UNDERSTANDING RECOGNITION OF PRIMARY MICRORNAS BY

HUMAN DROSHA/DGCR8

APPROVED BY SUPERVISORY COMMITTEE

Yunsun Nam, Ph.D.

Jose Rizo-Rey, Ph.D.

Kim Orth, Ph.D.

Michael Rosen, Ph.D.

UNDERSTANDING RECOGNITION OF PRIMARY MICRORNAS BY

HUMAN DROSHA/DGCR8

by

ALEXANDER CHRISTIAN PARTIN

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May 2020

ABSTRACT

MicroRNAs (miRNAs) are small noncoding RNAs that regulate gene expression by suppressing messenger RNAs. The RNase Drosha, as part of the Microprocessor complex, performs the initial processing step in miRNA biogenesis. Drosha, in complex with its partner protein DGCR8, must distinguish its RNA substrates (called primary microRNAs) and identify the proper positions within these substrates to perform ribonuclease activity. My thesis work focuses on how the Drosha/DGCR8 complex recognizes primary microRNA structures and determines their processing sites, thereby regulating the initiation of microRNA biogenesis.

Past biochemical and structural studies have left numerous unanswered questions about pri-miRNA processing. Using an *in vitro* system with human Drosha and DGCR8, I pinpointed the role of heme binding in DGCR8-mediated regulation of pri-miRNA processing, and revealed that heme activates DGCR8 to preferentially bind terminal loop structures, thereby orienting Drosha at the proper position for productive processing. Drosha tends to favor improper binding/processing sites on many pri-miRNAs, and heme-activated DGCR8 serves to correct this erroneous behavior and enable pri-miRNA processing at the appropriate site. Using information gained from these studies, I optimized a stable Drosha/DGCR8/pri-miRNA complex, which I then used as part of a collaboration to determine atomic structures through cryo-electron microscopy (cryo-EM). My cryo-EM structures revealed major new insights into how Drosha and DGCR8 ensure that pri-miRNAs possess the requisite structural features. A rigid assembly of double-stranded RNA-binding domains (dsRBDs) establishes a measuring system for pri-

miRNA stem length. In addition, two newly identified regions of Drosha, the Wedge and the Belt, clamp over the basal end of bound RNAs and check for structural features required for efficient processing. In conjunction with the heme-binding region, which identifies the terminal loop, these components work together to recognize pri-miRNAs and enable processing. Overall, my work reveals critical information about how Drosha and DGCR8 work together to identify pri-miRNAs.

Prior Publications

1. **Partin AC**, Zhang K, Jeong BC, Herrell E, Li S, Chiu W, Nam Y. Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. (*currently under review*)

2. Ngo TD, **Partin AC,** Nam Y. RNA specificity and autoregulation of DDX17, a modulator of microRNA biogenesis. *Cell Rep*. 29(12):4024-4035 (2019)

3. **Partin AC**, Ngo TD, Herrell E, Jeong BC, Hon G, Nam Y. Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs. *Nature Commun*. 8(1):1737 (2017)

4. De Solis CA, Morales AA, Hosek MP, **Partin AC**, Ploski JE. Is Arc mRNA Unique: A Search for mRNAs That Localize to the Distal Dendrites of Dentate Gyrus Granule Cells Following Neural Activity. *Front. Mol. Neurosci*. 10:314 (2017)

5. Selvaraj U, Ortega S, Hu R, Gilchrist R, **Partin A**, Plautz E, Klein R, Gidday J, Stowe A. Preconditioning-induced CXCL12 upregulation minimizes leukocyte infiltration after stroke in ischemia-tolerant mice. *J Cereb. Blood Flow Metab*. 37(3):801 (2016)

6. **Partin AC**, Hosek MP, Luong JA, Lella SK, Sharma SA, Ploski JE. Amygdala nuclei critical for emotional learning exhibit unique gene expression patterns. *Neurobiol Learn Mem.* 104:110-21. (2013)

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction—MicroRNA regulation and recognition

1.1. MicroRNA functions and association with disease

MicroRNAs are short (21-22 nt) noncoding RNAs that regulate gene expression. MiRNAs post-transcriptionally regulate nearly every biological process, through their partial complementarity to messenger RNAs (mRNAs). The interactions between a miRNA and its target mRNA initiates a process called RNA interference, in which translation of targeted mRNAs are suppressed through several mechanisms (1). The majority of human genes are regulated by miRNAs (2), and a single mirRNA may target dozens or hundreds of target mRNAs (3). Thus nearly all biological processes are regulated by miRNAs, including brain development (4), synaptic plasticity (5, 6), hematopoiesis (7), muscle development (8), immune development (9), viral infection (10), and apoptosis (11) (Figure 1). RNA interference is thought to be conserved across all eukaryotes (1), and miRNAs and miRNA-like RNAs appear to have evolved independently in other eukaryotes including plants and fungi (12-14).

Considering the pervasive roles of miRNAs, it is unsurprising that miRNA dysregulation is associated with the pathogenesis of many diseases. Many studies have established links between aberrant miRNA function and numerous disorders, including fragile X syndrome (6, 15), Rett syndrome (16), autism (17), depression (18), schizophrenia (19), drug addiction (20), Alzheimer's disease (21), cardiovascular disease (22), and psoriasis (23). In addition miRNA dysfunction has frequently been associated with cancer (24-27). Aberrant expression of microRNAs is found in virtually all forms of

cancer, and these expression patterns are frequently interpreted as "hallmarks" of cancer (28-30). Some cancers are associated with deletions or point mutations in miRNA genes (27, 31), while many others are linked to alterations in miRNA expression. Cancer can be driven by enhanced expression of oncogenic miRNAs (32, 33), or by downregulation of tumor suppressor miRNAs (34-36). In some cases, these alterations in miRNA expression are driven epigenetically, through promoter hypermethylation (37) and through histone modifications (38). Because of the high degree of involvement of miRNAs in disease, a more complete understanding of miRNA biogenesis would provide a foundation for development of improved therapeutics.

1.2. MicroRNAs are pervasive throughout animal species and play critical roles in humans

MiRNAs were first identified in *Caenorhabditis elegans* (39) as critical regulators of development, and soon thereafter were found to act through partial complementarity with specific mRNAs (39, 40). Later it was discovered that a large class of RNAs with this role exists (41-43). Identification of potential miRNAs has been facilitated by genomic analysis, and miRNA genes have been found throughout the animal kingdom. In the case of humans, one conservative estimate is that at least 60% of protein-coding genes are targets of microRNAs, although the number may be much higher (2). A very stringent analysis of potential human miRNAs found that at least 519 (44) miRNAs exist in humans. The number in lower vertebrates is smaller, and smaller still in simpler

organisms such as *Drosophila melanogaster* (164 miRNAs) (44, 45) and *Caenorhabditis elegans* (147 miRNAs) (44, 46).

MiRbase, the online database of microRNAs, contains annotations for approximately 2000 human miRNAs, but many of these may be false positives (47). The list of highly conserved human miRNAs is likely nearly complete, but more poorly conserved miRNAs are still being identified. A more thorough understanding of what constitutes a functional microRNA would facilitate identification of these poorly-conserved miRNAs.

1.3. Genomic sources of microRNAs

Most miRNAs are processed by RNases that cut out mature miRNAs from larger RNA sequences, in a process called miRNA maturation. The first step involves cutting a primary microRNA (pri-miRNA) transcript out of a larger piece of genomically encoded RNA. These larger RNAs often consist of spliced introns (>25% of conserved miRNAs and >50% of poorly conserved miRNAs in mammals (48, 49)), while many others are noncoding RNAs harboring clusters of microRNAs (50). Thus, pri-miR sizes can range from a few hundred nucleotides to several thousand nucleotides (45), are frequently 7-methylguanosine-capped and poly-adenylated (51), and are transcribed by RNA polymerase II (52). Therefore, the system by which miRNAs are processed must critically rely upon cues to ensure that miRNAs are properly distinguished from other RNAs, and processed accurately. Furthermore, the machinery involved in producing microRNAs must be flexible enough to accommodate many different sequences.

1.4. The canonical microRNA biogenesis pathway

A pri-miRNA consists of a hairpin loop structure formed by intramolecular folding (Figure 2). Human microRNA processing is carried out by two RNaseIII family enzymes, called Drosha and Dicer. RNaseIII enzymes are ribonucleases containing highly conserved RNase III domains (RIIIDs), which form inter- or intramolecular dimers and cleave double-stranded RNA (dsRNA) by hydrolyzing a phosphodiester bond on both strands, leaving a phosphate group on the 5' ends of the cut. The staggered orientation of RIIIDs on dsRNA leads to a characteristic 3' overhang containing 2 single-stranded nucleotides (53-55). Drosha, which serves as the catalytic component of a larger multi-protein complex called Microprocessor, initiates miRNA processing by cleaving pri-miRNAs on the end opposite the loop (the basal region). This event produces a product called a precursor miRNA (pre-miR)(56), which contains a mature miRNA duplex and a terminal loop opposite the Drosha cut sites (Figure 3).

Pre-miRNAs are then exported to the cytosol by Exportin-5/Ran-GTP (57, 58) and are cleaved again by Dicer (55, 59-61), yielding a mature miRNA duplex (Figure 3). Dicer associates with the RNA-induced silencing complex (RISC), which contains argonaute family proteins (62). Argonaute proteins utilize several mechanisms to separate the two miRNA strands, select a "target strand", and use the sequence of the target strand to identify target mRNAs and suppress their translation. The strand that is not selected is called a passenger strand or "star" strand and is discarded (63, 64). The strand selected to

become the target strand depends on the stability of the two ends, with preference going to the more weakly pairing end (65) and with a preference for 5' pU or pA (66, 67).

MiRNAs typically target the 3' untranslated region (UTR) of mRNAs at conserved sites (47). MiRNAs can be classified into families according to their target sequences, which are determined primarily through the 5' region of the miRNA (positions 2-8, called the "seed region") (68). It is likely that members of individual families arose from gene duplications (i.e., they are paralogous). The seed sequence is so critical to targeting that even a shift of a single nucleotide can alter the seed sequence in such a way as to allow targeting of different mRNAs (68-71). Because miRNA targeting is driven by sequence, it is critical that miRNA production is precise and tightly regulated.

1.5. Primary microRNA processing is carried out by the RNase Drosha

Drosha is an RNaseIII enzyme that initiates the pathway of microRNA biogenesis by processing primary microRNAs into precursor microRNAs. Through its two tandem RNaseIII domains (RIIIDa and RIIIDb), Drosha cleaves near the basal end of the pri-miR, releasing the stem-loop structure known as a precursor microRNA (pre-miR).

Human Drosha is a 160kDa polypeptide containing two RNaseIII domains, which are flanked by domains known to be critical for function (Figure 4). The disordered N-terminus (residues 1-352) is not involved in RNase activity (72), but is critical for nuclear localization (73). The Central Domain (CED) is N-terminally adjacent to the RIIIDs, and is essential for pri-miRNA processing activity. A crystal structure of human Drosha revealed that the CED shares some structural features with Dicer, including structures

resembling the "platform" and "connector" helix found in human and *Giardia intestinalis* Dicer, suggesting a common ancestral origin for the two RNaseIII proteins (74). In addition, a "PAZ-like" domain was proposed for a region of Drosha that was mostly disordered but had slight resemblance to the PAZ domain of Dicer. The RIIID domains are stabilized by the CED and the association of these three domains forms a binding surface for dsRNA. The two RIIID domains form an intramolecular dimer and each domain cleaves one of the two strands in the dsRNA stem (53, 55). Mutagenesis studies have shown that the N-terminal RIIID (RIIIDa) is responsible for the 3' cut, while the C-terminal RIIID (RIIIDb) makes the 5' cut (72). Located C-terminally to the RIIIDs is a double-stranded RNA binding domain (dsRBD), a common domain involved in (usually nonspecifically) recognizing double-stranded RNA. Interestingly, the dsRBD of Drosha contains substitutions at several residues normally associated with RNA binding, and these substitutions are critical for Drosha activity for unknown reasons (75, 76).

1.6. Downstream steps in microRNA biogenesis are directly influenced by Drosha activity

The steps following pri-miRNA processing rely upon Drosha activity. As mentioned previously, Drosha cleavage determines the boundaries of the basal end of the miRNA duplex. The 2nt 3' overhang produced by Drosha is directly recognized by Exportin-5 and is therefore critical for pre-miRNA export to the cytosol (77). During Dicer processing, the overhang is recognized by the PAZ domain of Dicer, and is critical for pre-miRNA processing (78-81). Furthermore, Dicer uses a "molecular ruler" to

determine placement of its catalytic centers, using the Drosha cut sites as reference points

(81, 82). Therefore, the activity of Drosha indirectly influences the Dicer cut sites,

thereby determining the sequences of both ends of the miRNA duplex. Understanding

recognition and cut site determination of pri-miRNAs by Drosha is critical to our

understanding of microRNA biogenesis.

1.7. A unique role for DGCR8 in dsRNA processing

Microprocessor contains numerous factors, including various RNA-binding

proteins, although the exact composition may vary under different circumstances (45, 72,

83). However, it has been established that the core components of Microprocessor are

Drosha and its partner protein, DiGeorge syndrome critical region 8 (or DGCR8).

Together, these two proteins are sufficient for pri-miRNA processing *in vitro*, and can

produce pre-miR products with the same sequences as is observed *in vivo*. Thus, gaining

a more complete understanding of these minimal Microprocessor components can

provide a wealth of knowledge on substrate recognition and specificity, as well as

regulation of Drosha function.

The requirement of Drosha for a partner protein is unique among RNaseIII family

members. For example, human Dicer does not directly require association with any other

proteins for stability or activity (49). RNaseIII proteins in lower species, such as the

Saccharomyces cerevisiae protein Rnt1p and the bacterial Ribonuclease III proteins, are

also capable of functioning without additional factors. In contrast, Drosha requires co-

expression with DGCR8 for stability, and Drosha activity is virtually zero without

DGCR8. Drosha is stabilized by a conserved C-terminal tail (CTT) of DGCR8, which binds to the RIIID domains of Drosha. DGCR8 itself has no catalytic activity; instead, it contributes to processing via its RNA-binding regions, including two tandem dsRBDs and a heme-binding region (HBR, also called Rhed or HBD). Drosha's dependence on DGCR8 is conserved in lower organisms such as Drosophila melanogaster, where it is called Pasha (45).

Human DGCR8 is a 90kDa polypeptide, containing an N-terminal region which is not involved in pri-miR processing but contains a nuclear localization signal (84), a central heme-binding region (HBR) (see section 3.2), two tandem dsRBDs, and a C-terminal tail that binds directly to Drosha. Previously it has been shown that DGCR8 adopts a dimeric stoichiometry in solution (85). Studies suggest that dsRBD1 and dsRBD2 of DGCR8 play different roles in processing. A crystal structure of the two tandem dsRBDs (called the "core") shows the two domains pack against each other, with their RNA-binding regions facing away from one another (86). This conformation would prevent both dsRBDs from simultaneously binding an RNA stem without major distortions in the dsRNA. FRET experiments from the same group indicated that a rearrangement of the core might occur upon RNA binding. Whether expressed together or separately, both DGCR8-dsRBDs bind dsRNA nonspecifically (87). Interestingly, the dsRBDs of DGCR8 show significant divergence from the dsRBD consensus sequence, including regions typically associated with folding and RNA binding (75, 88).

1.8. Heme binding regulates microRNA biogenesis through DGCR8

DGCR8 is a hemeprotein, and heme binding is essential for pri-miRNA processing. DGCR8 binds heme through its heme-binding region (HBR), a ~30-35kDa N-terminal region containing no known homology to other proteins, except for a small ~50 residue region thought to be a WW fold. Directly adjacent to the WW region is the Cys352 residue, which forms a coordination complex with heme B (85) through its thiol side chain. The stoichiometry of this assembly is believed to be one heme molecule for two copies of DGCR8 (i.e., one per DGCR8 dimer), and spectroscopic evidence suggests that the two axial ligands for heme are the thiolate groups from the two Cys352 side chains of each protomer in the DGCR8 dimer (89). Thus, DGCR8 is the only known example of two thiolates serving as axial ligands in a hemeprotein. Heme-bound DGCR8 exhibits a characteristic Soret peak at 450nm, as well as several other minor peaks at longer wavelengths.

In 2007 it was discovered that human DGCR8 binds to heme B. The initial report described two species of recombinant DGCR8 co-purified from overexpression: a heme-bound species with a larger hydrodynamic radius, and a heme-deficient species with a significantly smaller radius, as measured by size-exclusion chromatography. The authors concluded that heme binding is associated with DGCR8 dimerization. When they combined these DGCR8 samples with recombinant Drosha and reconstituted primiR processing *in vitro*, they found that only the larger, heme-bound species enabled pri-miRNA processing. Analysis of sequence conservation led them to identify Cys352 as the axial ligand for heme. Mutagenesis and methylmercury acetate titration confirmed that Cys352 serves as both axial ligands for heme in DGCR8, and a model was proposed

in which one heme molecule is bound by the DGCR8 dimer (89, 90). Another group (91) used electron paramagnetic resonance and electron nuclear double resonance (EPR/ENDOR) to provide further support to this model, by revealing that the coordination environment is highly symmetric and homogeneous (as would be the case with two Cys352 axial ligands). Another study presented EPR experiments suggesting the iron in heme-bound DGCR8 is homogeneously ferric (Fe^{3+}), and used *in vitro* activity assays to show that the ferric state is essential for pri-miRNA processing (92).

For many years, the generally accepted model was that heme enables dimerization of DGCR8, and the dimeric state is critical for pri-miR processing. DGCR8 constructs without the HBR exhibit nonspecific RNA-binding activities (87, 93). Expression constructs containing the HBR were dimers, and thus contained four dsRBDs rather than two; as expected, HBR-containing constructs bound RNA with higher affinity, and while specificity remained weak, HBR-containing constructs did exhibit a preference for single-stranded/double-stranded RNA (ss/dsRNA) junctions (the points on either side of the pri-miR at which the dsRNA stem branches apart into single-stranded RNA) (94). Thus, a model was proposed in which heme binding activates DGCR8 by promoting dimerization and enabling the HBR to sense junctions. Due to the poor specificity of DGCR8 for RNA, it was proposed that DGCR8 itself does not initially identify pri-miRNAs, and that instead, a pre-formed Drosha/DGCR8 complex likely recognizes pri-miRNAs(87).

Several key features make DGCR8 unique among hemeproteins. For example, the ferric configuration of the iron in DGCR8 is far less common than the ferrous

configuration in most hemeproteins. One notable example of a ferric system is cytochrome p450 oxidase, which exists in a Fe^{3+} state during one of its stages of catalysis (this transition leads to a Soret peak at 450nm, similar to DGCR8). Another unique feature of DGCR8 is its P351/C352 sequence at the coordination site, in contrast to the reversed "CP" motif, which is frequently observed in hemeproteins (95). Finally, DGCR8 can coordinate cobalt in place of iron (by binding to cobalt protoporphyrin IX rather than heme), and cobalt-containing DGCR8 can enable pri-miRNA processing at similar levels to the heme-bound version, *in vitro* and *in vivo* (96).

1.9. Structural features define pri-miRNAs

As mentioned previously, hundreds of pri-miRNAs exist in humans, and each contains a unique primary sequence. Thus, Drosha/DGCR8 must be flexible enough to accommodate numerous substrates, yet strict enough in its substrate selection so as to prevent off-target activity. Only a tiny fraction of cellular RNA consists of pri-miRNAs (97), and Drosha/DGCR8 must rely upon some combination of structural features and RNA sequence to distinguish their substrates from other RNAs.

Drosha/DGCR8 is hypothesized to recognize multiple features within the pri-miRNA molecule (Figure 5). This combination of recognition cues distinguishes pri-miRNAs from other cellular RNAs. Early *in vitro* work showed that Drosha/DGCR8 utilize several structural cues for pri-miRNA recognition, including the need for a double-stranded stem containing multiple helical turns (98), a terminal hairpin loop of sufficient size (98-100), and single-stranded flanking segments opposite the loop (100).

Critically, the stem length must be approximately 35 nt {Han, 2006 #4361, 101-103). To a certain extent, mismatches within the stem are tolerated, and a typical pri-miRNA stem contains 2-4 mismatches. The pri-miRNA is usually represented with the terminal loop being the "top" end; thus the junction near the terminal loop is called the apical junction, and the junction at the opposite end is called the basal junction. Both junctions, but primarily the basal junction, are thought to serve as major reference points for determining the cleavage sites (100, 102). It is believed that Drosha binds to the basal junction, in part because its cleavage sites are positioned closer to the basal end of the pri-miRNA. The stem length is also important for determining the cut site, and it has been shown that adding or removing base pairs from the basal (104, 105) or apical (99, 104) regions of a pri-miRNA can alter the cut sites.

1.10. Primary sequence can influence pri-miRNA processing

Several small primary sequence motifs have been identified in pri-miRNAs, and were shown to have an enhancing effect on processing *in vitro* (Figure 6). These motifs are found in pri-miRNAs more often than what would occur by chance. The motif that is thought to play the most pivotal role is the mismatched GHG motif (where H is A, C, or U), which is found on the 3' strand approximately 3nt basal to the 3' cut site (103). It is critical that the nucleotide in position H be mismatched (i.e., not engaging in base pairing). The dsRBD of Drosha has been proposed to be is involved in detecting the GHG motif (105). In addition, there are three other known primary sequence motifs: the basal UG motif, the 3' single-stranded CNNC motif, and the apical UGU motif (102, 106). In

human miRNAs, the frequency of these three motifs is much higher than that which would occur by chance. The basal UG motif is believed to be recognized by Drosha (106). Consistent with models that place DGCR8 at the apical stem, there is evidence that the apical UGU motif is recognized by DGCR8 in a heme-dependent manner (106). The CNNC motif has been associated with recognition by various regulatory proteins (102, 107, 108). No enrichment of these motifs were found in the microRNAs of *C. elegans* (102), suggesting that the human miRNA processing system has undergone specialization and developed unique processing cues. Taken together with the high number of miRNAs in humans relative to other animals, it is likely that microRNAs have evolved to perform unique and critical roles in human biology. Thus, a close examination of the components of the human microRNA biogenesis system is essential to furthering our understanding how microRNAs influence human health and disease.

1.11. Pri-miR processing poses a unique challenge for RNA recognition

Because it requires a balance of flexibility and stringency in substrate selection, Drosha faces a unique recognition problem in pri-miRNA processing. Many cellular RNases utilize primary sequence or simple secondary structure elements to identify substrates. In contrast to the direct recognition of the 3' overhang of pre-miRNAs by Dicer, Drosha cannot utilize simple recognition cues to identify substrates; and although primary sequence recognition cues (Section 1.10) may help to properly identify pri-miRNAs, many pri-miRNAs lack these motifs, and thus a more general set of recognition cues must be utilized. Furthermore, many non-miRNA transcripts contain ss/dsRNA

junctions and thus the mere presence of junctions is not necessarily a useful determinant for pri-miRNAs. Therefore, Microprocessor likely utilizes numerous protein modules to recognize a combination of features on potential pri-miRNAs, to ensure that only true pri-miRNAs are processed. To complicate this process further, because a pri-miRNA hairpin contains a ss/dsRNA junction at each end of the stem, Drosha is presented with two possible binding sites, and two possible orientations for placing its catalytic centers. Thus, Drosha must reliably discriminate between the two junctions, and bind the basal junction, in order to produce functional precursor microRNAs.

Figure 1. Roles of microRNAs in regulating gene expression

cancer hematopoiesis

neural disorders ← microRNAs → organ development

viral infection immune function apoptosis

Terminal loop

Dicer cleavage sites

double-stranded stem

Drosha cleavage sites

5' 3'

single-stranded flanking segments (arms)

Figure 2. Basic structure of a primary microRNA. The cleavage sites for Drosha and Dicer are shown in red and blue, respectively.

Figure 3. Canonical microRNA processing pathway. The cleavage sites for Drosha and Dicer are shown in red and blue, respectively.

Figure 4. Organization of Drosha and DGCR8. (A) Domain organization and construct diagrams for human Drosha and DGCR8. P-rich, proline-rich region; RS-rich, arginine/serine-rich region; NLS, nuclear localization signal; CED, central domain; RIIID, ribonuclease III domain; dsRBD double-stranded RNA-binding domain; HBR, heme-binding region, CTT C-terminal tail.

