 1 mM ATP was added to reactions with ATP. Reactions were stopped by addition of 1.2 volumes of loading buffer (95% formamide, 50 mM EDTA, 0.025% SDS, 0.1% xylene cyanole FF and 0.1% bromophenol blue). After heating at 95°C for 5 min, the samples were analyzed by electrophoresis with a 12.5% polyacrylamide-7M urea gel run in 0.5×TBE buffer and quantified using the Phosphorimager/ImageQuant (GE Healthcare).

3.2.4 ATP hydrolysis assay

Protein (1 μM) was incubated with 1 mM ATP, 10-100 nM α-^{32}P-ATP (3000Ci/mmol, PerkinElmer), and 4 μM nucleic acid substrate in a buffer consisting of 20 mM Tris-HCl (pH 6.5), 5 mM $MgCl_2$, 25 mM NaCl, 1 mM DTT and 1% glycerol. Reactions were stopped with the addition of 100 mM EDTA, spotted onto 20 × 20 cm PEI Cellulose F plates (EMD Millipore), and chromatographed in 1 M formic acid and 0.5 M LiCl until the solvent traveled ¾ of the plate. The plate was dried and quantified using the Phosphorimager/ImageQuant (GE Healthcare).

3.2.5 Electrophoretic mobility shift assay

Approximately 0.5-1 nM (500-1000 CPM) labeled and annealed RNA substrates were incubated with the indicated concentrations of StDicer constructs for 30 min at 4º C in 20 mM Tris-HCl (pH 7.5), 25 mM NaCl, 5 mM EDTA, 1 mM DTT and 1% glycerol. 1:2 serial dilutions were used to reach the indicated concentrations. Reactions were analyzed on a 6% native polyacrylamide gel and quantified using the Phosphorimager/ImageQuant. Percent bound RNA was plotted as a function of protein concentration. K_d was determined by global fitting to the equation fraction bound = A x [protein]/(K_d + [protein]), where A is the amplitude of the binding curve. Curve fitting was conducted with KaleidaGraph (Synergy Software, Reading, PA).

3.2.6 *S. pombe* strains, media, and constructs

The *S. pombe dcr1* gene was amplified from genomic DNA. All proteins were cloned into the pREP1 vector, carrying a *Leu* selection marker and *nmt1* promoter, using NdeI and BamHI. The *S. pombe* strain h- ∆dcr1+::kanMX6 ura4-D18 leu1-32 was used for functional complementation assays (MacRae *et al.*, 2006). Transformed cells were grown in minimal media with glutamate as nitrogen source (PMG, Sunrise Science), supplemented with 225 mg/L of adenine, histidine, lysine hydrochloride, and uracil. Once the cells reached an OD_{600} of 0.4, cells were either plated or grown in liquid culture containing yeast extract media (YES or rich media) or modified KsnoT media with 15 μg/mL thiabendazole (Santa Cruz Biotechnology) (Lindner *et al.*, 2002). Plates were imaged after about 4 days of growth at 30°C. Liquid culture measurements were

done in 96-well plates using a plate reader (Tecan Infinite F200 Pro). The liquid cultures were fit to either an exponential growth equation using Prism or a logistic growth equation using R.

3.2.7 qRT-PCR

RNA was purified from 1 mL of saturated *S. pombe* cultures using hot acid phenol and chloroform. Residual DNA was removed by DNase treatment (Promega), after which the RNA was purified again by ethanol precipitation. cDNAs were prepared with an Invitrogen Superscript III kit followed by the addition of RNase H (NEB). cDNAs were quantified with a Stratagene MX3000 quantitative PCR system using DyNAmo HS SYBR Green (Thermo). All primer-set amplification values were normalized to ACT1 amplification values. All primer sets used can be viewed in Table 3.2.

RESULTS

3.3.1 StDicer-1 and StDicer-2 display distinct cleavage preferences

We performed a phylogenetic analysis to search for smaller eukaryotic Dicers that might be tractable for biochemical investigation, leading to selection of the thermophilic fungus, *Sporotrichum thermophile* (synonymous with *Myceliophthora thermophile*) (4th *et al.*, 2011). This organism possesses two Dicers containing the domains that are common to most eukaryotic Dicers (Figure 3.1A). In particular, both the helicase and the tandem RNase III domains are well-conserved and contain the putative metal-coordinating residues required for enzymatic activity (Figure 3.1B). Both proteins are significantly smaller than human Dicer-1 due to two large deletions. One deletion occurs between the putative PAZ domain and the first RNase III domain and the second is within the first RNase III domain (Figure 3.1A, B). Both of these regions are poorly conserved among eukaryotic Dicers and the latter deletion in the RNaseIIIa domain is proposed to be an Argonaute binding site unique to vertebrates (Sasaki & Shimizu, 2007). However, we were unsure whether these deletions would affect the function of these Dicers and proceeded to investigate their catalytic activities.

We first assayed RNA cleavage by incubating purified StDicer-1 and StDicer-2 with two different radiolabeled RNAs in the presence or absence of ATP (Figure 3.2A, B). Since it was difficult to predict computationally the natural *S. thermophile* Dicer substrates due to poor sequence and structure conservation (Lee *et al.*, 2010), we first used two RNA substrates that have been well-characterized for other eukaryotic Dicers (Table 3.1) (Ma *et al.*, 2008; Chakravarthy *et al.*, 2010). One substrate, 35 bp RNA, mimics a pre-siRNA substrate and comprises two 37 nt RNA strands which hybridize to form a perfect RNA duplex with 2 nt, 3′ overhangs. The other substrate mimics a pre-miRNA hairpin and is derived from human pre-let7, containing an imperfectly base paired stem and a 27 nt loop. RNA cleavage was observed with both proteins, but only StDicer-1 was able to cleave the hairpin RNA to approximately 25 nt, which is a length consistent with possible loading into Argonaute and is the observed cleavage length for small RNAs in fungal RNAi (Lee *et al.*, 2010; de Nabanita *et al.*, 2013). The rate of RNA cleavage for StDicer-1 with the hairpin RNA substrate was the same in the presence and absence of ATP (Figure 3.2B).

StDicer-2 was only able to generate ~25 nt products from the 35 bp pre-siRNA substrate (Figure 3.2A, B). Its rate of cleavage was reduced compared to that observed

for StDicer-1 with the 35 bp pre-siRNA. Cleavage activity of StDicer-2 was unaffected by ATP, a behavior distinct from that of *D. melanogaster* Dicer-2 (Cenik *et al.*, 2011; Welker *et al.*, 2010), suggesting that additional factors may be required for efficient RNA cleavage.

3.3.2 StDicer-1 is a thermostable protein that cleaves pre-miRNAs

To further investigate the different RNA cleavage activities of StDicer-1 and StDicer-2 we tested additional pre-miRNA substrates and temperatures (Figure 3.2C-E) (Ma *et al.*, 2012; Lee & Doudna, 2012). The two RNA substrates tested were a pre-miRNA with a reduced loop size of 5 nt (smloop) and a related hairpin with a large loop but a completely perfect RNA stem (pStem hairpin). Both RNAs were rapidly cleaved by StDicer-1 and not by StDicer-2 (Figure 3.2C). This suggests that either the loop of the pre-miRNA inhibits cleavage by StDicer-2 or the RNA substrate must have a perfectly base-paired duplex longer than 25 bp to allow for StDicer-2 cleavage. Since StDicer-1 more efficiently and accurately cleaves RNA, the temperature dependent cleavage was tested for the smloop pre-miRNA and compared with human Dicer (HsDicer). Four different temperatures were assayed: 30°C, the optimal temperature for *S. pombe* growth, 37°C, the temperature at which all other assays were performed, 48°C, the optimal temperature for *S. thermophile* growth, and 68°C. While both enzymes were able to cleave the pre-miRNA efficiently at the highest temperature, HsDicer's cleavage rate plateaued at 37°C while the rate of cleavage for the thermophilic protein continued to increase (Figure 3.2D, E). This indicates that as anticipated, the thermophilic proteins may be more stable at the higher temperatures.

3.3.3 StDicer-2 displays robust ATP hydrolysis in the presence of dsRNA

The differences in cleavage activities of StDicer-1 and StDicer-2 suggest that these proteins have distinct functions, which are reminiscent of the proteins from *D. melanogaster*. To investigate this further, we tested the ATPase activity of the two *S. thermophile* proteins using a radiolabeled TLC assay (Figure 3.3A). There was a striking increase in ATP hydrolysis for the full-length StDicer-2 that warranted further investigation.

We isolated the helicase domain (HD) from both StDicer-1 and StDicer-2 and measured ATP hydrolysis rates in the presence and absence of nucleic acid substrates (Figure 3.3B). StDicer-1 HD displayed no activity even in the presence of 35 bp RNA; similar results were obtained for full-length StDicer-1. In contrast, StDicer-2 HD showed RNA-dependent ATP hydrolytic activity. Although both single-stranded RNA (ssRNA) and dsRNA stimulated ATP hydrolysis by StDicer-2 HD, a significantly higher increase was observed in the presence dsRNA. These results are similar to those obtained for other RIG-I family helicases, of which Dicer is a member, for which ATP hydrolysis is only observed when the protein is bound to nucleic acid (Gee *et al.*, 2008; Cenik *et al.*, 2011).

We further investigated this nucleic acid stimulated activity for full-length protein StDicer-2 (Figure 3.3C). We incubated the protein with single or double stranded DNA or RNA as well as a DNA/RNA heteroduplex. The slight stimulation of activity was once again observed in the presence of ssRNA. ATP hydrolysis further increased in the

presence of an A-form duplex such as the DNA/RNA heteroduplex and reached a maximum with dsRNA.

With *D. melanogaster* Dicer-2, the rate of ATP hydrolysis for dsRNA substrates of different lengths was comparable (Cenik *et al.*, 2011). In contrast, significant differences between the different nucleic acid substrates observed for *S. thermophile* ATPase activity led us to test whether changes in the RNA structure could alter the rate of ATP hydrolysis (Figure 3.3D). We incubated StDicer-2 with four different RNA substrates. Three of the RNAs were perfect RNA duplexes that ranged in size from 35 base pairs (bp) to 22 base pairs (bp), representing Dicer substrates and products respectively. The fourth RNA substrate was a hairpin RNA containing an imperfect 22 bp stem and a 27 nucleotide (nt) loop. There was no significant difference in the rates of ATP hydrolysis, similar to what was observed with *D. melanogaster* Dicer-2 (Cenik *et al.*, 2011).

3.3.4 RNA binding properties of StDicers

In light of the observed differences in cleavage activity, we investigated the binding affinities of corresponding RNA substrates. Affinities were measured by EMSA with 0.5-1 nM of RNA substrate and increasing protein concentrations (Figure 3.4A, Table 3.3). StDicer-1 had higher affinity for both hairpin and dsRNA substrates, in agreement with its observed rapid rates of cleavage. StDicer-2 had significantly weaker affinity to the RNA substrates tested, approximately 1-10 μM, with a slightly higher affinity for duplex RNA substrates (Figure 3.4B).

The observation that the StDicer helicase domains have different RNA binding and cleavage activities led us to test the RNA binding affinities of the isolated helicase domains (HD) from StDicer-1 and StDicer-2 (Figure 3.4C). While weaker than that observed for the full-length proteins, the measured affinities followed a similar trend. StDicer-1 HD has the highest affinity, with a 100-fold difference in affinity between hairpin and duplex RNA substrates. StDicer-2 had only a two-fold different for duplex RNA substrates over hairpin RNAs, which was approximately 1-10 μM. This difference in StDicer-2 affinities may be due to interactions with the ssRNA regions of the loop as previously proposed (Ma *et al.*, 2008; Tsutsumi *et al.*, 2011).

3.3.5 The helicase domain inhibits RNA binding and cleavage for StDicer-2

To determine the effect of the helicase domain on substrate recruitment to Dicer, we tested the affinity of a StDicer-2 truncation that lacked the helicase domain (ΔHel StDicer, Figure 3.5A). While the affinity for duplex RNAs only increased by 3-4 fold, the affinity for hairpin RNAs increased ~10^3-fold. These results indicate that the helicase domain inhibits pre-miRNA binding.

In accordance with the RNA binding data, removal of the helicase domain greatly increased the rates at which some RNA molecules were cleaved. This protein showed increased cleavage rates with the duplex RNA substrates (Figure 3.5B). Surprisingly, although the affinity for hairpin RNA substrates increased greatly the protein was still unable to cleave these RNA molecules to produce miRNA-length products. As with the full-length proteins, the addition of ATP did not change the rate of RNA cleavage. These results are consistent with studies of Drosophila and human Dicers and indicate there may be a universal mode of regulation that involves the helicase domain (Ma *et al.*,

2008; Cenik *et al.*, 2011). However, the catalytic core still retains some selectivity for RNA processing.

3.3.6 StDicer-1 and StDicer-2 complement growth defects *in vivo*.

To determine if StDicer-1 and StDicer-2 are functional *in vivo*, we used a genetic complementation assay in *S. pombe* which has been used successfully before to assay the function of human and *Giardia* Dicer (Provost *et al.*, 2002; MacRae *et al.*, 2006). *S. pombe* only contains one endogenous Dicer (SpDicer) and strains lacking Dicer (KO) exhibit slight growth impairment. In *S. pombe*, RNAi is needed for the production of small RNAs that are generated from the centromere and are necessary for proper chromosome segregation and microtubule attachment during mitosis (Provost *et al.*, 2002). These mitotic defects can be exacerbated by the addition of the microtubule-destabilizing drug thiabendazole (TBZ), nearly preventing the growth of the KO strains (Provost *et al.*, 2002).

This KO strain was transformed with different Dicer constructs and growth was assayed by plating serial dilutions on rich media or media containing TBZ (Figure 3.6A). On rich media, growth of strains containing full-length Dicer from *S. pombe* or *S. thermophile* were nearly indistinguishable from wild-type and empty vector growth. Surprisingly, the strains expressing the StDicer constructs lacking the helicase domain were quite sick and showed decreased growth on rich media. The mRNAs encoding the episomally expressed Dicer proteins were all at least 3-fold higher than the wild-type Dicer expression as determined by RT-qPCR (Figure 3.6B, C). Thus, gene expression levels of these constructs are likely to be similar. Previously it has been shown that *Giardia* Dicer (MacRae *et al.*, 2006), which lacks the helicase domain altogether, is able to partially complement the KO strain. We propose that the helicase domains of the StDicer enzymes help to restrict the access of small RNAs to the protein, preventing promiscuous RNA binding and cleavage. It is possible that the *Giardia* Dicer has evolved to not require this additional inhibition and has other ways to regulate substrate cleavage.

On media containing TBZ, only strains containing full-length Dicer were able to suppress the growth defects of the endogenous Dicer deletion. We noted that there was a slight difference in growth between strains expressing StDicer-1 versus StDicer-2, so we assayed the growth in liquid culture to quantify the changes in growth (Figure 3.6D). By fitting the growth data of StDicer-1 and StDicer-2 to exponential curves, we observed that the expression of StDicer-1 gave an approximately 2-fold increase in growth (t_d = 12.3 ± 0.4 hrs) over strains containing StDicer-2 (t_d = 22.3 ± 0.7 hrs). While both of these growth behaviors were significantly faster than that observed for the empty vector control, they were only a fraction of the growth rate observed for wild-type *S. pombe* (t_d = 1.7 ± 0.1 hrs) and KO complemented with *S. pombe* Dcr1 (t_d = 1.6 ± 0.1 hrs) when fit with a logistic growth equation. These results indicate that full-length Dicer from S. thermophile is a functional ortholog to previously described Dicer systems and can maintain RNAi in a heterologous system.

3.4 Discussion

Dicer is a large, multi-functional protein and previous biochemical studies have focused on how its noncatalytic domains affect RNA cleavage. We have found that the

helicase domain plays an evolutionarily conserved role in selecting RNAs for Dicer-mediated cleavage. Due to the technical challenges in purifying fragments from human Dicer, we turned to the thermophilic fungus *Sporotrichum thermophile*. We found that distinct biochemical activities of the isolated helicase domains from both Dicers contribute to the specialized functions of the full-length proteins. This provides insight into its evolution as a core enzyme of the RNAi machinery.

Conflicting models have emerged for the function of these two Dicers in fungi. A previous study on *Neurospora crassa* indicated that these proteins are redundant in function during quelling, a process in which long dsRNAs lead to decreased expression of corresponding genes (Catalanotto *et al.*, 2004). However, another study involving the filamentous fungus *Magnaporthe oryzae* showed that only one Dicer isoform was required for siRNA accumulation (Kadotani *et al.*, 2004). Our biochemical and *S. pombe* complementation data support the hypothesis that the two proteins are not redundant in the small RNA pathway. RNA cleavage assays showed that although both proteins can cleave duplex RNAs, only StDicer-2 accurately generated siRNAs of a defined length. The differences in the two proteins are more pronounced when tested for pre-miRNA processing, where only StDicer-1 was able to cleave hairpin RNAs. Only recently have functional miRNA-like mechanisms been discovered in fungi and the distinctions between individual Dicer protein have yet to be fully elucidated in the processing of hairpin RNAs (Lee *et al.*, 2010).

