

**Structure and Targeting of Precursor MicroRNA-31:  
From Mechanism to Application**

by

Sicong Ma

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Biophysics)  
in The University of Michigan  
2024

Doctoral Committee:

Assistant Professor Sarah C. Keane, Chair  
Dr. Aaron T. Frank, Arrakis Therapeutics  
Associate Professor Amanda L. Garner  
Assistant Professor Markos Koutmos  
Professor Janet L. Smith

Sicong Ma

[sicongm@umich.edu](mailto:sicongm@umich.edu)

ORCID iD: 0000-0002-3587-8249

© Sicong Ma 2024

## **DEDICATION**

This thesis is dedicated to my parents, my wife, and my little family.

## ACKNOWLEDGMENTS

Five years was a long journal for me. Finally, it came to an end and all the effort and time deserved. I learned a lot, experienced a lot, and harvested a lot during these five years. I want to thank all the people that helped me during these five years and all the things I have experienced these five years.

First, I want to thank my advisor Professor Sarah Keane for her instruction in science. The first time I met Sarah was seven years ago, when I was an undergraduate visiting student. Inspired by her passion for science, I decided to go back to the University of Michigan to pursue my PhD degree. Sarah was willing to help me grow from an awkward student to an independent researcher. During these years, she gave me millions of instructions on the things she was expert at and for the things she was not familiar with, she reached out everything she could to help me figure out. As an international student, writing in science is always hard. Sarah was willing to spend several hours with me to edit the papers and the proposals. Without her help, I could not be an independent researcher in the RNA field. I also want to thank Sarah for her help in my life during the pandemic period and the period when my daughter was born. Life with a newborn was always messy, and it was hard for me to get life and work balanced without any help from my parents. However, Sarah agreed that I could be in lab in a flexible time schedule to let me take care my baby. I am grateful that I met Sarah and chose her as my PI. I also want to thank all my committee members for their support. Thank you, Prof. Markos Koutmos, Prof. Janet L. Smith, Prof. Amanda L. Garner, and Dr. Aaron Frank.

I also want to thank all my labmates, including previous and current members. During these years, thanks for all your company and advice to help me out in some dilemma. I want to give my special thanks to previous co-worker Anita, Huaqun and current member Yaping. Anita was my mentor and she instructed me during my first year and second year in my study. Without her help, I may still be a beginner in research after several years of struggle. Huaqun was a previous postdoc researcher in the lab, he helped me a lot in my life the first time I came to the US and the time I was back to the lab. His several years' help built my life in the US. As a postdoc researcher in the lab, Yaping always gives helpful advice in science, and she also provided a lot of help in my life. Her help accelerated me to be an independent researcher.

I also want to thank the whole biophysics department. I would like to thank the biophysics department for making me feel like in home. I would like to give my special thanks to our coordinator, Sara Grosky. Thank you for helping me in these years, from enrolling in the program to graduating from the university.

Here, I want to give my thanks to all my friends, who experienced the fun life with me these years. Without you guys, life would be so boring. Without your support, my PhD life would be so hard to come to an end. Thank you Ziyuan, Zhijian, Guoming, Minjun, Jiameng, Zhenyu, and So on.

Finally, and most importantly, I want to thank my parents, my wife, my daughter, and my lovely cats. The love, the courage, and every support they gave to me are the things that strongly encouraged me to complete the PhD degree. Without them, I cannot complete my PhD degree.

## CONTENTS

<b>DEDICATION</b> .....	<b>ii</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>iii</b>
<b>LIST OF TABLES</b> .....	<b>viii</b>
<b>LIST OF FIGURES</b> .....	<b>x</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>xv</b>
<b>LIST OF APPENDICES</b> .....	<b>xx</b>
<b>ABSTRACT</b> .....	<b>xxi</b>
<b>CHAPTER I Introduction</b> .....	<b>1</b>
1.1 The history of microRNA .....	1
1.1.1 The discovery of microRNA .....	1
1.1.2 The role of miRNAs in diseases .....	2
1.2 miRNA biogenesis pathway and post-transcriptional regulation.....	4
1.2.1 General miRNA biogenesis pathway .....	5
1.2.2 Microprocessor processing and regulation .....	6
1.2.3 Exportin-5 transportation mechanisms .....	8
1.2.4 Dicer/TRBP processing and regulation .....	10
1.2.5 RISC assembly and regulation .....	15
1.3 miRNA-31 in cancers.....	18
1.4 Antisense oligo application and development.....	21
1.5 Reference.....	28
<b>CHAPTER II pH Dependence of C•A, G•A and A•A Mismatches in the Stem of Precursor MicroRNA-31</b> .....	<b>44</b>
2.1 Introduction .....	45
2.2 Material and methods .....	47
2.2.1 Cataloging of mismatches in human pre-miRNAs.....	47
2.2.2 Construct design and template preparation .....	48
2.2.3 RNA Preparation .....	49
2.2.4 NMR experiments.....	50
2.2.5 pH titrations and pKa measurements.....	51
2.2.6 Thermal denaturation of RNA and data analysis.....	52
2.3 Results .....	52

2.3.1 70% of human pre-miRNAs contain at least one base pair mismatch in the stem.....	52
2.3.2 NMR data reveal the pre-miRNA-31 base pair mismatches have different conformations .....	53
2.3.3 Stabilization of the structure in the C·A mismatch at lower pH.....	58
2.3.4 pK <sub>a</sub> values of adenosine nucleotides in mismatches differ .....	64
2.4. Discussion .....	68
2.5 Reference.....	73
<b>CHAPTER III Structure of Pre-miR-31 Reveals an Active Role in Dicer/TRBP Complex Processing .....</b>	<b>83</b>
3.1 Introduction .....	84
3.2 Results .....	87
3.2.1 The secondary structure of FL-pre-miR-31 contains three mismatches in the helical stem and three base pairs in the apical loop. ....	87
3.2.2 Tertiary structure of pre-miR-31 .....	90
3.2.3 TRBP inhibits Dicer processing of pre-miR-31. ....	92
3.2.4 Mismatches within the helical stem region have no impact on Dicer-TRBP cleavage.....	92
3.2.5 Structure at the cleavage site affects Dicer-TRBP processing. ....	94
3.2.6 Size and relative position of the apical loop regulates Dicer-TRBP processing.....	95
3.2.7 Junction residues function as critical control elements for Dicer processing.....	97
3.3 Discussion .....	100
3.4 Methods.....	104
3.4.1 Preparation of recombinant human Dicer.....	104
3.4.2 Expression and purification of TRBP.....	105
3.4.3 Dicer-TRBP complex formation .....	106
3.4.4 Preparation of DNA templates .....	106
3.4.5 Preparation of plasmid templates for in vitro transcription.....	106
3.4.6 Preparation of RNA .....	107
3.4.7 Dimethyl sulfate (DMS) modification of pre-miR-31 RNA .....	108
3.4.8 RT-PCR with DMS-modified RNA .....	109
3.4.9 DMS-MaPseq of pre-miR-31 RNA.....	109
3.4.10 Isotopic labeling of RNAs for NMR. ....	110
3.4.11 NMR experiments. ....	110
3.4.12 Structure calculations. ....	112
3.4.13 Small angle X-ray scattering (SAXS) .....	113
3.4.14 <sup>32</sup> P labeling of RNA.....	114
3.4.15 Dicer and Dicer-TRBP processing assay.....	114
3.5 Reference.....	117
<b>CHAPTER IV Use of Steric Blocking Antisense Oligonucleotides for the Targeted Inhibition of Junction Containing Precursor MicroRNAs .....</b>	<b>125</b>
4.1 Introduction .....	125
4.2 Materials and Methods .....	128

4.2.1 Preparation of recombinant human Dicer.....	128
4.2.2 Expression and purification of TRBP.....	129
4.2.3 Dicer/TRBP complex formation.....	130
4.2.4 Preparation of DNA templates for in vitro transcription.....	130
4.2.5 Preparation of RNA.....	131
4.2.6 <sup>32</sup> P labeling of RNA.....	131
4.2.7 Dicer/TRBP antisense oligo cleavage assays.....	132
4.2.8 Dicer/TRBP pre-miRNA competition assays.....	133
4.2.9 Dicer/TRBP pre-miRNA SYBR Gold cleavage assays.....	134
4.2.10 Cataloging of junction regions of human pre-miRNAs.....	135
4.2.11 pmirGLO plasmid construction.....	136
4.2.12 pCMV plasmid construction.....	137
4.2.13 In cell dual luciferase assays.....	137
4.2.14 Quantitative Real-Time PCR Analysis.....	137
4.3 Results.....	138
4.3.1 Antisense oligos inhibit pre-miR-31 cleavage by Dicer/TRBP.....	138
4.3.2 Pre-miR-144 is a targetable pre-miRNA containing a junction region.....	142
4.3.3 The junction region is a common structural feature within pre-miRNAs.....	146
4.3.4 Targeting junction containing pre-miRNAs is a reasonable approach for inhibiting Dicer/TRBP processing.....	147
4.3.5 ASO targeting of junction-containing pre-miRNAs inhibit Dicer cleavage in cell ...	150
4.4 Discussion.....	153
4.5 Reference.....	159
<b>CHAPTER V Conclusion and Future Directions.....</b>	<b>164</b>
5.1 Conclusion.....	164
5.2 Future directions.....	166
5.2.1 Junction region dynamics of pri-miR-31 and its role in Drosha/DGCR8 processing	166
5.2.2 pre-miR-31 structural role in regulating Dicer/TRBP binding affinity.....	169
5.2.3 Novel design of shRNA combined with junction region targeted ASOs.....	170
5.3 Reference.....	173
<b>Appendix A Supplemental Figure and Tables for Chapter III.....</b>	<b>177</b>
<b>Appendix B Dicer Only Data and Conclusions for Chapter III.....</b>	<b>222</b>