Figure 5. Structural recognition cues involved in pri-miRNA processing. Drosha cleavage sites are indicated by red arrows.

Figure 6. Primary sequence motifs involved in pri-miRNA processing. Drosha cleavage sites are indicated by red arrows. Pri-miRNA orientation is represented the same as in Figure 5.

Chapter 2. The Role of heme in pri-miRNA recognition

Authors:

Alexander C. Partin, Tri D Ngo, Emily Herrell, Byung-Cheon Jeong, Yunsun Nam

Author Contributions:

A.C.P. and Y.N. designed the experiments and analyzed data. A.C.P. conducted the experiments. T.D.N. developed workflows for *in vitro* RNA biochemistry approaches and performed Northing blotting and splinted ligation assays. E.M.H. assisted with cloning and RNA purification. B.J. established initial Microprocessor purification protocols.

2.1. Introduction

MicroRNAs (miRNAs) control expression of most protein-coding genes in higher eukaryotes at the post-transcriptional level. The canonical miR maturation pathway requires the RNase III enzyme Drosha and its essential cofactor, DGCR8 (together forming the core Microprocessor or MP) (56, 72, 83). Processing by Drosha requires a high degree of precision, because even a single-nucleotide deviation can affect which mRNAs are targeted (68, 71, 80, 82). Drosha activity is therefore highly regulated, and its dysfunction is frequently associated with disease (25, 49). Many mature miR levels do not correlate with the levels of pri-miRNAs (109-112), underscoring the prevalence of post-transcriptional regulation of miR biogenesis.

A Drosha substrate—a pri-miR—can vary in structure and sequence, but always includes a major stem-loop structure. At each end of the stem is a single-stranded/double-stranded RNA (ss-dsRNA) junction: the apical junction forming the terminal loop, and the basal junction at the opposite end. Drosha and DGCR8 have been suggested to cooperate to bind these two junctions, ensuring proper substrate recognition. Several primary sequence motifs have been identified in pri-miRNAs (102, 103), which may help orient the complex, but these motifs are only present in a subset of pri-miRNAs, leaving hundreds of others apparently lacking recognition sequences. Therefore, in order to reliably bind and process pri-miRNAs, MP is likely to use the structural features of pri-miRNAs, but how such recognition is accomplished is unclear (113). For many years, DGCR8 was associated with binding the basal junction, and the importance of the terminal loop was unclear due to conflicting results (98-100). More recently, a crystal

structure of Drosha has led to a revised model where Drosha binds to the basal junction, and recognizes a "UG" motif (74, 106). Assigning Drosha to the basal junction led to a model where DGCR8 recognizes a "UGU" motif in the terminal loop (94, 106).

For more than a decade, DGCR8 has been known to interact with heme, via a heme-binding region (HBR or RNA-binding heme domain)(85). Yet, how heme affects DGCR8 at the molecular level has remained unclear. Pri-miR processing is enhanced by heme both *in vitro* and *in vivo*, and variations of heme state (e.g. oxidation state and gas binding) can modulate DGCR8 activity, through unknown mechanisms (89, 92, 114, 115). Many heme-binding proteins utilize heme for chemical processes, such as catalysis, electron transfer, and gas transport (116); however, heme can also serve as a signaling molecule, regulating diverse functions such as transcription (117, 118), ion flux (119), and cell signaling (120). Thus, heme is a versatile regulator of biological function, and is expected to play a regulatory role in miR biogenesis. The HBR of DGCR8 has no known homologs, except for a WW motif (residues 298-352). Previous studies have proposed that heme serves to promote dimerization of DGCR8, where the thiol groups of Cys352 from two copies of DGCR8 serve as the axial ligands (85). Dimerization has been suggested to help substrate recognition and accuracy (106, 121). However, how dimerization, heme binding, and pri-miR processing are mechanistically related is unknown.

Here we show that heme ensures fidelity of pri-miR processing. We show that heme is necessary for Microprocessor to cleave certain substrates at the correct site, and the degree of heme dependence varies for different pri-miRNAs. Heme dependence arises

from failure of the basal junction to recruit Drosha; heme-bound DGCR8 provides the necessary specificity to direct Drosha to the correct binding site, the basal junction. We show that DGCR8 is dimeric independent of heme binding, but dimerization alone is insufficient for accurate processing. Furthermore, heme induces a conformational change that activates the HBR to recognize the substrate RNA. We reveal that heme activates DGCR8 to recognize the structural context of the terminal loop of pri-miRNAs, and the binding event induces a unique RNA conformation near its 3' end.

2.2. Methods

Protein expression and purification

Drosha/DGCR8 complexes were expressed using the Bac-to-Bac system (ThermoFisher) in HighFive cells. Drosha and DGCR8 fragments were cloned into pFastbacDual, with a hexahistidine tag, and FLAG tag, respectively. The expression constructs contained the following residues: Drosha, 353-1372; DGCR8$^{Heme/Apo/C352S}$, 223-751; DGCR8$^{\Delta HBR}$, 489-751; and DGCR8^{CTT}, 728-751. Bacterial DGCR8 constructs in pET21 were expressed fused to hexahistidine tags in BL21 (DE3) Rosetta cells. For heme-saturated bacterial DGCR8, 5 μM hemin Cl (Sigma) was added to the media at the time of inoculation. Drosha/DGCR8 complexes and bacterial DGCR8 were purified by Ni-NTA affinity chromatography. Lysis and washing were performed using a buffer containing 20 mM Tris pH 8.0, 1 M sodium chloride, 10% glycerol, and 1 mM DTT. Cation exchange chromatography was performed using a 100-800 mM NaCl gradient buffered with 20mM Bis-Tris pH 7.0, followed by a size-exclusion step. MP^{Apo}, the

heme-deficient species, was separated away from MP^{heme} through multiple rounds of cation-exchange chromatography.

RNA transcription and purification

RNA templates were cloned into pRZ vectors containing self-cleaving ribozymes on either end to produce homogenous ends (5' hammerhead, 3' hepatitis delta virus) (122). Transcription reactions contained 1 μg linearized DNA template, 4 mM rNTPs, 35 mM MgCl$_2$, 50 mM Tris pH 8.0, 1 mM spermidine, 0.01% Triton-X 100, 10 mM DTT, and 100 U RNAse inhibitor (Thermo). Target RNA was purified by denaturing polyacrylamide gel electrophoresis (PAGE).

In vitro pri-miR processing assays

5' end-labeling was performed using T4 Polynucleotide Kinase (NEB) and γ-^{32}P-ATP. 3' end labeling was performed using α-^{32}P-ATP and T4 RNA ligase 1 (NEB). Pri-miR processing assays were performed in 15 μL reactions containing 30 mM Tris pH 7.5, 67 mM sodium chloride, 5% glycerol, 10 mM MgCl$_2$, 5 mM DTT, 8 U RNase inhbitior, and 1.5 μg yeast tRNA. Reactions were incubated for 30 minutes at room temperature and analyzed by urea PAGE.

Gel-shift assay (Electrophoretic Mobility Shift Assay, EMSA)

5' end-labeled RNA was incubated with a dilution series of protein samples in 20 mM Tris pH 7.5, 140 mM sodium chloride, 10 mM DTT, 3 mM MgCl$_2$, 100 μg mL^{-1}

yeast tRNA, and 10% glycerol. The mixture was analyzed by native PAGE. For K_d determination, the band intensities in each lane were quantified using the program ImageLab (BioRad), and fraction of RNA bound was plotted as a function of protein concentration. The data were fitted to a logarithmic curve and the concentration at which 50% of RNA was bound was reported as the K_d.

Size-exclusion chromatography—Multiangle light scattering (SEC-MALS)

Purified protein samples were loaded onto a Superdex 200 Increase 10/300 column (GE) using Agilent 1200 Infinity series HPLC, in a buffer containing 20 mM Tris pH 8.0, 750 mM sodium chloride, 1 mM TCEP and 10% glycerol. DAWN HELIOS II and tREX detectors (Wyatt) were used to monitor scattering and differential refractometry. The results were processed using Astra 6.1 (Wyatt) and fitted using the Zimm model to obtain molar mass. For all samples, dn/dc was approximated to be 0.185 mL g^{-1}. Zimm R^2 values are shown in Table 1. Heme-saturated samples gave inconclusive results, likely due to interference from the heme signal, and are not reported.

Sedimentation Velocity Analytical Ultracentrifugation

Samples were prepared at a series of concentrations between 0.1 and 1 mg mL^{-1}, in a buffer containing 20mM Tris pH 8.0, 500 mM NaCl, and 1 mM TCEP, and then loaded onto double-sector analytical cells. Using a ProteomeLab XL-I ultracentrifuge (Beckman Coulter), the cells were spun for approximately 16 hours, at either 42,000 or 30,000 RPM, for DGCR8 and Drosha/DGCR8 complex, respectively, with UV and

interferometry scans approximately every 10 minutes per cell. The data were date-corrected using the Redate program, then analyzed using SedFit {Schuck, 2000 #4637} to obtain molar mass, sedimentation coefficient, and other parameters. For continuous c(S) distribution analysis, the simplex algorithm was used to fit the data, and the confidence level (F-ratio) was set to 0.68. The figures were created using the program GUSSI (123). RMSD values are reported in Table 1.

Electron microscopy

MP^{Heme} and MP^{C352S} complexes were crosslinked using 0.1% glutaraldehyde and a protein concentration of 1 μM. The crosslinked samples were purified by size-exclusion chromatography. Negative stain grids were prepared using 2% uranyl acetate and the images were obtained using a Tecnai G2 Spirit TEM (FEI) running at 120kV, and using a magnification of 30,000x.

Limited proteolysis

α-chymotrypsin (Hampton Research) was diluted to 0.01 mg mL^{-1} in 10 mM HEPES pH 7.5, 500 mM sodium chloride, then mixed 1:1 with DGCR8 protein (5mg mL^{-1}) and incubated at 37°C for 1 hour. The reactions were analyzed by SDS-PAGE and visualized by Stain-free imaging (Biorad).

Selective hydroxyl acylation followed by primer extension (SHAPE).

SHAPE reactions were performed on pre-assembled RNA-protein complexes. Protein and RNA samples were mixed at a final concentration of 312.5 nM and 65 nM, respectively, in 10 mM Tris pH 7.5, 130 mM sodium chloride, 50 μM $ZnCl_2$, 10 mM $MgCl_2$, 6.4 ng yeast tRNA, and 2 U Ribolock. SHAPE reactions were performed with 40 mM benzoyl cyanide (BzCN). Superscript III was used for reverse transcription using ^{32}P-labeled primers and the resultant DNA was subjected to sequencing gel electrophoresis. To generate ladders, a separate reverse transcription was performed containing 1 mM of either dideoxy (dd)ATP, ddCTP, ddGTP, or ddTTP added alongside standard 0.5 mM dNTPs. These four ddNTPs were used to assign their respective Watson:Crick partners. Capillary electrophoresis samples were prepared with 6FAM and HEX-labeled primers and analyzed using a 3730XL Genetic Analyzer. The scans (.fsa files) were visualized using the QuShape program (124), and the traces were manually aligned according to the ddCTP ladder. After visualization of SHAPE results, the ladder was shifted down by one nucleotide position, to account for the mechanism by which the SHAPE adduct interferes with reverse transcription. All secondary structure diagrams to locate SHAPE hotspots were generated by using mfold (125).

Cell culture and quantitative real-time PCR (qPCR)

293T cells (authenticated and tested by ATCC) were maintained using Dulbecco's modified eagle medium (DMEM) supplemented with 10% FBS and transient transfections were carried out using Lipofectamine3000 (Thermo) according to the manufacturer's instructions. 40 hours after transfection, total RNA and protein were

extracted using Trizol (Thermo) and SuperScript II (Thermo) was used for reverse transcription. Mature microRNA levels were determined using the Taqman microRNA assay (Thermo), according to the manufacturer's instructions. Quantitative PCR was performed using the Taqman PCR Universal Master Mix II (no UNG) on a LightCyclerII (Roche). Primary transcripts and mature levels of miR-9-1 and miR-21 were normalized to pri-miR-125a or miR-125a by using comparative C_T method. Statistical analyses were carried out for 3 biological replicates (3 independent transfections) using paired Student's t-test (two tailed). Four technical replicates were performed for each value used in the analysis. For Western blotting, protein samples from Trizol extraction were separated by SDS-PAGE and transferred to PVDF membranes. Anti-actin HRP (Sigma, A3854) dilution was 1:5000, Anti-FLAG M2 (Sigma, F1804) dilution was 1:1000. Goat anti-mouse HRP (Pierce, PI31430) dilution was 1:5000.

Hydroxyl radical footprinting

RNA-protein complexes were assembled as described in the SHAPE methods above, in a volume of 25 μL. To induce cleavage of RNA, a hydroxyl radical mixture (6 μL of 2% H_2O_2, 5.33 mM ammonium Fe(II) sulfate hexahydrate pre-mixed with 50 mM EDTA, and 63 mM ascorbic acid) was rapidly assembled and mixed into to the RNA-protein complex. After a 30-second incubation, the reactions were quenched, and reverse transcription was performed after acid phenol-chloroform extraction and ethanol precipitation, as described in the SHAPE methods above. The ^{32}P-labelled DNA products were visualized by sequencing gel electrophoresis. For quantitation, band intensities were

measured using ImageLab, and the differences in absolute intensities between buffer and protein lanes (I_{buffer}-$I_{protein}$) were reported for each nucleotide. The y-axis unit is defined as 25,000 volume units as measured by ImageLab. Nucleotides at positions 29 and 30 were excluded from the graph due to reverse transcription artifacts.

Splinted ligation small RNA detection assays

Splinted ligation reactions were performed as described previously (126). Briefly, 3 µg of total RNA was annealed with 1 pmol of ^{32}P-end-labelled ligation oligonucleotide and 1 pmol of a bridge nucleotide specific to the target miR, in a reaction containing 75 mM KCl and 20 mM Tris pH 7.5. The annealed sample was then incubated in a ligation reaction containing 30 mM KCl, 75 mM Tris pH 7.6, 10 mM $MgCl_2$, 1mM DTT, 0.5 mM ATP, 7.5% w/v PEG 6,000, and 2 U T4 DNA ligase (NEB). After a 1-hour incubation at room temperature, the reactions were stopped by adding 10 units of alkaline phosphatase and incubated at 37°C for 15 minutes. The ligated samples separated by denaturing PAGE, and visualized by autoradiography followed by quantitation using ImageLab. Statistics were carried out with three biological replicates for each group, using paired Student's t-test (two tailed).

Northern Blotting

Total RNA (10 µg) mixed with equal volume of 2x formamide RNA loading dye was resolved using a denaturing polyacrylamide-urea gel. RNA was transferred to Hybond-N membrane in TBE buffer using Bio-Rad Trans-Blot cell followed by UV-

crosslinking. Hybridization was performed in a buffer containing 0.5 M sodium phosphate pH 7.2, 7% SDS and 1mM EDTA with 10 ng of radiolabeled probe at 42°C overnight. The membrane was washed in buffer containing 60 mM sodium phosphate, 1 M sodium chloride, and 6 mM EDTA at 42°C for 30 minutes before exposing to a phosphor screen. The probe sequences used for U6, pre-miR-125a, pre-miR-21, and pre-miR-9-1 were: GCAGGGGCCATGCTAATCTTCTCTGTATCG, TCACAGGTTAAAGGGTCTCAGGGA, TCAACATCAGTCTGATAAGCTA and ACTTTCGGTTATCTAGCTTTAT, respectively.

2.3. Results

5.3.1. Heme binding reverses Microprocessor orientation on pri-miRNAs

In order to investigate the role of heme in miRNA biogenesis, we purified recombinant human Drosha and DGCR8 polypeptides co-expressed from insect cells. Heme-bound and heme-free Microprocessor can be separated slightly by cation-exchange chromatography (with heme-bound species eluting at slightly higher conductivity), and we took advantage of this separation to isolate heme-saturated (MP^{Heme}) and heme-deficient (MP^{Apo}) species. Our expression constructs contain all of the folded domains known to be required to cleave pri-miRNAs into canonical pre-miRNAs (Figure 7). We also purified the same complex with mutant DGCR8 (C352S)(85) that does not bind heme (MP^{C352S}) (Figure 8). When Drosha/DGCR8 complexes were deficient in heme content, we observed a dramatic change in how they process pri-miR-143 (Figure 9). In

contrast to MP^{Heme}, both MP^{Apo} and MP^{C352S} generate a distinct, secondary cleavage product (red asterisk). When the MP^{Apo} reaction is supplemented with 2.5 μM hemin, the proper cut site is reinstated. Adding heme to MP^{C352S} does not affect the outcome, suggesting that the corrective effect of heme requires a specific coordination via Cys-352. Using sequencing gels and 5' or 3' end-labeled substrates, we identified that the wrong cleavage site lies in the middle of the mature miR sequence (Figure 10). By combining the heme-binding mutation (DGCR8^{C352S}) with point mutations of the catalytic residues of Drosha, we determined the orientation of the Microprocessor complex in the absence of heme (Figure 11). The alternative cleavage event associated with heme loss is caused by a "flipping" of the complex (Figure 11B). Though reminiscent of "abortive processing" described previously for mutant substrates or domain deletions (100, 106), what is striking is that wild type Drosha and DGCR8 reverse their orientation on a wild type pri-miR, depending on the mere presence of heme.

2.3.2. Heme dependence varies among miRNAs

We investigated whether alternative processing in the absence of heme occurs for other miRNAs. Using MP^{C352S}, we screened a series of pri-miRNAs for any changes in processing associated with heme loss (Figure 12). The extent of processing defect with the heme-binding mutant varies for different pri-miRNAs. Some pri-miRNAs (e.g., 9-1 and 21) are highly dependent on heme for processing at the correct site, while other pri-miRNAs (e.g., miR-125a and let-7d) show little to no difference in processing patterns in the absence of heme. These results suggest that a subset of miRNAs, which we will refer

to as "heme-dependent," need heme for high-fidelity processing, whereas other miRNAs ("heme-independent") can be processed primarily at the correct site regardless of heme presence.

To assess how individual pri-miRs respond differently to heme in intact cells, we generated tandem pri-miR constructs for expression in mammalian cells (Figure 13). A heme-independent pri-miR (125a) and a heme-dependent pri-miR (9-1 or 21) were expressed as part of a single transcript, along with either wild type DGCR8 or the heme binding mutant (DGCR8^{C352S}), in 293T cells. We quantified the amounts of intracellular mature miR levels by using both quantitative PCR (Figure 14) and direct labeling via splinted ligation (Figure 15). In both assays, the DGCR8 mutation (C352S) is more detrimental to the production of mature miR-9 and miR-21 in comparison to the heme-independent control, miR-125a. Because the primary transcript levels of heme-dependent miRs are not lower in the cells with DGCR8^{C352S}, the heme-binding mutation is likely to affect processing of pri-miRs after transcription (Figure 16). To confirm that the mutation affects production of pre-miRs, we also quantified the pre-miR levels using Northern blotting (Figure 17). Similar to mature miR levels, we observe less pre-miRs in cells expressing mutant DGCR8, and that heme-dependent miRs are more sensitive to the loss of heme compared to the heme-independent miR. Therefore, heme-dependence varies among different miRs both *in vitro* and in cells, and is an intrinsic property encoded into each pri-miR sequence.

Additional *in vitro* processing assays revealed that, although heme-independent pri-miRNAs are not processed at different sites in the absence of heme, their processing

kinetics are nevertheless affected by heme. Titrating MP^{Heme} and MP^{C352S} revealed that even the most heme-independent pri-miRNA, pri-125a, is processed at slightly lower efficiency without heme (Figure 18). In the cases of pri-miR-9-1 and pri-miR-21, Microprocessor kinetics are only moderately affected by heme loss, and the primary effect is change in cleavage site.

2.3.3. Heme enables DGCR8 to correct Drosha

We then asked why heme-dependent pri-miRNAs are processed differently from heme-independent pri-miRNAs. One possibility is that heme-bound DGCR8 is required to guide Drosha to the correct binding site, while heme-free DGCR8 fails to do so. Another possibility is that heme-free DGCR8 actively directs Drosha to an incorrect site to give rise to processing errors. To determine the intrinsic behavior of Microprocessor lacking the HBR, we tested how a series of DGCR8 truncations affect processing of the two classes of pri-miRNAs (Figure 19). As long as heme-bound HBR is present, Drosha can process both classes of pri-miRNAs at the canonical sites. However, when the HBR is removed, the processing accuracy is lost for heme-dependent pri-miRNAs, similar to MP^{C352S}, indicating that C352S is a loss-of-function mutation. When all of the RNA-binding regions of DGCR8 are removed—leaving only the minimal 23-residue C-terminal tail required for Drosha solubility (MP^{CTT})(106)—Drosha cuts pri-miR-21, a heme-dependent substrate, predominantly at the wrong site. In contrast, Drosha cleaves pri-miR-125a, a heme-independent substrate, at the correct location regardless of which

domains of DGCR8 are present. Therefore, Drosha is inherently error-prone for heme-dependent pri-miRNAs, and heme-activated DGCR8 drives it to the correct cut site.

To determine if Drosha intrinsically prefers the wrong site in heme-dependent pri-miRNAs, we tested its *in vitro* binding affinities for different junctions. EMSAs reveal that MP$^{\Delta HBR}$ binds to the apical junction of heme-dependent pri-miR-21 more tightly than to the basal junction (Figure 20A). In contrast, a heme-independent substrate, pri-miR-125a, has a robust basal junction with higher affinity for MP$^{\Delta HBR}$ than its apical junction (Figure 20B). Screening additional pri-miRNAs shows that human pri-miRNA basal junctions have a broad range of favorabilities (Figure 21). Thus, our results reveal that junctions in wild type pri-miRNAs vary widely in their abilities to recruit Drosha, leading to a "tug-of-war" between the apical and basal ends. For heme-dependent pri-miRNAs where Drosha is drawn to the incorrect (i.e., apical) junction, heme-bound DGCR8 intervenes to position Drosha at the correct site. For heme-independent pri-miRNAs where Drosha finds the correct junction independently, DGCR8^{heme} may provide additional contact to reinforce the overall binding affinity, but is not necessary for processing fidelity. Therefore, heme-dependence arises from the inability of the basal junction to outcompete the apical junction in recruiting Drosha.

2.3.4. Heme binding induces a conformational change but not dimerization

Heme binding has been associated with homotypic interactions of DGCR8, but whether heme induces dimerization is unclear (74, 85). Using sedimentation velocity analytical ultracentrifugation (svAUC) (Figures 22A-B) and size-exclusion

chromatography with multiangle static light scattering (SEC-MALS) experiments (Figure 22C-D), we determined that heme-free DGCR8 is a stable dimer, both in isolation and in complex with Drosha (Table 1). Our results disagree with the previous report that heme promotes dimerization, possibly because our construct includes more residues at the amino terminus that may prevent degradation. Furthermore, we found that Drosha with DGCR8 lacking the HBR (MP$^{\Delta HBR}$) is a 1:1 heterodimer, suggesting that the HBR is necessary for forming heterotrimeric complexes. Thus, our biophysical studies show that DGCR8 dimerizes via the HBR, in isolation as well as in the presence of Drosha, regardless of heme presence. Combined with our results described above, we show that dimerization of DGCR8 cannot correct Drosha, but that heme is the key factor required for high fidelity processing.

Because heme binding affects DGCR8 function without changing its oligomerization state, we questioned whether it might assert an allosteric effect as seen in many other heme-binding proteins. Limited proteolysis of DGCR8 shows that the presence of heme affects its digestion profile (Figure 23A), suggesting a conformational change. Additionally, elution profiles from size-exclusion chromatography and the frictional ratios measured through svAUC suggest that heme-bound DGCR8 has a more compact conformation compared to the heme-free state (Table 1, Figure 23B). Using negative stain electron microscopy we show that MP^{Heme} exhibits distinct features compared to the heme-free MP^{C352S} particles (Figure 23C). As expected from a conformational change, we also observe a shift in the melting temperature of DGCR8 depending on heme occupancy (Figure 23D). Collectively, our results indicate that heme

alters DGCR8 conformation rather than stoichiometry. To detect the functional outcome of the heme-induced conformational change, we performed electrophoretic mobility shift assays (EMSAs) to compare RNA-binding activities of DGCR8^{Heme} and DGCR8^{C352S} (Figure 24). Upon binding heme, the affinity of DGCR8 for pre-miR-143 increases dramatically. Therefore, our biophysical and biochemical results suggest that heme acts a molecular switch to induce a conformational change in DGCR8, which in turn activates DGCR8 to recognize substrate RNA.