The distinct functions for eukaryotic Dicer can be attributed to the different activities of the helicase domain. Recent studies have shown that the helicase domain, which cannot hydrolyze ATP, preferentially interacts with the loops of pre-miRNAs and inhibit cleavage of pre-siRNAs (Ma *et al.*, 2012; Tsutsumi *et al.*, 2011). This is supported by structural studies with human Dicer, which has an inactive helicase domain, demonstrating that pre-siRNAs are held in a nonproductive conformation between the PAZ and helicase domain (Taylor *et al.*, 2013). In addition, pre-miRNAs induce structural changes involving the helicase domain that enable productive substrate recognition and rapid cleavage (Ma *et al.*, 2012; Taylor *et al.*, 2013). Our ATP hydrolysis assays and RNA binding experiments with StDicer-1 support this hypothesis, as only StDicer-1, which bears an inactive helicase domain, could cleave pre-miRNAs.

In contrast, the active helicase domains of other Dicers are found to be important to generate multiple siRNAs processively from a long, dsRNA precursor. This has been shown with *D. melanogaster* Dicer-2, where mutations to the helicase domain abolish RNA cleavage *in vivo*(Lee *et al.*, 2004b). Because *Drosophila* have two Dicer proteins, one hypothesis for the divergent function of Dicer-2 in flies is that this protein, which is one of the fastest evolving genes in the *Drosophila* genome, became a component of insect-specific antiviral defense within flies (Galiana-Arnoux *et al.*, 2006; de Jong *et al.*, 2009). An emerging alternative view holds that Dicer was duplicated much earlier in metazoan evolution and one copy was subsequently lost as alternative antiviral defenses were developed (Mukherjee *et al.*, 2013). Our results support the latter hypothesis, although we cannot rule out the possibility that the activity of StDicer-2 could have arisen by convergent evolution.

It is interesting to consider how Dicers with an active helicase domain may have evolved different modes of RNA recognition that promote interaction with dsRNA and more facile access to the catalytic center. The closest structure is that from RIG-I, and

the preferential binding of dsRNA is likely conserved with StDicer-2 but not with StDicer-1 (Kowalinski *et al.*, 2011; Luo *et al.*, 2011; Jiang *et al.*, 2011; Civril *et al.*, 2011). Understanding the detailed molecular changes between these two enzymes could provide further insight into how Dicer recognizes RNAs and how this helicase family has evolved to enable substrate selection. Aside from providing evolutionary insights, Dicers from a thermophilic fungus also exhibit increased stability over other eukaryotic Dicers, making them a great model system for further functional and structural studies of proteins in this family.

Table 3.1 Sequences of oligonucleotides.

	Oligonucleotide Sequence
hairpin	5′-UGA GGU AGU AGG UUG UAU AGU UUU AGG GUC ACA CCC ACC ACU GGG AGA UAA CUA UAC AAU CUA CUG UCU UAC C-3′
smloop hairpin	5′-UAG CUU AUC AGA CUG AUG UUG ACU GUU GAA UCU CAU GGC AAC ACC AGU CGA UGG GCU GUC-3′
pStem-a hairpin	5′-UGA GGU AGU AGG UUG UAU AGU UUG AUU AGG GUC ACA CCC ACC-3′
pStem-b hairpin	5′-P-ACU GGG AGA UUC AAA CUA UAC AAC CUA CUA CCU CAU U-3′
37a RNA	5′-UGA GGU AGU AGG UUG UAU AGU UUG AAA GUU CAC GAU U-3′
37b RNA	5′-UCG UGA ACU UUC AAA CUA UAC AAC CUA CUA CCU CAA A-3′
37b DNA	5′-TCG TGA ACT TTC AAA CTA TAC AAC CTA CTA CCT CAT T-3′
27a RNA	5′-GUC ACG CUG CCC AAG UCU CUG CUG AAA-3′
27b RNA	5′-UCA GCA GAG ACU UGG GCA GCG UGA CUU-3′
24a RNA	5′-GUC ACG CUG CCC AAG UCU CUG CAA-3′
24b RNA	5′-GCA GAG ACU UGG GCA GCG UGA CUU-3′
53a DNA	5′-GAT GGA GTT AGA AGC AGC ACT TGA TGC TAT CAA TGA TTG TAA TGT AGC TGT AC-3′
53b DNA	5′- GTA CAG CTA CAT TAC AAT CAT TGA TAG CAT CAA GTG CTG CTT CTA ACT CCA TC-3′

The sequence of the hairpin RNA is from human pre-let-7a and smloop is from human pre-miR-21 (Lee & Doudna, 2012). pStem-a hairpin and pStem-b hairpin can be hybridized to form pStem hairpin and were previously named 37a-loop and 37b-loop respectively (Ma *et al.*, 2012). Each a and b pair of RNAs can be hybridized to form a perfectly base-paired RNA substrate with 2 nt, 3′ overhangs. Since the duplexes are shorter, the names as also reduced by 2. For example, 37a RNA and 37b RNA hybridize to form 35 bp RNA. 53a DNA and 53b DNA hybridize to form completely complementary 53 bp DNA. P indicates a 5′ phosphate.

Table 3.2 RT-qPCR primers used in this study.

	Oligonucleotide Sequence
Pombe Dicer	forward, 5′-GAT TGT GAG TTG CCG AAG TAT G-3′ reverse, 5′-GCT TCG ACC ATA TCT GCT ATC C-3′
StDicer-1	forward, 5′-CTG CGG TGA TAT GGT AGA GAA C-3′ reverse, 5′-CGA AGA CCA AGA GAT TCA CCT G-3′
ΔHel StDicer-1	forward, 5′-AAG ATC CAC TTT CAG GGC G-3′ reverse, 5′-GTG GTA TTC AGA CTC GTC CTT G-3′
StDicer-2	forward, 5′-GAT CTG AGC CAA AAG GAC TAC C-3′ reverse, 5′-AGC AGA TGA CGA GAT TAC ACG-3′
ΔHel StDicer-2	forward, 5′-ATG CGA TCC TCA AGT TCT GTA C-3′ reverse, 5′-ACG ATC TTG TCC TTC AAC CG-3′
Act1+	forward, 5′-GGT TTC GCT GGA GAT GAT G-3′ reverse, 5′-ATA CCA CGC TTG CTT TGA G-3′

Table 3.3 K_ds for StDicer-1 and StDicer-2 for duplex and hairpin RNAs.

	35-bp duplex	Hairpin RNA
StDicer-1	76 ± 14 nM	33 ± 3 nM
StDicer-2	1,800 ± 100 nM	9,600 ± 300 nM
StDicer-1 HD	>10,000 nM	930 ± 80 nM
StDicer-2 HD	3,200 ± 400 nM	6,500 ± 2000 nM
ΔHel StDicer-2	900 ± 200 nM	33 ± 3 nM

There reported equilibrium dissociation constants were calculated from the gel shifts in Figure 3.4. The errors represent the standard fitting error. The concentration of the radiolabeled RNA used in these experiments was approximately 0.5-1 nM.

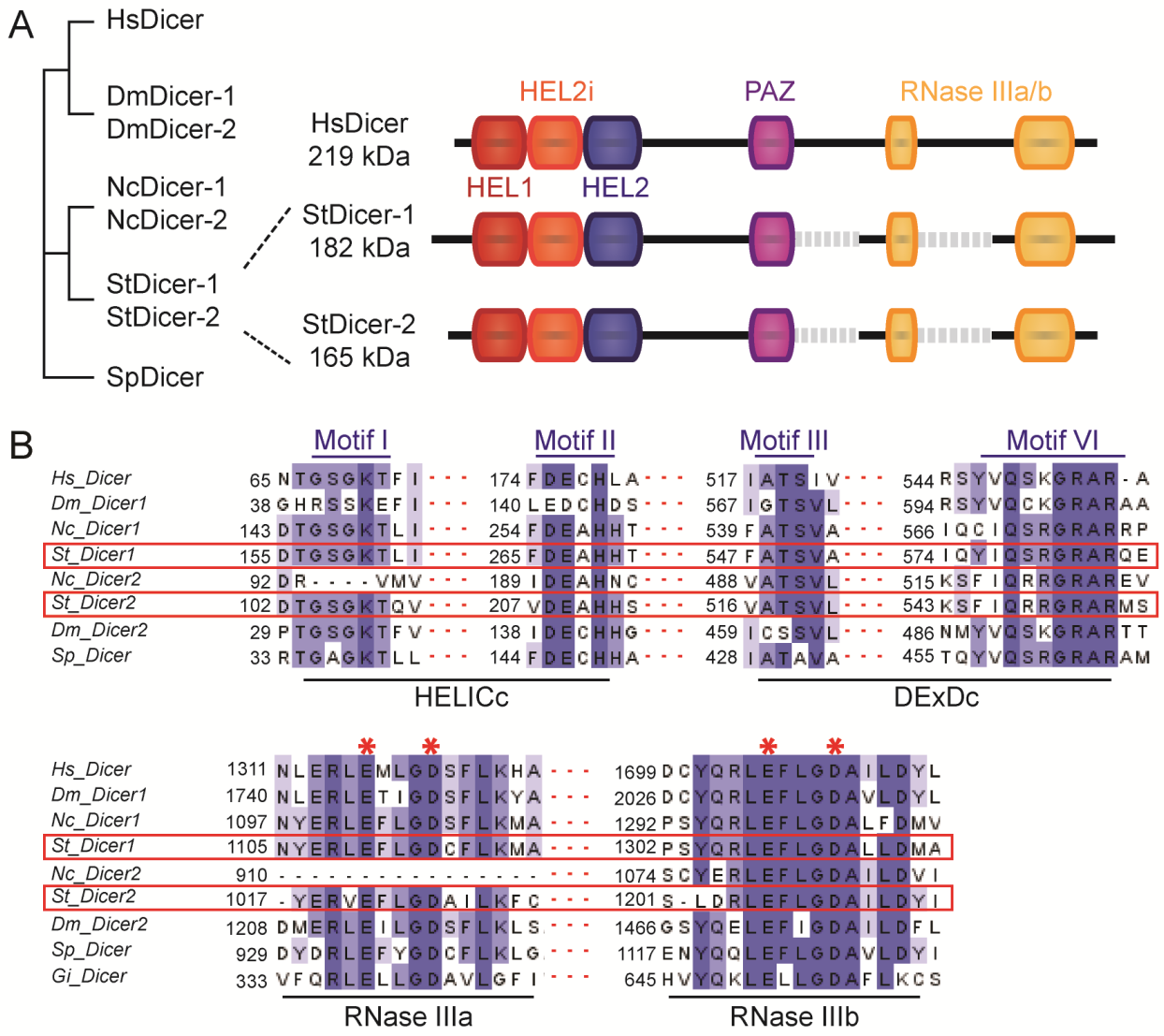

Figure 3.1 Domain architecture and alignment of StDicer-1 and StDicer-2 with other eukaryotic Dicers. (A) Schematic representation of the domain organization and phylogenetic relationship of StDicer-1 and StDicer-2 to other eukaryotic Dicers. The left side of the panel is the phylogenetic relationship between Dicers from *S. thermophile* (StDicer) and Dicers from humans (HsDicer), *D. melanogaster* (DmDicer), *N. crassa* (NcDicer), and *S. pombe* (SpDicer). On the right side of the panel is the domain architecture of human and StDicer. The helicase domain is organized into three lobes termed HEL1, HEL2i, and HEL2. The grey dashed lines represented large regions of HsDicer that are absent in StDicer. (B) Multiple sequence alignment output from ClustalX comparing the amino acid sequence of *S. thermophile* Dicer-1 (*St_Dicer1*) and Dicer-2 (*St_Dicer2*), which are outlined in red, and closely related proteins from Human (*Hs_Dicer*), *D. melanogaster* (*Dm_Dicer1*, *Dm_Dicer2*), *N. crassa* (*Nc_Dicer1*, *Nc_Dicer2*), *S. Pombe* (*Sp_Dicer*), and *G. intestinalis* (*Gi_Dicer*). Dark blue represents

positions that have a single, fully conserved residue with the two lighter blue colors indicating strongly conserved and weakly conserved residues. The helicase domain is one of the most conserved regions among full-length Dicers. The highlighted motifs are involved in binding an NTP, typically ATP, and the energy of hydrolysis is used to dynamically interact with RNA. *S. thermophile* Dicers contain intact RNaseIII domains and the residues highlighted by the red asterisks are involved in coordinating Mg^{2+} in the *G. intestinalis* structure.

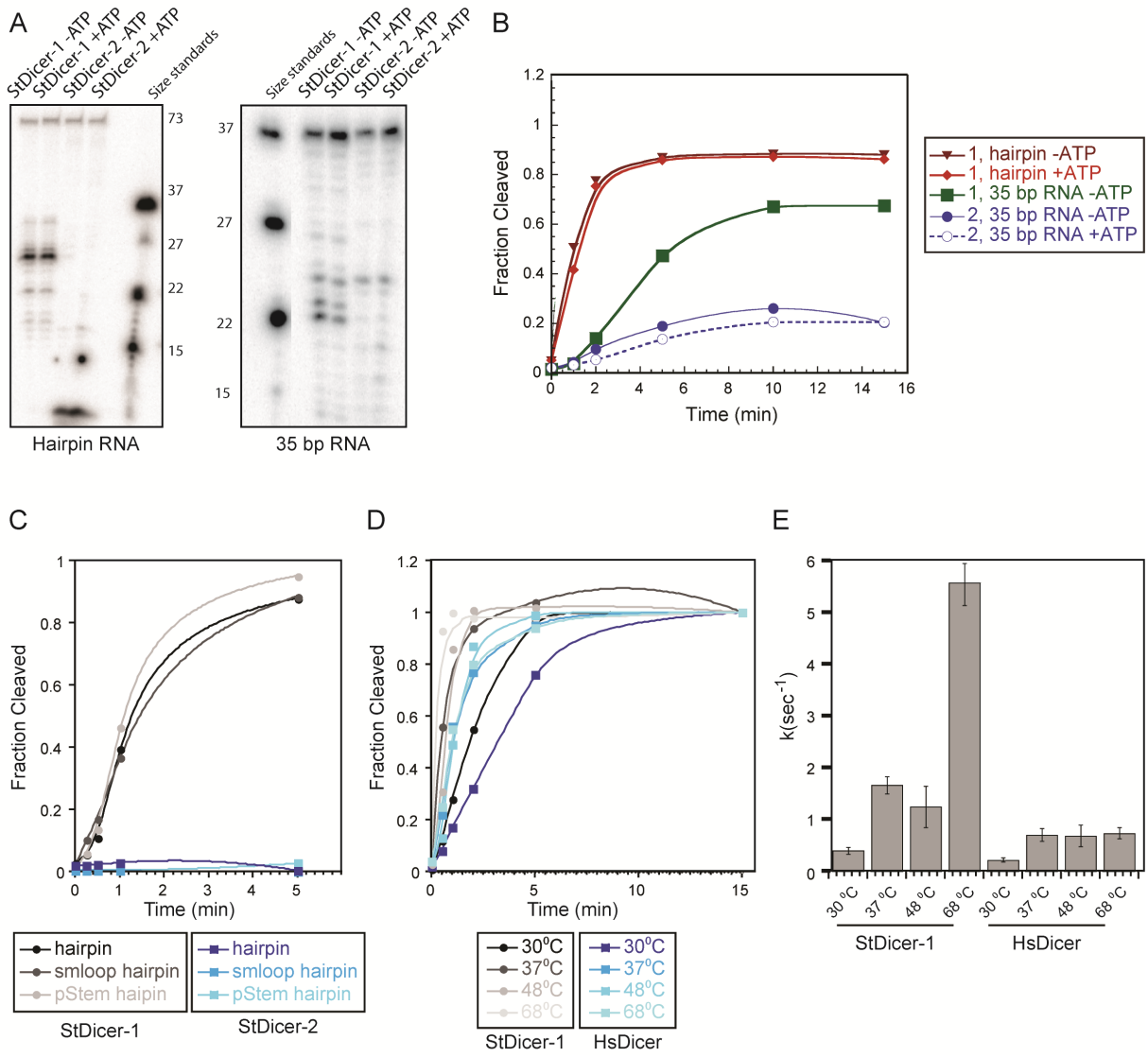

Figure 3.2 StDicer-1 and StDicer-2 have distinct cleavage properties *in vitro*.(A) Representative cleavage of StDicer-1 and StDicer-2 on hairpin RNA and 35 bp RNA. Radiolabeled RNA was analyzed with 12.5% denaturing PAGE with labeled ssRNA size standards. All size standards were run on the same gel with the cleavage assays. (B) StDicer-1cleaves hairpin RNAs more rapidly than duplex RNAs without the requirement for ATP. StDicer-2 cleavage is not influenced by ATP and only cleaves perfect duplex RNA accurately. (C) StDicer-1 is able to cleave different RNA hairpin structures. StDicer-1 and StDicer-2 were incubated with three different hairpin RNA substrates including a hairpin with a 5 nt loop (smloop hairpin) and one with a perfectly base-paired stem (pStem hairpin). The substrate labeled “hairpin” is the same as that in Figure 3.1C and is the main hairpin substrate used throughout the manuscript. (D) Temperature dependent cleavage of StDicer-1 and HsDicer. Cleavage of the smloop hairpin RNA was assayed at four different temperatures with StDicer-1 and HsDicer. (E) Quantified rates of Dicer cleavage at various temperatures. The curves from Figure 3.1E were fit with a one-phase association equation and the rate constants were plotted.