## LIST OF TABLES

Table 2.1. RNA constructs.....	49
Table 2.2. NMR experimental parameters.....	51
Table 2.3. Intermolecular NOE cross-peaks defining the A·A, G·A, and C·A mismatches. ....	58
Table 2.4. Thermal stability of BottomA and BottomB RNAs. ....	67
Table 4.1. Overlap extension PCR primers for generation of pre-miRNA templates. ....	155
Table 4.2. Amplification primers for template. ....	156
Table 4.3. RNA sequences.....	156
Table 4.4. Antisense oligo sequences. ....	157
Table 4.5. Templates for generating pmirGLO plasmid inserts. ....	158
Table 4.6. Overlap extension PCR primers for generation of pCMV miR31 plasmid insert. ....	158
Appendix A Table S1. Chemical shift completeness. <sup>a</sup> .....	203
Appendix A Table S2. NMR restraints and structural statistics for the FL pre-miR-31 structure. <sup>a</sup> .....	204
Appendix A Table S3. SEC-SAXS data acquisition, sample details, data analysis, model fitting, and software used.....	205
Appendix A Table S4. Dicer+TRBP cleavage of pre-miR-31 RNAs. ....	207
Appendix A Table S5. Dicer cleavage of pre-miR-31 RNAs.....	208
Appendix A Table S6. Thermal stability of pre-miR-31 RNAs. ....	210
Appendix A Table S7. Synthetic DNA templates and associated RNA constructs.....	211
Appendix A Table S8. DNA primers for pre-miR-31-tail (DMS) experiments. ....	212
Appendix A Table S9. DNA primers for generation of the pre-miR-31 (NMR) template.....	213

Appendix A Table S10. Amplification primers for template. ....	214
Appendix A Table S11. DNA primers for HH-pre-miR-31 template. ....	215
Appendix A Table S12. Mutation DNA primers for processing constructs. ....	216
Appendix A Table S13. RNA sequences used for structural and processing studies. ....	219
Appendix B Table 1. Dicer binding affinity of pre-miR-31 mutations. ....	236

## LIST OF FIGURES

Figure 1.1. Overview of miRNA biogenesis pathway.....	5
Figure 1.2 Structure of XPO5:RanGTP:pre-miRNA complex.....	9
Figure 1.3. Mechanistic insights into Drosophila Dicer-1 and human Dicer .....	11
Figure 1.4. Dicer specifically recognizes and cleaves pre-miRNAs .....	12
Figure 1.5. Protein regulation pathway for let-7 family pre-miRNAs.....	15
Figure 1.6. RISC assembly and function pathway.....	17
Figure 1.7. ASOs use different mechanisms to modulate gene expression .....	22
Figure 2.1. Statistics of base pair mismatches in human pre-miRNA stems.....	46
Figure 2.2. Predicted secondary structure of pre-miRNA-31 .....	54
Figure 2.3. Assigned chemical shifts of BottomA and BottomB RNAs .....	55
Figure 2.4. Overlay of the aromatic-anomeric region of the 1H-1H NOESY spectra.....	55
Figure 2.5. Sequence analysis and validation of BottomA and BottomB chemical shift assignments .....	56
Figure 2.6. pH dependence of chemical shifts.....	61
Figure 2.7. H2-N1 cross-peak of 100% 15N1 labeled A64 in 1H-15N HSQC spectra recorded at pH=7.5 and pH=5.2 .....	62
Figure 2.8. Aromatic-anomeric region of BottomB RNA 1H-1H NOESY spectrum .....	62
Figure 2.9. H2-N1 cross-peaks of A58 and A54 are much broader compared to A64. H2-N1 cross- peaks of 100% 15N1 labeled .....	63
Figure 2.10. Protonation of A54 is favored at low pH.....	63

Figure 2.11. H8-C8 correlations in 1H-13C HSQC spectra of 100% 13C8 labeled A53 and A54 at different pH values.....	65
Figure 2.12. Thermal stability of BottomA and BottomB RNAs.....	66
Figure 3.1. Conflicting secondary structure models for pre-miR-31 apical loop. ....	88
Figure 3.2. Tertiary structure of pre-miR-31. ....	91
Figure 3.3. Structure at the dicing site serves as an important feature for Dicer-TRBP processing. ....	94
Figure 3.4. Pre-miR-31 requires a greater than 4-nt apical loop for efficient processing. ....	96
Figure 3.5. The junction region is a regulatory element within pre-miR-31. ....	99
Figure 3.6. Secondary structure elements and their contribution to the regulation of pre-miR-31 processing. ....	103
Figure 4.1 Flowchart of junction region-containing pre-miRNAs determination .....	136
Figure 4.2. Antisense oligos that target the pre-miR-31 junction structure are potent and specific inhibitors of Dicer/TRBP processing in vitro .....	141
Figure 4.3. ASOs targeting the pre-miR-144 junction inhibit Dicer/TRBP processing in vitro	143
Figure 4.4. Anti-miR144 ASO are specific and do not affect the processing of pre-miR-31 or pre-let-7c .....	145
Figure 4.5. ASO 144-A4 inhibits Dicer/TRBP processing of pre-miR-144.....	145
Figure 4.6. Junction regions are a common feature of pre-miRNAs and are effective ASO targets .....	146
Figure 4.7. Secondary structure of pre-miR-19a and GG-pre-miR-19a .....	148
Figure 4.8. Representative GG-pre-miR-19a processing assay, visualized by SYBR gold staining .....	149

Figure 4.9. Representative pre-miR-143 processing assay, visualized by SYBR gold staining	149
Figure 4.10. Apical loop region targeted ASO does not affect pre-let-7c processing by Dicer/TRBP .....	150
Figure 4.11. Anti-junction-region ASOs can inhibit pre-miRNA maturation in cell .....	153
Figure 5.1. Two conformations of pre-miR-31 .....	166
Figure 5.2. Proposed mutants of pri-miR-31 junction region .....	169
Figure 5.3 New principle for designing shRNA .....	172
Appendix A Fig. S1. In vitro DMS-MaPseq of pre-miRNA-31 RNA .....	178
Appendix A Fig. S2. Oligo controls for NMR chemical shift assignment of FL-pre-miR-31 ...	179
Appendix A Fig. S3. Assigned chemical shifts of TopA RNA .....	180
Appendix A Fig. S4. Assigned chemical shifts of Top RNA .....	182
Appendix A Fig. S5. Deuterium labeling improves spectral quality by reducing overlap .....	183
Appendix A Fig. S6. Summary of the secondary structure, NOE connectivity, and chemical shift assignment validation for FL pre-miR-31.....	184
Appendix A Fig. S7. Imino proton NOESY spectrum of pre-miR-31 .....	185
Appendix A Fig. S8. Imino region of <sup>1</sup> H spectra of pre-miR-31 FL at different temperatures.	186
Appendix A Fig. S9. pH-dependence of unpaired adenosines .....	187
Appendix A Fig. S10. Solvent paramagnetic relaxation effect analysis of FL pre-miR-31 reveals solvent accessibility in the loop region.....	188
Appendix A Fig. S11. Sequence conservation of pre-miR-31.....	189
Appendix A Fig. S12. Schematic presentation of inter-residue NOEs defining the pre-miR-31 junction base pairs.....	190
Appendix A Fig. S13. Schematic presentation of inter-residue NOEs defines the apical and internal	

loops in pre-miR-31 .....	191
Appendix A Fig. S14. Correlation plot between measured and back-calculated RDCs for the lowest energy pre-miR-31 structure. ....	192
Appendix A Fig. S15. Validation of pre-miR_31 tertiary structure by SAXS .....	193
Appendix A Fig. S16. Structure of pre-miR-31 is not sensitive to the monovalent cation .....	194
Appendix A Fig. S17. Structure of pre-miR-31 is not sensitive to the divalent cation .....	195
Appendix A Fig. S18. Dicer processing co-factor, TRBP, inhibits pre-miR-31 processing .....	196
Appendix A Fig. S19. Mismatches in the stem of pre-miR-31 do not significantly impact Dicer-TRBP processing of the substrate RNA.....	197
Appendix A Fig. S20. Mismatches in the stem region have no impact on Dicer processing.....	198
Appendix A Fig. S21. Structure at the dicing site serves as a control element for Dicer processing. ....	199
Appendix A Fig. S22. Dicer processing assays for apical loop mutations.....	200
Appendix A Fig. S23. A two base-pair junction between the apical loop and dicing site recovers reduced Dicer processing efficiency due to large apical loop size .....	201
Appendix A Fig. S24. Thermal stability of junction mutants.....	202
Appendix B Figure 1. Structure at the dicing site serves as a control element for Dicer processing .....	225
Appendix B Figure 2. Apical loop size is optimized for efficient Dicer binding and processing .....	227
Appendix B Figure. 3. Extending the length of the helical junction between the apical loop and dicing site reduces processing efficiency and accuracy.....	228
Appendix B Figure 4. The junction region is a regulatory element within pre-miR-31.....	230

Appendix B Figure 5. Secondary structure elements and their contribution to the regulation of pre-miR-31 processing ..... 237

