이학박사학위논문

다이서의 마이크로RNA 생합성에 대한 분자생물학적 연구

Molecular insights into microRNA biogenesis by DICER

2023년 8월

서울대학교 대학원

생명과학부

이영윤

Molecular insights into microRNA biogenesis by DICER

Advisor: Professor V. Narry Kim

Submitting a doctoral thesis of philosophy
April, 2023

Graduate School of Seoul National University
School of Biological Sciences
Young-Yoon Lee

Confirming the doctoral thesis written by Young-Yoon Lee
June, 2023

Chair	Daehyun Baek	(seal)
Vice Chair	Vic Narry Kim	(seal)
Examiner	Jung Kyoon Choi	(seal)
Examiner	Soung Hun Roh	(seal)
Examiner	Jing-Hong Kim	(seal)

Abstract

Molecular insights into microRNA biogenesis by DICER

Young-Yoon Lee
School of Biological Sciences
The Graduate School
Seoul National University

RNA silencing relies on specific and efficient processing of double-stranded RNA (dsRNA) by DICER, which yields microRNAs (miRNAs) and small interfering RNAs (siRNAs). However, our current knowledge of DICER's specificity is limited to the secondary structures of its substrates: a dsRNA of approximately 22 bp with a 2-nucleotide (nt) 3′ overhang and a terminal loop. We here found evidence pointing to an additional sequence-dependent determinant beyond these structural properties. To systematically interrogate the features of precursor miRNAs (pre-miRNAs), we carried out massively parallel assays with over a million pre-miRNA variants and human DICER. Our analyses revealed a deeply conserved cis-acting element, termed the 'GYM' motif (paired G, paired pyrimidine, and mismatched C or A) near the cleavage site. The GYM motif strongly promotes processing at a specific position and can override the previously identified 'ruler'-like counting mechanisms from the 5′ and 3′ ends of pre-miRNA. Consistently, integrating the GYM motif into short hairpin RNA (shRNA) or DICER substrate siRNA (DsiRNA) potentiates RNA interference. Furthermore, we find that the C-terminal double-stranded RNA binding domain (dsRBD) of DICER recognizes the GYM motif. Mutations in the dsRBD reduce processing and alter cleavage sites in a motif-dependent fashion, which in turn affects the miRNA repertoire in cells. In particular, the cancer-associated R1855L mutation

in the dsRBD strongly impairs GYM motif recognition, offering insights into the molecular impact of this mutation.

While the mechanism of long dsRNA cleavage has been well documented, our understanding of pre-miRNA processing is limited due to the lack of the structure of human DICER (hDICER) in a catalytic state. Here we report the cryo-electron microscopy structure of hDICER bound to pre-miRNA in a dicing state, uncovering the structural basis for pre-miRNA processing. hDICER undergoes large conformational changes to achieve the cleavage-competent state. The helicase domain becomes flexible, allowing pre-miRNA binding to the catalytic valley. The dsRBD relocates and anchors pre-miRNA in a specific position through both sequence-independent and sequence-specific recognition of the newly identified GYM motif. The DICER-specific helix in the PAZ domain is also reoriented to accommodate the RNA. Furthermore, our structure unveils a previously uncharacterized configuration of the 5´ end of pre-miRNA inserted into a basic pocket. In this pocket, a group of arginines recognize the 5´ terminal base (disfavoring guanine) and terminal monophosphate group, explaining hDICER's specificity and cleavage site determination. We identify cancer-associated mutations in the 5´ pocket residues, which impair hDICER activity *in vitro* and miRNA biogenesis in cells.

Taken together, this study (1) uncovers an ancient principle of substrate recognition by metazoan DICER , (2) implicates its potential in RNA therapeutics design, (3) explains how hDICER recognizes pre-miRNAs with stringent specificity, and (4) allows a mechanistic understanding of hDICER-related diseases.

Keywords: MicroRNA; Small RNA; DICER; High-throughput sequencing; Cryo-EM; Gene silencing; RNA interference

Student ID: 2017-24011

Contents

List of Figures

List of Tables

List of Abbreviations

Aa	Aquifex aeolicus
ADAR	Adenine deaminases acting on RNA
C. elegans	Caenorhabditis elegans
Cryo-EM	Cryo-electron microscopy
DGCR8	DiGeorge syndrome critical region 8
dsRBD	Double-stranded RNA binding domain
dsRNA	Double-stranded RNA
ERH	Enhancer of rudimentary homolog
HBR	Heme-binding region
hDICER	Human DICER
KO	Knock-out
MB	Mobile basic
MID	Middle
miRNA	MicroRNA
mRNA	Messenger RNA

NMR	Nuclear Magnetic Resonance
nt	Nucleotide
P-rich	Proline-rich
PABPC	Poly(A)-binding protein
PACT	Protein kinase RNA activator
Pre-miRNA	Precursor microRNA
Pri-miRNA	Primary microRNA
R/S-rich	Arginine/Serine-rich
RISC	RNA-induced silencing complex
RMSD	Root-mean-square deviation
RNase	Ribonuclease
SAFB	Scaffold attachment factor B
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SRSF3	Serine/arginine-rich splicing factor
TNRC6	Trinucleotide repeat containing 6 protein
TRBP	TAR-RNA binding protein
UTR	Untranslated region

1. Introduction

1.1 The function of microRNA in gene regulation

MicroRNAs (miRNAs) are short non-coding RNAs that regulate the vast majority of genes, influencing essentially all developmental process and diseases. Earlier studies have discovered the first known miRNA genes—*lin-4* and *let-7*—in *Caenorhabditis elegans* (*C. elegans*), revealing their function as key developmental regulators (Lee et al., 1993; Reinhart et al., 2000). Interestingly, these genes produced small RNAs of ~22-nt in length, both of which had partial complementarity with conserved sites within the 3´ untranslated region (UTR) of messenger RNA (mRNA) targets—leading to a model in which small RNAs can mediate posttranscriptional repression of genes through sequencing matching (Lee et al., 1993; Moss et al., 1997; Olsen & Ambros, 1999; Reinhart et al., 2000; Wightman et al., 1993). It was later found that these small regulatory RNAs also exist in humans and other animals (Pasquinelli et al., 2000). High-throughput sequencing of small RNAs, together with stringent anotation criteria, identified that there are hundreds of such miRNAs (Lu et al., 2005; Ruby et al., 2006; Fromm et al., 2015). Extensive studies later found that the dominant mode of target recognition by miRNAs occurs through the 'seed sequence', which is the miRNA nucleotides 2–7 (Bartel, 2009). Imperfect base-pairing with the target sites of mRNAs in the 3´ UTR recruits trinucleotide repeat containing 6 protein (TNRC6), which interacts with the cytoplsamic poly(A)-binding protein (PABPC) to recruit two major deadenylase complexes (PAN2–PAN3 and CCR4–NOT complexes) (Jonas & Izaurralde, 2015). Shortening of the poly(A) tail by the deadenylase complexes leads to mRNA destabilization through decapping and 5´-to-3´ exonucleolytic decay (Chen & Shyu, 2011). In this manner, miRNAs shape the transcriptome

of the host, adding an important layer of posttranscriptional regulation that complements gene regulation at transcriptional level.

1.2 Metazoan microRNA biogenesis

While the mature miRNAs are very small (~22-nt), canonical miRNAs are first transcribed by RNA polymerase II as long primary miRNAs (pri-miRNAs). They are usually encoded in the intronic or exonic regions of host genes and form a long hairpin structure with a dsRNA stem of ~35 bp (Figure 1.1). This secondary structure is preferentially recognized by DROSHA, a class II ribonuclease (RNase), that makes staggered cuts by measuring a set distance from the basal or apical junction of pri-miRNAs (Figure 1.1). DROSHA forms a stable complex with its partner protein, DiGeorge syndrome critical region 8 (DGCR8), to recognize the key structural features of pri-miRNAs. While the N-terminal proline-rich (P-rich) and arginine/serine-rich (R/S-rich) domains are largely dispensable for pri-miRNA processing *in vitro*, the central domain and double-stranded RNA binding domain (dsRBD) are responsible for recognizing the basal junction and the dsRNA stem of pri-miRNA, respectively. More importantly, the RNase III domains form an intramolecular dimer, the architecture of which allows the precise processing of the dsRNA, leaving 2-nt 3´ overhang. Besides these structural features of pri-miRNAs, there are several position-specific sequence features (Figure 1.2)—apical UGU, mismatched GHG, basal UG, and flanking CNNC motif—that are recognized by DROSHA-DGCR8 complex facilitating efficient and accurate processing of pri-miRNAs (Fang & Bartel, 2015; Auyeung et al., 2013). Pri-miRNA processing by DROSHA-DGCR8 complex releases a much shorter hairpin, called precursor miRNA (pre-miRNA), which is exported from the nucleus to the cytoplasm through the action of Exportin 5.

In the cytoplasm, DICER further processes the pre-miRNAs to liberate miRNA duplexes. (Figure 1.3). DICER utilizes its two specialized domains—basic 5´ pocket (in the platform domain) and 3´ pockets (in the PAZ domain)—to recognize the 5´ and 3´ ends of pre-miRNAs (Figure 1.3) (Park et al., 2011).

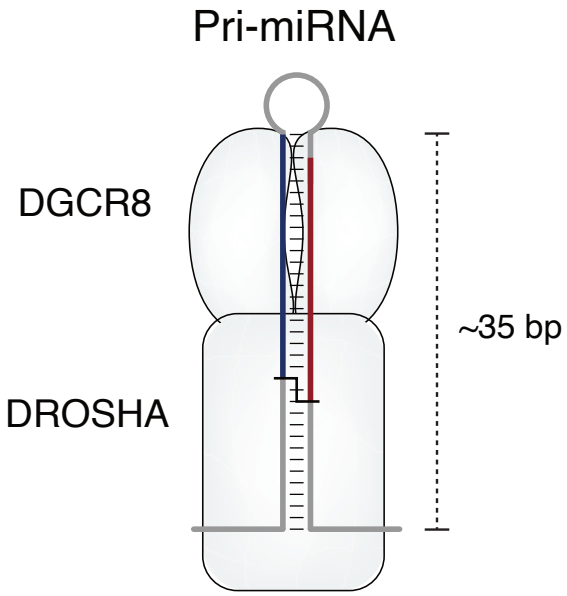

Figure 1.1 Pri-miRNA processing by DROSHA-DGCR8 complex.

DICER then measures ~22-nt from either 5′ or 3′ terminus to lop off the apical loop of pre-miRNAs. These mechanisms are called 5′ counting or 3′ counting rules (Park et al., 2011). The relative contribution of each of the end counting rules depends on the thermodynamic stability of pre-miRNA terminus (Park et al., 2011). That is, pre-miRNAs with unstable terminal structures (mismatch, G–U, or A–U pair) tend to be processed by the 5′ counting rule, whereas those with stable terminal structures (G–C pair) follow the 3′ counting rule. DICER also possesses a helicase domain, which has an autoinhibitory role in pre-miRNA processing (Ma et al., 2008). While non-essential for pre-miRNA processing, DICER also have partner proteins, such as TAR-RNA binding protein (TRBP), protein kinase RNA activator (PACT), and adenine deaminases acting on RNA (ADAR) proteins, that modulate the DICER's activity on certain pre-miRNA substrates (Kim et al., 2014).

After pre-miRNA processing by DICER, miRNA duplexes are released and associated with Argonaute (AGO) proteins (Khvorova et al., 2003; Schwarz et al.,

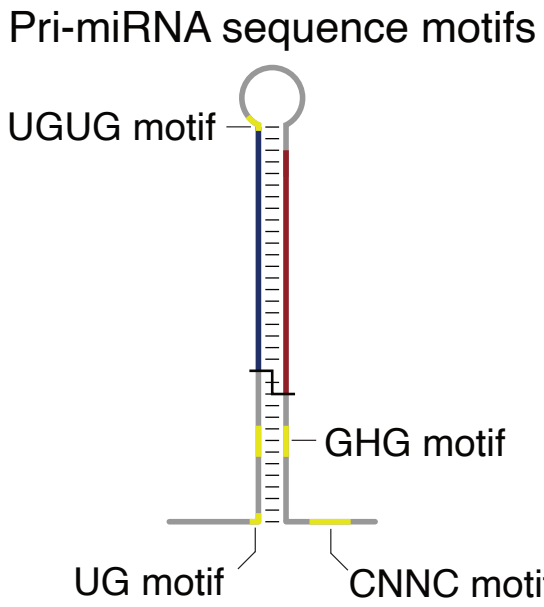

Figure 1.2 Sequence motifs of pri-miRNAs that facilitate DROSHA processing (Auyeung et al., 2013; Fang & Bartel, 2015).

2003). AGO proteins adopt a bilobal architecture composed of multiple protein domains (Schirle & MacRae, 2012; Schirle et al., 2014). Through the action of ATPase activity of molecular chaperones, miRNA duplex is loaded into AGO proteins, where only one strand (i.e., the guide strand) remains in the AGO proteins (Iwasaki et al., 2010). The other strand (i.e., passenger strand) is released and degraded (Kawamata & Tomari, 2010). AGO proteins possess a basic pocket in the middle (MID) domain that binds to the 5´ monophosphate of the guide strand (Schirle & MacRae, 2012; Schirle et al., 2014). Importantly, miRNA strands with uridine or adenine residue at the 5´ terminus are favorably selected by the AGO proteins (Frank et al., 2010). In addition, miRNA strands with thermodynamically unstable terminus (mismatch, G–U, or A–U pair) are more likely to be selected by the AGO strands and become the guide strands (Khvorova et al., 2003; Schwarz et al., 2003). AGO proteins with the guide strands then form RNA-induced silencing complex (RISC) to post-transcriptionally repress target mRNAs.

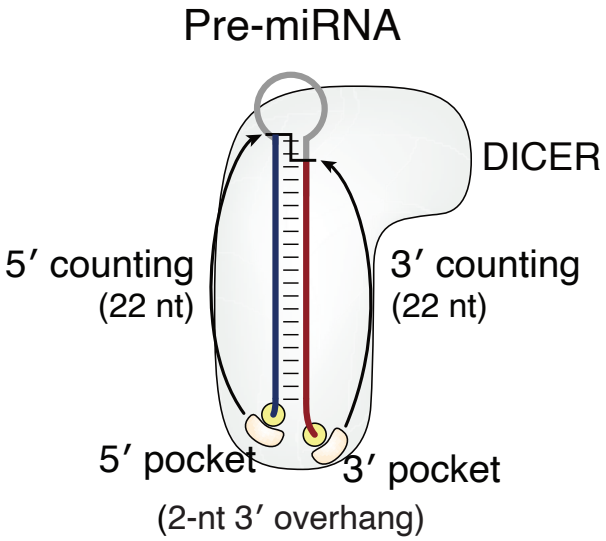

Figure 1.3 Pre-miRNA processing by DICER. DICER recognizes the 2-nt 3´ overhang to follow 5´ and 3´ counting rules (Park et al., 2011).

1.3 Structural insights into microRNA biogenesis

Structural biology is at the heart of modern biology and enables us to understand how proteins function at atomic resolution. In the past few decades, numerous efforts—using Nuclear Magnetic Resonance (NMR) spectroscopy, X-ray crystallography, and cryo-electron microscopy (cryo-EM)—were devoted to determining the structure of microRNA biogenesis factors. Here I describe the structures of DROSHA and DICER (Kwon et al., 2016; Partin et al., 2020; Jin et al., 2020; Liu et al., 2018) that have considerably advanced the understanding of miRNA maturation processes.

1.3.1 Primary microRNA processing by DROSHA

In 2016, the crystal structure of DROSHA in complex with the C-terminal DROSHA-interacting fragment of DGCR8 was solved at atomic resolution (Figure 1.4) (Kwon et al., 2016). In this structure, DROSHA forms an elongated shape with two RIIID domains intramolecularly dimerized on top of the central domain involving the platform and 'PAZ-like' domain (Figure 1.4). A long α-helix connects the RIIID dimer to the central region, hence referred to as the connector helix. Located in the C-terminal region is the dsRBD, connected to the RIIIDb via a hydrophilic linker, that may allow dsRBD to dissociate from RIIIDb and interact with dsRNA upon substrate binding. The PAZ-like domain and long loops of the central domain may form dynamic conformations at least without a pri-miRNA substrate, preventing us from building a reliable model. The architecture of the RIIIDa and RIIIDb dimer superimposes well onto the homodimer of *Aquifex aeolicus* (Aa) RNase III, which is the bacterial prototype of RNase III enzymes containing a single RIIID domain and a single dsRBD (Gan et al., 2006). The main chain root-mean-square deviation (RMSD) of important catalytic residues of DROSHA and Aa RNase III is very small (0.44 Å), suggesting that the overall catalytic mechanism is highly conserved. On the other hand, the RIIIDa of DROSHA has an extended insertion, predicted to form α-helices, that is highly conserved among DROSHA orthologs. These helices, referred to as Bump helix and mobile basic (MB) helix, may have roles in substrate binding. Interestingly,

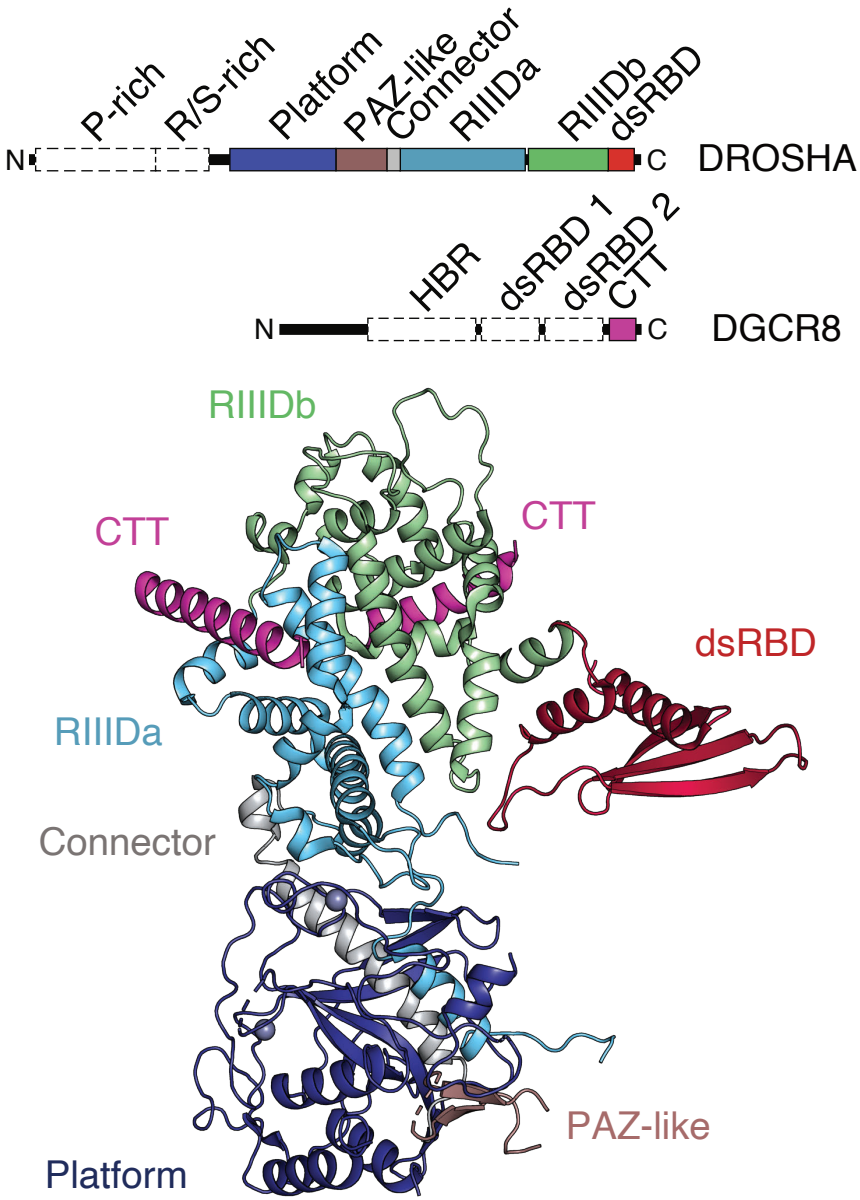

Figure 1.4 Structure of human DROSHA (Kwon et al., 2016). Domain compositions of human DROSHA and DGCR8 are shown. Colored boxes indicate regions resolved in the structure.

the structure shows that each of RIIIDa and RIIIDb domain also interacts with the C-terminal fragment (729-750 aa) of DGCR8, whose co-expression with DROSHA was essential to prevent heavy aggregation of DROSHA. This is in line with the previous observation that DROSHA interacts with DGCR8 molecules in a 1:2 stoichiometry (Nguyen et al., 2015). While the structure of DROSHA answers important questions regarding its evolutionary origin as well as the molecular basis of DROSHA-DGCR8 assembly, the lack of structure in complex with its RNA substrate leaves many questions unanswered as to how DROSHA interacts with pri-miRNA.

Two cryo-EM structures of DROSHA-DGCR8 in complex with pri-miRNA were reported in 2020 (Partin et al., 2020; Jin et al., 2020), revealing detailed interactions between DROSHA and the pri-miRNA (Figure 1.5). While most globular domains of RNA-bound DROSHA align well with the previous crystal structure of apo-DROSHA, DROSHA undergoes considerable conformational changes in some regions. For example, dsRBD shows a large conformational change towards the lower stem of pri-miRNA, consistent with the previous observation that DROSHA dsRBD recognizes the key sequence motif of pri-miRNA in the lower stem—the GHG motif—to facilitate efficient and accurate processing (Figure 1.2) (Fang & Bartel, 2015; Kwon et al., 2019). In addition, previously unresolved regions of the central domain are visible in the structure, forming intimate contact with pri-miRNA. These regions were referred to as 'Belt' and 'Wedge', which are positioned near the basal junction of pri-miRNA (Partin et al., 2020; Jin et al., 2020), revealing a mechanism by which DROSHA uses the basal junction to determine cleavage site (Nguyen et al., 2015). The structure also gives clue to how the basal sequence motif, the UG motif, facilitates efficient pri-miRNA processing (Fang & Bartel, 2015); the uridine of the UG dinucleotide flips out and is stabilized by interaction with the MB helix previosuly described in the crystal structure of DROSHA (Jin et al., 2020). The guanine base of the motif, on the other hand, forms a Watson-Crick base-pair with the cytosine of the 3p strand, marking the end of the stem (Jin et al., 2020). The structure also reveals that dsRBDs of DGCR8 cover the upper stem region of pri-miRNA, which along with DROSHA

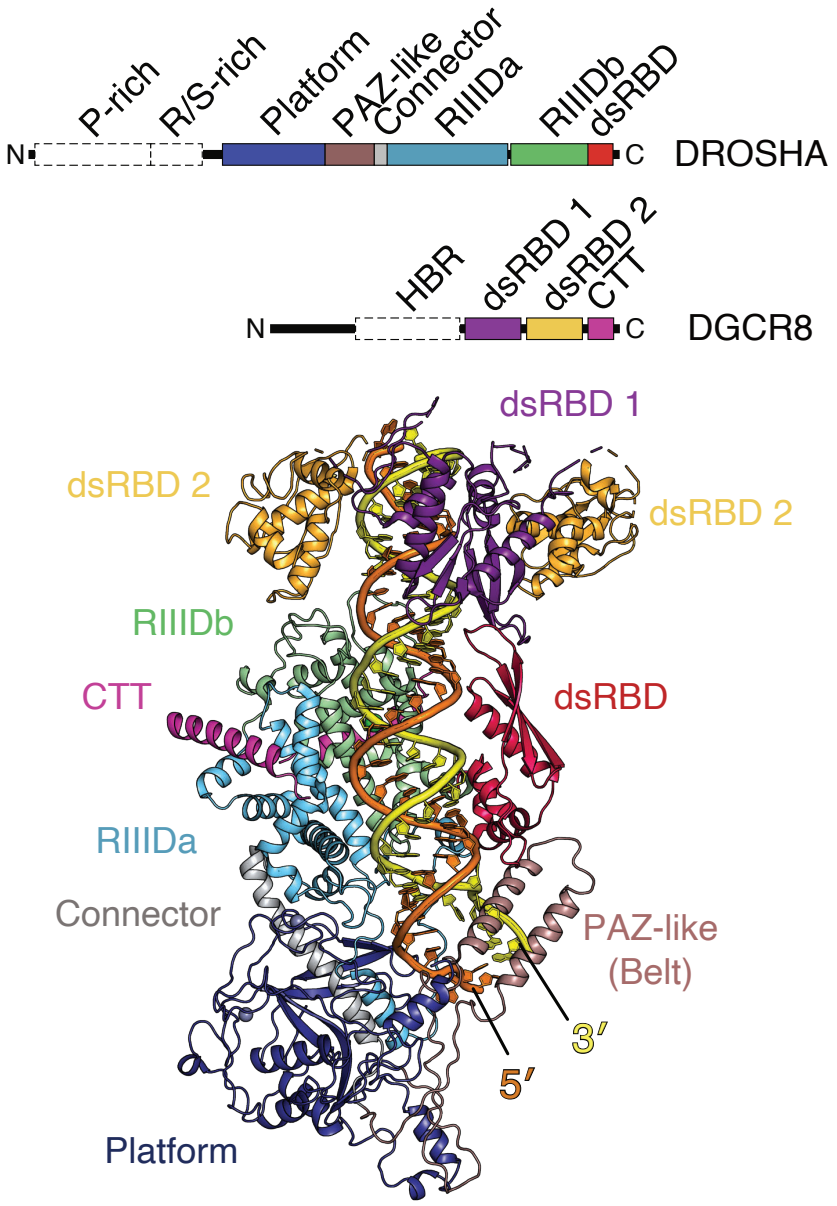

Figure 1.5 Structure of human DROSHA and DGCR8 in complex with pri-miRNA (Partin et al., 2020). Domain compositions of human DROSHA and DGCR8 are shown. Colored boxes indicate regions resolved in the structure.

measure the stem length between the basal and apical junctions. However, the majority of DGCR8 including the heme-binding region (HBR) and the apical loop of pri-miRNA could not be included into the model likely due to the structural flexibility. As a result, it remains unclear how DGCR8 recognizes the apical loop and the sequence motif embedded in it (such as the UGUG motif) (Auyeung et al., 2013). Future studies are therefore required to uncover the role of DGCR8 in pri-miRNA recognition by stabilizing the DGCR8-RNA interaction. Using an optimal pri-miRNA substrate would help determine the structure, as envisioned by the authors. It also remains to be investigated the role of other DROSHA and DGCR8 partners—such as serine/arginine-rich splicing factor (SRSF3), enhancer of rudimentary homolog (ERH), and scaffold attachment factor B (SAFB)—in pri-miRNA processing (Fang & Bartel, 2020; Shang et al., 2020; Kwon et al., 2020).

1.3.2 Precursor microRNA processing by DICER

Earlier attempts to determine the structure of human DICER at atomic resolution were confounded by the fact that metazoan Dicers are recalcitrant to overexpression and purification (Kidwell et al., 2014; Ma et al., 2012). In 2018, Liu and his colleagues reported the first full-length structure of human DICER-TRBP complex (Figure 1.6) (Liu et al., 2018). In this structure, DICER forms an elongated shape that resembles the letter 'L' with mostly globular domains, as previously observed with negative staining (Lau et al., 2009; Wang et al., 2009; Lau et al., 2012). At the bottom of the L-shaped structure is the helicase domain, which interacts with the C-terminal dsRBD of TRBP (dsRBD3). Note that the other two dsRBDs of TRBP are invisible in the structure due to structural flexibility. The RIIIDa and RIIIDb intramolecularly dimerize on top of the helicase domain, creating a processing center for pre-miRNA processing. A long connector helix connects the RIIIDa/b to the platform and PAZ domains, which are important for recognition of the dsRNA terminus. Like DROSHA, DICER possesses dsRBD that is connected to the RIIIDb domain via a short flexible linker—spatial arrangement that is very similar to that of DROSHA. Importantly, the structure of human DICER-

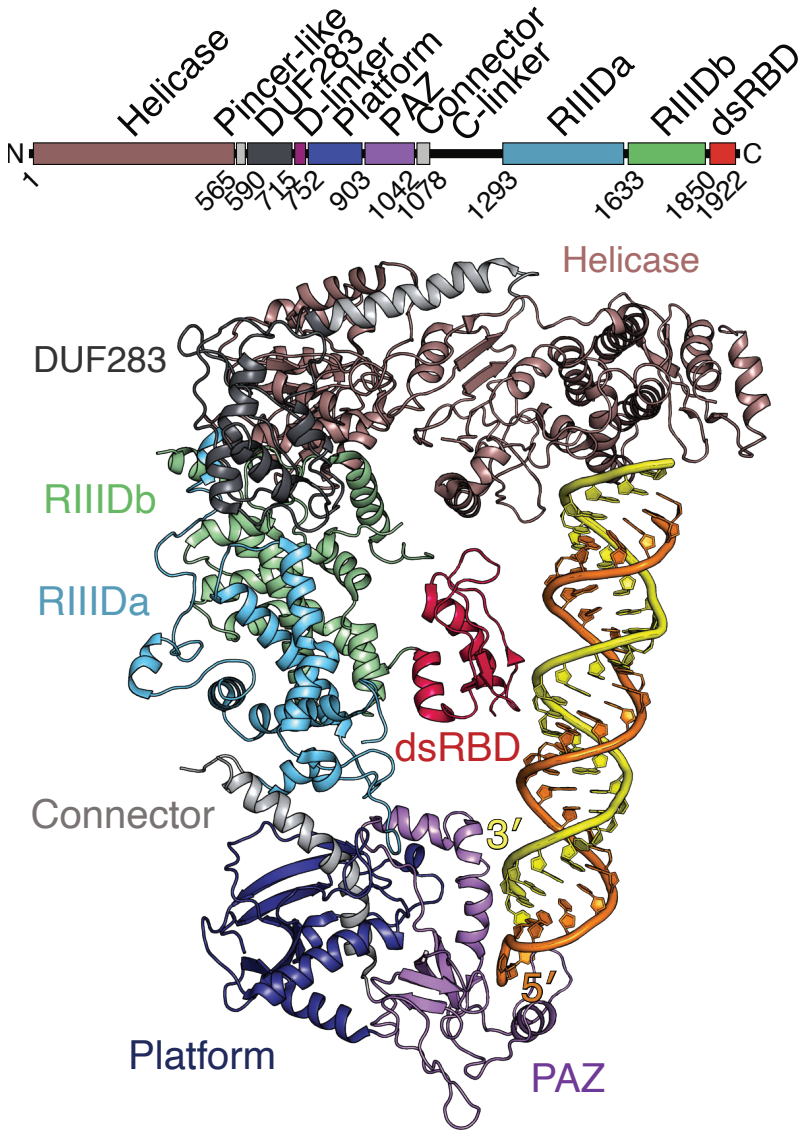

Figure 1.6 Structure of human DICER in complex with pre-miRNA (Liu et al., 2018). Domain composition of human DICER is shown.

TRBP complex with pre-let-7a-1 variant was also solved, providing insights into how DICER interacts with a pre-miRNA (Liu et al., 2018). In the RNA-bound structure, however, the pre-miRNA stem is distantly located from the catalytic center of DICER, suggesting that the structure rather represents a 'pre-dicing' state. The pre-dicing state structure shows the apical loop of pre-miRNA interacts with the helicase domain. The 3´ end of pre-miRNA is inserted into the basic 3´ pocket created by the PAZ domain. However, the 5´ end is exposed to the solvent without any prominent interactions. Overall, the structure shows that protein-RNA interactions are minimal in the pre-dicing state, leaving many questions unanswered. The structure of human DICER in dicing and other functional states are awaiting to be eludicated to advance our understanding of how DICER processes pre-miRNAs.

2. Sequence determinant of small RNA production by DICER

2.1 Background

DICER, a multidomain RNase III, serves as a key player in RNA silencing by cleaving dsRNA into small RNAs of 21–25 nt in length (Elbashir et al., 2001; Hutvagner et al., 2001). Endogenous siRNAs and miRNAs are produced from long RNA duplexes and RNA hairpins, respectively. The miRNA pathway in metazoa involves another RNase III, DROSHA, that cleaves primary miRNA (pri-miRNA) transcript to release pre-miRNA, a hairpin of ~70 nt with a characteristic 2-nt 3´ overhang (Lee et al., 2002, 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2004). DICER cleaves pre-miRNA to produce a ~22-nt duplex with 2-nt 3´ overhang at both ends (Hutvagner et al., 2001). After loading onto the AGO protein, one strand of the duplex remains as a mature miRNA that functions as a guide to base-pair with the cognate target (Khvorova et al., 2003; Schwarz et al., 2003). Targeting specificity relies on the precision of processing because even a small change in the processing site can alter the 'seed' sequence (2–7 nt region relative to the 5´ end of the guide RNA) critical for target binding (Liu et al., 2004; Lewis et al., 2003; Bartel, 2009).

DICER is known to recognize its substrates by relying solely on their secondary structural features, such as the 2-nt 3´ overhang, a dsRNA stem of ~22-bp, and a terminal loop (Macrae et al., 2006a,b; Park et al., 2011; Zhang et al., 2004, 2002; Gu et al., 2012; Tsutsumi et al., 2011; Feng et al., 2012; Liu et al., 2015). According to the current model, DICER acts as a 'molecular ruler' that measures 22-nt from the ends of pre-miRNA (Macrae et al., 2006a,b; Park et al., 2011; Zhang et al., 2004). The 3´ end is recognized by a conserved '3´ pocket' in the

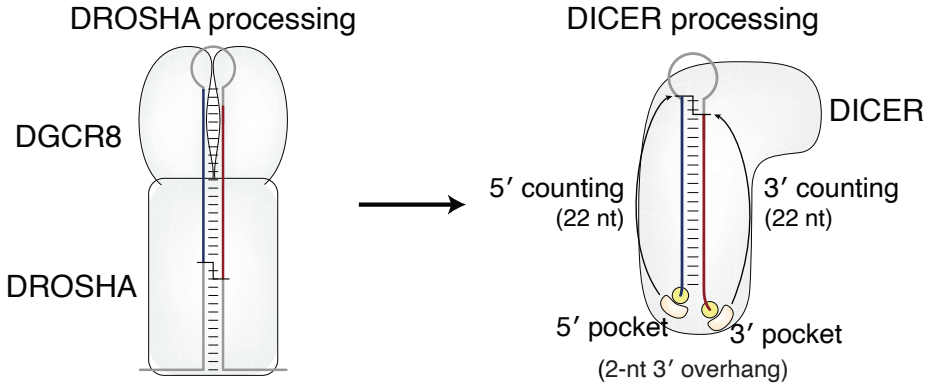

Figure 2.1 Illustration of the mechanism of cleavage site choice by DICER.

PAZ domain of DICER. DICER homologs in the metazoan miRNA pathway also have a '5´ pocket' in the platform domain to capture the 5´ end (Park et al., 2011; Tian et al., 2014; Liu et al., 2018). These pockets anchor the ends most effectively when the ends are in a 2-nt 3´ overhang arrangement (Macrae et al., 2006a,b; Park et al., 2011; Heo et al., 2012). Because the catalytic center of DICER is separated by a fixed distance from these pockets, DICER can measure a specified length (22 nt in the case of human DICER) from the 5´ end ('5´ counting rule') and the 3´ end ('3´ counting rule') (Macrae et al., 2006a,b; Park et al., 2011; Zhang et al., 2004). The relative contribution of the 5´ and 3´ counting is influenced by thermodynamic stability because an unstable 5´ end can be readily frayed and inserted into the 5´ pocket, facilitating the 5´ counting mechanism (Park et al., 2011).

Because the termini of pre-miRNAs are created by DROSHA, DICER is largely considered to play a passive role when it comes to the determination of miRNA sequences (Figure 2.1). For instance, pre-let-7a-1 follows the 5´ counting rule and is cleaved essentially at a single site that is measured from the 5´ end created at the DROSHA processing step (Figure 2.2) (Park et al., 2011). However, some pre-miRNAs do undergo alternative processing at the DICER level (Kim et al., 2019; Chiang et al., 2010). Most notably, pre-miR-324 is uridylated

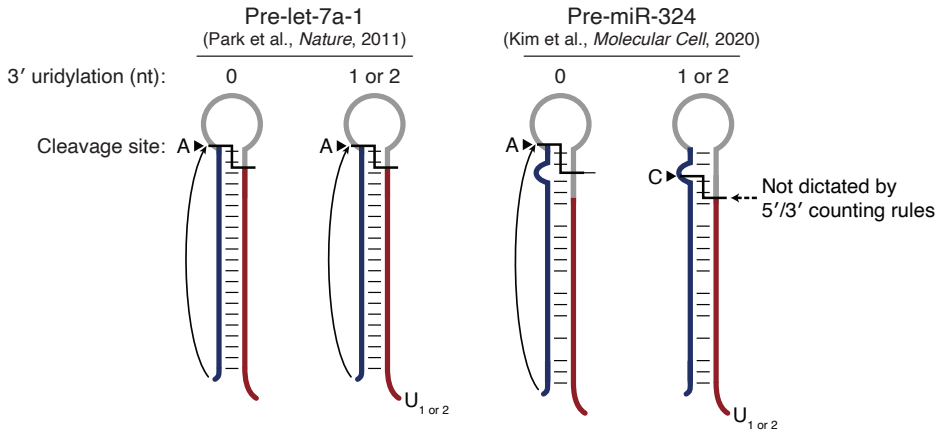

Figure 2.2 Cleavage site decision of pre-let-7a-1 and pre-miR-324.

frequently at the 3´ end, which results in a shift of the DICER cleavage site (Kim et al., 2020).

Curiously, we noticed that the end counting rules cannot fully explain the processing pattern of pre-miR-324 nor its variants (Figure 2.3), suggesting that there may be a yet-unknown mechanism by which DICER engages in processing. For example, when we used a pre-miR-324-derived substrate with a symmetric stem and a 3-nt 3´ overhang, it was cleaved at three sites (Figure 2.3, lane 3). The 5´ and 3´ counting rules explained products A and B, respectively, but not product C. Intriguingly, another variant with a mutation near the cleavage site failed to generate product C (Figure 2.3, lane 4), suggesting that there may be a critical element near the cleavage site. As previous studies focused only on secondary structures outside of this region and have not investigated the substrate specificity in a comprehensive manner, it remains unknown if and how DICER recognizes its substrates in a sequence-specific manner.

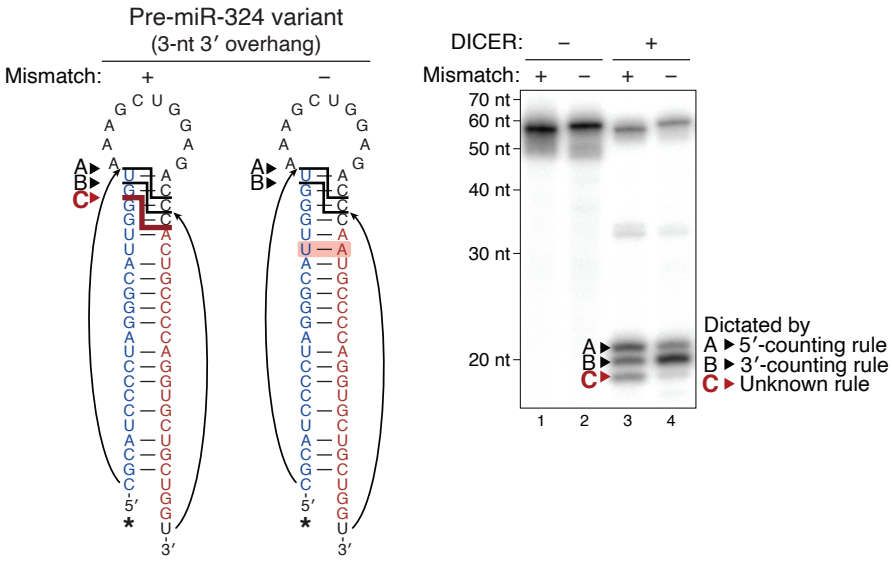

Figure 2.3 *In vitro* processing of a pre-miR-324 variant by DICER. "No-bulge pre-miR-324" was used for this assay to avoid the influence of the bulge. A mismatch near the cleavage was replaced with a base-pair marked in pink. Cleavage sites and their corresponding products are marked with arrowheads.

2.2 Result

2.2.1 Massively parallel assay to identify a sequence motif

To comprehensively interrogate the upper stem region, we implemented a massively parallel assay which enables quantitative testing of a large number of variants (Figure 2.4). In brief, we synthesized 1,048,576 pre-miRNA variants by randomizing the nucleotide identities within a 5-bp window in the upper stem region (−1 through +3 relative to the cleavage site in the 3p strand) which is predicted to contact DICER in our structural model (Figure 2.5–2.7). Pre-let-7a-1 was used as a parental backbone because it is cleaved relatively homogeneously (Park et al., 2011). After a brief incubation with purified human DICER protein (Figure 2.8), in which 5% of the substrate pool was cleaved, the uncleaved RNAs were gel-purified and sequenced by the AQ-seq method that allows efficient ligation of structured RNAs (Kim et al., 2019).[1] The processing efficiency was quantified by dividing the fraction of each variant in the input population by that of the uncleaved reads after reaction. We refer to this metric as 'cleavage score'.[2]

To identify the structural and sequence elements required for processing, we examined the pre-miRNA variants scoring within the top 0.1%. These top variants showed an overall tendency for base-pairing as expected (Figure 2.5, left panel). Intriguingly, however, a mismatch was enriched at position 1. In addition to this structural feature, we found some sequence preferences (Figure 2.5, middle and right panels). At positions −1 and 0, the 5′-C–G-3′ pair and 5′-G–C-3′ pair were

[1] AQ-seq library was generated by Dr. Haedong Kim.

[2] AQ-seq analysis was done by Dr. Haedong Kim.

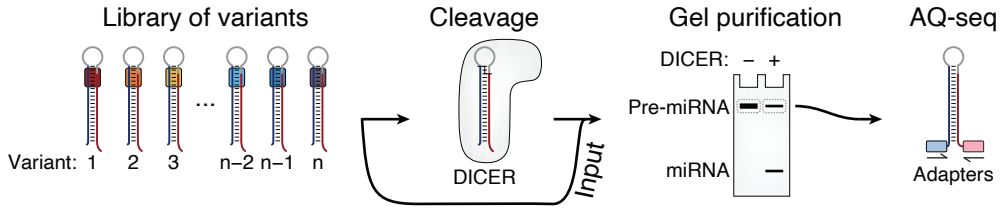

Figure 2.4 Schematic outline of the massively parallel assay.

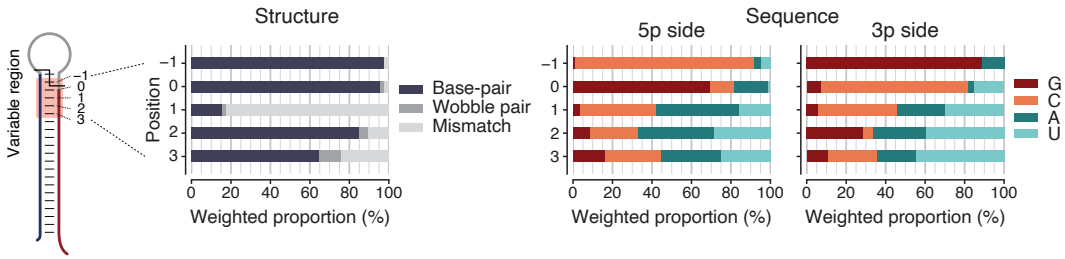

Figure 2.5 Massively parallel assay results from the 1st screening. Structural (left) and sequence (middle and right) preference of DICER substrates at the indicated positions are shown. Proportions of top 0.1% variants were weighted by their cleavage scores. Variants with over 100 read counts in the input were included in this analysis.