2.3.5. Heme enables DGCR8 to recognize the terminal loop structure

To dissect the molecular role of heme, we examined how heme affects specific protein/RNA interactions, using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). SHAPE analysis on assembled pri-miR/Drosha/DGCR8 complexes reveals a striking difference between heme-bound and heme-free states. In the presence of heme, we observe a hyperreactive acylation site in the terminal loop of pri-let-7d, which would suggest a remodeling of the RNA backbone due to direct interactions with protein (Figure 25). This intense "hotspot" is absent for MP^{C352S}, even if the reaction is supplemented with extra heme (Figure 26A). The same nucleotide exhibits hyperreactivity when we add full-length, wild type Drosha/DGCR8^{Heme} complexes purified from HEK293 cells (Figure 26B). We also tested whether isolated DGCR8 is sufficient to recognize the terminal loop. Indeed, we observe the same SHAPE signature for DGCR8^{Heme} but not for DGCR8^{C352S}, suggesting that DGCR8^{Heme} specifically binds and remodels the terminal loop, independent of Drosha (Figure 26C).

To determine whether the specific interaction between the terminal loop and DGCR8^{Heme} is conserved in other miRNAs, we tested for the presence of a SHAPE hotspot. We observe that many miRNAs exhibit intense levels of acylation at a confined site in the loop, but only in the presence of MP^{Heme} (Figures 27A-G). The SHAPE hyperreactivity tends to focus around a single nucleotide, and this hotspot usually lies on the 3' side of the ssRNA region in the terminal loop, according to the secondary structure predictions by mfold (125). We observe these hotspots regardless of whether or not the pri-miR contains the apical UGU motif described previously (102, 106). Furthermore, we note that the hotspots usually do not overlap with the UGU motifs whenever they are present. Aligning the terminal loop sequences according to the hotspot position reveals that many of them center around an adenosine (6 out of 8 pri-miRNAs) (Figure 27H). To investigate the significance of the base identity, we examined how point mutations affect the *in vitro* binding affinity. In some cases, the base identity may play a modest role, as we observe reduced affinity when the hyperreactive adenosine of pri-let-7d is mutated to a cytosine (Figure 28A-B). However, in other miRNAs such as pri-miR-9-1, we observe little to no change in affinity for DGCR8 when we mutate the hotspot (Figures 28C-D), suggesting that sequence may not play a dominant role. As A49 in pri-let-7d makes a significant contribution to affinity, we tested how its base identity affects the SHAPE reactivity. Indeed, the same mutation that affects the affinity for DGCR8^{Heme} also reduces the extent of hyperreactivity (Figure 29). Lack of sequence dependence in pri-miR-9-1 prompted us to test whether DGCR8^{Heme} recognizes the structural context of the terminal loop. We investigated whether removing the constrained nature of the loop would

eliminate the contribution of heme to DGCR8/miR interactions. A break in the terminal loop of pre-miR-9-1 reduces its *in vitro* affinity for DGCR8^{Heme} to undetectable levels, similar to that of the heme-binding mutant, DGCR8^{C352S} (Figure 30A). Therefore, heme confers upon DGCR8 the ability to detect the constricted loop structure, and upon binding induces a unique RNA backbone configuration near the 3' end of the loop.

The conserved SHAPE hyperreactive site in terminal loops is thus likely due to a specific contact with the heme-bound DGCR8. Probing the hydroxyl radical reactivity of the loop region in the absence and presence of DGCR8 shows increased protection around the SHAPE hotspot (Figure 31). Therefore, both SHAPE and hydroxyl radical footprinting data suggest that heme-bound DGCR8 directly binds the terminal loop.

2.3.6. Recognition of the terminal loop by DGCR8^{Heme} can be targeted to regulate Drosha processing

Since heme-bound DGCR8 makes critical contacts with pri-miRNAs to enhance Microprocessor activity, we asked whether loop binding proteins would target this interaction to modulate miR biogenesis. Lin28 is an oncofetal factor that inhibits processing of the let-7 family of miRNAs, including blocking the Drosha step (127-129). Lin28A and Lin28B are highly conserved paralogs with no known biochemical differences *in vitro*, but because Lin28B localizes to the nucleus, it is likely to be the isoform responsible for inhibiting pri-let-7 processing by Microprocessor (130). How Lin28B is capable of regulating Drosha activity via binding the terminal loop has

remained unclear. Heme-bound DGCR8 induces a SHAPE hotspot in pri-let-7d (Figure 32A) near the Lin28 binding site, namely where the CCHC zinc knuckles bind in the crystal structure (131). Therefore, we tested how Lin28B affects HBR interactions with the terminal loop. Similar to the SHAPE hotspot induced by DGCR8^{Heme}, Lin28B itself can also elicit a specific hyperreactive nucleotide at a nearby location in the same terminal loop. The crystal structure of Lin28 in complex with the terminal loop of let-7d shows that the Lin28 hotspot base (G47) makes extensive contact with the protein (Figure 32A). When both MP^{Heme} and Lin28 are added simultaneously in equimolar amounts, the effect on SHAPE reactivity resembles that of Lin28 alone, suggesting that Lin28 displaces the HBR to bind the terminal loop. In our *in vitro* binding assays, when pre-let-7d is pre-saturated with Lin28B, the ability of DGCR8 to bind the same RNA substrate does indeed decrease significantly (Figure 32C). Moreover, the cooperativity is reduced (Hill coefficient of 2.43 alone versus 1.83 in the presence of Lin28B), likely due to the loss of HBR/loop interactions while the dsRBD interactions remain intact. Many loop-binding proteins have been implicated in regulation of miR biogenesis, and they are likely to interfere with the HBR/loop interactions.

2.4 Discussion

2.4.1. Heme-driven bipartite recognition of pri-miRNAs

We present a model for how heme plays a critical role in reliable recognition of pri-miRs by Drosha and DGCR8 (Figure 33). A pri-miR contains two ss-dsRNA

junctions at each end of the stem, and Drosha must bind the basal junction to make the correct cut. The inherent structural symmetry near the branch points poses a unique challenge for Microprocessor to distinguish the basal junction from the apical junction. We reveal that heme-dependent miRNAs lack robust basal junctions to reliably recruit Drosha, resulting in Drosha reversing its orientation to bind the apical junction. In this tug-of-war between the junctions, heme tips the balance; heme confers upon DGCR8 the specificity for the terminal loop, thereby increasing the fidelity of the bipartite recognition. What is notable is that a small (<1 kDa) molecule—heme—is capable of globally altering the orientation of a large multiprotein complex (~300kDa).

2.4.2. Heme activates DGCR8 to recognize the terminal loop structure

For the previous decade, heme binding has been associated with dimerization of DGCR8 but its molecular role in pri-miR processing has been unclear. Here we show that heme does not change stoichiometry; rather, it converts DGCR8 to an active conformation capable of recognizing the terminal loop. The change of DGCR8 conformation is coupled with a remodeling of the terminal loop, manifested as a SHAPE hyperreactive hotspot. Previous studies identified a UGU motif near the terminal loop as a binding site for the HBR, and mutating the UGU motif in miR-30a was shown to increase processing errors (106). However, this experiment involved simultaneous mutation of 4 nucleotides, one of which is immediately adjacent to the hotspot we identify in this study. Since only < 30% of pri-miRNAs contain the UGU motif, how the HBR interacts with the remaining pri-miRNAs is unknown. All pri-miRNAs we have

tested exhibit a heme-dependent SHAPE hotspot, unless technical difficulties (such as reverse transcription artifacts) prevent detection. The unique SHAPE hyperreactivity caused by the HBR/terminal loop interaction identified in this study suggests a stable and specific protein/RNA complex with a distinct conformation.

Here we uncover the RNA substrate specificity of DGCR8. Upon binding heme, DGCR8 gains the ability to recognize the loop structure of pri-miRNAs. The binding specificity relies heavily on the constrained geometry of the terminal loop, because a break in the loop, which releases the tension, leads to a significant loss in affinity. As our results show, Drosha often fails to identify the correct ss-dsRNA junction. However, heme allows DGCR8 to distinguish the apical loop from the basal junction, effectively enabling Microprocessor to break the symmetry of pri-miRNAs and support high-fidelity processing.

2.4.3. MicroRNAs vary in their dependence on heme

We also show that heme dependence varies among different pri-miRNAs. Heme-dependent pri-miRNAs contain stronger Drosha binding sites in the apical junction rather than in the basal junction. Thus, not all wild type basal junctions are ideal binding sites for Drosha, and recognition of pri-miRNAs with suboptimal basal junctions is driven by heme-activated DGCR8. The extent of heme dependence is thus encoded into the pri-miR sequence, and may serve as a regulatory handle. Therefore, different heme states, such as changes in oxidation state or gas levels, are likely to alter the relative levels of mature miRNAs.

2.4.4. Heme provides an opportunity to regulate miRNA biogenesis

Heme can enhance processing of most pri-miRNAs, and we observe a more dramatic effect on heme-dependent pri-miRNAs containing weak basal junctions. The important role of heme in general pri-miR processing opens various avenues for miRNA regulation. One way to regulate the HBR function is through direct competition, such as through loop-binding proteins. We show that Lin28B, an oncogenic stem cell factor, inhibits production of mature let-7d—a tumor suppressor miRNA—by competing against the HBR for the terminal loop. Various other loop-binding proteins associated with different signaling pathways have been described to modulate Drosha activity, may also affect HBR function. The ability of heme to allosterically modulate protein function has been observed in other nucleic acid-binding factors such as the Rev-Erb family and Bach1 (117, 118), and the effect of heme binding can be influenced by heme availability, oxidation state, and gas binding. Our identification of different classes of heme-dependence will help future investigations on varying biological contexts might regulate specific miRNAs at the post-transcriptional level.

Figure 7. Recombinant Drosha and DGCR8. (A) Domain organization and construct diagrams for Drosha and DGCR8. Recombinant protein complexes were purified after coexpressing the highlighted regions (magenta). Coexpression of Drosha with DGCR8Heme, DGCR8Apo, and DGCR8C352S yields MPHeme, MPApo, and MPC352S, respectively. CED central domain, RIIIDribonuclease III domain, dsRBD double-stranded RNA-binding domain, HBR heme-binding region, CTT C-terminal tail. (B) SDS-PAGE showing MPHeme, MPApo, and MPC352S used in this study.

Figure 8. UV-VIS profiles of DGCR8 with and without heme. UV-VIS profile for DGCR8 containing heme (black) and DGCR8 containing the C352S mutation (red), which abrogates heme binding. Absorbance peaks associated with heme binding, at wavelengths 366 nm, 450 nm, and 557 nm, are indicated.

Figure 9. In vitro pri-miR processing assay with MP^{Heme}, MP^{Apo} and MP^{C352S}. Reactions involving MP^{Heme} (130 nM), MP^{Apo} (130 nM) and MP^{C352S} (260 nM) are shown. Correct product is indicated with a black asterisk, incorrect product is indicated with a red asterisk.

Figure 10. Sequence gels for mapping alternative cleavage sites. (A-B) Representative images of sequencing polyacrylamide/urea gels used to identify Microprocessor cleavage sites, using (A) 5' and (B) 3' end-labeled pri-miR-143. RNAse T1 (not shown) and alkaline hydrolysis ladders were used to map the cut sites for MP^{Heme} and MP^{C352S}. Cut site associated with MP^{Heme} and MP^{C352S} are shown in black and red, respectively.

Figure 11. Heme flips Microprocessor orientation on pri-miRNAs. (A) In vitro pri-miR processing assays used to orient the RIIID domains of Drosha on a pri-miR, with diagram (B) shown to the right. Substrate shown is pri-miR-21. (C) Model depicting the heme-induced reversal of Drosha/DGCR8 complex

Figure 12. In vitro screen for heme dependence. 5′ end-labeled in vitro processing assays comparing activities of MPHeme (130 nM) and MPC352S (130 nM) for a series of pri-miRs. Substrates are arranged from most heme independent (left) to most heme dependent (right).

Figure 13. Cotransfection scheme for cell-based processing assays shown in Figures 14-17. Constructs containing pri-miR-9-1 and pri-miR-125a in tandem, or pri-miR-21 and pri-miR-125a in tandem, were cotransfected with either full-length FLAG-DGCR8WT or FLAG-DGCR8C352S into 293T cells.

49

Figure 14. Mature miRNA levels are affected by the C352S mutation. (A) Taqman quantitative PCR assay results from transfections described in Figure 13. Mature miR levels were normalized to miR-125a expression levels. Data are represented as mean ± standard deviation from three biological replicates. (B-C) Western blots for transfections described in Figure 13, for (B) pri-21 and (C) pri-9-1 experiments.

Figure 15. Splinted ligation assays support a miR-specific role for heme. (A-B) Representative splinted ligation assay results detecting miR-125a and miR-21 (A) or miR-9-1 (B). U6 levels from Northern blots were used for normalization. (C-D) Statistical analyses of splinted ligation results, where miR levels with mutant DGCR8 were normalized to those with wild-type DGCR8. Data are represented as mean ± standard deviation, for a total of three biological replicates. *p < 0.05, **p < 0.005, n.s. (not significant), p > 0.05, Student's t-test (two-sided, paired)

Figure 16. Quantitative PCR of primary transcript levels for transfections in Figure 13. Levels were normalized to pri-miR-125a levels

Figure 17. Heme-dependent changes in pre-miR levels support a miR-specific role for heme. (A-B) Representative Northern blots for pre-miR-21 (A) and pre-miR 9-1 (B). U6 levels confirm normalized loading in each lane. (C-D) Quantitation of Northern blots as shown in (A) and (B). Data are shown as mean ± standard deviation, for a total of three biological replicates.

Figure 18. Processing assays reveal a kinetic role of heme in pri-miRNA processing. (**A-F**) 5' end-labeled in vitro pri-miR processing assays for (A-B) pri-miR-125a, (C-D) pri-miR-9-1, and (E-F) pri-miR-21, in which MP^{Heme} and MP^{C352S} were titrated into reactions containing pri-miR substrate. Processing efficiency was quantified by measuring the fraction of signal for the product indicated, compared to the total lane signal. Data are represented as mean ± standard deviation. Protein concentrations are (left to right): 0.5 nM, 1.0 nM, 2.0 nM, 4.1 nM, 8.1 nM, and 16.3 nM.

Figure 19. Alternative processing is driven by Drosha, not DGCR8. (A) SDS-PAGE of recombinant MP$^{\Delta HBR}$, MP^{CTT}, DGCR8^{Heme} and DGCR8^{C352S} stained with Coomassie Blue. (B-C) 5′ end-labeled in vitro processing assays of pri-miR-21 (B) and pri-miR-125a (C), using MPHeme (130 nM), MPC352S (130 nM), MPΔHBR (260 nM), and MPCTT (1.04 μM). Correct (black asterisk) and incorrect (red asterisk) products are indicated. DGCR8ΔHBR lacks the HBR but contains the tandem dsRBDs and C-terminal tail. Deletion of HBR and dsRBDs yields the C-terminal tail (DGCR8CTT).

Figure 20. Drosha preferentially binds to different junctions of different pri-miRNAs. (A-B) EMSAs comparing binding affinities of MPΔHBR to basal and apical junctions of pri-miR-21 (A) and pri-miR-125a (B). Schematic diagrams of the RNA are shown next to free RNA bands (arrow), with mature sequence shown in light blue. The basal junction RNAs are annealed RNA duplexes lacking the terminal loop. The apical junction RNAs are the pre-miR fragments. Protein concentrations are (left to right): 3.33, 6.67, 10, 13.32, and 16.67 nM.

Figure 21. Additional screening of pri-miRNAs with DGCR8 truncations. Additional 5' end-labeled in vitro processing assays comparing DGCR8 titrations (see Figure 3D). MP^Heme (130 nM), MP^C352S (130 nM), MP^ΔHBR (260 nM), and MP^CTT (1.04 μM) reactions are shown.

Figure 22. Biophysical analysis of DGCR8 stoichiometry. (A) Representative radial scans for MP^{Heme}, with residuals shown below for fit quality. Movement of the boundary region is indicated by color spectrum, with earliest time points in violet and later time points in red. Data curves were created using GUSSI. (B) Graph showing c(S) analysis and sedimentation coefficients for DGCR8^{Heme}, DGCR8^{C352S}, MP^{Heme} and MP^{C352S}. Curves were created using GUSSI. (C) SEC-MALS analysis reports a molecular mass of 121.6 kDa for the heme-binding mutant DGCR8^{C352S}, indicating stable dimeric state (calculated mass of dimer: 120.9 kDa). (D) SEC-MALS analysis reports a molecular mass of 252.7 kDa for the heme-binding mutant MP^{C352S}, indicating stable 1:2 Drosha:DGCR8 stoichiometry (calculated mass of 2:1 complex: 239.8 kDa)

Figure 23. Heme induces an alternative conformation of DGCR8 dimers. (A) Stain-free SDS-PAGE showing proteolytic digestion of DGCR8Heme and DGCR8C352S. Differential degradation pattern (red arrows) suggests an altered conformation. (C) Representative negative stain electron micrographs of a MPHeme and b MPC352S after crosslinking with glutaraldehyde, with zoomed-in view of the selected particles shown underneath. (D) Thermal denaturation of DGCR8 in heme-saturated vs. heme-deficient states shows that the melting temperature (Tm) increases by 1.5 °C in the presence of heme.

Figure 24. Heme activates DGCR8 to bind pri-miRNAs. EMSA with 5′ end-labeled pre-miR-143 shows that DGCR8Heme binds with an increased affinity compared to that of the DGCR8C352S mutant. "−" lane contains no protein. Protein concentrations are (left to right): 0.13, 0.26, 0.52, 1.04, and 2.08 μM

Figure 25. Heme induces SHAPE hyperreactivity in pri-let-7d. (A) SHAPE analysis of pri-let-7d, showing hyperreactivity associated with MPHeme binding. Mature sequence is shown in light blue. Control reactions containing no SHAPE reagent (benzoyl cyanide) are shown in lanes 1–3. (B) SHAPE analysis of pri-let-7d with MPHeme by capillary electrophoresis. (C) Secondary structure diagram (right) depicts the position of the SHAPE "hotspot" (red).

Figure 26. Heme-bound DGCR8 is responsible for SHAPE hyperreactivity. SHAPE analysis of pri-let-7d, indicating that addition of 1 µM hemin chloride to Microprocessor incapable of binding heme does not restore hyperreactivity observed with heme-saturated Microprocessor. (B) SHAPE analysis of pri-let-7d, with full-length Drosha and full-length DGCR8 purified from 293T cells. Control reactions containing no SHAPE reagent (BzCN) are shown (lanes 1-2). (C) SHAPE analysis of pri-let-7d with DGCR8. Only the reactions containing the SHAPE reagent are shown, and (unshifted) nucleotide marker lanes are shown to the right.

A pri-miR-30a

B pri-miR-143

C pri-miR-18a

D pri-miR-125a

E pri-miR-9-1

F pri-miR-21

G pri-miR-93

H

Alignment of SHAPE hotspots

Sequence	miR
GUGGUGUGGAGUCUUC	miR-9-1
UGAAGUAGAUUAGCAUCU	miR-18a
CUGUUGAAUCUCAUGG	miR-21
CUGUGAAGCCACAGAUGGG	miR-30a
UGUGAUUACCCAACCU	miR-93
GGACAUCCAGGGUC	miR-125a
CAGUUGGGAGUC	miR-143
UUAGGGCAGGGAUUUUGCCCACAAGGAGGUAA	let-7d

Figure 27. Pri-miRNA screen for heme-induced SHAPE hyperreactivity. SHAPE analysis of pri-miR (A) -30a, (B) -143, (C) -18a, (D) -125a, (E) 9-1, (f) -21, (G) -93, (H) -30a, and (I) -143 with secondary structure diagrams shown to the right. SHAPE hotspot associated with wild type MP^Heme is marked in red. Mature sequence is shown in light blue. All secondary structure diagrams were designed using mfold. (H) Diagram summarizing SHAPE results for a series of pri-miRs, aligned according to the site of highest reactivity marked with red font on yellow background. The sequence of the apical fragment released by Dicer is shown for each miR.

Figure 28. DGCR8/loop interaction may be partially sequence-specific. (A-B) EMSAs comparing binding affinities of DGCR8[Heme] for (A) wild-type pre-let-7d and (B) a pre-let-7d point mutant in which the "hotspot" adenine has been mutated to a cytosine. (C-D) EMSAs comparing binding affinities of DGCR8[Heme] for (C) wild-type pre-miR-9-1 and (D) a pre-miR-9-1 point mutant in which the "hotspot" adenine has been mutated to a uracil. Protein concentrations were (left to right): 33.3 nM, 66.7 nM, 100 nM, 133.3 nM, 166.7 nM, and 200 nM.

Figure 29. Altering the base identity of the SHAPE hotspot affects reactivity. Sequencing gel showing SHAPE hyperreactivities for the let-7d hotspot in the wild type (left) or A49C mutant (right) variants. Mutating the hotspot to a cytosine leads to a moderate decrease in hyperreactivity that correlates with a decrease in binding affinity for DGCR8.

Figure 30. Heme-activated DGCR8 recognizes loop structure. EMSAs comparing the binding affinities of DGCR8Heme and DGCR8C352S for (A) pre-miR-9-1 and (B) pre-miR-9-1 containing a central break in the terminal loop. Diagrams of RNA sequences are shown to the left. Protein concentrations are (left to right): 33, 50, 67, 83, 100, 117, 133 , 150, 167, 183, 200, 217, 233, and 250 nM. All secondary structure diagrams were designed using mfold

Figure 31. Footprinting supports a DGCR8/loop interaction in pri-miR-9-1. Quantitation of hydroxyl radical footprinting experiments for pri-miR-9-1 and DGCR8[Heme], at a final protein concentration of 32.5 nM (light blue) and 130 nM (dark blue). From left to right, the bars represent nucleotides in the 5' to 3' direction. The hyperreactive nucleotide from SHAPE, A33, is indicated with an asterisk.

Figure 32. Lin28 may modulate pri-miRNA processing by targeting the HBR. (A) SHAPE analysis comparing hyperreactive sites for MP[Heme] and Lin28B on pri-let-7d. Lin28B produces a hotspot at a guanosine (G47, orange), two nucleotides 5' to the MP[Heme] hotspot (A49, red). Assembling a MP[Heme]/Lin28B/pri-let-7d complex (equimolar concentrations of MP[Heme] and Lin28B) results in the hotspot from Lin28B dominating. (B) Cartoon illustration showing Lin28/preE-let-7d complex (PDB ID: 3TRZ). Lin28 (purple) engages in extensive interactions with G47 (orange) of pre-let-7d fragment (gray). The black arrow indicates the 2' hydroxyl group of G47. The SHAPE hotspot associated with MPHeme (A49) is shown in red. (C) EMSA comparing binding affinity of DGCR8[Heme] for pre-let-7d, in the absence (left) or presence (right) of Lin28B. Lin28B concentration was 521 nM. Concentrations of DGCR8 were (left to right): 33 nM, 65 nM, 130 nM, 260 nM, 521 nM.

Figure 33. Model for heme-dependent pri-miRNA recognition by Drosha and DGCR8. Each pri-miR contains two ssRNA/dsRNA junctions where one is a stronger Drosha-binding site (indicated by "strong" or "weak"). Heme induces a conformational switch in DGCR8 that enables it to bind the terminal loop with higher specificity and affinity. Pri-miRs containing weak basal junctions are more heme dependent because placement of Drosha at the proper junction is driven by heme-bound DGCR8. For pri-miRs containing strong basal junctions, the interaction between the HBR and terminal loop can reinforce Drosha/pri-miR interactions to enhance processing efficiency

| | predicted mass for monomeric DGCR8 (kDa) | predicted mass for dimeric DGCR8 (kDa) | svAUC | | | | SEC-MALS | | |
			Sed. coeff. (S)	Molar mass (kDa)	Frictional ratio	RMSD	Molar mass by RI (kDa)	Zimm R^2	DGCR8 state
DGCR8^{Heme}	61.1	121.5	4.39	119	1.73	0.0049	ND	ND	dimer
DGCR8^{C352S}	59.8	120.9	4.05	120	1.94	0.0069	121.6	0.9894	dimer
DGCR8^{Apo}	60.5	120.9	ND	ND	ND	ND	125.3	0.9995	dimer
MP^{Heme}	179.4	240.5	7.50	230	1.59	0.0054	ND	ND	dimer
MP^{C352S}	179.4	239.8	7.31	215	1.57	0.0053	252.7	0.9962	dimer
MP$^{\Delta HBR}$	149.2	178.5	ND	ND	ND	ND	151.5	0.9125	monomer

Table 1. svAUC and SEC-MALS results for DGCR8 and Drosha/DGCR8.
Table summarizing results from svAUC and SEC-MALS experiments.