## LIST OF ABBREVIATIONS

miRNA: microRNA

miR-31: microRNA-31

ASO: antisense oligonucleotide

pre-miRNA: precursor microRNA

siRNA: small interfere RNA

ncRNA: noncoding RNA

nt: nucleotide

UTR: untranslated region

mRNA: messenger RNA

CCL: chronic lymphocytic leukemia

qPCR: quantitative real-time polymerase chain reaction

SOCS-3: suppressor of cytokine signaling 3

pri-miRNA: primary miRNA

pre-miRNA: precursor miRNA

RISC: RNA-induced silencing complex

bp: base pair

DGCR8: DiGeorge Syndrome Critical Region 8

XPO5: exportin-5

TRBP: transactivation responsive RNA-binding protein

AGO: Argonaute



TNRC6: Trinucleotide Repeat Containing 6

CED: central domain

dsRBD: double strand RNA binding domain

dsRNA: double strand RNA

ssRNA: single strand RNA

hDicer: human Dicer

dmDicer-1: Drosophila Dicer-1

CSD: cold shock domain

ZKD: zinc knuckle domain

TUTase: terminal uridine transferase

DIS3L2: DIS3 like 3'-5' Exoribonuclease 2

ILF3: Interleukin Enhancer Binding Factor 3

AMP: Adenosine monophosphate

UMP: uridine monophosphate

PABP: poly(A)-binding protein

CCR4-NOT: Carbon Catabolite Repression-Negative On TATA-less

4PH: Type I collagen prolyl 4-hydroxylase

AKT3: RAC $\gamma$  Serine/Threonine protein kinase

E2F2: E2F Transcription Factor 2

STK40: Serine threonine kinase 40

RB: retinoblastoma protein

AR: androgen receptor

TNBC: Triple negative breast cancer

PKCε: protein kinase C epsilon

NSCLC: non-small-cell lung carcinomas

ESCC: Esophageal squamous cell carcinoma

SATB2: Special AT-rich Sequence-binding protein 2

ERK5: Extracellular Signal-regulated Kinase 5

LATS2: Large tumor suppressor kinase 2

ADC: Adenocarcinoma

ssASO: single stranded ASO

dsASO: double strand ASO

NMD: Non-sense-mediated decay

EJC: exon–exon junction complex

PTC: premature termination codon

CMV: cytomegalovirus retinitis

Apo B: Apolipoprotein B

TTR: Transthyretin

DMD: Duchenne muscular dystrophy

SMN2: survival of motor neuron 2

DTT: dithiothreitol

NTP: ribonucleoside triphosphate

MgCl<sub>2</sub>: magnesium chloride

DMSO: dimethyl sulfoxide

EDTA: ethylenediaminetetraacetic acid

TBE: Tris/Borate/EDTA

DMT: 4,4'-dimethoxytrityl

NOESY:  $^1\text{H}$ - $^1\text{H}$  nuclear Overhauser effect spectroscopy

TOCSY:  $^1\text{H}$ - $^1\text{H}$  total correlation spectroscopy

HMQC:  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear multiple quantum correlation

WT: wild type

shRNA: short hairpin RNA

FL: full length

SHAPE: selective 2' hydroxyl acylation analyzed by primer extension

DMS-MaPseq: dimethyl sulfate mutational profiling with sequencing

sPRE: solvent paramagnetic relaxation enhancement

NOE: Nuclear Overhauser effect

RDC: residual dipolar coupling

SAXS: small angle X-ray scattering

TCEP: tris(2-carboxyethyl) phosphine

OE: overlap-extension

PCR: polymerase chain reaction

HH: hammerhead

TCI: three channel inverse

SEC: size exclusion chromatography

MALS: multiangle light scattering

DLS: dynamic light scattering

RI: refractive index

uORF: upstream open reading frame

cDNA: complementary DNA

DEME: Dulbecco's Modified Eagle Medium

FBS: fetal bovine serum

CEST: Chemical exchange saturation transfer

CPMG: Carr–Purcell–Meiboom–Gill spin echo

## LIST OF APPENDICES

Appendix A Supplemental Figure and Tables for Chapter III.....**Error! Bookmark not defined.**

Appendix B Dicer Only Data and Conclusions for Chapter III.....**Error! Bookmark not defined.**

## ABSTRACT

As an essential post-transcriptional regulator of gene expression, microRNA (miRNA) levels must be strictly maintained. The biogenesis of many miRNAs is regulated by trans-acting protein partners, which exert control through a variety of mechanisms, including remodeling of the RNA structure and recruitment of RNA processing or modifying enzymes. MicroRNA-31 (miR-31) functions as an oncogene in numerous cancers, and interestingly, its biogenesis is not known to be regulated by protein factors. Therefore, I investigated if the intrinsic structural and dynamical properties of the miR-31 precursor element, pre-miR-31, can provide a mechanism by which its biogenesis is regulated. Base pair mismatches are a common feature of primary and precursor miRNAs. In this thesis, I characterized the base pair mismatches within pre-miR-31 and found the C•A mismatch within the stem of pre-miR-31 to be strongly pH sensitive and, stabilizing the RNA structure at near physiological pH. Next, I investigated the role of distinct structural elements within pre-miR-31 in regulating processing by the Dicer/TRBP complex. I found that both the apical loop size and structure at the Dicing site are key elements for discrimination by the Dicer/TRBP complex. Interestingly, our NMR-derived structure revealed the presence of a triplet of base pairs, or junction region, that link the Dicer cleavage site and the apical loop. My mutational analysis in this region revealed that the stability of the junction region strongly influenced processing by the Dicer/TRBP complex. Based on these findings, I developed a new type of antisense oligonucleotide (ASO) that specifically targets the junction region of pre-miR-

31 to inhibit Dicer/TRBP processing. Furthermore, I demonstrated that this new type of ASO design is broadly applicable to other junction containing pre-miRNAs, which account for ~20 % of human pre-miRNAs, and function to reduce the Dicer/TRBP cleavage of this family of pre-miRNAs. These studies enhance our understanding of RNA structure based ASO design and development. The results in this thesis enrich our understanding of the active role that RNA structure plays in regulating miRNA biogenesis, which has direct implications for the control of gene expression. This study further points out that RNA structure is not a passive element in the protein enzymatic steps. Rather, the RNA structural elements play important roles in regulating processing by Dicer/TRBP. Furthermore, my thesis work provides a new approach for antisense oligonucleotide design by targeting microRNA biogenesis at a step upstream of the traditional anti-microRNA antisense oligonucleotide design strategy. This new type of antisense oligonucleotide allows intervention at an early stage of miRNA biogenesis and may lead to a novel treatment by selectively inhibiting disease-related pre-miRNAs.

## CHAPTER I Introduction

This dissertation defines key structural and dynamical features that are important for regulating precursor microRNA (pre-miRNA) maturation and develops a novel type of antisense oligonucleotide (ASO) that can specifically inhibit the maturation of a sub population of pre-miRNA. In this chapter, I review the history of microRNA discovery and the biogenesis of microRNAs inside the cell. I detail how microRNAs are enzymatically processed and how each processing step is post-transcriptionally regulated. I discuss the role of miR-31, the main miRNA target in these studies, in multiple cancers and detail the application of antisense oligos therapeutics including microRNAs and small interfere RNAs (siRNAs) therapeutics.

### 1.1 The history of microRNA

#### 1.1.1 The discovery of microRNA

MicroRNAs (miRNA) are a family of small noncoding RNA (ncRNA), which are approximately 22 nucleotides (nt) in length. Three decades ago, the importance and existence of miRNAs were unknown. In 1993, the first microRNA, *lin-4*, was discovered by the Ambros and Ruvkun laboratories [1-3]. In the nematode *C. elegans*, *lin-4* is a gene that controls the timing of larval development and interestingly, this gene produces a pair of short RNAs rather than encoding a protein [1]. Previous studies reported that *lin-14* translation was repressed by *lin-4* gene product



[4]. Two RNA transcripts were identified from the *lin-4* gene; one RNA that was 22 nt in length and a second that was 61 nt long [1]. The longer transcript was proposed to be a precursor element of the shorter RNA and this precursor RNA was predicted to fold into a stem loop structure. The short *lin-4* RNA was shown to bind to the 3' untranslated region (UTR) of the *lin-14* messenger RNA (mRNA) [1, 5]. The Ruvkun lab demonstrated that this binding can downregulate *lin-14* protein expression level without reducing *lin-14* mRNA level [5]. These studies revealed the important role of this short noncoding RNA in regulating cell function in *C. elegans*.

Seven years later, a second miRNA, let-7, was shown to control the L4-to-adult transition of larval development in *C. elegans* [6]. Let-7 is complementary to the 3' UTR of *lin-41* and both the deletion of the *lin-41* 3' UTR and let-7 mutations were shown to diminish the repression of *lin-41* protein translation, indicating the importance of regulation between let-7 and *lin-41* [6-8]. Unlike *lin-4*, the let-7 sequence is conserved across multiple species, from flies to humans [9]. In humans, the expression levels of let-7 varied across different tissues, including the brain, heart, and kidney [9]. The let-7 family of miRNAs shares the same seed sequence, which is defined as position 2-7 from the miRNA 5'-end. The human let-7 family includes 12 miRNAs while in *C. elegans*, only four miRNAs (let-7, miR-48, miR-84 and miR-241) are classified as members of the let-7 family [9-12].

The discovery of let-7 triggered a revolution in the small noncoding RNA research area. More and more miRNAs were discovered over the past three decades and databases detailing miRNA sequences in different species, such as the miRbase database [13-17], were established.

### **1.1.2 The role of miRNAs in diseases**

Over the past three decades, not only were more and more miRNAs discovered, but also a

link between the dysregulation of miRNAs levels and diseases was revealed. The dysregulation of miRNA levels was associated with different diseases, including cancer, cardiovascular diseases, autoimmune disease and neurodegenerative diseases [3].

In 2002, miR-15a and miR-16-1 were first proposed as tumor suppressors in chronic lymphocytic leukemia (CLL) [18]. The miR-15a and miR-16-1 genes were frequently deleted or downregulated in more than 60% of human B-cell CLL. These findings suggested the tumor suppressor role that miR-15 and miR-16 may play in CLL. The first biological function of miRNAs in cancer development was revealed in 2005 [19]. Different microRNAs can function as tumor suppressors or to promote tumorigenesis (also known as oncomiRs) in different types of cancers and can impact numerous biopathways [19]. With the development of novel technologies, such as quantitative real-time polymerase chain reaction (qPCR) [20], high-throughput array-based Klenow enzyme assays [21], and deep sequencing [22], the alternation of miRNA expression profile has been discovered in almost all cancers [23, 24].