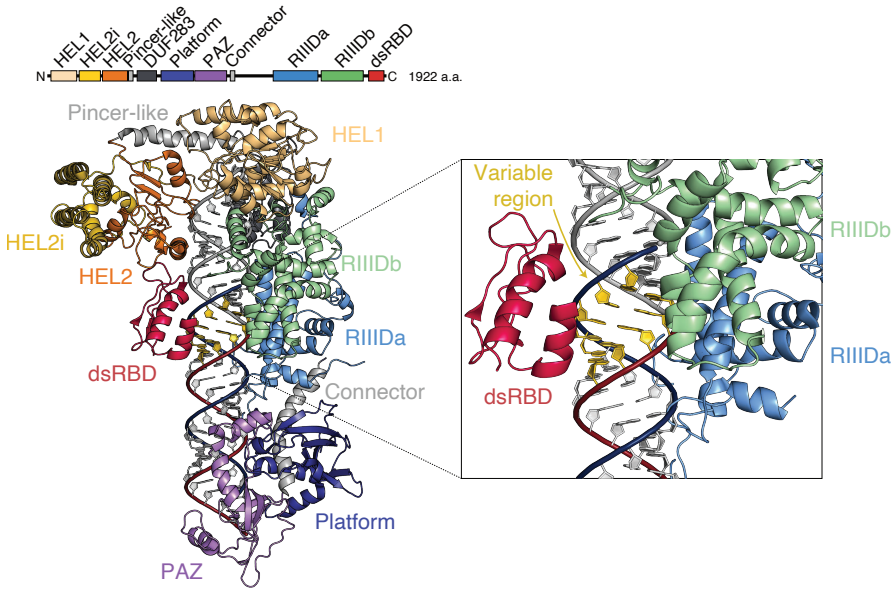

Figure 2.6 A structural model of human DICER in a dicing state. The dsRNA was modeled into the cryo-EM structure of human DICER, based on the the crystal structure of dsRNA-bound *Aquifex aeolicus* (Aa) RNase III. DICER dsRBD was then superimposed with that of Aa RNase III to predict its position in a dicing state.

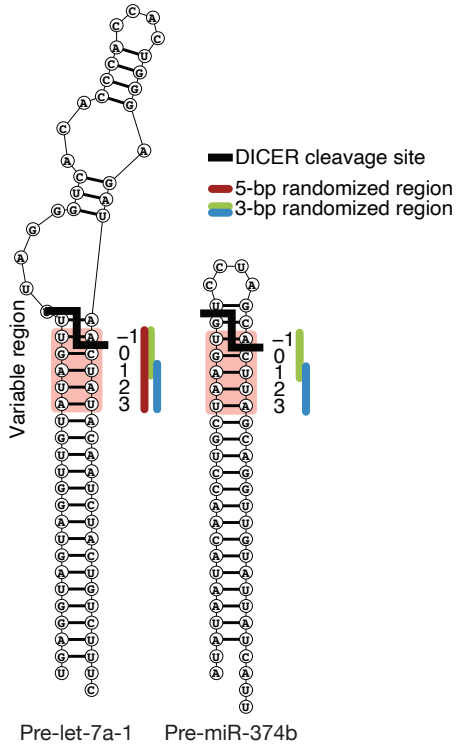

Figure 2.7 Pre-miRNAs used in the massively parallel assay. The 5-bp and 3-bp windows (–1-to-3, –1-to-1, 1-to-3 positions relative to the starting position of 3p miRNA) were targeted for randomization based on the structural model. Secondary structures of pre-miRNAs were obtained using (Bellaousov et al., 2013).

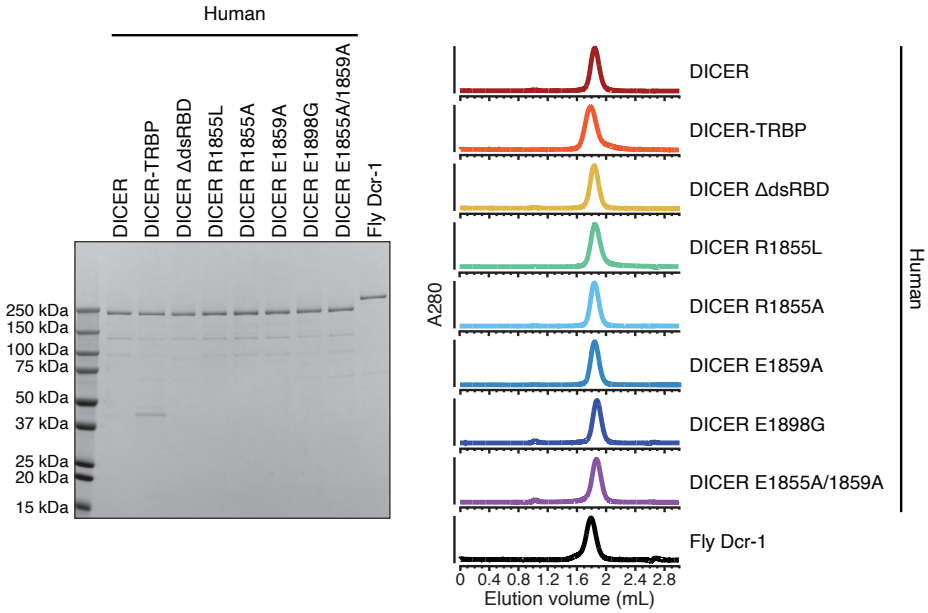

Figure 2.8 SDS-PAGE and Size-exclusion chromatography of purified proteins.

strongly favored, respectively. At position 1, G is depleted while C is enriched on both strands. Sequence preference was less prominent at positions 2 and 3.

These observations prompted further evaluation of the base-pairing and base identity at each position. To increase the sequencing depth, we performed the second round of massively parallel assays, with two 3-bp windows (randomizing positions −1-to-1 and 1-to-3) to generate 4,096 variants per each window (Figure 2.7, 2.9). In addition, we employed another pre-miRNA backbone, pre-miR-374b, that has a relatively simple structure (Figure 2.7). The cleavage reaction was done briefly under conditions where 10%, 20%, or 30% of the substrate pool was cleaved (Figure 2.10). Cleavage scores increased with incubation time but showed strong correlations between conditions (Figure 2.11), indicating that the experiments were within dynamic ranges.

The variants in the −1-to-1 window showed a broad distribution in their cleavage scores (Figure 2.9). Interestingly, two different pre-miRNAs showed clear correlations, particularly among the top scoring sequences (Figure 2.9). This

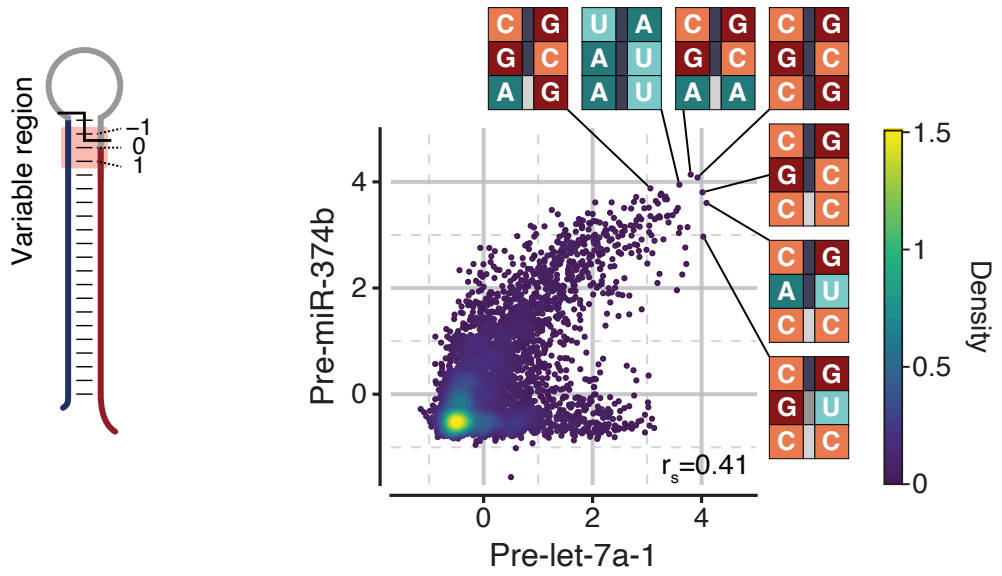

Figure 2.9 Distribution of the cleavage scores measured from the 2nd screening. Top 5 motifs in each backbone are shown. Analyses were done using data from the condition where ~20% of substrates were cleaved.

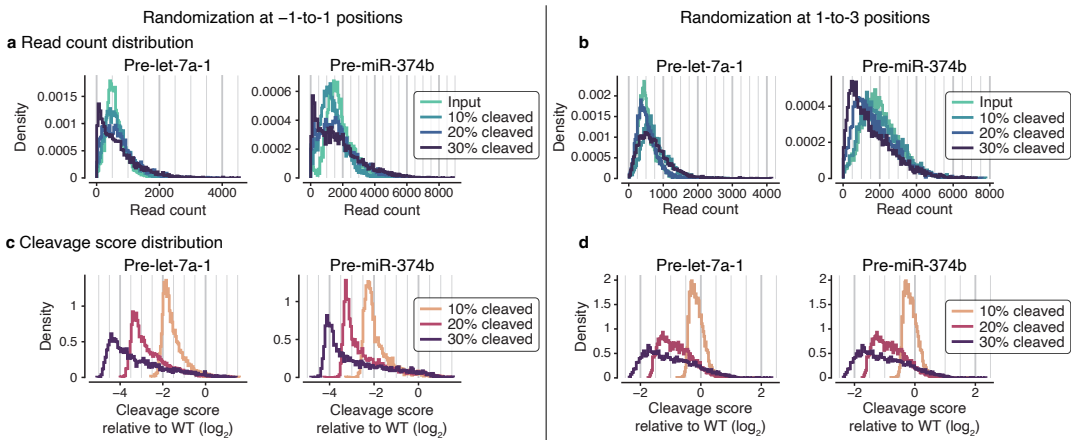

Figure 2.10 Distribution of read counts and cleavage scores of variants.

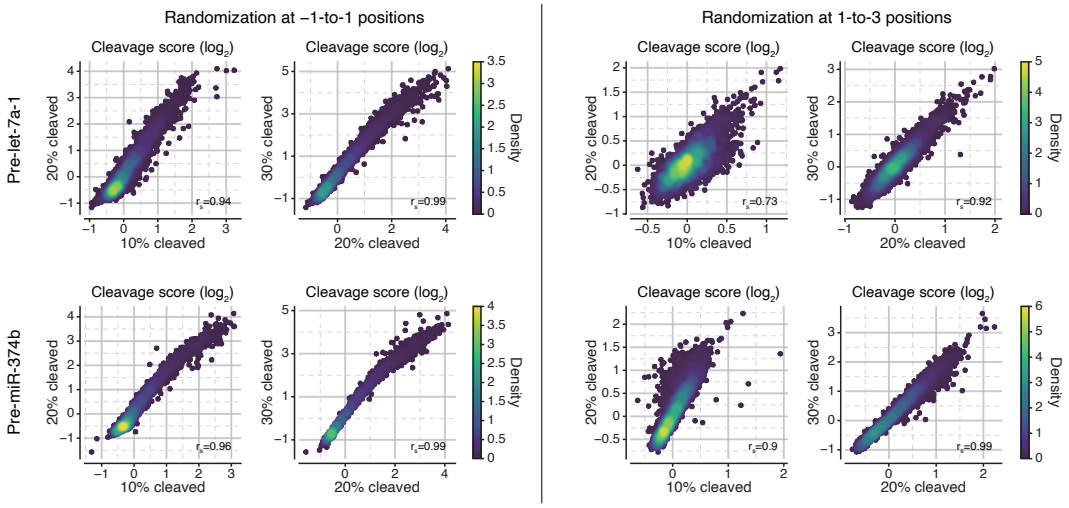

Figure 2.11 Correlation of cleavege scores of variants between different conditions of varying reaction time.

result suggests that the −1-to-1 region plays a significant role in the control of pre-miRNA processing, independently of the backbone. With the other window randomizing the 1-to-3 region, the cleavage scores of the variants did not show large differences in both pre-miRNA datasets, suggesting this region is less impactful than the −1-to-1 region (Figure 2.12).

In line with the initial screening data, we could identify sequence preferences at positions −1, 0, and 1 (Figure 2.13). At position −1, the 5´-C–G-3´ pair is strongly enriched. At position 0, purines (mainly G) are enriched in the 5p strand while pyrimidines prevail in the 3p strand, forming a base-pair. At position 1, A or C are favored while G or U are depleted.

We also found that base-pairing is overall beneficial to processing, but a mismatch at position 1 is associated with high cleavage scores (Figure 2.14, 2.15). This unexpected preference for a mismatch at position 1 was also found among the variants at positions 1-to-3 (Figure 2.16). Some base combinations at position 1 are favored over others, with the C–C mismatch generally exerting the strongest effects (Figure 2.17, 2.18). These sequence/structure preferences at position 1

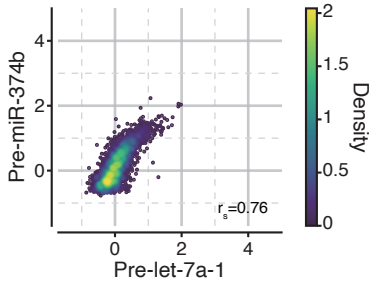

Figure 2.12 Distribution of the cleavage scores measured from the 2nd screening with 1-to-3 randomization.

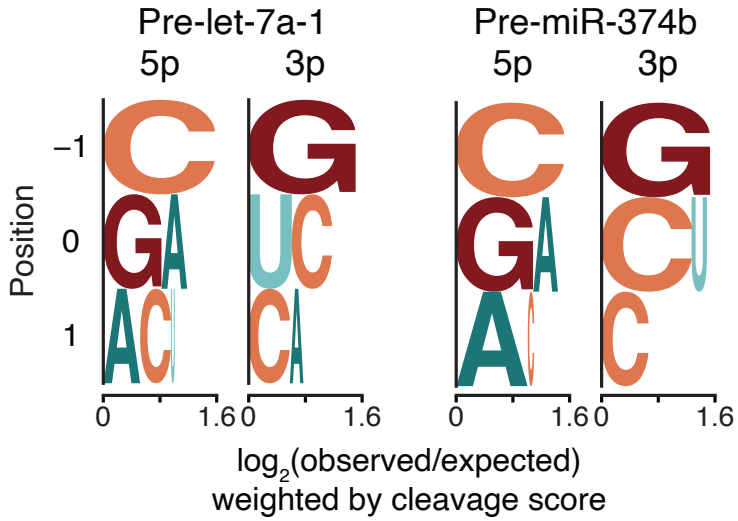

Figure 2.13 Enrichment of sequences of the top 1% variants. Expected proportions were calculated by assuming that 4 bases have the equal contributions to the total cleavage score of the top 1% variants.

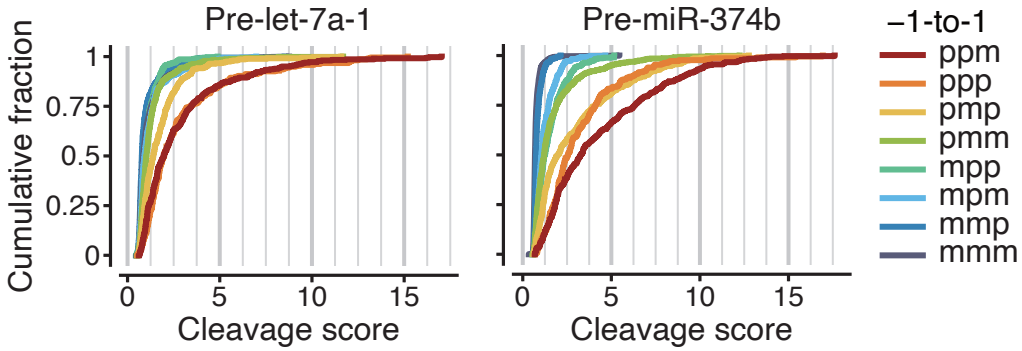

Figure 2.14 Structural impact on cleavage scores. G–U pair was considered as mismatch only when it is in between mismatches. p, pair; m, mismatch.

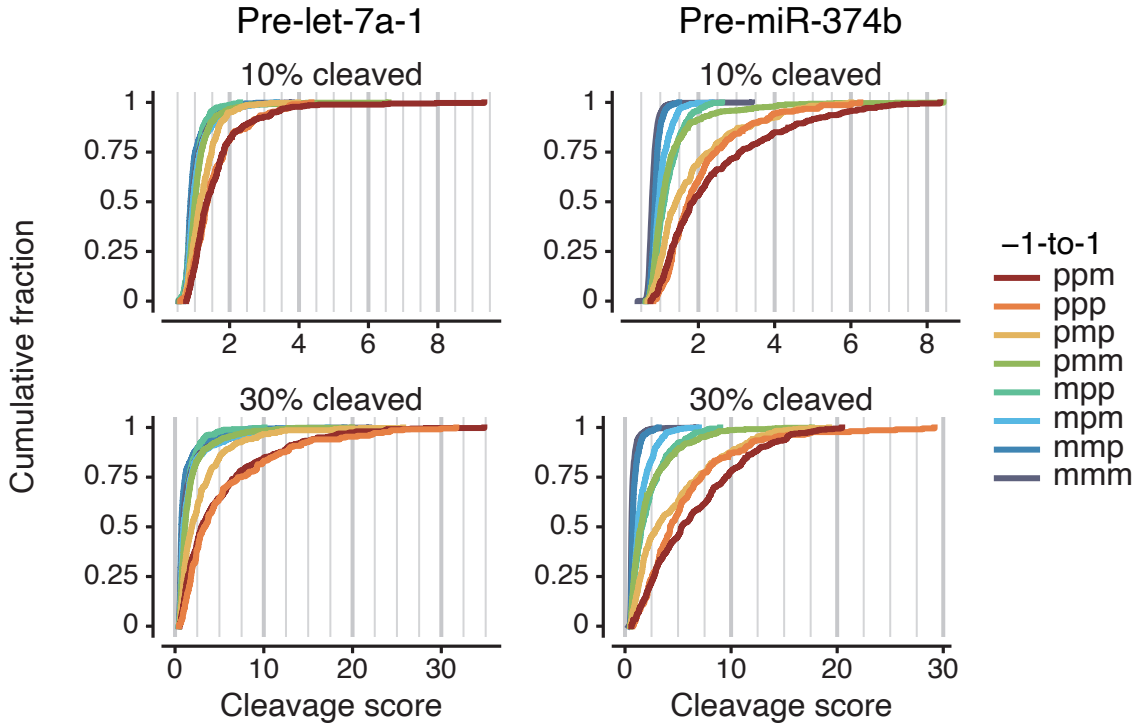

Figure 2.15 Structural impact on cleavage scores of variants (positions −1-to-1) with 10% or 30% cleavage. G–U pair was considered as mismatch only when it is in between mismatches. p, pair; m, mismatch.

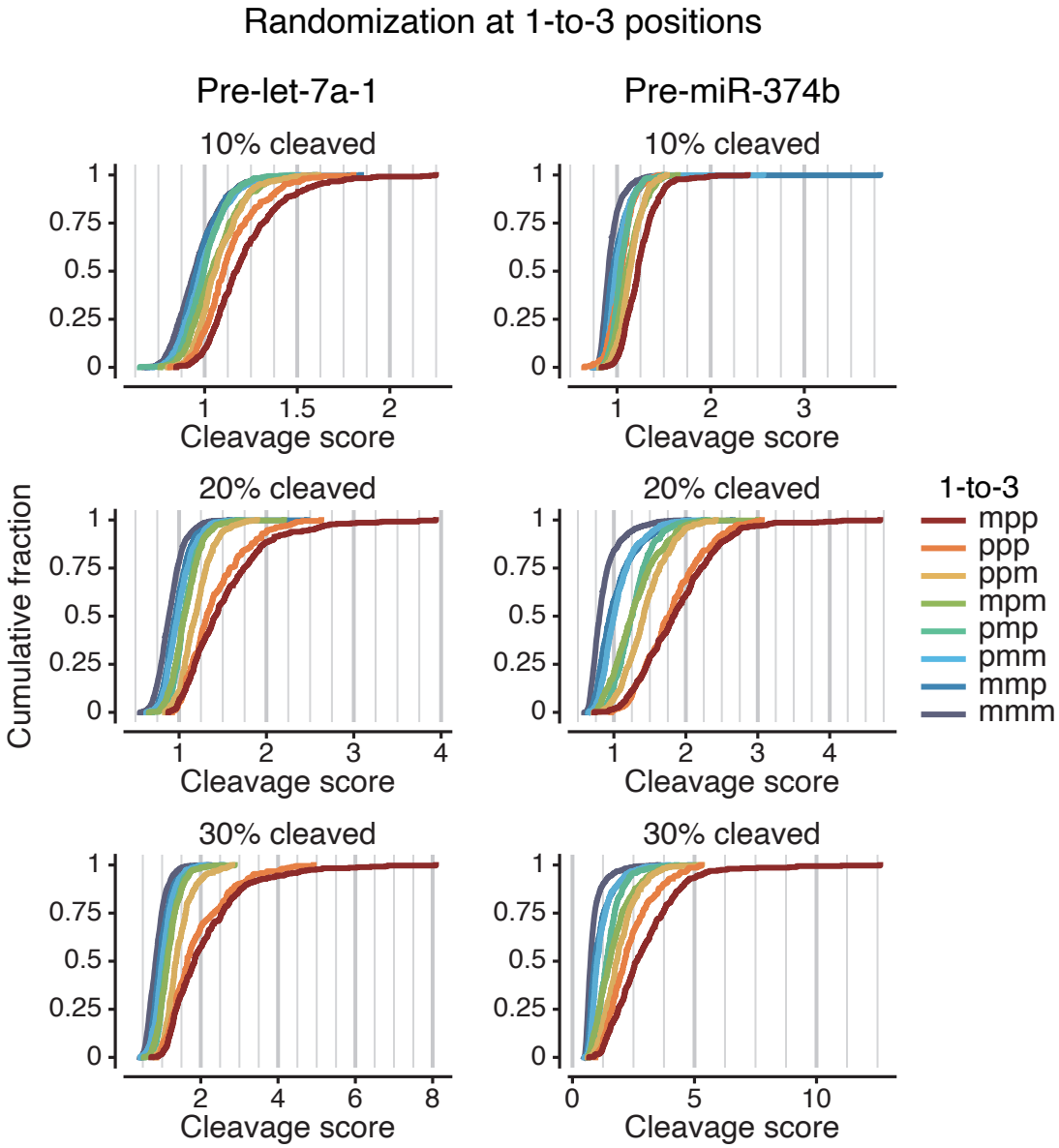

Figure 2.16 Structural impact on cleavage scores of variants (positions 1-to-3) with 10%, 20%, or 30% cleavage. G–U pair was considered as mismatch only when it is in between mismatches. p, pair; m, mismatch.

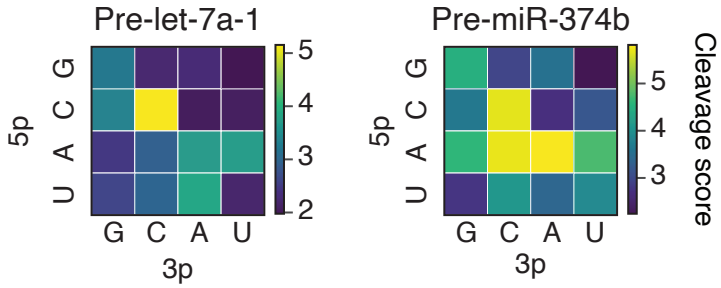

Figure 2.17 Impact of the base combinations at position 1 on cleavage scores with 20% cleavage. Variants with base-pairs at all but position 1 were included in this analysis.

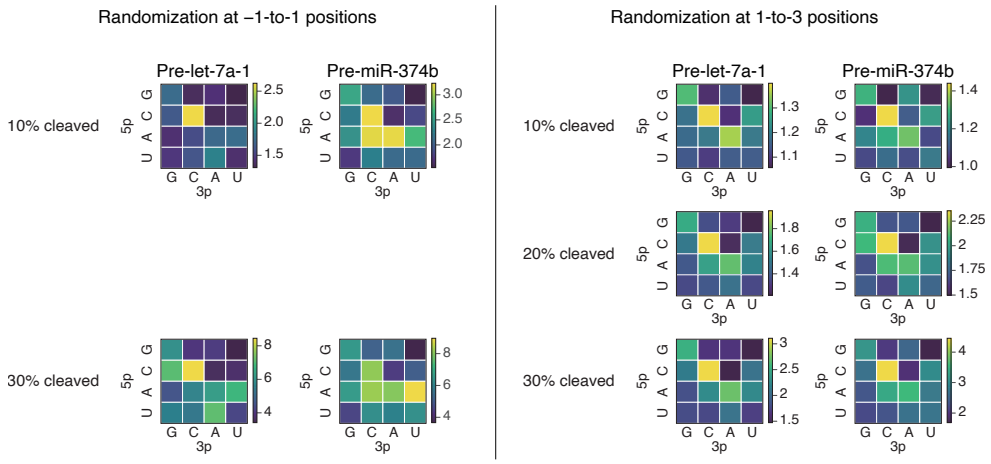

Figure 2.18 Impact of the base combinations at position 1 on cleavage scores with 10%, 20%, or 30% cleavage. Variants with base-pairs at all but position 1 were included in this analysis.

were detectable in almost all contexts examined, regardless of the sliding windows, backbones, and reaction times.

Collectively, this analysis revealed a position-dependent 3-bp motif that strongly promotes processing: a paired G, a paired pyrimidine (Y), and a mismatched C or less favorably A (M) at positions −1, 0, and 1 on the 3p side, respectively (with DICER cleaving between G and Y). We therefore named this element the 'GYM motif'. It is important to note that although there is a clear consensus sequence for the GYM motif, the cleavage scores of the variants are actually a continuum. Thus, the GYM motif should be perceived as a quantitative feature (based on the cleavage scores) rather than being defined in a binary manner (based on the sequence consensus). Therefore, we devised a metric termed the 'GYM score' to indicate the strength of the motif. The GYM score was assigned to each 3-bp combination by averaging its cleavage scores measured in the contexts of pre-let-7a-1 and pre-miR-374b with 20% cleavage, and by normalizing the average (0–100).

2.2.2 The GYM motif enhances dsRNA processing

To validate the high-throughput data, we performed *in vitro* processing assays using pre-let-7a-1 with various GYM motifs (Figure 2.19). A representative variant with a high-score motif ('GCm', GYM score 94) was processed more efficiently than the wild-type (GYM score 43) (Figure 2.19, 2.20). We further tested a variant with a G–C pair at position 1 ('GCp' scoring 38) and that with a UA dinucleotide at positions −1 and 0 ('UAm' scoring 30) which were processed less efficiently than the GCm variant. Finally, the variant replacing both parts ('UAp' scoring 14) performed least efficiently, indicating that both the paired 'GY' and the mismatched 'M' parts are important and confer additive effects on processing. We made similar observations in the presence of TRBP, a dsRNA binding protein that associates with DICER, suggesting that TRBP does not influence the GYM motif recognition by DICER (Figure 2.21).

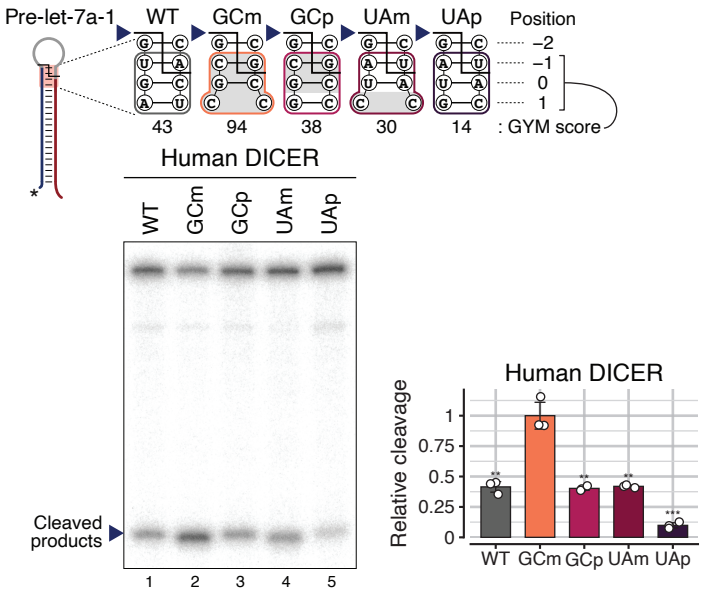

Figure 2.19 *In vitro* processing of pre-let-7a-1 variants by human DICER. Bars indicate mean ± SD (n = 3). **p < 0.01, ***p < 0.001 by two-sided Student's t test compared to GCm. Relative cleavage was calculated by quantifying the band intensities (1−uncleaved/input).

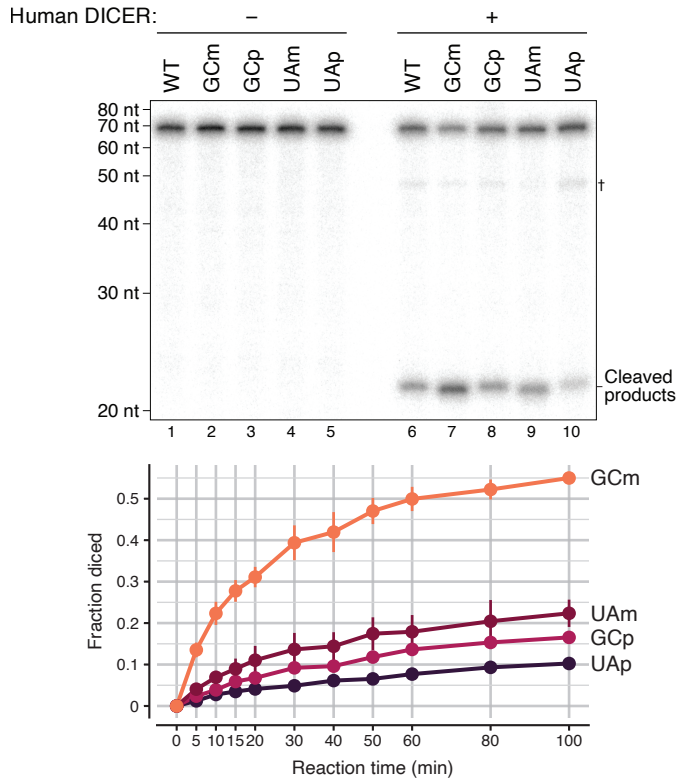

Figure 2.20 *I*Time-course *in vitro* processing of pre-let-7a-1 variants by human DICER. †, nicked products at the 3p positions.

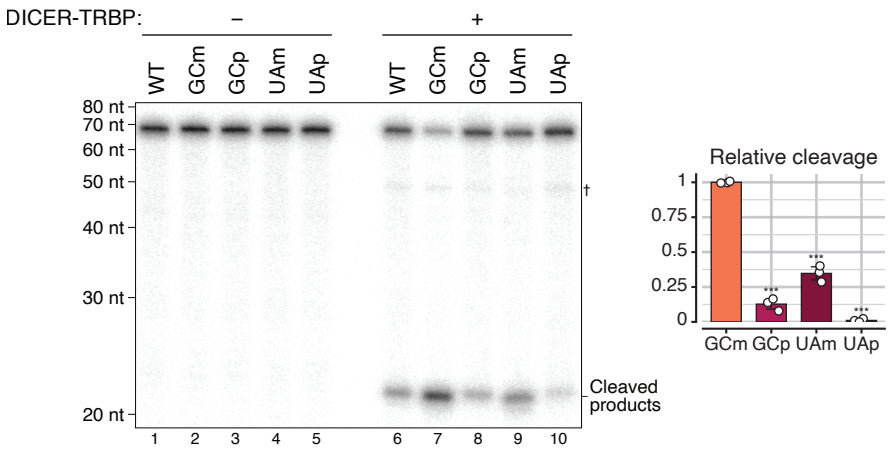

Figure 2.21 *In vitro* processing of pre-let-7a-1 variants by human DICER-TRBP complex. Substrates were radiolabeled at their 5´ ends. †, nicked products at the 3p positions.**p < 0.01, ***p < 0.001 by two-sided Student's t test compared to GCm. Relative cleavage was calculated by quantifying the band intensities (1−uncleaved/input). Bars indicate mean ± SD (n = 3). ***p < 0.001 by two-sided Student's t test compared to GCm.

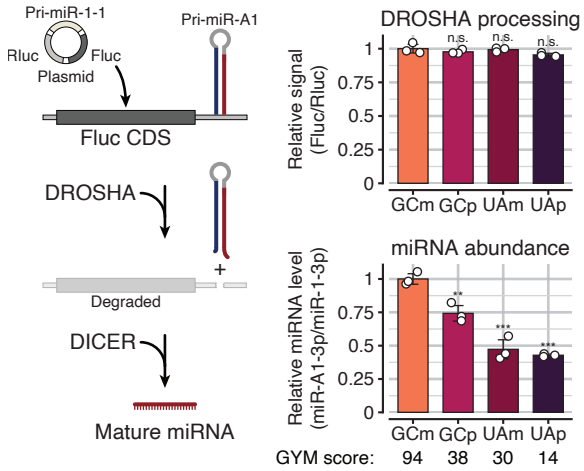

Figure 2.22 DROSHA processing assay and miRNA abundance measurement of pre-miR-A1 variants in HEK293T cells. Left: Schematic outline of this experiment. Right top: Luciferase assay. Firefly luciferase signals were normalized to Renilla luciferase (Rluc) signals. Right bottom: miRNA levels measured by qRT-PCR. The TaqMan probe was designed to target the common sequence of variants. Bars indicate mean ± SD (n = 3, biological replicates). **p < 0.01, ***p < 0.001 by two-sided Student's t test compared to GCm.

Next, to examine the motif's activity in cells, the GYM variants were incorporated into an artificial pri-miRNA hairpin (Fang & Bartel, 2015) embedded in the 3´ UTR of a luciferase reporter construct (Figure 2.22). This construct has been used to monitor DROSHA's activity because the hairpin cleavage reduces luciferase expression (Han et al., 2014). The luciferase level was not changed by the GYM motif, indicating that the motif does not impact DROSHA processing. In contrast, the mature miRNA levels from these constructs correlated with the GYM score. This is consistent with the above *in vitro* results, and further shows that the GYM motif operates regardless of the backbone.

We next questioned if this motif also contributes to cleavage site selection. We designed pre-miRNA-like duplexes with the high-score GYM motif either properly positioned or relocated by 1 bp towards the terminus. Interestingly, the relocation

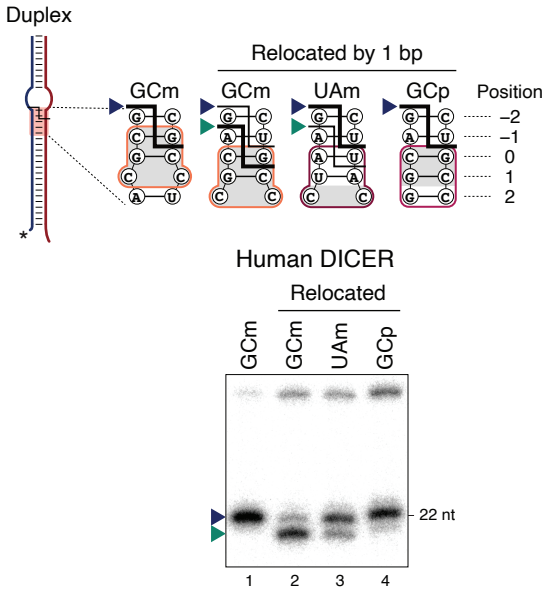

Figure 2.23 *In vitro* processing of duplex variants by human DICER. Cleavage products and their corresponding cleavage sites are marked with arrowheads.

resulted in a cleavage site shift by 1 bp (Figure 2.23). The motif lacking the GC dinucleotide ('UAm') also affected the cleavage site selection, albeit moderately. In contrast, the motif lacking the mismatch ('GCp') did not significantly alter the cleavage site. Therefore, the major impact of the GYM motif on cleavage site selection is exerted by the mismatched M, while the GC dinucleotide reinforces the influence of the mismatch and increases processing efficiency (Figure 2.24). Collectively, these results suggest that the GYM motif not only promotes efficient processing, but also serves as an important determinant of cleavage site.

To gain further insights into the functional relevance of the GYM motif in other species, we examined *Drosophila* Dicer-1 (Dcr-1) using our substrates. Like human DICER, *Drosophila* Dcr-1 exhibited a preference for the high-score GYM motif (Figure 2.25). Consistently, repositioning the motif was sufficient to alter the cleavage site of Dcr-1 (Figure 2.26). Thus, the GYM motif may play a deeply conserved role in pre-miRNA processing by metazoan DICER homologs.

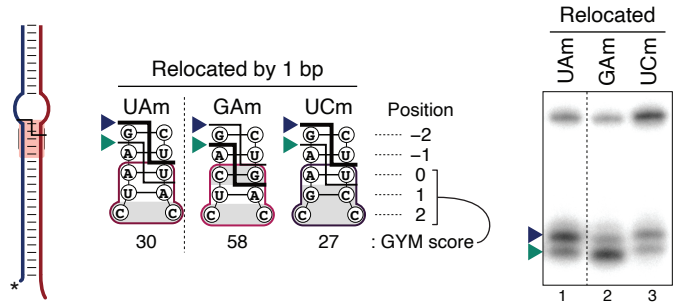

Figure 2.24 *In vitro* processing of duplex variants by human DICER. Cleavage products and their corresponding cleavage sites are marked with arrowheads.

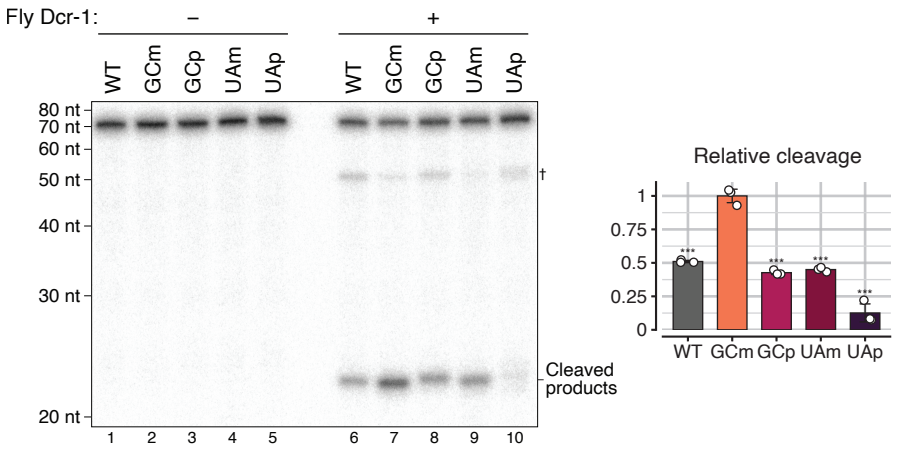

Figure 2.25 *In vitro* processing of pre-let-7a-1 variants by fly Dcr-1. Substrates were radiolabeled at their 5´ ends. †, nicked products at the 3p positions.

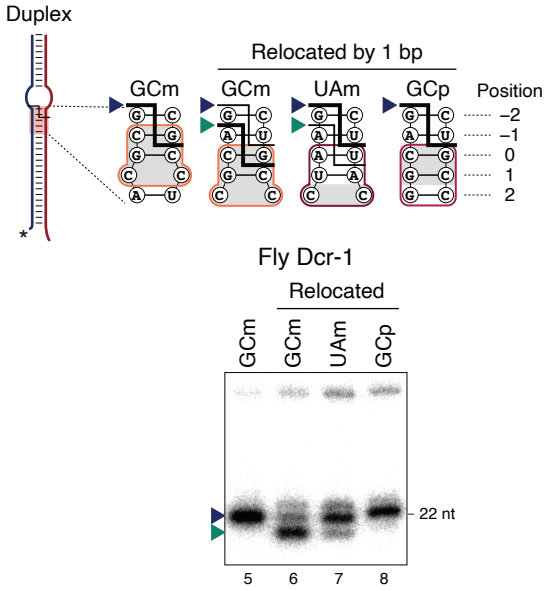

Figure 2.26 *In vitro* processing of duplex variants by fly Dcr-1. Cleavage products and their corresponding cleavage sites are marked with arrowheads.

The current model of dsRNA processing posits that DICER follows the '5′ counting' and '3′ counting' rules (Macrae et al., 2006a,b; Park et al., 2011; Zhang et al., 2004). To investigate how the GYM motif interplays with these end counting rules, we prepared a series of RNA duplexes with a low-score GYM motif ('UAp'), which shows a mixture of products from 5′ counting and 3′ counting according to the varying 1–3 nt 3′ overhang (Figure 2.27, purple and green arrowheads, respectively). However, in the presence of a high-score GYM motif ('GCm'), a single site was chosen, yielding a homogeneous product (Figure 2.27, blue arrowheads). Similarly, another set of substrates with a terminal base-pair (so as to promote 3′ counting), the high-score GYM motif not only increased the cleavage efficiency but also abolished the 3′ end counting mechanism (Figure 2.28) Thus, the optimal GYM motif is a dominating determinant of cleavage site decision, overriding the 5′ and 3′ counting rules when in conflict.

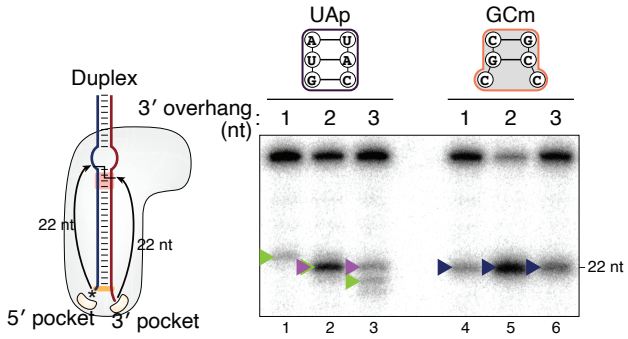

Figure 2.27 *In vitro* processing of duplex variants with a terminal mismatch by human DICER. The duplex had a mismatch at its terminus (marked in orange) so that the 5′ end can be incorporated readily into the 5′ pocket. The duplex RNAs have varying 3′ overhang (1–3 nt).

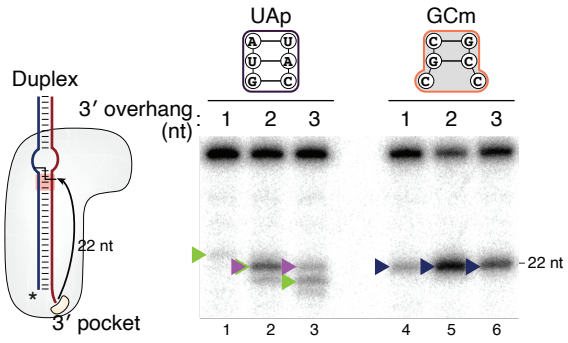

Figure 2.28 *In vitro* processing of duplex variants with a terminal base-pair by human DICER. The duplex had a base-pair at its terminus (marked in orange) so that the 5′ end cannot be incorporated into the 5′ pocket.

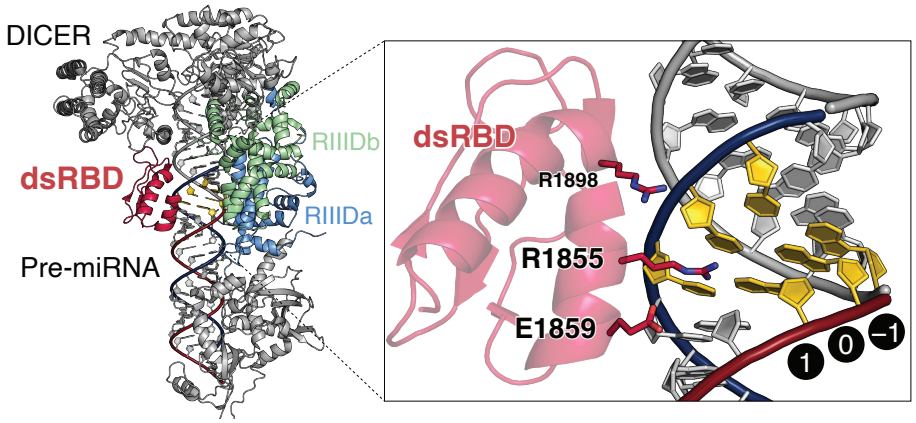

Figure 2.29 A structural model of human DICER in a dicing state. The dsRBD (marked in red) is in the vicinity of the GYM motif (marked in yellow).