Chapter 3. Structural basis for pri-miRNA recognition by Microprocessor

Authors:

Alexander C. Partin, Kaiming Zhang, Byung-Cheon Jeong, Emily Herrell, Shanshan Li, Wah Chiu, Yunsun Nam

Author Contributions

Y.N. conceptualized the project; W.C. supervised the cryo-EM study. A.C.P prepared the samples, performed *in vitro* assays, and analyzed XLMS data; K.Z. and A.C.P. optimized grid preparation; K.Z. collected cryo-EM data, processed images and performed 3D reconstruction; A.C.P., K.Z., and S.L. built and refined the models; B.J. established initial purification protocols; E.H. assisted with sample preparation; S.L. assisted with cryo-EM data collection.

3.1 Introduction

MicroRNAs (miRNAs) constitute a major class of small non-coding RNAs that regulate gene expression throughout normal development and also in many pathological processes. To generate functional ~22nt-long metazoan miRNAs, primary microRNA transcripts (pri-miRNAs) undergo specific cleavage events by Drosha to generate precursor microRNAs (pre-miRNAs), which are further processed by Dicer (56, 59, 61). The initial step by Drosha requires another RNA binding protein, DGCR8 (45, 72, 83, 132). Active Microprocessor is known to contain one copy of Drosha and two copies of DGCR8 (106, 133). Both proteins contain domains that typically bind RNA, including the RNase III domains (RIIIDs) of Drosha and dsRNA binding domains (dsRBDs) of both proteins. DGCR8 also contains a heme-binding region that gains specificity and affinity for terminal loops of pri-miRNAs upon binding heme (133, 134). Short RNA motifs have been observed to affect cleavage efficiency and/or location, and the length of the stem has also been shown to be an important determinant of a suitable pri-miRNA substrate (101-104). Although partial structures of isolated Drosha and DGCR8 have been determined by X-ray crystallography (74), how the polypeptides work together to recognize and process pri-miRNAs accurately remains unclear.

3.2 Methods

Protein expression and purification

Drosha/DGCR8 complexes were expressed using the Bac-to-Bac system (Thermo Fisher Scientific) in HighFive cells. Human Drosha (353-1372, wild type sequence for fully docked and with E1045Q/E1222Q mutations for the partially docked structure) and human DGCR8 (223-751 for fully docked, and 175-751 for partially docked structures) fragments were cloned into pFastbacDual, with a hexahistidine tag and FLAG tag, respectively. Drosha/DGCR8 complexes were purified by Ni-NTA affinity chromatography. Lysis and washing were performed using a buffer containing 20 mM Tris (pH 8.0), 1 M NaCl, 10% glycerol, and 1 mM DTT. Cation exchange chromatography was performed using a 100-800 mM NaCl gradient buffered with 20mM Bis-Tris (pH 7.0), followed by size-exclusion chromatography. Human DGCR8 (223-751) containing an N-terminal hexahistidine tag was cloned into pET21 and expressed in BL21 (DE3) Rosetta cells. DGCR8 was purified using the same strategy as for Drosha/DGCR8 complexes.

RNA transcription and purification

RNA templates were cloned into pRZ vectors containing self-cleaving ribozymes on either end to produce homogenous ends (5' hammerhead, 3' hepatitis delta virus), and *in vitro* transcribed RNA fragments were purified from denaturing PAGE, as described previously (122).

Preparation of cryo-EM samples

RNA and proteins were mixed at 2:1 ratio and dialyzed overnight at 4°C against a dialysis buffer containing 20 mM HEPES (pH 7.1), 75 mM NaCl, 5 mM DTT, and 2 mM $CaCl_2$ ($CaCl_2$ was added for the fully docked state only). The RNA-protein complex was purified away from extra RNA, using a Superdex200 size-exclusion column in dialysis buffer. Fractions from the RNA-protein peak were pooled and concentrated to obtain a measurement of protein concentration using A_{450nm}. The RNA/protein complex was diluted to 1 μM in dialysis buffer. The complex was then crosslinked by adding 3 mM DSG dissolved in DMSO and then quenched with 75 mM glycine. The samples were then purified by an additional round of SEC in dialysis buffer, and concentrated for grid preparation.

Cryo-EM data collection

Three microliters of the Drosha/DGCR8/pri-miR-16-2 complex (WT at 4.2 mg/ml, mutant at 3.7 mg/ml) were applied onto 200-mesh R3.5/1 Quantifoil grids, which were glow discharged for 35 seconds using PELCO easiGlow (TED PELLA, INC.) at a plasma current of 15 mA. The blotting paper was standard Vitrobot filter paper Ø55/20mm, Grade 595 (TED PELLA, INC.). To improve particle distribution, 0.02% Nonidet P-40 substitute and 0.05% octyl-Beta glucoside were added prior to freezing. The grids were blotted for 4 s and rapidly cryocooled in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) at 4°C and 100% humidity. The samples were screened using a Talos Arctica cryo-electron microscope (Thermo Fisher Scientific) operated at 200 kV and then imaged in a Titan Krios cryo-electron microscope (Thermo Fisher Scientific) with GIF

energy filter (Gatan) at a magnification of 130,000x (corresponding to a calibrated sampling of 1.06 Å per pixel) using Stanford-SLAC Cryo-EM Facilities. Micrographs were recorded using EPU software (Thermo Fisher Scientific) with a Gatan K2 Summit direct electron detector, where each image was composed of 30 individual frames with an exposure time of 6 s and a dose rate of 7.8 electrons per second per $Å^2$. A total of 12,681 movie stacks for fully docked state and 6,070 movie stacks for partially docked state were collected, with defocus ranges of -1.3 to -3 μm and -1.5 to -3.6 μm, respectively.

Image processing

All micrographs were motion-corrected using MotionCor2(135) and the contrast transfer function (CTF) was determined using CTFFIND4(136). All particles were autopicked using the NeuralNet option in EMAN2(137) and manually checked, yielding 1,385,678 particles from selected 12,455 micrographs for fully docked state and 1,063,710 particles from selected 5,994 micrographs for partially docked state. The particle coordinates were imported into Relion(138), where multiple rounds of 2D classification were performed to remove poor quality 2D class averages. The initial model was built using the "ab-initio 3D" program in cryoSPARC(139). A total of 937,668 particles for fully docked state and 842,037 particles for partially docked state were used for 3D classification in Relion to remove the poor 3D classes. Next, 505,640 particles for fully docked state and 381,468 particles for partially docked state were imported into cryoSPARC, where the final refinements were performed using the "Non-uniform refinement" program, achieving the 3.7-Å map with mask (4.6 Å without mask)

for fully docked state and 4.4-Å map with mask (5.8 Å without mask) for the partially docked state. The resolution for the final maps was estimated by the 0.143 criterion of the FSC curve (See more information in Table 2).

Model Building, Refinement and Validation

The crystal structures of Drosha (PDB entry 5B16) and the DGCR8 core (PDB entry 2YT4) were fitted into the density of the fully docked state structure by rigid body fitting using Chimera(140). The dsRBDs of Drosha and DGCR8-2 were rearranged to fit the density. The *de novo* model building of RNA and Drosha (residues 459-462, 501-521, 668-674, 712-849, 930-957, and 1334-1356) was performed using Coot(141). Additionally, the RNA building was guided by mfold secondary structure predictions(142). Real-space refinement was then performed using PHENIX, with secondary structure restraints for protein as well as RNA. Model geometries were assessed using Molprobity(143). Maps and structures shown in the figures were generated with PyMOL, Chimera and Coot. After docking the Drosha crystal structure, the density near the active site was closely analyzed to identify the scissile phosphate and calcium ions. In particular, the RIIIDb active site contains larger additional density in the region typically associated with magnesium binding. This density was modeled with a calcium ion, and the final models including the calcium ion were evaluated by Q-score(144). To compare the RNA-protein interactions in fully docked and partially docked MP/RNA structures, the script github/pdbfairy was used. For partially docked state, the model from the fully docked complex was rigid-body fitted into the map of the

partially docked complex using Chimera, further optimized in Coot, and followed by real space refinement in Phenix with geometry restraints.

Negative stain electron microscopy

Drosha/DGCR8/pri-miR-16-2 complexes were crosslinked using the crosslinkers indicated in the figures, for 10 minutes at room temperature, at a protein concentration of 1 μM. Negative stain grids were prepared using 2% uranyl acetate (sample concentration 0.03 mg/mL) and the images were obtained using a Tecnai G2 Spirit TEM (FEI) running at 120kV, and using a magnification of 49,000x.

In vitro pri-miRNA processing assays

5' end-labeling was performed using T4 Polynucleotide Kinase (NEB) and γ-^{32}P-ATP. Unless otherwise indicated, pri-miRNA processing assays were performed in 15 μL reactions containing approximately 1 nM end-labeled RNA substrate, 30 mM Tris (pH 7.5), 67 mM NaCl, 5% glycerol, 10 mM $MgCl_2$, 5 mM DTT, 8 U RNase inhibitor, and 1.5 μg yeast tRNA. Unless otherwise indicated, reactions were incubated for 10 minutes (except for assays comparing arm truncations, which were incubated for 15 minutes) at room temperature, stopped with 0.7% SDS and 24mM EDTA, and treated with 0.38mg/mL proteinase K at 50° C for 30 minutes. Reactions were analyzed by denaturing PAGE using either a phosphorimager (for end-labeled) or SYBR Green II (Thermo Fisher Scientific)(for unlabeled). For quantified reactions, the assays were performed in triplicate, and the gels were analyzed using ImageLab software (BioRad) to quantify

percent processed by densitometry. Quantified results were plotted and statistically analyzed using Prism 8 (Graphpad). For sequencing gel analysis, the reactions were run on an 8% denaturing sequencing PAGE for approximately 3.5 hours at a constant current of 30 mA. For activity assays on pre-assembled RNA-protein complexes, a reaction containing 1.7 μM MP-RNA complex and $MgCl_2$ (concentration as indicated) was incubated at 37° C for 20 minutes. The reactions were then processed similar to the end-labeled reactions, and visualized using SYBR Green II.

Gel-shift assay (Electrophoretic Mobility Shift Assay, EMSA)

5' end-labeled RNA (approximately 1 nM) was incubated with a dilution series of protein samples in 20 mM Tris (pH 7.5), 67 mM NaCl, 10 mM DTT, 1000 μg/mL yeast tRNA, and 10% glycerol. The mixture was analyzed by native PAGE and visualized with a phosphorimager.

Crosslinking Mass Spectrometry (XLMS)

RNA/protein complexes were prepared as described above. RNA/protein complexes were diluted to 1 μM by measuring absorbance at 450 nm. Samples without RNA were prepared similarly but in a buffer containing 20 mM HEPES (pH 7.1), 1 M NaCl, and 5 mM DTT. 1:1 mixtures of DSS and deuterated DSS (DSS-d4, Proteochem) were prepared in DMSO. Crosslinking reactions were performed at room temperature for 10 minutes, and quenched with 75 mM glycine. The crosslinked samples were concentrated, run on SDS PAGE, and stained with Coomassie Blue. Crosslinked bands were cut and

submitted for mass spectrometry analysis by the UTSW Proteomics core facility. The samples were analyzed on a Fusion Lumos (Thermo Fisher Scientific) mass spectrometer coupled to a Dionex UltiMate 3000 RSLCNano LC system. The data were analyzed using the xQuest/xProphet pipeline(145). The results were then sorted by ID-score and false discovery rate (FDR). Links shown in the crosslink maps consist of all inter-peptide Drosha-Drosha hits with an FDR $\leq$ 0.05, and Id-Score $\geq$ 20.

Quantification and Statistical Analysis

Cryo-EM validation was performed and is shown in Figures 40-41 and 65-66. Gold standard FSC plots (calculated in CryoSPARC), Euler angle distributions (calculated in CryoSPARC), Q-scores for individual residues of Drosha and RNA, and per-residue cross-correlation coefficient (calculated in Phenix) are shown for each structure.

Crosslinking mass spectrometry was performed using xQuest/xProphet, which utilizes the target-decoy strategy (146) to determine the False Discovery Rate (FDR) for each hit. Only hits with FDR scores < 0.05 were considered, and three individual replicates of each experiment was performed. Quantified *in vitro* processing assays were performed in triplicate. To compare the processing of each Drosha mutant compared to the wild type control, unpaired t-tests were performed.

3.3 Results

3.3.1 Development of an optimal MP/RNA complex for cryo-EM studies

Based on our findings from Chapter 2, we focused on identifying an optimal substrate by characterizing the apical junction and basal junction separately. We screened apical junctions by performing EMSAs with pre-miRNAs and DGCR8, with the goal of identifying the tightest binders (Figure 34). We then screened pri-miRNAs for heme-dependence, with the understanding that heme-independent substrates have optimal Drosha-binding sites as the basal junction (Figure 35) (133). Assessing basal junction by processing rather than binding assays was necessary, because measuring the basal junction/Drosha interaction by EMSA involves assembling an RNA duplex, which often fails to produce homogeneous products.

We chose pri-miR-16-2 as the substrate RNA for structure determination, because it not only binds Drosha/DGCR8 tightly enough to copurify during size-exclusion chromatography, but is also processed correctly regardless of heme presence (Figure 36A). Heme-independent processing suggests that pri-miR-16-2 has an optimal Drosha-binding site at the basal dsRNA-ssRNA junction, as described in Chapter 2 (133). We included 20 nucleotides flanking the Drosha cleavage sites, as they were enough to maintain robust processing. Furthermore, to capture the pre-catalytic state while inhibiting RNA cleavage we replaced the magnesium ions required for RNase III activity with calcium (Figure 36B).

To derive a structural model for an active Microprocessor/pri-miRNA (MP/RNA) assembly, we aimed to prepare a homogenous protein/RNA complex (Figure 37). We purified recombinant protein complexes using the truncated constructs of Drosha and DGCR8 that are sufficient for processing pri-miRNAs *in vitro* in the correct orientation

when fully saturated with heme. Because the complex dissociates readily during grid preparation, we treated the complex with a crosslinking agent to enable structure determination by cryo-electron microscopy (cryo-EM). We chose DSG after screening numerous crosslinkers. DSG was highly reactive, and produced more intact particles on negative stain compared to other crosslinkers (Figure 38). Later screens using cryo-EM with our collaborator further supported the finding that short amine-amine crosslinkers, especially DSG, were the best choice for MP/RNA complexes. Numerous attempts at obtaining intact particles without crosslinking were unsuccessful; this included screening a multitude of protein and RNA constructs, additives such as poly-L-lysine and detergents, numerous types of grids, buffer conditions and freezing conditions.

3.3.2. Structure Determination and Modeling

Using the MP/RNA sample as described above, we collected cryo-EM datasets. Despite severe particle heterogeneity, we determined a cryo-EM structure of an active human Microprocessor in complex with pri-miR-16-2 (Figure 39-41). The overall resolution is approximately 3.7 Å (Table 2). The quality of the map varies for different regions, with higher resolution for Drosha and the majority of the pri-miRNA, and lower resolution for DGCR8 and the apical loop of the substrate RNA (Figure 40E). We used crystal structures of isolated Drosha and DGCR8 dsRBDs to aid with model building (74, 86). The regions resolved in the crystal structure of Drosha in isolation aligns well with the cryo-EM density, except for the dsRBD, which swings toward the bound RNA (Figures 42-43). In addition, many regions of Drosha (221 residues out of 908 included in

our final model) were modeled *de novo*, because they could not be resolved in the isolated Drosha structure (74). The cryo-EM map for the apical region, especially for the heme-binding region (HBR) of DGCR8, is resolved to lower resolution. However, using the crystal structures of the dsRBDs (residues 489-583 and 592-701 for dsRBD1 and dsRBD2, respectively) we could perform initial rigid-body docking followed by refinement to reveal their spatial disposition with respect to Drosha and the bound pri-miRNA.

We were also able to visualize the density connectivity and thus model *de novo* the majority of pri-miR-16-2, except for the apical loop and a few nucleotides near the 5' and 3' ends (Figure 44). Determining the register of the RNA was guided by the secondary structure calculated by mfold (142), as the location of the mismatches are obvious in the cryo-EM map (Figure 45A). The location of the RIIID active sites coincides with the known Drosha cleavage site on pri-miR-16-2, according to miRBase (147). We used the Q-score to evaluate the density resolvability and the quality of the model per residue, for both Drosha and RNA, and it is consistent with the expected scores for the reported resolution (144) (Figures S2H-S2L). Furthermore, we observe extra density with high resolvability Q score (144) near the active sites, especially in the RIIIDb domain, most likely due to ordered calcium ions (Figure 40G) (54, 148).

3.3.3. Overall Structure

As suggested by MALS and other data (133, 149)(Table 1), the overall structure consists of a heterotrimeric protein complex consisting of one copy of Drosha and two

copies of DGCR8, and one pri-miRNA hairpin. We were unable to model the HBR, due to poor resolution and lack of available structures for docking. However, the HBR is positioned at the terminal loop of pri-miR-16-2, which is consistent with the currently accepted model. The two copies of the DGCR8 "core" adopt different dsRBD conformations on RNA, leading to 3 dsRBDs engaging the same helical turn of RNA. The first copy (DGCR8-1, yellow-orange) adopts the same conformation observed in the crystal structure, in which the two dsRBDs form a stable interface against one another. Only the N-terminal dsRBD forms contacts with pri-miR-16-2, while the C-terminal dsRBD is pointed outward. The second copy (DGCR8-2, purple) shows the two dsRBDs split apart from one another and engaging in direct interactions with adjacent faces of the apical RNA stem. The different conformations of the two copies of DGCR8 indicates that the dsRBD core is dynamic, which may explain the FRET results observed by the authors of the crystal structure (PDB: 2YT4). The linkers connecting the dsRBD cores to their C-terminal tails is not visible, likely due to high flexibility.

The RNA stem is positioned near the catalytic RIIID sites of Drosha, in a pre-catalytic state. It appears that internal mismatched bases do not significantly disrupt the A-type helical structure of the dsRNA. As mentioned above, the calcium density is much stronger near RIIIDb compared to RIIIDa. This asymmetry in metal binding may be because calcium binds the RIIIDs differently than the natural ligand, magnesium. RIIIDb may also be innately better than RIIIDa at binding divalent cations. Interestingly, RIIIDa contains a less common asparagine (N1042) at a position occupied by aspartate in other RNaseIII domains. The latter hypothesis is intriguing since E1147—one of the key

coordinating residues in RIIIDb—is a mutation hotspot for Wilms tumor patients, while RIIIDa mutations are much less frequently found in patients (25).

The cryo-EM structure reveals the detailed interactions between Drosha and the bound pri-miRNA, and various regions are likely to undergo RNA-induced folding, as they could not be modeled in the isolated crystal structure (Figure 43). Most of the RNA stem docks against the globular core of Drosha composed of the RIIIDs and the Central domain (CED). On the opposite side of the core domains, various parts of Drosha rearrange or become folded to wrap around the RNA, consequently increasing direct contact with the RNA. The dsRBD of Drosha swings inward to bind the basal stem region of the pri-miRNA . Previously unresolved regions in CED, such as a long segment that includes two antiparallel helices (757-848, green) (Figure 45B), and a stretch with two shorter helices (930-957, light green) are visible in our cryo-EM complex structure (Figure 46), likely because of the intimate contact with RNA to rigidify them. Moreover, additional newly visualized regions such as the basal tip of the CED and the C-terminal tail of Drosha are likely to depend on the regions that become ordered with RNA for folding. Together, our cryo-EM structure of Microprocessor with fully docked pri-miRNA elucidates how several distinct polypeptide regions are involved in coordination of binding the RNA substrate to ensure stable and specific interactions.

The largest consecutive region of Drosha that we modeled *de novo* includes a two-stranded coiled coil (753-848) that we refer to as the "Belt" (Figure 46). Previously this region of Drosha was proposed to fold into a "PAZ-like" domain similar to Dicer (74), but instead it forms a two-helix segment that binds the basal junction. The Belt acts

like a seatbelt that straps across the passenger RNA and buckles into a 4-way junction consisting of the dsRBD, the C-terminal tail, the Wedge and the Belt itself (Figure 47). The Belt mostly interacts with the ssRNA immediately next to the dsRNA, to clamp down the separated RNA strands. When the Belt crosses over the unwound RNA immediately next to the dsRNA, the tip of the helices interacts favorably with the dsRBD and the C-terminal tail of Drosha (Figure 48). In this conformation, the α-helices of the Wedge and the C-terminal tail bundle with the last helical segment of the RIIIDb. The linker between the dsRBD and the RIIIDb is flexible in the absence of RNA, but with such multi-domain interactions, the connection between the dsRBD and the RIIIDs becomes more rigidified. Thus, when the Belt locks into the 4-way junction, the RNA-binding modules become more rigidly coupled, strengthening the interactions with substrates that meet the structural requirement. Consequently, formation of the multi-domain buckle provides specificity for the ssRNA/dsRNA junction by positive reinforcement. To validate our structural model, we used crosslinking mass spectrometry to probe the Microprocessor conformation in solution (150) (Figure 49). Few crosslinks are observed for the MP/RNA complex, but two of the most robust are between the Belt and the dsRBD, which come in close contact in the observed conformation in the cryo-EM structure (Figure 49C, Table 3). The same interdomain linkages are not observed when the crosslinking reagents are added to MP without RNA, suggesting that the Belt/dsRBD interaction as observed in the structure is specific to the state in which the pri-miRNA is properly docked.

Analysis of the MP/RNA complex structure allows us to visualize the relative locations of various domains of Drosha and DGCR8 and thus to interpret their functional roles in recognition of a pri-miRNA, as the direct contacts can be mapped for almost the entire length of the stem (Figure 44).

3.3.4. The Helical Belt is important for pri-miRNA processing

The cryo-EM structure of the MP/RNA complex provides a physical framework to understand how various features of the pri-miRNA might affect processing. Most of the dsRNA is well-resolved in our cryo-EM structure, and we have modeled a total of 33 base-pairs (bp): 21 out of 23 total bp on the apical side and 12 bp on the basal side of the 5' cut site. The required lengths of dsRNA on both sides of the cleavage sites determined by previous studies are thus explained by the RNA-binding protein domains that cover such distances (100, 102-104, 106). To measure the stem length consistently, both ends are marked as dsRNA-ssRNA junctions. On the apical end, the terminal loop is capped by the HBR and the four dsRBDs surround the apical end of the stem. The basal side is clamped down via the Belt, which locks into the 4-way junction. Between the two clearly marked ends, the dsRBDs from Drosha and DGCR8-1 interact to form a contiguous, elongated protein mass that measures the entire length of a stem of about 35 bp, acting as a "dsRBD ruler" (Figure 50). If we model a longer dsRNA protruding basally, the double-helix structure is likely to clash with the Belt when the stem is only 2 bp longer (Figure 51A). For a typical pri-miRNA with a dsRNA-ssRNA junction, the Belt keeps the 5' flanking arm away from the 3' strand. Previously a "Bump helix" (residues

910-919) was proposed to block unbranched dsRNA from binding Drosha (74). Although Arg914 of the Bump helix makes contact with the phosphate backbone of the 3' ssRNA (Figure 51A), it is directed away from the RNA. Thus, while the Bump helix may support the unwound conformation by stabilizing the hold on the 3' ssRNA arm, it does not appear to present a "block" against dsRNA as previously suggested. In summary, the cryo-EM structure of the MP/RNA complex highlights the roles of the dsRBD ruler and the 4-way junction (including the Belt) in recognizing the correct stem-loop for processing.