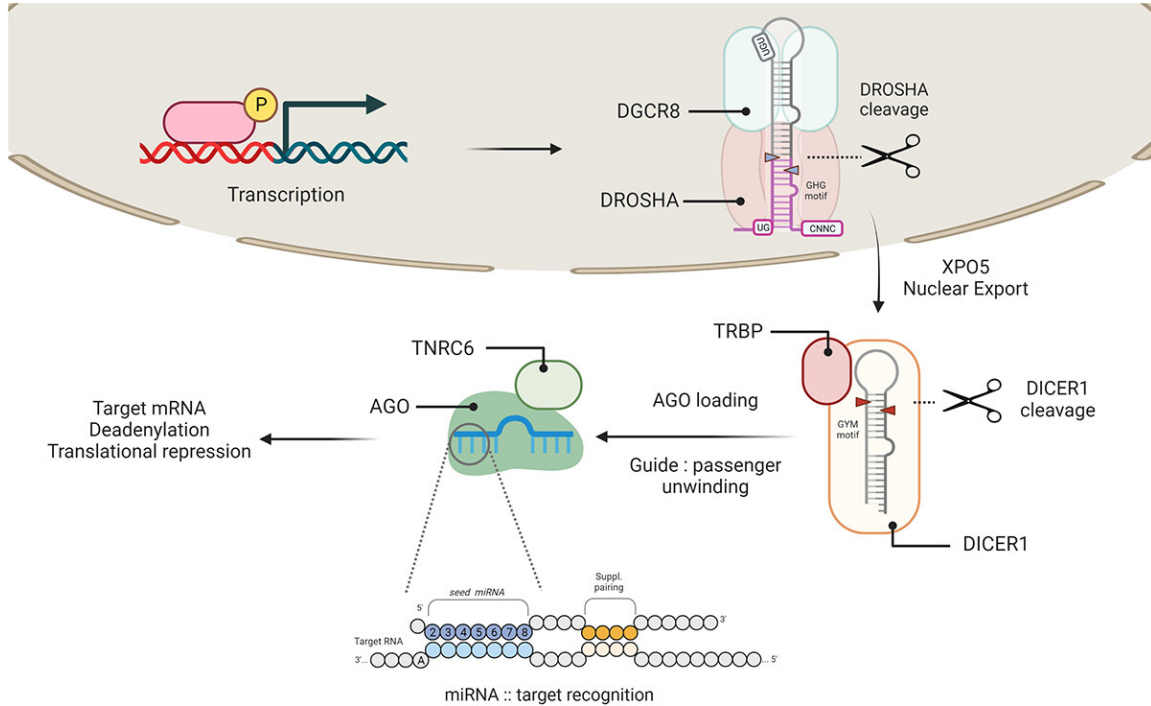
In 2005, the possible role of miR-1-1 and miR-1-2 in cardiovascular diseases was revealed [25]. Zhao and Samal revealed that miR-1-1 and miR-1-2 were specifically expressed in the cardiac and skeletal muscle precursor cells, and that miR-1 overexpression resulted in a reduced population of actively dividing ventricular cardiomyocytes in the developing heart of a mouse model [25]. miR-1 was also identified as an important gene that regulates the *Drosophila* cardiogenesis [26] and was shown to promote myogenesis in concert with miR-133 in mice [27]. The first report that strongly correlated the microRNAs with heart failure and cardiac hypertrophy was published in 2006 [28]. In this paper, van Rooij et al. showed that more than 12 miRNAs were dysregulated in the heart failure and cardiac hypertrophy and overexpression of one miRNA, miR-195, was enough to cause heart failure and cardiac hypertrophy in the mouse model [28, 29].

MicroRNAs have also been implicated in autoimmune disease development. In 2007, Sonkoly et al. discovered that miR-203 is overexpressed in psoriasis, which is the most common chronic inflammatory skin disease [30]. The miR-203 targets suppressor of cytokine signaling 3 (SOCS-3), an important negative regulator of cytokine signaling, and the downregulation of SOCS-3 resulted in spontaneous development of psoriasis in mice [30-32]. MiR-146 is another psoriasis associated miRNA. MiR-146 was also overexpressed in psoriasis patients and the upregulation of miR-146 was associated with inflammation in psoriasis patients [30, 33]. The relationship between other autoimmune diseases and miRNAs continues to be uncovered. More recently, the overexpression of miR-155 and miR-146 were found in the rheumatoid arthritis synovial tissue and this overexpression causes the dysregulation inflammatory immune response [34].

Neurodegenerative diseases involve the progressive degeneration of neurons in the nervous system [35]. Some miRNAs play important roles in the neuronal development [36], therefore, the dysregulation of these miRNAs was linked to neurodegenerative diseases. In 2007, Lukiw et al. found that miR-9, miR-25b and miR-128 were overexpressed while miR-124a was underexpressed in brain samples of the Alzheimer's patients when compared with healthy age-matched samples [37, 38]. Eight miRNAs are known to be consistently dysregulated in Alzheimer's disease experimental models, including miR-29, miR-15, miR-107, miR-146, miR-9, miR-101 and the miR-212/132 cluster [35]. MiR-133b was shown to be downregulated in Parkinson's disease [39], which is the neurodegenerative disease with the second highest prevalence, affecting approximately 1% of the population over 55 years old [35].

## **1.2 miRNA biogenesis pathway and post-transcriptional regulation**

### 1.2.1 General miRNA biogenesis pathway



**Figure 1.1. Overview of miRNA biogenesis pathway [40].** (Figure adopted from Bofill-De Ros, X. and U.A. Vang Ørom, ‘Recent progress in miRNA biogenesis and decay’, [CC BY-NC 4.0 Deed Attribution-NonCommercial 4.0 International](https://doi.org/10.1080/15476286.2023.2288741). The figure source data is from: <https://doi.org/10.1080/15476286.2023.2288741>) The primary miRNA (pri-miRNA) is transcribed from the genome. The pri-miRNA basal helix is cleaved by the Drosha/DGCR8 complex, producing the precursor miRNA (pre-miRNA). The pre-miRNA is then exported into cytoplasm where the terminal loop is processed by Dicer/TRBP. Dicer/TRBP processing yields a mature miRNA, which is bound by the AGO protein to form the RNA-induced silencing complex (RISC) which functions to regulate gene expression.

MicroRNA genes are initially produced as longer transcripts which undergo two sequential enzymatic processing steps to generate mature miRNAs. The canonical miRNA biogenesis pathway is the most common biogenesis pathway that miRNAs use for maintaining expression levels and sequence integrity (**Fig. 1.1**) [41, 42]. The canonical miRNA biogenesis pathway begins with the transcription of the miRNA gene by RNA polymerase II [41-43]. The first and longest transcription product is named the primary microRNA (pri-miRNA). The pri-miRNA contain a stem-loop structure which sequesters the mature miRNA sequences within the helical stem (**Fig.**

**1.1)** [42]. A typical pri-miRNA has a 33-35 base pair (bp) stem, a terminal or apical loop, and long overhangs at both the 5' and 3'-ends, which forms a basal junction region. Pri-miRNAs are enzymatically processed by Drosha in complex with its co-factor, DiGeorge Syndrome Critical Region 8 (DGCR8), to generate a medium length product, known as a precursor microRNA (pre-miRNA) [44-47]. Pre-miRNAs are ~60-70 nt long and adopt a stem-loop structure with 2 nt overhang at the 3'-end. Pre-miRNAs are transported from nucleus to cytoplasm by the exportin-5 (XPO5)/RanGTP complex [48-50]. After transport into cytoplasm, pre-miRNAs are recognized and cleaved by the Dicer protein in complex with transactivation responsive RNA-binding protein (TRBP) [51-55]. Dicer/TRBP cleaves off the pre-miRNA terminal loop and releases the miRNA duplex (**Fig. 1.1**). The miRNA duplex is loaded into the Argonaute (AGO) protein, and the passenger strand of the miRNA duplex is discarded by AGO protein with guide strand remaining inside AGO protein. The guide strand and AGO protein forms the RNA-induced silencing complex (RISC) (**Fig. 1.1**) [56, 57]. RISC recognizes target mRNA based on sequence complementarity between the loaded miRNA guide strand and the mRNA 3' UTR. This binding represses mRNA translation and/or degrades mRNA by recruiting Trinucleotide Repeat Containing 6 (TNRC6) protein (**Fig. 1.1**) [56-59].

### **1.2.2 Microprocessor processing and regulation**

The Microprocessor is comprised of one Drosha protein and dimer of DGCR8 proteins. DGCR8 is indispensable for Microprocessors to maintain both efficient and accurate cleavage activity [46]. Drosha contains an intramolecular heterodimeric RNase III processing center, a PAZ-like domain and Platform domain inside the central domain (CED), and two double strand RNA binding domains (dsRBD), which bind double strand RNA (dsRNA) [60, 61]. The PAZ-like

and Platform domain binds to the stem region of the pri-miRNA while DGCR8 proteins bind the terminal loop of the pri-miRNA. The Belt and Wedge domains, which are inside CED, clamp over the single strand RNA (ssRNA) and this clamping helps Drosha to recognize the basal junction by forming a four-way intramolecular junction [61]. The two dsRBDs measure the stem length between dsRNA-ssRNA junctions, which is around 35 bp, and RNase III domains cleave pri-miRNAs to form pre-miRNAs, and processing occurs approximately 11 bps from the basal junction in the pri-miRNA [60, 61].

Multiple structural and sequence features in pri-miRNAs are recognized by Microprocessor to determine the cleavage site (blue arrow in **Fig. 1.1**). Sequence elements within the pri-miRNAs play a role in regulating the efficiency and precision of microprocessor processing. The upstream UG motif and the downstream CNNC motif, located within the 5' and 3' flanking sequence region, respectively, have been shown to be important for efficient Microprocessor processing (**Fig. 1.1**). These two motifs are vital for Serine and Arginine Rich Splicing Factor 3 (SRSF3) protein binding and promotes Drosha/DGCR8 processing [62]. The GHG mismatch motif (where H represents a mismatched A, C or U, positioned between two Gs that are base paired) is present in the lower stem and is important for Drosha/DGCR8 processing accuracy (**Fig. 1.1**). The structural studies on Drosha/DGCR8 have shown that GHG motif helps the pri-miRNA to be recognized by Drosha dsRBDs and ensures precise cleavage by the Drosha protein [61, 63, 64]. The UGU motif, which is located within the apical loop of some pri-miRNAs, interacts with the heme-containing domains of DGCR8 and promotes the accurate and efficient processing of pri-miRNAs by microprocessor [40, 65].