2.2.3 The C-terminal dsRBD of DICER recognizes the motif

To gain mechanistic insights into the recognition of the GYM motif, we built a structural model of the DICER-dsRNA complex in a dicing state. In this predicted structure, the C-terminal dsRBD is located near the GYM motif, suggesting a role for the dsRBD (Figure 2.29). Indeed, a DICER mutant lacking the dsRBD ('ΔdsRBD') cannot distinguish a mismatch at position 1 (Figure 2.30). Contrariwise, the GC dinucleotide at positions −1 and 0 was still favored over the UA dinucleotide. Thus, the dsRBD has a critical role in sensing the mismatch, while bases at positions −1 and/or 0 may be recognized by other domain(s).

Next, we examined cleavage site selection and found that, unlike wild-type DICER, ΔdsRBD is no longer influenced by the position of the GYM motif (Figure 2.31). These results support the role of the dsRBD in the recognition of the GYM motif for cleavage site decision.

These results prompted a closer inspection on the dsRBD-dsRNA interface to identify the residues responsible for the motif recognition. Among the amino acids lining the first α-helix of the dsRBD in the minor groove of dsRNA, a highly conserved arginine residue (R1855) is located in very close proximity to

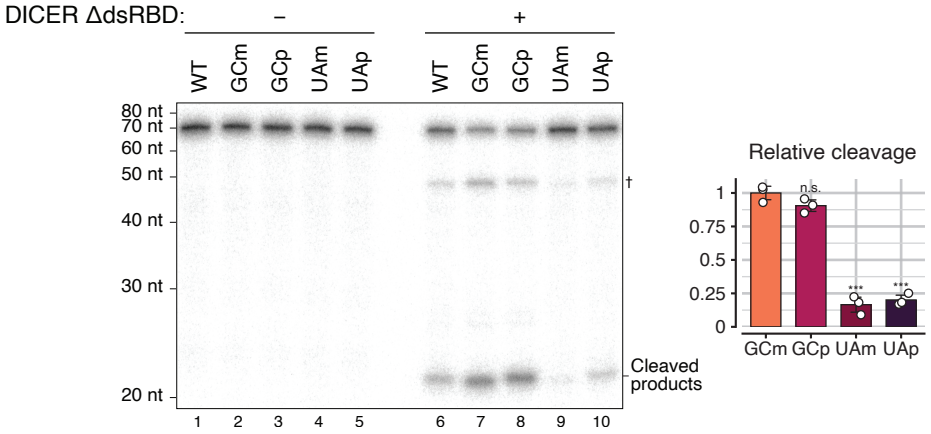

Figure 2.30 *In vitro* processing of pre-let-7a-1 variants by DICER ΔdsRBD. Bars indicate mean ± SD (n = 3). ***p $<$ 0.001 by two-sided Student's t test compared to GCm.

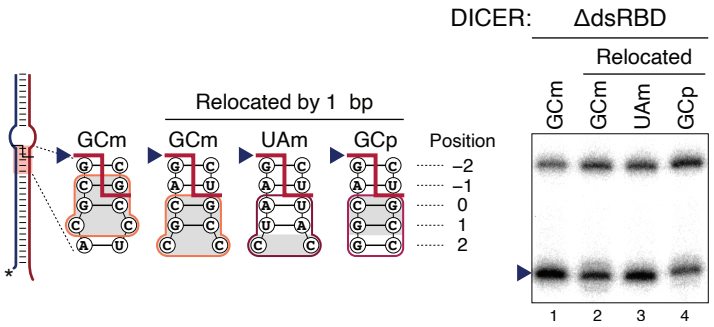

Figure 2.31 *In vitro* processing of duplex variants by either DICER ΔdsRBD or R1855L mutant. The cleavage product and its corresponding cleavage site marked with the arrowhead are largely unaffected by the GYM motif variations, which contrasts the result from WT DICER shown in panels b, d.

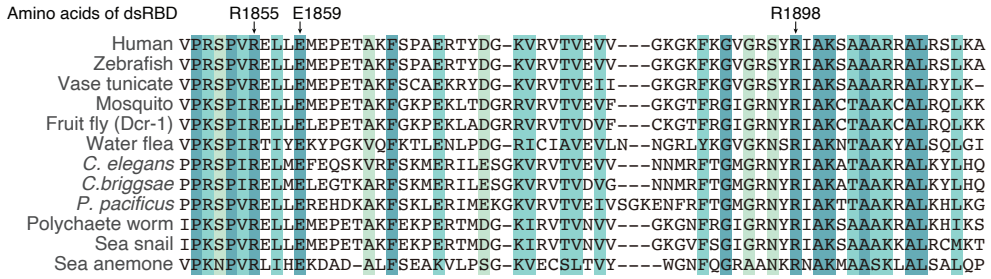

Figure 2.32 Amino acid sequence alignment of dsRBDs of metazoan DICERs.

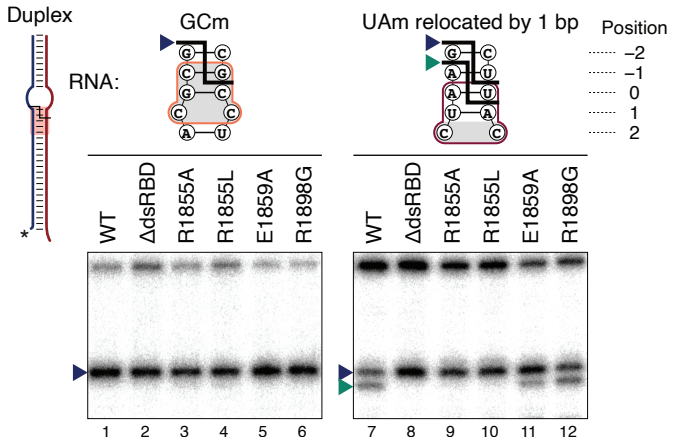

Figure 2.33 *In vitro* processing of duplex variants by human DICER point mutants at the indicated position. Cleavage products and their corresponding cleavage sites are marked with arrowheads.

the base at position 1 (Figure 2.29, 2.32). Intriguingly, a point mutation in this arginine residue to leucine (R1855L) was found in cancer based on The Cancer Genome Atlas (TCGA) (Cerami et al., 2012; Gao et al., 2013), implying its defective role in pre-miRNA processing. Indeed, mutations of this arginine to leucine ('DICER^{R1855L}') or alanine ('DICER^{R1855A}') were both sufficient to reduce the site-specific recognition of the mismatch in the GYM motif (Figure 2.33). We further interrogated the other two highly conserved residues—E1859 and R1898—situated near the mismatch and found that E1859 also contributes to the motif recognition (Figure 2.29, 2.32, 2.33). The R1855L mutant as well

as the double mutant DICER^{R1855A,R1859A} ('AA') lost their ability to favor a mismatch at position 1 (Figure 2.34, 2.35). Taken together, DICER recognizes the mismatched 'M' mainly using its highly conserved R1855 and E1859 residues of the dsRBD so as to ensure efficient and precise processing.

2.2.4 miRNA processing relies on the interaction between the GYM motif and dsRBD

We next asked if the GYM motif is biologically relevant in the context of endogenous miRNA maturation. The wild-type or mutant DICER proteins were ectopically expressed in DICER knockout (KO) HCT116 or HEK293T cells (Figure 2.36). Small RNAs were sequenced by the AQ-seq protocol for bias-minimized quantification which allows reliable comparison between miRNA isoforms (isomiRs) (Figure 2.37).

The miRNA abundance was reduced when the dsRBD was deleted or when the R1855 and E1859 residues were mutated (DICER^{R1855L} and DICER^{R1855A,R1859A} ('AA')) (Kim et al., 2019) (Figure 2.37, 2.38).[3] Some conserved and abundant miRNAs were strongly affected, prompting us to investigate whether these miRNAs are produced in a GYM-dependent manner. *In vitro* processing assays showed that mutations in the dsRBD reduced the DICER activity on pre-miR-27b, pre-miR-21, pre-let-7d, pre-let-7f-1, and pre-let-7i, thus requiring longer incubation time (Figure 2.39–2.41). Weakening variations in the GYM motif of the pre-miRNAs decreased processing efficiency, and this GYM motif effect was diminished when the dsRBD was mutated. Collectively, our data demonstrate that GYM motif recognition by the DICER dsRBD promotes processing of pre-miRNAs both in cells and *in vitro* .

We also detected substantial alterations in the cleavage sites. The most notable example was miR-324-3p which is known to be dependent on the dsRBD of DICER *in vitro* (Kim et al., 2020) (Figure 2.42). In addition, we found many miRNAs whose processing was affected by the dsRBD mutations (Figure 2.43,

[3]AQ-seq analysis was done by Dr. Haedong Kim.

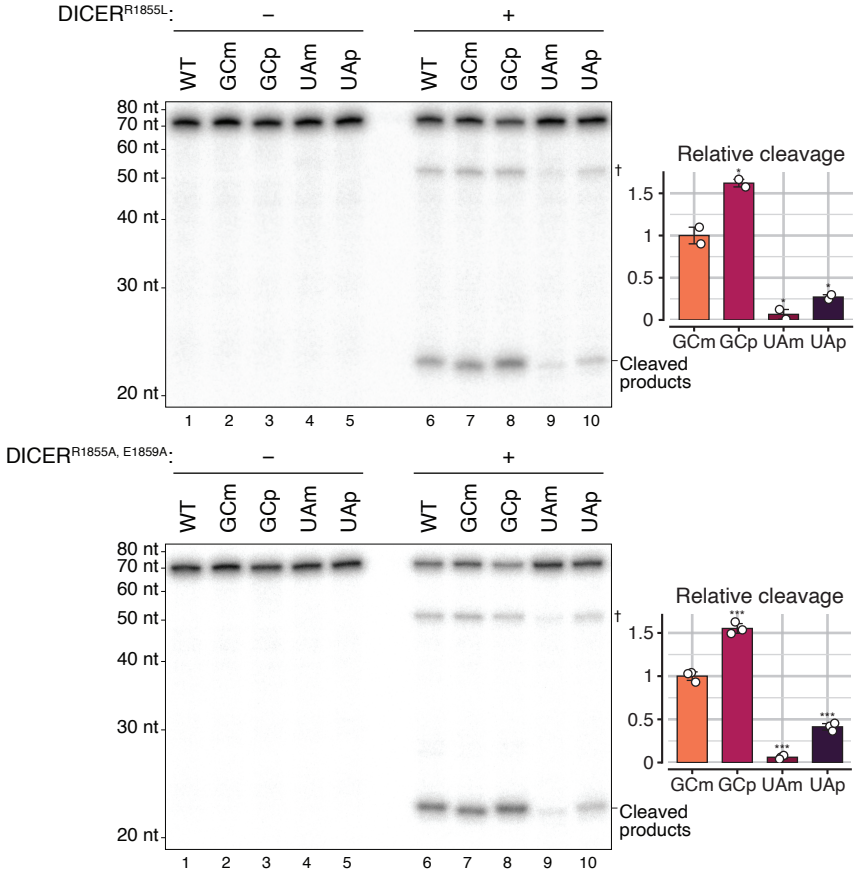

Figure 2.34 *In vitro* processing of pre-let-7a-1 variants by either DICER R1855L or R1855A/E1859A (AA). Bars indicate mean ± SD (n = 2 or 3, respectively). *p < 0.05, ***p < 0.001 by two-sided Student's t test compared to GCm. †, nicked products at the 3p positions.

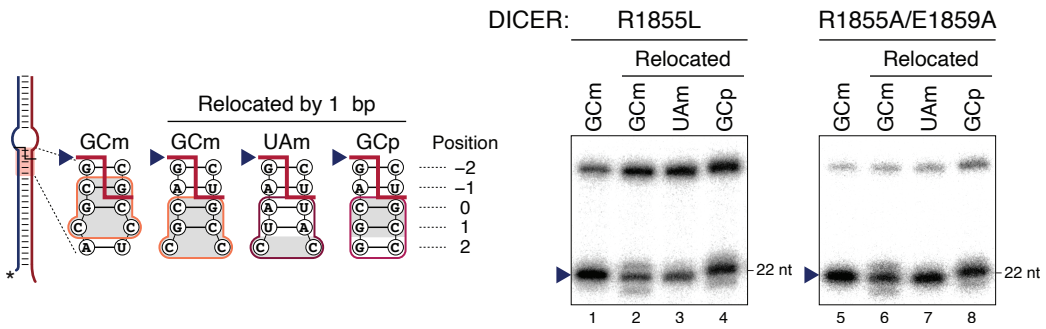

Figure 2.35 *In vitro* processing of duplex variants by either DICER R1855L or R1855A/E1859A (AA). The cleavage product and its corresponding cleavage site marked with the arrowhead are largely unaffected by the GYM motif variation.

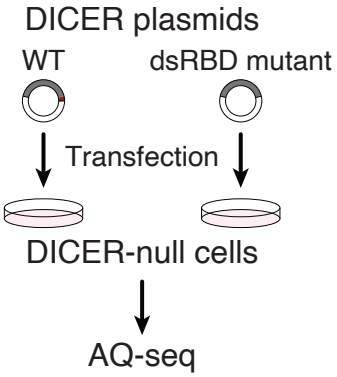

Figure 2.36 Schematic outline of the rescue experiment.

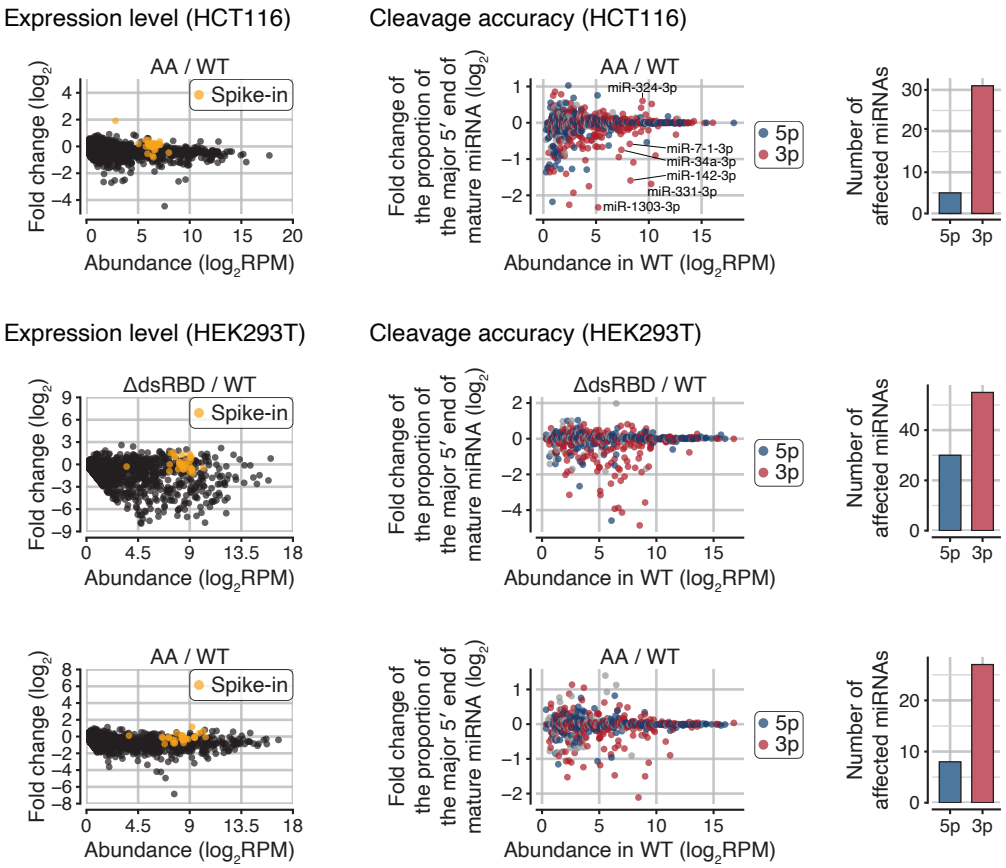

Figure 2.37 Comparison of miRNA expressions and cleavage accuracy in either HCT116 or HEK293T. Spike-ins were used for normalization. RPM, reads per million. For a given miRNA, the most abundant 5′-isomiR was identified in the WT sample. Then the fold change of its proportions in each sample was measured as cleavage accuracy. Grey, unannotated strand. Bar graphs show the number of miRNAs whose major 5′-isomiR was significantly affected by the mutation ($p < 0.01$ by two-sided Student's t test).

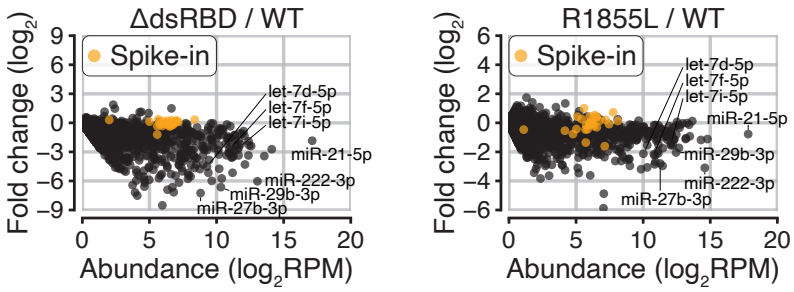

Figure 2.38 Comparison of miRNA abundance. Spike-ins were used for normalization. RPM, reads per million.

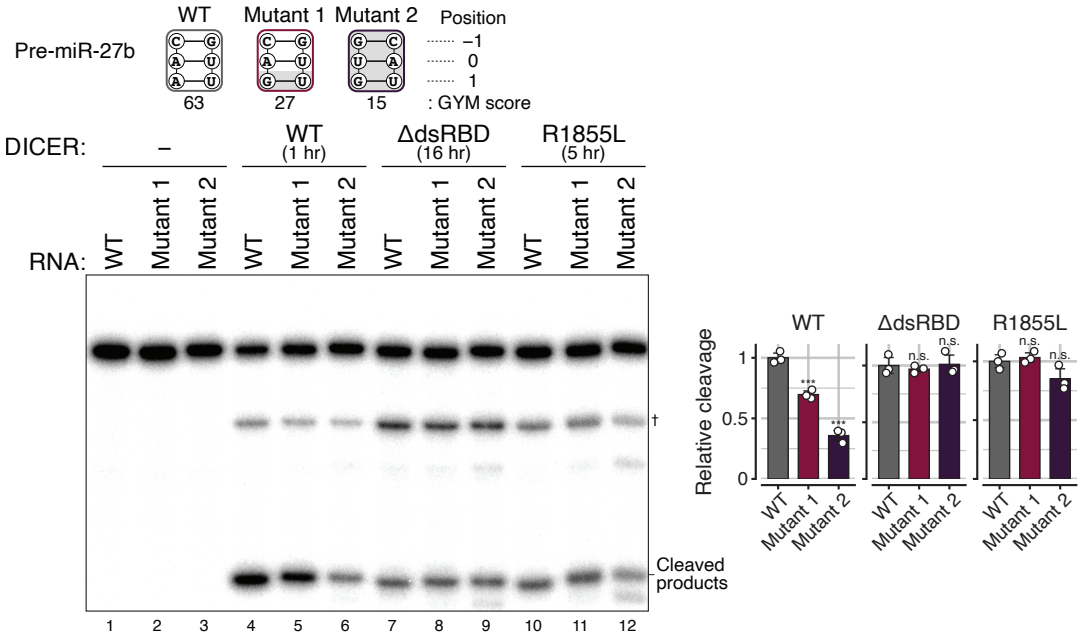

Figure 2.39 *In vitro* processing of variants of pre-miR-27b by either DICER WT or ΔdsRBD. Pre-miRNAs were radiolabeled at their 5´ end. Reactions were performed with different time points as indicated. Bars indicate mean ± SD (n = 3). *p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001 by two-sided Student's t test compared to the WT substrate. †, nicked products at the 3p positions.

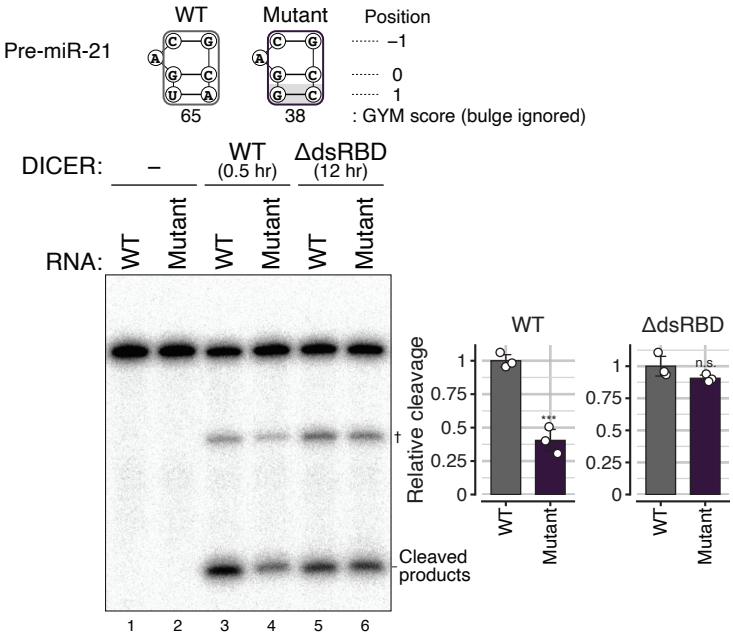

Figure 2.40 *In vitro* processing of variants of pre-miR-21 by either DICER WT or ΔdsRBD. Pre-miRNAs were radiolabeled at their 5´ end. Reactions were performed with different time points as indicated. Bars indicate mean ± SD (n = 3). *p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001 by two-sided Student's t test compared to the WT substrate. †, nicked products at the 3p positions.

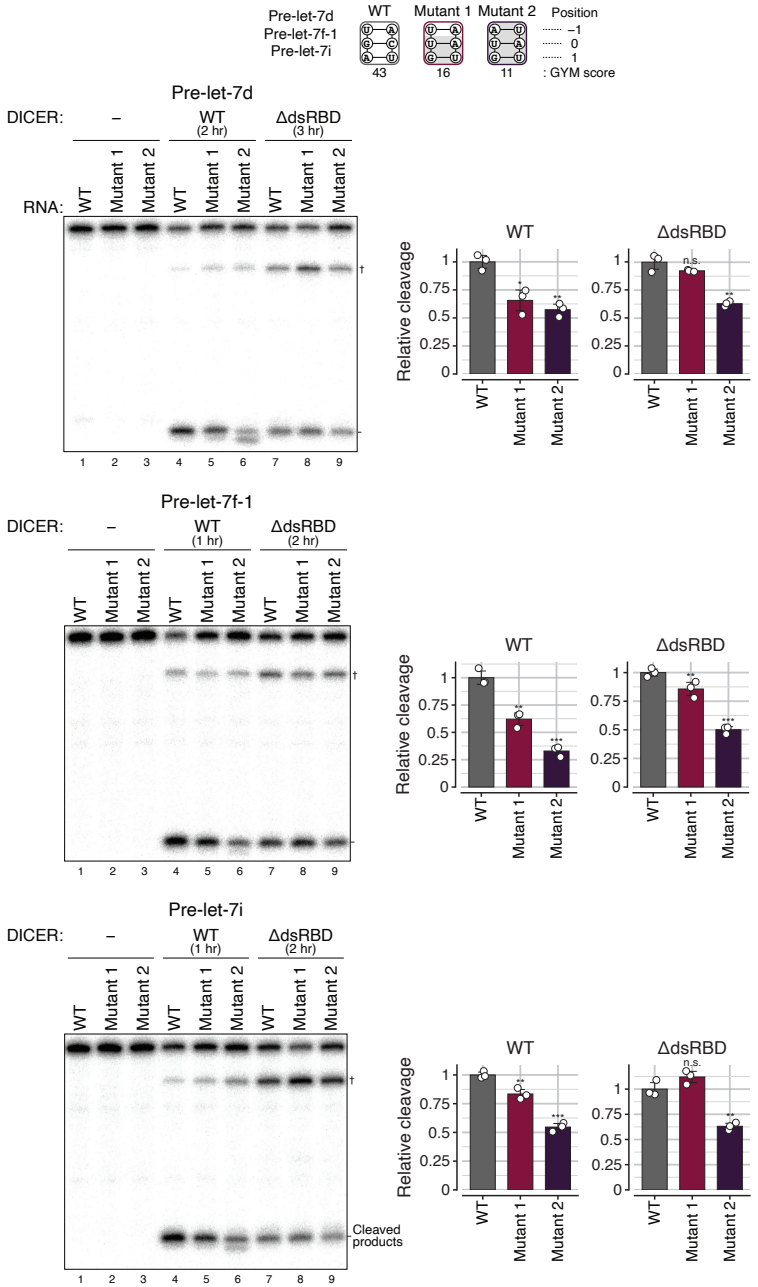

Figure 2.41 *In vitro* processing of variants of pre-let-7d/f-1/i by either DICER WT or ΔdsRBD. Reactions were performed with different time points as indicated. Bars indicate mean ± SD (n = 3). *p < 0.05, **p < 0.01, ***p < 0.001 by two-sided Student's t test compared to the WT substrate. †, nicked products at the 3p positions.

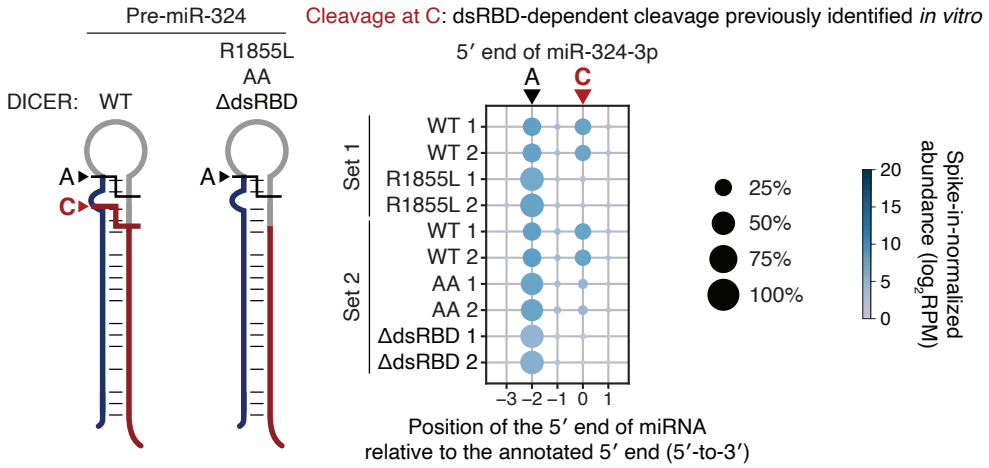

Figure 2.42 Alternative processing of pre-miR-324.

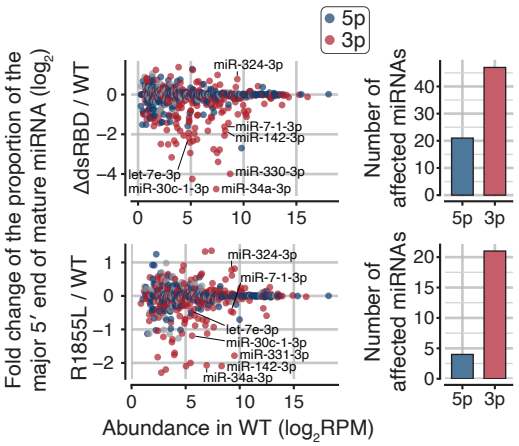

Figure 2.43 Comparison of cleavage accuracy. For a given miRNA, the most abundant 5´-isomiR was identified in the WT sample. Then the fold change of its proportions was quantified to estimate cleavage accuracy. Grey, unannotated strand. Bar graphs show the number of miRNAs whose major 5´-isomiR was significantly affected by the mutation ($p < 0.01$ by two-sided Student's t test).

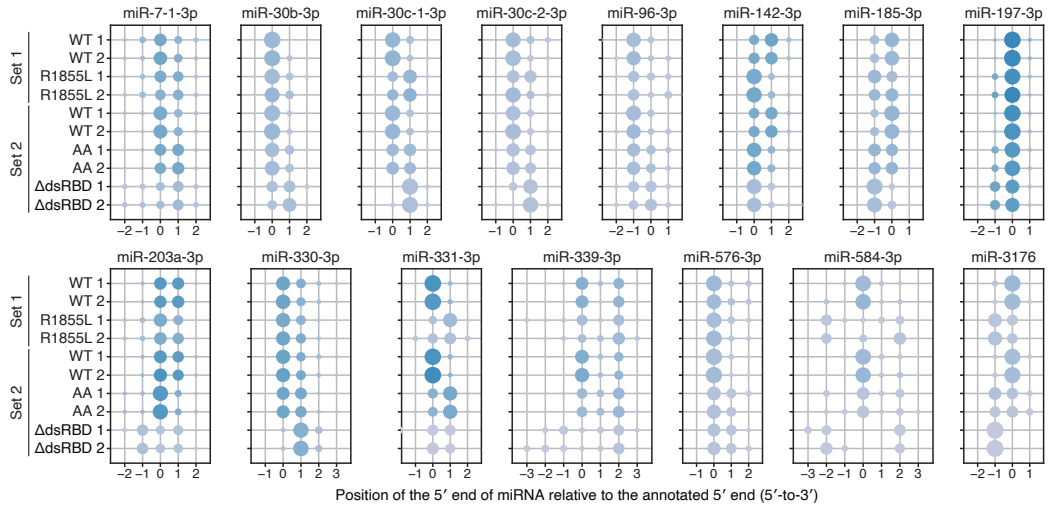

Figure 2.44 The usage of 5´ ends of miRNAs in the DICER-null HCT116 cells rescued with indicated DICER. The annotation in miRBase release 21 was used as a reference. Cleavage sites and their corresponding positions are marked with arrowheads. RPM, reads per million.

2.44). To infer the cleavage sites by DROSHA and DICER, we examined the 5´ ends of 5p miRNAs (blue dots) and 3p miRNAs (red dots), respectively. For quantification, we measured the proportion of the major 5´-isomiR (i.e., isomiRs with the same 5´ end). The DICER dsRBD mutations affected mainly the DICER cleavage sites, leading to seed alterations and/or strand switches of many miRNAs (Figure 2.43, 2.44) (Kim et al., 2020). We observed similar changes in HEK293T cells (Figure 2.37). For validation, we performed *in vitro* assays with pre-miR-34a and pre-let-7e and found that their cleavage patterns drastically changed when the dsRBD was deleted, consistently with the sequencing data (Figure 2.37, 2.43, 2.45). The GYM motif mutations altered processing sites in these pre-miRNAs, confirming the importance of the GYM motif in cleavage site choice. Importantly, these GYM variants cannot be distinguished by the ΔdsRBD mutant, indicating that the dsRBD-GYM interaction is critical for the cleavage site decision in human pre-miRNA processing. It is noteworthy that the GYM motif seems to have differential effects on different miRNAs, either increasing (e.g.,

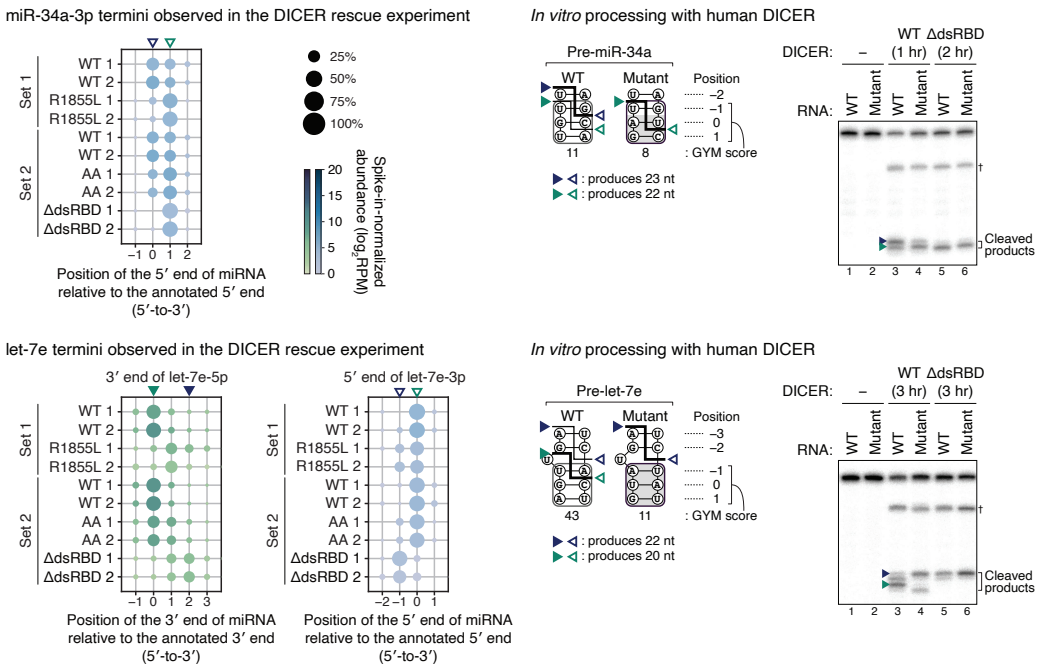

Figure 2.45 GYM motif-mediated cleavage site decision of miR-34a and let-7e. The usage of 5´ ends of miR-34a-3p or 3´ ends of let-7e-5p and 5´ ends of let-7e-3p in the DICER-null HCT116 cells rescued with indicated DICER. The annotations in miRBase release 21 were used as references. Corresponding positions of the major cleavage sites are marked with arrowheads. RPM, reads per million. *In vitro* processing of pre-miR-34a variants or pre-let-7e variants by either DICER WT or ΔdsRBD. Major cleavage products and their corresponding cleavage sites are marked with arrowheads. Reactions were performed with different time points as indicated because DICER ΔdsRBD has reduced activity.

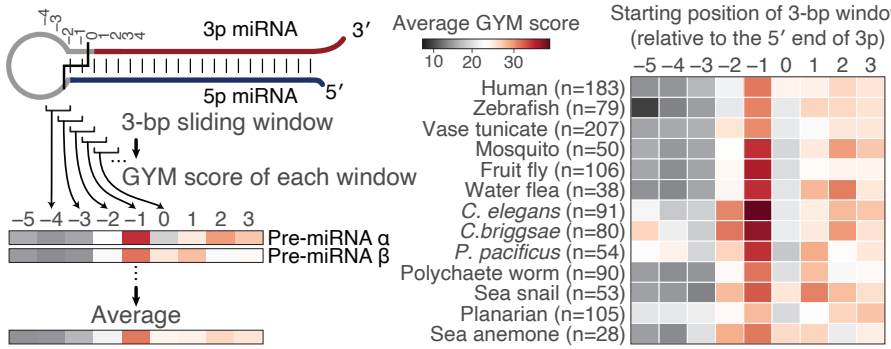

Figure 2.46 Distribution of the GYM motif at the indicated positions in natural pre-miRNAs across species.

pre-miR-7-1 and pre-miR-30c) or decreasing (e.g., pre-miR-34a and pre-let-7e) the cleavage homogeneity (Figure 2.44, 2.45), which may be because the GYM motif cooperates or competes with the end recognition mechanisms depending on its relative position.

2.2.5 Evolutionary implications of the GYM motif

If the GYM motif plays a significant and general role in miRNA maturation, one would expect that GYM motifs are pronounced at position −1-to-1 across multiple pre-miRNAs and various species. Indeed, in human pre-miRNAs, the GYM scores are higher at positions −1-to-1 compared to those at neighboring positions (Figure 2.46, human).[4] Remarkably, such position-specific enrichment was observed in all metazoan species examined, that is, from sea anemone to human. Together with our biochemical data that fly Dcr-1 is facilitated by the GYM motif *in vitro* (Figure 2.25, 2.26), this positional conservation pattern suggests that the GYM motif is deeply rooted in eumetazoan evolution.

Notably, homogeneously processed miRNAs show stronger positional enrichment at a single site than alternatively processed miRNAs do (Figure 2.47). This observation suggests that while GYM motifs that are 'prominent' relative to the

[4]Analysis was done by Dr. Haedong Kim.

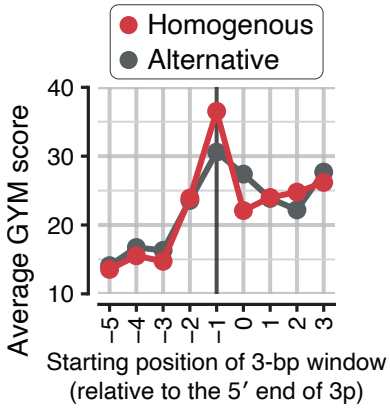

Figure 2.47 The association between the GYM score and alternative processing. miRNAs registered in miRGeneDB were included in this analysis. For a given miRNA, if the proportion of the most abundant 5´-isomiR is over 80%, it was classified into the 'homogenous processing' group (n = 203). Otherwise, it was classified into the 'alternative processing' group (n = 76).

neighboring positions facilitate the production of single miRNA species, 'not-so-prominent' GYM motifs allow alternative processing, yielding multiple isomiRs and diversifying miRNA functions, which possibly explains why the majority, but not all human pre-miRNAs evolved to have prominent GYM motifs (Figure 2.48).

To experimentally test this hypothesis, we selected pre-miR-9 which is alternatively processed by DICER in cells to produce major 22-nt and minor 21-nt isoforms (Figure 2.49, left panel, blue and green arrowheads, respectively). Importantly, pre-miR-9 lacks a prominent GYM motif, with comparable GYM scores between two 3-bp sliding windows corresponding to the two isomiRs. We introduced a weaker GYM variant specifically to the lower window, without altering the GYM motif in the upper window, which in effect made the upper motif more 'prominent' without increasing its score. In this situation, pre-miR-9 was no longer alternatively processed, indicating that the comparable GYM motifs in two consecutive windows allowed alternative processing of pre-miR-9 (Figure 2.49).

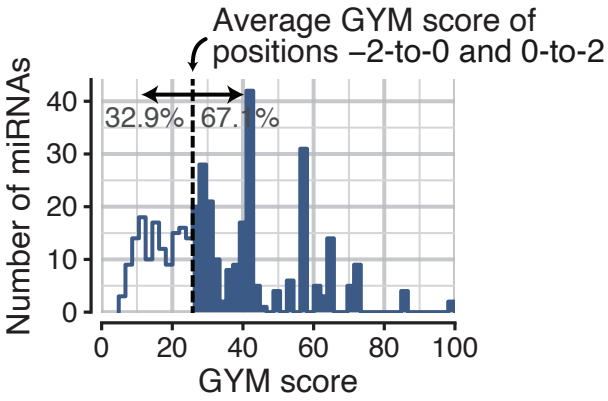

Figure 2.48 The GYM scores at the position −1 of human pre-miRNAs. miRNAs registered in miRGeneDB (n = 383) were included in this analysis. The dashed line indicates the average of GYM scores of the surrounding positions (−2 and 0).

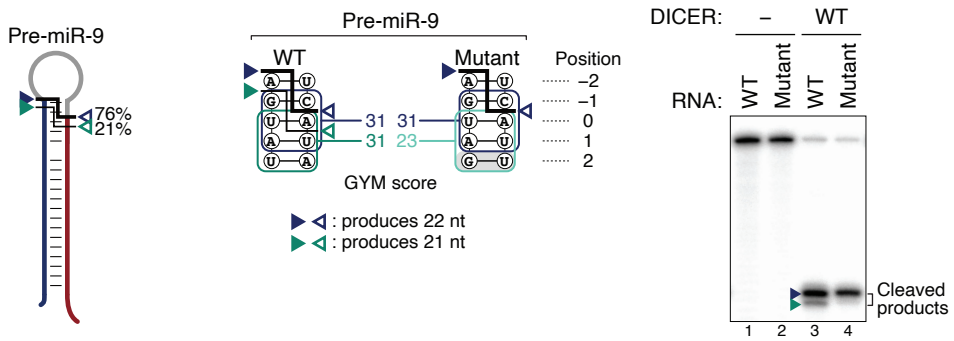

Figure 2.49 The GYM scores at the position −1 of human pre-miRNAs. miRNAs registered in miRGeneDB (n = 383) were included in this analysis. The dashed line indicates the average of GYM scores of the surrounding positions (−2 and 0).

Thus, depending on their relative strength and position, GYM motifs can be used to facilitate either homogeneous or alternative processing of pre-miRNAs.

2.2.6 The GYM motif improves the design of shRNA and DsiRNA

Having identified the GYM motif as a strong determinant of DICER-mediated processing, we next sought to investigate its effect on RNA interference by shRNA (short hairpin RNA) and DsiRNA (DICER-substrate siRNA) that are dependent on cellular DICER to yield siRNAs. shRNA mimics pre-miRNA but contains a 5´ triphosphate group and a 3´ overhang with 3–5 uridines originating from the RNA polymerase III termination signal (Gao et al., 2018). This non-optimal terminal structure impairs DICER processing, posing a drawback of shRNA-mediated RNAi. To test if the GYM motif can override the end-dependency, we constructed shRNAs with GYM motifs of variable scores (Figure 2.50). The shRNA with a high-scoring GYM motif yielded the highest-level of siRNA products and the most potent gene silencing activity.[5]

We also evaluated the GYM motif in the context of DsiRNA—synthetic 27-mer siRNA duplex designed to be processed by DICER to yield 22-mer siRNA (Amarzguioui et al., 2006). DsiRNAs exhibit more potent activity than synthetic 21-mer siRNA duplexes which bypass DICER processing, suggesting a coupling between DICER processing and AGO loading (Kim et al., 2005; Snead et al., 2013). DsiRNAs can be designed in two different configurations, in which the antisense guide sequence is placed in either strand. Strong GYM motifs increased the knockdown efficiencies, regardless of whether the antisense is placed in the 5p or 3p strand (Figure 2.51). Taken together, the identified GYM motif promotes both shRNA- and DsiRNA-mediated RNAi, allowing effective and rapid knockdown of target genes.

[5]shRNA experiments were performed by Dr. Haedong Kim.

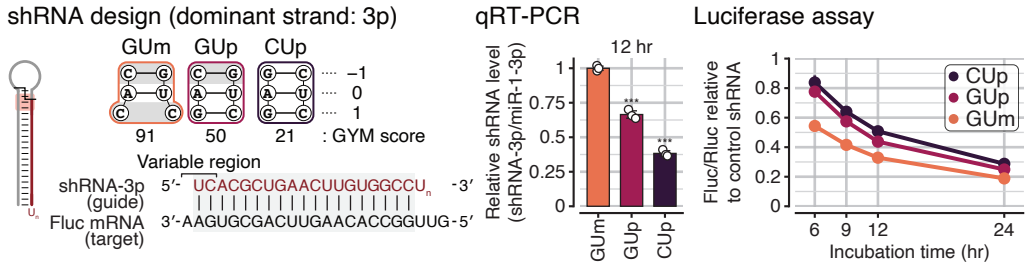

Figure 2.50 The effect of GYM motif in RNA interference by shRNA. Design of shRNAs with optimal (GUm), suboptimal (CUp), and poor (CUp) GYM motifs. Un indicates varying number of uridines which correspond to the Pol III transcription terminator. Note that all variants are expected to produce identical 3p siRNA. shRNA levels measured by qRT-PCR. shRNA was encoded in a plasmid which also encoded pri-miR-1-1 under an independent promoter. Bars indicate mean ± SD (n = 3, biological replicates). ***p $<$ 0.001 by two-sided Student's t test compared to GUm. Luciferase assay. Firefly luciferase (Fluc) signals were normalized to Renilla luciferase (Rluc) signals. A control shRNA which does not target Fluc was used to measure gene silencing activities. Points and bars indicate mean ± SD (n = 3, biological replicates).