When we insert base pairs into the stem of pri-miR-16-2 near the basal junction, we observe a precipitous decline in processing efficiency when more than 1 bp is inserted (Figures 51B-C). The endonucleolytic cleavage that occurs for the mutant pri-miRNAs is at the same location as the wild type pri-miRNA, whether the flanking regions are short (20nt) or long (60-70nt), suggesting that additional mechanisms are in place for choosing the cut site in pri-miR-16-2 (Figure 51D). When the stem is shortened by a mere 1 bp, the enzymatic activity is also dramatically reduced (Figure 51C). Shorter stems would result in the dsRBD ruler hanging over the edge of the dsRNA or deform, thereby preventing the formation of a proper 4-way multi-domain junction. Together, these results suggest that the Belt/dsRBD connection acts as a positive checkpoint; in the presence of a dsRNA-ssRNA junction 12-13 bp away from the 5' cut site, the Belt clamps down and forms a productive 4-way junction to stabilize the tightly bound conformation of Drosha on pri-miRNA.

Given its critical contact with the basal dsRNA-ssRNA junction, we investigated whether the Belt affects the processing of pri-miR-16-2. When the two helical segments are deleted (ΔBelt), the processing efficiency is dramatically reduced (Figure 52). The first helix (H1) interacts with the groove formed by the most basal portion of the stem, and the basic side chains (K792, K795, and K799) of H1 face the phosphate backbone to support favorable interactions. Mutating all three to glutamates (K3E3) also reduces the Microprocessor activity. Moreover, mutating the side chains of N806 and Q1336 to Asp and Glu, respectively, is also detrimental to pri-miRNA processing, likely due to charge repulsion in the Belt/C-terminus contact area. The mutant Microprocessor complexes were purified over multiple chromatography steps and have similar K_D values for pri-miR-16-2 as the wild type Microprocessor, except for the MP^{K3E3}, which results in ~ 2 fold decrease in RNA affinity (Figure 53). When the enzyme concentration is increased by 2 fold, MP^{K3E3} shows similar activity as MP^{WT}, but the other mutant complexes still exhibit largely impaired enzymatic activity (Figure 52E). Moreover, when the substrate concentration is increased by 2 orders of magnitude, $MP^{\Delta Belt}$ remains less efficient than MP^{WT} (Figure 54). Therefore, missing the Belt is likely to debilitate MP activity on pri-miR-16-2 in a way that affects more than the overall affinity for pri-miRNA. We also investigated the possibility of the Belt affecting stem length specificity. Inserting or deleting base pairs in the stem did not improve the activity of $MP^{\Delta Belt}$ over MP^{WT}, over a wide range of substrate concentrations (Figure 55). Thus, the Belt is critical for detecting the basal RNA structure, and strengthening the 4-way junction upon binding appropriate substrates, thereby enhancing processing.

Many protein or RNA features have been shown to affect processing of distinct miRNAs differently (133, 134). Therefore, we tested the importance of the Belt in processing of other miRNAs by using the Microprocessor lacking the two helices (ΔBelt). For most miRNAs, Drosha cleavage is less efficient in the absence of the Belt, though to varying degrees (Figure 56A). For miR-107 and miR-106b, we also observe a shift in cut site location (Figure 56B). Interestingly, some pri-miRNAs (pri-miR-30a and pri-miR-125a) are processed similarly with and without deletion of the Belt, and this independence from the Belt is observed even at lower enzyme concentrations (Figure 56C). It is possible that defects associated with Belt loss for pri-miR-30a and pri-miR-125a are mitigated by other RNA features such as the presence of primary sequence motifs (102, 103), as well as their heme-independence (133), which suggests the presence of a favorable basal junction. Therefore, the Belt helices of Drosha generally contribute to both efficiency and accuracy of Microprocessor, and the specific effect on each miRNA may vary.

3.3.5. The Wedge and Belt of Drosha form a narrow tunnel for the ssRNA at the basal junction

When the basal junction is clamped down by the Belt, it forms a narrow tunnel with another region of CED which we call the "Wedge" (928-957) (Figure 57). The hollow is lined with basic residues to mediate interactions with the acidic phosphate backbone. The Wedge is disordered in the isolated Drosha structure, but it is induced to fold in our complex as it wedges into the crevice between the stem and the unwound ssRNA near the

5' end. Moreover, the Wedge also contacts the "GHG" motif (GUG in pri-miR-16-2) of pri-miRNAs that has been proposed to affect cleavage efficiency and accuracy, although the interactions are through the RNA backbone (Figure 57C) (103, 105). The mismatched U may be recognized via a unique change in the backbone structure. The GUG sequence is located near the basal end of the dsRNA, and the dsRBD of Drosha is also situated proximally, as predicted from modeling from typical mode of binding for dsRBDs (75, 105). However, the distance between the GUG nucleobases and the dsRBD is relatively large, and there is little evidence for sequence-specific binding. The only interaction we observe is a weak contact between the Uracil base in the "H" position and Q1266 of Drosha. Therefore, the GHG motif is likely recognized through its impact on the RNA structure, rather than through direct contact with the nucleobases. As part of the 4-way junction (Figure 47), the Wedge lies directly adjacent to the GHG motif, with the C-terminal tail of Drosha packed over its surface. A mismatch at the H position would likely enable local deformability and could lead to more favorable interactions with the Wedge, thereby strengthening the 4-way junction and thus facilitating processing. Although the GHG motif may represent a special case of processing enhancement, general RNA flexibility in the area basal to the Drosha cleavage sites has been associated with processing efficiency (151).

The Wedge, together with the Belt, forms intimate contact with the first few nucleotides in the ssRNA region (Figure 58A). On each ssRNA strand, one nucleotide (Ura7 on 5' and Ade99 on 3') is best modeled in a flipped-out conformation (Figures 58B-C). Drosha stabilizes the two nucleobases in pockets with pi-pi stacking with various

side chains, and these pockets are formed at the interface between the Belt and the Wedge. As a narrow groove is formed between the dsRNA and the 5' ssRNA arm, the concentrated negative charge from the phosphates is also mitigated by the basic side chains from the Wedge (R938, K939, and K940). When mutations are introduced to disrupt a nucleobase binding pocket or the electrostatic interactions with the backbone, there is a markedly reduced processing without affecting the overall affinity for pri-miR-16-2 (Figures 59-60). Therefore, formation of a narrow tunnel is likely to ensure the presence of ssRNA flanking regions, to help Drosha to distinguish pri-miRNAs from other dsRNAs or dsRNA-containing RNAs.

Given that the length of the tunnel is not long, we investigated how many single-stranded flanking nucleotides are needed for processing of pri-miR-16-2. For both of the flanking arms, we determined that 13 nucleotides (counting from the Drosha cut site) are necessary for robust processing (Figure 61). This means that only 1 or 3 single-stranded nucleotides are critical on 5' and 3' ends, respectively. Although we observe such short essential arms for pri-miR-16-2, longer arms may be necessary for other miRNAs, especially if the basal junction is not a strong binding site for Drosha. Interestingly, the flipped nucleotides mark where the seemingly less critical regions of ssRNA begin (Figure 58). Since pri-miRNAs normally have longer arms *in vivo,* the ability of the Belt and the Wedge to stabilize the flipped nucleobases may be important to maintain separation of the ssRNA strands, thus making them easier to clamp down.

3.3.6. Cryo-EM structure of Drosha/DGCR8 with partially docked pri-miRNA

Most of the Drosha/RNA contacts are between the pri-miRNA stem and the two RIIIDs, and divalent cations are known to be important for mediating RIIID interactions with RNA (54). However, even in the presence of EDTA or mutations of key metal-coordinating residues (E1045Q/E1222Q), the overall affinity for pri-miRNA is not reduced significantly (Figure 62). This is surprising, because divalent cations serve as a bridge to enable the RNA to be brought close to the highly acidic "catalytic valley" by mitigating charge-charge repulsion (54). Drosha RIIID residues involved in magnesium coordination have also been found to be frequently mutated in Wilms tumor patients (25). To gain insight into the architecture of the inactive assembled MP/RNA complex, we determined a cryo-EM structure of Microprocessor containing mutant Drosha (E1045Q/E1222Q), at ~ 4.4 Å-resolution (Figures 63-66, Table 2). Altering the acidic side chains abrogates the ability to coordinate divalent cations that are needed to mediate contacts with the pri-miRNA stem and thus enable hydrolysis. The two cryo-EM structures are best compared by superimposing the dsRBDs (2 from each DGCR8 and 1 from Drosha); the five dsRBDs exhibit similar relative orientations in the two structures, and aligning the domains also results in good agreement between the two RNA conformations (Figure 67A). Therefore, the arrangement of all of the dsRBDs on pri-miR-16-2 is preserved in both structures. However, the core domains of Drosha—the RIIIDs and CED—move away from the RNA, reducing direct interactions with the substrate. When the two structures are aligned by Drosha core domains, there are two major conformational changes: the Belt swings inward to where the RNA usually binds, and the dsRBD also swings away from the core, likely to retain its contact with the

dsRNA (Figure 67B). Thus, without a properly docked pri-miRNA, the helical Belt can no longer interact with the 4-way junction across the RNA; instead, it leans against the core of Drosha on the other side of RNA. Comparing the two structures provides a model for how the Belt buckles into the 4-way junction. In the fully bound state, the C-terminus of Drosha rearranges to pack against the Belt and Wedge, establishing the 4-way junction and stabilizing an alpha-helical structure in the C-terminus. The Wedge, which is not visible in the crystal structure, favors a random coil structure in the partially docked state. Upon basal junction docking, the Wedge moves basally to contact the basal junction as well as the major groove adjacent to the basal junction, and rigidifies against the RNA to adopt a more alpha-helical structure. As described above, the dsRBD of Drosha maintains a similar footprint on the RNA stem in both cryo-EM structures, despite a significant rearrangement relative to the rest of Drosha. The dsRBD in the crystal structure forms extensive crystal contacts and likely adopts a different conformation in solution prior to RNA binding. We tested if the observed change in Belt conformation is also evident in solution, using crosslinking mass spectrometry. When we use catalytically inactive Drosha to form MP/RNA complexes, we do not detect the Belt-dsRBD crosslinks observed in the active state (Figure 68). Therefore, even when RNA is associated with MP, the RIIIDs require properly coordinated divalent ions to support the activated conformation where the Belt reaches across the ssRNA to interact with the rest of Drosha.

Despite the conformational changes in Drosha, DGCR8 and the dsRBD of Drosha maintain remarkably similar conformations in both cryo-EM structures (Figure 69). The

footprints of the dsRBDs on pri-miR-16-2 (conserved in both structures) span nearly the entire length of the stem region (Figure 70). Consistent with such extensive contact, the affinity of DGCR8 for pri-miR-16-2 is high (Figures 71A, C). In contrast, when we test a Microprocessor complex lacking any RNA-binding regions of DGCR8, (Drosha^{+CTT}, containing only the C-terminal tails of DGCR8 to increase Drosha solubility (106)), we observe no detectable affinity for the RNA substrate without added divalent ions (Figures 71B, D). Therefore, in the partially docked (calcium-free) structure, what we observe is likely a state in which both specificity and affinity for RNA are driven by the HBRs and the dsRBDs, mostly from the DGCR8 polypeptides. After we crosslink the wild type MP/RNA complex in the absence of calcium (similar to the mutant MP/RNA complex), adding magnesium can still activate its ability to cleave the pri-miRNA similar to the uncrosslinked complex (Figure 72). Thus, the partially docked conformation is likely capable of transitioning to the fully docked conformation. To drive the complex to the fully docked conformation, the coordinated divalent cations are necessary to switch on the specificity of Drosha for dsRNA (Figure 73). Similar experiments involving a Drosha^{+CTT} variant lacking the Belt (Drosha$^{+CTT/\Delta Belt}$) showed that calcium cannot activate pri-miRNA binding if the Belt is removed. Rather, the Belt itself, and its association with the other components of the 4-way junction, enables pri-miRNA binding by Drosha (provided that the charge-charge repulsion between the catalytic valley and the RNA is first mitigated by calcium or magnesium). Surprisingly, we observed the same trend for a pri-miRNA that was dependent on the Belt for processing (pri-miR-16-2, Figures 74A-B), and a pri-miRNA that was processed similarly with and without the Belt (pri-miR-16-

2, Figures 74C-D) However, even in the presence of calcium, Drosha^{+CTT} fails to bind RNA lacking the appropriate basal junction geometry. Although Drosha^{+CTT} binds well to WT and +1bp variants of pri-miR-16-2, the affinity is dramatically lower for +2bp, -1bp and -2bp variants (Figure 75). This is consistent with the effects of stem insertions on processing efficiency (Figures 51C, 55), and suggests that Drosha cannot fully engage substrates without proper belt closure.

In spite of the binding observed with Drosha^{+CTT}, it is noteworthy that we observe little difference in overall RNA-binding affinity between MP^{WT} with calcium, MP^{WT} without calcium, and MP$^{E1045Q/E1222Q}$ (Figure 62). The presence of numerous RNA-binding domains makes dissecting individual affinity contributions challenging. The partially docked structure provides an opportunity to examine the RNA/protein interactions at the basal and apical regions separately. Therefore, by comparing the affinities of the individual components (Figure 71) and analyzing conserved contacts in both structures (Figure 70), we conclude that the HBRs and the dsRBDs have specificity and affinity for pri-miRNAs independent of the conformational state of Drosha. Furthermore, comparing the two structures also reveals the extent of flexibility in interdomain linkers (eg. to dsRBD and Belt of Drosha) that permit large domain movements within a MP/RNA complex.

3.4 Discussion

6.4.1. A model for understanding Drosha/DGCR8 assembly on pri-miRNAs

Our results allow us to propose a physical model for how a pri-miRNA interacts with the Drosha/DGCR8 complex (Figure 75). The robust specificity of Microprocessor for the apical junction is driven by DGCR8; the HBRs bind the terminal loop and the dsRBDs of DGCR8 surround the apical half of RNA stem, as suggested by biochemical studies described in Chapter 2. The dsRBD of Drosha recognizes the dsRNA on the basal half, and in tandem with the first dsRBD of DGCR8-1, forms a ruler to measure ~35 bp of RNA stem. Similar RNA binding modes are observed for the HBRs and the dsRBDs in both the fully docked and partially docked structures, and they are sufficient for nanomolar binding affinity. In the presence of divalent cations such as magnesium or calcium, the catalytic domains RIIIDa and RIIIDb also engage with the pri-miRNA stem. Proper docking of the pri-miRNA in the active sites positions the basal dsRNA-ssRNA junction near the two-helix Belt, which can then establish favorable interactions with other regions of Drosha (dsRBD, C-terminal tail, and Wedge) to form the 4-way junction. The productive Belt rearrangement stabilizes Drosha in a conformation that favorably contacts the dsRNA-ssRNA junction, thus reinforcing substrate RNA specificity. The specific and tight contacts via the HBRs and the dsRBDs near the apical junction, combined with the molecular clamp imposed by the helical Belt and the Wedge at the basal end, together ensure that the bound RNA has two junctions separated by a stem of particular length (35 ± 1 bp (103)). Therefore, our physical model reveals many features of the Microprocessor that are important to ensure that a proper substrate is loaded and poised for hydrolysis at the specific nucleotide linkages.

In addition to providing a detailed view of an active MP/RNA complex, the fully docked structure can be compared to the partially docked state to provide insight into how the macromolecules assemble. Furthermore, examining the isolated Drosha crystal structure also shows the flexible nature of the important RNA-contacting regions (e.g., dsRBD, Wedge, Belt), most of which are disordered or conformationally distinct from our MP/RNA structures. Although the structures provide static pictures, the ability of the dsRBDs to maintain the same contact with the pri-miRNA, even in the absence of Drosha core domain binding, suggests that the dsRBDs have robust affinity and independent specificity. Thus, the dsRBDs (and the HBR domains) are likely able to drive the initial loading of the pri-miRNA, especially in the case of miR-16-2. This is supported by our previous findings (Chapter 2), which showed that the HBR specifically engages the terminal loop, even in the absence of Drosha (Figure 26C).

Since the discovery of Drosha, many models have been proposed to describe how pri-miRNAs are distinguished from other RNAs. The characteristic features of pri-miRNAs have been investigated in various ways. First, the basal (lower) and apical (upper) stem lengths have both been suggested to affect substrate selection and cut site determination (100, 102-106). As shown in the current study with $MP^{\Delta Belt}$, a lack of proper basal junction detection can manifest as a loss in efficiency, shift in cut site, or both. Although the resultant processing defect may vary for each pri-miRNA, the biochemical data converge on the importance of the stem length. Our cryo-EM structures provide a framework to explain why the length of the dsRNA between the junctions is

physically measured by the size of the polypeptide domains and their arrangement along the RNA.

3.4.2 The role of primary sequence of pri-miRNAs

Short sequence motifs in pri-miRNAs have also been shown to impact processing efficiency, although deriving a universal model for their importance was difficult (102, 103). A "GHG" motif has been highlighted for its effect on processing efficiency, and our structural observations show that the GHG motif is at the center of the 4-domain junction— the "buckle". Although we do not observe base-specific contacts, the unique RNA backbone geometry near the motif suggests that it contributes to the RNA structure that is important to support the formation of the 4-way junction. A "UG" motif in the 5' arm has been proposed to play a role in determining processing efficiency. The unpaired Ura in this motif is in the same position as the 5' flipped Ura (Ura7) in our fully docked model. The binding pocket for this uracil base may favor Ura over other nucleotides. However, the pockets for both of the flipped bases rely mostly on ring stacking interactions and may not be highly discriminating for the nucleobase identity. In the 3' arm, the "CNNC" motif has been proposed to contribute to the processing of some pri-miRNAs, and DEAD-box helicases may act through this motif to remodel the RNA to make the cleavage event more efficient (102, 107, 108). For certain pri-miRNAs, the additional help in unwinding the flanking ssRNAs may be necessary for Drosha to then clasp the flexible RNA in the tunnel formed by the Belt and the Wedge. Variations in stem length, primary sequence motifs, and structural features in the flanking arms are

some of the elements that could influence the processing efficiency and accuracy of a pri-miRNA, as well as the degree to which it is sensitive to deletion of the Belt. The built-in plasticity of the Microprocessor complex shown through our cryo-EM structures is likely critical for its ability to accommodate and utilize the diversity among the pri-miRNAs, while simultaneously distinguishing them from other dsRNAs.

Figure 34. Screening of pre-miRNAs for optimal binding to DGCR8. K_D values derived from EMSAs for DGCR8 and 18 different pri-miRNAs. Values represent an average of 4 replicates. Error bars represent ± S.D.

Figure 35. Expanded in vitro processing assay screen for selection of ideal cryo-EM substrate. 5′ end-labeled in vitro processing assays comparing activities of MP^Heme (130 nM) and MP^C352S (130 nM) for a series of pri-miRs. Reactions containing a single major product, with similar results between MP^Heme and MP^C352S, were considered potential candidates for cryo-EM sample preparation.

Figure 36. Design of a Drosha/DGCR8/pri-miRNA assembly. (A) *In vitro* pri-miRNA processing assay of pri-miR-16-2 with either wild type Microprocessor (left) or containing the heme-free mutant DGCR8[C352S]. Protein concentrations are (left to right): 4, 16, and 65 nM. (B) *In vitro* pri-miRNA processing assays of pri-miR-16-2, demonstrating the inhibitory effects of calcium on Drosha activity. [MP] is 16 nM, except for the first lane. CaCl$_2$ concentrations are (left to right): 0, 0, 1, 2, 3, 4, 5, and 6 mM. MgCl$_2$ concentration is 8 mM.

Figure 37. Drosha/DGCR8 constructs used for cryo-EM structure determination. (A) Domain organization of Drosha and DGCR8, with the DGCR8 sequence depicted twice to represent its dimeric state in the complex. CED, central domain of Drosha. RIIID, RNaseIII domain. dsRBD, double-stranded RNA-binding domain. HBR, heme-binding region. CTT, C-terminal tail. The sequences used for cryo-EM structure determination are marked underneath. (B) Representative size-exclusion chromatogram for a MP/RNA complex. Fractions pooled for grid sample preparation are highlighted in blue. (C) SDS-PAGE of the concentrated MP/RNA complex

Figure 38. Selection of crosslinking agent for Drosha/DGCR8/pri-miRNA. Representative negative stain micrographs for MP/RNA complex with (A) 1 mM SPB followed by 5 min UV treatment, (B) 3 mM DSG, (C) 1 mM sulfo-SMPB, (D) 1 mM Sulfo-EGS, (E) 0.85 mM BMOE, and (F) 0.1% glutaraldehyde. Magnification 49,000x.

Figure 39. Cryo-EM maps (A) and models (B) of Drosha/DGCR8/pri-miR-16-2. Illustrations are colored by domain as shown in Figure 34A.

Figure 40. Workflow and data processing for Drosha/DGCR8/pri-miR-16-2 structure. (A) Workflow of the data processing. Total of 12,681 movie stacks were collected, motion-corrected using MotionCor2 and CTF-determined using CTFFIND4. All particles were autopicked and then checked manually from the selected 12,455 micrographs in EMAN2. Then a total of 1,385,678 particle images were derived and four rounds of 2D classification were performed to remove poor 2D class averages in Relion. The initial model was built in cryoSPARC. Afterwards, a total of 937,668 particles were used for 3D classification in Relion. A final dataset with 505,640 particles was used for 3D refinement in cryoSPARC. (B) Representative motion-corrected micrograph of the crosslinked active MP/RNA complex. (C) Reference-free 2D class averages computed in Relion. (D) Gold standard FSC plot for the 3D reconstruction of the active complex calculated in CryoSPARC. (E) Resolution map for the 3D reconstruction of the active complex. (F) Euler angle distribution of the particle images calculated in CryoSPARC. (G) Proposed calcium location near the active site of RIIIDb of WT Drosha with the Q score calculated. Calcium was only modeled in RIIIDb due to lack of signal near RIIIDa.

Figure 41. Analysis of map quality using Q scores. (A-B) Q-score for each residue in the model and map of Drosha (A) and RNA (B). The gray line represents the expected Q-score at 3.7 Å-resolution (0.46) based on the correlation between Q-scores and map resolution. (C-D) Per-residue cross-correlation coefficient between the model and map of Drosha (C) and RNA (D) calculated in Phenix. (E) FSC plot between the refined atomic model and 3D reconstruction of the active complex calculated in Phenix.

Figure 42. Model and map fit for Drosha/DGCR8/pri-miR-16-2. (A) Model of the MP/RNA complex shown separately and assembled. Nucleotides on either side of the cut sites are colored in magenta. (B) Individual cryo-EM densities and models of Drosha, two copies of DGCR8, and RNA from the MP/RNA structure. The N- and C-terminal dsRBDs of DGCR8 are labelled as "1" or "2", respectively.

Figure 43. Comparison of cryo-EM structure and previous crystal structure. (A) Domain organization of Drosha colored to show regions either newly identified or differently oriented from the crystal structure (PDB: 5b16). CED, central domain of Drosha. RIIID, RNaseIII domain. dsRBD, double-stranded RNA-binding domain. RIIIDs are colored dark gray for orientation. Positions of the active site residues E1045 and E1222 are marked by asterisks. (B) Superimposition of Drosha+CTT crystal structure (beige, PDB: 5b16) onto Drosha+CTT from MP/RNA structure, with the same color code as panel A.

Figure 44. RNA-protein interactions observed in MP/RNA structure. Secondary structure diagram of pri-miR-16-2 used for cryo-EM. Protein/ RNA contacts (distance < 4.5 Å) are highlighted according to the protein domain or region, with the same color scheme used in Figures 1A-C. Nucleotides that were not modeled due to poor resolution or disorder are colored gray. Drosha cut sites are indicated with black arrowheads. Pri-miR-16-2 in the model is numbered from 1-105 from 5' to 3'.

Figure 45. Representative density and model fitting for MP/RNA. (A) Cryo-EM density and model for pri-miR-16-2, showing views of the mismatched C15-U89 base pair in the lower stem (left) and the basal junction (right). (B) Cryo-EM density for Helix-1 of the Belt with model.

Figure 46. RNA-induced rearrangements in Drosha observed in cryo-EM structure. (A-B) Overall views of MP/RNA with front (A) and back (B) views. Colored regions highlight major conformational changes of Drosha when compared to the isolated Drosha structure, except for RIIIDs (dark blue) which are colored for orientation. Cleavage sites are colored magenta.