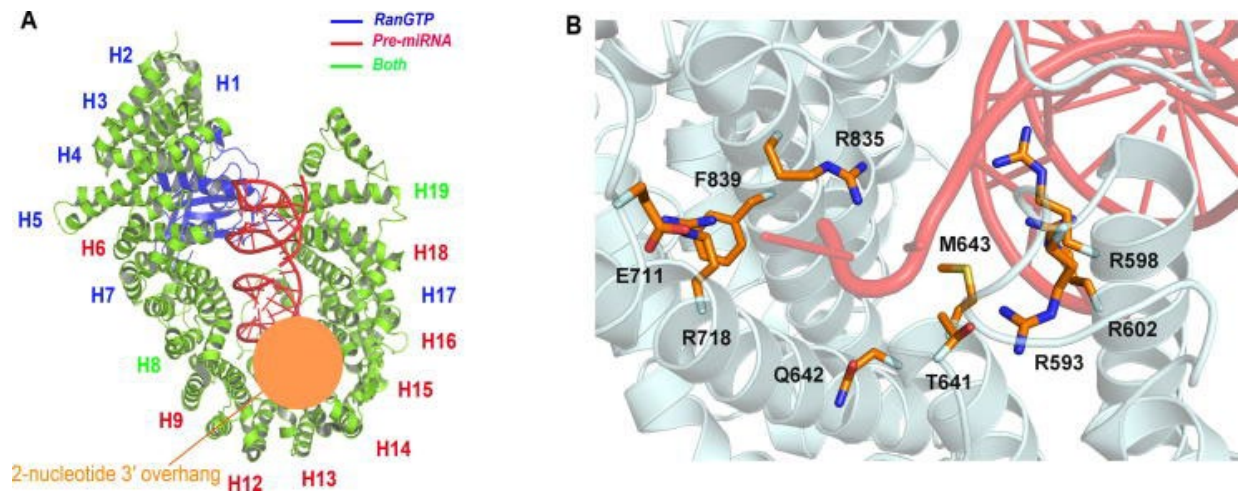
Not all pri-miRNA transcripts contain the sequence features described above. These “suboptimal” pri-miRNAs are poorly processed by Microprocessor. However, the cleavage of

suboptimal pri-miRNAs can be enhanced with the help of neighboring pri-miRNA. This phenomenon is called cluster assistance [66]. One example is the miR-144-451 cluster. Pri-miR-451 is a poor substrate for Drosha/DGCR8 cleavage when it is isolated as a single pri-miRNA. However, pri-miR-451 and pri-miR-144 are co-transcribed to form a miRNA cluster. The presence of pri-miR-144 in this cluster recruits the Microprocessor to the cluster and helps nearby pri-miR-451 to be efficiently processed by Drosha/DGCR8 [67-69]. The miR-15a-16-1 cluster is another miRNA cluster that has a similar cluster assistance phenomenon. With the help of a protein co-factor, Scaffold Attachment Factor B2 (SAFB2), pri-miR-16-1 motivates the cleavage of pri-miR-15a by Drosha/DGCR8 [70].

Protein binding partners can also promote processing of Drosha/DGCR8 cleavage on some pri-miRNAs. XPO5 was previously thought to only mediate the transportation of the pre-miRNA from the nucleus to the cytoplasm. However, recent studies found that XPO5 can enhance Microprocessor cleavage efficiency on the miR-17-92a cluster [71]. Heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) is another protein that can enhance the Drosha/DGCR8 processing efficiency on pri-miR-18a by remodeling the apical loop region [72].

### **1.2.3 Exportin-5 transportation mechanisms**

As a member of the karyopherin  $\beta$  family of transport factors, exportin-5 shuffles from nucleus to the cytoplasm, using a Ran-GTPase to control cargo association [73]. XPO5 can transport a variety of substrates out of nucleus, including pre-miRNAs, tRNAs, some RNA-binding proteins and ribosomal subunits [73-76]. XPO5 can also stabilize the pre-miRNAs and regulate cell cycle entry by increasing global miRNA expression [77].



**Figure 1.2 Structure of XPO5:RanGTP:pre-miRNA complex [73].** (Figure adopted from Wu, K et. al., ‘The Role of Exportin-5 in MicroRNA Biogenesis and Cancer’, [CC BY-NC 4.0 Deed | Attribution-NonCommercial 4.0 International](https://doi.org/10.1016/j.gpb.2017.09.004). The figure source data is from: <https://doi.org/10.1016/j.gpb.2017.09.004>) **A.** Structure showing how XPO5 (green), RanGTP (blue) and a human pre-miRNA (red) interact to form a complex. The stem region of the pre-miRNA is mostly embedded in the XPO5 tunnel-like structure, and the 2-nt overhang at the pre-miRNA 3'-end (orange circle) makes many interactions with residues in XPO5. ‘H’ indicates HEAT repeats of XPO5. **B.** Detailed intermolecular interactions between the pre-miRNA 3' overhang and XPO5 HEAT repeats 12-15. PDB ID: 3A6P.

The transport of pre-miRNAs involves several steps. XPO5 binds to pre-miRNAs in the nucleus in a complex with RanGTP. The newly formed complexes go through the nuclear pore complex into cytoplasm. The pre-miRNAs are released from the complex by hydrolysis of GTP. The free XPO5 then returns to the nucleus and start another round of pre-miRNA transport [73]. XPO5 could be a rate-limiting step in the miRNA biogenesis as the overexpression of XPO5 was found to enhance mature miRNA expression [78].

According to structural studies, XPO5 recognizes the dsRNA stem structure of pre-miRNAs using its inherent tunnel-like structure (**Fig. 1.2 A**) [79, 80]. XPO5 is comprised of 19 HEAT repeats (a tandem repeat protein structural motif composed of two alpha helices linked by a short loop) and HEAT motifs 8, 9, 12-16, 18 and 19 interact with the pre-miRNA and stabilize the pre-miRNA (**Fig. 1.2 A**) [79, 80]. An important recognition features of pre-miRNAs is the 2-



nt 3'-end overhang. This overhang interacts with R593, R598, R602, T641, Q642, M643, E711, R718, R835, and F839 by hydrogen bonding and salt bridge formation (**Fig. 1.2 B**).


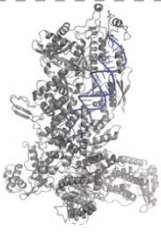




#### **1.2.4 Dicer/TRBP processing and regulation**

In the cytoplasm, pre-miRNAs need to be cleaved by Dicer/TRBP to produce the mature miRNA duplex (**Fig. 1.1**) [40]. Unlike the Drosha/DGCR8 complex, Dicer can cleave pre-miRNAs efficiently without its partner protein, TRBP [42]. However, TRBP plays an important role in promoting selective Dicer cleavage of pre-miRNAs in a complex RNA milieu [81], enhancing Dicer and pre-miRNAs binding affinity [81], enhancing Dicer cleavage efficiency and accuracy on some pre-miRNAs [82, 83] and rescuing RNA interference (RNAi) functions [84]. Many studies have been conducted to investigate the structure of the Dicer protein, in species ranging from flies to humans, to elucidate the cleavage mechanism. As more and more structural and biochemical studies are being completed on the human Dicer (hDicer) protein, how hDicer cleaves pre-miRNAs is gradually being revealed.

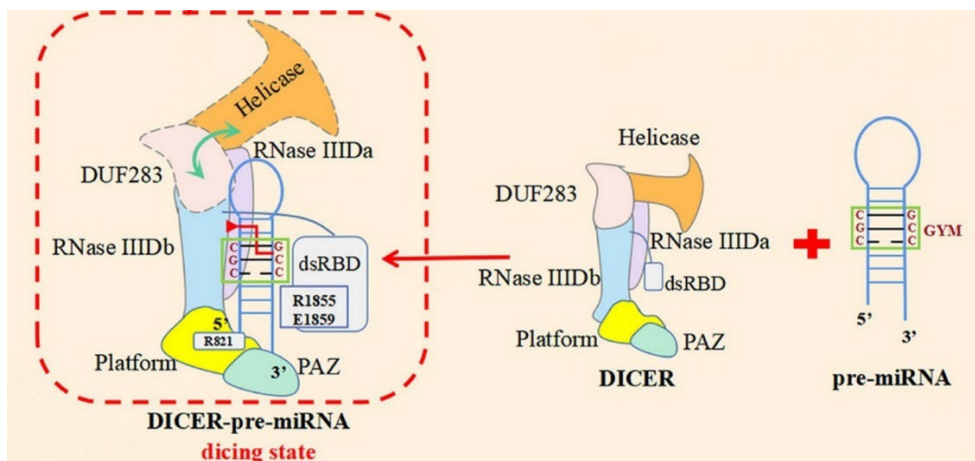
Human Dicer includes a helicase domain, a DUF283 domain, a dsRNA-binding domain, a PAZ domain, a platform domain and two RNase III domains (**Fig. 1.3, 1.4**) [85]. A number of models have been put forth which define the “rules” by which Dicer processes pre-miRNAs. The first model is called the ‘end counting rule’ [86, 87]. This model proposed that hDicer binds the pre-miRNA 3'-end 2-nt overhang in the PAZ domain and 5'-end phosphate group in the platform domain. The hDicer strictly follows the ruler-like counting mechanism from 3'-end or 5'-end, which two different binding pockets recognizes two ends individually [86, 87]. In this ruler-like counting mechanism, the distance between the 5'-end binding pocket and processing center of the RNase IIIb domain and the distance between the 3'-end binding pocket and processing center of

the RNase IIIa are  $\sim 58$  Å, which is equivalent to the length of  $\sim 22$  bps. These inherent structure features ensures hDicer cleaves both strands to generate products that are  $\sim 22$  nt long when counting from either the 5' or 3'-end of a pre-miRNA [88, 89]. Later, a new model called the 'loop counting rule' was proposed [91]. In this model, the helicase domain recognizes the single stranded region of the pre-miRNAs, either within the apical loop or within an internal bulge, to determine the cleavage site 2 nt upstream of ssRNA region [91, 92]. However, there is no dominant rules in hDicer cleavage, which was revealed by recent *in vivo* RNA chemical probing assays [92].