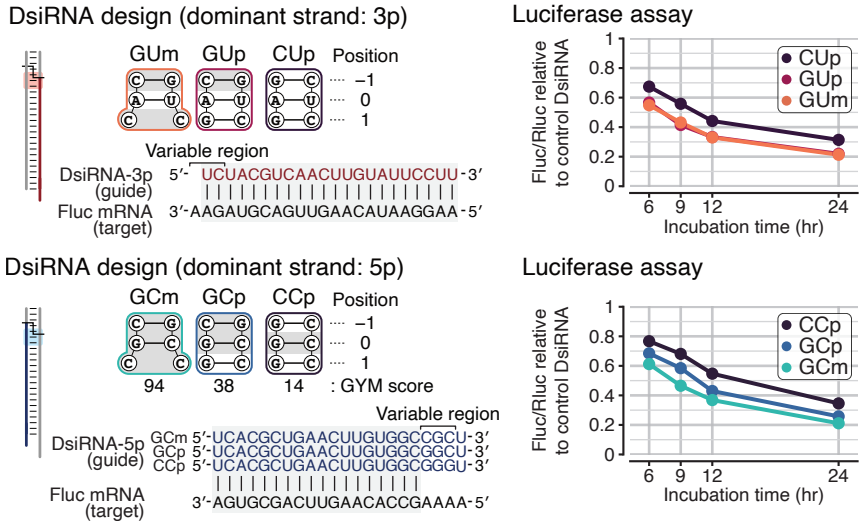

Figure 2.51 Points and bars indicate mean ± SD (n = 3, biological replicates).

2.3 Discussion

Here we employed massively parallel assays to identify a position-dependent sequence motif for DICER processing, namely the GYM motif. The GYM score measured in this study offers a way to quantitatively define optimal substrates of DICER. The GYM score generally peaks at positions −1-to-1 relative to the cleavage site and this positional pattern is conserved throughout animal miRNAs. Both fly Dcr-1 and human DICER show similar regulation *in vitro*. These findings strongly suggest that the GYM-mediated mechanism is conserved across metazoan species. It remains to be investigated if a similar mechanism is at work in plants and fungi and if this also applies to siRNA-specific DICER homologs such as fly Dcr-2. Moreover, it will be also intriguing to interrogate the remaining regions of pre-miRNAs—the loop and the lower part of the stem—to examine additional sequence determinants. However, it is noteworthy that because the lower part of the stem becomes mature miRNAs, the sequence changes would be constrained by target interaction.

Central to the GYM-mediated mechanism is the DICER dsRBD that recognizes the GYM motif. The DICER dsRBD was known to play a role in dsRNA binding and hence enhance cleavage efficiency (Zhang et al., 2002; Gan et al., 2006). Our study reveals how the dsRBD achieves its function by recognizing the substrate RNA in not only sequence-independent but also sequence-dependent manners. Furthermore, this study unveils the function of the dsRBD in the cleavage site decision which even can override the other domains' function. In light of our finding, it is noteworthy that dsRBDs of bacterial RNase III (class I RNase III) and human DROSHA (class II RNase III) also contribute to substrate specificity by recognizing the sequence and structural features near the cleavage sites although the positions and sequences of the motifs are different (Fang & Bartel, 2015; Kwon et al., 2019; Kim et al., 2017). Therefore, together with the previous reports, the current study expands our understanding of dsRBDs by illustrating its high amenability conferring specificity to a wide array of sequence/structural compositions in dsRNA.

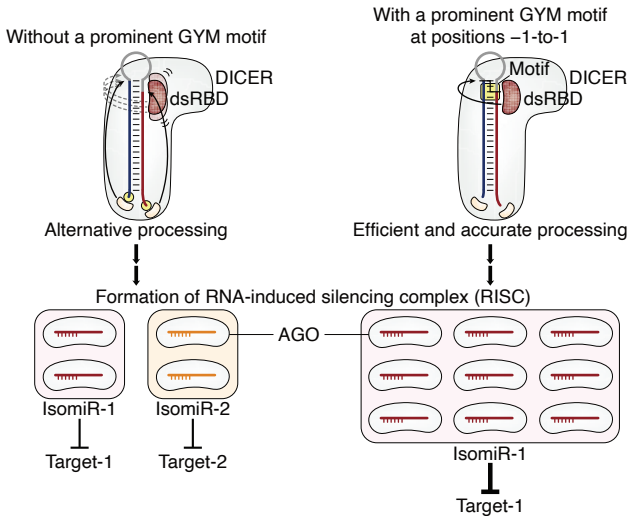

Figure 2.52 A proposed model for the role of GYM motif in DICER processing.

The GYM motif is a decisive factor specifying the cleavage site, which can override the 5´ and 3´ end counting rules when in conflict. However, when consistently positioned relative to the termini, the GYM motif can cooperate with the termini to ensure accurate and homogenous processing. Taking our results together, we propose a model for the role of the GYM motif in small RNA biogenesis and its effects on RNA silencing (Figure 2.52). For a pre-miRNA with a not-so-prominent GYM motif, the dsRBD of DICER does not form a tight interaction at a specific position (top left model). In this scenario, the processing accuracy is compromised, generating multiple isoforms. This mechanism can offer a useful basis for regulation via alternative processing that can expand and/or alter the target repertoire by changing the seed sequence and sometimes even switching the functional strand (a phenomenon called 'arm switching') as seen with miR-9 and miR-324 (Kim et al., 2020).

In situations where there is a prominent GYM motif, the dsRBD latches onto the GYM motif to accommodate the dsRNA in a fixed position (Figure 2.52, top right model). In this case, the impact of the end counting rule can be negated. Thus, this study amends the previous notion that the DICER cleavage

site is passively determined by DROSHA that generates pre-miRNA termini. Accordingly, the GYM motif can compensate for the imprecise (alternative) processing of pri-miRNAs by DROSHA (Kim et al., 2019; Chiang et al., 2010; Kim et al., 2017, 2021; Bofill-De Ros et al., 2019; Wu et al., 2009; Burroughs et al., 2010) or negate the effect of promiscuous end modifications (such as 3´ tailing and trimming)—which would otherwise affect DICER cleavage sites (Kim et al., 2019; Chiang et al., 2010; Martin, 2011). This allows DICER to precisely process the substrate to produce miRNAs/siRNAs with high accuracy and efficiency, which consequently contributes to the fidelity and potency of RNA silencing.

Altogether, our study reveals an integral and conserved mechanism of DICER's substrate recognition and provides a framework to understand how DICER produces small RNAs for biological and therapeutic regulations. Moreover, our finding of a cancer-associated DICER mutation (R1855L) which disrupts the GYM motif recognition offers insights into the molecular impact of this mutation.

2.4 Methods

Plasmid construction

To construct plasmids for expressing human DICER, the coding sequence of the human DICER reference (RefSeq NM_030621) was amplified and subcloned into the pX vector (for protein purification) containing N-terminal His_{10}-eYFP-SUMOstar-Strep sequence, or pCK vector (for ectopic expression in DICER KO cells) with a PGK promoter substituted for the CMV promoter. Other DICER constructs, including DICER ΔdsRBD (1–1848 aa), were generated via site-directed mutagenesis. For the plasmid expressing human TRBP, the coding sequence of the human TRBP reference (RefSeq NM_134323) was amplified and subcloned into the same pX vector. For the plasmid expressing *Drosophila* Dcr-1, the coding sequence of the *Drosophila* Dcr-1 reference (RefSeq NM_079729.3) was amplified from *Drosophila* cDNA and subcloned into the same pX vector. To construct plasmids for dual reporter luciferase assays, synthetic DNA oligos containing the pri-miRNA sequences or the target sites of shRNAs and DsiRNAs

were inserted into the pmirGLO vector (Promega), downstream of the firefly luciferase gene. To construct plasmids for shRNA expression, synthetic DNA oligos containing the shRNA sequences and pri-miR-1-1 sequence were inserted into a vector, downstream of U6 and UbC promoters, respectively.

Protein purification

N-terminal His_{10}-eYFP-SUMOstar-Strep tagged DICER expression plasmids were transiently transfected into HEK293E cells, derived from human embryonic kidneys, that were grown in suspension culture (Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal bovine serum) in a 8% CO_2 humidified shaking incubator at 37°C. At a cell density of ~7×10^6 cells per mL, each plasmid was transfected at a final concentration of 0.3 µg ml^{-1} with linear polyethylenimine (PEI) and dimethyl sulfoxide (DMSO) sequentially added at a final concentration of 3 µg ml^{-1} and 1%, respectively. Transfected cells were incubated at 33°C for 72 hours.

All purification procedures were carried out at 4°C. Cell pellet was resuspended in buffer containing 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10% glycerol, and 2 mM "mercaptoethanol (BME), supplemented with 0.1 mM phenylmethylsulfonyl fluoride (PMSF), EDTA-free Pierce Protease Inhibitor Mini Tablets (Thermo Fisher Scientific), 20 µl ml^{-1} micrococcal nuclease, and 5 mM $CaCl_2$ Cells were lysed by sonication and centrifuged at 35,000 g for 1 hr. The supernatant was applied to Ni-NTA Superflow resin (Qiagen) equilibrated with buffer containing 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10% glycerol, 2 mM BME, and 0.1 mM PMSF. The resin was washed with 5 column volumes of the equilibrium buffer supplemented with 40 mM imidazole. The bound proteins were eluted with the equilibrium buffer supplemented with 200 mM imidazole. For proteolytic cleavage of the N-terminal His10-eYFP-SUMOstar tag, the sample was treated with SUMOstar protease (LifeSensors) at 4°C for 1 hr. The sample was then applied to Strep-Tactin Superflow (IBA Lifesciences) resin equilibrated with buffer containing 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM BME, and 0.1 mM PMSF. The resin was washed sequentially with 5 column volumes of the

equilibrium buffer containing 2 mM EDTA and no EDTA. The bound protein was eluted with buffer containing 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM BME, and 50 mM biotin. The eluate was applied onto Ni-NTA Superflow resin (Qiagen) to remove uncleaved fusion protein. The unbound fraction was concentrated to ~20 µM using an Amicon® Ultra-15 Centrifugal Filter Unit (100 kDa cutoff) and loaded onto Superose 6 Increase 5/150 GL (GE Healthcare) equilibrated with 50 mM Tris (pH 8.0), 100 mM NaCl, 0.5 mM TCEP. The fractions containing the protein were pooled, concentrated to ~5 µM, snap frozen in liquid nitrogen, and stored at −80°C until use. All human DICER constructs and *Drosophila* Dcr-1 were purified using the same purification method. Protein concentrations were quantified by UV absorbance at 280 nm (NanoDrop), using a molecular extinction coefficient calculated for each protein.

Massively parallel assay to identify DICER processing determinants

In order to predict the structure of dsRNA-bound human DICER, the cryo-EM structure of human DICER in the cleavage-incompetent state was superimposed with the crystal structure of Aqufiex aeolicus RNase III (Aa RNaseIII) bound to dsRNA (Gan et al., 2006) using PyMol. Based on the structural model, the dsRNA region likely to be recognized by the dsRBD was predicted (Figure 2.6). Synthetic pre-miRNAs (Integrated DNA Technologies), pre-let-7a-1 and pre-miR-374b, containing random bases within the specified window were denatured at 80°C for 5 min and annealed by slowly cooling the temperature down to 4°C at a -1°C min^{-1}, rate in a buffer containing 20 mM Tris (pH 7.5), 80 mM NaCl, 1 mM EDTA. Massively parallel assays were carried out by mixing randomized pre-miRNA pool with purified DICER at final concentrations of 0.1 µM and 0.2 µM, respectively, in a reaction buffer containing 50 mM Tris (pH 7.5), 100 mM NaCl, 0.5 mM TCEP, 2 mM $MgCl_2$. The reaction was carried out at 37°C for predetermined time periods to achieve approximately 10, 20, or 30% cleavage of the pre-miRNA pool. The reaction was stopped by adding an equivalent volume of 2X RNA Loading Dye (NEB) supplemented with additional 20 mM EDTA. Along

with the input RNA, the mixture was resolved on a 15% urea-polyacrylamide gel. The input RNA and the uncleaved pre-miRNA pool were gel-purified. The uncleaved pre-miRNAs were subsequently subjected to AQ-seq as previously described (Kim et al., 2019), with some modifications as follows. After 3´ adapter ligation, the ligated RNAs were purified on a 10% urea-polyacrylamide gel along with Century-Plus RNA Markers (Thermo Fisher) as a size marker. After reverse transcription, cDNAs were amplified with NEBNext Ultra II Q5 Master Mix (NEB). The libraries were sequenced on the NovaSeq 6000 platform.

Sequencing reads were first pre-processed by clipping the 3´ adapter using cutadapt (Martin, 2011). Next, 4-nt degenerate sequences were further removed with FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and short, low-quality, and artifact reads were filtered out using FASTX-Toolkit. Only reads without mutations outside the variable region were selected for subsequent analysis. Cleavage scores of individual pre-miRNA variants were calculated by calculating the proportion of a given variant with a pseudocount 1 and then dividing it in the input by that in the uncleaved substrate sample. The sequences of pre-miRNAs are listed in Table 2.1.

In vitro processing of RNA substrates

Pre-miRNA substrates were prepared by ligating two single-stranded RNA fragments (Integrated DNA Technologies) as previously described (Heo et al., 2009). For preparation of dsRNA substrates, singlestranded RNAs (Bioneer) of the 5´-arm were 5´-end-radiolabeled with γ-ATP by T4 polynucleotide kinase (Takara) and purified using Oligo Clean & Concentrator (Zymo Research) according to manufacturer's instructions. Next, they were annealed to the complementary synthetic ssRNAs of the 3´-arm in a buffer containing 20 mM Tris (pH 7.5), 80 mM NaCl, 1 mM EDTA, by heating at 80°C for 5 min and slowly cooling the temperature down to 4°C at a -1°C min^{-1}, rate. *In vitro* processing assays were carried out by mixing the RNA substrates with the 0.5 µM purified proteins in a reaction buffer containing 50 mM Tris (pH 7.5), 100 mM NaCl, 0.5 mM TCEP, 2 mM $MgCl_2$, 5 µg ml^{-1} yeast RNA. The reaction was carried out at 37°C

and stopped by adding an equivalent volume of 2X RNA Loading Dye (NEB) supplemented with 0.5 mg ml^{-1} proteinase K (Roche). The mixture was resolved on a 15% urea-polyacrylamide gel along with 5´-end-radiolabeled synthetic mature miRNAs and Decade Markers System (Ambion) as size markers and visualized by the phosphorimager, Typhoon FLA 7000 (GE Healthcare). The sequences of synthetic oligos are listed in Table 2.1.

DICER rescue experiment and data analysis

DICER KO HCT116 and HEK293T cells were maintained in DMEM and McCoy's 5A media (WELGENE), respectively, supplemented with 10% fetal bovine serum (WELGENE) (Kim et al., 2016; Bogerd et al., 2014). Both cell lines were authenticated using the ATCC short tandem repeat profiling. DICER KO HCT116 and HEK293T cells were transiently transfected with DICER expression plasmids using FuGENE HD (Promega) or Lipofectamine 3000 (Thermo Fisher), respectively. Total RNAs were extracted using TRIzol (Thermo Fisher) 48 hrs post-transfection and subjected to AQ-seq (Kim et al., 2019) except that 100 µg of total RNAs were used for library construction, and that cDNAs were amplified with NEBNext Ultra II Q5 Master Mix (NEB). The libraries were sequenced on the NovaSeq 6000 platform.

Data processing was performed as previously described (Kim et al., 2019). Briefly, the 3´ adapter and 4-nt degenerate sequences were removed from the reads using cutadapt (Martin, 2011) and FASTX-Toolkit (http://hannonlab.cshl.edu/fastxtoolkit/), respectively. Next, short, low-quality, and artifact reads were filtered out using FASTX-Toolkit. The output reads were first mapped to the spike-in sequences and then unmapped reads were aligned to the human genome (hg38) using BWA (Li & Durbin, 2009a). Corresponding miRNA annotations were found with miRBase release 21 using the intersect tool in BEDTools (Kozomara & Griffiths-Jones, 2014a; Quinlan & Hall, 2010). For the analysis of miRNA abundance, read counts of miRNAs were normalized with the average of read counts of spike-ins. For the analysis of processing accuracy, we first identified the most abundant 5´-isomiR for a given mature miRNA in WT DICER-expressing

samples. Then the proportions of the 5´-isomiR were measured for all of the samples. Fold changes of the proportions between samples were calculated to examine processing accuracy defects. In order to reduce the influence of decay intermediates, 5´-isomiRs with 5´ ends within a range of −5 to 5 positions relative to the annotated one and miRNAs with a position where more than 20% of reads account for were used for subsequent analyses.

Analysis on the GYM motif of natural miRNAs

Representative animal miRNAs from diverse species have been used in the primary miRNA motif conservation analysis (Fang & Bartel, 2015; Fromm et al., 2020). We adopted the curated lists (Fang & Bartel, 2015; Fromm et al., 2020) for our analysis and further curated as follows: 1) miRNA annotations and sequences were manually curated based on miRBase release 22 (Kozomara & Griffiths-Jones, 2014a). 2) Pre-miRNA sequences were retrieved based on their 5p and 3p annotations in miRBase release 22 and added to the lists. 3) Pre-miRNA sequences from miRGeneDB release 2.0 were added to the lists (Auyeung et al., 2013). 4) 3p sequences from miRBase release 22 and miRGeneDB release 2.0 were added to the lists. 5) Starting positions of 3p miRNAs were identified for each miRNA if 3p sequences were available. If not, 3p sequences and starting positions were manually curated based on its orthologs in close species.

Pre-miRNA secondary structures were obtained using Fold in RNAstructure version 6.3 with default settings (Bellaousov et al., 2013). Based on the predicted structure for a given miRNA, 3-bp composition of each position relative to the starting position of 3p was identified (Fig. 3g). Next, its corresponding cleavage score was obtained by calculating average GYM scores of the 3-bp measured in the massively parallel assay of two pre-miRNA backbones with −1-to-1 randomization performed under the condition where approximately 20% of substrates were cleaved. Pre-miRNA sequences were referred to those of miRGeneDB, miRBase, and to those listed by Fang and Bartel, in order of priority. 3p positions were referred to those of miRGeneDB, miRBase, and our manual curation, in order of

priority. If there was a bulge or mismatch for a given position, the position was excluded. If there was no available 3p position, the miRNA was excluded.

Dual luciferase reporter assay for gene silencing activity

For shRNA, HEK293T cells were co-transfected with pmirGLO vector containing a target site of shRNA-3p along with a plasmid encoding both shRNA and pri-miR-1-1 under the control of independent promoters using Lipofectamine 3000 reagent (Thermo Fisher). shRNAs were optimally designed following the published guidelines (Bofill-De Ros & Gu, 2016). For DsiRNA, HEK293T cells were cotransfected with pmirGLO containing a target site of DsiRNA along with DsiRNA (Integrated DNA Technologies) using Lipofectamine 2000 reagent (Thermo Fisher). Cells were harvested at different time points post-transfection and the luciferase signals were measured by DualLuciferase Reporter Assay System according to the manufacturer's instructions (Promega) on Spark Multimode Microplate Reader (TECAN). The sequences of DsiRNAs are listed in Table 2.1.

Quantitative real-time PCR (RT-qPCR)

HEK293T cells were transfected as described above. RNAs were isolated at different time points post-transfection using TRIzol (Thermo Fisher) and Direct-zol RNA Miniprep Kit (Zymo Research) according to the manufacturer's instructions. cDNAs were synthesized using the TaqMan miRNA Reverse Transcription Kit (Applied Biosystems) and subjected to quantitative real-time PCR with the TaqMan MicroRNA Assay (Applied Biosystems) on StepOnePlus RealTime PCR System (Thermo Fisher). Given that miR-1-3p is little expressed in HEK293T (~8 RPM in the AQ-seq result) (Kim et al., 2019), miR-1-3p was used as an internal control.

Description	Sequence
hsa-pre-mir-324 variant (5' half)	CGCAUCCCCUAGGGCAUUGGGUAAAGCUG
hsa-pre-mir-324 variant (stem mismatch) (3' half) + U	GAGACCCACUGCCCCAGGUGCUGCUGGU
hsa-pre-mir-324 variant (stem base-pair) (3' half) + U	GAGACCCA**A**UGCCCCAGGUGCUGCUGGU
hsa-pre-let-7a-1 (5-bp randomized) (positions −1-to-3)	UGAGGUAGUAGGUUGU**NNNNN**UUUAGGGUCACACCCACCACUGGGAGAUA**NNNNN**ACAAUCUACUGUCUUUC
hsa-pre-let-7a-1 (3-bp randomized) (positions −1-to-1)	UGAGGUAGUAGGUUGUAU**NNN**UUUAGGGUCACACCCACCACUGGGAGAUA**NNN**AUACAAUCUACUGUCUUUC
hsa-pre-let-7a-1 (3-bp randomized) (positions 1-to-3)	UGAGGUAGUAGGUUGU**NNN**GUUUUAGGGUCACACCCACCACUGGGAGAUAAC**NNN**ACAAUCUACUGUCUUUC
hsa-pre-miR-374b (3-bp randomized) (positions −1-to-1)	AUAUAAUACAACCUGCUA**NNN**GUCCUAGC**NNN**UAGCAGGUUGUAUUAUCAUU
hsa-pre-let-7a-1 (5! half)	UGAGGUAGUAGGUUGUAUAGUUUUAGGGUCACACCCACC
hsa-pre-let-7a-1 (3! half)	ACUGGGAGAUAACUAUACAAUCUACUGUCUUUC
hsa-pre-let-7a-1 variant (5! half) (GCm)	UGAGGUAGUAGGUUGUAU**CGC**UUUAGGGUCACACCCACC
hsa-pre-let-7a-1 variant (5! half) (GCp)	UGAGGUAGUAGGUUGUAU**GGC**UUUAGGGUCACACCCACC
hsa-pre-let-7a-1 variant (3! half) (GCm or GCp)	ACUGGGAGAUA**GCC**AUACAAUCUACUGUCUUUC
hsa-pre-let-7a-1 variant (5! half) (UAm)	UGAGGUAGUAGGUUGUAU**CUA**UUUAGGGUCACACCCACC
hsa-pre-let-7a-1 variant (5! half) (UAp)	UGAGGUAGUAGGUUGUAU**GUA**UUUAGGGUCACACCCACC
hsa-pre-let-7a-1 variant (3! half) (GCm or GCp)	ACUGGGAGAUA**UAC**AUACAAUCUACUGUCUUUC
Duplex RNA (5! half) (GCm)	CGAGGUAGUAGGUUGUAACGCGAAAGGACAAAGAG
Duplex RNA (3! half) (GCm) (terminal mismatch) (1-nt overhang)	CUCUUUGUCCAAACGCCUUACAACCUACUACCUCAU
Duplex RNA (3! half) (GCm) (terminal mismatch) (2-nt overhang)	CUCUUUGUCCAAACGCCUUACAACCUACUACCUCAUU
Duplex RNA (3! half) (GCm) (terminal mismatch) (3-nt overhang)	CUCUUUGUCCAAACGCCUUACAACCUACUACCUCAUUU
Duplex RNA (3! half) (GCm) (terminal base-pair) (1-nt overhang)	CUCUUUGUCCAAACGCCUUACAACCUACUACCUCGU
Duplex RNA (3! half) (GCm) (terminal base-pair) (2-nt overhang)	CUCUUUGUCCAAACGCCUUACAACCUACUACCUCGUU
Duplex RNA (3! half) (GCm) (terminal base-pair) (3-nt overhang)	CUCUUUGUCCAAACGCCUUACAACCUACUACCUCGUUU
5! half (UAp)	CGAGGUAGUAGGUUGUAAGUAGAAAGGACAAAGAG
3! half (UAp) terminal mismatch (1-nt overhang)	CUCUUUGUCCAAACUACUUACAACCUACUACCUCAU
3! half (UAp) terminal mismatch (2-nt overhang)	CUCUUUGUCCAAACUACUUACAACCUACUACCUCAUU
3! half (UAp) terminal mismatch (3-nt overhang)	CUCUUUGUCCAAACUACUUACAACCUACUACCUCAUUU
3! half (UAp) terminal base-pair (1-nt overhang)	CUCUUUGUCCAAACUACUUACAACCUACUACCUCGU
3! half (UAp) terminal base-pair (2-nt overhang)	CUCUUUGUCCAAACUACUUACAACCUACUACCUCGUU
3! half (UAp) terminal base-pair (3-nt overhang)	CUCUUUGUCCAAACUACUUACAACCUACUACCUCGUUU
Duplex RNA (5! half) (GCm) (1-bp shifted)	CGAGGUAGUAGGUUUAACGCAGAAAGGACAAAGAG
Duplex RNA (5! half) (GCp) (1-bp shifted)	CGAGGUAGUAGGUUUAAGGCAGAAAGGACAAAGAG
Duplex RNA (3! half) (GCm or GCp) (1-bp shifted)	CUCUUUGUCCAAACU**GCC**UUAAACCUACUACCUCAUU
Duplex RNA (5! half) (UAm) (1-bp shifted)	CGAGGUAGUAGGUUUAA**CUA**AGAAAGGACAAAGAG
Duplex RNA (3! half) (UAm) (1-bp shifted)	CUCUUUGUCCAAACU**UAC**UUAAACCUACUACCUCAUU
Duplex RNA (5! half) (GAm) (1-bp shifted)	CGAGGUAGUAGGUUUAA**CUC**AGAAAGGACAAAGAG
Duplex RNA (3! half) (GAm) (1-bp shifted)	CUCUUUGUCCAAACU**GAC**UUAAACCUACUACCUCAUU
Duplex RNA (5! half) (UCp) (1-bp shifted)	CGAGGUAGUAGGUUUAA**GGA**AGAAAGGACAAAGAG
Duplex RNA (3! half) (UCp) (1-bp shifted)	CUCUUUGUCCAAACU**UCC**UUAAACCUACUACCUCAUU
hsa-pre-mir-27b (5! half)	AGAGCUUAGCUGAUUGGUGAACAGUGAUUGGU
hsa-pre-mir-27b (3! half)	UUCCGCUUUGUUCACAGUGGCUAAGUUCUGC
hsa-pre-mir-27b variant 1 (5! half)	AGAGCUUAGCUGAUUGGUG**G**ACAGUGAUUGGU
hsa-pre-mir-27b variant 1 (3! half)	UUCCGCUUUGU**C**CACAGUGGCUAAGUUCUGC
hsa-pre-mir-27b variant 2 (5! half)	AGAGCUUAGCUGAUUGGUG**GUG**AGUGAUUGGU
hsa-pre-mir-27b variant 2 (3! half)	UUCCGCUUU**CA**UCACAGUGGCUAAGUUCUGC
hsa-pre-mir-34a (5! half)	UGGCAGUGUCUUAGCUGGUUGUUGUGAGCAA

Table 2.1 Oligonucleotides used in this study.

hsa-pre-mir-34a (3! half)	UAGUAAGGAAGCAAUCAGCAAGUAUACUGCCCUA
hsa-pre-mir-34a variant (5! half)	UGGCAGUGUCUUAGCUGGU**GA**UUGUGAGCAA
hsa-pre-mir-34a variant (3! half)	UAGUAAGGAAG**UC**AUCAGCAAGUAUACUGCCCUA
hsa-pre-mir-21 (5! half)	UAGCUUAUCAGACUGAUGUUGACUGUUGAA
hsa-pre-mir-21 (3! half)	UCUCAUGGCAACACCAGUCGAUGGGCUGUC
hsa-pre-mir-21 variant (5! half)	UAGCUUAUCAGACUGAUGU**G**GACUGUUGAA
hsa-pre-mir-21 variant (3! half)	UCUCAUGGC**C**ACACCAGUCGAUGGGCUGUC
hsa-pre-let-7e (5! half)	UGAGGUAGGAGGUUGUAUAGUUGAGGAGGAC
hsa-pre-let-7e (3! half)	ACCCAAGGAGAUCACUAUACGGCCUCCUAGCUUUCC
hsa-pre-let-7e variant (5! half)	UGAGGUAGGAGGUUGUAU**GUA**UGAGGAGGAC
hsa-pre-let-7e variant (3! half)	ACCCAAGGAGAUC**UAU**AUACGGCCUCCUAGCUUUCC
hsa-pre-let-7d (5! half)	AGAGGUAGUAGGUUGCAUAGUUUUAGGGCAGGGAUUU
hsa-pre-let-7d (3! half)	UGCCCACAAGGAGGUAACUAUACGACCUGCUGCCUUUCU
hsa-pre-let-7d variant (5! half)	AGAGGUAGUAGGUUGCAU**GU**UUUUAGGGCAGGGAUUU
hsa-pre-let-7d variant (3! half)	UGCCCACAAGGAGGUAA**A**UAUACGACCUGCUGCCUUUCU
hsa-pre-let-7d variant (5! half)	AGAGGUAGUAGGUUGCAU**GUA**UUUAGGGCAGGGAUUU
hsa-pre-let-7d variant (3! half)	UGCCCACAAGGAGGUA**UA**UAUACGACCUGCUGCCUUUCU
hsa-pre-let-7f-1 (5! half)	UGAGGUAGUAGAUUGUAUAGUUGUGGGGUAGUGA
hsa-pre-let-7f-1 (3! half) + U	UUUUACCCUGUUCAGGAGAUAACUAUACAAUCUAUUGCCUUCCU
hsa-pre-let-7f-1 variant (5! half)	UGAGGUAGUAGAUUGUAU**GU**UUGUGGGGUAGUGA
hsa-pre-let-7f-1 variant (3! half) + U	UUUUACCCUGUUCAGGAGAUAA**A**UAUACAAUCUAUUGCCUUCCU
hsa-pre-let-7f-1 variant (5! half)	UGAGGUAGUAGAUUGUAU**GUA**UGUGGGGUAGUGA
hsa-pre-let-7f-1 variant (3! half) + U	UUUUACCCUGUUCAGGAGAUA**UA**UAUACAAUCUAUUGCCUUCCU
hsa-pre-let-7i-1 (5! half)	UGAGGUAGUAGUUUGUGCUGUUGGUCGGGUUGUGACAU
hsa-pre-let-7i-1 (3! half) + U	UGCCCGCUGUGGAGAUAACUGCGCAAGCUACUGCCUUGCU
hsa-pre-let-7i-1 variant (5! half)	UGAGGUAGUAGUUUGUGC**GU**UUGGUCGGGUUGUGACAU
hsa-pre-let-7i-1 variant (3! half) + U	UGCCCGCUGUGGAGAUAA**A**UGCGCAAGCUACUGCCUUGCU
hsa-pre-let-7i-1 variant (5! half)	UGAGGUAGUAGUUUGUGC**GUA**UGGUCGGGUUGUGACAU
hsa-pre-let-7i-1 variant (3! half) + U	UGCCCGCUGUGGAGAUA**UA**UGCGCAAGCUACUGCCUUGCU
hsa-pre-mir-9-1 (5! half)	UCUUUGGUUAUCUAGCUGUAUGAGUGGUGUGGAGU
hsa-pre-mir-9-1 (3! half)	CUUCAUAAAGCUAGAUAACCGAAAGUA
hsa-pre-mir-9-1 variant (5! half)	UCUUUGGUUAUCUAGCUG**G**AUGAGUGGUGUGGAGU
hsa-pre-mir-9-1 variant (3! half)	CUUCAU**U**AAGCUAGAUAACCGAAAGUA
DsiRNA 5p strand (dominant strand: 5p) (GCm)	UCACGCUGAACUUGUGGC**CGC**UCTA
DsiRNA 3p strand (dominant strand: 5p) (GCm)	UAGA**GCC**GCCACAAGUUCAGCGUCCAA
DsiRNA 5p strand (dominant strand: 5p) (GCp)	UCACGCUGAACUUGUGGC**GGC**UCTA
DsiRNA 3p strand (dominant strand: 5p) (GCp)	UAGA**GCC**GCCACAAGUUCAGCGUCCAA
DsiRNA 5p strand (dominant strand: 5p) (CCp)	UCACGCUGAACUUGUGGC**GGG**UCTA
DsiRNA 3p strand (dominant strand: 5p) (CCp)	UAGA**CCC**GCCACAAGUUCAGCGUCCAA
DsiRNA 5p strand (dominant strand: 3p) (GUm)	GGAAUACAAGUUGACGUA**CAC**AGTG
DsiRNA 3p strand (dominant strand: 3p) (GUm)	CACU**GUC**UACGUCAACUUGUAUUCCUU
DsiRNA 5p strand (dominant strand: 3p) (GUp)	GGAAUACAAGUUGACGUA**GAC**AGTG
DsiRNA 3p strand (dominant strand: 3p) (GUp)	CACU**GUC**UACGUCAACUUGUAUUCCUU
DsiRNA 5p strand (dominant strand: 3p) (CUp)	GGAAUACAAGUUGACGUA**GAG**AGTG
DsiRNA 3p strand (dominant strand: 3p) (CUp)	CACU**CUC**UACGUCAACUUGUAUUCCUU

Table 2.2 Oligonucleotides used in this study.

3. Structure of the human DICER-pre-miRNA complex in a dicing state

3.1 Background

Small regulatory RNAs serve as guide molecules in RNA interference (RNAi) by inducing translational repression and destabilization of the cognate mRNAs (Hutvagner & Zamore, 2002; Mourelatos et al., 2002; MacRae et al., 2007). Central to the pathway is the ribonuclease (RNase) III enzyme Dicer that cleaves long double-stranded RNAs (dsRNAs) or short hairpin RNAs to generate small RNAs in length of 21-25 nucleotides (nt) (Elbashir et al., 2001; Hutvagner et al., 2001) Dicer homologs are found throughout eukaryotes and show substantial diversity in their substrate specificity and action mechanism. Some Dicer proteins are specific to long dsRNAs as observed in the structural and biochemical studies on Giardia Dicer, plant Dicer-like proteins, and fly Dicer2 (Macrae et al., 2006a; MacRae et al., 2007; Xie et al., 2004; Mosher et al., 2008; Ronemus et al., 2006; Vazquez et al., 2004; Zhang et al., 2007; Wang et al., 2021; Sinha et al., 2018; Su et al., 2022; Welker et al., 2011; Cenik et al., 2011; Lee et al., 2004). In contrast, other homologs such as fly Dicer-1 are highly selective to hairpin-shaped pre-miRNAs (Lee et al., 2004; Tsutsumi et al., 2011). Human DICER (hDICER) can cleave both types of substrates, with a clear preference for short hairpins over long dsRNAs (Ma et al., 2008; Chakravarthy et al., 2010)

hDICER recognizes several features of its substrates: a dsRNA stem with approximately 22 bp, a 2-nt 3´ overhang, and a flexible loop next to the cleavage site (Macrae et al., 2006a,b; Park et al., 2011; Zhang et al., 2002; Gu et al., 2012; Feng et al., 2012). The flexible loop is known to be sensed by the helicase domain19,20, while the 5´ phosphorylated end and the 3´ overhang are recognized

by basic pockets in the platform and PAZ domains, respectively (Park et al., 2011; Zhang et al., 2002). By anchoring the termini, DICER can act as a 'molecular ruler' to measure ~22 nt away from the 5´ end ('5´ counting rule') and 3´ end ('3´ counting rule') of the substrate (Macrae et al., 2006a,b; Park et al., 2011; Zhang et al., 2002). In addition, a newly identified 'GYM' motif at the cleavage site enables precise determination of the cleavage site. However, the structural basis of the substrate specificity of hDICER remains largely unknown due to the lack of an active state structure.

Early electron microscopic (EM) analyses of hDICER revealed its overall L-like shape (Lau et al., 2009; Wang et al., 2009; Lau et al., 2012). Crystal structures of a partial fragment containing the platform-PAZ domain showed the 5´ and 3´ pockets which recognize the respective ends of a small RNA duplex (Tian et al., 2014). A more recent cryo-electron microscopy (cryo-EM) study displayed the overall topology of full-length hDICER in apo and RNA-bound states (Liu et al., 2018). However, in this structure, the pre-miRNA is situated distant from the catalytic valley, likely representing a 'pre-dicing' state. Thus, we still lack the structural understanding of how hDICER recognizes pre-miRNAs in an active state.

Here, we aimed to determine the cryo-EM structure of hDICER with pre-miRNA in a cleavage-competent state. The structure reveals dynamic spatial rearrangements of multiple domains of hDICER during the transition to a catalytic state and explains how hDICER selects its substrates with specificity.

3.2 Results

3.2.1 Structural determination

To reconstitute the enzyme-substrate complex, we employed mono-uridylated pre-let-7a-1, which has an optimal 2-nt 3´ overhang, stem length and terminal loop. Meanwhile, in our concurrent study, we carried out massively parallel assays using over a million pre-miRNAs with random sequences near the cleavage

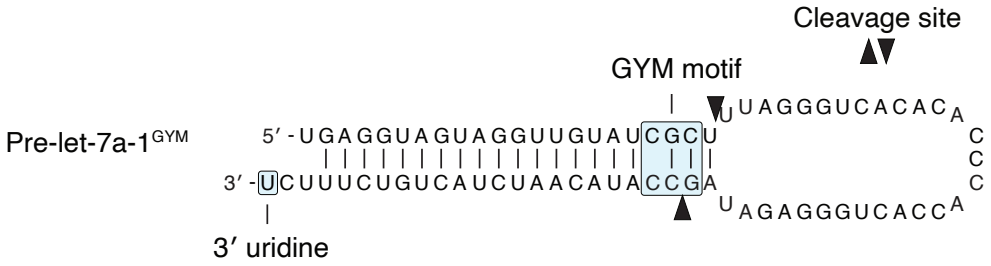

Figure 3.1 The sequence of pre-let-7a-1^{GYM} used for structural determination.

sites. Analysis of the top variants revealed that nucleotide sequences at positions −1, 0, and 1 relative to the 3p DICER cleavage site are enriched with a paired guanine (G), a paired pyrimidine (Y), and a mismatched cytosine or adenine (M), respectively. We therefore termed this motif the 'GYM' motif. The GYM motif robustly enhances dsRNA processing at a specific site, suggesting its role in the catalytic step. The mismatched 'M' is particularly important for cleavage site decision. In order to design an optimal substrate for structural determination as well as to obtain the mechanistic insights into the GYM motif-mediated processing, we incorporated the highest-scoring GYM motif (5′-CGC/GCC-3′) into pre-let-7a-1 (Figure 3.1), which substantially increased the processing rate compared to the wild-type sequence.

We overexpressed and purified the full-length hDICER protein, which cleaved the pre-let- 7a-1^{GYM} precisely, yielding the expected 22-nt fragment (Figure 3.2, 3.3). Previous cryo-EM structures of hDICER were determined with its accessory protein, TRBP, which exhibits flexible and heterogenous conformations (Liu et al., 2018). Since the vast majority of human miRNAs including let-7 do not require TRBP both *in vitro* and in cells (Macrae et al., 2006b; Kim et al., 2014; Wilson et al., 2015), we set out to determine the cryo-EM structure of hDICER alone so as to reduce the structural heterogeneity.

We first validated the architecture of the purified apo-hDICER at 4.0 Å resolution by cryo-EM (Figure 3.4–3.7, Table 3.1).[1] This structure exhibits a

[1]Cryo-EM experiments and analyses were performed by Dr. Hansol Lee.

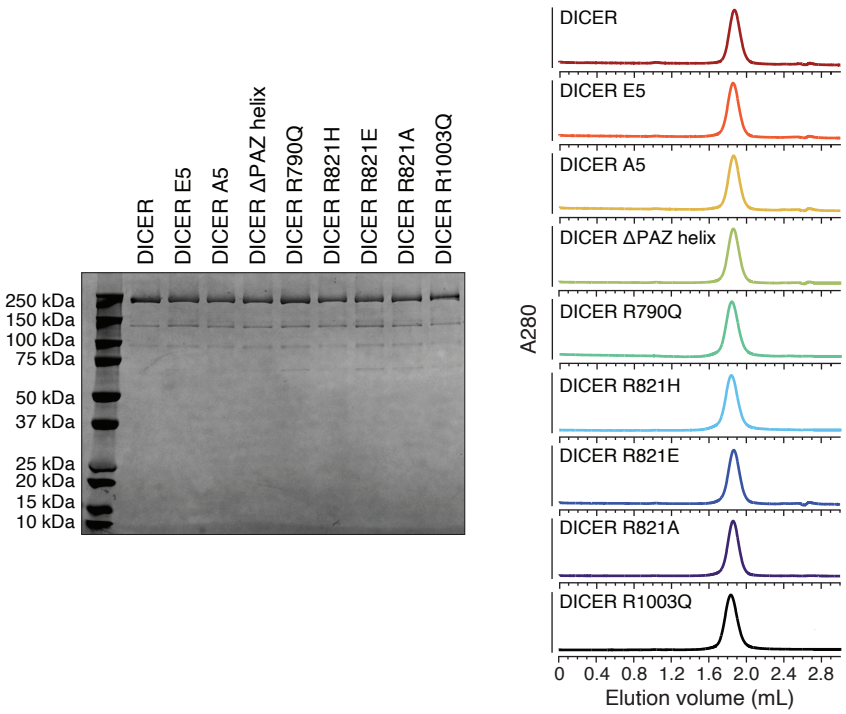

Figure 3.2 SDS-PAGE and size-exclusion chromatography of purified proteins.

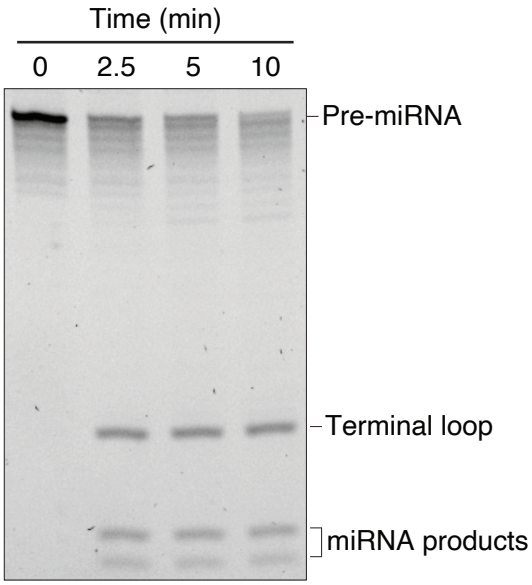

Figure 3.3 In vitro processing of pre-let-7a-1 with human DICER.

compact 'hatchet-like' (or L-shaped) architecture (Lau et al., 2009; Wang et al., 2009; Lau et al., 2012) with mostly globular domains, as observed in the previous cryo-EM structure of the hDICER-TRBP complex (Liu et al., 2018) Our map clearly displays overall domains with defined helical features (Figure 3.8, 3.9). Notably, the helicase domain shows lower local resolution suggesting its intrinsically flexible behavior (Figure 3.7). Next, to build a model of apo-hDICER, we docked the cryo-EM model of hDICER from hDICER-TRBP into our map and refined it by flexible fitting (Liu et al., 2018). Our model covers the 3D position of most folded domains and partial interdomain loops including DUF283 linker (D-linker) that connects DUF283 to platform (Figure 3.8, 3.9). Notably, D-linker which was mostly undetermined in the previous hDICER-TRBP structure forms a short beta sheet with the N-terminal region of RIIIDa.

Next, we reconstituted the hDICER-pre-miRNA complex (Figure 3.10, 3.11, Table 3.1) and were able to reconstruct the map of the complex at 3.0 Å resolution

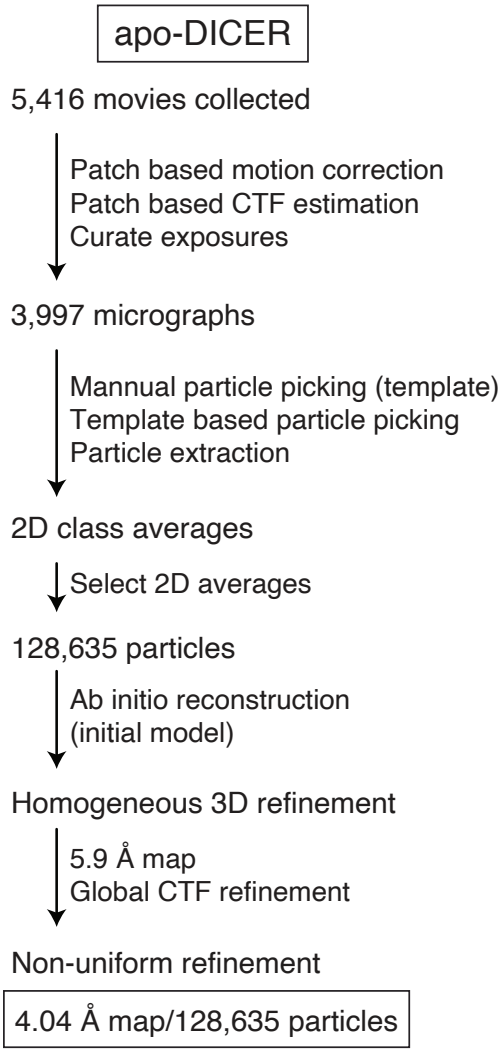

Figure 3.4 Overview of the image processing procedure for structural determination of apo-DICER.