Figure 47. The 4-way junction in Drosha assembles over pri-miR-16-2. (A-B) Cartoon illustration of MP/RNA showing the RNA-induced 4-way junction (A), and rotated to show the interaction with RIIIDb (B). Same color scheme as Figure 2A. 4-way junction is indicated by a red circle

Figure 48. View of the interactions of the Belt with the rest of Drosha and RNA. Same color scheme as in Figure 44. The N-terminal and C-terminal helix of the Belt are labeled as H1 and H2, respectively

Figure 49. Crosslinking mass spectrometry analysis of MP/RNA. (A) Representative SDS PAGE showing crosslinking mass spectrometry sample before (left) and after (right) extraction of the crosslinked band ("Target"). (B) Top-scoring intra-Drosha crosslinks in MP/RNA complex (top) and MP alone (bottom). Red lines represent hits identified in 3 out of 3 replicates. Gray lines represent hits found in 2 of 3 replicates. (C) Residues involved in red crosslinks from MP/RNA in Figure 2G are shown with distances.

Figure 50. A dsRBD ruler measures stem length. (A-B) Cartoon and surface representation of MP/RNA showing distances from cut site to apical and basal branch points with front (A) and back (B) views.

Figure 51. The Belt is involved in measuring stem length. (A) Pri-miRNA structure compared to a model with an elongated stem. The Belt (green) clashes with the modeled dsRNA (right). (B) RNA secondary structure diagrams of pri-miRNA constructs used for Figure 3D, and S3. Insertions and deletions are colored in pink, and Drosha cut sites are indicated with red arrows. (C) *In vitro* pri-miRNA processing assays of pri-miR-16-2 with the RNA sequences shown in Figure S3A. Protein concentrations for each series are (left to right): 8, 32, and 130 nM. (D) Sequencing gel analysis of 1bp and 2bp insertions showing 1nt and 2nt shifts in product size, respectively, indicating that the cut sites are conserved. [MP] is 65 nM. 5' and 3' flanking segments were extended to 70 and 60 nt, respectively, to facilitate visualization. Results are not normalized to probe intensity and are presented solely to show shift in product size.

Figure 52. The Belt of Drosha is critical for pri-miRNA processing. (A) SDS-PAGE for Drosha mutants used in Figure 2. (B) Representative *in vitro* pri-miRNA processing assay of pri-miR-16-2 with a series of belt mutants. [MP] is 8 nM. A total of three replicates were performed, and quantified results are shown in Figure 3E. (C) Quantified *in vitro* pri- miRNA processing assays with a series of Belt mutants. Data are from three replicates, and error bars represent ± S.D. [MP] is 8 nM. *p < 0.001 (unpaired t-test for WT v. each mutant). (D) *In vitro* pri-miRNA processing assay of pri-miR-16-2 with a series of belt mutants, performed at 37° C. [MP] is 8 nM. (E) *In vitro* pri-miRNA processing assay of pri-miR-16-2 with a series of Belt mutants, performed at room temperature. [MP] is 16 nM.

Figure 53. Belt mutations do not significantly impair binding affinities to pri-miR-16-2. Electrophoretic mobility shift assays (EMSAs) showing the binding of mutant Drosha/DGCR8 complexes to pri-miR-16-2. Protein concentrations are (left to right): 2, 4, 8, 16, and 33 nM. RNA alone is shown on the top-left. Estimated dissociation constants (K_d) are listed below each gel.

Figure 54. MP^{ΔBelt} fails to properly process pri-miR-16-2. Representative unlabeled *in vitro* pri-miRNA processing assays of wild type pri-miR-16-2 with either wild type (left) or ΔBelt (right) Microprocessor. A total of three replicates were performed. RNA concentrations are 200 nM. Protein concentrations for each group are (left to right): 33, 65, 130, 260, and 521 nM. (B) Quantification of results shown in Figure 53A. A total of three replicates were performed.

Figure 55. MP$^{\Delta Belt}$ defects are not rescued by stem mutations in pri-miR-16-2. (A) 5' end-labeled *In vitro* pri-miRNA processing assays of pri-miR-16-2 with the RNA sequences shown in Figure S3A, using MP$^{\Delta Belt}$. Protein concentrations for each series are (left to right): 8, 32, and 130 nM. (B) *In vitro* pri-miRNA processing assays of unlabeled pri-miR-16-2 with the RNA sequences shown in Figure S3A, using MP$^{\Delta Belt}$. RNA concentrations are 200 nM. Protein concentrations for each series are (left to right): 32, 130, and 521 nM.

Figure 56. Deletion of the Belt produces a range of effects in different pri-miRNAs. (A) *In vitro* processing assays using wild type Microprocessor or the ΔBelt mutant on various pri-miRNAs. [MP] is 65 nM. (B) Sequencing gel analysis of the products of pri-miR-107 and pri-miR-106b from Figure 3F, to indicate the shifts in Drosha cut sites caused by deletion of the Belt. (C-D) *In vitro* pri-miRNA processing assays of (C) pri-miR-30a and (D) pri-miR-125a with increasing concentrations of MP^{WT} and MP^{ΔBelt}. Protein concentrations are (left to right): 1, 2, 4, 8, and 16 nM. Unprocessed substrate is shown on the left.

Figure 57. The Belt and Wedge clamp down on the basal portion of pri-miR-16-2. (A) Vacuum electrostatic potential surface depicting basal side of the tunnel formed by the Belt and Wedge. RNA is shown in orange. (B) Cartoon representation of MP/RNA showing pathway of the basal junction between the Belt and Wedge. (C) Same as Figure 4B, but rotated 20° to show interactions between the Wedge and GHG motif (black). Nucleotides of the GHG motif (G88, U89, and G90) are indicated by white circles.

Figure 58. Examination of the single-stranded flanking segments of pri-miR-16-2. (A) Overall view of Wedge and single- stranded nucleotides in flipped conformation near the basal junction. (B) Close-up views of interactions between Drosha and single-stranded A99. (C) Close-up views of interactions between Drosha and single-stranded U7

Figure 59. Mutagenesis suggests a critical role for the Wedge and other regions of CED in processing. (A) SDS-PAGE for Drosha mutants used in Figures S4B-F. RKK-AAA, R938/K939/K940A. (B) Representative *in vitro* pri-miRNA processing assays of pri-miR-16-2 with a series of Drosha mutants. [MP] is 8 nM. A total of three replicates were performed. (C) Quantified *in vitro* pri-miRNA processing assays with a series of Drosha mutants. Data are from three replicates, and error bars represent ± S.D. [MP] is 8 nM. *p < 0.001 (unpaired t-test for WT v. each mutant). (D) *In vitro* pri-miRNA processing assay of pri-miR-16-2 with a series of Drosha mutants. [MP] is 16 nM. (E) *In vitro* pri-miRNA processing assay of pri-miR-16-2 with a series of Drosha mutants, performed at 37° C. [MP] is 8 nM.

Figure 60. Wedge and CED mutants do not significantly impair binding affinities to pri-miR-16-2. Electrophoretic mobility shift assay (EMSA) results showing the binding of Drosha mutants to pri-miR-16-2. Drosha/DGCR8 to pri-miR-16-2. Protein concentrations are (left to right): 0, 2, 4, 8, 16, and 33 nM. RNA alone is shown on the left. Estimated dissociation constants (K_d) are listed below each gel.

Figure 61. Short regions of single-stranded flanking segments are critical for processing. Representative denaturing RNA gels for the pri-miR-16-2 variants with different lengths of flanking arms. The 5' and 3' arm lengths are indicated above each gel. Protein concentrations are (left to right): 2, 4, 8, 16 nM. Quantified results from three replicates are shown in Figure 4G and 4H. (B-C) Quantification of *in vitro* pri-miRNA processing results on pri-miR-16-2 variants containing shortened 5' (B) or 3' (C) flanking segments. [MP] is 8 nM. Data are from three individual replicates, and error bars represent ± S.D. (D) RNA secondary structure diagram showing the essential ssRNA nucleotides in orange and the nonessential ones in gray, with the same number system as in Figure 1D.

Figure 62. Presence of divalent cations at the active sites of Drosha do not affect affinity for pri-miRNA. (A-C) Electrophoretic mobility shift assays (EMSA) showing the binding of pri-miR-16-2 to either wild type (MP^{WT}) Microprocessor with 2mM CaCl$_2$ (A), wild type Microprocessor with 1mM EDTA (B), and the catalytically dead E1045Q/E1222Q mutant Microprocessor (MP$^{E1045Q/E1222Q}$) in the absence of divalent ions (C). Protein concentrations are (left to right): 2, 4, 8, 16, 33, and 65 nM. Free RNA is shown to the left on C, Estimated dissociation constants (K$_d$) are listed below each gel.

Figure 63. Assembly of a partially-docked MP/RNA complex. (A) Representative size-exclusion chromatogram for the catalytically inactive MP/RNA complex. Fractions pooled for grid sample preparation are highlighted in blue. (B) Stain-free SDS-PAGE of the concentrated partially docked MP/RNA complex.

Figure 64. Map, model, and fit of a partially-docked MP/RNA structure. Cryo-EM map (A) and model (B) of the partially docked Drosha/DGCR8/pri-miR-16-2 complex. The HBR regions of both DGCR8 monomers are gray in the map and not modeled. Each polypeptide of the cryo-EM map was segmented in Chimera, the contour levels for the display are 0.5 for Drosha, 0.4 for both DGCR8, 0.4 for pri-miR-16-2, and 0.4 for HBR. (C) Individual cryo-EM densities and models of Drosha, two copies of DGCR8, and RNA in the partially docked state.

Figure 65. Workflow and data processing for a partially-docked Drosha/DGCR8/pri-miR-16-2 structure. (A) Workflow of the data processing. Total of 6,070 movie stacks were collected, motion-corrected using MotionCor2 and CTF-determined using CTFFIND4. All particles were autopicked and then checked manually from the selected 5,994 micrographs in EMAN2. Then a total of 1,063,710 particle images were derived and four rounds of 2D classification were performed to remove poor 2D class averages in Relion. The initial model was built in cryoSPARC. Afterwards, a total of 842,037 particles were used for 3D classification in Relion. A final dataset with 381,468 particles was used for 3D refinement in CryoSPARC. (B) Representative motion-corrected micrograph of the crosslinked partially docked state of the MP/RNA complex. (C) Reference-free 2D class averages computed in Relion. (D) Resolution map for the 3D reconstruction of the partially docked complex. (E) Euler angle distribution of the particle images calculated in cryoSPARC. (F) Gold standard FSC plot for the 3D reconstruction of the partially docked complex calculated in CryoSPARC.
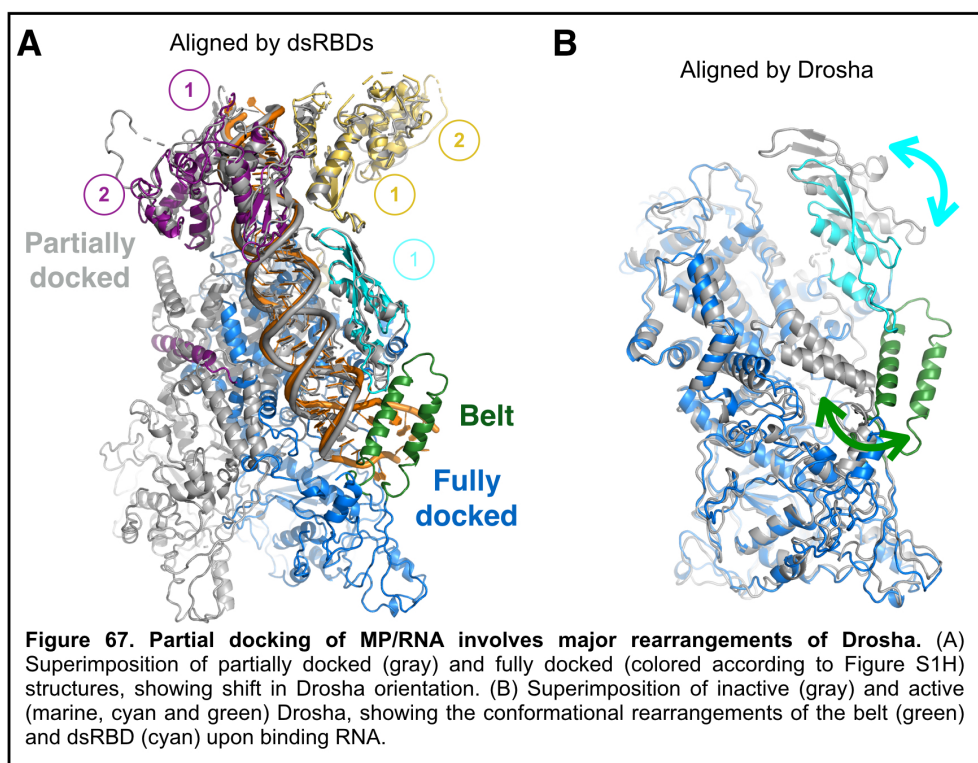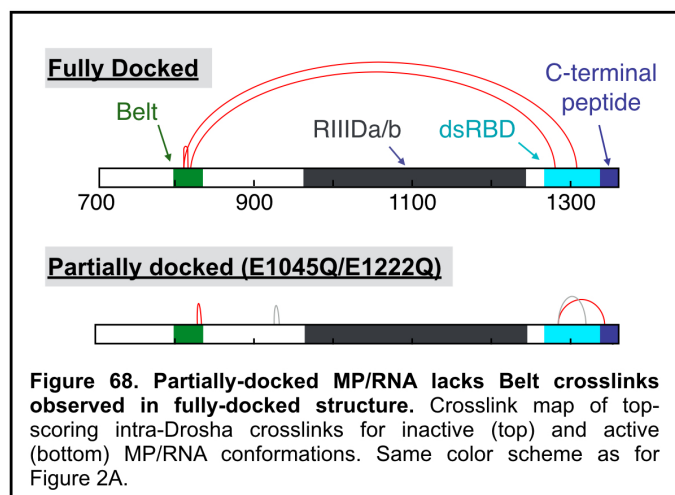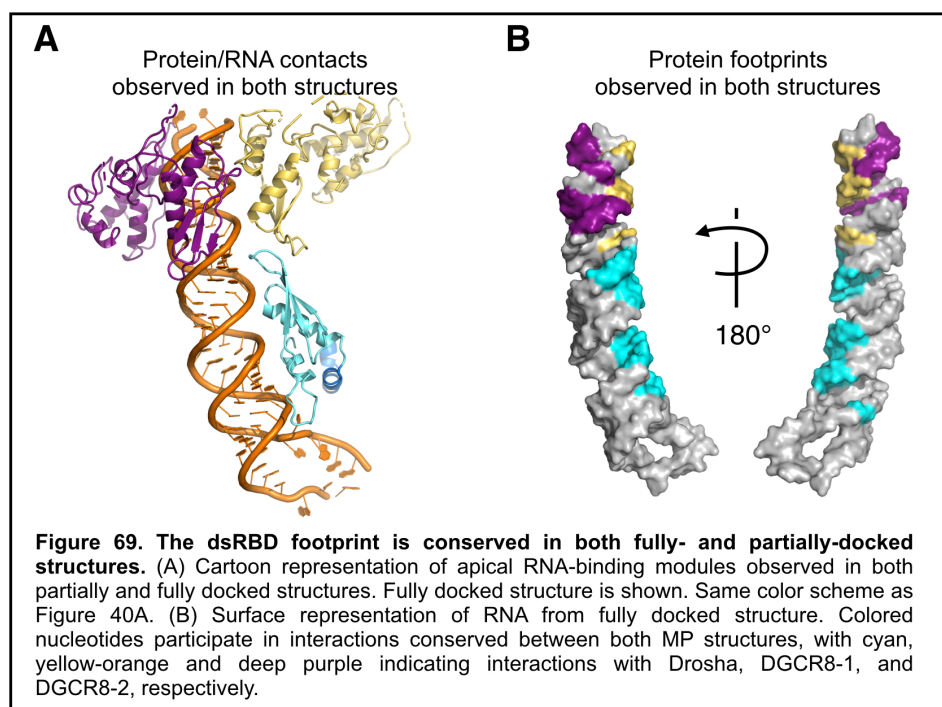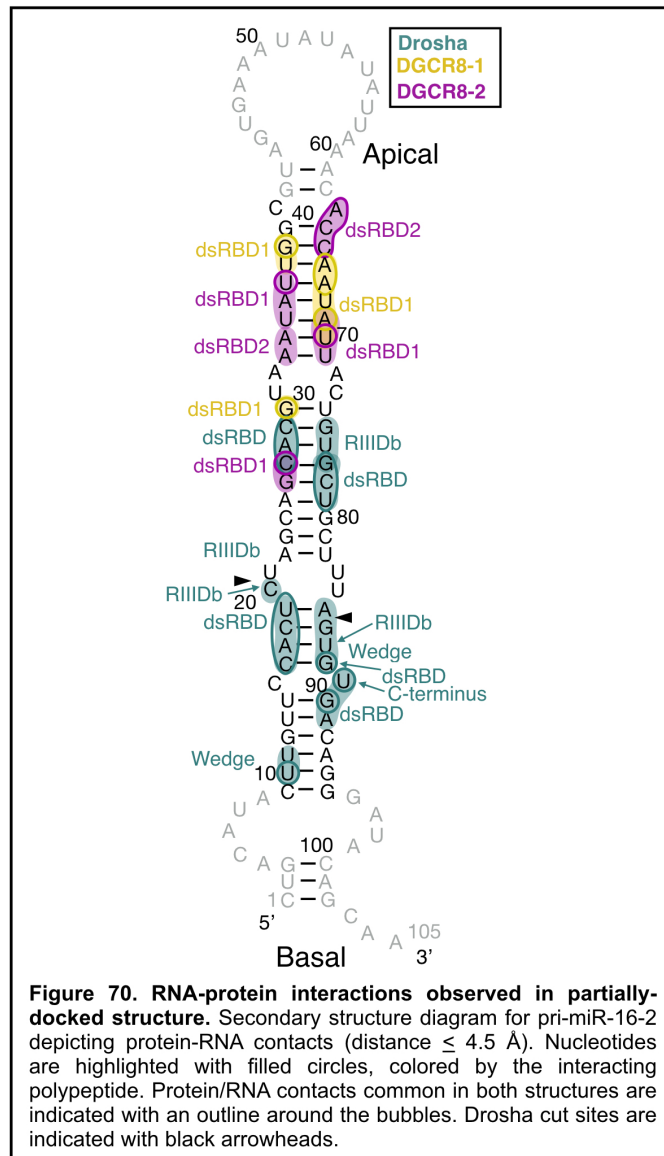
Figure 66. Analysis of a partially-docked map quality using Q-scores. (A-B) Q-score for each residue in the model and map of Drosha (A) and RNA (B). The gray line represents the expected Q-score at 4.4 Å-resolution (0.38) based on the correlation between Q-scores and map resolution. (C-D) Per-residue cross-correlation coefficient between the model and map of Drosha (C) and RNA (D) calculated in Phenix. (E) FSC plot between the refined atomic model and 3D reconstruction of the partially docked complex calculated in Phenix.

Figure 67. Partial docking of MP/RNA involves major rearrangements of Drosha. (A) Superimposition of partially docked (gray) and fully docked (colored according to Figure S1H) structures, showing shift in Drosha orientation. (B) Superimposition of inactive (gray) and active (marine, cyan and green) Drosha, showing the conformational rearrangements of the belt (green) and dsRBD (cyan) upon binding RNA.

Figure 68. Partially-docked MP/RNA lacks Belt crosslinks observed in fully-docked structure. Crosslink map of top-scoring intra-Drosha crosslinks for inactive (top) and active (bottom) MP/RNA conformations. Same color scheme as for Figure 2A.

Figure 69. The dsRBD footprint is conserved in both fully- and partially-docked structures. (A) Cartoon representation of apical RNA-binding modules observed in both partially and fully docked structures. Fully docked structure is shown. Same color scheme as Figure 40A. (B) Surface representation of RNA from fully docked structure. Colored nucleotides participate in interactions conserved between both MP structures, with cyan, yellow-orange and deep purple indicating interactions with Drosha, DGCR8-1, and DGCR8-2, respectively.

Figure 70. RNA-protein interactions observed in partially-docked structure. Secondary structure diagram for pri-miR-16-2 depicting protein-RNA contacts (distance $\leq$ 4.5 Å). Nucleotides are highlighted with filled circles, colored by the interacting polypeptide. Protein/RNA contacts common in both structures are indicated with an outline around the bubbles. Drosha cut sites are indicated with black arrowheads.

Figure 71. Pri-miRNA binding is mostly driven by DGCR8. (A-B) SDS-PAGE gels for human DGCR8 (A) and Drosha[+CTT] (B), which contains Drosha and a small C-terminal peptide of DGCR8. (C-D) Electrophoretic mobility shift assay (EMSA) results showing the binding of DGCR8 (C) and Drosha[+CTT] (D) to pri-miR-16-2. Protein concentrations are (left to right): 0, 0.13, 0.26, 0.52, 10.4, and 2.08 μM. Estimated dissociation constants (Kd) are listed below each gel.

Figure 72. The partially-docked state of MP/RNA is reversible. *In vitro* activity assays performed on pre-assembled wild type MP/RNA complexes, assembled and crosslinked in the presence of 1 mM EDTA. A DMSO-treated control complex, and a complex treated with 3mM of the crosslinker DSG, are shown. Purified pre-miR-16-2 is shown on the left as a reference. Mg^{2+} concentrations for each titration are: 0, 10, and 20 mM.

Figure 73. Drosha/pri-miRNA binding is dependent on divalent cations. Electrophoretic mobility shift assay (EMSA) results showing the binding of catalytically active Drosha[+CTT] to pri-miR-16-2, in the absence (left) or presence (right) of 2 mM $CaCl_2$. Protein concentrations for each series are (left to right): 0, 8, 16, 32, 65, 130, and 260 nM.

A pri-miR-16-2 (Belt-sensitive)

No divalent ions

B

+ 2mM CaCl$_2$

Drosha^{+CTT} Drosha$^{+CTT/\Delta Belt}$

Drosha^{+CTT} Drosha$^{+CTT/\Delta Belt}$

C pri-miR-125a (Belt-insensitive)

No divalent ions

D

+ 2mM CaCl$_2$

Drosha^{+CTT} Drosha$^{+CTT/\Delta Belt}$

Drosha^{+CTT} Drosha$^{+CTT/\Delta Belt}$

Figure 74. Drosha/pri-miRNA binding is dependent on Belt closure. (**A-B**) EMSA results showing the binding of Drosha with (Drosha^{+CTT}) and without (Drosha$^{+CTT/\Delta Belt}$) the Belt to pri-miR-16-2, in the absence (**A**) or presence (**B**) of 2 mM CaCl$_2$. (**C-D**) EMSAs for pri-miR-125a, with the same arrangement as panels A and B. Protein concentrations for each series are (left to right): 8, 16, 32, 65, 130, and 260 nM.

Figure 75. Belt closure and pri-miRNA binding is dependent on stem length. Electrophoretic mobility shift assay (EMSA) results showing the binding of catalytically active Drosha^{+CTT} to wild type pri-miR-16-2, and the +1bp, +2bp, -1bp, and -2bp variants shown in Figure 49B. Protein concentrations for each series are (left to right): 0, 16, 32, 65, 130, 260, and 521 nM. Reactions were assembled in the presence of 2 mM CaCl$_2$.

Figure 76. Proposed model for pri-miRNA recognition by Drosha. Different protein modules recognize specific structural features of pri-miRNAs, with the HBR recognizing the terminal loop, dsRBDs recognizing the stem, and the Belt and Wedge recognizing the basal junction. Flexibility between the modules enables recognition of diverse substrates that meet the structural requirements. The dsRBDs, the Belt and the Wedge, together form a molecular ruler that selects stem-loops with a length of approximately 35 bp. DGCR8 CTTs are removed for clarity.