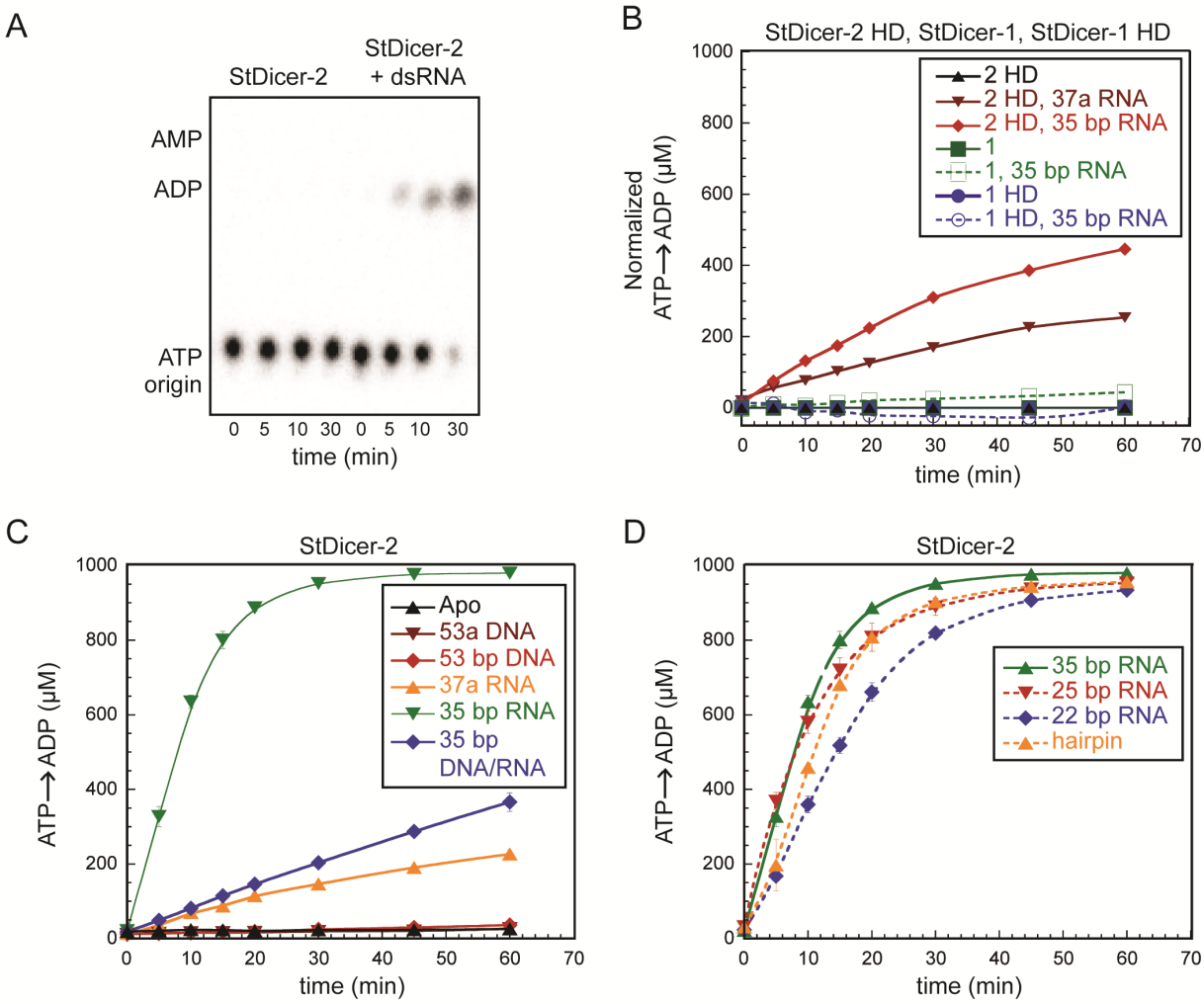

Figure 3.3 StDicer-2 and its isolated helicase domain show robust ATP hydrolysis in the presence of dsRNA. (A) ATP hydrolysis by StDicer-2 was monitored by thin-layer chromatography. StDicer-2 was incubated with [α-^{32}P]ATP and 35 bp dsRNA.(B) Isolated helicase domains (HD) recapitulate the ATP hydrolysis of the full-length protein. ATP hydrolysis of StDicer-2 HD, StDicer-1, and StDicer-1 HD was measured with different nucleic acid substrates. The 37a RNA was a representative ssRNA while the 35 bp RNA was a representative dsRNA. The lines for StDicer-2 HD without substrate, StDicer-1, and StDicer-1 HD all overlap. StDicer-1 and StDicer-2 were abbreviated as 1 and 2 respectively. (C) StDicer-2 displays robust ATP hydrolysis in the presence of dsRNA. ATP hydrolysis by StDicer-2 was measured with different nucleic acid substrates: 53 nt single-stranded (ss) DNA (53a), 53 bp DNA, 37 nt ssRNA (37a), 35 bp RNA, and 35 bp DNA/RNA heteroduplex. The lines for the protein alone (Apo), ssDNA, and dsDNA all overlap. (D) ATP hydrolysis by StDicer-2 is not sensitive to the length and structure of dsRNA. ATP hydrolysis by StDicer-2 was measure with different nucleic acid substrates: 35 bp RNA, 25 bp RNA, 22 bp RNA, and hairpin RNA. Error bars for each time point represent the standard deviation of three replicates in every graph in this figure and are not always visible.

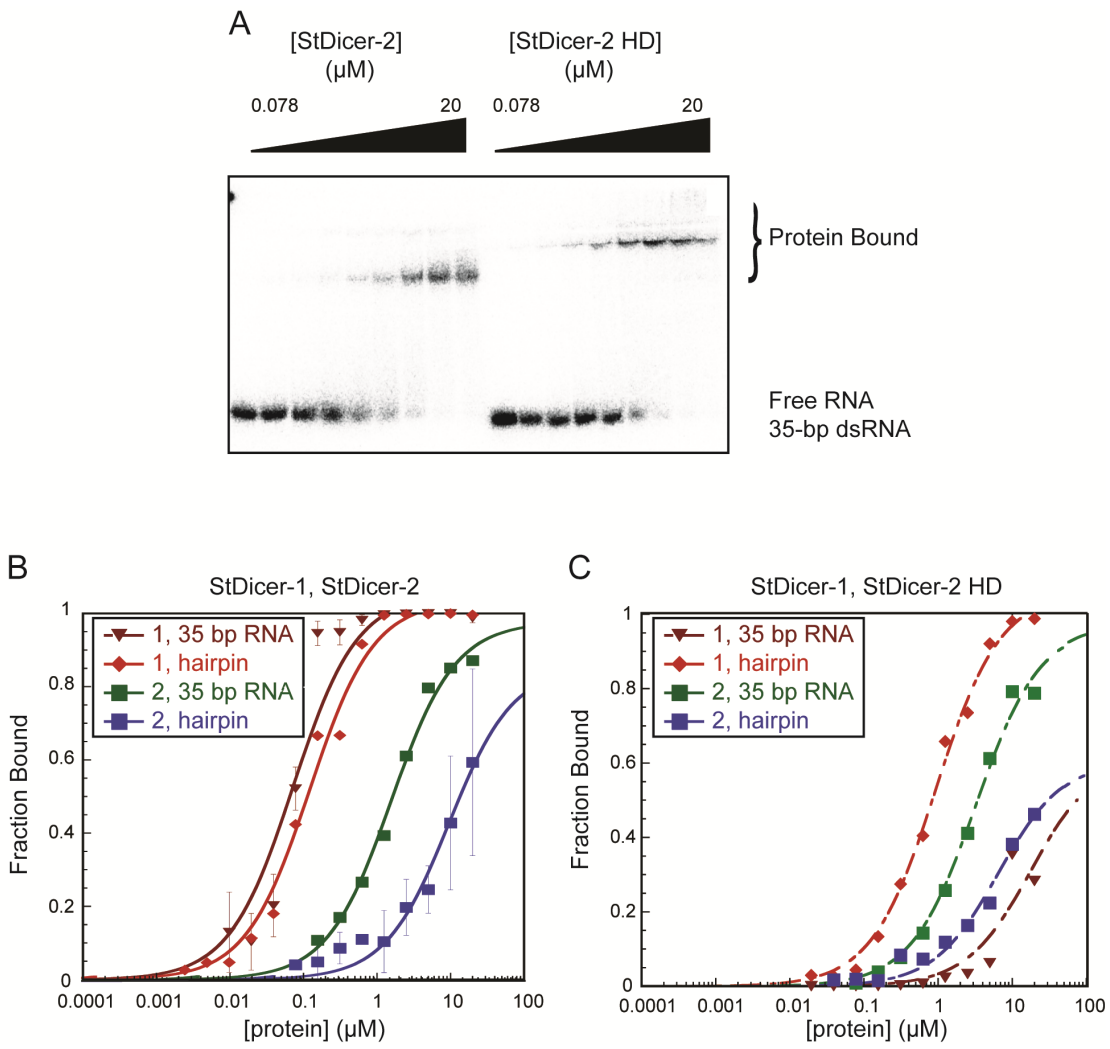

Figure 3.4 RNA binding properties of StDicer-1 and StDicer-2. (A) Representative gel shift for the proteins. Approximately 0.5-1 nM of radiolabeled RNA was incubated with protein varying in concentration from 0.078 μM to 20 μM. (B) StDicer-1 has a higher affinity than StDicer-2 for both duplex and hairpin RNA substrates. The affinity for two different RNAs, 35 bp and hairpin RNA, was measured for both StDicer-1 and StDicer-2 which were abbreviated as 1 and 2 respectively. (C) The affinity of the helicase domain for different substrates mirrors the affinity of their respective full-length proteins. The affinity for two different RNAs, 35 bp and hairpin RNA, was measured for the isolated helicase domains from StDicer-1 and StDicer-2 which were abbreviated as 1 and 2 respectively.

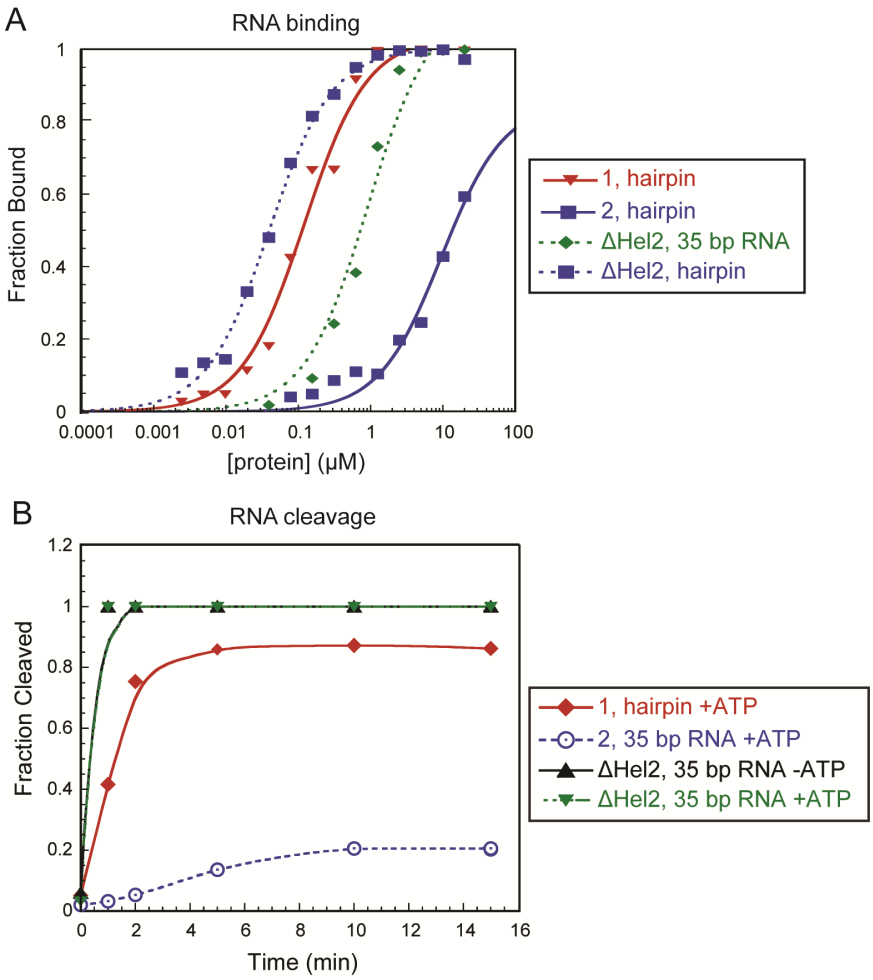

Figure 3.5 Removal of the helicase domain increases RNA affinity and the rate of RNA cleavage. (A) Removal of the helicase domain for StDicer-2 greatly increases the affinity for RNAs. The affinity for two different RNAs, 35 bp and hairpin RNA, was measured for ΔHel StDicer-2 (ΔHel2). The data for StDicer-1 (1) and StDicer-2 (2) are reproduced from Figure 3.2Bfor comparison. (B) Removal of the helicase domain increases the rate of RNA cleavage. The fraction of RNA cleaved is plotted for a 35 bp RNA with and without the presence of ATP and the graphs for the lines overlap. The data for StDicer-1 (1) and StDicer-2 (2) are reproduced from Figure 3.1B for comparison.

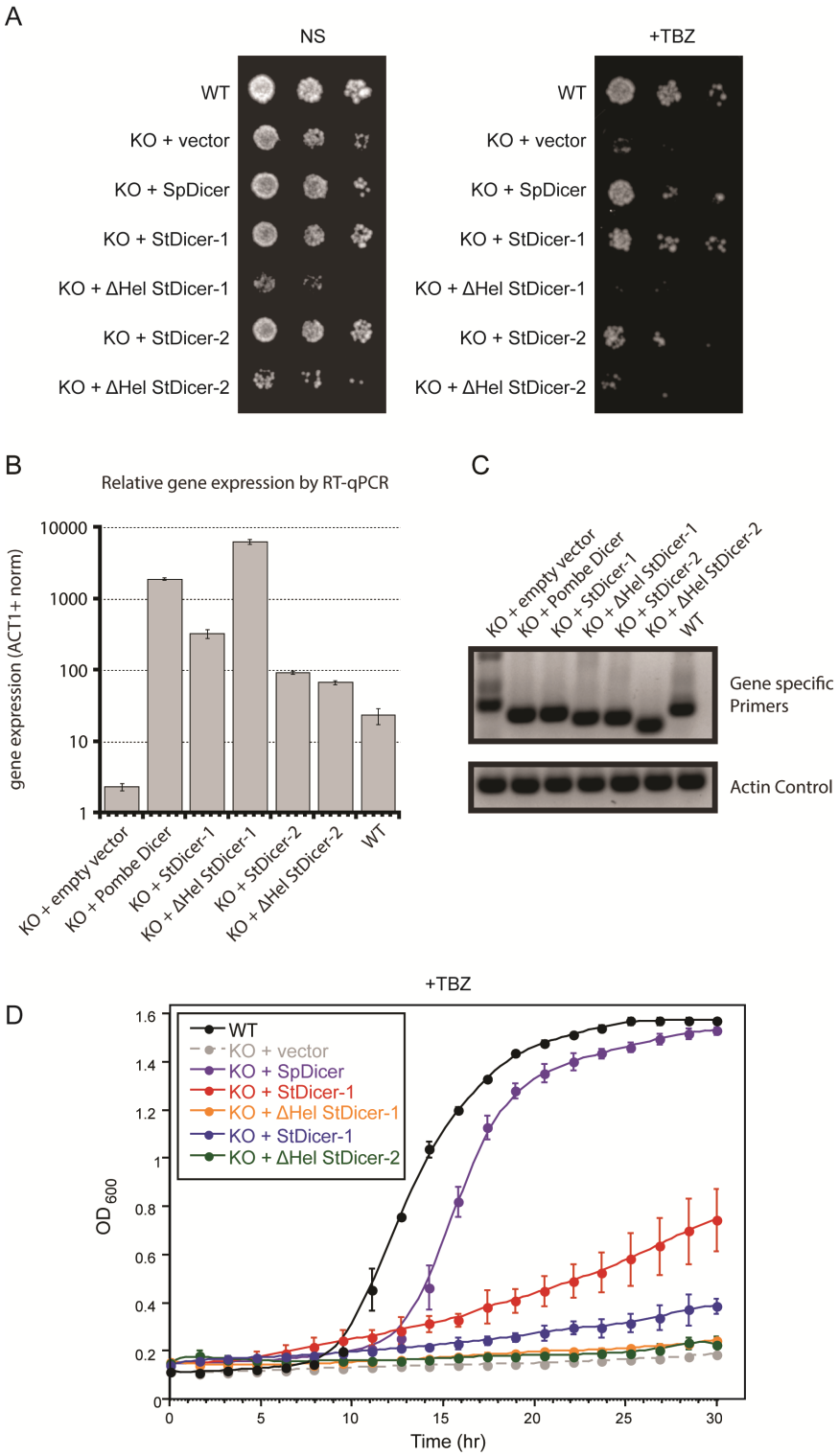

Figure 3.6 StDicer-1 and StDicer-2 complement growth defects *in vivo*. (A) Overexpression of StDicer-1 and StDicer-2 rescues the TBZ sensitivity of the *S. pombe* Dicer delete (KO). Knock-out cells were transformed with a vector expressing the indicated proteins and grown at 30°C. Growth was assayed by spotting 5-fold serial dilutions of the cultures indicated and plating on nonselective medium (NS) or medium

supplemented with 15 µg/ml TBZ. (B) Relative gene expression by RT-qPCR. Primers from Table 3.2 were used to quantify the relative levels of mRNA transcript that was either expressed from the plasmid or endogenously expressed. The gene specific primers were used for each StDicer construct and Pombe Dicer primers were used for KO + empty vector, KO + Pombe Dicer, and WT. All values were normalized to ACT+ levels and the standard deviations were derived from three replicates. (C) PCR products from RT-qPCR. The products from the RT-qPCR were run on a 1.5% agarose gel at the correct sizes were obtained for all reactions except for the empty vector control. The product obtained for the empty vector control was larger than expected and is 10-fold less than that obtained from the WT control. (D) Knock-out cells expressing StDicer-1 have higher growth rates than those expressing StDicer-2. Cells were grown in liquid culture containing 15 µg/ml TBZ and time points were taken every 1.5 hrs for 30 hrs. Error bars for each time point represent the standard deviation from three replicates.