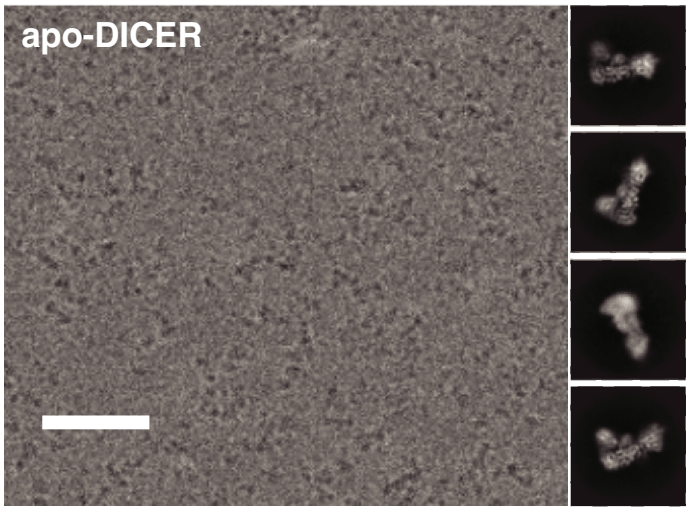

Figure 3.5 Representative micrograph and 2D class averages of the apo-DICER. Scale bar represents 50 nm.

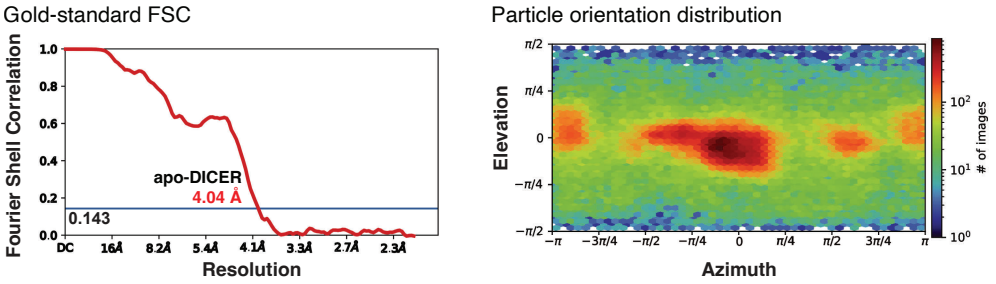

Figure 3.6 Gold standard Fourier Shell Correlation (FSC) and angular particle distribution heatmaps of the apo-DICER.

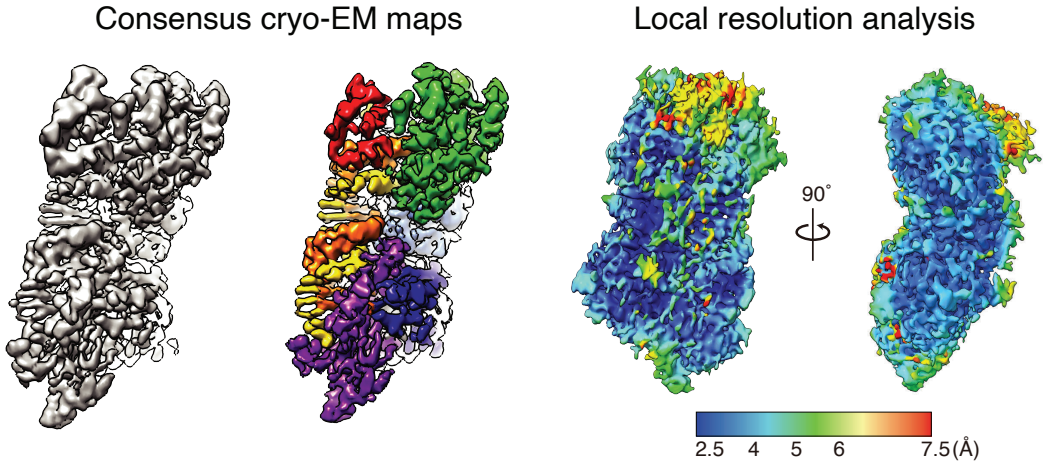

Figure 3.7 Consensus maps and local resolution analysis of apo-DICER. Each domain is indicated in different colors.

using a cryo-EM dataset of ~1,210K particle images.[2] In this map, hDICER embraces the helix of pre-miRNA within the catalytic center of the enzyme through an extensive surface contact (Figure 3.12–3.14). Our map displays the side chain details of the platform, PAZ, RIIIDa/b domains and the secondary structural features for the dsRBD and interdomain linkers while we could not identify the featured densities from helicase (residues 1–564), DUF283 (residues 590–714) and several loops in RIIID domains (residues 1392–1545 and 1595–1684) (Figure 3.15, 3.16). We referenced the model of apo-hDICER and pre-let-7a-1 structures and built an atomic model of hDICER-pre-let-7a-1^{GYM} (Figure 3.8, 3.9).

3.2.2 Overall structure of hDICER in a dicing state

The structure of the hDICER-pre-let-7a-1^{GYM} complex shows that the pre-miRNA is fully docked and poised for cleavage within the catalytic center formed by intramolecular dimerization of RIIIDa and RIIIDb (Figure 3.17). The catalytic center has two clusters of conserved acidic amino acid residues in the RIIIDa

[2]Cryo-EM experiments and analyses were performed by Dr. Hansol Lee.

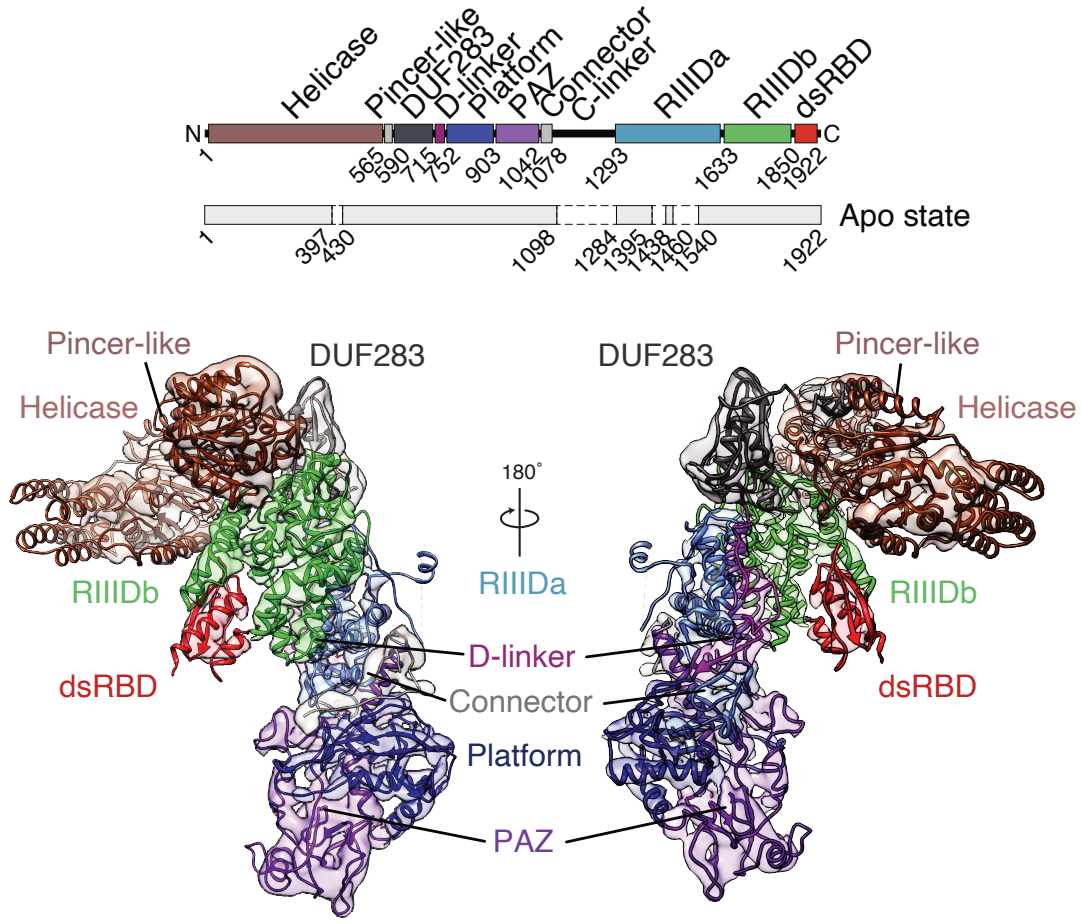

Figure 3.8 Domain organization of hDICER. Schematics for the apo state shows amino acid residues included in the model (solid lines) and not included in the model (dashed lines).

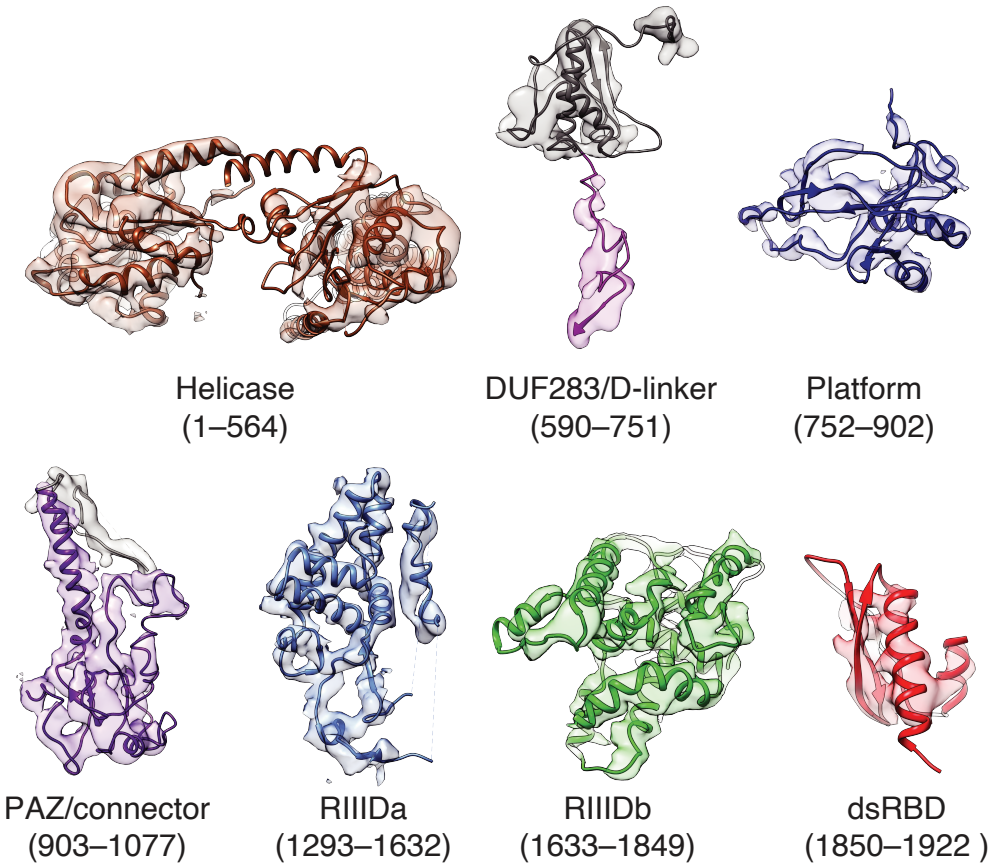

Figure 3.9 3D maps of individual domains of the apo-DICER.

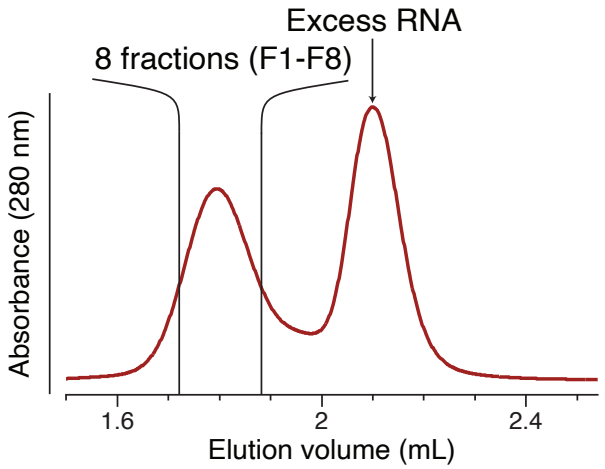

Figure 3.10 Size-exclusion chromatography of the DICER-RNA complex.

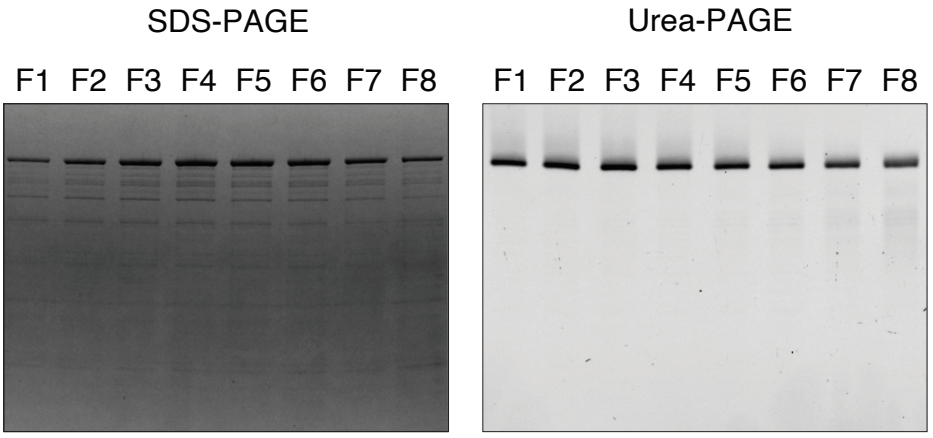

Figure 3.11 SDS-PAGE and urea-PAGE of the DICER-RNA complex.

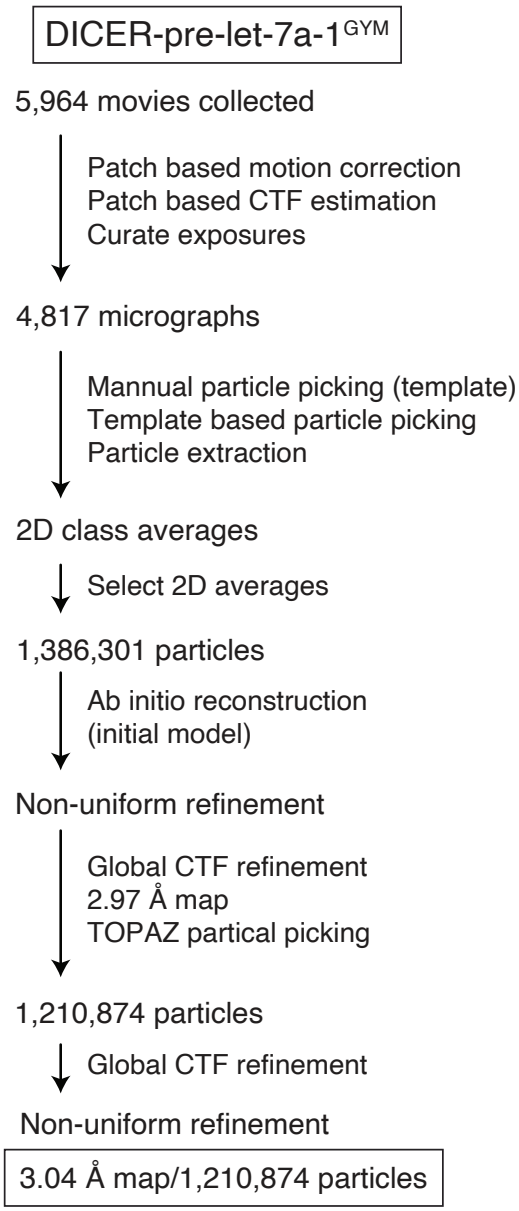

Figure 3.12 Overview of the image processing procedure for structural determination of DICER in a dicing state.

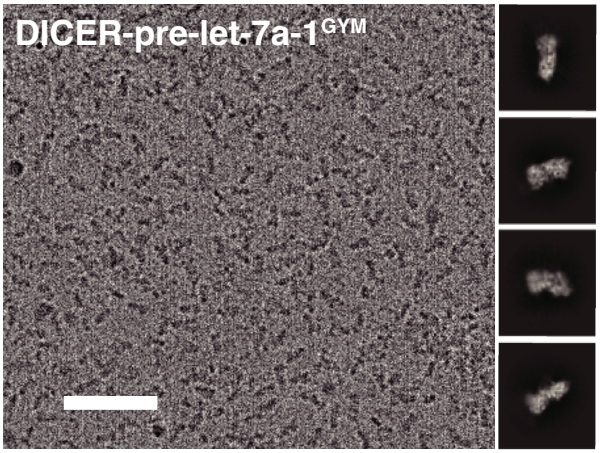

Figure 3.13 Representative micrograph and 2D class averages of the DICER-pre-let-7a-1^{GYM}. Scale bar represents 50 nm.

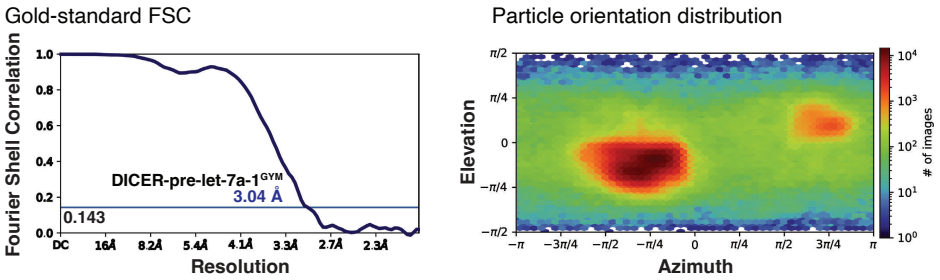

Figure 3.14 Gold standard Fourier Shell Correlation (FSC) and angular particle distribution heatmaps of the DICER-pre-let-7a-1^{GYM}.

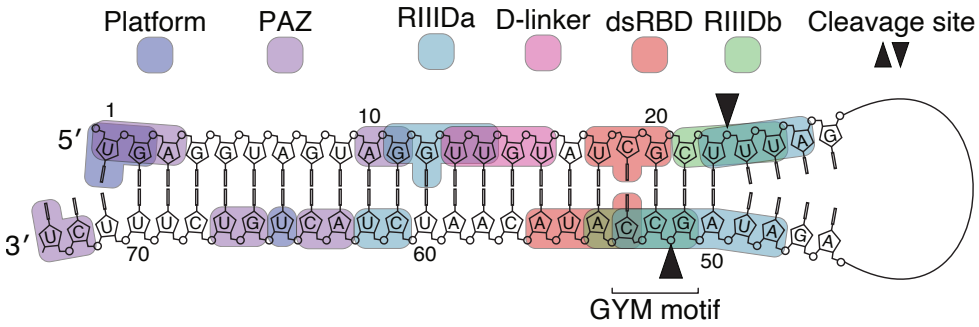

Figure 3.15 Protein-RNA interactions in the dicing state at the domain level. Sequences of pre-let-7a-1^{GYM} that are not included in the model are not shown.

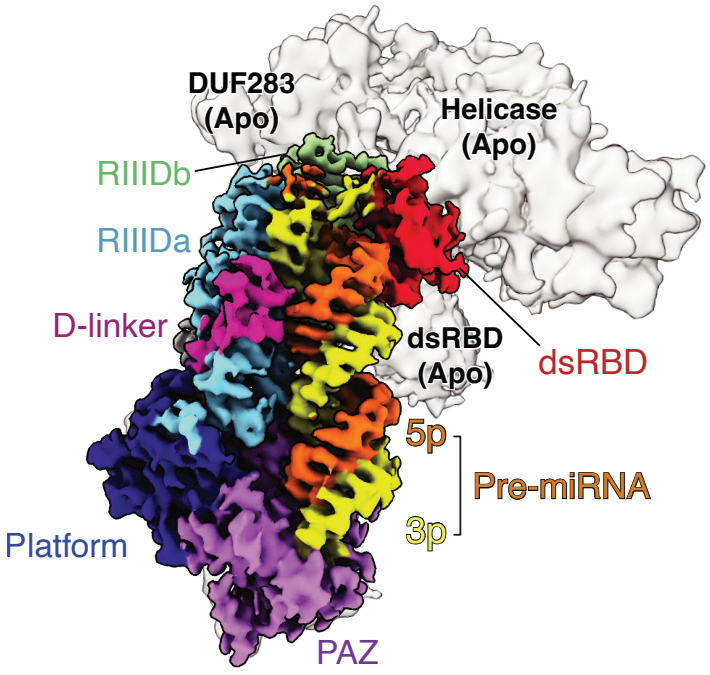

Figure 3.16 Cryo-EM map of hDICER in a dicing state (color, 3.0 Å) overlaying that of hDICER in an apo state (grey, 4.0 Å).

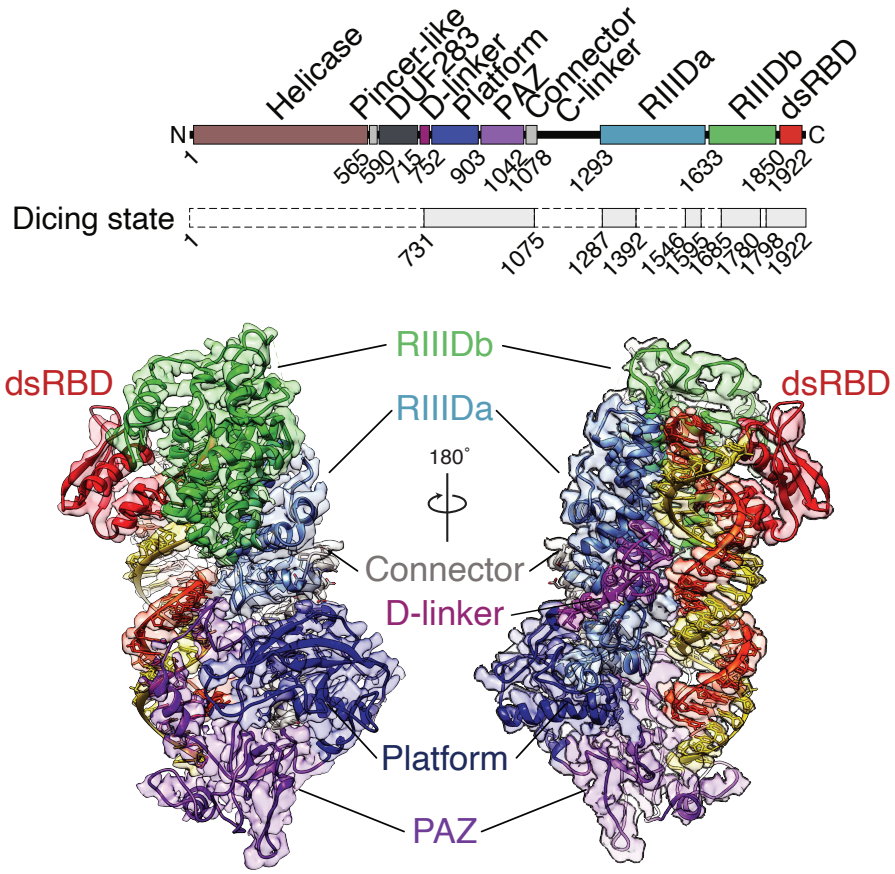

Figure 3.17 Overall structure of the hDICER with a pre-miRNA in a cleavage-competent state. Black arrowheads point to the DICER cleavage sites.

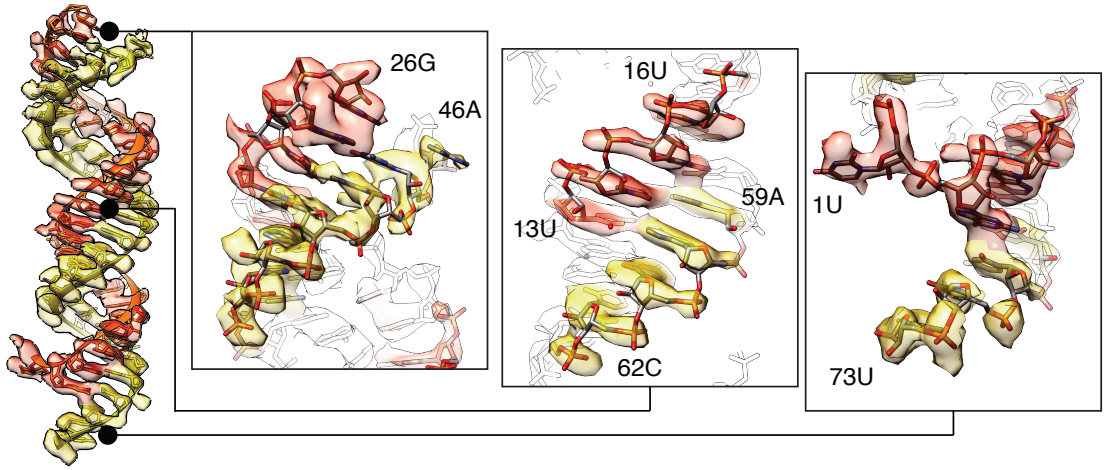

Figure 3.18 3D maps of the pre-let-7a-1^{GYM}.

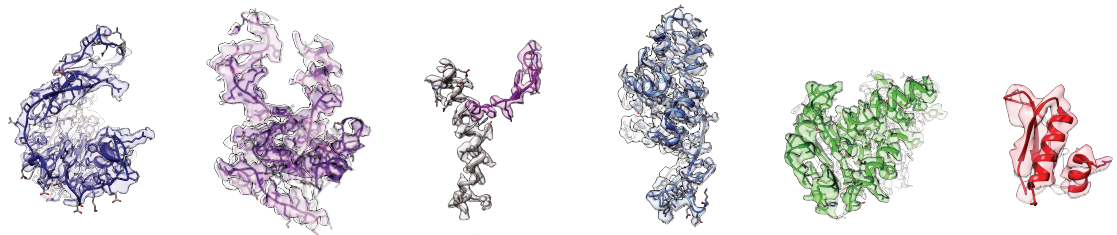

Figure 3.19 3D maps of the individual domains of DICER-pre-let-7a-1^{GYM}.

(E1316, D1320, D1561, E1564) and RIIIDb (E1705, D1709, D1810, E1813) (Figure 3.18–3.21). Our map also displays extra densities in the catalytic core, which are attributed to calcium ions used to substitute magnesium ions to prevent hydrolysis (Provost et al., 2002). The calcium ions are situated near the oxygen atoms of the scissile phosphodiester bonds in the 5p strand (between U22 and U23) and 3p strand (between 51G and 52C) (Figure 3.22, 3.23), whose position coincides with the actual cleavage sites of pre-let-7a-1. This spatial arrangement is highly homologous to that of other RNase III type enzymes including human DROSHA and *Aquifex aeolicus* RNase III (Aa RNase III) (Figure 3.24) (Partin et al., 2020; Jin et al., 2020; Gan et al., 2006).

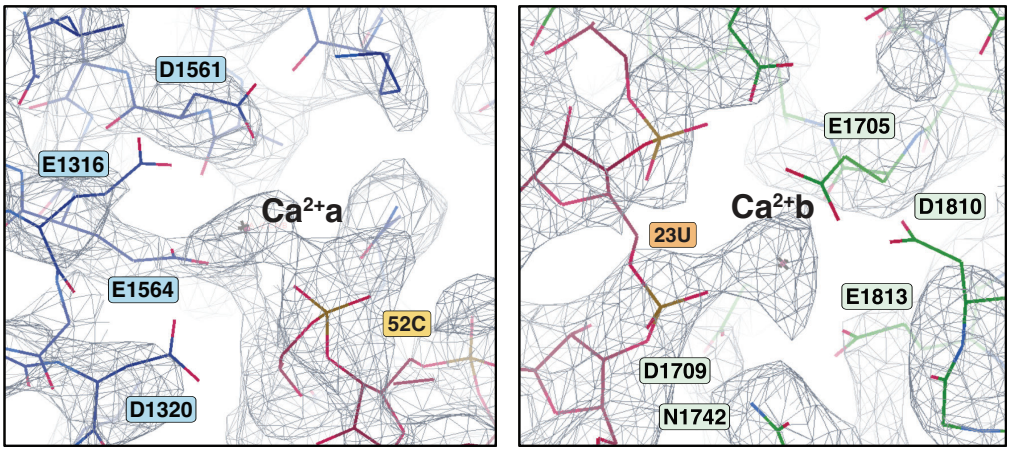

Figure 3.20 Cryo-EM map of the catalytic site created by RIIIDa and RIIIDb.

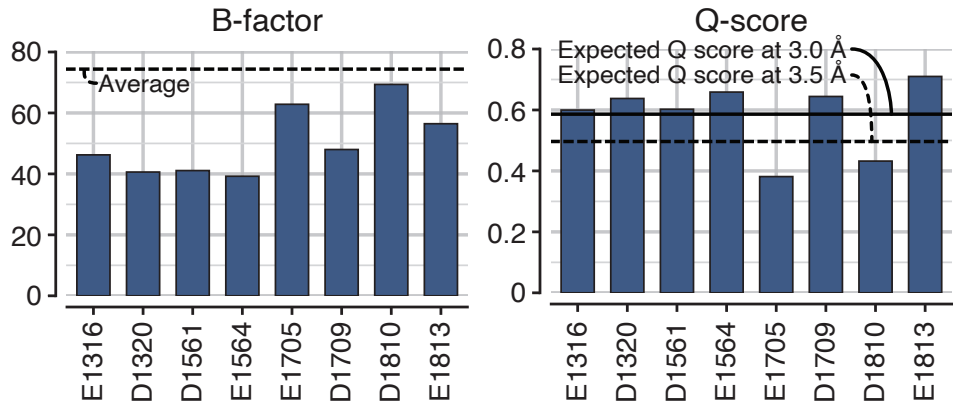

Figure 3.21 B-factor and Q-score plots for active site residues in the hDICER-let-7a-1^{GYM} complex structure. Q-scores for each residue were derived from MapQ of Segger tool plugged in Chimera v1.15. B-factor values were derived from real space refinement in Phenix ISOLDE v1.1.0.

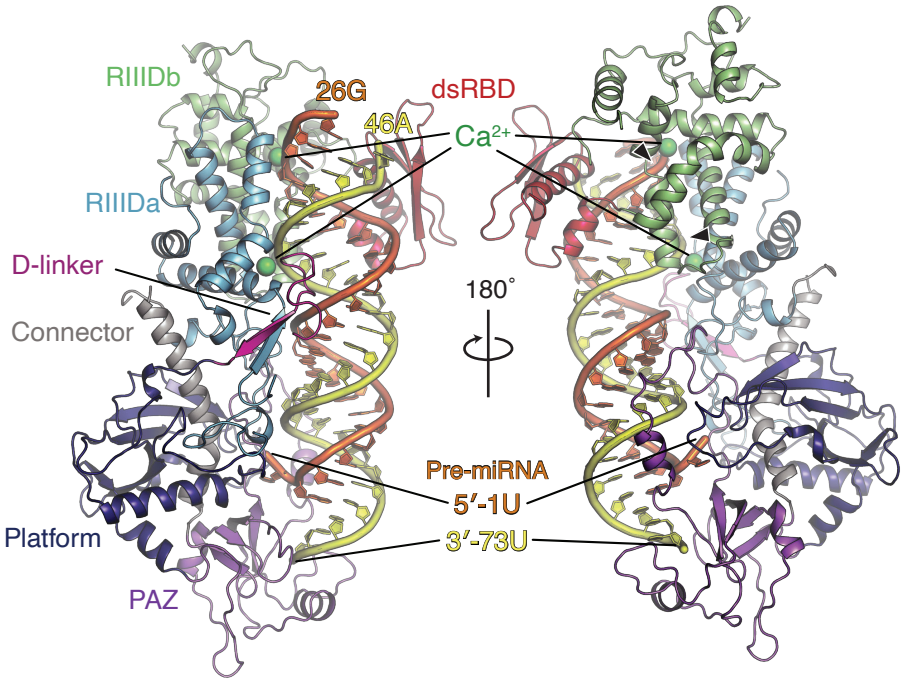

Figure 3.22 Overall structure of the hDICER with a pre-miRNA in a cleavage-competent state. Black arrowheads point to the DICER cleavage sites.

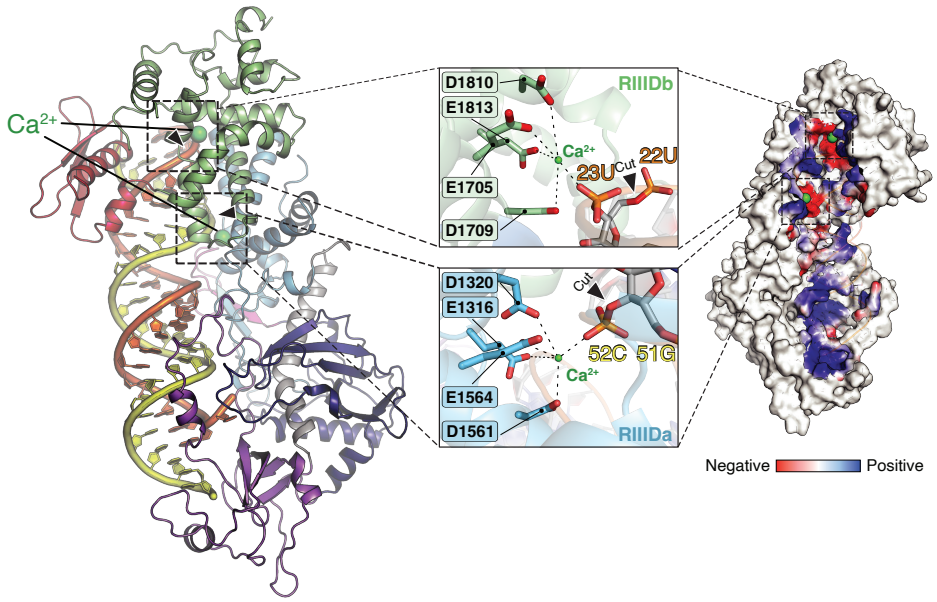

Figure 3.23 Close-up views of the catalytic sites in RIIIDa and RIIIDb domains and electrostatic potential surface model of the catalytic valley along the protein-RNA interface.

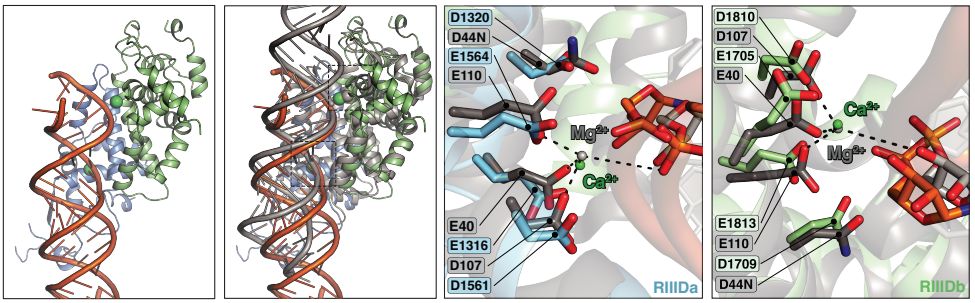

Figure 3.24 Superposition of RIIID domains of hDICER and Aa RNase III (left) and active sites of hDICER and Aa RNase III (right).

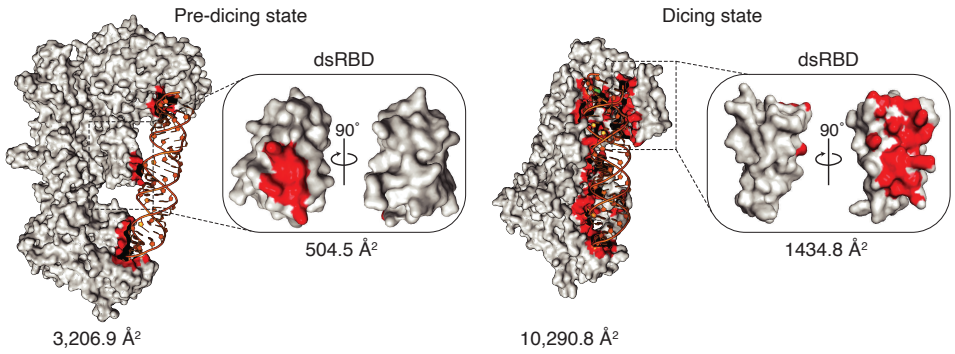

Figure 3.25 Buried surface area of hDICER in a pre-dicing state and a dicing state.

We could build the 3D model for the majority of the pre-miRNA, including the stem region and the additional 4-nt and 6-nt beyond the cleavage sites in the 5p and 3p strands, respectively (Figure 3.15, 3.16). The rest of the terminal loop could not be modeled, affirming the flexible nature of the terminal loop. Our structure shows considerable contacts between hDICER and pre-miRNA, with the total buried surface area of 10,290.8 Å^2 (Figure 3.25). The RMSD between the apo and dicing states was 2.9 Å, mostly accounting for differences in the dsRBD and PAZ domain, with RMSDs of 17.7 Å and 13.6 Å, respectively (Figure 3.26). Compared to the pre-dicing state (Liu et al., 2018), the dicing state structure shows large differences both in protein domain organization and RNA interaction. Note that in the pre-dicing state, there is only limited interaction with pre-miRNA mainly through its termini and loop (with total buried surface area of 3,206.9 Å^2).

One of the prominent changes observed during the transition between apo and dicing states was the fade-out of the density of helicase domain. In the apo-state structure, there appears to be interdomain interactions among the helicase domain, dsRBD, and RIIIDb (Figure 3.27). These interactions likely support the overall architecture of these domains in a stable "closed" conformation. However, superposition of the apo-hDICER structure, for which the helicase domain could be modeled into, shows a steric clash between the helicase domain and the pre-miRNA loop (Figure 3.28). Consistently, the N-terminal helicase-DUF283 domain

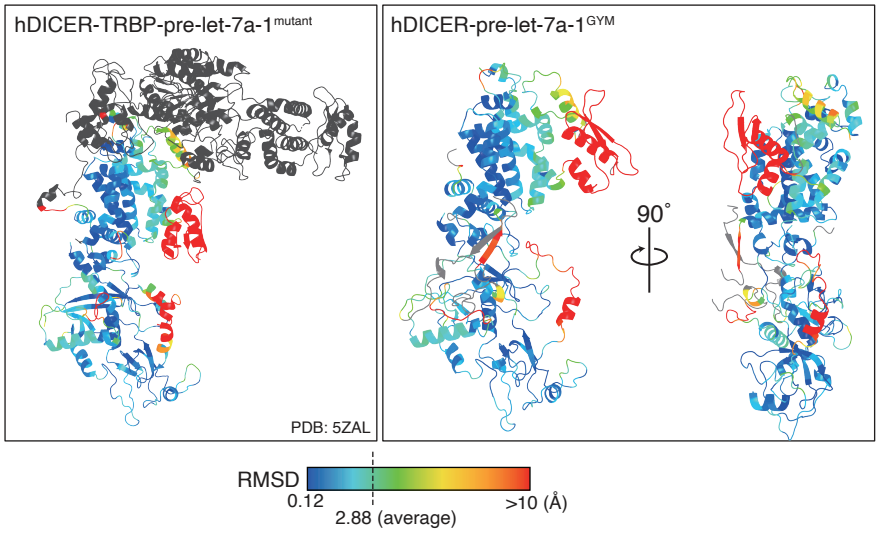

Figure 3.26 RMSD of hDICER-pre-let-7a-1^{GYM} (this study) compared to hDICER-TRBP-pre-let-7a-1^{mutant} (PDB: 5ZAL). Residues not resolved in the dicing state are colored in grey.

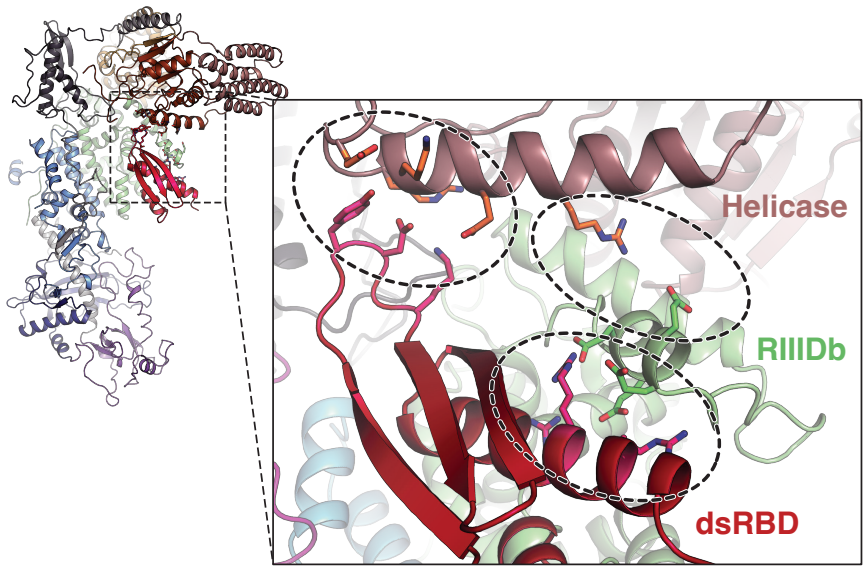

Figure 3.27 Interdomain interactions in apo-hDICER.