More and more structural features of the hDicer protein are being uncovered with the advancement of structural biology methods, namely cryoEM (**Fig. 1.3**). The first intact structure of hDicer was determined in 2018 in both the apo state and a 'pre-dicing' or pre-cleavage state[88]. More recently, the Kim group determined the structure of hDicer in an active cleavage conformation, adding to our understanding of hDicer structure and function [89].

	Apo state	Pre-cleavage	Cleavage-competent	Post-cleavage	Strand Release
Drosophila Dicer-1		Structurally Uncharacterized			Structurally Uncharacterized
PDB ID	8DGI	—	8DG5	8DGA	—
Human Dicer				Structurally Uncharacterized	Structurally Uncharacterized
PDB ID	5ZAK	5ZAL	7XW2	—	—

**Figure 1.3. Mechanistic insights into Drosophila Dicer-1 and human Dicer.** The RNA substrates are colored blue and proteins are colored gray. Conformations yet to be unknown are shown as 'Structurally Uncharacterized'. The figure templates are adopted from Torrez et. al [90].



**Figure 1.4. Dicer specifically recognizes and cleaves pre-miRNAs [85].** (Figure adopted from Weng et. al., ‘New discoveries on how DICER efficiently processes pre-miRNA’, [CC BY-NC 4.0 Deed | Attribution-NonCommercial 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/). The figure source data is from: <https://doi.org/10.1002/mco2.430>) Some specific features of pre-miRNAs are recognized by Dicer protein, including but not limited to 3'-end 2-nt overhang, 5'-end phosphate group and ‘GYM’ motif. The recognition of these motifs helps Dicer accurately and efficiently process pre-miRNAs [86, 93, 94].

In the ‘pre-dicing’ state, hDicer binds a pre-miRNA and anchors the 3'-end within the 3'-end binding pocket. In this conformation, the hDicer protein stabilizes the stem region of the pre-miRNA, but the RNase III domain of hDicer is not positioned near the dicing site within the bound pre-miRNA [85]. In 2023, the cleavage-competent structure of hDicer was determined by Lee et. al [89]. Relative to the pre-dicing state of hDicer, in the cleavage competent structure, the pre-miRNA substrate is positioned closer to the platform and RNase III domains and additional conformational changes were observed in the helicase, DUF293, dsRBD, and PAZ domains [88, 89]. In the dicing state, the DUF283 domain and helicase domain become more flexible relative to the apo hDicer structure. In addition, the C-terminus of the dsRBD moves away from the cleavage site to avoid the steric clashes between the dsRBDs and the pre-miRNA. The A-helical structure of the pre-miRNA substrate is distorted so that it fits into the cleavage site. Meanwhile, in the dicing state of hDicer, two  $\alpha$ -helices within the RNase III domains make electrostatic interactions with the upper stem region of the pre-miRNA to promote dsRNA recognition (**Fig 1.4**) [89].

Additional sequence features promote Dicer cleavage efficiency and accuracy. For example, the ‘GYM’ (paired G, paired pyrimidine (Y), and unpaired C or A (M)) motif is a sequence motif that is commonly found near the dicing site on pre-miRNAs [93]. hDicer uses residues R1855 and E1859 inside the dsRBD to identify the mismatch within the ‘GYM’ motif, which assists in cleavage site determination (**Fig 1.4**) [89].

Structural studies on the homologous *Drosophila* Dicer-1 (dmDicer-1) provide additional mechanistic insight [95]. In a recent dmDicer-1 structural study that used an artificial and highly stabilized pre-let-7 RNA substrate, Karina et. al captured three distinct structural states; an apo state, a cleavage competent state, and a post-cleavage state. They found that in dmDicer-1, the platform-PAZ domain is dynamic when not bound to a pre-miRNA substrate. Additionally, they observed minimal conformational changes in helicase domain of dmDicer-1 when comparing the cleavage competent and post-cleavage states.

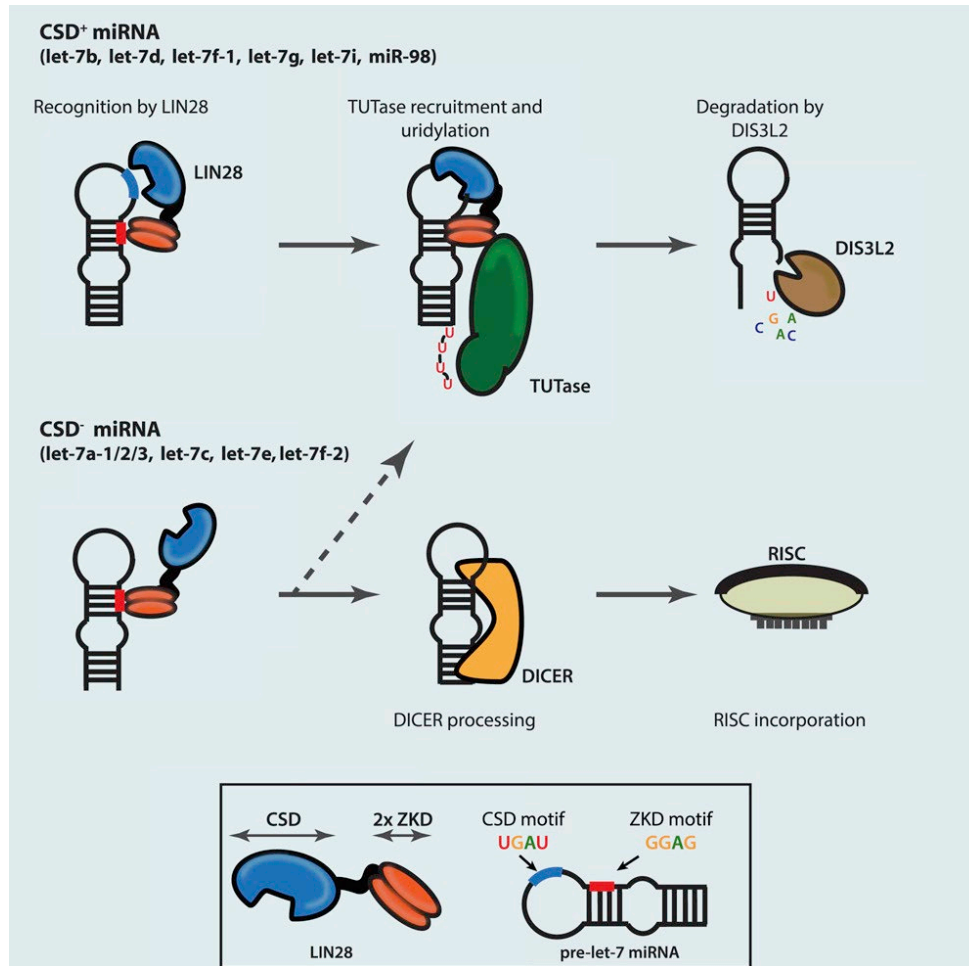
Although the cleavage state of hDicer has been structurally resolved, other conformational states within the enzymatic cycle remain uncharacterized (**Fig. 1.3**). More structural studies are needed to reveal how the helicase helps to recognize the pre-miRNA and how hDicer releases the mature miRNA duplex.

hDicer plays a crucial role in miRNA maturation. Therefore, it is important to identify and characterize the mechanisms by which hDicer is post-transcriptionally regulated. Both protein binding partners and RNA structure are important regulators for pre-miRNA processing by hDicer [41, 96-98]. Lin28 is a protein that regulates pre-miRNAs maturation [99-101]. Lin28 is composed of one cold shock domain (CSD) and two zinc knuckle domains (ZKD), which can bind UGAU and GGAG motifs, respectively (**Fig. 1.5**). The let-7 family is composed of 12 different miRNAs, and these 12 miRNAs can be divided into two sub-groups. One sub-group comprises those let-7

pre-miRNAs that contain a CSD binding motif (CSD<sup>+</sup>). The other sub-group is the CSD<sup>-</sup> group, comprised of let-7 pre-miRNAs that lack the UGAU CSD recognition motif (**Fig. 1.5**). The CSD<sup>+</sup> pre-let-7 RNAs contain two binding motifs, which ensures strong binding by Lin28. The binding of Lin28 will recruit terminal uridine transferase (TUTase) to add a poly-U tail at the 3'-ends of the RNAs. The polyuridylated RNAs are recognized and degraded by DIS3 like 3'-5' Exoribonuclease 2 (DIS3L2) exonuclease [101]. The weak binding between LIN28 and CSD<sup>-</sup> pre-let-7 RNAs will not recruit TUTase to trigger further degradation. Therefore, CSD<sup>-</sup> pre-let-7 RNAs can be cleaved by Dicer protein and further form RISC to regulate target mRNAs [99-101].

Other protein regulators were found to regulate pre-miRNAs maturation. For example, Interleukin Enhancer Binding Factor 3 (ILF3) was found to play an important role in pre-miR-144 maturation by a proposed model that ILF3 protein can reshape the apical loop region of pre-miR-144 to enhance Dicer cleavage efficiency on pre-miR-144 [102].

An additional layer of regulation can be found within the pre-miRNA structure. The pH-mediated structural change in pre-miR-21 reveals that pre-miR-21 can shift the stem region of pre-miR-21 to promote its cleavage by Dicer [97]. Other studies also identified a different conformational change in pre-miR-21 and demonstrated that these two structures were processed by Dicer differently. Based on these findings, Shortridge et al. found a small molecule that can stabilize the poor substrate conformation to inhibit mature miR-21 production *in cells* [98, 103]. Recent work from our lab (Chapter III) identifies the stability of the pre-miR-31 junction region as a regulatory element for control of miR-31 maturation [96]. The GYM motif is also one of the examples that shows the importance of context in pre-miRNA, which is detailed in the previous paragraph [93].