Chapter 4

Summary

The overarching goal of this work has been to elucidate the molecular mechanisms by which miRNAs and siRNAs are recognized by Dicer and correctly cleaved in RNAi biogenesis. The essential and widespread effect of RNAi underscores the need to understand the function of these enzymes and will help guide future engineering of this pathway.

Dicer is a large, multi-functional protein and previous biochemical studies have focused on how its noncatalytic domains affect RNA cleavage. We have found that the helicase domain plays an evolutionarily conserved role in selecting RNAs for Dicer-mediated cleavage. We found that distinct biochemical activities of the isolated helicase domains from both Dicers contribute to the specialized functions of the full-length proteins. Recent studies have shown that the helicase domain, which cannot hydrolyze ATP, preferentially interacts with the loops of pre-miRNAs and inhibit cleavage of pre-siRNAs (Ma *et al.*, 2012; Tsutsumi *et al.*, 2011). This is supported by structural studies with human Dicer, which has an inactive helicase domain, demonstrating that pre-siRNAs are held in a nonproductive conformation between the PAZ and helicase domain (Taylor *et al.*, 2013). In addition, pre-miRNAs induce structural changes involving the helicase domain that enable productive substrate recognition and rapid cleavage (Ma *et al.*, 2012; Taylor *et al.*, 2013). Our ATP hydrolysis assays and RNA binding experiments with StDicer-1 support this hypothesis, as only StDicer-1, which bears an inactive helicase domain, could cleave pre-miRNAs.

Dicer does not act in isolation and in many systems requires partner proteins to accurately cleave its substrates. TRBP, PACT, and R3D1-L, with multiple dsRBDs, diffuse exclusively on dsRNA whereas Dicer with only one dsRBD shows transient static binding without diffusion on dsRNA. This diffusion activity arising from multiple tandem dsRBDs, could at least in part, unravel the unsolved biological questions regarding the role of dsRBPs. *In vitro*, TRBP facilitates Dicer's cleavage of pre-siRNA or pre-miRNA (Haase *et al.*, 2005; Chakravarthy *et al.*, 2010), but how it contributes to the function of Dicer has remained unknown. Based on the correlation between TRBP-driven diffusion of Dicer-TRBP complex and its cleavage activity, we propose that TRBP-induced diffusion could aid in positioning Dicer at the proper cleavage site (Figure 2.12F), resulting in an enhanced cleavage rate of Dicer-TRBP compared to Dicer alone (Haase *et al.*, 2005; Chakravarthy *et al.*, 2010). Dicer by itself will find the cleavage site through multiple trials of random binding and dissociation while Dicer-TRBP complex can find it more efficiently through a diffusion mechanism. In other words, Dicer-TRBP diffusion can serve to scan pre-miRNA and pre-siRNA substrates and to facilitate Dicer's catalytic activity by locating the complex at the RNA cleavage site with an improved efficiency and precision (MacRae *et al.*, 2006; MacRae *et al.*, 2007). It is also likely that TRBP's diffusion plays another role in RNA-mediated gene silencing, for instance, a recruitment of Dicer-TRBP complex to Ago2 or guide strand selection by dsRNA scanning (Noland *et al.*, 2011; Gredell *et al.*, 2010).

Together, these two stories have helped advance the understanding of substrate selection and cleavage by Dicer and its partner proteins.

BIBLIOGRAPHY

4th WTB, Iavarone AT, Hausmann CD, Cate JHD & Marletta MA (2011) Extracellular aldonolactonase from Myceliophthora thermophila. *Applied and environmental microbiology***77:** 650-6

Addo-Quaye C, Eshoo TW, Bartel DP & Axtell MJ (2008) Endogenous siRNA and miRNA Targets Identified by Sequencing of the Arabidopsis Degradome. *Current Biology***18:** 758-762

Adilakshmi T, Bellur DL & Woodson SA (2008) Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature***455:** 1268-1272

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G & Tuschl T (2003) A uniform system for microRNA annotation. *RNA***9:** 277-9

Arava Y, Wang Y, Storey JD, Liu CL, Brown PO & Herschlag D (2003) Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences***100:** 3889-94

Auyeung V, Ulitsky I, McGeary S & Bartel D (2013) Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell***152:** 844-858

Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, Doudna JA, Arkin AP & Pachter L (2011) Modeling and automation of sequencing-based characterization of RNA structure. *Proceedings of the National Academy of Sciences***108:** 11069-74

Babiarz JE, Ruby JG, Wang Y, Bartel DP & Blelloch R (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes & Development***22:** 2773-2785

Bartel DP (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell***136:** 215-233

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

Bentele K, Saffert P, Rauscher R, Ignatova Z, Bl&uuml N, Bl N, uuml, thgen & Blüthgen N (2013) Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology***9:**

Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RHA & Cuppen E (2005) Phylogenetic Shadowing and Computational Identification of Human microRNA Genes. *Cell***120:** 21-24

Berg OG & Ehrenberg M (1982) Association kinetics with coupled three- and one-dimensional diffusion. Chain-length dependence of the association rate of specific DNA sites. *Biophysical chemistry***15:** 41-51

Berg OG, Winter RB & Hippel von PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry***20:** 6929-48

Bernstein E, Caudy AA, Hammond SM & Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature***409:** 363-6

Blainey PC, van Oijen AM, Banerjee A, Verdine GL & Xie XS (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proceedings of the National Academy of Sciences***103:** 5752-7

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:** 1225-36

Bohnsack MT, Czaplinski K & Gorlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA***10:** 185-91

Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E & Basquin J (2009) The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. *Cell***139:** 547-559

Borchert GM, Lanier W & Davidson BL (2006) RNA polymerase III transcribes human microRNAs. *Nature structural & molecular biology***13:** 1097-1101

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R & Hannon GJ (2007) Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell***128:** 1089-1103

Bullock SL, Ringel I, Ish-Horowicz D & Lukavsky PJ (2010) A′-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. *Nature structural & molecular biology***17:** 703-709

CLOTE P (2005) Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. **11:** 578-591

Cai X, Hagedorn CH & Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA***10:** 1957-66

Castanotto D & Rossi JJ (2009) The promises and pitfalls of RNA-interference-based therapeutics. *Nature***457:** 426-433

Catalanotto C, Pallotta M, ReFalo P, Sachs MS, Vayssie L, Macino G & Cogoni C (2004) Redundancy of the Two Dicer Genes in Transgene-Induced Posttranscriptional Gene Silencing in Neurospora crassa. *Molecular and Cellular Biology***24:** 2536-2545

Cech T & Steitz J (2014) The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell***157:** 77-94

Cenik E, Fukunaga R, Lu G, Dutcher R, Wang Y, Hall TT & Zamore P (2011) Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease. *Molecular Cell***42:** 172-184

Cerutti H & Casas-Mollano JA (2006) On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics***50:** 81-99

Chakravarthy S, Sternberg SH, Kellenberger CA & Doudna JA (2010) Substrate-Specific Kinetics of Dicer-Catalyzed RNA Processing. *Journal of molecular biology***404:** 392-402

Chartrand P, Meng XH, Huttelmaier S, Donato D & Singer RH (2002) Asymmetric sorting of ash1p in yeast results from inhibition of translation by localization elements in the mRNA. *Molecular Cell***10:** 1319-30

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

Chen J, Murchison EP, Tang R, Callis TE, Tatsuguchi M, Deng Z, Rojas M, Hammond SM, Schneider MD, Selzman CH, Meissner G, Patterson C, Hannon GJ & Wang D (2008) Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure. *Proceedings of the National Academy of Sciences***105:** 2111-6

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

Chi SW, Zang JB, Mele A & Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature***479:** 479-486

Chowdhury S, Maris C, Allain FH & Narberhaus F (2006) Molecular basis for temperature sensing by an RNA thermometer. *The EMBO Journal***25:** 2487-97

Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND, Wolfe SA & Giraldez AJ (2010) A Novel miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. *Science***328:** 1694-1698

Civril F, Bennett M, Moldt M, Deimling T, Witte G, Schiesser S, Carell T & Hopfner K (2011) The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO reports***12:** 1127-1134

Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E, Krishnan K, Gardiner B, Wang X, Nones K, Steen JA, Matigian NA,

Wood DL, Kassahn KS, Waddell N, Shepherd J, et al. (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biology***12:** R126

Cook AG & Conti E (2010) Nuclear export complexes in the frame. *Current Opinion in Structural Biology***20:** 247-252

Cruz JA & Westhof E (2009) The Dynamic Landscapes of RNA Architecture. *Cell***136:** 604-609

Current protocols in nucleic acid chemistry

Dai X, Zhuang Z & Zhao PX (2011) Computational analysis of miRNA targets in plants: current status and challenges. *Briefings in Bioinformatics***12:** 115-121

Dana A & Tuller T (2012) Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells. *PLoS Computational Biology***8:** e1002755

Das R, R. das, Kudaravalli M, Jonikas M, Fong R, Schwans JP, Baker D, Piccirilli JA, Altman RB & Herschlag D (2008) Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proceedings of the National Academy of Sciences***105:** 4144-4149

Daviet L, Erard M, Dorin D, Duarte M, Vaquero C & Gatignol A (2000) Analysis of a binding difference between the two dsRNA-binding domains in TRBP reveals the modular function of a KR-helix motif. *FEBS Journal***267:** 2419-31

Denli AM, Tops BBJ, Plasterk RHA, Ketting RF & Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature***432:** 231-5

Dienstbier M, Boehl F, Li X & Bullock SL (2009) Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes & Development***23:** 1546-1558

Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC & Assmann SM (2013) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature***505:** 696-700

Disney MD, Childs JL, Schroeder SJ, Zuker M & Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences***101:** 7287-7292

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.

Dorin D, Bonnet MC, Bannwarth S, Gatignol A, Meurs EF & Vaquero C (2003) The TAR RNA-binding Protein, TRBP, Stimulates the Expression of TAR-containing RNAs in Vitro and in VivoIndependently of Its Ability to Inhibit the dsRNA-dependent Kinase PKR. *Journal of Biological Chemistry***278:** 4440-4448

Doyle M & Jantsch MF (2003) Distinct in vivo roles for double-stranded RNA-binding domains of the Xenopus RNA-editing enzyme ADAR1 in chromosomal targeting. *The Journal of cell biology***161:** 309-19

Elkayam E, Kuhn C, Tocilj A, Haase A, Greene E, Hannon G & Joshua-Tor L (2012) The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell***150:** 100-110

Faller M, Matsunaga M, Yin S, Loo JA & Guo F (2006) Heme is involved in microRNA processing. *Nature structural & molecular biology***14:** 23-29

Fang X, Wang J, O'Carroll I, Mitchell M, Zuo X, Wang Y, Yu P, Liu Y, Rausch J, Dyba M, Kjems J, Schwieters C, Seifert S, R, Winans A, Watts N, Stahl S, Wingfield P, Byrd R, Le Grice SJ, et al. (2013) An Unusual Topological Structure of the HIV-1 Rev Response Element. *Cell***155:** 594-605

Ferré-D'Amaré AR & Doudna JA (1996) Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Research***24:** 977-8

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:** 806-11

Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA & Imler J (2006) Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. *Nature Immunology***7:** 590-7

Gantier MP & Williams BRG (2007) The response of mammalian cells to double-stranded RNA. *Cytokine & Growth Factor Reviews***18:** 363-371

Garneau NL, Wilusz J & Wilusz CJ (2007) The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology***8:** 113-126

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D & Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell***11:** 4241-57

Gatignol A, Buckler C & Jeang KT (1993) Relatedness of an RNA-binding motif in human immunodeficiency virus type 1 TAR RNA-binding protein TRBP to human P1/dsI kinase and Drosophila staufen. *Molecular and Cellular Biology***13:** 2193-202

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:** 1597-600

Gee P, Chua PK, Gevorkyan J, Klumpp K, Najera I, Swinney DC & Deval J (2008) Essential Role of the N-terminal Domain in the Regulation of RIG-I ATPase Activity. *Journal of Biological Chemistry***283:** 9488-9496

German MA, Pillay M, Jeong D, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, de Paoli E, Lu C, Schroth G, Meyers BC & Green PJ (2008) Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnology***26:** 941-946

Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature***425:** 737-41

Goodman DB, Church GM & Kosuri S (2013) Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science***342:** 475-479

Gorman J, Plys AJ, Visnapuu M, Alani E & Greene EC (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nature structural & molecular biology***17:** 932-938

Graneli A, Yeykal CC, Robertson RB & Greene EC (2006) Long-distance lateral diffusion of human Rad51 on double-stranded DNA. *Proceedings of the National Academy of Sciences***103:** 1221-1226

Gredell JA, Dittmer MJ, Wu M, Chan C & Walton SP (2010) Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry***49:** 3148-55

Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH & Ecker JR (2008) A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. *Developmental Cell***14:** 854-866

Gregory RI, Chendrimada TP, Cooch N & Shiekhattar R (2005) Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell***123:** 631-640

Grundhoff A, Sullivan CS & Ganem D (2006) A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. *RNA***12:** 733-50

Gu W, Zhou T & Wilke CO (2010) A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLoS Computational Biology***6:** e1000664

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 reports***6:** 961-967

Halvorsen M, Martin JS, Broadaway S & Laederach A (2010) Disease-Associated Mutations That Alter the RNA Structural Ensemble. *PLoS Genetics***6:** e1001074

Hamilton AJ & Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science***286:** 950-2

Han J, Lee Y, Yeom K, Nam J, Heo I, Rhee J, Sohn SY, Cho Y, Zhang B & Kim NV (2006) Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell***125:** 887-901

Han Y, David A, Liu B, Magadán JG, Bennink JR, Yewdell JW & Qian S (2012) Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proceedings of the National Academy of Sciences***109:** 12467-12472

Hauptmann J, Dueck A, Harl S, er, Pfaff J, Merkl R, Meister G & Harlander S (2013) Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nature structural & molecular biology***20:** 814-817

(2007) A microRNA component of the p53 tumour suppressor network. *Nature***447:** 1130-1134

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ & Hammond SM (2005) A microRNA polycistron as a potential human oncogene. *Nature***435:** 828-833

Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Research***31:** 3429-3431

Hohng S, Joo C & Ha T (2004) Single-Molecule Three-Color FRET. *Biophysical Journal***87:** 1328-1337

Holt CE & Bullock SL (2009) Subcellular mRNA Localization in Animal Cells and Why It Matters. *Science***326:** 1212-1216

Hur JK, Zinchenko MK, Djuranovic S & Green R (2013) Regulation of Argonaute Slicer Activity by Guide RNA 3' End Interactions with the N-terminal Lobe. *Journal of Biological Chemistry***288:** 7829-7840

Hutvagner G & Simard MJ (2008) Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology***9:** 22-32

Hwang H, Kim H & Myong S (2011) Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proceedings of the National Academy of Sciences***108:** 7414-8

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. *Molecular Cell***39:** 282-291

Ingolia NT, Ghaemmaghami S, Newman JRS & Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science***324:** 218-23

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. *Molecular Cell***39:** 292-299

Jain A, Liu R, Ramani B, Arauz E, Ishitsuka Y, Ragunathan K, Park J, Chen J, Xiang YK & Ha T (2011) Probing cellular protein complexes using single-molecule pull-down. *Nature***473:** 484-488

Jambhekar A & DeRisi J (2007) Cis-acting determinants of asymmetric, cytoplasmic RNA transport. *RNA***13:** 625-642

Jiang F, An, Ramanathan, Miller MT, Tang G, Gale M, Patel SS, Marcotrigiano J & Ramanathan A (2011) Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature***479:** 423-427

Jiang F, Ye X, Liu X, Fincher L, McKearin D & Liu Q (2005) Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes & Development***19:** 1674-9

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

Kadotani N, Nakayashiki H, Tosa Y & Mayama S (2004) One of the two Dicer-like proteins in the filamentous fungi Magnaporthe oryzae genome is responsible for hairpin RNA-triggered RNA silencing and related small interfering RNA accumulation. *The Journal of biological chemistry***279:** 44467-74

Katz L (2003) Widespread Selection for Local RNA Secondary Structure in Coding Regions of Bacterial Genes. *Genome Research***13:** 2042-2051

Kelly BN, Howard BR, Wang H, Robinson H, Sundquist WI & Hill CP (2006) Implications for viral capsid assembly from crystal structures of HIV-1 Gag(1-278) and CA(N)(133-278). *Biochemistry***45:** 11257-11266

Kertesz M, Iovino N, Unnerstall U, Gaul U & Segal E (2007) The role of site accessibility in microRNA target recognition. *Nature Genetics***39:** 1278-1284

Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY & Segal E (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature***467:** 103-107

Kielpinski LJ & Vinther J (2014) Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility.