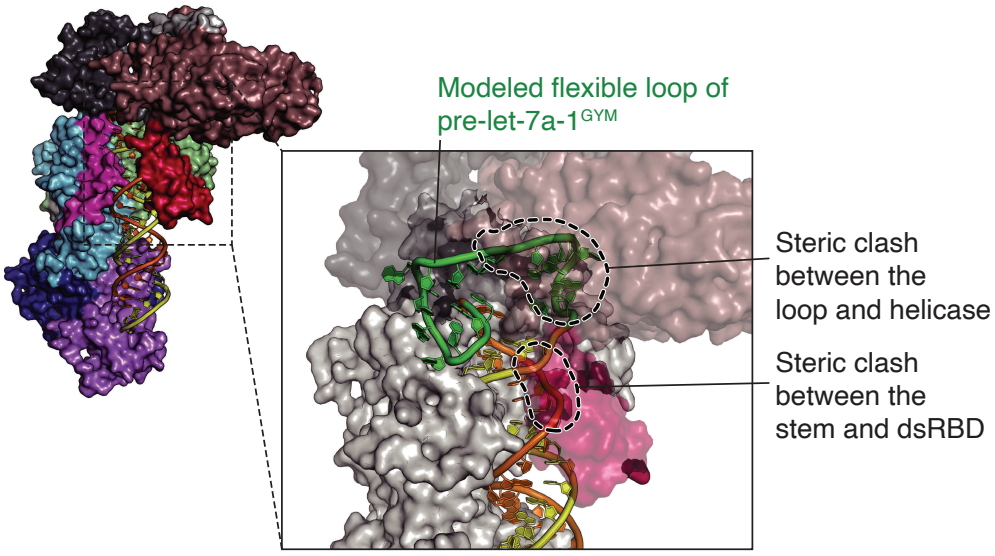

Figure 3.28 Steric clash between pre-let-7a-1^{GYM} and apo-hDICER.

exhibits significant flexibility in a dicing state in our analysis (Figure 3.16). We observed the same result with pre-miR-3121^{GYM} that has a small loop (11 nt), suggesting that the helicase domain becomes flexible generally during the dicing step regardless of the terminal loop size (Figure 3.29). Further in-depth particle classification of ~4,000 particles revealed the extended map of the helicase domain, dislocated from other domains (Figure 3.30). To test whether this missing density is due to chemical integrity, we performed another cryo-EM imaging of the same specimens, after incubating with 2 mM $MgCl_2$ which allowed cleavage reaction (Figure 3.31). We obtained particles at multiple structural states including dicing-state (25%), some intermediates (51%) and apo-like structures (34%) (Figure 3.32). This structural analysis indicates that the helicase domain is chemically intact and that the structural heterogeneity of the helicase domain is induced transiently in the dicing state. After the cleavage reaction, the protein returns to the original apo-state conformation. Collectively, these analyses suggest that the conformational change in the helicase domain is required to allow productive interaction with the substrate. These observations contrast to the recent structures of fly Dcr-2—specialized in the siRNA pathway—that showed a fixed orientation

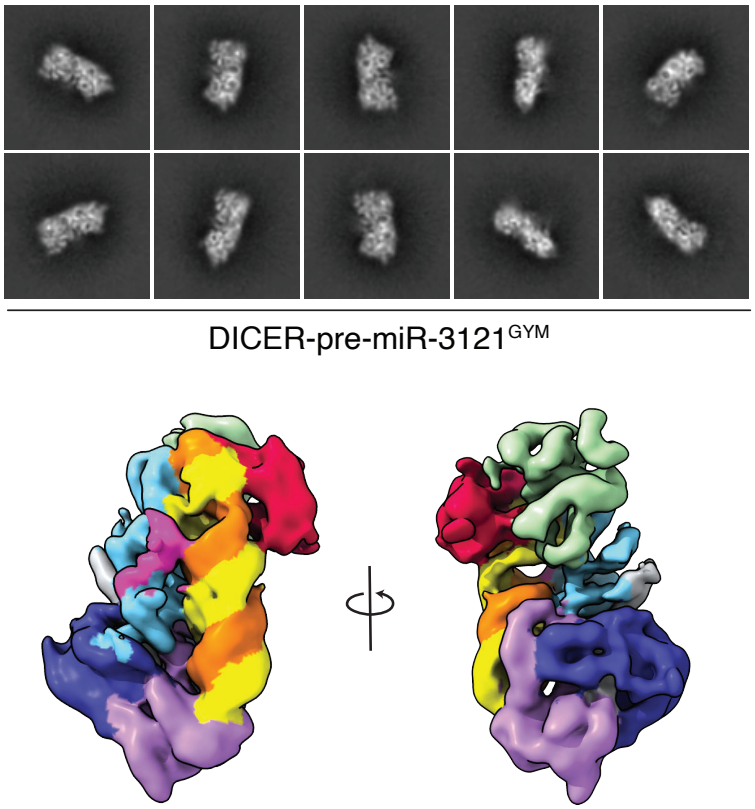

Figure 3.29 Cryo-EM map of the hDICER-pre-miR-3121^{GYM} complex in a dicing state.

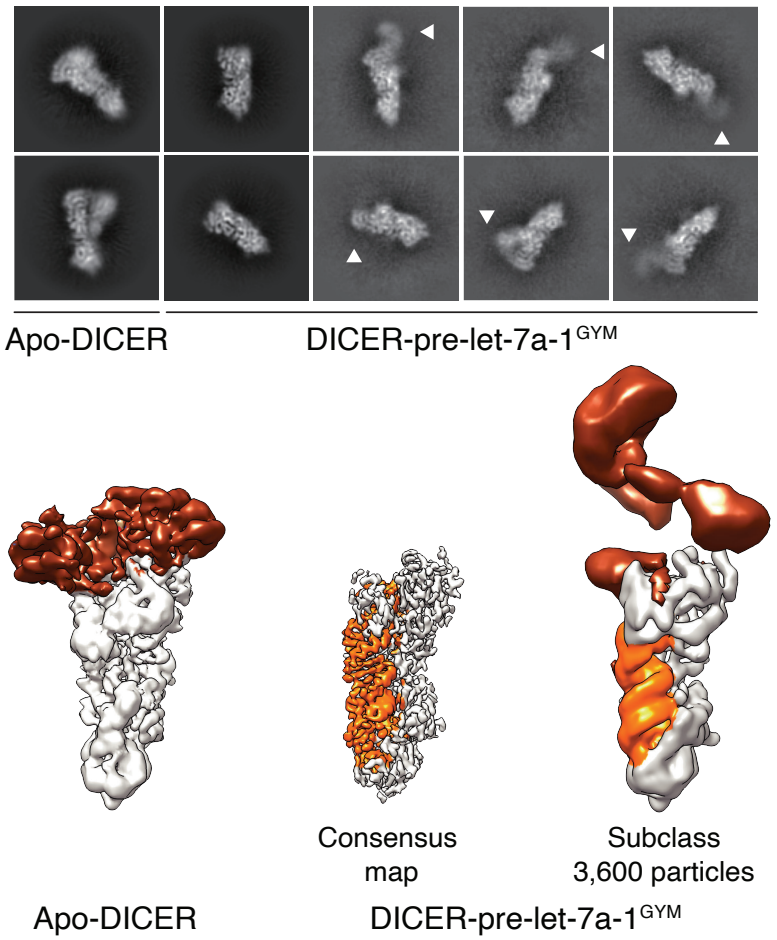

Figure 3.30 Selected 2D class averages and 3D maps showing heterogeneity in the helicase domain. White arrowhead indicates the location of the helicase domain in 2D averages. Bound RNA density is indicated in orange color.

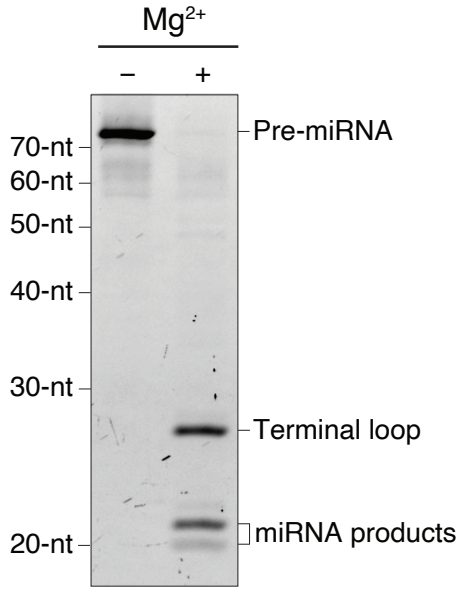

Figure 3.31 Urea-PAGE of hDICER-pre-let-7a-1^{GYM} complex incubated with or without $MgCl_2$ for 10 min at RT, visualized by SYBR gold staining.

of the helicase domain which is engaged in a ATP-dependent translocation of the long dsRNA (Su et al., 2022).

3.2.3 Stem recognition by dsRBD and RIIID

Near the catalytic sites in the upper stem of pre-miRNA, we observed a pronounced movement of the C-terminal dsRBD (Figure 3.33), mediated by the flexible linker that connects to the RIIIDb. This RNA-induced conformational switching orients the dsRBD away from the inner core of the catalytic valley and relieves the steric clash between dsRBD and dsRNA (Figure 3.34) so as to permit dsRNA recognition. With respect to its original position in the apo and pre-dicing states, the dsRBD swings about 12.6 Å and 16.5 Å, respectively (Figure 3.33, 3.34).

Intriguingly, close to the dsRBD-dsRNA interface, we observed a large change in the helical structure of dsRNA, which deviates from the ideal A-form dsRNA structure (Figure 3.35). This conformational distortion expands the width of

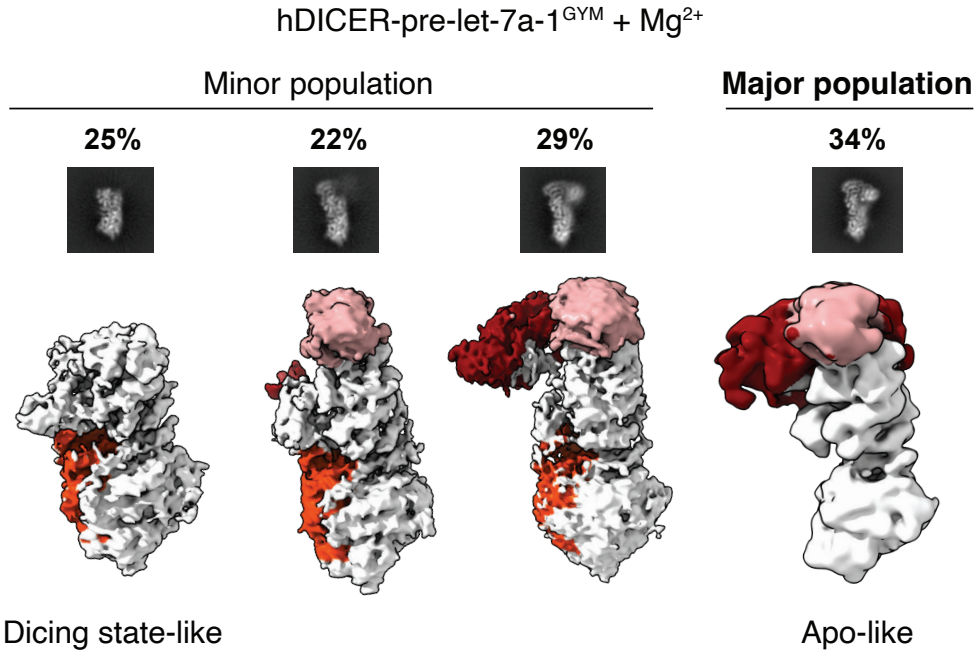

Figure 3.32 Selected 2D class averages and 3D maps showing heterogeneity of the helicase domain of hDICER-pre-let-7a-1^{GYM} complex.

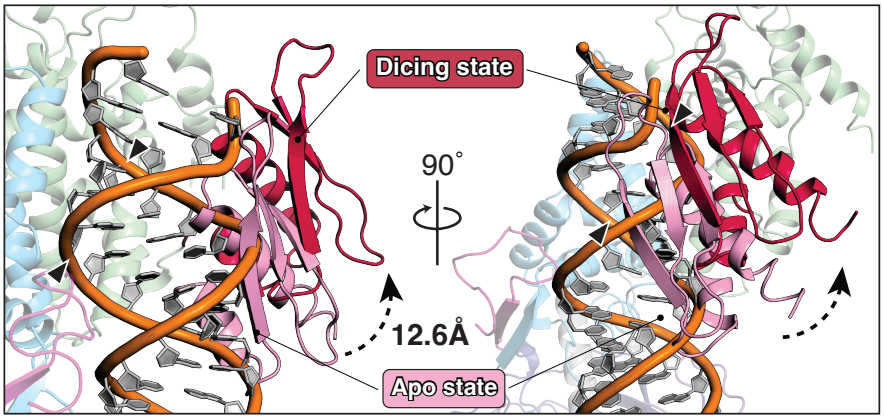

Figure 3.33 Conformational change in dsRBD during the transition from an apo state to a dicing state. Black arrowheads indicate cleavage sites.

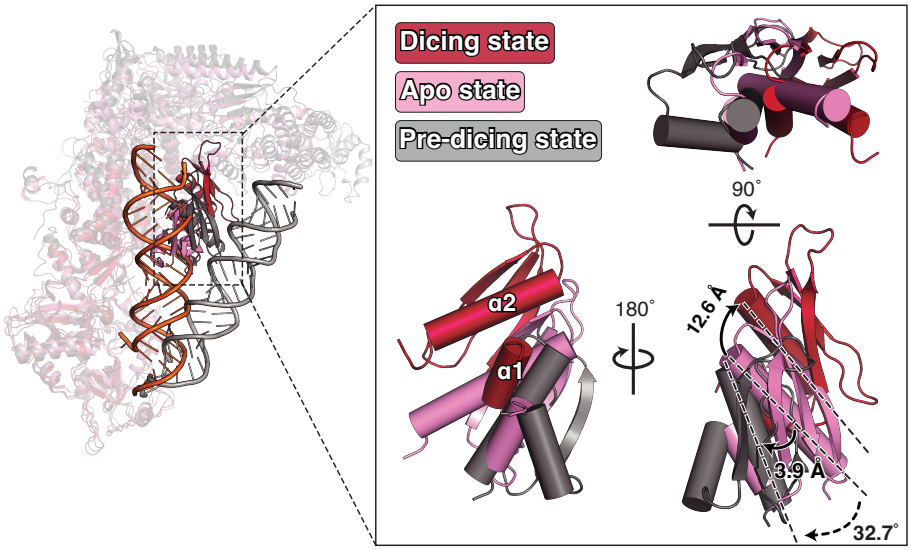

Figure 3.34 Conformational changes of the dsRBD in the apo, dicing and pre-dicing states (PDB: 5ZAL).

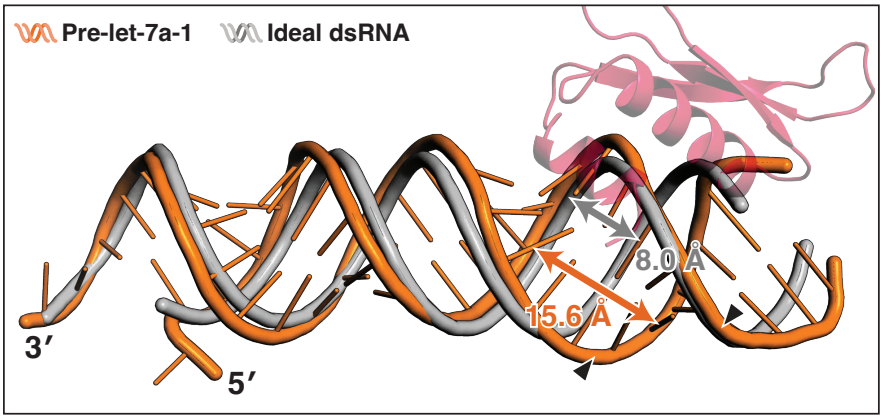

Figure 3.35 Comparison between the structures of ideal A-form dsRNA helix and pre-let-7a-1^{GYM}.

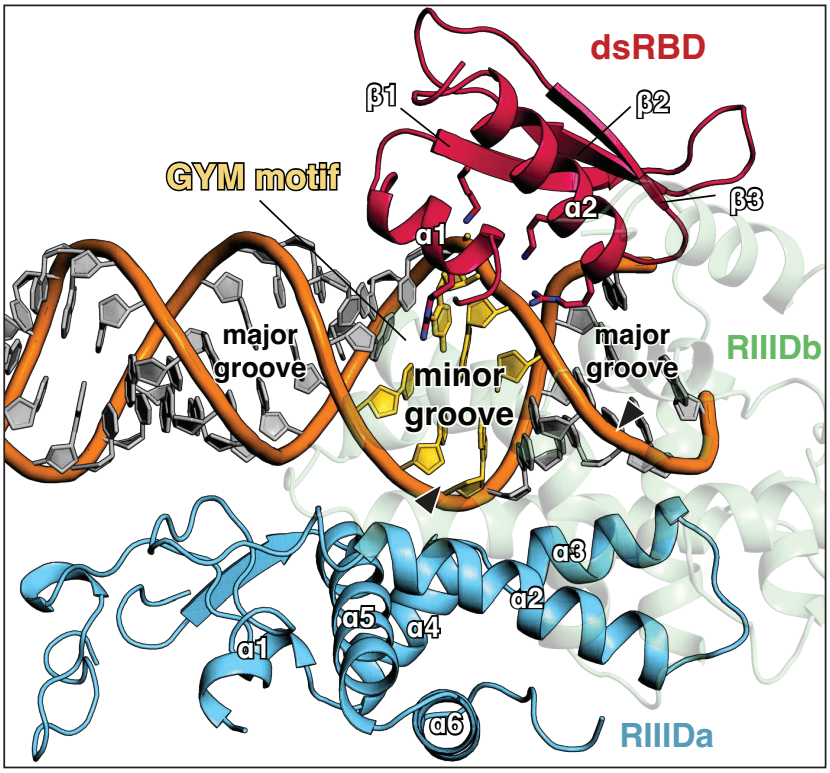

Figure 3.36 Protein-RNA interactions near the cleavage sites in the minor groove.

the major groove of pre-let-7a-1^{GYM} to 15.6 Å, compared to 8.0 Å in the ideal A-form RNA. The successive major and minor grooves across the region are sandwiched between dsRBD and RIIIDa, and form extensive interactions with both domains (Figure 3.36), suggesting a possible basis for the local distortion in the dsRNA structure. Similar observations were made in the high-resolution cryo-EM structure of Arabidopsis DCL-3 (AtDCL3), in complex with a pre-siRNA (Wang et al., 2021) (Figure 3.37), implying that the conformational distortion in dsRNA helix is not specific to pre-let-7a-1 sequence, but induced by protein-RNA interactions unique to a certain group of Dicer homologs.

In addition to the dsRBD, the RIIIDa/b domains wrap around the RNA, forming extensive electrostatic interactions with the upper stem region of the pre-miRNA. Besides the contacts at the catalytic core of the DICER cleavage

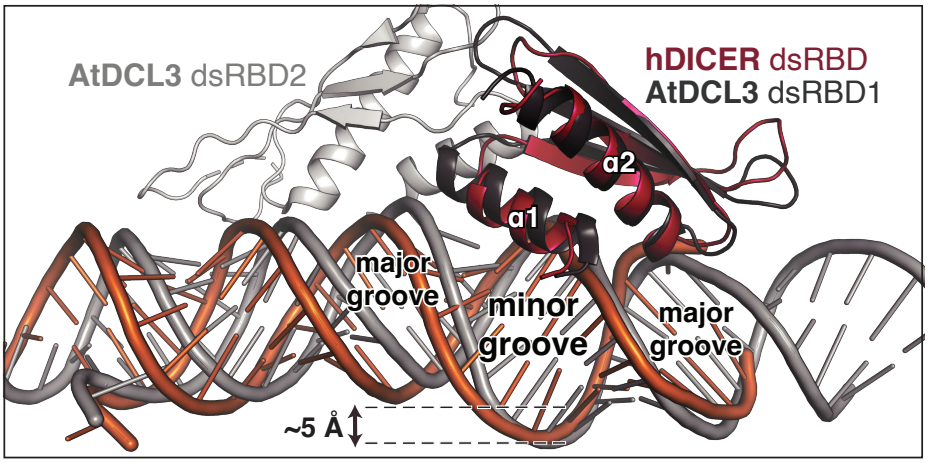

Figure 3.37 Superposition of the dsRBDs of hDICER and Arabidopsis DCL3 (AtDCL3) (PDB: 7VG2).

sites, the RIIIDa, situated on the opposite face to the dsRBD-binding site, makes tight interactions with the RNA in the minor groove (Figure 3.36). We observed that the α-helix 2 and 3 of RIIIDa potentially interact with the ribose sugars and internucleotide phosphate groups (Figure 3.38, left panel). The symmetrically located α-helix 2 and 3 in the RIIIDb may also participate in dsRNA recognition (Figure 3.38, right panel).

Interacting with the distorted dsRNA, the dsRBD adopts a canonical $\alpha\beta\beta\beta\alpha$ topology to cover the dsRNA across the minor and major grooves (Figure 3.36). Basically, the reoriented dsRBD interacts with RNA backbone via its mostly basic patch on the surface (Figure 3.39). For instance, in the major groove, the positively charged residues contact with RNA backbone near 19C (Figure 3.40, 3.41). In the minor groove, however, we observed a sequence-specific interaction between the dsRBD and the RNA. The α-helix 1 of the dsRBD is situated in the vicinity of the GYM motif (Figure 3.36, 3.42), the cis-acting element that significantly improves the fidelity of processing. An arginine (R1855) in the α-helix 1 protrudes into the minor groove and may form hydrogen bonds with the C–C mismatch (Figure 3.42). This is consistent with the observation that mutating this

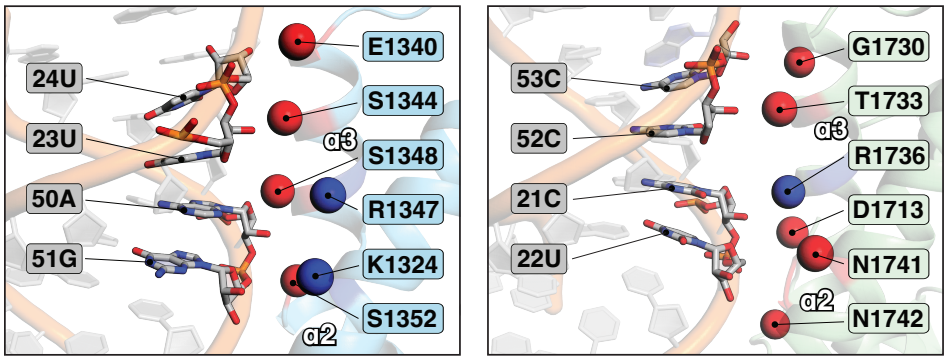

Figure 3.38 Protein-RNA interactions in the interface between RIIID domains and RNA backbon. Left: RIIIDa, Right: RIIIDb.

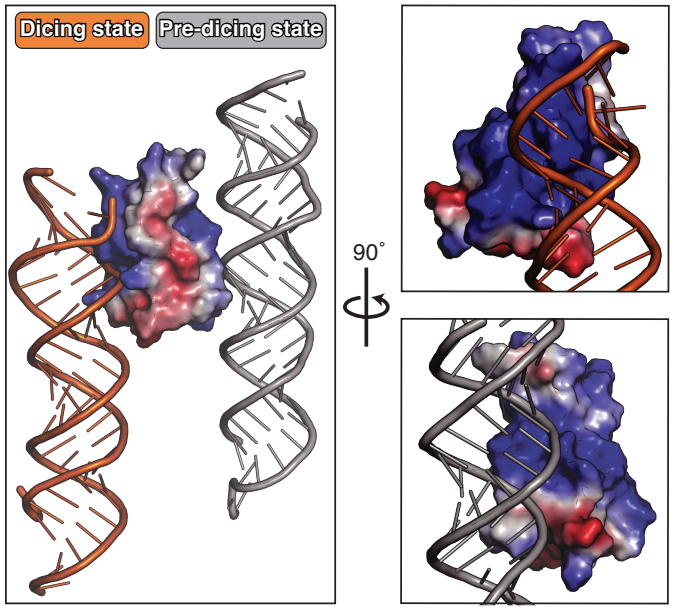

Figure 3.39 Surface charge of the dsRBD, with dsRNA-dsRBD interface in dicing and pre-dicing states (PDB: 5ZAL).

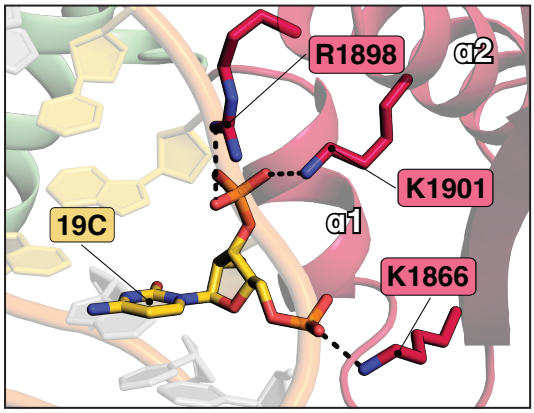

Figure 3.40 Non-sequence-specific interactions between dsRBD and the RNA phosphate backbone.

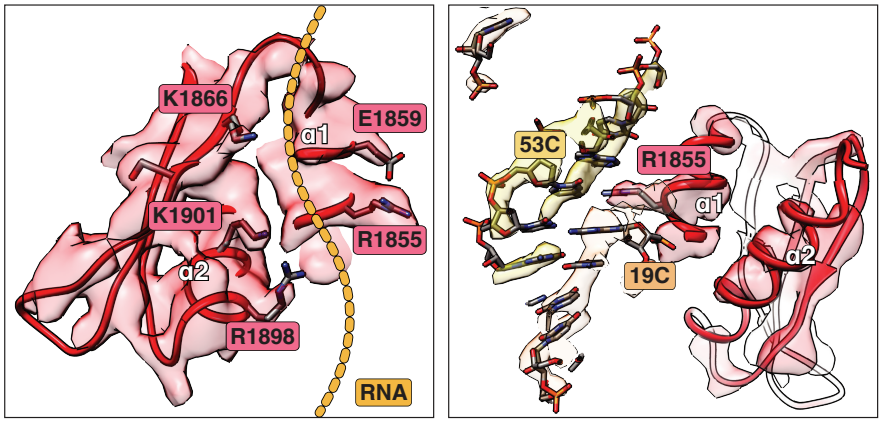

Figure 3.41 Cryo-EM map and model of the hDICER dsRBD with dsRNA.

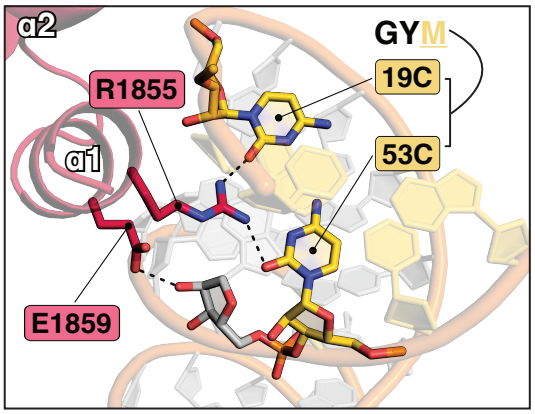

Figure 3.42 Sequence-specific interactions between dsRBD and the C–C mismatch of the GYM motif.

arginine residue abolishes the effect exerted by the mismatch. Thus, in contrast to the previous papers claiming the auxiliary function of the hDICER dsRBD (Ma et al., 2012; Zhang et al., 2004), our structure provides the structural basis for its predominant role in cleavage site selection, overriding the effects of 5´ and 3´ counting mechanisms. Taken altogether, the dsRBD and the RIIIDa/b domains anchor the upper region of the pre-miRNA and induce local distortion of the RNA, facilitating the recognition of not only the RNA backbone but also the GYM motif.

3.2.4 PAZ helix reorients to accommodate RNA in a dicing state

An earlier structural study on a hDICER fragment containing the platform-PAZ domain showed a knob-like protrusion with a small alpha helical segment within the PAZ domain, which is specifically found in DICER (Tian et al., 2014). This 'PAZ helix' (also known as hDICER-specific helix) separates the 5´ and 3´ pockets and orients the bound RNA away from the surface of hDICER, which is thought to occur in product-release state and/or pre-dicing state (Tian et al., 2014) (Figure 3.43, middle panel). This helix, however, may be dynamic given that an additional 'melted' conformation of the PAZ helix was observed in the platform-PAZ-small

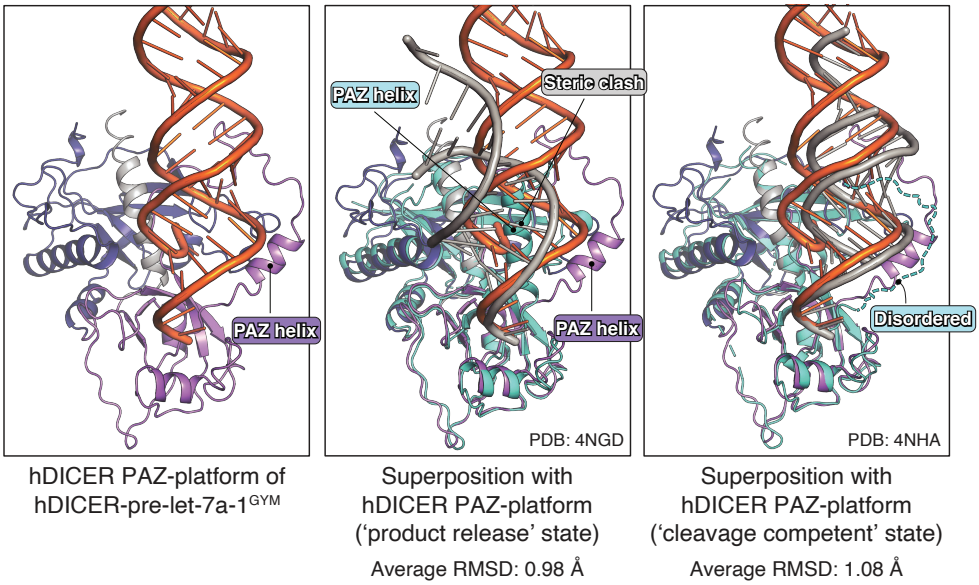

Figure 3.43 Superposition of hDICER PAZ-platform domain in the cryo-EM structure and in the crystal structure (PDB: 4NHA, grey).

RNA duplex complex (Tian et al., 2014) (Figure 3.43, right panel). Indeed, we found a large conformational change in the PAZ helix resulting in a tilt angle of ~54° from its position in the pre-dicing state (Figure 3.44). The PAZ helix is consequently located near the lower stem region of pre-miRNA (Figure 3.44). Superposition with the structure in a pre-dicing state showed that the pre-miRNA moves ~30° towards the catalytic valley during the transition to a dicing state (Figure 3.45). Taken together, the spatial rearrangement of the PAZ helix is necessary to allow the pre-miRNA to be aligned parallel to the catalytic valley for subsequent cleavage.

In addition, this conformational change puts the short stretch of positively charged amino acids ($_{1019}$KRKAK$_{1023}$) in the vicinity of the negatively charged backbone of the 3p strand (Figure 3.46). To assess the significance of the observed interaction between the PAZ helix and the dsRNA, we introduced mutations by replacing the positively charged residues with five glutamate ('E5') or alanine ('A5') residues, or by deleting the helix ('ΔPAZh') (Figure 3.47). The PAZ helix

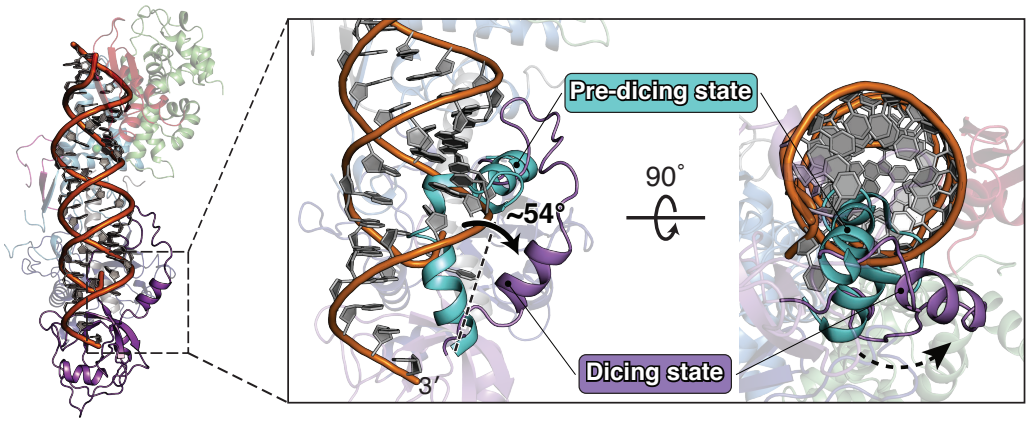

Figure 3.44 Conformational change of the PAZ helix between a dicing state and a pre-dicing state.

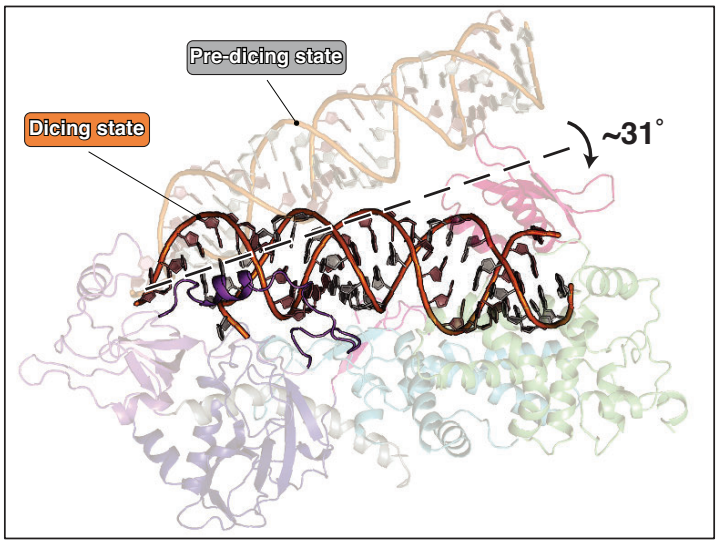

Figure 3.45 Changes in the position of the pre-miRNA in a dicing state and a pre-dicing state.

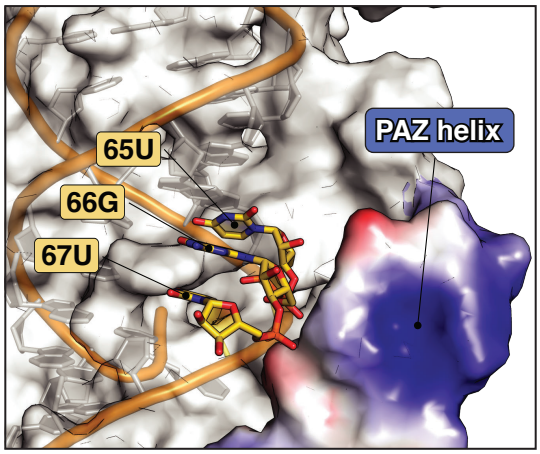

Figure 3.46 Electrostatic interactions between the positively charged PAZ helix and the negatively charged RNA phosphate backbone.

mutant proteins were purified (Figure 3.2) and incubated with pre-let-7a-1 to quantitatively measure their effects on the cleavage efficiency *in vitro*. The mutant proteins showed reduced cleavage efficiencies, regardless of the 3´ overhang lengths (Figure 3.47–3.49). PAZ helix^{E5}, showed a more severe effect compared to PAZ helix^{A5}, presumably due to the electrostatic repulsive forces created between the PAZ helix and the RNA backbone. Intriguingly, the deletion of the PAZ helix led to a modest but significant reduction in cleavage efficiency. This result, with the structural observations, implies that the PAZ helix may have an autoinhibitory effect on the transition to a dicing state besides its contribution to RNA binding affinity, once the dicing state is achieved.

We next sought to investigate the role of the PAZ helix in miRNA biogenesis by transiently expressing the mutant DICER in DICER KO HCT116 cells, and then performed AQ-seq for accurate quantification of miRNAs (Figure 3.50) (Kim et al., 2019).[3] The mutations resulted in a global reduction in the abundance of miRNAs (Figure 3.50), corroborating the *in vitro* results. Our data collectively

[3]AQ-seq analysis was done by Dr. Haedong Kim.

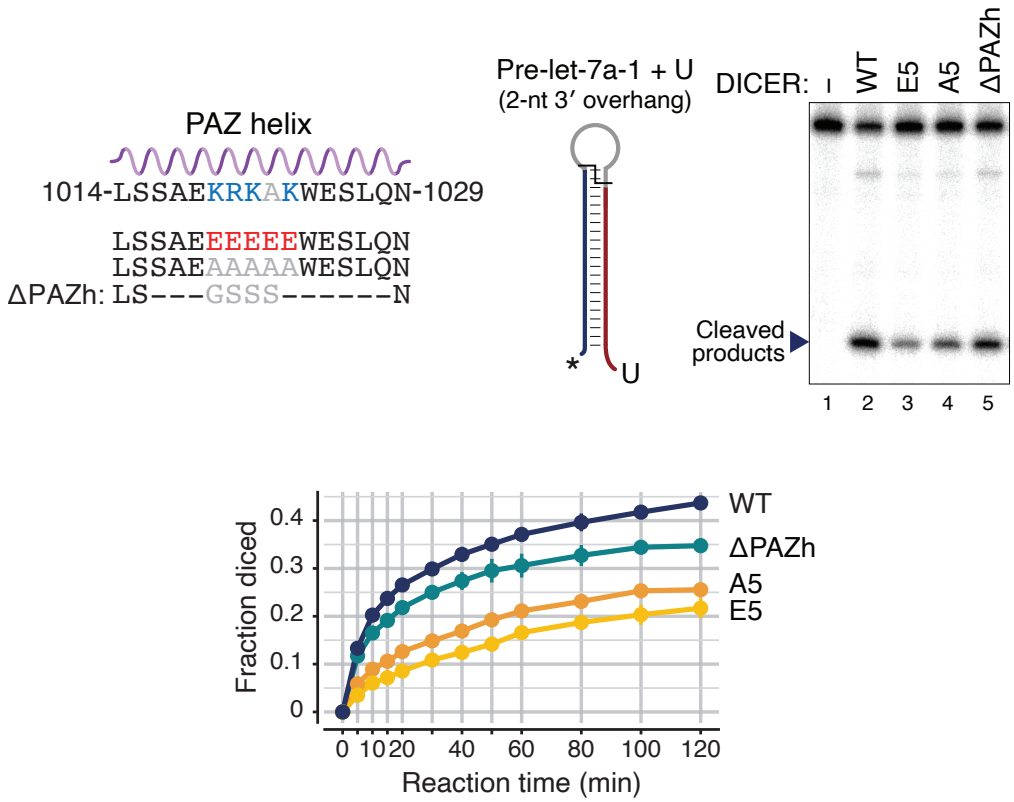

Figure 3.47 *In vitro* processing of pre-let-7a-1 with a 2-nt 3´ overhang. Points and bars indicate mean ± SD (n = 2). Relative cleavage was calculated by quantifying the band intensity (1−uncleaved/input).

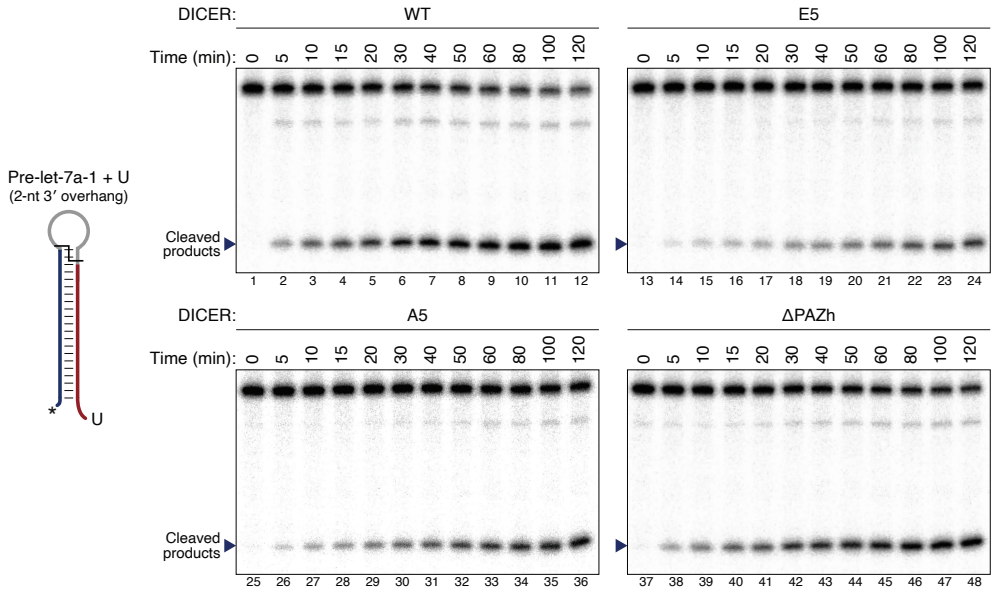

Figure 3.48 Time-course *in vitro* processing of pre-let-7a-1 with a 2-nt 3′ overhang.

suggest that the conformational change in the PAZ helix and its subsequent interaction with the RNA backbone are important for pre-miRNA processing.

3.2.5 Architecture of the 5′ and 3′ end pockets

Consistent with the idea that hDICER recognizes the pre-miRNA termini for accurate processing, our structure illustrates both 5′ and 3′ ends stably anchored within the platform and PAZ domains, respectively (Figure 3.51–3.53). In the 3′ pocket, the last phosphodiester linkage makes close interactions with a cluster of four conserved tyrosine residues (Y936, Y971, Y972, Y976) and an arginine residue (R937) through potential hydrogen bonding, which is in line with previous literatures (Tian et al., 2014; Liu et al., 2018) (Figure 3.52).

The 5′ end of pre-miRNA is in a unique kinked conformation (Figure 3.51), which is very different from the previous structures of hDICER with small RNA duplexes or in the pre-dicing state or the other Dicer homologs in the dicing state

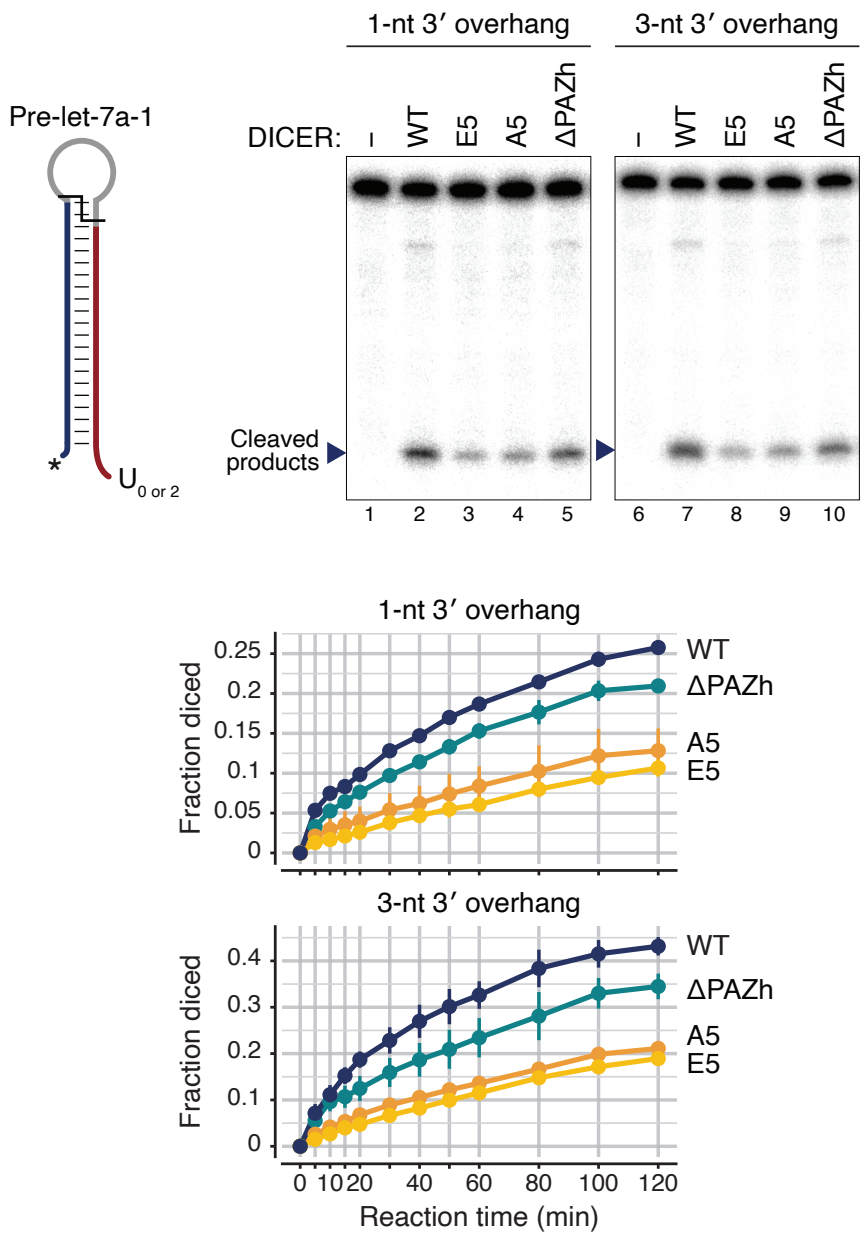

Figure 3.49 *In vitro* processing of pre-let-7a-1 with a 1-nt 3´ overhang or a 3-nt 3´ overhang. Points and bars indicate mean ± SD (n = 2). Relative cleavage was calculated by quantifying the band intensity (1−uncleaved/input).