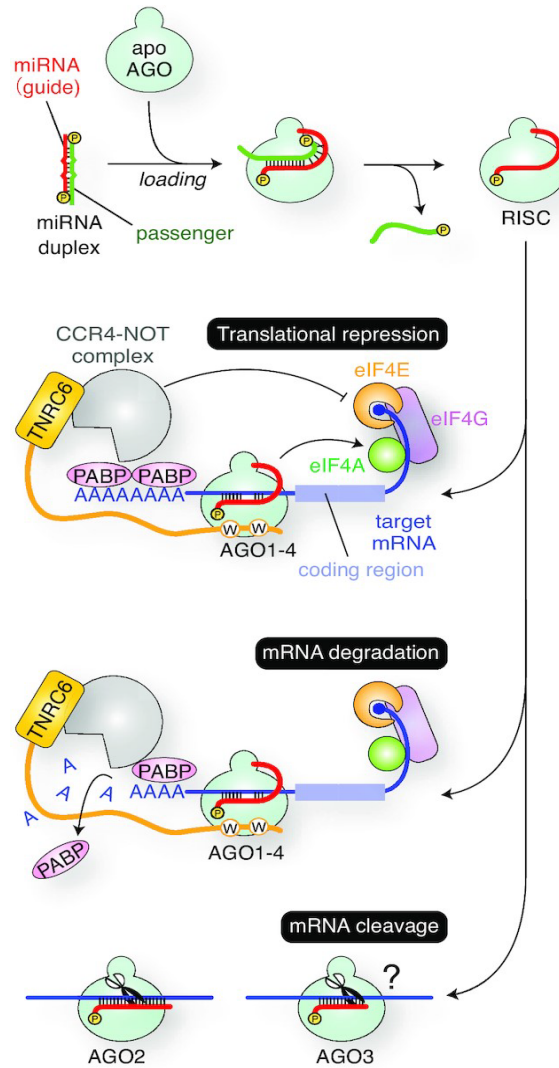


**Figure 1.5. Protein regulation pathway for let-7 family pre-miRNAs [101].** (Figure from Ustianenko et. al., ‘LIN28 Selectively Modulates a Subclass of Let-7 MicroRNAs’, [CC BY-NC-ND 4.0 DEED Attribution-NonCommercial-NoDerivs 4.0 International](https://doi.org/10.1016/j.molcel.2018.06.029). Source: <https://doi.org/10.1016/j.molcel.2018.06.029>) Let-7 family pre-miRNAs that contain a cold shock motif (UGAU) in their apical loops can be bound by the Lin28 protein. Lin28 recruits TUTase to polyuridylylate the 3'-end of the pre-miRNA. The polyuridylylated let-7 pre-miRNA is recognized and degraded by the DIS3L2 protein. Other let-7 pre-miRNA that lack the cold shock motif are not bound strongly by Lin28 and will undergo normal Dicer cleavage, generating a mature miRNA duplex and ultimately a functional RISC.

### 1.2.5 RISC assembly and regulation

After cleavage by Dicer/TRBP, the mature miRNA duplexes are loaded into the AGO protein to form the RISC (**Fig. 1.1, Fig. 1.6**). The AGO protein can be classified into four different stages: apo (no substrate), pre-RISC (with miRNA duplex), RISC (with guide strand of miRNA duplex) and target complex (with guide strand and target mRNA). The strand selection by the

AGO protein is known as the asymmetry rule. This rule establishes that the strand with the less thermodynamically stable 5'-end (A or U end) is selected as the guide strand and the other strand is ejected by the AGO protein [104, 105]. The MID domain of the AGO protein recognizes the 5' monophosphate and base moiety at position 1 of guide strand, further supporting guide strand selection [106, 107]. Adenosine monophosphate (AMP) and uridine monophosphate (UMP) at position 1 bind to the MID domain with a 10-fold higher affinity compared to cytidine and guanosine monophosphates (CMP and GMP, respectively) [106, 107]. Structural studies of the MID domain with AMP and UMP revealed that the protein main-chain carbonyl and amino groups are organized to form hydrogen bonds with uracil and adenine while the structure of the MID domain in complex with GMP and CMP revealed the base moieties were completely disordered [104, 107, 108]. TNRC6 proteins are essential proteins for RISCs to inhibit protein translation and degrade mRNA. TNRC6 proteins preferentially bind mature RISCs over apo AGO proteins [109, 110]. This preference ensures that TNRC6 proteins only bind RISCs that are ready to participate in gene silencing steps. The N-terminal Ago-binding domain of the TNRC6 protein binds RISC while the C-terminal region interacts with poly(A)-binding protein (PABP) (**Fig. 1.6**) [111-113]. The TNRC6 C-terminal domains also recruit Carbon Catabolite Repression-Negative On TATA-less (CCR4-NOT) complex, which is responsible for cap-dependent translation repression and mRNA decay acceleration (**Fig. 1.6**) [105, 114-116].



**Figure 1.6. RISC assembly and function pathway [105].** (Figure from Kotaro Nakanishi, ‘Anatomy of four human Argonaute proteins’, [CC BY-NC-ND 4.0 DEED Attribution-NonCommercial-NoDerivs 4.0 International](https://doi.org/10.1093/nar/gkac519). Source: <https://doi.org/10.1093/nar/gkac519>) AGO protein loads miRNA duplex, selects the guide strand (red) as template, and ejects the passenger strand (green) to form RISC. Within a RISC complex, all AGO proteins function to repress translation or degrade target mRNA. RISC complexes containing AGO2 or AGO3 proteins can slice target mRNA, though the mechanism by which AGO3-RISC slices mRNA is unknown.

RISC can also function to slice target mRNAs. Among the four different human AGOs, AGO2 and AGO3 can function as scissors to slice the target mRNA. The slicing activity of AGO2 proteins is activated only when the guide strand sequence is fully complementary to the mRNA sequence [117, 118]. AGO3 also can be catalytically activated by some miRNAs, such as miR-20a,



however, the guide strand sequence requirement for AGO3 protein remains unknown [105, 119].

Posttranslational modifications can modulate AGO protein function and lifetime. Type I collagen prolyl 4-hydroxylase (4PH) can hydroxylate proline residues in human AGO2, increasing AGO2 protein stability [120]. Phosphorylation at Ser387 position of AGO2 by the RAC $\gamma$  Serine/Threonine protein kinase (AKT3) enhances translational repression and decreases cleavage [121]. Ubiquitylation of AGO has been shown to promote degradation by the proteasome under multiple conditions [42, 122-125].

### **1.3 miRNA-31 in cancers**

MicroRNA-31 (miR-31) is one of the miRNAs that is known to be abnormally expressed in different diseases, including cancers, autoimmune diseases, and allergies [126]. MiR-31 is a highly conserved microRNA in metazoan species and is involved in various biological processes, which includes myogenesis, bone formation, and fertility [126]. Therefore, the dysregulation of miR-31 leads to cell dysregulation and disease. Research on miR-31 shows that in different types of tumors, the miR-31 expression level is altered and its function is context dependent [127].

Interestingly, miR-31 can play dual roles in tumorigenesis. When miR-31 is downregulated, it often functions as a tumor suppressor and when it is overexpressed, cancer cell proliferation and invasion are promoted. MiR-31 is found to play tumor suppressive role in several cancers, including ovarian cancer, prostate cancer, and breast cancer [127].

In serous ovarian cancers, miR-31 is downregulated and involved in suppressing E2F Transcription Factor 2 (E2F2), Serine threonine kinase 40 (STK40), and tyrosine-protein kinase MET expression [128-130]. The E2F pathway is an important cell cycle progression regulatory pathway, which participates in the G1/S transition in the cell cycle. The p53 pathway is a canonical pathway for cancer repression, using pocket protein retinoblastoma protein (RB) to bind E2F2,

one of the essential transcription factors in the E2F pathway, to mask the transcription activation [129]. miR-31 can downregulate E2F2 to block the cell cycle especially for serous ovarian cancer cells that lack a functional p53 pathway [127]. STK40 is a p53 mediated transcription repressor, therefore, by downregulating STK40, miR-31 can partially restore the p53 pathway [127]. Furthermore, miR-31 can counteract the chemical resistance in ovarian cells by inhibiting translation of MET, a receptor tyrosine kinase for hepatocyte growth factor. Taxanes-resistant ovarian cancer cells express low miR-31 and restoration of miR-31 can reduce MET protein and increase cell sensitivity [127, 128].

In prostate cancer, androgen receptor (AR) signaling is a crucial way for prostate cancer to grow and mutations within the AR is the main cause of drug resistance [127]. MiR-31 can inhibit expression of AR by binding to the coding sequence of mRNA rather than to its 3' UTR. AR can also inhibit miR-31 expression by binding and hypermethylating the miR-31 promoter sequence to reduce expression of the transcript [131]. Thus, the occurrence of prostate cancer may be linked to the hyper-methylation of miR-31's promoter which reduces miR-31 levels which in turn upregulates AR expression, generating a positive feedback loop that ultimately promotes prostate cancer progression.[127].

Tripe negative breast cancer (TNBC) is a subtype of breast cancer that is extremely aggressive, resistant to drugs, and has a particularly poor prognosis. In the TNBC tissue and cell lines, miR-31 expression levels are downregulated [132]. In the cell lines, miR-31 targets protein kinase C epsilon (PKC $\epsilon$ ) in the NF- $\kappa$ B pathway. MiR-31 binds to three binding sites within the 3' UTR of the PRKCE mRNA to downregulate the translation of PKC $\epsilon$  without decreasing mRNA level [133]. This leads to a deficiency in p65, followed by a decrease in PKC $\epsilon$ , which results in the elimination of the anti-apoptotic BCL2 protein. Consequently, miR-31 can indirectly regulate

BCL2 to initiate apoptosis and increase sensitivity to chemotherapy and radiation in TNBC [127].

MiR-31 also functions as an oncogene in some type of cancers, including colorectal cancer, non-small-cell lung carcinomas (NSCLC) and Esophageal squamous cell carcinoma (ESCC) [126, 127]. The overexpression of miR-31 in these cancers promotes tumorigenesis and migration.