Cryo-EM Map	Active state	Inactive state
Data collection and processing		
Microscope	Titan Krios	Titan Krios
Voltage (kV)	300	300
Camera	Gatan K2 Summit	Gatan K2 Summit
Magnification	130,000x	130,000x
Pixel size (Å)	1.06	1.06
Total exposure (e-/Å^2)	46.8	46.8
Exposure time (s)	6	6
Number of frames per exposure	30	30
Energy filter slit width (keV)	20	20
Data collection software	EPU	EPU
Defocus range (μm)	-1.3 - -3	-1.5 - -3.6
Number of micrographs	12,681	6,070
Number of initial particles	1,385,678	1,063,710
Number of particles for 3D analyses	937,668	842,037
Symmetry	C1	C1
Number of final particles	505,640	381,468
Resolution (0.143 gold standard FSC, Å) masked/unmasked	3.7/4.6	4.4/5.8
Local resolution range (Å)	3 - 7	4 - 8
Atomic model refinement		
Software	Phenix	Phenix
Number of protein residues	1341	1336
Number of nucleotide residues	78	66
Number of atoms	12593	12254
Geometric parameters (r.m.s.d.)		
Bond length (Å) (# > 4σ)	0.007	0.006
Bond angle (°)(# > 4σ)	1.061	1.113
Ramachandran statistics		
Favored (%)	88.35	86.48
Allowed (%)	11.50	13.52
Disallowed (%)	0.15	0
Validation		
MolProbity score	2.09	2.25
Clashscore	9.48	11.32
Rotamer outliers (%)	0.33	0.75
Crosscorrelation	0.83	0.81
B factor (Protein/Nucleotide)	132.85/127.69	184.94/200.75
C-beta deviation	0	0
EMRinger score	1.52	0.69
CaBLAM outliers (%)	6.77	8.36
Nucleic acid geometry		
Probably wrong sugar puckers	0	0
Bad bonds	0	0
Bad angles	0	0

Table 2. Cryo-EM data collection, processing, and model validation.

	Replicate ID	Site 1	Site 2	Id-Score
MP/RNA Fully docked (WT)	3	1277	817	40.8
	13	1277	817	38.9
	11	809	815	38.5
	11	1277	817	38.2
	13	809	815	37.9
	3	809	815	37.7
	13	827	831	36
	13	748	460	34.9
	3	1277	811	31.8
	13	827	817	31.5
	13	1305	809	31.3
	13	399	388	30.7
	3	1277	809	29
	11	817	799	28.3
	11	1305	809	24.7
	3	1305	809	24.3
	11	399	382	21.1
MP (No RNA) (WT)	17	940	1277	35.5
	17	1285	940	35.2
	18	809	817	33.7
	18	1250	940	33.6
	19	809	817	33
	17	512	831	32.9
	18	512	831	32.9
	17	809	817	32.3
	17	1285	795	31.9
	17	827	817	30.8
	19	817	799	30.7
	17	827	831	30.2
	17	827	831	29.6
	19	1285	382	29.1
	17	817	799	28.3
	18	827	831	28
	19	809	815	27.8
	18	809	799	27.7
	18	399	388	27.7
	19	827	817	27.6
	18	827	817	26.4
	18	399	388	26.1
	19	827	831	24.4
	19	1285	799	23.4
MP/RNA Partially Docked (E1045Q/E1222Q)	16	1277	1305	36.8
	15	1250	460	35
	16	931	926	33.9
	15	1277	382	33.8
	15	1277	1305	33.8
	16	1250	382	32.3
	10	827	831	31.6
	15	931	1341	30.4
	15	1277	1341	30.4
	15	399	926	30.1
	15	827	388	29.9
	16	827	831	29.5
	15	1313	831	28.3
	16	827	831	28.1
	15	1277	1207	25.9
	10	1277	1341	24.9
	16	1277	1341	24.7
	16	1250	1341	24.6
	10	390	390	17.6

Table 3. Top inter-peptide intra-Drosha crosslinks from XLMS. Top intra-Drosha crosslinks from XLMS on fully docked MP-RNA (WT), MP without RNA (WT), and partially docked MP/RNA (E1045Q/ E1222Q) complexes are shown. Inter-peptide intra-Drosha crosslinks with FDR scores > 0.05 were excluded, and the hits are sorted according to Id-Score.

<u>**Chapter 4. Conclusions and Future Directions**</u>

4.1 Introduction

Microprocessor processes hundreds of pri-miRNA substrates, each of which possesses a unique sequence and structure. Identification of these substrates requires a highly adaptable system, but one that also follows strictly defined parameters for recognition. Although many questions remain about Microprocessor regulation, this work provides a structural foundation for understanding how flexibility and stringency can coexist in this system. Overall conclusions and future directions are discussed below.

4.2. Structure determination guided by in-depth in vitro characterization

The major achievement of this work was the determination of a precatalytic Drosha/DGCR8/pri-miRNA complex, which led to substantial insight into pri-miRNA recognition. The design and optimization of a complex suitable for cryo-EM required drawing from the findings described in Chapter 2. Specifically, a critical insight was that two independent protein modules drive pri-miRNA recognition on either side of the RNA stem, and that each module must be examined separately from one another in order to properly assess its contribution. The contribution of the apical junction to recognition by Microprocessor can be assessed through direct measurement of binding (Figure 34), while the basal junction must be measured by comparing processing efficiency/accuracy by heme-bound versus heme-free Drosha/DGCR8 (Figure 35). Thus, our structure determination efforts relied upon thorough in vitro characterization of Drosha/DGCR8.

In addition, our model for pri-miRNA recognition in Chapter 3 (Figure 76) borrowed heavily from the findings from Chapter 2, due to the relatively low resolution of the HBR. The discovery that the HBR recognizes terminal loop structures, and can serve as an anchor for repositioning Drosha, complemented the structural information and allowed us to develop a general model for the recognition events occurring on both sides of the stem.

4.3. Reexamining the role of heme in a structural context

One major question is whether heme serves a regulatory role in pri-miRNA processing. As discussed in Chapter 2, heme binding is critical for the processing of most pri-miRNAs, and in Chapter 3 it was shown that the heme-bound HBR defines the apical boundary of the dsRBD ruler. The preserved footprint between partially-bound and fully-bound Microprocessor (Figure 69) suggests that the initial contact between Microprocessor and pri-miRNA involves the HBR. Therefore, a loss of heme binding likely leads to a failure to engage pri-miRNAs. However, pri-miR-16-2, which is heme-independent, contains an ideal GHG motif, which we observed to be serving an anchoring function for the active sites of the RIIID domains. This feature may serve a compensatory role in the absence of heme, allowing processing to occur efficiently regardless of heme availability. For pri-miR-16-2, the dominance hierarchy places the basal GHG motif at a higher priority than the HBR/loop interaction. As suggested in Chapter 2, the secondary role of the HBR in some pri-miRNAs presents an opportunity for differential regulation of subclasses of pri-miRNAs depending on heme levels, by

altering the expression of GHG-lacking substrates while leaving the GHG-containing substrates relatively unaffected. Thus, our structure may offer a structural explanation for one type of heme-independence.

One of the most intriguing questions remaining about pri-miRNA processing is whether heme serves a regulatory role in microRNA biogenesis. It has been proposed that heme may modulate DGCR8 through gas binding (115), sensing oxidative stress (92), or sensing changes in heme levels (85). Results from a previous study implied that DGCR8 may not be fully heme-saturated in vivo, under certain conditions (96), suggesting that DGCR8 may be capable of responding to increased intracellular heme levels by increasing the production of microRNAs. Despite these speculations, there remains no conclusive evidence for a regulatory role for heme in pri-miRNA processing, and more work is needed to fully understand how heme might modulate microRNA biogenesis.

4.4. New modes of RNA recognition

In recent years it has become clear that noncoding RNAs have diverse and crucial roles in human biology. Thus, there is great interest in understanding how noncoding RNAs are recognized, processed and trafficked, as well as how they perform their effector functions. Concurrently, the number of identified RNA-binding proteins has exploded through the use of newly developed molecular tools (152-154); previous studies failed to identify these factors, because many of these proteins do not contain classical RNA-binding domains.

The work described here provides examples of previously unidentified RNA-

binding modes. The strategies that Microprocessor has developed for recognizing and measuring RNA structure, and how sequence preference can be layered on top of these structure-specific interactions, may represent a general paradigm for how other noncoding RNAs are recognized. Microprocessor constructs a stem-detection system by coupling two junction-sensing modules connected by a rigid bridge formed by dsRBDs. In the case of Drosha/DGCR8, these RNA-recognizing modules are coupled to RNase activity, whereas other RNA-binding proteins may utilize complex structure recognition modules coupled to RNA-modification, remodeling, or transport.

4.5. A need for additional structures

Although our structures have provided a wealth of knowledge about the general rules of pri-miRNA recognition, many questions remain about Microprocessor function. Previously proposed mechanisms for pri-miRNA recognition have been based upon mutagenesis (insertions, deletions, or substitutions) of individual pri-miRNAs; in some cases, these proposals conflict with one another (94, 100, 106), perhaps due to the fact that the recognition events vary between substrates.

As seen with our ΔBelt screen (Figure 56A), it is clear that individual pri-miRNAs are subject to different forms of regulation by the Belt. It is generally accepted that the stem length is crucial to proper recognition, but this requirement is meaningless if the boundaries are not clearly defined. In pri-miR-16-2, the basal junction is recognized in part by the Belt and its interactions with other regions of Drosha, and unsurprisingly, deletion of the Belt leads to a severe processing defect. In the case of pri-miR-125a, there

is no defect observed without the contribution of the Belt. For both pri-miRNAs, stem length is presumably measured by the dsRBD ruler, but the basal junctions are apparently defined through different mechanisms. Interestingly, Belt closure still serves as a major source of binding energy for Drosha (Figure 74), supporting a model in which a compensatory mechanism for basal junction recognition exists in pri-miR-125a. One possibility is that the GHG motif, which is present in its "ideal" form with the sequence GCG in pri-miR125a, is serving as an anchoring site for positioning the RIIIDs. This is further complicated by the fact that pri-miR-16-2, as discussed in Chapter 3, also contains an ideal GHG motif, yet is Belt-sensitive. Thus, there may be additional recognition events that occur with some substrates, but not others. This hypothesis is further supported by the MP^{C352S} screen (Figures 12 and 35), which showed that the overall basal junction "strength" — a function of belt-sensitivity, GHG region composition, and likely other factors — varies widely between pri-miRNAs.

The best approach to understanding these differences is to determine MP/RNA structures containing other substrates. A structure of Microprocessor with a Belt-independent substrate could provide clues about how other modules mitigate the loss of a functional 4-way junction. Structures with ideal basal UG motifs and apical UGUG motifs could reveal how these sequences can serve to reposition Microprocessor. These structures would form the basis for new hypotheses on pri-miRNA recognition, and would spur new biochemical and biophysical studies on the general mechanisms of Microprocessor function, and the dominance patterns of the numerous RNA/protein interactions involved.

4.6. Structure-guided therapeutic applications for pri-miRNAs

Previous studies have explored the design of therapeutics that act through the RNA interference pathway, and it has been shown that utilizing the microRNA biogenesis pathway with an artificial pri-miRNA can enhance efficacy and regulation of therapies (103, 156, 157). Therefore, past efforts have focused on the design of an optimized artificial pri-miRNA chassis to deliver siRNAs. Because the structural and sequence-based features that modulate Drosha activity reside outside of the mature miRNA region, a more complete understanding of pri-miRNA processing could enable the design of a new generation of improved RNAi therapeutics. Recognition modules could be combined in specific configurations to achieve the precisely desired amount of siRNA. Introducing features in the loop that are targeted by regulatory proteins such as Lin28 (Figure 32) would enable these artificial substrates to be temporally or spatially regulated.

4.7. Balancing stringency and promiscuity in biomolecule recognition

In many biological systems, biomolecules must be capable of exhibiting flexibility in selecting substrates or partners. MHC Class I/II proteins exhibit promiscuous binding specificity to a multitude of antigen peptides, through noncovalent interactions, and recognition patterns are dependent on allelic variation in the MHC genes (155). For MHC I molecules, allowable antigen length is restricted to a narrow range (~8-10 residues) due to the pockets that grip the ends of the antigen within the peptide-binding groove. MHC II proteins do not have this size limitation and can bind much longer antigens. In both

cases, there is preference for certain side chains in particular positions (hydrophobic, basic, etc.) due to the specific interactions with the peptide-binding groove. An enormous number of potential antigens meet these general requirements, and thus MHC proteins can present a strong diversity of peptides for antigen recognition, which is critical for proper immune function.

Nuclear trafficking of proteins also uses a set of structural determinants to bind targets with a certain degree of promiscuity. CRM1/XPO1 exports a multitude of proteins from the nucleus, by recognizing their nuclear export sequence (NES). Similar to MHC protein targets, these NES sequences are restricted in size, and require the presence of hydrophobic side chains in certain positions. Within these requirements exist a diverse array of NES sequences that bind CRM1/XPO1 and undergo nuclear export (158).

There are many other examples of biological systems that utilize structural determinants while maintaining a certain degree of promiscuity. This balance is essential to proper biological function, because many systems rely upon complex interactions with other systems involving a multitude of substrates. These substrates have upstream and downstream interactions with more systems, each of which may involve their own unique structural determinants; thus the processing or trafficking of individual biomolecules may share common intermediate steps involving a single enzyme or partner. This is certainly the case for Microprocessor, which processes hundreds of different pri-miRNAs. The different layers of regulation in each pri-miRNA lead to a range processing efficiencies, thereby regulating biological systems to different extents.

Acknowledgments

First and foremost I must thank my family, especially my mom Leslie, my dad Brian, and my sisters Lauren and Haley; my brother-in-law Michael, my stepmother Tara, stepfather Todd, my grandmothers Linda and Brenda, my grandfathers Leon and Michael, and my bonus siblings Hannah, Ellyn and Cade; Scott and Sarah MacFarlane, Casey Partin, Natalie Snow and her family, and Valerie Tabor. These people gave me enormous support during difficult times at UT Southwestern, and I will always be grateful for their love and encouragement.

I am grateful to all of my lab members for our shared experiences. The core group during my time in the Nam lab consisted of Tri Ngo, Katelyn Doxtader, Emily Herrell, and Ping Wang. For many years, we provided companionship and encouragement to each other every day, and I spent more time with them than anyone else over these years. They are like a second family to me and I am extremely fortunate to have worked and grown alongside them. Jessica Parker, who arrived more recently, has brought an enormous amount of energy and positivity to our group. Initially I worked with Byung-Cheon Jeong, who was very helpful in teaching me to navigate the lab. I also thank Weronika Chojnacka, Oliwia Koczy and Prem Prakash for their kindness and encouragement. I am very fortunate to have worked with the new generation of the Nam Lab, consisting of Chi Zhang, Luiza Tunes, and Rishi Raj, all of who are excellent scientists and will undoubtedly achieve great things.

I would also like to thank my advisor, Yunsun Nam, for all of her advice and training, and for teaching me the importance of personal growth. Her relentless

enthusiasm for science is a powerful inspiration to those who work with her, and it is clear to me that she does science for the right reason: not for personal gain, but for improving the world. She has always been supportive and positive, sustaining my motivation and enthusiasm even during the most discouraging times. Throughout my training, she fostered stability, resiliency, mutual trust, compassion, and openness in her lab. Most importantly, I was greatly inspired by her willingness to help fellow scientists simply for the sake of science.

I thank my thesis committee—Jose Rizo-Rey, Kim Orth and Michael Rosen, for their support and advice over the years. Their encouragement helped me to endure through some very challenging times, and to persist all the way through to the end of my project. I also need to thank Jose for inviting me to rotate in his lab, where his enthusiasm and kindness led me to develop many of the core interests that I took with me to the Nam lab.

I am grateful for my collaborators, Wah Chiu and Kaiming Zhang, who persisted with me through years of obstacles in order to achieve our goals. This work would not have been possible without their contributions.

I also thank Amy Haughey and Nancy Street for their support through the DBS Graduate Program; faculty members Xiochen Bai, Ryan Hibbs, Luke Rice, Xiochun Li, Gary Hon, and Lee Kraus for their support and advice; and Green Center staff Connie Grafer, Dawn Dale, Kimberly Buggs, and Elisabeth Henderson for all of their assistance over the years.

I have made a lot of amazing friends while at UT Southwestern. In particular, I

Bibliography

1.   Shabalina SA, Koonin EV. Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol. 2008;23(10):578-87. doi: 10.1016/j.tree.2008.06.005. PubMed PMID: 18715673; PMCID: PMC2695246.

2.   Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19(1):92-105. doi: papers3://publication/doi/10.1101/gr.082701.108.

3.   Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005;433(7027):769-73. doi: 10.1038/nature03315. PubMed PMID: 15685193.

4.   Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF. MicroRNAs regulate brain morphogenesis in zebrafish. Science. 2005;308(5723):833-8. doi: 10.1126/science.1109020. PubMed PMID: 15774722.

5.   Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME. A brain-specific microRNA regulates dendritic spine development. Nature. 2006;439(7074):283-9. doi: 10.1038/nature04367. PubMed PMID: 16421561.

6.   Im HI, Kenny PJ. MicroRNAs in neuronal function and dysfunction. Trends Neurosci. 2012;35(5):325-34. doi: 10.1016/j.tins.2012.01.004. PubMed PMID: 22436491; PMCID: PMC3565236.

7.   Chen CZ, Lodish HF. MicroRNAs as regulators of mammalian hematopoiesis. Semin Immunol. 2005;17(2):155-65. doi: 10.1016/j.smim.2005.01.001. PubMed PMID: 15737576.

8.   Lee CT, Risom T, Strauss WM. MicroRNAs in mammalian development. Birth Defects Res C Embryo Today. 2006;78(2):129-39. doi: 10.1002/bdrc.20072. PubMed PMID: 16847889.

9.   Cobb BS, Hertweck A, Smith J, O'Connor E, Graf D, Cook T, Smale ST, Sakaguchi S, Livesey FJ, Fisher AG, Merkenschlager M. A role for Dicer in immune regulation. J Exp Med. 2006;203(11):2519-27. doi: 10.1084/jem.20061692. PubMed PMID: 17060477; PMCID: PMC2118134.

10.   Roberts AP, Lewis AP, Jopling CL. The role of microRNAs in viral infection. Prog Mol Biol Transl Sci. 2011;102:101-39. doi: 10.1016/B978-0-12-415795-8.00002-7. PubMed PMID: 21846570.

11.   Vecchione A, Croce CM. Apoptomirs: small molecules have gained the license to kill. Endocr Relat Cancer. 2010;17(1):F37-50. doi: 10.1677/ERC-09-0163. PubMed PMID: 19815577.

12.   Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol. 2006;57:19-53. doi: 10.1146/annurev.arplant.57.032905.105218. PubMed PMID: 16669754.

13.   Molnar A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC. miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii.

Nature. 2007;447(7148):1126-9. doi: 10.1038/nature05903. PubMed PMID: 17538623.

14. Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, Lewis ZA, Freitag M, Selker EU, Mello CC, Liu Y. Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell. 2010;38(6):803-14. doi: 10.1016/j.molcel.2010.04.005. PubMed PMID: 20417140; PMCID: PMC2902691.

15. Jin P, Zarnescu DC, Ceman S, Nakamoto M, Mowrey J, Jongens TA, Nelson DL, Moses K, Warren ST. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. Nat Neurosci. 2004;7(2):113-7. doi: 10.1038/nn1174. PubMed PMID: 14703574.

16. Klein ME, Lioy DT, Ma L, Impey S, Mandel G, Goodman RH. Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. Nat Neurosci. 2007;10(12):1513-4. doi: 10.1038/nn2010. PubMed PMID: 17994015.

17. Abu-Elneel K, Liu T, Gazzaniga FS, Nishimura Y, Wall DP, Geschwind DH, Lao K, Kosik KS. Heterogeneous dysregulation of microRNAs across the autism spectrum. Neurogenetics. 2008;9(3):153-61. doi: 10.1007/s10048-008-0133-5. PubMed PMID: 18563458.

18. Gururajan A, Naughton ME, Scott KA, O'Connor RM, Moloney G, Clarke G, Dowling J, Walsh A, Ismail F, Shorten G, Scott L, McLoughlin DM, Cryan JF, Dinan TG. MicroRNAs as biomarkers for major depression: a role for let-7b and let-7c. Transl Psychiatry. 2016;6(8):e862. doi: 10.1038/tp.2016.131. PubMed PMID: 27483380; PMCID: PMC5022079.

19. Beveridge NJ, Cairns MJ. MicroRNA dysregulation in schizophrenia. Neurobiol Dis. 2012;46(2):263-71. doi: 10.1016/j.nbd.2011.12.029. PubMed PMID: 22207190.

20. Wu Q, Law PY, Wei LN, Loh HH. Post-transcriptional regulation of mouse mu opioid receptor (MOR1) via its 3' untranslated region: a role for microRNA23b. FASEB J. 2008;22(12):4085-95. doi: 10.1096/fj.08-108175. PubMed PMID: 18716031; PMCID: PMC2614608.

21. Basavaraju M, de Lencastre A. Alzheimer's disease: presence and role of microRNAs. Biomol Concepts. 2016;7(4):241-52. doi: 10.1515/bmc-2016-0014. PubMed PMID: 27505094; PMCID: PMC5035151.

22. Wojciechowska A, Braniewska A, Kozar-Kaminska K. MicroRNA in cardiovascular biology and disease. Adv Clin Exp Med. 2017;26(5):865-74. doi: 10.17219/acem/62915. PubMed PMID: 29068585.

23. Bostjancic E, Glavac D. Importance of microRNAs in skin morphogenesis and diseases. Acta Dermatovenerol Alp Pannonica Adriat. 2008;17(3):95-102. PubMed PMID: 18853072.

24. Lujambio A, Lowe SW. The microcosmos of cancer. Nature. 2012;482(7385):347-55. doi: papers3://publication/doi/10.1038/nature10888.

25. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. Nat Rev Cancer. 2015;15(6):321-33. doi: papers3://publication/doi/10.1038/nrc3932.

26.   Yu X, Li Z, Chan MT, Wu WK. The roles of microRNAs in Wilms' tumors. Tumour Biol. 2016;37(2):1445-50. doi: 10.1007/s13277-015-4514-8. PubMed PMID: 26634744.

27.   Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(24):15524-9. doi: 10.1073/pnas.242606799. PubMed PMID: 12434020; PMCID: PMC137750.

28.   Olson P, Lu J, Zhang H, Shai A, Chun MG, Wang Y, Libutti SK, Nakakura EK, Golub TR, Hanahan D. MicroRNA dynamics in the stages of tumorigenesis correlate with hallmark capabilities of cancer. Genes Dev. 2009;23(18):2152-65. doi: 10.1101/gad.1820109. PubMed PMID: 19759263; PMCID: PMC2751988.

29.   Garzon R, Marcucci G, Croce CM. Targeting microRNAs in cancer: rationale, strategies and challenges. Nat Rev Drug Discov. 2010;9(10):775-89. doi: papers3://publication/doi/10.1038/nrd3179.

30.   Van Roosbroeck K, Calin GA. Cancer Hallmarks and MicroRNAs: The Therapeutic Connection. Adv Cancer Res. 2017;135:119-49. doi: 10.1016/bs.acr.2017.06.002. PubMed PMID: 28882220.

31.   Kotani A, Ha D, Schotte D, den Boer ML, Armstrong SA, Lodish HF. A novel mutation in the miR-128b gene reduces miRNA processing and leads to glucocorticoid resistance of MLL-AF4 acute lymphocytic leukemia cells. Cell Cycle. 2010;9(6):1037-42. doi: 10.4161/cc.9.6.11011. PubMed PMID: 20237425; PMCID: PMC3096720.

32.   O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839-43. doi: 10.1038/nature03677. PubMed PMID: 15944709.

33.   Dews M, Homayouni A, Yu D, Murphy D, Sevignani C, Wentzel E, Furth EE, Lee WM, Enders GH, Mendell JT, Thomas-Tikhonenko A. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. Nat Genet. 2006;38(9):1060-5. doi: 10.1038/ng1855. PubMed PMID: 16878133; PMCID: PMC2669546.

34.   Blandino G, Fazi F, Donzelli S, Kedmi M, Sas-Chen A, Muti P, Strano S, Yarden Y. Tumor suppressor microRNAs: a novel non-coding alliance against cancer. FEBS Lett. 2014;588(16):2639-52. doi: 10.1016/j.febslet.2014.03.033. PubMed PMID: 24681102.

35.   Asghari F, Haghnavaz N, Baradaran B, Hemmatzadeh M, Kazemi T. Tumor suppressor microRNAs: Targeted molecules and signaling pathways in breast cancer. Biomed Pharmacother. 2016;81:305-17. doi: 10.1016/j.biopha.2016.04.011. PubMed PMID: 27261608.

36.   Hammond SM. MicroRNAs as tumor suppressors. Nat Genet. 2007;39(5):582-3. doi: 10.1038/ng0507-582. PubMed PMID: 17460676.