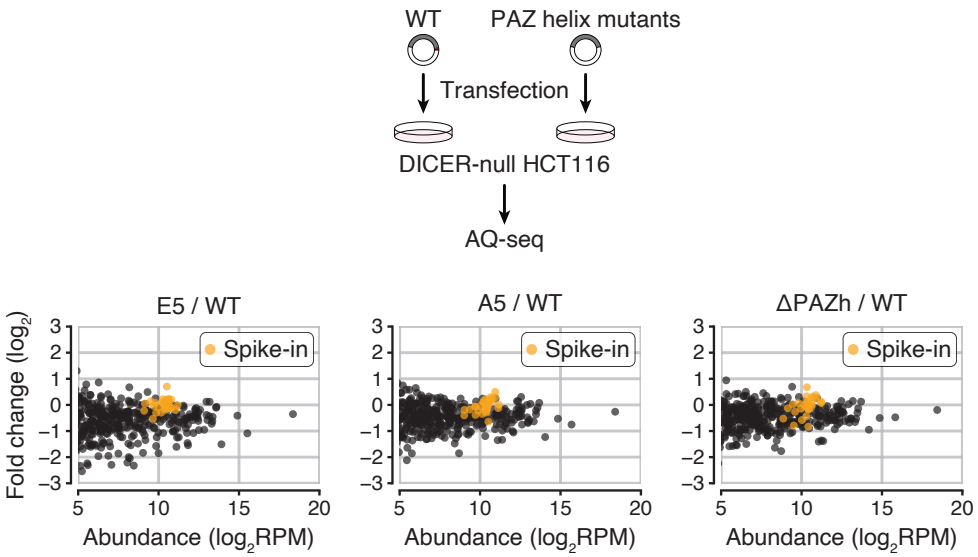

Figure 3.50 Schematic outline of the rescue experiment and comparison of miRNA abundance. Spike-ins were used for normalization. RPM, reads per million.

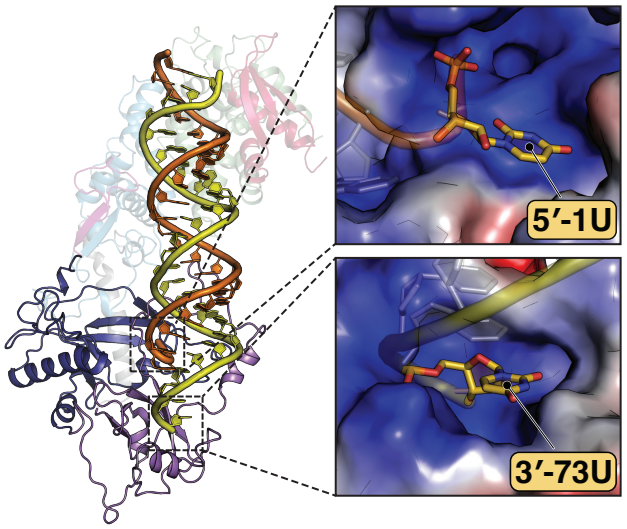

Figure 3.51 The 5´ and 3´ ends of pre-miRNA anchored in the basic 5´ and 3´ pockets in the platform and PAZ domains, respectively.

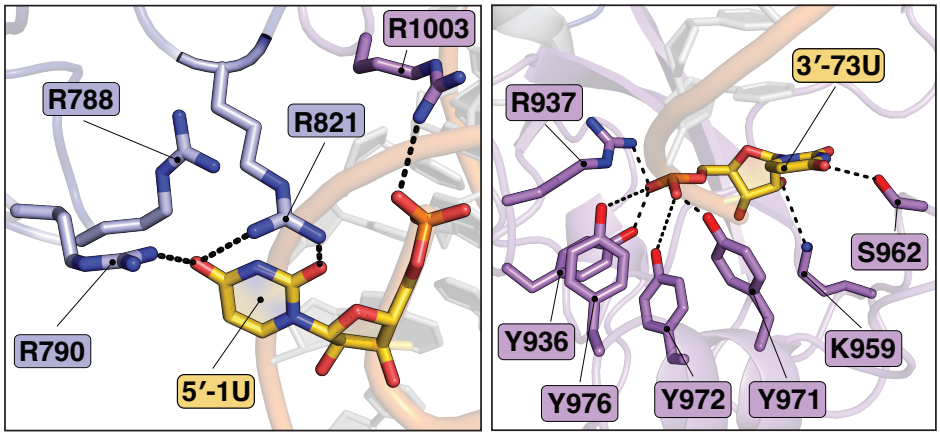

Figure 3.52 Pre-miRNA end recognition by the 5´ and 3´ pocket via hydrogen bonding.

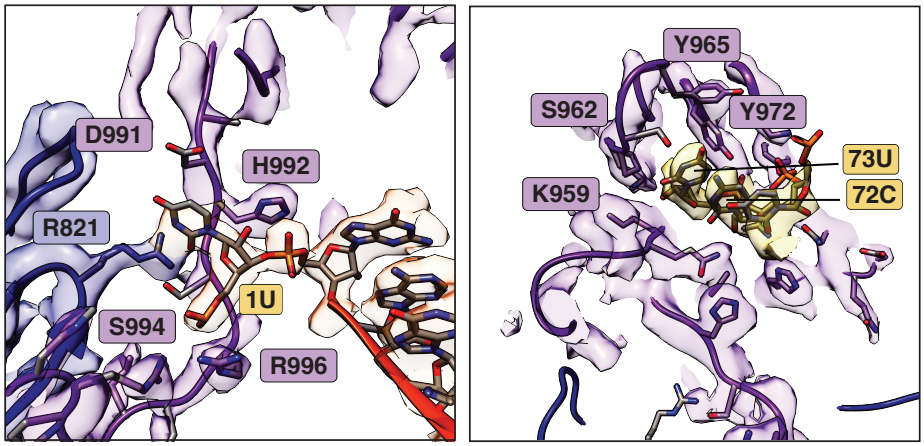

Figure 3.53 Cryo-EM map of the 5´ pocket (left) and 3´ pocket (right).

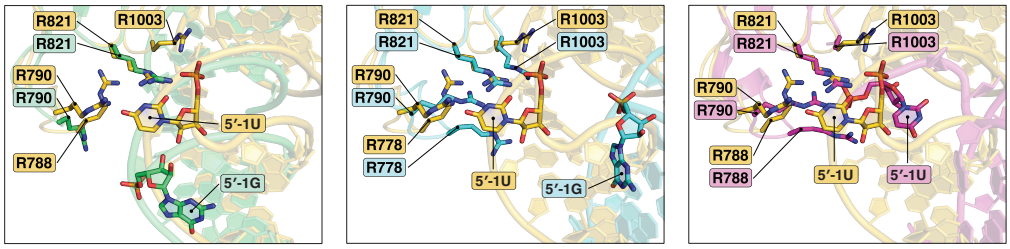

Figure 3.54 Superposition of hDICER-pre-let-7a-1^{GYM} and Platform-PAZ-Connector Helix (Tian et al., 2014).

(Wang et al., 2021; Tian et al., 2014; Liu et al., 2018; Nagano et al., 2014) (Figure 3.54–3.56). The conformation in our structure allows the 5′ monophosphate to be inserted into the 5′ pocket and possibly interact via hydrogen bonds with the main chain amide and the amine group of two arginine residues, R996 and R1003, respectively (Figure 3.52, 3.53). In addition, the 5′ base unexpectedly flips out to interact with a cluster of three arginine residues—R788, R790, and R821—that make hydrogen bonds with the base (Figure 3.52). These results collectively suggest that the 3′ pocket is conserved while the 5′ pocket may have emerged more recently to meet the needs of individual Dicer homologs with different substrate types.

3.2.6 5′ end recognition is disrupted in cancer-associated DICER mutations

The ‘5′ counting’ is important for miRNA biogenesis because it ensures accurate and efficient processing regardless of the frequent 3′ end trimming and tailing in cells (Park et al., 2011). Interestingly, three amino acids in the 5′ pocket are found to be mutated in diverse cancer types according to the cancer databases including The Cancer Genome Atlas (TCGA): R790Q (in colon adenocarcinoma), R821H (in colorectal adenocarcinoma), and R1003Q (in uterine endometrioid carcinoma and rectal adenocarcinoma) (Cerami et al., 2012; Gao et al., 2013).

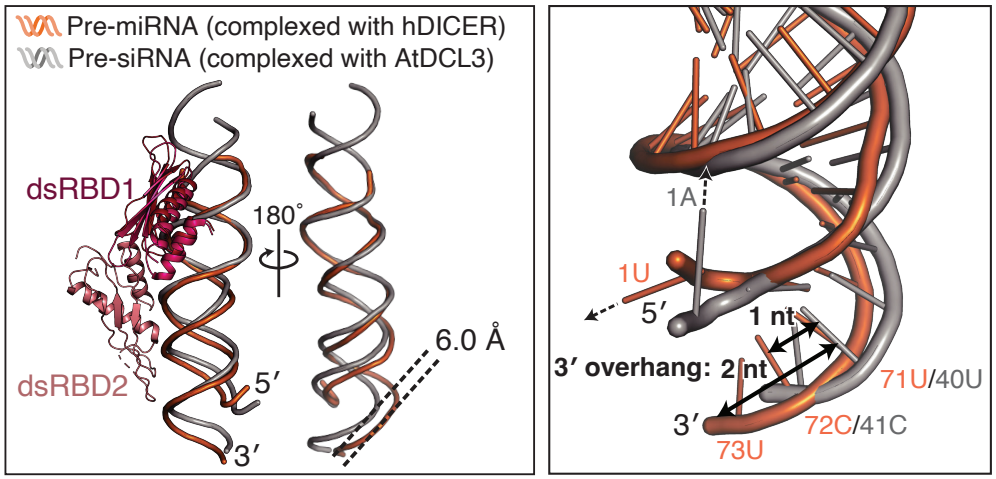

Figure 3.55 Superposition of dsRNAs complexed with hDICER and AtDCL3 (PDB: 7VG2).

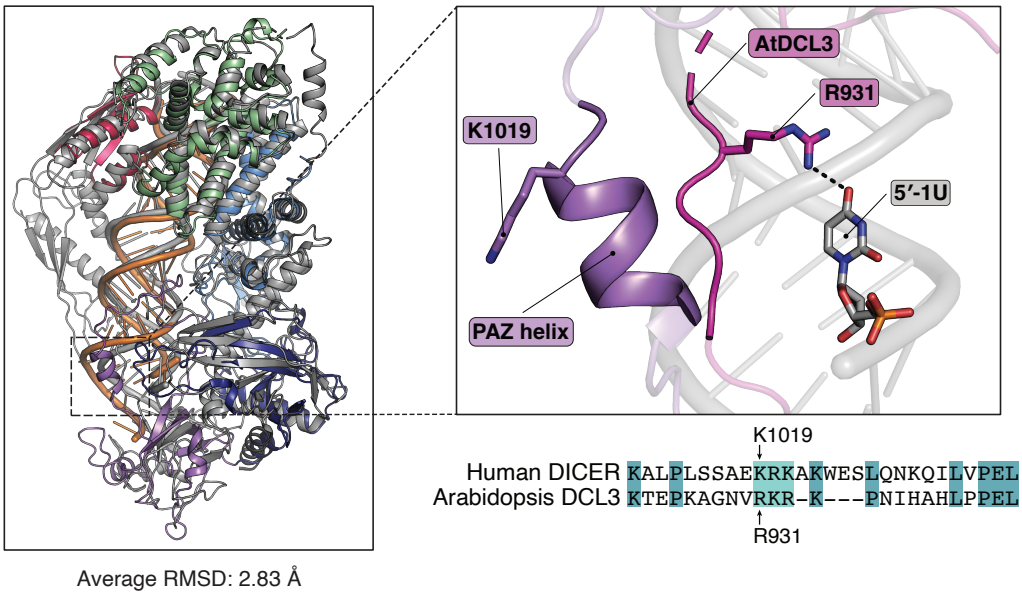

Figure 3.56 Superposition of hDICER in a dicing state and AtDCL3-pre-siRNA complex (PDB: 7VG3, magenta).

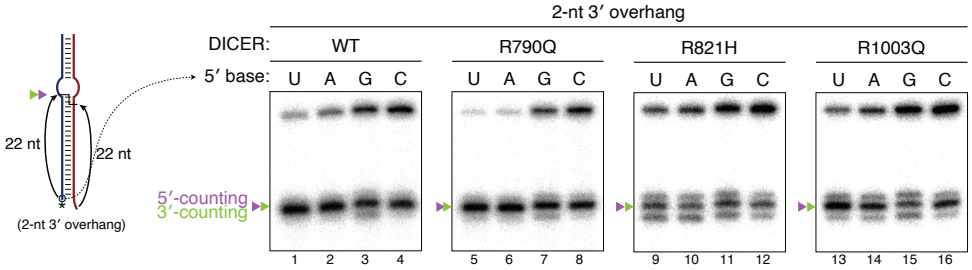

Figure 3.57 *In vitro* processing of duplex RNAs with either a 2-nt overhang and a varying sequence at the 5′ end. The nucleotide in the 3p strand opposite to the varying sequence is A.

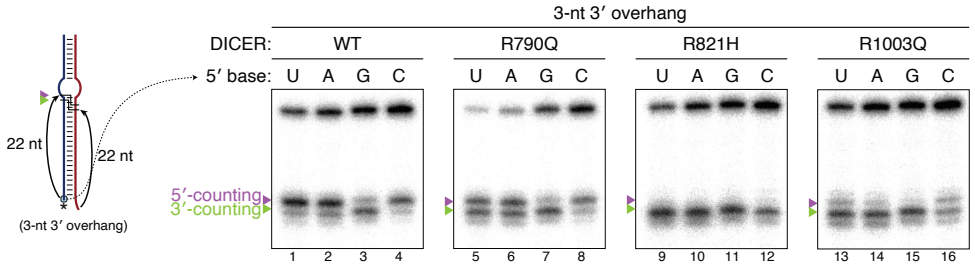

Figure 3.58 *In vitro* processing of duplex RNAs with either a 3-nt overhang and a varying sequence at the 5′ end. The nucleotide in the 3p strand opposite to the varying sequence is A.

In vitro, the R821H and R1003Q mutants show defects in cleavage site selection; they produced multiple products (21–23 nt) from dsRNA with a canonical 2-nt 3′ overhang (Figure 3.57). Note that wild-type hDICER cleaves this substrate homogeneously because the 5′ and 3′ counting mechanisms corroborate to ensure the generation of a 22-nt product (Figure 3.57). To differentiate the 5′ counting and 3′ counting mechanisms more clearly, we next used dsRNA with a 3-nt 3′ overhang (Figure 3.58, lanes 9–16). In this assay, all three mutants showed significant reduction of the 5′ counting capability, validating our structural observation that these residues are in close contact with the 5′ end of RNA. The R821H mutation as well as two other mutations on R821 (R821E and R821A) resulted in a complete loss of 5′ counting (Figure 3.58, 3.59).

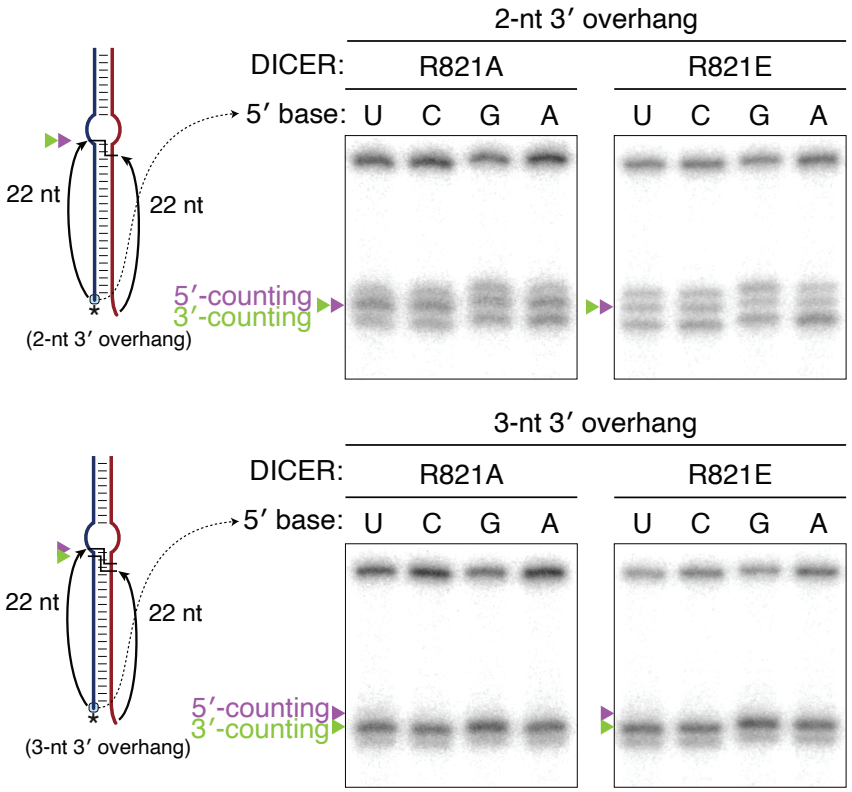

Figure 3.59 *In vitro* DICER processing of pre-miRNA-like duplex with a 2-nt or 3-nt 3′ overhang. The base opposite to the varying sequence is A on the 3p strand.

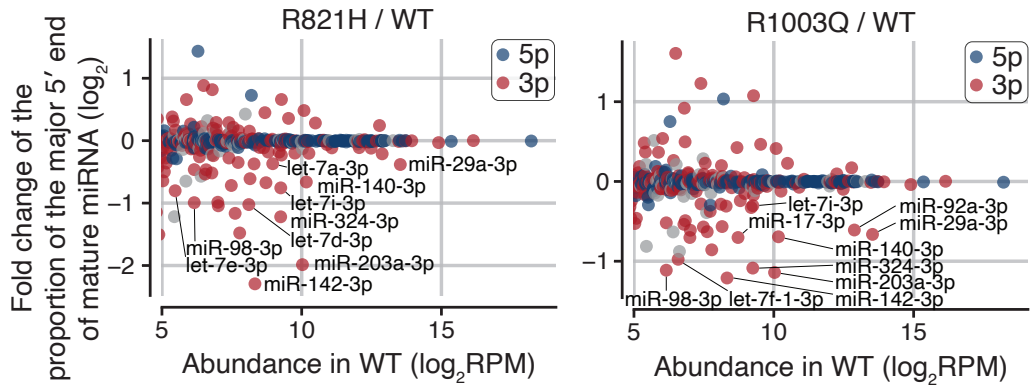

Figure 3.60 The effect of cancer-associated 5´ pocket mutations in cleavage accuracy. Changes in cleavage accuracy was estimated with the fold change of the proportion of the major 5´-isomiR. For a given miRNA, the most abundant 5´-isomiR was identified in the WT sample. Grey, unannotated strand.

To interrogate the impact on miRNA biogenesis, we performed rescue experiments with wild-type and mutant hDICER proteins (Figure 3.60, 3.61).[4] We observed dramatic alterations in the DICER cleavage sites of many miRNAs (Figure 3.60, 3.61) as indicated by the 5´ end of 3p miRNAs, which is determined at the DICER processing step (Figure 3.62). The DROSHA processing sites (5´ ends of 5p mature miRNAs) were largely unaffected, as expected. The most notable examples are let-7-3p and miR-324-3p which were previously shown to be dependent on the recognition by the 5´ pocket (Figure 3.60, 3.61) (Park et al., 2011; Kim et al., 2020). Other 3p miRNAs (such as miR-29a, miR-92a, miR-140, miR-142, and miR-203a) also showed substantial changes in the processing sites (Figure 3.60, 3.61), which alter the miRNA isoform landscape and their target repertoire.

Moreover, abundance of many miRNAs was decreased by the mutations (Figure 3.63). Notably, the mutations led to a significant reduction in the levels of tumor-suppressive let-7 family members—let-7a, let-7d, let-7e, let-7f, let-7i, and miR-98—which were shown to be downregulated in various cancers (Figure

[4]AQ-seq analysis was done by Dr. Haedong Kim.

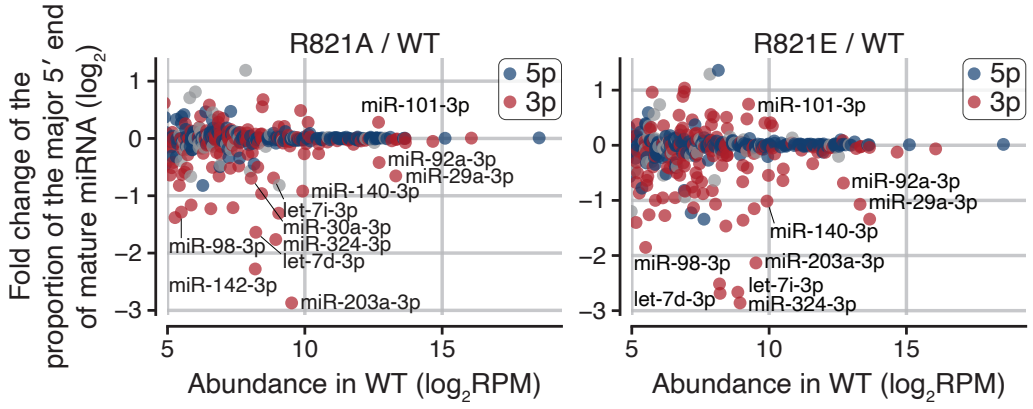

Figure 3.61 The effect of 5´ pocket mutations in cleavage accuracy. Changes in cleavage accuracy was estimated with the fold change of the proportion of the major 5´-isomiR. For a given miRNA, the most abundant 5´-isomiR was identified in the WT sample. Grey, unannotated strand.

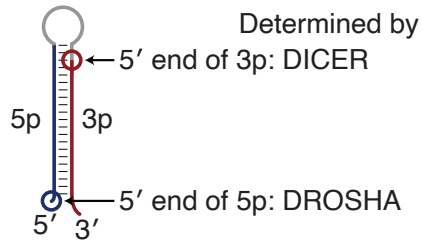

Figure 3.62 DROSHA/DICER cleavage sites dictated by 5´ ends of mature miRNAs.

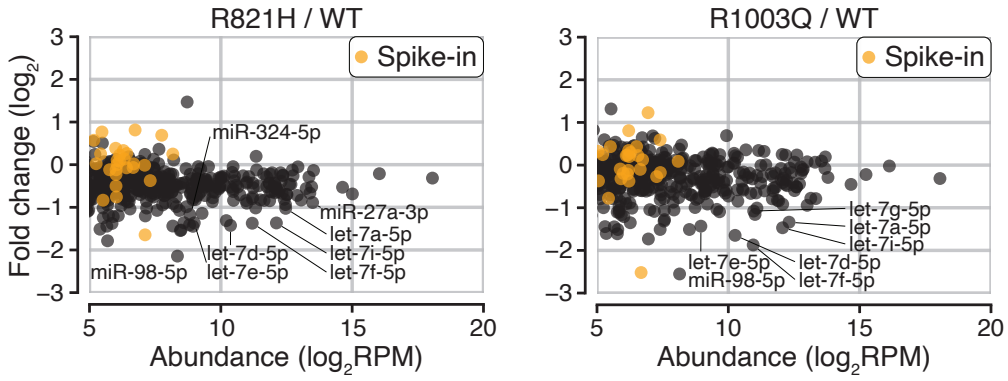

Figure 3.63 Abundance change of endogenous miRNAs. Spike-ins were used for normalization.

3.63) (Takamizawa et al., 2004; Sampson et al., 2007; Johnson et al., 2005; Frank et al., 2010). Our results indicate the significance of the 5′ end recognition and implicates the oncogenic potential of the 5′ pocket mutations.

3.2.7 Identity of the 5′ end base affects the 5′ counting mechanism

In the 5′ pocket, we noticed that R821 potentially forms hydrogen bonds with the O2 carbonyl of the 5′ terminal uracil (Figure 3.52), hinting at a previously unknown base specificity. To examine the base specificity, we modeled other bases—cytosine, adenine, and guanine—into the corresponding position within the 5′ pocket structure (Figure 3.64). The guanine, unlike other bases, has a 2-amino group that sterically overlaps with the guanidino group of R821, prompting us to experimentally test the effect of the 5′ terminal sequence on the 5′ counting rule.

In *in vitro* processing assays (Figure 3.58, lanes 1–4), dsRNAs with 5′-U, 5′-C, and 5′-A base are cleaved predominantly by the 5′ counting mechanism, while the substrate with 5′-G is cleaved mainly by the 3′ counting mechanism . Even in the presence of a 2-nt 3′ overhang, the dsRNA substrate with 5′-G was imprecisely cleaved, yielding 21–22 nt products (Figure 3.57, lanes 1–4), which

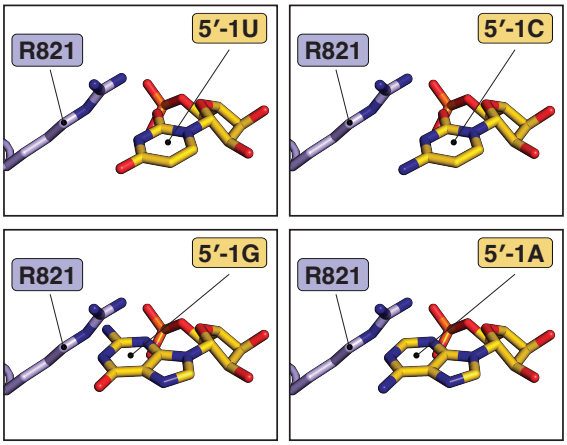

Figure 3.64 Predicted structural impact of the 5′ end base substitutions on the interaction with the 5′ pocket.

is consistent with earlier observation that defects in the 5′ counting mechanism give rise to multiple products (Park et al., 2011). Of note, these observations are not explained by the previous notion that pre-miRNAs with high thermodynamic stability at the 5′ terminus do not follow the 5′ counting rule, because among our substrates, only the dsRNA starting with 5′-U forms a base-pair at the terminus. Taken together, 5′-G may be inefficiently inserted to the 5′ pocket compared to other bases, compromising the 5′-end recognition.

To investigate whether R821 is involved in the recognition of the 5′ end, we performed the same *in vitro* processing assays with R821 mutants (Figure 3.2). The R821 mutants completely lost the 5′ counting capability, and were no longer influenced by the 5′ terminal base regardless of the 3′ overhang length (Figure 3.57–3.59). Thus, R821 plays a key role in recognizing the 5′ terminal sequence, disfavoring 5′-G.

To further inquire about the 5′ terminal base identity, we examined the sequences of human pre-miRNAs. We found a significant enrichment of uridine and adenine—bases known to facilitate AGO loading (Figure 3.65) (Frank et al., 2010; Suzuki et al., 2015). Intriguingly, we found a marked depletion of guanine

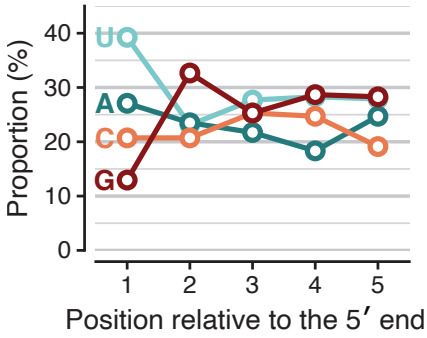

Figure 3.65 The nucleotide composition of human pre-miRNAs. The proportion of each base at the indicated position relative to the 5´ end of pre-miRNAs is shown. miRNAs whose 5p is registered in miRGeneDB (n = 502) were included in this analysis.

compared to cytosine at the 5´ end relative to other neighboring positions (Figure 3.65), despite the fact that both bases are equally disfavored by the AGO protein. This suggests an evolutionary pressure against the 5´-G in natural pre-miRNAs.

Finally, by examining miRNAs that are significantly affected by the R821 mutations (Figure 3.60, 3.61, 3.66), we found that the G base is completely depleted at the 5´ end of affected miRNAs, in contrast to those unaffected (Figure 3.67). This is in line with the structural prediction (Figure 3.64) that U, C, or A can be inserted into the 5´ pocket while G is incompatible with R821 in the 5´ pocket. Thus, owing to the structural restriction posed by R821, hDICER has a nucleotide preference for H (any base but G) at the 5´ terminal position for the recognition of the 5´ end.

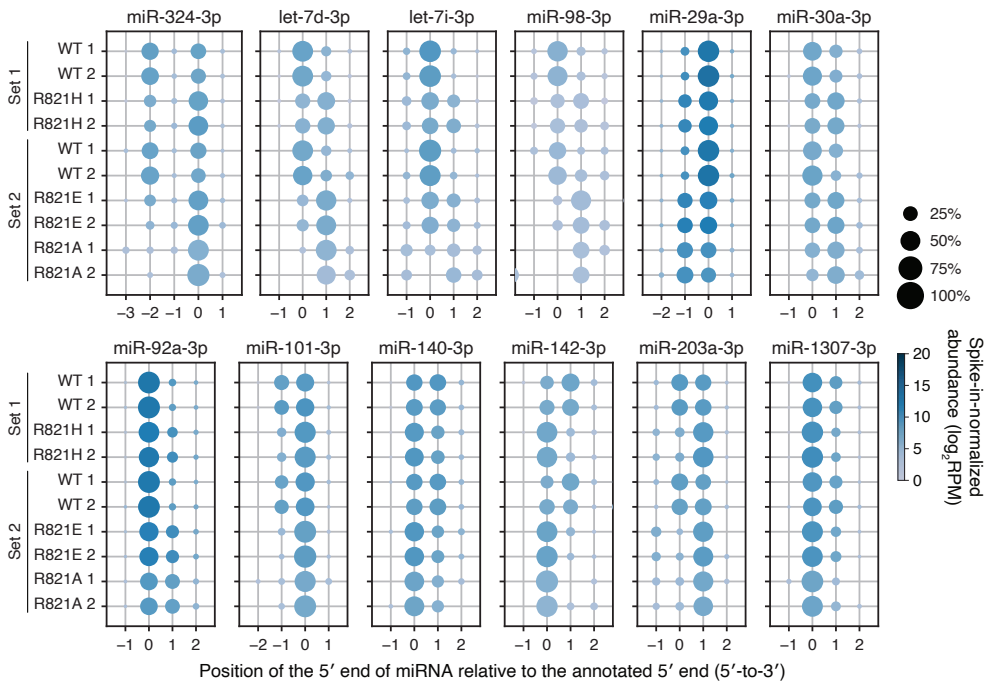

Figure 3.66 Examples of altered processing sites observed in the rescue experiments. Note that the DICER cleavage sites can be inferred from the 5´ end of 3p miRNAs. miRNA isoforms beginning at the indicated position are plotted with circles, with the size of the circle reflecting the proportion of the cleavage site usage at the given position.

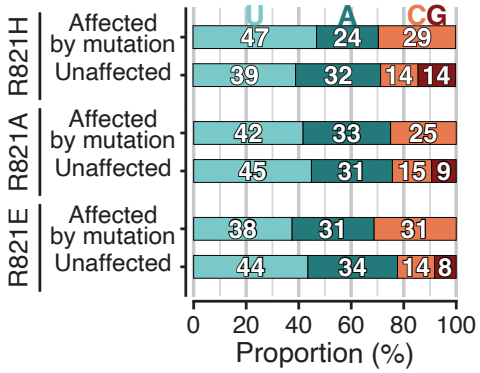

Figure 3.67 The 5´-terminal base identity of pre-miRNAs. Compare the miRNA groups whose major 5´-isomiRs were affected by the R821 mutation with those that are unaffected. Note that G is not found among the pre-miRNAs affected by R821 mutations. Affected miRNAs ($p < 0.05$, $\log_2$ fold change > 0.6 or < -0.6); Unaffected miRNAs ($p > 0.05$).

3.3 Discussion

We here captured the cryo-EM structure of hDICER-pre-let-7a-1^{GYM} in a cleavage-competent state. Our structure reveals how hDICER can achieve its substrate specificity via extensive contacts (Figure 3.68). Most notably, we discover the frayed 5´ terminal nucleotide other than G ('5´-H') is inserted into the 5´ pocket in the platform domain. Thermodynamically unstable 5´-H contributes to efficient and accurate processing. We also find that the dsRBD not only interacts with the RNA backbone but also actively engages in a sequence-specific interaction by recognizing the mismatch of the GYM motif. The dsRBD, via the flexible linker, may scan the upper stem of dsRNA, locate the motif, and facilitate the cleavage 2-nt away from the mismatch.

Combined with the apo and pre-dicing states (Liu et al., 2018), our dicing-state structure completes the structural landscape of the pre-miRNA processing cycle (Figure 3.68 for a cartoon model). In the apo state, hDICER exhibits a compact architecture in which the helicase domain, the PAZ helix, and the dsRBD limit the access of RNA to the catalytic valley. During the transition to pre-dicing

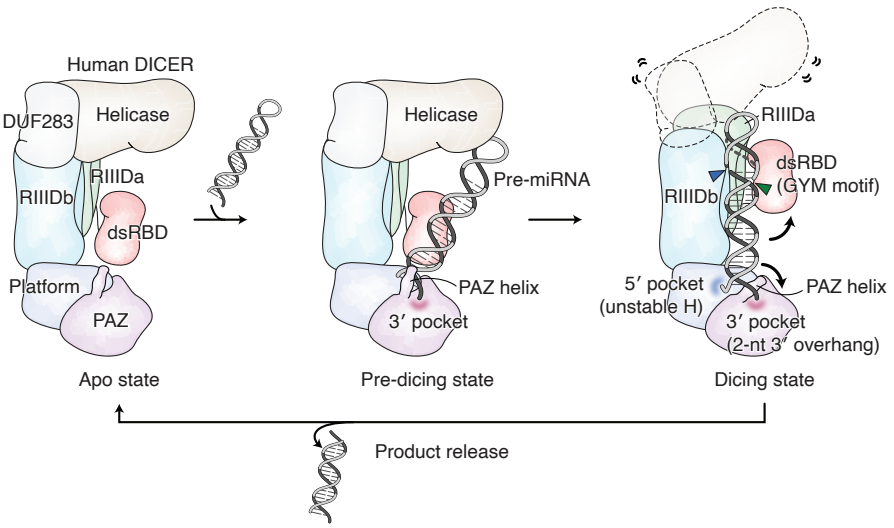

Figure 3.68 Model of the structural transition and substrate recognition of hDICER during the pre-miRNA processing cycle. In the apo state (this study, PDB: 5ZAK16), the N-terminal helicase domain, the PAZ helix, and the C-terminal dsRBD are in confor mations that block the entrance of the pre-miRNA substrate. In the pre-dicing state (PDB: 5ZAL), the 3´ end of the pre-miRNA is stably inserted into the 3´ pocket and the terminal loop contacts the helicase domain while no major conformational changes occur, limiting the access of RNA to the catalytic center. In transition to the dicing state, three major conformational changes occur, with the helicase domain in highly flexible conformations, the PAZ helix reorienting to allow a simultaneous recognition of the 5´ and 3´ termini, and the dsRBD swinging out to accommodate pre-miRNA in the catalytic center. After cleavage, the product (miRNA duplex) is released and the enzyme returns to the original closed conformation. Note that the cis-elements are recognized by the specific domains of hDICER to allow RNA binding and dictate cleavage sites: the GYM motif by the dsRBD and RIIIDa, the thermo dynamically unstable H by the 5´ pocket, and the 2-nt 3´ overhang by the 3´ pocket.

state, the protein conformation remains largely unchanged, creating only limited protein-RNA contacts through the 3´ pocket, the outer surface of the dsRBD, and the helicase domain (Liu et al., 2018). Next, the transition from the pre-dicing to dicing state induces major changes in the helicase domain, the PAZ helix, and the dsRBD to fully dock and poise pre-miRNA for cleavage within the catalytic center. After cleavage, the product is subsequently released from DICER, and transferred to AGO. Understanding of the AGO loading step and the role of TRBP in the process warrants further structural investigation.

Comparison between Dicer homologs provide valuable insights into the mechanism and evolution of eukaryotic small RNA pathways. Some primitive homologs such as Giardia Dicer do not have the helicase, the PAZ helix, or the dsRBD, implying that these domains may have emerged relatively recently for regulatory purposes. The helicase domains of Dicer homologs seem to be particularly diverse even among higher eukaryotes. Our structure shows that the helicase domain of hDICER becomes largely flexible in a dicing state, potentially serving as the regulatory barrier that contributes to substrate specificity. The helicase domain is also known to interact with TRBP, but the mutations or deletions of the helicase domain do not impair or even enhance the dicing activity, suggesting an autoinhibitory function (Ma et al., 2008; Soifer et al., 2008). Unlike Dicer homologs processively cleaving long dsRNAs using an ATP-dependent translocation by the helicase domain, such as *Drosophila* Dicer-2 and plant Dicer-like proteins (Zamore et al., 2000; Nykanen et al., 2001) animal Dicer whose main substrates are short hairpin-shaped pre-miRNAs, do not require ATP and cannot perform processive dicing on long dsRNAs (Zhang et al., 2002; Feng et al., 2012) (Figure 3.69). Based on sequence similarity, vertebrate Dicer proteins form a separate group from the other Dicer homologs in the miRNA pathway (such as fly Dcr-1) as well as those in the long dsRNA pathway (including fly Dcr-2), suggesting their unique properties (Figure 3.70). hDICER prefers hairpins with a flexible terminal loop but it can also process long dsRNAs unlike fly Dcr-1 which is highly specialized in pre-miRNA processing (Tsutsumi et al., 2011; Gu et al., 2012; Nguyen et al., 2022; Zapletal et al., 2022). Together with these earlier

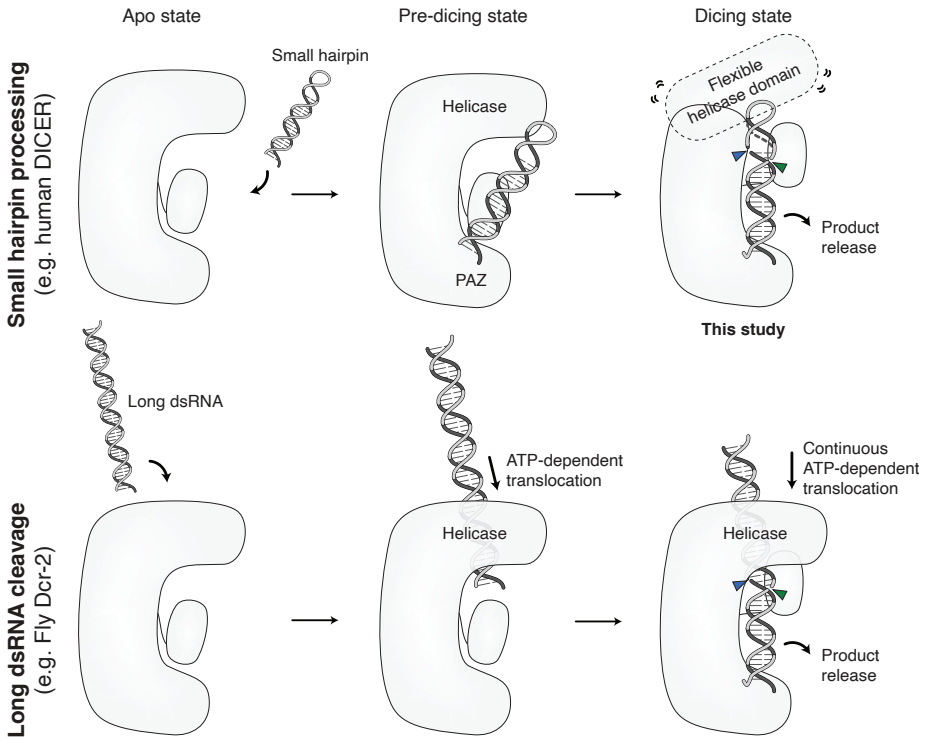

Figure 3.69 Comparison of the substrate RNA movement during DICER processing between two small RNA pathways. In the miRNA pathway, a hairpin-shaped small RNA (pre-miRNA) is bound to DICER by the helicase and PAZ domains. For cleavage, the helicase domain becomes flexible to accomodate the pre-miRNA into the catalytic center. In contrast, in the siRNA pathway, a long dsRNA comes into DICER by passing through the helicase domain. The ATP-dependent translocation by the helicase domain leads to processive cleavage of long dsRNAs.

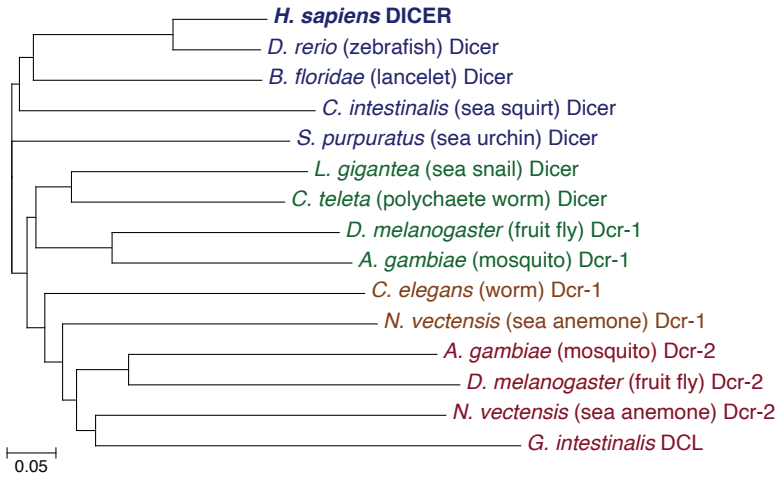

Figure 3.70 A phylogenetic tree of Dicer homologs. The scale bar indicates the length for the indicated frequency of amino acid variation.

observations, our current structure suggests that the helicase domain of hDICER may contribute to processing by interacting with a single-stranded RNA region at the initial binding step, but is not required for the catalytic step and needs to move flexibly to transit to an open conformation. Also supporting this evidence is the recent structure of oocyte-specific mouse ΔHEL1 Dicer isoform in a dicing state, whose helicase and DUF283 domains were difficult to be modeled due to flexibility (Zapletal et al., 2022). Recent structures of *Drosophila* Dcr-1 and its partner Loqs-PB (Jouravleva et al., 2022) shows an overall similar configuration to that of hDICER, but the helicase part seems to be more rigid and it is unclear if there is any base-specificity in RNA recognition. Together, these structures suggest conserved yet unique features of vertebrate Dicer proteins which have evolved to act mainly in the miRNA pathway while still maintaining a role in the siRNA pathway.

Collectively, our work on hDICER will provide practical benefits to the rational design of shRNAs and Dicer substrate RNAs (DsiRNAs) to improve RNA interference (Paddison et al., 2002; Amarzguioui et al., 2006; Kim et al., 2005). Moreover, our structure offers a molecular framework for understanding

of DICER-related diseases in humans. Based on the structure, we identified hDICER mutations in the 5´ pocket, which had not been characterized previously. These mutations, R821H and R1003Q, reduce well-known tumor suppressive miRNAs (e.g. let-7 members, miR-324, and miR-142). Given that various DICER mutations have been described in cancers and in a genetic disorder called DICER1 syndrome (Hill et al., 2009; Witkowski et al., 2013; Wu et al., 2013; Seki et al., 2014; Rakheja et al., 2014; Torrezan et al., 2014), functional analyses based on the structure will help us understand how the mutations in various parts of hDICER differentially affect the miRNA population and shape the transcriptome in favor of tumorigenesis.

3.4 Methods

Plasmid construction

The DNA encoding human DICER (RefSeq NM_030621) was PCR-amplified by a human cDNA library and subcloned into either pX vector (for protein purification) containing N-terminal His_{10}-eYFP-SUMOstar-Strep or pCK vector (for ectopic expression in DICER KO cells) with its original CMV promoter replaced with PGK promoter. Site-directed mutagenesis was performed to introduce mutations for functional studies.

Cell culture, protein expression and purification

Suspension HEK293E cells were subcultured in a 37°C shaking incubator with humidified atmosphere of 8% CO_2, in Dulbecco's modified Eagle's medium (DMEM, WELEGENE) supplemented with 5% fetal bovine serum (FBS, WELEGENE). For a half-liter culture, 0.15 mg of the plasmid encoding the full-length human DICER, N-terminally tagged with His_{10}-eYFP-SUMOstar-Strep tag, was transiently transfected using 1.5 mg of linear polyethylenimine (PEI) and 1% DMSO. After transfection, the cells were incubated at 33°C.