Kim Y & Kim VN (2007) Processing of intronic microRNAs. *The EMBO journal***26:** 775-83

Kladwang W, VanLang CC, Cordero P & Rhiju das (2011) A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nature Chemistry***3:** 954-962

Komar AA (2009) A pause for thought along the co-translational folding pathway. *Trends in Biochemical Sciences***34:** 16-24

Kortmann J & Narberhaus F (2012) Bacterial RNA thermometers: molecular zippers and switches. *Nature Reviews Microbiology***10:** 255-265

Kowalinski E, Lunardi T, McCarthy A, 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 (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene***361:** 13-37

Kozomara A & Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research***42:** D68-73

Krovat BC & Jantsch MF (1996) Comparative Mutational Analysis of the Double-stranded RNA Binding Domains of Xenopus laevis RNA-binding Protein A. *Journal of Biological Chemistry***271:** 28112-28119

Kudla G, Murray AW, Tollervey D & Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science***324:** 255-8

Kwak PB & Tomari Y (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature structural & molecular biology***19:** 145-151

Kwok CK, Ding Y, Tang Y, Assmann SM & Bevilacqua PC (2013) Determination of in vivo RNA structure in low-abundance transcripts. *Nature Communications***4:** 2971

Laederach A, Shcherbakova I, Jonikas MA, Altman RB & Brenowitz M (2007) Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proceedings of the National Academy of Sciences***104:** 7045-7050

Lau P, Guiley KZ, de De N, Potter CS, Carragher B, MacRae IJ & de Nabanita (2012) The molecular architecture of human Dicer. *Nature structural & molecular biology***19:** 436-440

Lee H, Li, Li E, Gu W, Xue Z, Crosthwaite SK, Alex, Pertsemlidis E, Lewis ZA, Freitag M, Selker EU, Mello CC, Liu Y, Li L & Pertsemlidis A (2010) Diverse Pathways Generate MicroRNA-like RNAs and Dicer-Independent Small Interfering RNAs in Fungi. *Molecular Cell***38:** 803-814

Lee HY & Doudna JA (2012) TRBP alters human precursor microRNA processing in vitro. *RNA***18:** 2012-2019

Lee HY, Zhou K, Smith AM, Nol CL, Noland CL & Doudna JA (2013) Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Research***41:** 6568-76

Lee J, Kladwang W, Lee M, Cantu D, Azizyan M, Kim H, Limpaecher A, Yoon S, Treuille A & R. das (2014) RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences***111:** 2122-2127

Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ & Gasch AP (2011) A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Molecular Systems Biology***7:**

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:** 843-54

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S & Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature***425:** 415-9

Lee Y, Hur I, Park S, Kim Y, Suh MR & Kim VN (2006) The role of PACT in the RNA silencing pathway. *The EMBO journal***25:** 522-32

Lee Y, Kim M, Han J, Yeom K, Lee S, Baek SH & Kim VN (2004a) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal***23:** 4051-60

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:** 69-81

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

Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, Aiyer S, Valladares O, Yang J, Bambina S, Sabin L, Murray J, Lamitina T, Raj A, Cherry S, Wang L & Gregory B (2012a) Global Analysis of RNA Secondary Structure in Two Metazoans. *Cell Reports***1:** 69-82

Li J, Liu S, Zhou H, Qu L & Yang J (2014) starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research***42:**

Lindner K, Gregán J, Montgomery S & Kearsey SE (2002) Essential role of MCM proteins in premeiotic DNA replication. *Molecular biology of the cell***13:** 435-44

Liu S, Abbondanzieri EA, Rausch JW, Le Grice SFJ & Zhuang X (2008) Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science***322:** 1092-7

Low JT, Knoepfel SA, Watts JM, Brake ter O, Berkhout B & Weeks KM (2012) SHAPE-directed Discovery of Potent shRNA Inhibitors of HIV-1. *Molecular Therapy***20:** 820-828

Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W & Cui Q (2008) An analysis of human microRNA and disease associations. *PloS one***3:** e3420

Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA & Arkin AP (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proceedings of the National Academy of Sciences***108:** 11063-8

Lund E, Güttinger S, Calado A, Dahlberg JE & Kutay U (2004) Nuclear export of microRNA precursors. *Science***303:** 95-8

Luo D, Ding S, Vela A, Kohlway A, Lindenbach B & Pyle A (2011) Structural Insights into RNA Recognition by RIG-I. *Cell***147:** 409-422

Lécuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P & Krause HM (2007) Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell***131:** 174-187

Ma E, MacRae IJ, Kirsch JF & Doudna JA (2008) Autoinhibition of Human Dicer by Its Internal Helicase Domain. *Journal of molecular biology***380:** 237-243

Ma E, Zhou K, Doudna JA & Kidwell MA (2012) Coordinated Activities of Human Dicer Domains in Regulatory RNA Processing. *Journal of molecular biology***422:** 466-476

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:** 512-7

MacRae IJ, Zhou K & Doudna JA (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nature structural & molecular biology***14:** 934-940

Macfarlane L & Murphy PR (2010) MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics***11:** 537-61

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:** 195-8

Manche L, Green SR, Schmedt C & Mathews MB (1992) Interactions between double-stranded RNA regulators and the protein kinase DAI. *Molecular and Cellular Biology***12:** 5238-48

Maniataki E & Mourelatos Z (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes & Development***19:** 2979-90

Mao Y, Wang W, Cheng N, Li Q & Tao S (2013) Universally increased mRNA stability downstream of the translation initiation site in eukaryotes and prokaryotes. *Gene***517:** 230-235

Martin KC & Ephrussi A (2009) mRNA Localization: Gene Expression in the Spatial Dimension. *Cell***136:** 719-730

Mauger DM, Siegfried NA & Weeks KM (2013) The genetic code as expressed through relationships between mRNA structure and protein function. *FEBS Letters***587:** 1180-1188

McManus CJ & Graveley BR (2011) RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development***21:** 373-379

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G & Tuschl T (2004) Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. *Molecular Cell***15:** 185-197

Mercer TR & Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology***20:** 300-307

Merino EJ, Wilkinson KA, Coughlan JL & Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society***127:** 4223-31

Meyer IM & Miklós I (2005) Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Research***33:** 6338-48

Meyer M, Plass M, Pérez-Valle J, Eyras E & Vilardell J (2011) Deciphering 3′ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. *Molecular Cell***43:** 1033-1039

Micklem DR, Adams J, Grünert S & Johnston DS (2000) Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. *The EMBO journal***19:** 1366-77

Moran Y, Praher D, Fredman D & Technau U (2013) The Evolution of MicroRNA Pathway Protein Components in Cnidaria. *Molecular Biology and Evolution***30:** 2541-2552

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ & Marra MA (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research***18:** 610-621

Mortimer SA & Weeks KM (2007) A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. **129:** 4144-4145

Mukherjee K, Campos H & Kolaczkowski B (2013) Evolution of Animal and Plant Dicers: Early Parallel Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants. *Molecular Biology and Evolution***30:** 627-641

Myong S, Cui S, Cornish PV, Kirchhofer A, Gack MU, Jung JU, Hopfner K- & Ha T (2009) Cytosolic Viral Sensor RIG-I Is a 5'-Triphosphate-Dependent Translocase on Double-Stranded RNA. *Science***323:** 1070-1074

Myong S, Rasnik I, Joo C, Lohman TM & Ha T (2005) Repetitive shuttling of a motor protein on DNA. *Nature***437:** 1321-1325

Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W & Diatchenko L (2006) Human Catechol-O-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure. *Science***314:** 1930-1933

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

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. *The Plant Cell***2:** 279-289

Nejepinska J, Malik R, Filkowski J, Flemr M, Filipowicz W & Svoboda P (2012) dsRNA expression in the mouse elicits RNAi in oocytes and low adenosine deamination in somatic cells. *Nucleic Acids Research***40:** 399-413

Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL & Weissman JS (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature***441:** 840-846

Noland C, Ma E & Doudna J (2011) siRNA Repositioning for Guide Strand Selection by Human Dicer Complexes. *Molecular Cell***43:** 110-121

Novikova IV, Hennelly SP, Tung C & Sanbonmatsu KY (2013) Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs. *Journal of molecular biology***425:** 3731-3746

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

Ouellet DL, Plante I, Landry P, Barat C, Janelle M-, Flamand L, Tremblay MJ & Provost P (2008) Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. *Nucleic Acids Research***36:** 2353-2365

Ouyang Z, Snyder MP & Chang HY (2013) SeqFold: Genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Research***23:** 377-387

Palazzo AF, Palazzo EF, Springer M, Shibata Y, Lee C, Dias AP & Rapoport TA (2007) The signal sequence coding region promotes nuclear export of mRNA. *PLoS biology***5:** e322

Panniers R (1994) Translational control during heat shock. *Cell***76:** 737-747

Parker GS, Maity TS & Bass BL (2008) dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *Journal of molecular biology***384:** 967-79

Petri S, Dueck A, Lehmann G, Putz N, Rudel S, Kremmer E & Meister G (2011) Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. *RNA***17:** 737-749

Provost P, Silverstein RA, Dishart D, Walfridsson J, Djupedal I, Kniola B, Wright A, Samuelsson B, Radmark O & Ekwall K (2002) Dicer is required for chromosome segregation and gene silencing in fission yeast cells. *Proceedings of the National Academy of Sciences***99:** 16648-53

Riggs AD, Bourgeois S & Cohn M (1970) The lac repressor-operator interaction. 3. Kinetic studies. *Journal of molecular biology***53:** 401-17

Romano N & Macino G (1992) Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular microbiology***6:** 3343-53

Rouskin S, Zubradt M, Washietl S, Kellis M & Weissman JS (2013) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature***505:** 701-705

Roy R, Alex, Kozlov EG, Lohman TM, Ha T & Kozlov AG (2009) SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature***461:** 1092-1097

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP & Lai EC (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Research***17:** 1850-1864

Ryter JM & Schultz SC (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *The EMBO journal***17:** 7505-13

Saini HK, Griffiths-Jones S & Enright AJ (2007) Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences***104:** 17719-24

Saito K, Ishizuka A, Siomi H & Siomi MC (2005) Processing of Pre-microRNAs by the Dicer-1–Loquacious Complex in Drosophila Cells. *PLoS biology***3:** e235

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

Saunders LR & Barber GN (2003) The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB***17:** 961-83

Scharff LB, Childs L, Walther D & Bock R (2011) Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. *PLoS Genetics***7:** e1002155

Schirle NT & MacRae IJ (2012) The Crystal Structure of Human Argonaute2. *Science***336:** 1037-1040

Schmedt C, Green SR, Manche L, Taylor DR, Ma Y & Mathews MB (1995) Functional characterization of the RNA-binding domain and motif of the double-stranded RNA-dependent protein kinase DAI (PKR). *Journal of molecular biology***249:** 29-44

Schopman NCT, Willemsen M, Liu YP, Bradley T, van Kampen A, Baas F, Berkhout B & Haasnoot J (2012) Deep sequencing of virus-infected cells reveals HIV-encoded small RNAs. *Nucleic Acids Research***40:** 414-27

Schwarz DS, Hutvágner G, Tingting du, Xu Z, Aronin N & Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell***115:** 199-208

Schürmann N, Trabuco LG, Bender C, Russell RB & Grimm D (2013) Molecular dissection of human Argonaute proteins by DNA shuffling. *Nature structural & molecular biology***20:** 818-826

Seetin M, Kladwang W, Bida J & R (2014) Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol. Biol.***1086:** 95-117

Shabalina SA, Ogurtsov AY & Spiridonov NA (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Research***34:** 2428-37

Shah P, Ding Y, Niemczyk M, Kudla G & Plotkin J (2013) Rate-Limiting Steps in Yeast Protein Translation. *Cell***153:** 1589-1601

Sharp PA (2009) The Centrality of RNA. *Cell***136:** 577-580

Silverman IM, Li F & Gregory BD (2013) Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants. *Plant Science***205-206:** 55-62

Smalheiser NR (2003) EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biology***4:** 403

Smit M & Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. In *Proceedings of the National*

Smith MA, Gesell T, Stadler PF & Mattick JS (2013) Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Research***41:** 8220-8236

Sohn SY, Bae WJ, Kim JJ, Yeom K, Kim VN & Cho Y (2007) Crystal structure of human DGCR8 core. *Nature structural & molecular biology***14:** 847-853

Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET & Chang HY (2012) RNA SHAPE analysis in living cells. *Nature Chemical Biology***9:** 18-20

Starega-Roslan J, Krol J, Koscianska E, Kozlowski P, Szlachcic WJ, Sobczak K & Krzyzosiak WJ (2011) Structural basis of microRNA length variety. *Nucleic Acids Research***39:** 257-68

Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY & Batey RT (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure***18:** 787-97

Svoboda P (2014) Renaissance of mammalian endogenous RNAi. *FEBS Letters***588:** 2550-6

Taylor DW, Ma E, Shigematsu H, Cianfrocco MA, Nol CL, Noland CL, Nagayama K, Nogales E, Doudna JA & Wang H (2013) Substrate-specific structural rearrangements of human Dicer. *Nature structural & molecular biology***20:** 662-70

Tijerina P, Mohr S & Russell R (2007) DMS footprinting of structured RNAs and RNA–protein complexes. *Nature Protocols***2:** 2608-2623

Tsutsumi A, Kawamata T, Izumi N, Seitz H & Tomari Y (2011) Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nature structural & molecular biology***18:** 1153-1158

Tuller T, Waldman YY, Kupiec M & Ruppin E (2010) Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences***107:** 3645-3650

Tyrrell J, McGinnis JL, Weeks KM & Pielak GJ (2013) The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry***52:** 8777-8785

Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR & Haussler D (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. **7:** 995-1001

Vandivier L, Li F, Zheng Q, Willmann M, Chen Y & Gregory B (2013) Arabidopsis mRNA secondary structure correlates with protein function and domains. *Plant Signaling & Behavior***8:** e24301

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS & Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science***297:** 1833-7

Wan Y, Kertesz M, Spitale RC, Segal E & Chang HY (2011) Understanding the transcriptome through RNA structure. *Nature Reviews Genetics***12:** 641-655

Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, Tibshirani R, Makino D, Nutter R, Segal E & Chang H (2012) Genome-wide Measurement of RNA Folding Energies. *Molecular Cell***48:** 169-181

Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E & Chang HY (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature***505:** 706-709

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

Warf MB & Berglund JA (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends in Biochemical Sciences***35:** 169-178

Waterhouse PM, Wang MB & Lough T (2001) Gene silencing as an adaptive defence against viruses. *Nature***411:** 834-42

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM & Bess JW Jr (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature***460:** 711-716

Weeks KM (2010) Advances in RNA structure analysis by chemical probing. *Current Opinion in Structural Biology***20:** 295-304

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

Wells SE, Hughes JM, Igel AH & Ares M Jr (2000) Use of dimethyl sulfate to probe RNA structure in vivo. *Methods in enzymology***318:** 479-93

Wen J, Lancaster L, Hodges C, Zeri A, Yoshimura SH, Noller HF, Bustamante C & Tinoco I (2008) Following translation by single ribosomes one codon at a time. *Nature***452:** 598-603

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:** 855-62

Willmann MR, Berkowitz ND & Gregory BD (2013) Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes 2.0. **:** 1-10

Willmann MR, Berkowitz ND & Gregory BD (2014) Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes--GMUCT 2.0. *Methods***67:** 64-73

Wilson RC & Doudna JA (2013) Molecular Mechanisms of RNA Interference. *Annual Review of Biophysics***42:** 217-239

Winter RB, Berg OG & Hippel von PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. *Biochemistry***20:** 6961-77

Wolin SL & Walter P (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO Journal***7:** 3559-69

Worthylake DK, Wang H, Yoo S, Sundquist WI & Hill CP (1999) Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr D Biol Crystallogr***55:** 85-92

Xie M, Li M, Vilborg A, Lee N, Shu M, Yartseva V, Šestan N & Steitz J (2013) Mammalian 5′-Capped MicroRNA Precursors that Generate a Single MicroRNA. *Cell***155:** 1568-1580

Yamashita S, Nagata T, Kawazoe M, Takemoto C, Kigawa T, Güntert P, Kobayashi N, Terada T, Shirouzu M, Wakiyama M, Muto Y & Yokoyama S (2011) Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein science***20:** 118-30

Yang J, Li J, Shao P, Zhou H, Chen Y & Qu L (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research***39:** D202-9

Yi R, Qin Y, Macara IG & Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development***17:** 3011-6

Zeng Y & Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Research***32:** 4776-85

Zeng Y & Cullen BR (2005) Efficient Processing of Primary microRNA Hairpins by Drosha Requires Flanking Nonstructured RNA Sequences. *Journal of Biological Chemistry***280:** 27595-27603

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 Q, Ryvkin P, Li F, Dragomir I, Valladares O, Yang J, Cao K, Wang L & Gregory BD (2010) Genome-Wide Double-Stranded RNA Sequencing Reveals the Functional Significance of Base-Paired RNAs in Arabidopsis. *PLoS Genetics***6:** e1001141

Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE & Yeo GW (2010) Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. *Nature structural & molecular biology***17:** 173-179

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research***31:** 3406-3415

Zur H & Tuller T (2012) Strong association between mRNA folding strength and protein abundance in S. cerevisiae. *EMBO reports***13:** 272-277

de Jong D, Eitel M, Jakob W, Osigus H-, Hadrys H, DeSalle R & Schierwater B (2009) Multiple Dicer Genes in the Early-Diverging Metazoa. *Molecular Biology and Evolution***26:** 1333-1340

de Nabanita, Young L, Lau P, Meisner N, Morrissey DV & MacRae IJ (2013) Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. *Molecular Cell***50:** 344-55

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. *The Plant Cell***2:** 291-9

Appendix I

Insights into RNA structure and function from genome-wide studies

*A portion of the work presented in this chapter has been previously published as part of the following paper: Mortimer SA**, Kidwell MA**& Doudna JA (2014) Insights into RNA structure and function from genome-wide studies. *Nature Reviews Genetics***15(7):** 469-79.