37.   Lujambio A, Calin GA, Villanueva A, Ropero S, Sanchez-Cespedes M, Blanco D, Montuenga LM, Rossi S, Nicoloso MS, Faller WJ, Gallagher WM, Eccles SA, Croce CM, Esteller M. A microRNA DNA methylation signature for human cancer metastasis.

Proceedings of the National Academy of Sciences of the United States of America. 2008;105(36):13556-61. doi: 10.1073/pnas.0803055105. PubMed PMID: 18768788; PMCID: PMC2528872.

38.   Guil S, Esteller M. DNA methylomes, histone codes and miRNAs: tying it all together. Int J Biochem Cell Biol. 2009;41(1):87-95. doi: 10.1016/j.biocel.2008.09.005. PubMed PMID: 18834952.

39.   Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843-54. doi: papers3://publication/uuid/E0CD8422-C770-4734-A3BD-738F91BF73DA.

40.   Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. 1993;75(5):855-62. PubMed PMID: 8252622.

41.   Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science. 2001;294(5543):853-8. doi: 10.1126/science.1064921. PubMed PMID: 11679670.

42.   Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science. 2001;294(5543):858-62. doi: 10.1126/science.1065062. PubMed PMID: 11679671.

43.   Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. Science. 2001;294(5543):862-4. doi: 10.1126/science.1065329. PubMed PMID: 11679672.

44.   Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, Peterson KJ. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet. 2015;49:213-42. doi: 10.1146/annurev-genet-120213-092023. PubMed PMID: 26473382; PMCID: PMC4743252.

45.   Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. Nature. 2004;432(7014):231-5. doi: 10.1038/nature03049. PubMed PMID: 15531879.

46.   Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature. 2011;469(7328):97-101. doi: 10.1038/nature09616. PubMed PMID: 21085120; PMCID: PMC3057491.

47.   Bartel DP. Metazoan MicroRNAs. Cell. 2018;173(1):20-51. doi: 10.1016/j.cell.2018.03.006. PubMed PMID: 29570994; PMCID: PMC6091663.

48.   Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010;24(10):992-1009. doi: papers3://publication/doi/10.1101/gad.1884710.

49.   Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24. doi: papers3://publication/doi/10.1038/nrm3838.

50.   Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. 2002;21(17):4663-70. PubMed PMID: 12198168; PMCID: PMC126204.

51.   Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004;10(12):1957-66. doi: 10.1261/rna.7135204. PubMed PMID: 15525708; PMCID: PMC1370684.

52.   Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051-60. doi: papers3://publication/doi/10.1038/sj.emboj.7600385.

53.   Blaszczyk J, Tropea JE, Bubunenko M, Routzahn KM, Waugh DS, Court DL, Ji X. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure (London, England : 1993). 2001;9(12):1225-36. doi: 10.1016/s0969-2126(01)00685-2. PubMed PMID: 11738048.

54.   Court DL, Gan J, Liang Y-H, Shaw GX, Tropea JE, Costantino N, Waugh DS, Ji X. RNase III: Genetics and Function; Structure and Mechanism. Annu Rev Genet. 2013;47:405-31. doi: papers3://publication/doi/10.1146/annurev-genet-110711-155618.

55.   Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. Single processing center models for human Dicer and bacterial RNase III. Cell. 2004;118(1):57-68. doi: 10.1016/j.cell.2004.06.017. PubMed PMID: 15242644.

56.   Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003;425(6956):415-9. doi: papers3://publication/doi/10.1038/nature01957.

57.   Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003;17(24):3011-6. doi: papers3://publication/doi/10.1101/gad.1158803.

58.   Bohnsack MT, Czaplinski K, Gorlich D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA. 2004;10(2):185-91. doi: papers3://publication/uuid/649CD287-F0BC-4042-BCDF-21EB0388E9B0.

59.   Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001;409(6818):363-6. doi: papers3://publication/doi/10.1038/35053110.

60.   Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell. 2001;106(1):23-34. doi: papers3://publication/uuid/6D47C810-FC3F-44C8-89D1-A9544998A15F.

61.   Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001;293(5531):834-8. doi: 10.1126/science.1062961. PubMed PMID: 11452083.

62.   Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol. 2008;9(1):22-32. doi: papers3://publication/doi/10.1038/nrm2321.

63.   Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell. 2010;39(2):292-9. doi: 10.1016/j.molcel.2010.05.015. PubMed PMID: 20605501.

64.   Kawamata T, Tomari Y. Making RISC. Trends Biochem Sci. 2010;35(7):368-76. doi: papers3://publication/doi/10.1016/j.tibs.2010.03.009.

65.   Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003;115(2):209-16. doi: 10.1016/s0092-8674(03)00801-8. PubMed PMID: 14567918.

66.   Frank F, Sonenberg N, Nagar B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature. 2010;465(7299):818-22. doi: papers3://publication/doi/10.1038/nature09039.

67.   Suzuki HI, Katsura A, Yasuda T, Ueno T, Mano H, Sugimoto K, Miyazono K. Small-RNA asymmetry is directly driven by mammalian Argonautes. Nat Struct Mol Biol. 2015;22(7):512-21. doi: 10.1038/nsmb.3050. PubMed PMID: 26098316.

68.   MicroRNAs: target recognition and regulatory functions., (2009).

69.   Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol. 2013;14(8):475-88. doi: papers3://publication/doi/10.1038/nrm3611.

70.   Wu H, Ye C, Ramirez D, Manjunath N. Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA. PloS one. 2009;4(10):e7566. doi: papers3://publication/doi/10.1371/journal.pone.0007566.

71.   Prediction of mammalian microRNA targets., (2003).

72.   Han J, Lee Y, Yeom K-H, Kim Y-K, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 2004;18(24):3016-27. doi: papers3://publication/doi/10.1101/gad.1262504.

73.   Tang X, Zhang Y, Tucker L, Ramratnam B. Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. Nucleic Acids Res. 2010;38(19):6610-9. doi: 10.1093/nar/gkq547. PubMed PMID: 20554852; PMCID: PMC2965249.

74.   Kwon SC, Nguyen TA, Choi Y-GG, Jo MH, Hohng S, Kim VN, Woo J-SS. Structure of Human DROSHA. Cell. 2016;164(1-2):81-90. doi: 10.1016/j.cell.2015.12.019.

75.   Masliah G, Barraud P, Allain FHT. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. Cell Mol Life Sci. 2013;70(11):1875-95. doi: papers3://publication/doi/10.1007/s00018-012-1119-x.

76.   Kranick JC, Chadalavada DM, Sahu D, Showalter SA. Engineering double-stranded RNA binding activity into the Drosha double-stranded RNA binding domain results in a loss of microRNA processing function. PloS one.

2017;12(8):e0182445. doi: 10.1371/journal.pone.0182445. PubMed PMID: 28792523; PMCID: PMC5549741.

77.   Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. Science. 2004;303(5654):95-8. doi: papers3://publication/doi/10.1126/science.1090599.

78.   Siolas D, Lerner C, Burchard J, Ge W, Linsley PS, Paddison PJ, Hannon GJ, Cleary MA. Synthetic shRNAs as potent RNAi triggers. Nat Biotechnol. 2005;23(2):227-31. doi: 10.1038/nbt1052. PubMed PMID: 15619616.

79.   Song J-J, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol. 2003;10(12):1026-32. doi: papers3://publication/doi/10.1038/nsb1016.

80.   Tian Y, Simanshu DK, Ma J-B, Park J-E, Heo I, Kim VN, Patel DJ. A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. Mol Cell. 2014;53(4):606-16. doi: papers3://publication/doi/10.1016/j.molcel.2014.01.003.

81.   Park J-E, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature. 2011;475(7355):201-5. doi: papers3://publication/doi/10.1038/nature10198.

82.   Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. Structural basis for double-stranded RNA processing by Dicer. Science. 2006;311(5758):195-8. doi: papers3://publication/doi/10.1126/science.1121638.

83.   Gregory RI, Yan K-p, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. The Microprocessor complex mediates the genesis of microRNAs. Nature. 2004;432(7014):235-40. doi: 10.1038/nature03120.

84.   Yeom K-H, Lee Y, Han J, Suh MR, Kim VN. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Research. 2006;34(16):4622-9. doi: papers3://publication/doi/10.1093/nar/gkl458.

85.   Faller M, Matsunaga M, Yin S, Loo JA, Guo F. Heme is involved in microRNA processing. Nat Struct Mol Biol. 2007;14(1):23-9. doi: papers3://publication/doi/10.1038/nsmb1182.

86.   Sohn SY, Bae WJ, Kim JJ, Yeom K-H, Kim VN, Cho Y. Crystal structure of human DGCR8 core. Nat Struct Mol Biol. 2007;14(9):847-53. doi: papers3://publication/doi/10.1038/nsmb1294.

87.   Roth BM, Ishimaru D, Hennig M. The core Microprocessor component DiGeorge syndrome critical region 8 (DGCR8) is a non-specific RNA-binding protein. J Biol Chem. 2013. doi: papers3://publication/doi/10.1074/jbc.M112.446880.

88.   Wostenberg C, Quarles KA, Showalter SA. Dynamic origins of differential RNA binding function in two dsRBDs from the miRNA "microprocessor" complex. Biochemistry. 2010;49(50):10728-36. doi: papers3://publication/doi/10.1021/bi1015716.

89.   Barr I, Smith AT, Senturia R, Chen Y, Scheidemantle BD, Burstyn JN, Guo F. DiGeorge critical region 8 (DGCR8) is a double-cysteine-ligated heme protein. J Biol Chem. 2011. doi: papers3://publication/doi/10.1074/jbc.M110.180844.

90.   Barr I, Guo F. Pyridine Hemochromagen Assay for Determining the Concentration of Heme in Purified Protein Solutions. Bio Protoc. 2015;5(18). doi: 10.21769/bioprotoc.1594. PubMed PMID: 27390766; PMCID: PMC4932910.

91.   Analysis of Heme Iron Coordination in DGCR8: The Heme-Binding Component of the Microprocessor Complex, (2016).

92.   Barr I, Smith AT, Chen Y, Senturia R, Burstyn JN, Guo F. Ferric, not ferrous, heme activates RNA-binding protein DGCR8 for primary microRNA processing. P Natl Acad Sci Usa. 2012;109(6):1919-24. doi: papers3://publication/doi/10.1073/pnas.1114514109.

93.   Quarles KA, Chadalavada D, Showalter SA. Deformability in the cleavage site of primary MicroRNA is not sensed by the double-stranded RNA binding domains in the microprocessor component DGCR8. Proteins. 2015. doi: papers3://publication/doi/10.1002/prot.24810.

94.   Quick-Cleveland J, Jacob JP, Weitz SH, Shoffner G, Senturia R, Guo F. The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin. Cell reports. 2014. doi: papers3://publication/doi/10.1016/j.celrep.2014.05.013.

95.   Heme Protein Assemblies, (2004).

96.   Barr I, Weitz SH, Atkin T, Hsu P, Karayiorgou M, Gogos JA, Weiss S, Guo F. Cobalt(III) Protoporphyrin Activates the DGCR8 Protein and Can Compensate microRNA Processing Deficiency. Chem Biol. 2015;22(6):793-802. doi: 10.1016/j.chembiol.2015.05.015.

97.   Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? Front Genet. 2015;6:2. doi: 10.3389/fgene.2015.00002. PubMed PMID: 25674102; PMCID: PMC4306305.

98.   Zeng Y. Sequence requirements for micro RNA processing and function in human cells. RNA. 2003;9(1):112-23. doi: papers3://publication/doi/10.1261/rna.2780503.

99.   Zeng Y, Yi R, Cullen BR. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. EMBO J. 2005;24(1):138-48. doi: papers3://publication/doi/10.1038/sj.emboj.7600491.

100.   Han J, Lee Y, Yeom K-H, Nam J-W, Heo I, Rhee J-K, Sohn SY, Cho Y, Zhang B-T, Kim VN. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006;125(5):887-901. doi: papers3://publication/doi/10.1016/j.cell.2006.03.043.

101.   Zeng Y, Cullen BR. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. J Biol Chem. 2005;280(30):27595-603. doi: 10.1074/jbc.M504714200.

102.   Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell. 2013;152(4):844-58. doi: 10.1016/j.cell.2013.01.031.

103. Fang W, Bartel DP. The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. Mol Cell. 2015;60(1):131-45. doi: papers3://publication/doi/10.1016/j.molcel.2015.08.015.

104. Ma H, Wu Y, Choi J-G, Wu H. Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. P Natl Acad Sci Usa. 2013;110(51):20687-92. doi: papers3://publication/doi/10.1073/pnas.1311639110.

105. Kwon SC, Baek SC, Choi YG, Yang J, Lee YS, Woo JS, Kim VN. Molecular Basis for the Single-Nucleotide Precision of Primary microRNA Processing. Mol Cell. 2019;73(3):505-18 e5. doi: 10.1016/j.molcel.2018.11.005. PubMed PMID: 30554947.

106. Nguyen TA, Jo MH, Choi Y-G, Park J, Kwon SC, Hohng S, Kim VN, Woo J-S. Functional Anatomy of the Human Microprocessor. Cell. 2015;161(6):1374-87. doi: papers3://publication/doi/10.1016/j.cell.2015.05.010.

107. Mori M, Triboulet R, Mohseni M, Schlegelmilch K, Shrestha K, Camargo FD, Gregory RI. Hippo signaling regulates microprocessor and links cell-density-dependent miRNA biogenesis to cancer. Cell. 2014;156(5):893-906. doi: 10.1016/j.cell.2013.12.043.

108. Ngo TD, Partin, Alexander C., Nam, Y. RNA Specificity and Autoregulation of DDX17, a Modulator of MicroRNA Biogenesis. Cell Rep. 2019;29(12):4024-35. Epub 12/17/2019.

109. Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. Genes Dev. 2006;20(16):2202-7. doi: papers3://publication/doi/10.1101/gad.1444406.

110. Tang GQ, Maxwell ES. Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. Genome Res. 2008;18(1):104-12. doi: 10.1101/gr.6539108. PubMed PMID: 18032731; PMCID: PMC2134782.

111. Conrad T, Marsico A, Gehre M, Orom UA. Microprocessor activity controls differential miRNA biogenesis In Vivo. Cell reports. 2014;9(2):542-54. doi: 10.1016/j.celrep.2014.09.007.

112. Feng Y, Zhang X, Song Q, Li T, Zeng Y. Drosha processing controls the specificity and efficiency of global microRNA expression. Biochim Biophys Acta. 2011. doi: papers3://publication/doi/10.1016/j.bbagrm.2011.05.015.

113. Roden CA, Gaillard J, Kanoria S, Rennie W, Barish S, Cheng J, Pan W, Liu J, Cotsapas C, Ding Y, Lu J. Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. Genome Res. 2017. doi: 10.1101/gr.208900.116. PubMed PMID: 28087842.

114. Weitz SH, Gong M, Barr I, Weiss S, Guo F. Processing of microRNA primary transcripts requires heme in mammalian cells. P Natl Acad Sci Usa. 2014. doi: papers3://publication/doi/10.1073/pnas.1309915111.

115. Hines JP, Smith AT, Jacob JP, Lukat-Rodgers GS, Barr I, Rodgers KR, Guo F, Burstyn JN. CO and NO bind to Fe(II) DiGeorge critical region 8 heme but do not restore primary microRNA processing activity. Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry. 2016;21(8):1021-35. doi: 10.1007/s00775-016-1398-z.

116. Poulos TL. Heme enzyme structure and function. Chem Rev. 2014;114(7):3919-62. doi: 10.1021/cr400415k.

117. Raghuram S, Stayrook KR, Huang P, Rogers PM, Nosie AK, McClure DB, Burris LL, Khorasanizadeh S, Burris TP, Rastinejad F. Identification of heme as the ligand for the orphan nuclear receptors REV-ERBalpha and REV-ERBbeta. Nat Struct Mol Biol. 2007;14(12):1207-13. doi: 10.1038/nsmb1344.

118. Ogawa K, Sun J, Taketani S, Nakajima O, Nishitani C, Sassa S, Hayashi N, Yamamoto M, Shibahara S, Fujita H, Igarashi K. Heme mediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. EMBO J. 2001;20(11):2835-43. doi: 10.1093/emboj/20.11.2835.

119. Tang XD, Xu R, Reynolds MF, Garcia ML, Heinemann SH, Hoshi T. Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels. Nature. 2003;425(6957):531-5. doi: 10.1038/nature02003.

120. Derbyshire ER, Marletta MA. Structure and regulation of soluble guanylate cyclase. Annu Rev Biochem. 2012;81:533-59. doi: 10.1146/annurev-biochem-050410-100030.

121. Faller M, Toso D, Matsunaga M, Atanasov I, Senturia R, Chen Y, Zhou ZH, Guo F. DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. RNA. 2010;16(8):1570-83. doi: papers3://publication/doi/10.1261/rna.2111310.

122. Walker SC, Avis JM, Conn GL. General plasmids for producing RNA in vitro transcripts with homogeneous ends. Nucleic Acids Research. 2003;31(15):e82. doi: papers3://publication/uuid/765F962D-5F68-4B3B-89AA-694E68D28AD5.

123. Brautigam CA. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. Meth Enzymol. 2015;562:109-33. doi: 10.1016/bs.mie.2015.05.001.

124. Karabiber F, McGinnis JL, Favorov OV, Weeks KM. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. RNA (New York, NY). 2013;19(1):63-73. doi: 10.1261/rna.036327.112.

125. Markham NR, Zuker M. DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Research. 2005;33(Web Server issue):W577-81. doi: papers3://publication/doi/10.1093/nar/gki591.

126. Maroney PA, Chamnongpol S, Souret F, Nilsen TW. Direct detection of small RNAs using splinted ligation. Nat Protoc. 2008;3(2):279-87. doi: 10.1038/nprot.2007.530. PubMed PMID: 18274530.

127. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. Science (New York, NY). 2008;320(5872):97-100. doi: 10.1126/science.1154040.

128.	Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell. 2008;32(2):276-84. doi: papers3://publication/doi/10.1016/j.molcel.2008.09.014.

129.	Shyh-Chang N, Daley GQ. Lin28: primal regulator of growth and metabolism in stem cells. Cell stem cell. 2013;12(4):395-406. doi: 10.1016/j.stem.2013.03.005. PubMed PMID: 23561442; PMCID: PMC3652335.

130.	Piskounova E, Polytarchou C, Thornton JE, Lapierre RJ, Pothoulakis C, Hagan JP, Iliopoulos D, Gregory RI. Lin28A and Lin28B Inhibit let-7 MicroRNA Biogenesis by Distinct Mechanisms. Cell. 2011;147(5):1066-79. doi: papers3://publication/doi/10.1016/j.cell.2011.10.039.

131.	Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P. Molecular Basis for Interaction of let-7 MicroRNAs with Lin28. Cell. 2011. doi: 10.1016/j.cell.2011.10.020.

132.	Landthaler M, Yalcin A, Tuschl T. The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol. 2004;14(23):2162-7. doi: papers3://publication/doi/10.1016/j.cub.2004.11.001.

133.	Partin AC, Ngo TD, Herrell E, Jeong B-CC, Hon G, Nam Y. Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs. Nat Commun. 2017;8(1):1737. doi: 10.1038/s41467-017-01713-y.

134.	Nguyen TA, Park J, Dang TL, Choi YG, Kim VN. Microprocessor depends on hemin to recognize the apical loop of primary microRNA. Nucleic Acids Res. 2018;46(11):5726-36. doi: 10.1093/nar/gky248. PubMed PMID: 29750274; PMCID: PMC6009577.

135.	Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods. 2017;14(4):331-2. doi: 10.1038/nmeth.4193. PubMed PMID: 28250466; PMCID: PMC5494038.

136.	Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015;192(2):216-21. doi: 10.1016/j.jsb.2015.08.008. PubMed PMID: 26278980; PMCID: PMC6760662.

137.	Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. 2007;157(1):38-46. doi: 10.1016/j.jsb.2006.05.009. PubMed PMID: 16859925.

138.	Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012;180(3):519-30. doi: 10.1016/j.jsb.2012.09.006. PubMed PMID: 23000701; PMCID: PMC3690530.

139.	Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14(3):290-6. doi: 10.1038/nmeth.4169. PubMed PMID: 28165473.

140.	Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12. doi: 10.1002/jcc.20084. PubMed PMID: 15264254.

141. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica Section D, Biological crystallography. 2010;66(Pt 4):486-501. doi: 10.1107/S0907444910007493. PubMed PMID: 20383002; PMCID: PMC2852313.

142. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406-15. doi: 10.1093/nar/gkg595. PubMed PMID: 12824337; PMCID: PMC169194.

143. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica Section D, Biological crystallography. 2010;66(Pt 1):12-21. doi: 10.1107/S0907444909042073. PubMed PMID: 20057044; PMCID: PMC2803126.

144. Pintilie G, Zhang K, Su Z, Li S, Schmid MF, Chiu W. Measurement of Atom Resolvability in CryoEM Maps with Q-scores. bioRxiv. 2019:722991. doi: 10.1101/722991.

145. Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Identification of cross-linked peptides from large sequence databases. Nat Methods. 2008;5(4):315-8. doi: 10.1038/nmeth.1192. PubMed PMID: 18327264; PMCID: PMC2719781.

146. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol. 2010;604:55-71. doi: 10.1007/978-1-60761-444-9_5. PubMed PMID: 20013364; PMCID: PMC2922680.

147. Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004;32(Database issue):D109-11. doi: 10.1093/nar/gkh023. PubMed PMID: 14681370; PMCID: PMC308757.

148. Nicholson AW. Ribonuclease III mechanisms of double-stranded RNA cleavage. Wiley Interdiscip Rev RNA. 2013. doi: papers3://publication/doi/10.1002/wrna.1195.

149. Herbert KM, Sarkar SK, Mills M, Delgado De la Herran HC, Neuman KC, Steitz JA. A heterotrimer model of the complete Microprocessor complex revealed by single-molecule subunit counting. RNA. 2016;22(2):175-83. doi: 10.1261/rna.054684.115. PubMed PMID: 26683315; PMCID: PMC4712668.

150. Leitner A, Walzthoeni T, Aebersold R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. Nat Protoc. 2014;9(1):120-37. doi: 10.1038/nprot.2013.168. PubMed PMID: 24356771.

151. Quarles KA, Sahu D, Havens MA, Forsyth ER, Wostenberg C, Hastings ML, Showalter SA. Ensemble analysis of primary microRNA structure reveals an extensive capacity to deform near the Drosha cleavage site. Biochemistry. 2013;52(5):795-807. doi: papers3://publication/doi/10.1021/bi301452a.

152. Butter F, Scheibe M, Mörl M, Mann M. Unbiased RNA-protein interaction screen by quantitative proteomics. P Natl Acad Sci Usa. 2009;106(26):10626-31. doi: papers3://publication/doi/10.1073/pnas.0812099106.

153.   Scherrer T, Mittal N, Janga SC, Gerber AP. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PloS one. 2010;5(11):e15499. doi: 10.1371/journal.pone.0015499. PubMed PMID: 21124907; PMCID: PMC2988813.

154.   Castello A, Horos R, Strein C, Fischer B, Eichelbaum K, Steinmetz LM, Krijgsveld J, Hentze MW. System-wide identification of RNA-binding proteins by interactome capture. Nat Protoc. 2013;8(3):491-500. doi: 10.1038/nprot.2013.020. PubMed PMID: 23411631.

155.   Wieczorek M, Abualrous ET, Sticht J, Alvaro-Benito M, Stolzenberg S, Noe F, Freund C. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front Immunol. 2017;8:292. doi: 10.3389/fimmu.2017.00292. PubMed PMID: 28367149; PMCID: PMC5355494.

156.   Bhaskaran V, Yao Y, Bei F, Peruzzi P. Engineering, delivery, and biological validation of artificial microRNA clusters for gene therapy applications. Nat Protoc. 2019;14(12):3538-53. doi: 10.1038/s41596-019-0241-8. PubMed PMID: 31748752.

157.   Zhang N, Zhang D, Chen SL, Gong BQ, Guo Y, Xu L, Zhang XN, Li JF. Engineering Artificial MicroRNAs for Multiplex Gene Silencing and Simplified Transgenic Screen. Plant Physiol. 2018;178(3):989-1001. doi: 10.1104/pp.18.00828. PubMed PMID: 30291175; PMCID: PMC6236610.

158. Fung HY, Fu SC, Brautigam CA, Chook YM. Structural determinants of nuclear export signal orientation in binding to exportin CRM1. Elife. 2015; 10.7554/eLife.10034; PMCID: PMC4596688.