The entire protein purification steps were carried out at 4°C. The cells were collected after 72 hours, washed with cold PBS, and resuspended in buffer A (100

mM Tris-HCl (pH 8.0), 150 mM NaCl, 10% glycerol, 2 mM "-mercaptoethanol (BME), 0.1 mM phenylmethylsulfonyl fluoride (PMSF)) supplemented with EDTA-free Pierce Protease Inhibitor Mini Tablets (Thermo Fisher Scientific), 20 µg mL^{-1} micrococcal nuclease, and 5 mM $CaCl_2$. Cells were subsequently sonicated and the lysate was clarified by centrifugation at 35,000 g for an hour. Clarified lysate was loaded onto a column packed with Ni-NTA Superflow resin (Qiagen) that was pre-equilibrated with buffer A. The resin with the bound protein was washed with 5 column volumes of buffer A supplemented with 40 mM imidazole. The protein was eluted with buffer A supplemented with 200 mM imidazole and incubated with SUMOstar protease (LifeSensors) at 4°C to cleave the N-terminal tag. The cleavage was confirmed by SDS-PAGE. The protein was loaded onto a column packed with Strep-Tactin Superflow (IBA Lifesciences) resin that was pre-equilibrated with buffer A. The resin was then washed with 5 column volumes of buffer A with 2 mM EDTA and subsequently with 5 column volumes of buffer A without EDTA. The protein was eluted with buffer A with 50 mM biotin. To remove uncleaved fusion protein, the eluate was loaded onto Ni-NTA Superflow resin (Qiagen). The unbound protein was concentrated in a 100 K molecular weight cut off Amicon Ultra-15 Centrifugal Filter Unit (Merck). The concentrated protein was subjected to size-exclusion chromatography on a Superose 6 Increase 5/150 GL (GE Healthcare) pre-equilibrated with 50 mM Tris (pH 8.0), 100 mM NaCl, 0.5 mM TCEP. For the apo structure, the protein was concentrated to ~1 mg ml^{-1}, snap frozen in liquid nitrogen, and stored at −80°C. The same purification method was used for mutant DICER proteins.

In vitro reconstitution of DICER and pre-miRNA complex

Synthetic pre-let-7a-1^{GYM} (Integrated DNA Technologies) (5´-UGAGGUAGUAG GUUGUAUCGCUUUAGGGUCACACCCACCACUGGGAGAUAGCCAUACAA UCUACUGUCUUUCU- 3´) was resuspended in 20 mM Tris (pH 7.5), 80 mM NaCl, 1 mM EDTA. The RNA was heated at 80°C for 5 min and annealed by slowly decreasing the temperature to 4°C at a rate of −1°C min^{-1}, Complex formation was carried out by mixing DICER and RNA at final concentrations

of 15 µM and 45 µM, respectively, in a buffer containing 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.5 mM TCEP, 3 mM $CaCl_2$. The mixture was incubated on ice for 30 min and loaded onto a Superose 6 Increase 5/150 GL (GE Healthcare) pre-equilibrated with 50 mM Tris (pH 8.0), 100 mM NaCl, 0.5 mM TCEP, 2 mM $CaCl_2$ The fractions containing the complex were pooled, snap frozen in liquid nitrogen, and stored at −80°C. The concentration of the complex was evaluated on a SDS-PAGE gel and 15% urea-polyacrylamide gel.

Cryo-EM specimen preparation and data collection

Apo-hDICER: An aliquot (2.5 µl) of ~1.6 µM apo-hDICER sample was applied to glow discharged 300-mesh UltraAuFoil R1.2/1.3 holey-gold grids (Quantifoil), blotted for 2s with blot force 5 at 15 °C in 100% humidity and vitrified using Vitrobot Mark IV (Thermo Fisher Scientific) at Center for Macromolecular and Cell imaging (CMCI) in Seoul National University (SNU). Total 5,416 movies (1,390 movies at 1.1 Å/pixel and 4,026 movies at 0.87 Å/pixel) were collected in electron-event representation (EER) mode with a defocus range of −1.0 to −2.75 µm on a 200 kV Glacios (Thermo Fisher Scientific, CMCI in SNU) equipped with a Falcon 4 direct electron camera. The movies at 0.87 Å/pixel were then resampled into 1.1 Å/pixel and combined for further process.

hDICER-pre-let-7a-1^{GYM}: An aliquot (2.5 µl) of ~1.0 µM DICER-pre-let-7a-1^{GYM} complex sample was applied to glow-discharged 300-mesh UltraAuFoil R1.2/1.3 holey-gold grids, blotted for 2s with blot force 5 at 15 °C in 100% humidity and vitrified using Vitrobot Mark IV in SNU CMCI. A total of 5,964 micrographs were collected at a magnification of ×105,000 (corresponding to a calibrated pixel size of 0.849 Å) and a defocus range of −0.9 to −2.2 µm on a 300kV Titan Krios (Thermo Fisher Scientific, Institute of Basic Science, Daejeon, Korea) equipped with a Gatan K3 BioQuantum (Gatan) detector.

hDICER-pre-let-7a-1^{GYM} + Mg^{2+}: ~1.0 µM DICER-pre-let-7a-1^{GYM} complex was incubated with $MgCl_2$ added at a final concentration of 2 mM for >10 min at room temperature. The cryo grid was prepared as described above for

hDICER-pre-let-7a-1^{GYM}. Total 2,882 movies at at 1.1 Å/pixel were collected in electron-event representation (EER) mode with a defocus range of −1.1 to −2.3 µm on a 200 kV Glacios (Thermo Fisher Scientific, CMCI in SNU) equipped with a Falcon 4 direct electron camera.

Data processing and 3D refinement

All image processing was done in cryoSPARC v3.2 (Punjani et al., 2017). Computing resources were utilized in CMCI at SNU.

Apo-hDICER: Movies were aligned in 5 x 5 patches in MotionCor2 (Zheng et al., 2017), and CTF parameters were estimated with GCTF (Zhang, 2016) Utilizing template-based autopicking in cryoSPARC, raw particles were initially picked and extracted. After 2D classification and removing bad particles, 128,635 particles were subjected to 3D heterogeneous refinement using ab initio model in cryoSPARC. After further 3D classification, global CTF refinements and non-uniform refinement were performed using 128,635 particles yielding a 4.04 Å map of apo-hDICER based on the gold-standard Fourier shell correlation (FSC) at 0.143.

hDICER-pre-let-7a-1^{GYM}: Motion correction and CTF estimation was performed in the same process of apo-hDICER. For further automated particle-picking process, an initial particle template was roughly generated based on manually picked 189 particles. Using 3,762,043 particles from an initial template-based picking method, 2D classification and 2D selection was performed. After excluding bad particles and a heterogeneous portion especially containing partial flexible helicase domains, 1,386,301 particles were used for ab initio modeling and 3D classification yielding a homogeneous particle population. Using non-uniform refinement and local refinement, hDICERpre-let-7a-1 was reconstructed at a resolution of 3.04 Å map based on the gold-standard FSC at 0.143.

hDICER-pre-let-7a-1^{GYM} + Mg^{2+}: Motion correction and CTF estimation was performed in the same process of hDICER-pre-let-7a-1^{GYM}. An initial particle template was roughly generated based on manually picked 87 particles, and

2,857,379 particles were automatedly picked by Blob tuner with blob diameter of 30 Å to 300 Å. For further automated particle-picking process, an initial particle template was roughly generated based on the manually picked 189 particles. After excluding bad particles and a heterogeneous portion especially containing partial flexible helicase domains through 2D classification and 2D selection, 279,763 particles were used for ab initio modeling and 3D classification yielding a homogeneous particle population. Further subclassification was performed by Hetero refinement based on 3D templates generated with structural models of apo-like and dicing state-like DICER in this study.

Model building

Apo-hDICER: Model-building started from an initial protein model from the previously reported hDICER-TRBP structure (PDB: 5ZAK) (Liu et al., 2018). The initial model was fitted into the density map by Dock in map tool from Phenix v1.18.1 and manually refined in Coot (Liebschner et al., 2019; Emsley et al., 2010.) Then, the model was refined on the Namdinator server using MDFF and Phenix real space refinement default options (Kidmose et al., 2019). After a few rounds, the model was further corrected using Phenix and Coot (Afonine et al., 2018).

hDICER-pre-let-7a-1^{GYM}: An initial protein model from our apo-hDICER structure and an initial pre-let-7a-1 model from previously reported hDICER structure (PDB: 5ZAL) (Liu et al., 2018) were rigidly fit into the density by rigid body fitting with Fit in Map tool from Chimera v1.14. This fitted model was inspected and manually adjusted in Coot. Further refined with phenix.real_space_refine in Phenix, ISOLDE v1.1.0.

All models were validated by phenix.validation_cryoem (Afonine et al., 2018). All figures in the manuscripts were illustrated by ChimeraX (Goddard et al., 2018) and PyMol (Schrödinger, LCC). For the residual validation analysis, Q-scores for each residue were derived from MapQ of Segger tool (Pintilie et al., 2020) plugged

in Chimera v1.15. B-factor values were derived from real space refinement in Phenix ISOLDE v1.1.0.

in vitro DICER processing assay

All of the RNA oligos described below were chemically synthesized (Integrated DNA Technologies) and gel-purified before use. To prepare pre-miRNA substrates with 1-nt, 2-nt and 3-nt 3′ overhang lengths, RNA oligos (5′-UGAGGUAGUAGG UUGUAUAGUUUUAGGGUCACACCCACCACUGGGAGAUAACUAUACAAU CUACUGUCUUUC(U)(U)-3′) were radiolabeled at their 5′ ends with γ-ATP by T4 polynucleotide kinase (Takara). The RNA oligos were then purified using Oligo Clean & Concentrator (Zymo Research) according to the manufacturer's instructions. The eluted RNAs were annealed in a buffer containing 20 mM Tris (pH 7.5), 80 mM NaCl, 1 mM EDTA, by slowly decreasing the temperature to 4°C at a rate of -1°C min^{-1}. To prepare pre-miRNA-like duplex RNA substrates, RNA oligos of the 5′-arm with varying 5′ terminal base (5′-(N)GAGGUAGUAGGUUGUAAGUAGAAAGGACA AAGAG-3′) were radiolabeled, purified, and annealed to the complementary RNA oligos of the 3′-arm (5′-CUCUUUGUCCAAACUACUUACAACCUACUACCUUAUU(U)-3′ as described above.

The RNA substrates were cleaved with the purified DICER proteins added at a final concentration of 0.5 µM, in a buffer containing 50 mM Tris (pH 7.5), 100 mM NaCl, 0.5 mM TCEP, 2 mM $MgCl_2$ 5 µg ml^{-1} yeast RNA. After incubation at 37°C, the reaction was halted by addition of an equivalent volume of 2X RNA Loading Dye (NEB) supplemented with 0.5 mg ml^{-1} proteinase K (Roche). The RNA was then resolved on a 15% urea-polyacrylamide gel, along with 5′-end-radiolabeled synthetic miRNAs and Decade Markers System (Ambion). The product was visualized by phosphorimaging with Typhoon FLA 7000 (GE Healthcare).

DICER rescue experiments and data analysis

DICER KO HCT116 was cultured in McCoy's 5A media (WELGENE), supplemented with 10% FBS (WELGENE) (Kim et al., 2016). Cells were authenticated

by short tandem repeat profiling (ATCC). To ectopically express DICER proteins, the cells were transiently transfected with DICER-expressing vectors using FuGENE HD (Promega). The cells were treated with TRIzol (Thermo Fisher) 24/48 hrs post-transfection to extract total RNAs. For library construction, AQ-seq was performed following the protocol (Kim et al., 2019), except that 100 µg of total RNAs were used and that NEBNext Ultra II Q5 Master Mix (NEB) was used to amplify the cDNAs. The libraries were sequenced on the NovaSeq 6000 platform.

Data was preprocessed as previously described (Kim et al., 2019). Briefly, the 3´ adapter with 5´ 4-nt degenerate sequences was removed using cutadapt75 and FASTX-Toolkit (http://hannonlab.cshl. edu/fastx toolkit/). Next, short, low-quality, and artifact reads were filtered out using FASTX-Toolkit. The output reads were mapped to the spike-in reference first and then the unmapped reads were aligned to the human genome (hg38) using BWA (Li & Durbin, 2009b). miRNA annotations corresponding to each alignment were retrieved with miRBase release 21 using the intersect tool in BEDTools (Kozomara & Griffiths-Jones, 2014b).

	#1 apo-hDICER (EMDB-33490) (PDB 7XW3)	#2 hDICER-let7a-1^{GYM} (EMDB-33489) (PDB 7XW2)
Data collection and processing		
Magnification	120,000	105,000
Voltage (kV)	200	300
Electron exposure (e–/Å^2)	44.163	40
Defocus range (μm)	-1.0 ~ -2.75	-0.9 ~ -2.2
Pixel size (Å)	1.102	0.849
Symmetry imposed	C1	C1
Initial particle images (no.)	189,908	1,386,301
Final particle images (no.)	128,635	1,210,874
Map resolution (Å)	4.04	3.04
FSC threshold	0.143	0.143
Map resolution range (Å)	3.0-10.0	2.5-7.5
Refinement		
Initial model used (PDB code)	5ZAK	7XW3
Model resolution (Å)	4.04	4.03
FSC threshold	0.143	0.143
Model resolution range (Å)	3.0-10.0	2.5-7.5
Map sharpening B factor (Å^2)	-120.8	-157.2
Model composition		
Non-hydrogen atoms	12087	6997
Protein residues	1502	725
Nucleotides	-	54 (RNA)
Ligands	--	2 (Ca)
B factors (Å^2)		
Protein	98.43	74.38
Nucleotides	--	83.33 (RNA)
Ligands	--	30.00 (Ca)
R.m.s. deviations		
Bond lengths (Å)	0.009	0.007
Bond angles (°)	1.238	1.058
Validation		
MolProbity score	2.49	1.77
Clashscore	15.50	7.50
Poor rotamers (%)	2.44	0.92
Ramachandran plot		
Favored (%)	91.54	95.10
Allowed (%)	8.39	4.90
Disallowed (%)	0.07	0.00

Table 3.1 Cryo-EM data collection, refinement and validation statistics.

4. Conclusion

This study uniquely offers mechanistic insights into how metazoan DICER achieves stringent substrate specificity towards pre-miRNAs and contributes to miRNA biogenesis. Previous studies focused on the role of secondary features of pre-miRNAs in DICER processing, leading to the prevailing notion that DICER relies primariliy on the 2-nt 3´ overhang to measure ~22-nt from the terminal end created by DROSHA to determine cleavage sites. The sequence of a miRNA was hence thought to be predetermined at the DROSHA processing step, with DICER simply playing a passive role. However, we found evidence suggesting that there is yet unknown mechanism by which DICER determines cleavage sites. By massively parallel assays of randomized pre-miRNA variants, we revealed that DICER recognizes another cis-acting element, named the GYM motif, which can play a dominant role in terms of both cleavage efficiency and accuracy. This newly identified mechanism of DICER is functionally important as it can compensate the imprecise cleavage of pri-miRNAs by DROSHA and 3´ terminal modifications of pre-miRNAs by enzymes (either nucleotidyl addition or trimming)—which alter the terminal structures of pre-miRNAs the end counting rules rely on. Our findings show that DICER actively engages in cleavage site determination mediated by the GYM motif, which adds an additional layer of miRNA biogenesis that shapes the seed and strand seletion of miRNAs—exerting a profound effect on target recognition. The GYM motif is recognized by DICER dsRBD and more specifically by an arginine residue (R1855). Accordingly, deletion of dsRBD and R1855 affects many miRNAs in terms of both abundance and accuracy in a GYM-motif-dependent fashion. We also found that more than 2/3 of endogenous miRNAs benefit from the 'prominent' GYM motif in human and other eumetazoan species, suggesting that the GYM motif is deeply rooted in eumetazoan evolution. Importantly, newly identified principles apply not only to miRNA but also to other

small RNAs such as shRNA and siRNA that are widely used for loss-of-function studies and RNA therapeutics. Taken together, our study unveils a dominant mechanism that DICER utilizes for miRNA production, which is deeply rooted in metazoan lineage.

This study also reports the first active-state structure of human DICER in complex with pre-miRNA. The high-resolution structure allows detailed inspection of RNA-DICER interaction for the first time. We found that DICER undergoes large conformational changes to accommodate pre-miRNA in a fixed position. Importantly, it revealed that R1855 of DICER dsRBD recognizes the newly identified GYM motif via hydrogen bonding interactions. The PAZ helix of DICER also plays an important role in pre-miRNA processing through electrostatic interactions. Our structure also reveals that the architecture of the 5´ pocket allows accomodation of all bases except for guanine. Consistently, 5´-guanine impairs 5´ counting mechanism by DICER. Our study also show sthat cancer-associated mutations in the 5´ pocket significantly affects miRNA biogenesis in cells. This study provides a framework for the design of RNAi therapeutics and for an understanding of cancers and genetic disorders caused by mutations in the human DICER gene.

국문초록

다이서의 마이크로RNA 생합성에 대한 분자생물학적 연구

마이크로 RNA는 세포내의 비번역 RNA 중에서도 가장 짧은 RNA 로서 전령 RNA에 염기서열 특이적으로 결합하여 유전자 발현을 조절하는 필수적인 역할을 수행하는 RNA입니다. DICER는 마이크로 RNA를 생성하는 데 매우 중요한 효소로서, 마이크로 RNA 전구체를 약 22개의 염기서열로 효율적이고 정확하고 절단하는 기능을 가지고 있습니다. DICER는 마이크로RNA의 말단과 같은 2차구조에 의존하여 절단위치를 결정한다고 알려져 있습니다.하지만 최근 연구에서 DICER가 2차 구조뿐만 아니라 마이크로RNA줄기의 구조, 염기서열을 특이적으로 인지하여 절단할 가능성을 보았습니다. 이를 바탕으로 본 연구에서는 DICER가 마이크로 RNA의 어떠한 염기서열과 구조를 특이적으로 인지하는지 분자적인 수준에서 주도 면밀하게 알아보고자 대용량 차세대 염기서열 분석을 이용하였습니다. 이를 이용하여 DICER의 구조 및 서열 특이적 마이크로RNA절단 기작을 규명하였습니다. 이 특이적 서열을 'GYM' 모티프라고 명명하였고, 구조 모델링과 생화학적 실험, 그리고 세포실험을 통해서 이중 가닥RNA와 결합하는 DICER의 도메인 (dsRBD)과 GYM 모티프의 결합이 마이크로RNA 절단 효율과 정밀성에 매우 중요한 역할을 한다는 것을 알 수 있었습니다. 이러한 기작은 진정후생동물에서 깊게 보존되어 있음을 알 수 있습니다. 또한 GYM모티프가 유전자 발현 억제 기술로 사용되고 있는 short hairpin RNA와 Dicer-substrate siRNA에 응용될 수 있음을 보여주었습니다.

다음으로, DICER가 어떠한 기작으로 마이크로 RNA의 염기서열을 특이적으로 인지하여 절단하는 지에 구조적 연구가 매우 부족하고, 따라서 암환자에서 관찰되는 DICER의 유전적 변이가 어떠한 상호작용 을 통해 발달과 질병 등의 생명 현상에 영향을 미치는지에 대한 연구가 결여되어 있어, DICER와 마이크로 RNA 복합체의 구조를 초저온전자현미경으로 규명하였습니다. 높은 해상도의 구조로 RNA와 DICER의 상호작용을 면밀히 조사하였고, DICER가 마이크로 RNA 전구체를 절단

하기위해서는 여러 도메인의 움직임이 동반되어야 한다는 사실을 알 수 있었습니다. 그 중에서도 dsRBD가 GYM 모티프 근처에 위치하여 수소결합을 하여 절단 위치를 결정할 수 있음을 관찰하였습니다. 중요하게도, DICER가 마이크로RNA의 5′ 말단을 서열 특이적으로 인지함을 알 수 있었고, 특히 5′ 말단이 구아닌일 때 절단 정밀도가 떨어지는 것을 발견하였습니다. 이렇게 5′ 말단의 서열을 인지하는 아미노산이 암환자에게서 관찰된다는 사실을 알 수 있었고, 이는 세포내에서 전반적인 마이크로 RNA 생합성에 크게 영향을 줄 뿐만 아니라, 암억제에서 기능한다고 알려져 있는 중요한 마이크로RNA의 양을 크게 줄인다는 사실을 알 수 있었습니다.

종합하면, 본 연구는 DICER가 마이크로 RNA를 GYM 모티프를 이용하여 절단 위치를 결정해, 마이크로 RNA 기능을 정하는데 큰 역할을 한다는 것을 알 수 있었습니다. 또한 DICER와 마이크로 RNA의 복합체의 고해상도 구조를 통해, 새롭게 발견된 기작의 중요성과 DICER의 변이가 어떻게 암을 유발할 수 있는 지를 밝혔습니다.

Bibliography

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., & Adams, P. D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol*, 74(Pt 6), 531–544.

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

Auyeung, V. C., Ulitsky, I., McGeary, S. E., & Bartel, D. P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*, 152(4), 844–58.

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

Bellaousov, S., Reuter, J. S., Seetin, M. G., & Mathews, D. H. (2013). RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res*, 41(Web Server issue), W471–4.

Bofill-De Ros, X. & Gu, S. (2016). Guidelines for the optimal design of miRNA-based shRNAs. *Methods*, 103, 157–66.

Bofill-De Ros, X., Kasprzak, W. K., Bhandari, Y., Fan, L., Cavanaugh, Q., Jiang, M., Dai, L., Yang, A., Shao, T. J., Shapiro, B. A., Wang, Y. X., & Gu, S. (2019). Structural Differences between Pri-miRNA Paralogs Promote Alternative Drosha Cleavage and Expand Target Repertoires. *Cell Rep*, 26(2), 447–459 e4.

Bogerd, H. P., Whisnant, A. W., Kennedy, E. M., Flores, O., & Cullen, B. R. (2014). Derivation and characterization of Dicer- and microRNA-deficient human cells. *RNA*, 20(6), 923–37.

Burroughs, A. M., Ando, Y., de Hoon, M. J., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y., & Daub, C. O. (2010). A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res*, 20(10), 1398–410.

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

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, 2(5), 401–4.

Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A., & Doudna, J. A. (2010). Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J Mol Biol*, 404(3), 392–402.

Chen, C.-Y. A. & Shyu, A.-B. (2011). Mechanisms of deadenylation-dependent decay. *Wiley interdisciplinary reviews. RNA*, 2, 167–183.

Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. C., Spies, N., Baek, D., Johnston, W. K., Russ, C., Luo, S., Babiarz, J. E., Blelloch, R., Schroth, G. P., Nusbaum, C., & Bartel, D. P. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev*, 24(10), 992–1009.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., & Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014), 231–5.

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

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4), 486–501.

Fang, W. & Bartel, D. P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. *Mol Cell*, 60(1), 131–45.

Fang, W. & Bartel, D. P. (2020). MicroRNA Clustering Assists Processing of Suboptimal MicroRNA Hairpins through the Action of the ERH Protein. *Molecular cell*, 78, 289–302.e6.

Feng, Y., Zhang, X., Graves, P., & Zeng, Y. (2012). A comprehensive analysis of precursor microRNA cleavage by human Dicer. *RNA*, 18(11), 2083–92.

Frank, F., Sonenberg, N., & Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 465(7299), 818–22.

Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., King, B. L., Newcomb, J. M., Sempere, L. F., Flatmark, K., Hovig, E., & Peterson, K. J. (2015). A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annual review of genetics*, 49, 213–242.

Fromm, B., Domanska, D., Hoye, E., Ovchinnikov, V., Kang, W., Aparicio-Puerta, E., Johansen, M., Flatmark, K., Mathelier, A., Hovig, E., Hackenberg, M., Friedlander, M. R., & Peterson, K. J. (2020). MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Res*, 48(D1), D1172.

Gan, J., Tropea, J. E., Austin, B. P., Court, D. L., Waugh, D. S., & Ji, X. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, 124(2), 355–66.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, 6(269), pl1.

Gao, Z., Herrera-Carrillo, E., & Berkhout, B. (2018). Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. *Mol Ther Nucleic Acids*, 10, 36–44.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*, 27(1), 14–25.

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., & Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014), 235–40.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P. N., & Kay, M. A. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell*, 151(4), 900–911.

Han, C., Liu, Y., Wan, G., Choi, H. J., Zhao, L., Ivan, C., He, X., Sood, A. K., Zhang, X., & Lu, X. (2014). The RNA-binding protein DDX1 promotes primary microRNA maturation and inhibits ovarian tumor progression. *Cell Rep*, 8(5), 1447–60.

Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., & Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev*, 18(24), 3016–27.

Heo, I., Ha, M., Lim, J., Yoon, M. J., Park, J. E., Kwon, S. C., Chang, H., & Kim, V. N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell*, 151(3), 521–32.

Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., Yeom, K. H., Han, J., & Kim, V. N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell*, 138(4), 696–708.

Hill, D. A., Ivanovich, J., Priest, J. R., Gurnett, C. A., Dehner, L. P., Desruisseau, D., Jarzembowski, J. A., Wikenheiser-Brokamp, K. A., Suarez, B. K., Whelan, A. J., Williams, G., Bracamontes, D., Messinger, Y., & Goodfellow, P. J. (2009). DICER1 mutations in familial pleuropulmonary blastoma. *Science*, 325(5943), 965.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., & Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, 293(5531), 834–8.

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

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.

Jin, W., Wang, J., Liu, C. P., Wang, H. W., & Xu, R. M. (2020). Structural Basis for pri-miRNA Recognition by Drosha. *Mol Cell*, 78(3), 423–433 e5.

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., & Slack, F. J. (2005). RAS is regulated by the let-7 microRNA family. *Cell*, 120(5), 635–47.

Jonas, S. & Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nature reviews. Genetics*, 16, 421–433.

Jouravleva, K., Golovenko, D., Demo, G., Dutcher, R. C., Hall, T. M. T., Zamore, P. D., & Korostelev, A. A. (2022). Structural basis of microRNA biogenesis by Dicer-1 and its partner protein Loqs-PB. *Mol Cell*, 82(21), 4049–4063 e6.

Kawamata, T. & Tomari, Y. (2010). Making RISC. *Trends in biochemical sciences*, 35, 368–376.

Khvorova, A., Reynolds, A., & Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115(2), 209–16.

Kidmose, R. T., Juhl, J., Nissen, P., Boesen, T., Karlsen, J. L., & Pedersen, B. P. (2019). Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ*, 6(Pt 4), 526–531.

Kidwell, M. A., Chan, J. M., & Doudna, J. A. (2014). Evolutionarily conserved roles of the dicer helicase domain in regulating RNA interference processing. *The Journal of biological chemistry*, 289, 28352–28362.

Kim, B., Jeong, K., & Kim, V. N. (2017). Genome-wide Mapping of DROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates. *Mol Cell*, 66(2), 258–269 e5.

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

Kim, H., Kim, J., Kim, K., Chang, H., You, K., & Kim, V. N. (2019). Bias-minimized quantification of microRNA reveals widespread alternative processing and 3' end modification. *Nucleic Acids Res*, 47(5), 2630–2640.

Kim, H., Kim, J., Yu, S., Lee, Y. Y., Park, J., Choi, R. J., Yoon, S. J., Kang, S. G., & Kim, V. N. (2020). A Mechanism for microRNA Arm Switching Regulated by Uridylation. *Mol Cell*, 78(6), 1224–1236 e5.

Kim, K., Baek, S. C., Lee, Y. Y., Bastiaanssen, C., Kim, J., Kim, H., & Kim, V. N. (2021). A quantitative map of human primary microRNA processing sites. *Mol Cell*, 81(16), 3422–3439 e11.

Kim, Y., Yeo, J., Lee, J. H., Cho, J., Seo, D., Kim, J. S., & Kim, V. N. (2014). Deletion of human tarbp2 reveals cellular microRNA targets and cell-cycle function of TRBP. *Cell Rep*, 9(3), 1061–74.

Kim, Y. K., Kim, B., & Kim, V. N. (2016). Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc Natl Acad Sci U S A*, 113(13), E1881–9.

Kozomara, A. & Griffiths-Jones, S. (2014a). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 42(Database issue), D68–73.

Kozomara, A. & Griffiths-Jones, S. (2014b). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 42(Database issue), D68–73.

Kwon, S. C., Baek, S. C., Choi, Y. G., Yang, J., Lee, Y. S., Woo, J. S., & Kim, V. N. (2019). Molecular Basis for the Single-Nucleotide Precision of Primary microRNA Processing. *Mol Cell*, 73(3), 505–518 e5.

Kwon, S. C., Jang, H., Shen, S., Baek, S. C., Kim, K., Yang, J., Kim, J., Kim, J.-S., Wang, S., Shi, Y., Li, F., & Kim, V. N. (2020). ERH facilitates microRNA maturation through the interaction with the N-terminus of DGCR8. *Nucleic acids research*, 48, 11097–11112.

Kwon, S. C., Nguyen, T. A., Choi, Y.-G., Jo, M. H., Hohng, S., Kim, V. N., & Woo, J.-S. (2016). Structure of Human DROSHA. *Cell*, 164, 81–90.

Lau, P. W., Guiley, K. Z., De, N., Potter, C. S., Carragher, B., & MacRae, I. J. (2012). The molecular architecture of human Dicer. *Nat Struct Mol Biol*, 19(4), 436–40.

Lau, P. W., Potter, C. S., Carragher, B., & MacRae, I. J. (2009). Structure of the human Dicer-TRBP complex by electron microscopy. *Structure*, 17(10), 1326–32.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75, 843–854.

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

Lee, Y., Jeon, K., Lee, J. T., Kim, S., & Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J*, 21(17), 4663–70.

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

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7), 787–98.

Li, H. & Durbin, R. (2009a). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–60.

Li, H. & Durbin, R. (2009b). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–60.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., & Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, 75(Pt 10), 861–877.

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

Liu, Z., Wang, J., Cheng, H., Ke, X., Sun, L., Zhang, Q. C., & Wang, H. W. (2018). Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate. *Cell*, 173(6), 1549–1550.

Liu, Z., Wang, J., Li, G., & Wang, H. W. (2015). Structure of precursor microRNA's terminal loop regulates human Dicer's dicing activity by switching DExH/D domain. *Protein Cell*, 6(3), 185–93.

Lu, C., Tej, S. S., Luo, S., Haudenschild, C. D., Meyers, B. C., & Green, P. J. (2005). Elucidation of the small RNA component of the transcriptome. *Science (New York, N.Y.)*, 309, 1567–1569.

Ma, E., MacRae, I. J., Kirsch, J. F., & Doudna, J. A. (2008). Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol*, 380(1), 237–43.

Ma, E., Zhou, K., Kidwell, M. A., & Doudna, J. A. (2012). Coordinated activities of human dicer domains in regulatory RNA processing. *J Mol Biol*, 422(4), 466–76.

Macrae, I. J., Li, F., Zhou, K., Cande, W. Z., & Doudna, J. A. (2006a). Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb Symp Quant Biol*, 71, 73–80.

MacRae, I. J., Zhou, K., & Doudna, J. A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat Struct Mol Biol*, 14(10), 934–40.

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

Martin, M. (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet. journal*, 17.

Mosher, R. A., Schwach, F., Studholme, D., & Baulcombe, D. C. (2008). PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A*, 105(8), 3145–50.

Moss, E. G., Lee, R. C., & Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell*, 88, 637–646.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., & Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev*, 16(6), 720–8.

Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H., & Fukuhara, T. (2014). Distinct substrate specificities of Arabidopsis DCL3 and DCL4. *Nucleic Acids Res*, 42(3), 1845–56.

Nguyen, T. A., Jo, M. H., Choi, Y.-G., Park, J., Kwon, S. C., Hohng, S., Kim, V. N., & Woo, J.-S. (2015). Functional Anatomy of the Human Microprocessor. *Cell*, 161, 1374–1387.

Nguyen, T. D., Trinh, T. A., Bao, S., & Nguyen, T. A. (2022). Secondary structure RNA elements control the cleavage activity of DICER. *Nat Commun*, 13(1), 2138.

Nykanen, A., Haley, B., & Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, 107(3), 309–21.

Olsen, P. H. & Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental biology*, 216, 671–680.

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

Park, J. E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., Patel, D. J., & Kim, V. N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 475(7355), 201–5.

Partin, A. C., Zhang, K., Jeong, B. C., Herrell, E., Li, S., Chiu, W., & Nam, Y. (2020). Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. *Mol Cell*, 78(3), 411–422 e4.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., & Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408, 86–89.

Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M. F., & Chiu, W. (2020). Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat Methods*, 17(3), 328–334.

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

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, 14(3), 290–296.

Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–2.

Rakheja, D., Chen, K. S., Liu, Y., Shukla, A. A., Schmid, V., Chang, T. C., Khokhar, S., Wickiser, J. E., Karandikar, N. J., Malter, J. S., Mendell, J. T., & Amatruda, J. F. (2014). Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat Commun*, 2, 4802.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., & Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403, 901–906.

Ronemus, M., Vaughn, M. W., & Martienssen, R. A. (2006). MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by argonaute, dicer, and RNA-dependent RNA polymerase in Arabidopsis. *Plant Cell*, 18(7), 1559–74.

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H., & Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell*, 127, 1193–1207.

Sampson, E., Brierley, J. D., Le, L. W., Rotstein, L., & Tsang, R. W. (2007). Clinical management and outcome of papillary and follicular (differentiated) thyroid cancer presenting with distant metastasis at diagnosis. *Cancer*, 110(7), 1451–6.

Schirle, N. T. & MacRae, I. J. (2012). The crystal structure of human Argonaute2. *Science (New York, N.Y.)*, 336, 1037–1040.

Schirle, N. T., Sheu-Gruttadauria, J., & MacRae, I. J. (2014). Structural basis for microRNA targeting. *Science (New York, N.Y.)*, 346, 608–613.

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

Seki, M., Yoshida, K., Shiraishi, Y., Shimamura, T., Sato, Y., Nishimura, R., Okuno, Y., Chiba, K., Tanaka, H., Kato, K., Kato, M., Hanada, R., Nomura, Y., Park, M. J., Ishida, T., Oka, A., Igarashi, T., Miyano, S., Hayashi, Y., Ogawa, S., & Takita, J. (2014). Biallelic DICER1 mutations in sporadic pleuropulmonary blastoma. *Cancer Res*, 74(10), 2742–9.

Shang, R., Baek, S. C., Kim, K., Kim, B., Kim, V. N., & Lai, E. C. (2020). Genomic Clustering Facilitates Nuclear Processing of Suboptimal Pri-miRNA Loci. *Molecular cell*, 78, 303–316.e4.

Sinha, N. K., Iwasa, J., Shen, P. S., & Bass, B. L. (2018). Dicer uses distinct modules for recognizing dsRNA termini. *Science*, 359(6373), 329–334.

Snead, N. M., Wu, X., Li, A., Cui, Q., Sakurai, K., Burnett, J. C., & Rossi, J. J. (2013). Molecular basis for improved gene silencing by Dicer substrate interfering RNA compared with other siRNA variants. *Nucleic Acids Res*, 41(12), 6209–21.

Soifer, H. S., Sano, M., Sakurai, K., Chomchan, P., Saetrom, P., Sherman, M. A., Collingwood, M. A., Behlke, M. A., & Rossi, J. J. (2008). A role for the Dicer helicase domain in the processing of thermodynamically unstable hairpin RNAs. *Nucleic Acids Res*, 36(20), 6511–22.

Su, S., Wang, J., Deng, T., Yuan, X., He, J., Liu, N., Li, X., Huang, Y., Wang, H. W., & Ma, J. (2022). Structural insights into dsRNA processing by Drosophila Dicer-2-Loqs-PD. *Nature*, 607(7918), 399–406.

Suzuki, H. I., Katsura, A., Yasuda, T., Ueno, T., Mano, H., Sugimoto, K., & Miyazono, K. (2015). Small-RNA asymmetry is directly driven by mammalian Argonautes. *Nat Struct Mol Biol*, 22(7), 512–21.

Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y., Mitsudomi, T., & Takahashi, T. (2004). Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res*, 64(11), 3753–6.

Tian, Y., Simanshu, D. K., Ma, J. B., Park, J. E., Heo, I., Kim, V. N., & Patel, D. J. (2014). A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. *Mol Cell*, 53(4), 606–16.

Torrezan, G. T., Ferreira, E. N., Nakahata, A. M., Barros, B. D., Castro, M. T., Correa, B. R., Krepischi, A. C., Olivieri, E. H., Cunha, I. W., Tabori, U., Grundy, P. E., Costa, C. M., de Camargo, B., Galante, P. A., & Carraro, D. M. (2014). Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nat Commun*, 5, 4039.

Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., & Tomari, Y. (2011). Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat Struct Mol Biol*, 18(10), 1153–8.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., Hilbert, J. L., Bartel, D. P., & Crete, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell*, 16(1), 69–79.

Wang, H. W., Noland, C., Siridechadilok, B., Taylor, D. W., Ma, E., Felderer, K., Doudna, J. A., & Nogales, E. (2009). Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol*, 16(11), 1148–53.

Wang, Q., Xue, Y., Zhang, L., Zhong, Z., Feng, S., Wang, C., Xiao, L., Yang, Z., Harris, C. J., Wu, Z., Zhai, J., Yang, M., Li, S., Jacobsen, S. E., & Du, J. (2021). Mechanism of siRNA production by a plant Dicer-RNA complex in dicing-competent conformation. *Science*, 374(6571), 1152–1157.

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

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

Wilson, R. C., Tambe, A., Kidwell, M. A., Noland, C. L., Schneider, C. P., & Doudna, J. A. (2015). Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Mol Cell*, 57(3), 397–407.

Witkowski, L., Mattina, J., Schonberger, S., Murray, M. J., Choong, C. S., Huntsman, D. G., Reis-Filho, J. S., McCluggage, W. G., Nicholson, J. C., Coleman, N., Calaminus, G., Schneider, D. T., Arseneau, J., Stewart, C. J., & Foulkes, W. D. (2013). DICER1 hotspot mutations in non-epithelial gonadal tumours. *Br J Cancer*, 109(10), 2744–50.

Wu, H., Ye, C., Ramirez, D., & Manjunath, N. (2009). Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA. *PLoS One*, 4(10), e7566.

Wu, M. K., Sabbaghian, N., Xu, B., Addidou-Kalucki, S., Bernard, C., Zou, D., Reeve, A. E., Eccles, M. R., Cole, C., Choong, C. S., Charles, A., Tan, T. Y., Iglesias, D. M., Goodyer, P. R., & Foulkes, W. D. (2013). Biallelic DICER1 mutations occur in Wilms tumours. *J Pathol*, 230(2), 154–64.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E., & Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, 2(5), E104.

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

Zapletal, D., Taborska, E., Pasulka, J., Malik, R., Kubicek, K., Zanova, M., Much, C., Sebesta, M., Buccheri, V., Horvat, F., Jenickova, I., Prochazkova, M., Prochazka, J., Pinkas, M., Novacek, J., Joseph, D. F., Sedlacek, R., Bernecky, C., O'Carroll, D., Stefl, R., & Svoboda, P. (2022). Structural and functional basis of mammalian microRNA biogenesis by Dicer. *Mol Cell*, 82(21), 4064–4079 e13.

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

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

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol*, 193(1), 1–12.

Zhang, X., Henderson, I. R., Lu, C., Green, P. J., & Jacobsen, S. E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci U S A*, 104(11), 4536–41.

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*, 14(4), 331–332.

감 사 의 글

학위 과정을 마무리하며, 모든 과정을 걸어오면서 도움을 주신 수많은 분들께 받은 은혜와 감사를 담아 표현하고 싶습니다.제 졸업 논문의 완성은 많은 분들의 도움과 헌신 없이는 불가능했을 것입니다.

먼저, 학위 과정을 무사히 마칠 수 있도록 항상 저의 편에 서서 끊임없이 지도와 조언을 해주신 김빛내리 교수님께 깊은 감사의 마음을 표합니다. 교수님의 가르침은 항상 저에게 큰 영감과 동기부여를 주었고, 특히 어려운 문제에 부딪혔을 때, 교수님께서 항상 저에게 문제를 다르게 바라보고 해결책을 찾는 데 도움을 주셨습니다. 어려운 시기에도 저를 격려하시고, 저의 아이디어를 지지해주신 교수님께 감사의 인사를 드리며, 교수님의 가르침을 잊지 않고 평생에 걸쳐 간직하겠습니다.

그리고, 우재성 교수님의 세심한 지도와 열정적인 가르침에 깊은 감사의 말씀을 드립니다. 학위 과정 1년차 때 가르쳐 주신 것들이 6년 동안의 학위 과정에 많은 영향을 끼쳤습니다. 특히 가르쳐 주셨던 단백질 정제 기술이 제 학위 과정에 아주 큰 도움이 되었습니다. 감사합니다.

심사를 위해 시간을 내 주신 교수님들께도 깊은 감사를 드립니다. 백대현 교수님, 노성훈 교수님, 김진홍 교수님, 그리고 최정균 교수님의 꼼꼼한 지적과 조언 덕분에 졸업 논문을 잘 마무리할 수 있었습니다.

저의 실험실 생활에 도움을 주신 선배, 동료들에게도 감사드립니다. 인턴 면접을 본 날에 같이 얘기하고 조언해주신 영석이 형과 재철이 형, 그리고 인턴 생활 동안 열정적으로 가르쳐주신 권성철 박사님, 웨스턴 블랏 실험을 꼼꼼하게 알려주신 혜림 누나, 백규형, 현준 박사님이 인상적인 기억으로 남아있습니다. 실험실에 적응하는 데 옆자리에서 도움을 많이 준 기준이 형의 기억도 좋게 남아있습니다. 항상 웃으며 대해주신 윤석이 형, 지미 박사님, 진아 누나, 성률 박사님, 그리고 모르는 게 없는 연이 형과의 대화도 좋은 기억으로 남아있습니다. 그리고 같은 시기에 입학하고, 모르는 게 있으면 제일 먼저 물어볼 수 있었던 부연이와 수민이가 있어서 좋았습니다.

그리고 항상 질문하고 배우고 싶어하는 동완이 형도 여러모로 제게 정말 많은 도움을 준 사람입니다. 또한 오랜 시간 같이 마이크로 RNA를 연구하며 서로 조언해준 동료분들께도 감사 인사를 드리고 싶습니다. 모든 실험을 최적화하는 기준이 형, 드로샤를 사랑하는 승찬이 형, 실험을 꼼꼼하게 잘 하는 수민이, 혼자서도 묵묵히 자기 일을 잘하는 하림이가 기억에 남습니다. 그리고 구조 연구를 성공적으로 마칠 수 있게 도와주신 노성훈 교수님과 한솔 박사님께도 감사의 인사를 전합니다.

연구실에서의 생활은 새로운 지식뿐만 아니라, 함께하는 동료들과의 소중한 인연 덕분에 풍성했습니다. 제가 이 연구실에 들어왔을 때부터 여러 가지로 영향을 많이 준 해동이와의 많은 토론은 저의 연구를 다양한 시각에서 바라볼 수 있게 해주었고, 연구실 밖에서의 시간도 저에게 귀중한 추억입니다. 특히 해동이와 같이 두 연구를 수행하고 성공적으로 마무리할 수 있었던 것이 좋았습니다. 학위 과정 동안 연구는 같이 하지 못했지만, 저와 많은 시간을 같이 보낸 명환이도 학위 과정을 잘 마무리할 수 있었던 원동력입니다. 연구에 항상 최선을 다하고 주말에도 열심히 연구하는 명환이의 모습이 그리울 것 같습니다. 그 외에도 저를 항상 진심으로 응원해준 수진 누나, 영란 박사님도 제 학위 과정 동안 큰 힘을 주었습니다. 또한, 새로운 분야에서 빠르게 적응하신 연우박사님, MS 전문가 종우, 실험과 분석을 동시에 해내는 제니, 실험하는 체력이 대단한 성한이와 도일이, 분석을 잘 해내는 혜준이와 단백질 정제를 벌써 잘 해내는 민석이는 본받을 점이 많은 동료입니다.

그 외에도 실험 지원을 해주신 지혜 누나, 연길 누나, 은지 누나, 다은 누나, 선영 선생님과 행정팀 성미주 실장님, 은주 선생님, 효진 선생님께도 꼭 감사하다는 말씀 전하고 싶습니다.

마지막으로, 항상 저를 응원해주고, 제가 추구하는 것을 뒷받침해준 가족과 승미 누나에게 감사의 말씀을 전합니다. 이분들이 없었다면 저는 이 시점에 있지 못했을 것입니다.

이 논문을 통해 받은 모든 도움과 지지에 감사의 인사를 전합니다.