**Stefanie Mortimer and Mary Anne Kidwell contributed equally to this work.

I.1 Introduction

RNA molecules influence all steps of gene expression and regulation (Sharp, 2009). Once thought to be only a cellular messenger between DNA and proteins, RNA is now known to participate in many aspects of cellular physiology through activities attributable to its secondary structures and tertiary structures. The ability of RNA strands to fold back on themselves to form stable three-dimensional architectures are fundamental to RNA transcription, splicing, translation, localization and turnover (Cruz & Westhof, 2009; Warf & Berglund, 2010; McManus & Graveley, 2011; Kozak, 2005; Martin & Ephrussi, 2009; Garneau *et al.*, 2007; Mauger *et al.*, 2013).

Classic techniques for biochemical interrogation of RNA structure focus on probing one type of RNA molecule at a time and have proved highly valuable to our understanding of RNA in biology and engineering applications (Merino *et al.*, 2005; Laederach *et al.*, 2007; Adilakshmi *et al.*, 2008; Weeks, 2010). These methods use small molecules or enzymes that react with the RNA in a structure-specific fashion and are read out using gel electrophoresis. However, given the number and types of coding and non-coding RNAs that comprise the transcriptome, these methods have lagged far behind the demand for knowledge about RNA structure and its role in cellular activities. To meet these demands, computational methods have been developed to accurately predict the formation of loops and simple helices (Hofacker, 2003; Zuker, 2003; Disney *et al.*, 2004; Das *et al.*, 2008; Ouyang *et al.*, 2013; Smith *et al.*, 2013) and the recent advent of crowdsourcing has further aided the prediction of more complex motifs (Lee *et al.*, 2014). While computational approaches have helped to increase the speed of RNA structure modelling, the computationally predicted structures still need to be validated by analysis of individual molecules.

Recently, several different high-throughput techniques have been developed that enable more facile and accurate structure prediction of large numbers of transcripts (Kertesz *et al.*, 2010; Zheng *et al.*, 2010; Underwood *et al.*, 2010; Lucks *et al.*, 2011; Li *et al.*, 2012a; 2012b; Wan *et al.*, 2012; Ding *et al.*, 2013; Rouskin *et al.*, 2013; Wan *et al.*, 2014). The field of high-throughput RNA structure probing and determination is moving at a rapid pace; we have highlighted these methods in Table I.1, and further information can be found in recent reviews (Wan *et al.*, 2011; Silverman *et al.*, 2013). In simple terms, these high-throughput techniques involve the marriage of traditional probes of RNA structure with new advances in next-generation sequencing (NGS). The application of these methods to transcriptomes has yielded the first glimpse of the 'RNA structurome' and has highlighted on a genome-wide scale how structured regions control RNA functions and gene expression.

With genome-wide RNA secondary structure analysis, it is possible to identify structural patterns and motifs in coding and noncoding RNAs from various organisms. This Review focuses on new insights into the regulation of gene expression as determined from investigations of the nature and abundance of structured regions within the transcriptomes of cells and viruses. We focus on the roles of these RNA structures in RNA translation, localization, stability, microRNA-mediated regulation, and splicing. We also explore the future of these techniques for examining structures *in vivo*.

I.2 RNA structures in translational control

It has long been postulated that RNA structure helps regulate each step of translation, from initiation to termination (Kozak, 2005). With recent methods using high-throughput sequencing, it has become possible to closely examine the relationship between RNA structure and translational control in eukaryotic systems. In this section, we discuss the contribution of RNA structure to various aspects of translation including efficiency, initiation, ribosome pausing, and termination.

I.2.1 Overall structure in the coding and non-coding regions of mRNAs

Systematic differences between the coding and non-coding regions of mRNAs can include patterns in sequence content, phylogenetic conservation and predicted secondary structures. Such evolutionarily conserved structural patterns in coding sequences imply functional importance that in at least some cases has been linked to translation regulation (Katz, 2003). In the first whole-genome structure analysis, which was performed on genomic RNA extracted from HIV-1 virions (Watts *et al.*, 2009), global themes began to emerge. Using selective 2'-hydroxyl acylation analysed by primer extension (SHAPE) (Mortimer & Weeks, 2007; Merino *et al.*, 2005) together with traditional Sanger sequencing methods read out through capillary electrophoresis, local nucleotide flexibility was assessed and used to derive a near complete secondary structure map of the single-stranded RNA (ssRNA) HIV-1 genome. In this study, both the 5' and 3' untranslated regions (UTRs) were found to be highly structured whereas the coding region contained periodic sections of highly structured and relatively unstructured RNA (Figure I.1A). This periodicity of structure in the coding region has important potential implications for ribosomal pausing and consequent effects on protein folding (see below).

A variety of NGS-based methods have been applied to eukaryotic cells to analyse the structural content of their transcriptomes. These strategies employ either combined nucleases — to cleave at ssRNA or double-stranded RNA (dsRNA) bases (Kertesz *et al.*, 2010; Zheng *et al.*, 2010; Underwood *et al.*, 2010; Li *et al.*, 2012a; 2012b; Wan *et al.*, 2012; 2014) — or chemical probes, with NGS (Lucks *et al.*, 2011; Ding *et al.*, 2013; Rouskin *et al.*, 2013; Seetin *et al.*, 2014)(Table I.1). Using parallel analysis of RNA structure (PARS), ds/ssRNA-Seq, FragSeq, and chemical probing, transcriptomes from a variety of organisms and cell types have been analysed to determine whether particular regions of the mRNAs contained characteristic structural features (Kertesz *et al.*, 2010; Zheng *et al.*, 2010; Underwood *et al.*, 2010; Lucks *et al.*, 2011; Li *et al.*, 2012a; 2012b; Wan *et al.*, 2012; Ding *et al.*, 2013; Rouskin *et al.*, 2013; Wan *et al.*, 2014).

In results from separate PARS and ds/ssRNA-Seq experiments, the structural content of the 5' and 3' UTRs relative to the coding regions was found to vary from organism to organism. The 5' and 3' UTRs were less structured than the coding regions on average for *Saccharomyces cerevisiae* and *Arabidopsis thaliana,* whereas opposite results were obtained for *Drosophila melanogaster*, *Caenorhabditis elegans*, and HIV-1 mRNAs (Figure I.1) (Watts *et al.*, 2009; Kertesz *et al.*, 2010; Li *et al.*, 2012a; 2012b). The lack of a universally conserved pattern could reflect functional differences in these regions, possibly influenced by the diversity of RNA binding proteins and microRNAs (miRNAs) that have evolved within each organism. While there is no obvious trend in the overall architecture of the UTRs, structures at the end of the 5' UTR and beginning

of the 3' UTR are well conserved. For all four eukaryotic species, there is a significant decrease in mRNA secondary structure near both the start and stop codons of the coding DNA sequence (CDS) and the implications of these observations are discussed below. Notably, the data from these NGS methods are in agreement with previous whole-genome computational predictions for both mouse and human mRNAs (Shabalina *et al.*, 2006).

Analysis of the coding regions revealed striking trends that warrant further exploration. In the PARS study using >3000 mRNAs from *S. cerevisiae*, a periodic structure signal across the coding regions with a cycle of three nucleotides was detected (Figure I.1B) (Kertesz *et al.*, 2010). This periodicity has been interpreted to mean that, on average, the first nucleotide of each codon is the least likely, and the second nucleotide is most likely, to be involved in secondary structural interactions. The amplitude of this periodicity is associated with ribosome density(Ingolia *et al.*, 2009) and suggests a functional role in translation. Also, similar to the results from the HIV-1 analysis, larger secondary structures were found to have a regulatory role in protein translation. Within the HIV-1 genome, these larger RNA structures correspond to inter-protein linkers in polyprotein precursors and unstructured peptide loops that link protein domains which could affect ribosomal pausing (Watts *et al.*, 2009).

I.2.2 Structure around start codons impacts translational efficiency

A relationship between mRNA structure around the translation start site and translation efficiency has been shown previously using a combination of experimental and computational approaches in bacteria. From a synthetic library, over 100 synonymous variants of the green fluorescent protein (GFP) gene were expressed in *Escherichia coli*. In this study, mRNA structure near the start codon (nt -4 through +37) correlated with more than half of the observed variations in protein levels within this library (Kudla *et al.*, 2009). Another study using a similar approach showed that mRNAs lacking ribosome-binding sites (Shine-Dalgarno sequences) in α- and γ-proteobacteria, cyanobacteria and plastids also exhibit a pronounced lack of mRNA secondary structure at the translation start codon. This finding, functionally validated using reporter gene constructs in *E. coli*, suggested that start codon accessibility is a major factor for even Shine–Dalgarno sequence (SD)-independent translation initiation (Scharff *et al.*, 2011). However, while the expression levels in these studies were experimentally determined, 5' mRNA folding energies were computationally predicted and hence experimental validation for these folding energies is required.

Transcriptome-wide RNA structure probing has provided some experimental support for the proposed correlation between mRNA structure and translation efficiency. In *S. cerevisiae*, a small but significant anti-correlation was observed between RNA structure in the region located 10 base pairs (bp) upstream from the translation start site and ribosome density throughout the transcript (Kertesz *et al.*, 2010). This finding was determined by ribosome profiling (Ingolia *et al.*, 2009), which is an indicator of translational efficiency. Notably, genes that exhibited less structure in their 5' UTR than in the beginning of their coding region also tended to have a higher ribosome density (Figure I.2A) (Kertesz *et al.*, 2010). This is reflected in the level of codon usage, where AT-rich codons are preferentially selected over GC-rich codons (Gu *et al.*, 2010; Bentele *et al.*, 2013; Goodman *et al.*, 2013). These results provided the first

comprehensive experimental evidence to validate the hypothesis that mRNA secondary structure around the start codon reduces translational efficiency.

An extension of the PARS method, Parallel Analysis of RNA Structures with Temperature Elevation (PARTE), has allowed for the genome-wide measurement of RNA folding energies of *S. cerevisiae*mRNAs by probing the secondary structure at temperatures ranging from 23 to 75 degrees Celsius (Wan *et al.*, 2012). The PARTE profile of *S. cerevisiae* mRNAs confirms that the two local regionswith the weakest structural pairings are at the start and stop codons. Regions ~20 nucleotides upstream of the start codon and ~10 nucleotides downstream of the stop codon have significantly more stable RNA structures (Wan *et al.*, 2012). This polarity of structural stability may be a consequence of secondary or higher-order structures of RNAs and could function in regulating translation initiation, elongation or termination.

In addition to structure around the start codon, evolutionarily conserved RNA structures and codon usage directly after the translation start site can impact protein expression (Mao *et al.*, 2013). For example, genome-wide studies of RNA structure have revealed an increase in RNA structure downstream of the start codon for genes that localize to distinct cellular domains (Kertesz *et al.*, 2010). With codon usage, rare codons are frequently found at the beginning of the gene and perhaps slowing ribosome translation to help with ribosomal allocation and prevent jamming (Tuller *et al.*, 2010). However, this may be more a result of RNA structure where rare codons in *E. coli* are frequently AT-rich at the third position (Goodman *et al.*, 2013). Overall, these results may be useful for optimizing gene expression levels when designing expression constructs for various applications and further experiments with controlled synthetic libraries could help resolve differences.

I.2.3 Structure in coding regions: ribosomal pausing and translational efficiency

Many proteins are known to fold co-translationally, and folding intermediates may be formed for almost every protein (Komar, 2009). The rate of translation, which can vary by several orders of magnitude across an mRNA transcript, can influence the protein folding pathway (Wolin & Walter, 1988; Shah *et al.*, 2013). Aside from physical obstacles to translation, such as RNA-binding proteins and pausing due to rare codon usage or low tRNA abundance, RNA structure can have a profound effect on the rate of translation (Meyer & Miklós, 2005; Dana & Tuller, 2012). Previous studies have focused on individual transcripts, highlighting how mutations can change the structure of an mRNA leading to downstream changes in protein levels (Nackley *et al.*, 2006). A recurring pattern of RNA structure located near or after the regions of the genome encoding autonomously-folding protein domains was discovered in the HIV-1 genome. This observation led to the proposal that protein structure is encoded by both the primary sequence and higher-order structure of an mRNA (Watts *et al.*, 2009). Ribosomal pausing due to highly structured RNA could facilitate the folding of individual domains within a single-protein or a polyprotein. Analysis of the ribosomal pause sites in the HIV-1 genome revealed that structured regions of the RNA were over-represented at pause sites. This co-translational pausing could be a common feature among eukaryotic mRNAs (Figure I.2B) (Wen *et al.*, 2008; Han *et al.*, 2012).

While there are variations in overall coding-region structure in the mRNAs from different organisms, in both *S. cerevisiae* and *A. thaliana* there is a correlation between

highly structured features in mRNAs and increased protein translation (Kertesz *et al.*, 2010; Li *et al.*, 2012b). A study comparing *S. cerevisiae* PARS-determined mRNA folding strength with protein abundance came to the surprising conclusion that mRNAs with highly structured features in their coding regions were more highly expressed (Zur & Tuller, 2012). To determine global protein abundance in *S. cerevisiae*, this study used the average of four previously published quantitative data sets (Ghaemmaghami *et al.*, 2003; Newman *et al.*, 2006; Lee *et al.*, 2011), and for ribosome densities on mRNAs, two genome-wide measurements were considered, one generated by ribosome-profiling (Ingolia *et al.*, 2009) and the other by microarray (Arava *et al.*, 2003). With the exception of codon bias, RNA secondary structure stability was the feature with the highest correlation to protein abundance and remained significant when controlling for mRNA levels. This study also found a strong positive correlation between RNA folding strength and ribosome density. Therefore, it appeared that highly structured RNAs might be more highly expressed — an observation that is not easily accounted for by mRNA half-lives because the correlation between mRNA folding strengths and half-lives is low. One explanation for this observation could be that RNA structure assists in protein folding, therefore enabling more productive expression levels. More experiments on a genome-wide scale with detailed biochemical follow-up are needed to fully understand the effect of increased mRNA structure, and other important factors, on protein expression levels.

I.3 RNA structures guiding RNA localization

Once thought to be a specialized mechanism limited to a small number of transcripts, it is now known that over 70% of mRNAs can be asymmetrically localized in certain cell types (Lécuyer *et al.*, 2007). RNA localization is an important aspect of gene expression, providing both temporal and spatial control that is crucial for cell migration and differentiation (Holt & Bullock, 2009). One method to control RNA localization is through *cis*-acting RNA elements known as zipcodes, which are RNA segments found primarily within the 3' UTRs of mRNAs (Martin & Ephrussi, 2009). Transported mRNAs can contain multiple zipcodes, ranging in size from a few nucleotides to over 1 kilobase, that act together to recruit a combination of proteins (Jambhekar & DeRisi, 2007). These proteins then associate with cytoskeletal motors and regulate transportation throughout the cell (Chartrand *et al.*, 2002). Thus, levels of structure across the transcriptome could reveal whether genes with shared cellular localization sites tend to have similar degrees of structure. From the PARS study in *S. cerevisiae*, increased RNA structure was found in transcripts whose proteins localize to specific cellular domains (Kertesz *et al.*, 2010), suggesting the presence of distinct structural localization elements. Surprisingly, this increased structure was prevalent within the coding regions. While less common, localization elements can be found within the coding region of mRNAs, such as for the well-characterized ASH1 mRNA from *S. cerevisiae* (Chartrand *et al.*, 2002), and this study suggests that we may need to rethink where localization elements are typically found.

Some mRNAs, particularly those encoding secreted proteins, contain sequences in the 5' end that affect RNA localization (Palazzo *et al.*, 2007). The amino-terminal hydrophobic amino acid sequence (the signal peptide) in a nascent protein guides the protein and mRNA co-translationally to the ER; however, the mRNA region that encodes the signal peptide also contains important information at the nucleotide level

for mRNA localization. The 5'UTRs and the first 30 coding nucleotides of these mRNAs are important for nuclear export (Palazzo *et al.*, 2007). From the PARS data, these same regions in mRNAs encoding secreted proteins were less structured on average than those in other mRNAs (Kertesz *et al.*, 2010).

In many cases, three-dimensional RNA structure may be critical for recognition. Several prominent examples support this hypothesis, such as the region of RNA recognized by the RNA-binding protein Egalitarian and the unusual A-form helix recognized for transport by dynein (Dienstbier *et al.*, 2009; Bullock *et al.*, 2010). Further analysis of structure, combined with new proteomic and high-resolution cell imaging data, could provide valuable insights for the future.

I.4 RNA thermometers and RNA stability

RNA structure and folding energies can influence both the translation and stability of an mRNA under stress conditions *in vivo*. During periods of stress, protein translation is sustained by mechanisms that facilitate translational initiation or elongation (Shah *et al.*, 2013). An early example of this phenomenon comes from bacteria, where temperature-sensitive structures known as "RNA thermometers" affect mRNA translation (Smit & Duin, 1990; Kortmann & Narberhaus, 2012). Under low or physiological temperatures, these RNA structures are stable and block ribosome binding to the 5'UTR. At elevated temperatures, such structures unfold to expose normally occluded ribosome binding sites, promoting the expression of genes that stabilize unfolded proteins (Figure I.3A)(Chowdhury *et al.*, 2006). While RNA thermometers have been found predominantly in bacteria, this mechanism may also exist in eukaryotes (Meyer *et al.*, 2011; Vandivier *et al.*, 2013). The folding energies of the entire *S. cerevisiae* transcriptome were determined using PARTE, revealing thermodynamically unstable structures to be enriched in ribosomal binding sites of mRNA 5'UTRs (Wan *et al.*, 2012). This suggested that RNA thermometers can function as an evolutionarily-conserved heat shock mechanism within eukaryotes; however, specific examples that affect translation have yet to be verified biochemically.

While the 5'UTR contains thermodynamically unstable structures to allow for translation initiation in general, mRNAs that were thermodynamically more stable across the entire transcript were more abundant during heat shock *in vivo* (Gasch *et al.*, 2000; Wan *et al.*, 2012). These structures could prevent degradation by the exosome complex, which is the major 3' to 5' exonuclease in eukaryotes and requires a 3' ssRNA region of approximately 30 nts. The exosome was inhibited if the 3' end of the RNA was structured or contained only a short 3' overhang (Figure I.3B)(Bonneau *et al.*, 2009). Mutational inactivation of the exosome in *S. cerevisiae* decreased the degradation of all mRNAs, and especially those with unstable RNA structures (Wan *et al.*, 2012). These results were recapitulated in an *in vitro* assay, wherein a reconstituted exosome was able to rapidly degrade an mRNA encoding a ribosomal protein, which has less thermodynamically stable structures overall. Generally, many of these housekeeping genes are repressed during heat shock to prevent the transcription, translation, and accumulation of misfolded proteins (Panniers, 1994). In contrast, anmRNA encoding a key activator of the unfolded protein response with a stable 3' UTR structure exhibited a much longer lifetime (Wan *et al.*, 2012). The exosome complex thus appeared to discriminate between structured and unstructured substrates for decay by selecting

unfolded RNA structures for degradation in a temperature-dependent fashion (Figure I.3B).

The situation differs in *A. thaliana*, where a significant negative correlation between RNA secondary structure and mRNA abundance was found by comparing RNA-seq and ds/ssRNA-seq data (Li *et al.*, 2012b). RNAs with low levels of secondary structure were on average more abundant in the transcriptome and vice versa. This finding was contrary to the observed relationship in *S. cerevisiae* mRNAs and could be explained by the increased tendency of structured RNAs to be processed into small regulatory RNAs in plants (Li *et al.*, 2012b). Since viral RNAs tend to be highly structured (Watts *et al.*, 2009), there could be selective pressure in *A. thaliana* to help differentiate between foreign and endogenous RNA molecules.

I.5 RNA structure in gene regulation by small RNAs

It is becoming increasingly clear that RNA structure contributes substantially to gene regulation by miRNAs: structure in miRNA precursor transcripts controls their processing, and structure within mRNAs influences susceptibility to miRNA targeting. In the canonical miRNA biogenesis pathway, double stranded regions of precursor RNAs are recognized by Dicer (an RNase III type enzyme) and are processed into 20-30 nt miRNAs. These small RNA duplexes are loaded into an Argonaute protein, which is part of an RNA-induced silencing complex (RISC), and the miRNA guides RISC to complementary regions in target mRNAs. Once bound by RISC, the translation of these mRNAs is reduced and the mRNA may be cleaved for complete translational repression (Wilson & Doudna, 2013). Because most regulatory small RNAs are produced from dsRNA precursors, mRNA secondary structure could, in principle, affect the total levels of small RNAs generated by the cell's processing machinery.

In *C. elegans*, *D. melanogaster* and *A. thaliana*, the highly base-paired regions within all interrogated RNA categories, including functional RNAs (e.g. rRNAs, snRNAs, snoRNAs, and tRNAs) and pre-mRNAs, were found to be preferentially processed into small RNAs (Zheng *et al.*, 2010; Li *et al.*, 2012a; 2012b). This finding suggested that transcripts containing long based-paired stem-loops, either isolated or in clusters, are recognized as small-RNA precursors (Figure I.4A). Additionally, long dsRNAs (which are common in the life cycles of various viruses) are known to activate an antiviral innate immune response in mammals, which induces cleavage of these non-self RNAs (Gantier & Williams, 2007). Thus, the need to avoid cleavage by cellular machineries is one reason why the protein-coding regions in many mRNAs maintain low levels of secondary structure.

In addition to influencing the biogenesis of these small regulatory RNAs, structure also affects the accessibility and stability of miRNA target binding sites. Until recently, the utility of incorporating thermodynamics into algorithms that predict miRNAs and their targets was limited by a lack of experimental data (Kertesz *et al.*, 2007). Now, thanks to the availability of methods to probe mRNA structure on a genome-wide scale, strategies for predicting and validating functional miRNA target sites are becoming increasingly accurate (Low *et al.*, 2012). Investigating the structure within 3'UTRs of *C. elegans* and *D. melanogaster* mRNAs on a transcriptome-wide scale was the first interrogation of the structural patterns within miRNA binding sites (Li *et al.*, 2012a). Overall, increased secondary structure in the 3'UTRs of animal mRNAs compared to *S.*

cerevisiae (Kertesz *et al.*, 2010)transcripts led to the hypothesis that such structures might serve to mask miRNA binding sites to prevent or reduce silencing of transcripts, since miRNA-mediated gene regulation is active in multicellular animals but not budding yeast (Bartel, 2009). This hypothesis was tested by determining the structure at predicted miRNA binding sites and the 50 bp of sequence immediately upstream and downstream (Ruby *et al.*, 2007; Dai *et al.*, 2011). This analysis revealed an overall decrease in the tendency to detect structured RNA across the length of miRNA binding sites in *C. elegans* target mRNAs (Figure I.4B). However, the opposite trend was observed for *D. melanogaster* miRNA target sites. Here, a significant increase in secondary structure specifically within the seed-pairing region of miRNA binding sites was observed. Together, these finding suggest that miRNA regulatory complexes encounter different mRNA structural contexts in these two animals; nonetheless, the findings for *D. melanogaster* appear to be more the exception than the rule (Kertesz *et al.*, 2007). Recent findings from additional genome-wide structure probing studies in humans provide strong experimental evidence that miRNA binding sites tend to be less structured than their flanking regions, allowing efficient targeting by small regulatory RNAs (Wan *et al.*, 2014).

Currently miRNAs and their targets are difficult to predict *de novo* due to lack of conservation in the sequence of small RNAs and competing structural elements in their targets. Degradome sequencing, which detects RNAs in the process of being degraded, has helped to facilitate the discovery of novel miRNA–target RNA pairs. Degradome sequencing takes advantage of the fact that for ligation during the library preparation step of Illumina sequencing, RNAs with a free 5'-monophosphate are required, which occurs naturally through cellular pathways of RNA decay. RNA intermediates of mRNA turnover can therefore be ligated and sequenced to find novel miRNA–target pairs by working in reverse from the cleaved transcripts (Addo-Quaye *et al.*, 2008; German *et al.*, 2008; Gregory *et al.*, 2008; Willmann *et al.*, 2013). In *A. thaliana,* where the miRNA binding sites are completely complementary to the target sites, strong signals are observed for sites of small-RNA guided cleavage (Figure I.4C)(Willmann *et al.*, 2013). A complementary method for organisms where small-RNA guided cleavage sites are more challenging to determine, is Argonaute HITS-CLIP (high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation) and its variants (Chi *et al.*, 2009; Zisoulis *et al.*, 2010; Yang *et al.*, 2011; Li *et al.*, 2014); this approach can simultaneously isolate miRNAs and mRNA target sites bound by Argonaute proteins. Linking both methods with transcriptome-wide RNA structure predictions may help accurately predict new small RNAs and their targets.

I.6 RNA structure *in vivo*

Until recently, high-throughput cellular RNA structural studies have used *in vitro* re-folded RNA, which may not reflect the true structure *in vivo* where RNAs interact with other cellular components. An important challenge is to adapt current technologies for applications *in vivo*, using reagents that can penetrate cell membranes and provide readouts with enhanced sensitivity. Various compounds have been developed to examine intracellular RNA structures (Wells *et al.*, 2000;; Spitale *et al.*, 2012; Tyrrell *et al.*, 2013; Kwok *et al.*, 2013),and recently the first global *in vivo* probing was performed using dimethyl sulfate (DMS) on the *A. thaliana*, *S. cerevisiae*, and *H. sapiens*

transcriptomes (Ding *et al.*, 2013; Rouskin *et al.*, 2013). DMS probing methylates adenine and cytosine bases in non Watson-Crick conformations in unprotected RNAs which are then detected by reverse transcription and NGS (Tijerina *et al.*, 2007). A caveat with this method, as well as other chemical probing methods, is that dsRNA and ssRNA protected by proteins may appear the same in the analysis. A complementary approach to these studies was performed using PARS under near *in vivo* conditions. This was achieved by deproteinizing natively folded RNA molecules from lymphoblastoid cells of a human parent–offspring trio (Wan *et al.*, 2014).

A primary observation from the *A. thaliana* transcriptome study (Ding *et al.*, 2013) was the presence of a three-nucleotide periodic repeat pattern within the coding region of mRNAs. This pattern was especially prominent for highly translated genes and is the first *in vivo* demonstration of triplet periodicity in a multicellular organisms. Notably, these results are consistent with *in vitro* observations from *S. cerevisiae* and what has been predicted computationally in humans (Shabalina *et al.*, 2006; Kertesz *et al.*, 2010), suggesting a universal feature of translated mRNAs. Two additional features revealed from the study on *A. thaliana* were a pattern of strong and weak RNA structures at alternative polyadenylation cleavage sites, and stable RNA structures at the 5' splice sites of unspliced pre-mRNAs, which suggested that secondary structure inhibits the first step of splicing. These results were corroborated by PARS on human transcripts, where decreased RNA structure was found at the 5' splice site (Wan *et al.*, 2014). The *A. thaliana* study also observed increased struture at the 3' splice site, supporting the influence that RNA structure has on mRNA maturation *in vivo*

A future direction for global RNA structure studies is to examine how genetic variation alters RNA structure to control gene regulation, thus providing insight into how RNA structure contributes to inter-individual differences in gene expression. An advantage of using cells from a family trio of humans is to evaluate how single nucleotide variants (SNVs) influence RNA structure(Wan *et al.*, 2014). SNVs that alter RNA structure have been termed riboSNitches (Halvorsen *et al.*, 2010). Analysing the genomic distribution of these riboSNitches provides information on how evolution suppresses SNVs that distrupt functionally important RNA structure. For example, riboSNitches are depleted in 3' UTRs, particularily around miRNA binding sites. This study was the first global evaluation of riboSNitches, and further *in vivo* studies can explore more detailed mechanisms for how riboSNitches lead to regulation of gene expression at the RNA level.

Finally, a surprising result from the pair of in vivo DMS probing studies was the mRNAs *in vivo* were more modified by DMS than expected, suggesting they were less structured or more dynamic than has been observed *in vitro* (Rouskin *et al.*, 2013; Ding *et al.*, 2013). In particular, *A. thaliana* mRNAs encoding proteins involved in stress responses tended to have more single-stranded regions when the structure was probed *in vivo* (Li *et al.*, 2012b). There are several hypotheses for why RNAs *in vivo* are less structured than predicted from *in vitro* data. First is that cellular RNAs are remodeled by energy-dependent processes such as protein translation. Second, this *in vivo* snapshot is an average of multiple RNA structures that represent difference stages in an RNA's life cycle. Specific examples of how RNA structure changes in response to protein binding and other cellular factors still need to be validated to clarify this global picture of the *in vivo* RNA structure landscape.

I.7 Tertiary RNA structure

It has been difficult to assess levels of RNA tertiary structure on a genome-wide scale. This aspect of structure determination lags behind secondary structure detection due to inherent challenges in accurately decoding long-range interactions by chemical and enzymatic probing methods. Methods such as X-ray crystallography, NMR and cryo-electron microscopy provide high-resolution information but require preparation of biochemically well-behaved samples, limiting their throughput. Small angle X-ray scattering (SAXS) can provide information about conformational states and changes in large sets of RNAs, analysed individually, but does not produce atomic-resolution structural models (Stoddard *et al.*, 2010; Fang *et al.*, 2013). Recent strategies for the detection of 3D structures of RNA have melded experimental and computational methods to allow for a much higher throughput approach to atomic-level RNA models (Kladwang *et al.*, 2011), but their output remains to be validated experimentally (Kielpinski & Vinther, 2014).

Tertiary RNA structure determination may give new insights into the growing class of long non-coding RNAs (lncRNAs) which have been largely omitted from genome-wide structural studies to date (Mercer & Mattick, 2013; Novikova *et al.*, 2013). Unlike mRNAs, lncRNAs are low in abundance and are difficult to enrich for sequencing. In the PARTE study (Wan *et al.*, 2012), non-coding RNAs such as rRNA, tRNA, snoRNA, and snRNA, had a higher average melting temperature per nucleotide than mRNAs suggesting that non-coding RNAs contain more higher-order structures (CLOTE, 2005). With improved sensitivity and modelling, lncRNAs and other poorly understand RNAs families can be explored to gain insight into function.

I.8 Conclusions and future perspectives

The rapid pace of new transcript discovery has driven efforts to detect and analyse the molecular structures of RNA more quickly than has been possible using traditional biophysical approaches. Classic RNA chemical probing methods coupled to advances in sequencing technology have enabled large-scale studies on whole transcriptomes as well as on intact viral genomes, providing comprehensive insights about RNA structure (Watts *et al.*, 2009; Ding *et al.*, 2013; Rouskin *et al.*, 2013; Wan *et al.*, 2014). Three major findings have emerged from such studies to date. The first is that unlike overcooked spaghetti, many mRNAs include regions of stable secondary and possibly tertiary structure. This is true regardless of organism, and although these regions are not necessarily conserved evolutionarily, existing evidence supports their role in regulating access to ribosomes, miRNAs and RNA degradation machinery (Addo-Quaye *et al.*, 2008; German *et al.*, 2008; Kertesz *et al.*, 2010; Dai *et al.*, 2011; Li *et al.*, 2012a; 2012b). Second, the levels of detected structure correspond to levels of precursor RNA processing. Thus, plant transcripts that are not substrates for double-stranded RNA-processing enzymes, and hence are less structured overall, tend to be more abundant in the transcriptome (Addo-Quaye *et al.*, 2008; Zheng *et al.*, 2010; Li *et al.*, 2012b). The opposite situation occurs in *S. cerevisiae*, which do not have as much double-stranded RNA processing – since they lack miRNAs – and also have a more highly structured transcriptome (Kertesz *et al.*, 2010). The third major finding is that RNAs are less structured, or at least more structurally dynamic, *in vivo* than *in vitro*

(Ding *et al.*, 2013; Rouskin *et al.*, 2013). Plant, yeast and mammalian transcripts share this property, possibly due to intracellular protein binding kinetics and/or association and rearrangements within large multi-component complexes.

Table I.1 Methods for High-Throughput RNA Structure Determination. Recent experimental approaches for RNA structure determination have melded classical nuclease and chemical probing of RNA and next-generation sequencing (NGS) technologies. To date, there have been several iterations of these methods, each with their own advantages and disadvantages, which have either combined nucleases (Kertesz *et al.*, 2010; Zheng *et al.*, 2010; Underwood *et al.*, 2010; Li *et al.*, 2012a; 2012b; Wan *et al.*, 2012; 2014) or chemical probes with NGS (Lucks *et al.*, 2011; Ding *et al.*, 2013; Rouskin *et al.*, 2013; Seetin *et al.*, 2014). All have provided the ability for more information-rich experiments involving RNA structure than has been previously possible. Listed below is a summary of the recent high-throughput methods for RNA structure probing highlighting the choice of structure probe, RNA population probed, and NGS platform. The nucleases that were used to cleave at either double stranded (ds) or single stranded RNA (ssRNA) are indicated.

Name	**Probe**	**RNA pool**	**RNA folding**	**Organism**	**NGS platform**
PARS (Kertesz *et al.*, 2010; Wan *et al.*, 2014), PARTE (Wan *et al.*, 2012)	RNase VI/SI (ds/ssRNA)	Poly(A) selection	*In vitro*	*S. cerevisiae, H. sapiens*	ABI SOLiD, Illumina HiSeq2000
ds/ssRNA-Seq (Zheng *et al.*, 2010; Li *et al.*, 2012a; 2012b)	RNase VI/I (ds/ssRNA)	rRNA depleted	*In vitro*	*A. thaliana, D. melanogaster, C. elegans*	Ilumina GAIIx, HiSeq2000
FragSeq (Underwood *et al.*, 2010)	RNase PI (ssRNA)	Nuclear RNA	*In vitro*	Mouse	ABI SOLiD3
SHAPE-Seq (Lucks *et al.*, 2011; Aviran *et al.*, 2011)	1M7	*In vitro*	*In vitro*	*B. subtilis*	Illumina GAIIx
DMS chemical probing (Ding *et al.*, 2013; Rouskin *et al.*, 2013)	DMS	Poly(A) selection	*In vivo*	*A. Thaliana, S. cerevisiae,H. sapiens*	Illumina HiSeq2000
MAP-Seq (Seetin *et al.*, 2014)	DMS,CMCT, and 1M7	*In vitro*	*In vitro*	Synthetic	Illumina MiSeq

1M7, 1-methyl-7-nitroisatoic anhydride; DMS, dimethyl sulphate; CMCT, 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate.

Table I.2 Comparison of RNA structures in translational control. The genome-wide structure studies have generated many important, but occasionally conflicting, observations about how RNAs are folded. This table summarizes six features that have emerged these studies. Structure within the 5' or 3' UTRs varies widely and structure within these regions likely represents regulatory elements, such as the pseudoknot in HIV-1 which binds to $tRNA^{Lys3}$ primer (Watts *et al.*, 2009). Generally, there is a lack of structure surrounding the start codon, which increases translation efficiency, and a corresponding lack of structure around the stop codon. Interestingly, a periodic three nucleotide signal in the coding has been observed for a few species and it may facilitate translation (Kertesz *et al.*, 2010; Ding *et al.*, 2013; Wan *et al.*, 2014). Overall, mRNAs have a complex landscape which varies from organism to organism and reflects the distinct regulation pathways that have evolved for each system.

	5'UTR more structured than coding region	Structure near start codon	Three nucleotide periodicity	Structure near stop codon	3'UTR more structured than coding region
HIV-1 (Watts *et al.*, 2009)	Yes	NA	NA	NA	Yes
S. cerevisiae(Kertesz *et al.*, 2010; Wan *et al.*, 2012)	No	No	Yes	No	No
A. thaliana(Li *et al.*, 2012b; Ding *et al.*, 2013)	No	No	Yes	No	No
D. melanogaster(Li *et al.*, 2012a)	Yes	No	NA	No	Yes
C. elegans(Li *et al.*, 2012a)	Yes	No	NA	No	Yes
H. sapiens(Shabalina *et al.*, 2006; Wan *et al.*, 2014)	Yes	No	Yes	No	Yes

"NA" = not evaluated in current genome wide studies

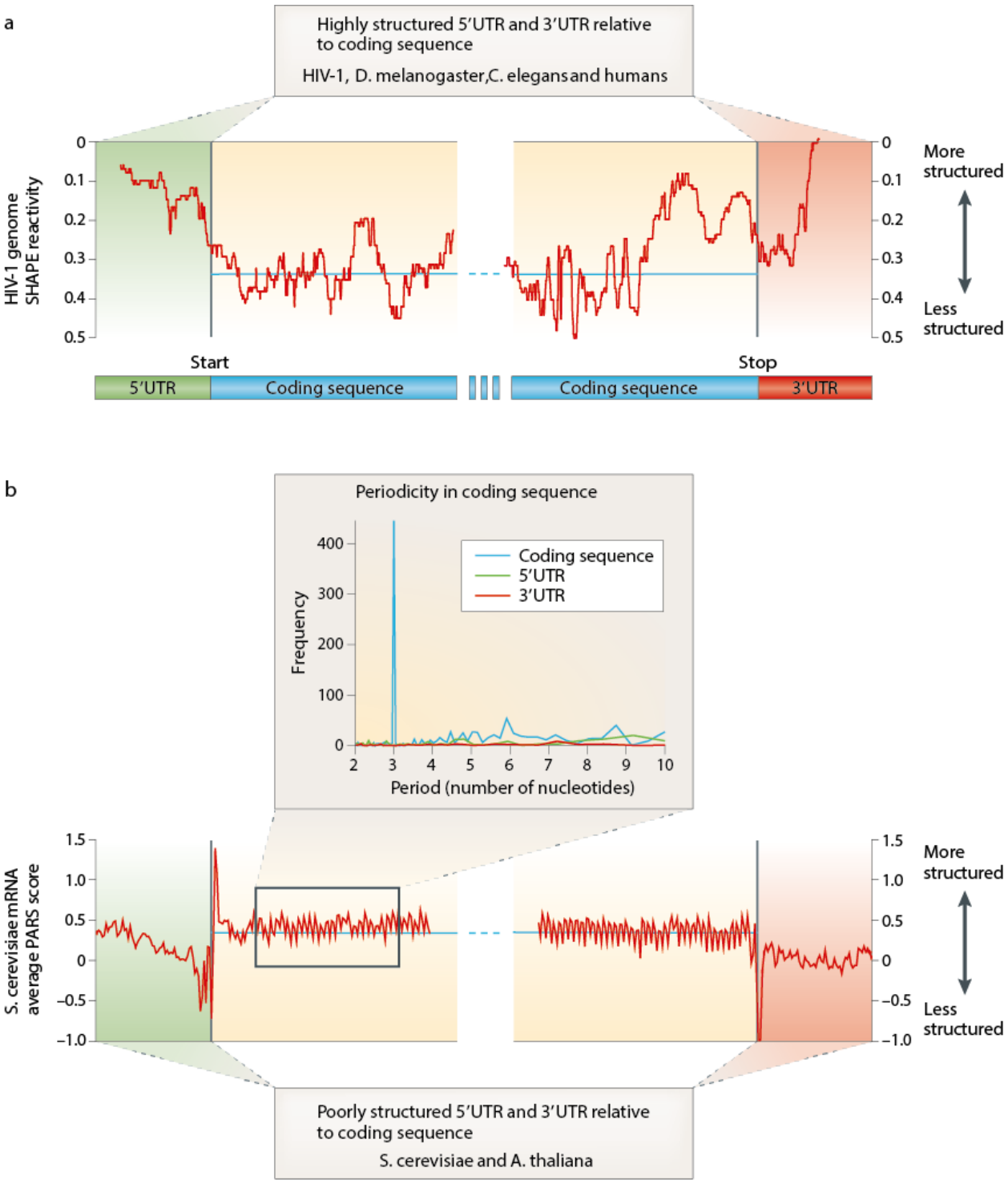

Figure I.1Structural motifs within the transcriptome. (A) RNA structure of the ssRNA HIV-1 genome (Watts *et al.*, 2009). The SHAPE reactivity of the beginning and end of the genome is indicated in red with higher SHAPE reactivity correlating with less structure. The highly structured UTRs are highlighted and similar trends are apparent for the *Drosophila* and *C. elegans* transcriptome (Watts *et al.*, 2009; Li *et al.*, 2012a). (B)

PARS structure score across the 5'UTR, the coding region and the 3'UTR average across >3000 *S. cerevisiae* mRNA transcripts (Kertesz *et al.*, 2010). Horizontal blue bars denote the average PARS score for the coding region. The poorly structured UTRs are highlighted and similar trends are apparent for the *Arabidopsis* transcriptome (Kertesz *et al.*, 2010; Li *et al.*, 2012b). The inset shows the three nucleotide periodicity that is only found in the coding regions of mRNAs (Kertesz *et al.*, 2010). Part A is adapted from (Watts *et al.*, 2009). Part B is adapted from (Kertesz *et al.*, 2010).

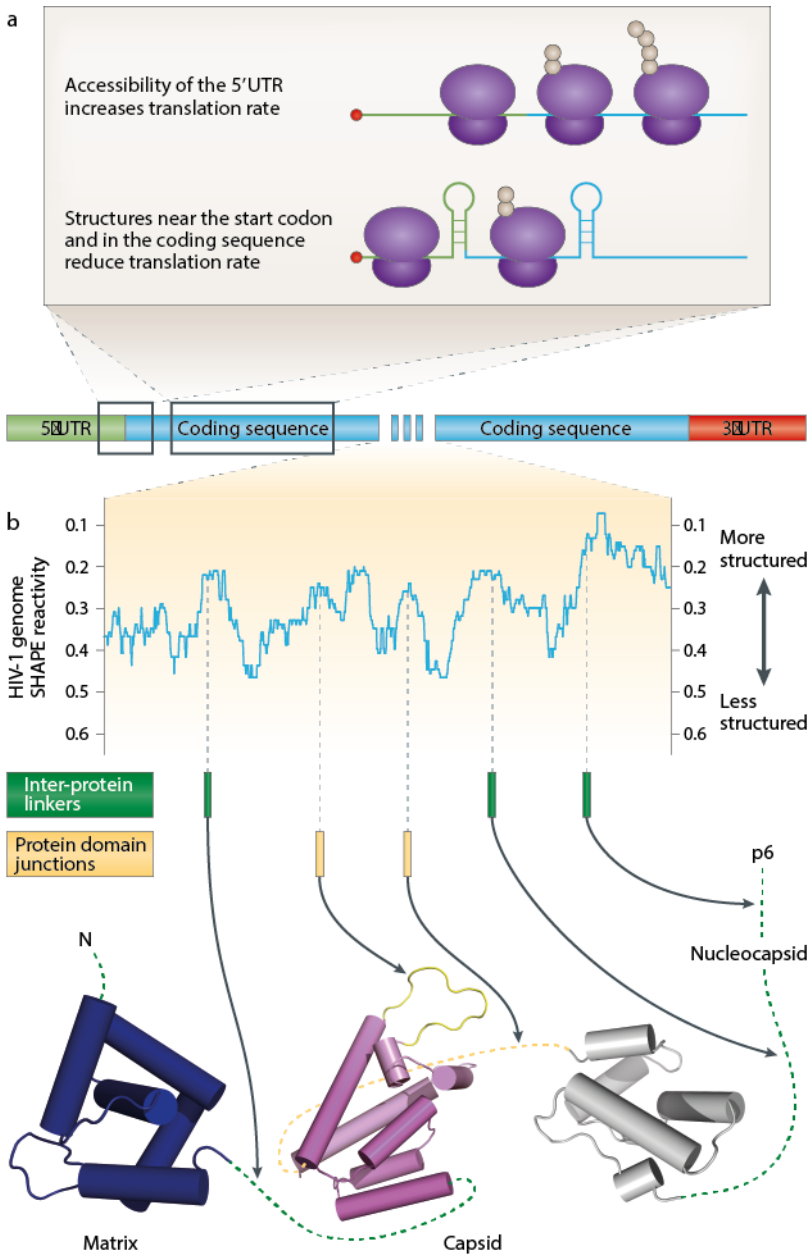

Figure I.2 Structure around start codons and translational efficiency. (A) Accessibility of 5' UTR increases translation rate. Structure around the 5'UTR directly upstream of the start codon has a large influence on initiation rates and protein abundance. Decreased structure in this region allows for efficient ribosome binding and start codon scanning. (B) Correlation between the RNA folding energy of different segments of the coding region and protein structure (Watts *et al.*, 2009). Increased RNA structure within the coding region correlates with inter-protein linkers and protein domain junctions. This structure promotes ribosomal pausing and assists in protein domain folding [PDB 2GOL, 1A43] (Kelly *et al.*, 2006; Worthylake *et al.*, 1999). Part A is adapted from (Wan *et al.*, 2011). Part B is adapted from (Watts *et al.*, 2009).

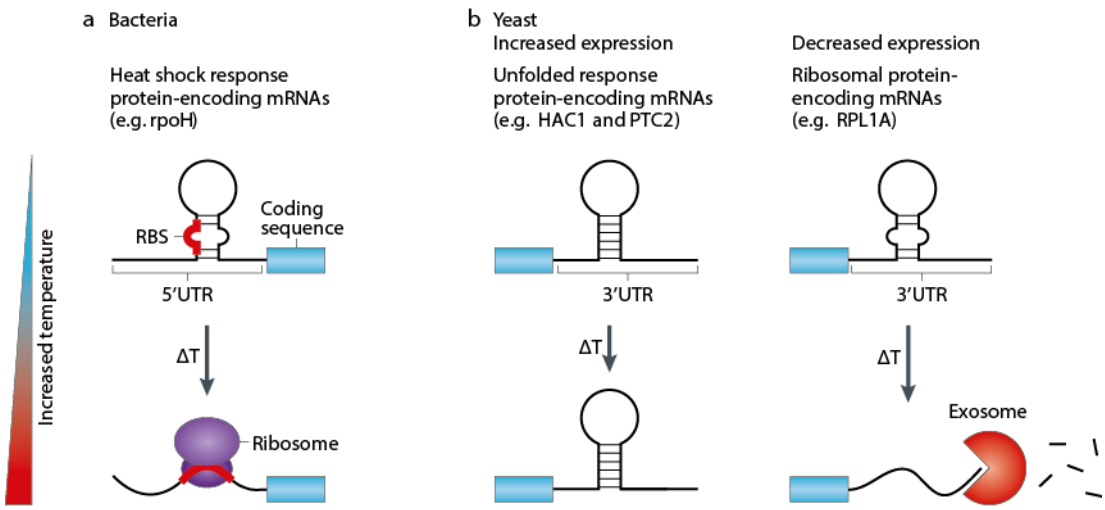

Figure I.3 RNA structure and stability under heat shock. (A) RNA thermometers are in equilibrium between folded and unfolded states. At low temperatures the RNA is predominantly in the folded state causing the ribosome binding site (RBS, shown in red) to be obscured, thus inhibiting translation. As the temperature increases, the RNA adopts a less structured conformation allowing ribosome binding and translation of heat shock response proteins. RNA thermometers are well characterized in bacteria, such as in the *rpoH* heat shock gene (Kortmann & Narberhaus, 2012). Within *S. cerevisiae*, bases that melt between 30–37°C in the 5'UTR are found most densely -10 and -20 nt upstream of the start codon, indicating that a similar mechanism might operate in eukaryotes (Wan *et al.*, 2012). B) Structured mRNAs are stabilized in *S. cerevisiae*. RNA in unstable structures becomes unfolded at higher temperatures allowing degradation by the exosome. This promotes the expression of chaperones and unfolded response proteins (ex. *HAC1*, *PTC2*) while shutting down other processes (for example, transcription and translation) to conserve energy expenditure (Wan *et al.*, 2012). *ΔT* indicates the change in temperature.

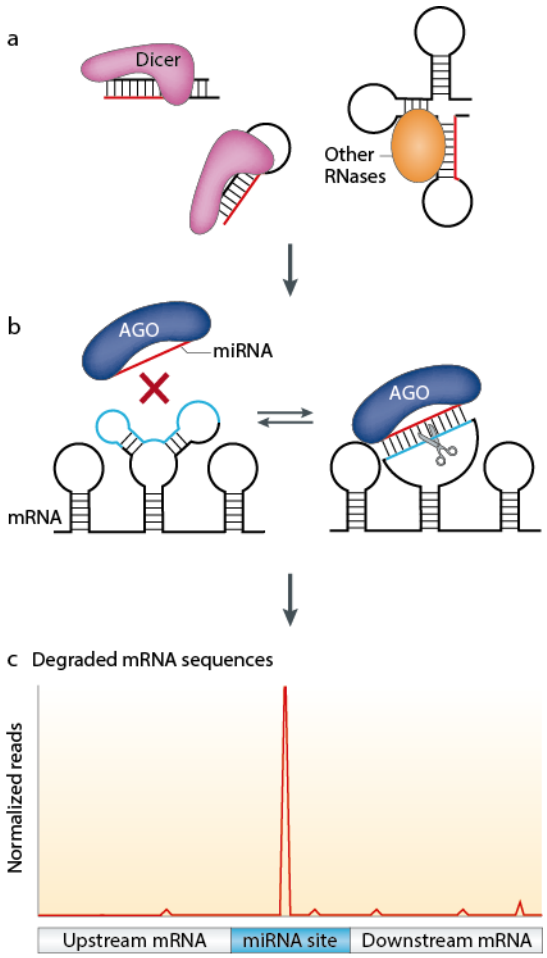

Figure I.4 mRNA structure involved in the regulation of translation by small RNAs. (A) RNA structure involved in the processing of small RNAs. Long dsRNAs or hairpin RNAs are recognized by Dicer, an RNase III type enzyme, and are processed into 20-25 nt small RNAs. Other RNases can also bind these RNA structures to generate additional small RNAs for gene regulation. (B) Schematic diagram indicating how structure affects miRNA binding to target sites. Small RNAs are loaded into Argonaute, a key component of the RNA induced silencing complex, which binds to target mRNAs to silence expression. Structure around the Argonaute binding sites in target mRNAs prevents this post-transcriptional regulation. (C) Degradome sequencing reveals novel miRNA binding sites in mRNAs. Reads mapping to the cleavage sites are enriched for miRNA binding sites (Willmann *et al.*, 2014).