In colorectal cancer cell lines and tissues, miR-31 promotes cancer cell invasion, migration and proliferation [134]. Special AT-rich Sequence-binding protein 2 (SATB2) is an essential tumor-suppressive protein in cancer cells. SATB2 can inhibit the phosphorylation of Extracellular Signal-regulated Kinase 5 (ERK5) to inhibit tumorigenicity. ERK5 is activated by the phosphorylation of threonine and tyrosine residues, which are in the kinase domain of ERK5. After activation, ERK5 can auto phosphorylate itself and activate its substrate, including c-Myc, an oncogene. By inhibiting ERK5 phosphorylation, SATB2 can downregulate oncogenic c-Myc expression. However, miR-31 targets SATB2 mRNA and downregulates both mRNA levels and protein levels of SATB2, which reactive the MEK/ERK5 signaling [132, 135, 136]. The RAS pathway plays a crucial role in cell proliferation, differentiation and gene expression [137]. A constitutively active RAS pathway can push the cell into a cancer-inducing state [138]. RASA1 regulates RAS function by inactivating the GTP-bound RAS. As RASA1 protein level is downregulated by the overexpressed miR-31, the dysregulated RAS pathway promotes colorectal cancer cell proliferation [139].

The hippo pathway is a key regulator for controlling organ size and cell proliferation, and disruption in this pathway can lead to tumorigenesis. Large tumor suppressor kinase 2 (LATS2) expression is under post-transcriptional control by miR-31. In ESCC tissues, upregulated miR-31 downregulates LATS2 expression and disrupts hippo pathway. This may account for the ESCC tumorigenesis [140].

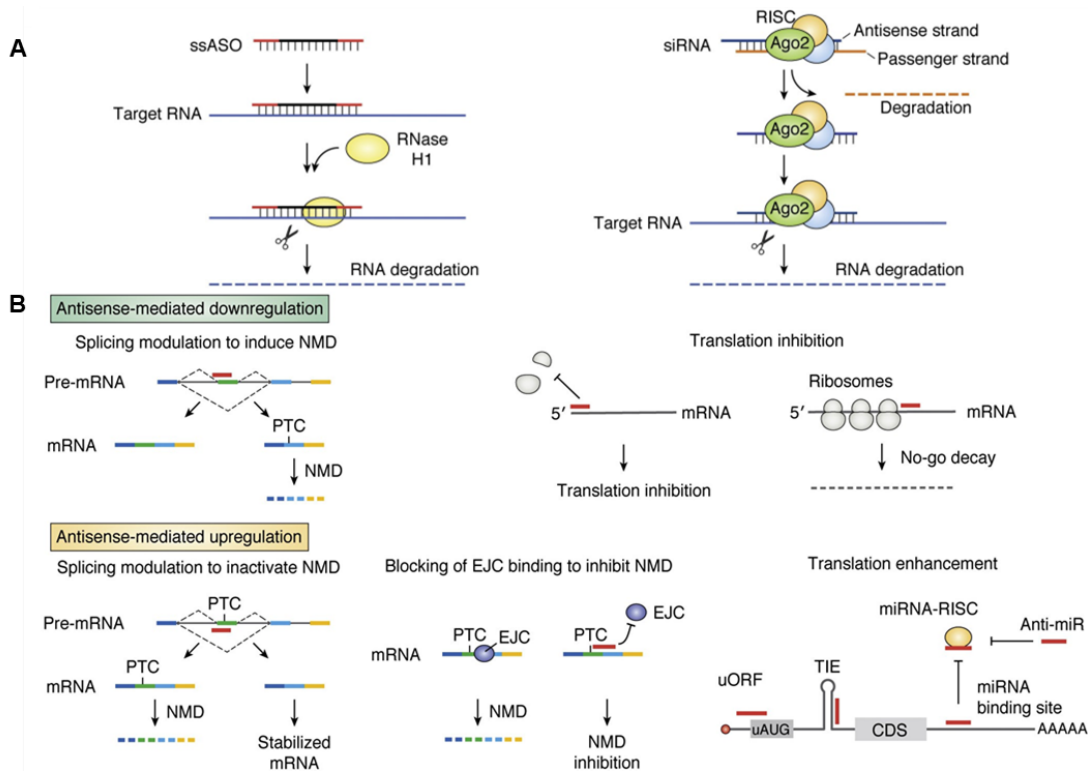
Adenocarcinoma (ADC) is the most common subtype of NSCLC. Abnormal expression of miR-31 has been identified in most stages of ADC, from its formation to its spread [141]. The expression level of miR-31 is higher in ADC than healthy lungs and as the disease progresses, especially during metastasis. MiR-31 promotes cell growth, and its levels rise with tumor size. Two mechanisms explain miR-31's role in promoting cell growth. Firstly, miR-31 interacts with the mutant KRAS, a common driver mutation in NSCLC, facilitating the transformation and growth of lung epithelial cells. Secondly, miR-31 targets six negative regulators of the RAS/MAPK signaling pathway, accelerating ERK activation and promoting cell proliferation. Experiments have confirmed that miR-31 can induce tumorigenesis and invasion in a manner dependent on ERK1/2 signaling. For instance, in smooth muscle cells, miR-31 targets LATS2, which is under the control of ERK1/2 signaling. This suggests that the ERK/miR-31/LATS2 axis may be common. Activated ERK1/2 can increase miR-31 expression, which can then downregulate LATS2 expression [127, 142, 143].

In summary, miR-31 plays a key role in regulating normal cell functions. The dysregulation of miR-31 leads to aberration of different signaling pathways, ultimately leading to tumorigenesis. For these reasons, the miR-31 expression level should be tightly controlled. It is therefore important to have a detailed understanding of the regulatory mechanisms by which miR-31 is produced in the cell.

#### **1.4 Antisense oligo application and development**

The 'antisense' concept was introduced for the first time in 1978 and the idea involved designing oligonucleotides complementary to target RNA sequences to alter the behavior of the target RNAs [144, 145]. Antisense oligos (ASOs) are divided into two types based on their form;

one is single stranded ASOs (ssASO), and the other is double strand ASOs (dsASO), which refer to small interfere RNAs (siRNAs) [146, 147]. You can also group ASOs based on their functional mechanism; ASOs can function as RNA degraders or use an occupancy-only (or steric blocking) mechanism [148].



**Figure 1.7. ASOs use different mechanisms to modulate gene expression [144].** (Figure adopted from Crooke, Stanley T., et al. ‘Antisense technology: A review’, [CC BY-NC-ND 4.0 DEED Attribution-NonCommercial-NoDerivs 4.0 International](https://doi.org/10.1016/j.jbc.2021.100416). The figure source data is from: <https://doi.org/10.1016/j.jbc.2021.100416>) **A.** Commonly used ASO-mediated RNA degradation mechanism, which includes RNase H1-mediated degradation (left) and AGO2-mediated degradation (right). **B.** ASOs can use an occupancy-only mechanism to affect gene expression. These mechanisms include the downregulation of gene expression by inducing non-sense-mediated decay (NMD) through changes in splicing (upper left), and the inhibition of translation initiation by binding to the cap region or triggering mRNA no-go decay by binding to the coding region of mRNAs (upper right panels). ASOs can also enhance expression levels by changing the splicing site to skip a exon which contains a premature termination codon (PTC) (lower left panel), or by preventing NMD through the inhibition of exon–exon junction complex (EJC) binding (lower middle panel). By masking translation inhibitory elements, including upstream open reading frame (uORF), translation inhibitory element (TIE), miRNA-binding sites or even miRNAs themselves (lower right panel) can also boost the protein expression.

For functionality in the cell, ASOs must be chemically modified because ASOs containing a native phosphodiester backbone are fragile and are quickly degraded by nucleases and the binding affinity to target sequences is relatively low [144]. To overcome these two limitations, in the past decades, modifications have been introduced at every site on oligonucleotides to evaluate the effectiveness of modification on stability or binding affinity *in vitro* or *in vivo* [144, 147, 149]. These modifications include conjugating other groups at 5' or 3' termini to enhance tissue targeting or cellular uptake [147]. The modification of sugar rings, such as replacing a 2'-hydroxyl with a 2'-methoxyethyl, can enhance binding affinity to RNA, increase ASO stability and decrease pro-inflammatory properties [147]. Backbone modifications, including replacing the natural phosphodiester backbone with a phosphorothioate backbone, can enhance stability against nucleases and modulate protein binding [147]. The base modification, such as 5-methylcytosine, leads to a decrease in pro-inflammatory properties and slightly enhances the binding affinity for the target RNA [144, 148].

ASOs designed to degrade target RNAs mainly use one of two different systems; the RNase H1 mediated system or the AGO2 mediated system [149] (**Fig. 1.7 A**). RNase H1 is a wide-spread endogenous RNase inside mammalian cells, which can be found in the nucleus, cytoplasm, and mitochondria [150]. RNase H1 recognizes RNA-DNA heteroduplexes and degrades the target RNA (**Fig. 1.7 A**). RNase H1 has special requirements for ASO design. As RNase H1 recognizes RNA-DNA duplexes for cleavage, a fully modified ASO will not be recognized as DNA in the ASO-duplex. Therefore, RNase H1 based ASO design requires a minimum of five consecutive DNA bases for RNA cleavage activation, however, seven-ten DNA nucleotides is optimal for cleavage [149]. siRNAs go through a similar biogenesis pathway as miRNAs, described in **section 1.2**, however, because of their design, siRNAs normally only require Dicer/TRBP for cleavage

and AGO2 for final function [149]. siRNAs are transfected into cells as duplex RNAs or as modified short hairpin RNAs (shRNAs). The duplex RNA can directly bind to the AGO2 protein, while shRNAs need Dicer/TRBP cleavage to form the siRNA duplex that can be loaded into the AGO2 protein. After loading into AGO2 protein, the passenger strand of the duplex is released from the complex and degraded in the cytoplasm [151, 152]. Once AGO2 loads the guide strand, the RISC will bind to the 3' UTR of the target messenger RNA (mRNA) and cleave the mRNA, as occurs in miRNA biogenesis [153, 154] (**Fig. 1.7 A**).

Occupancy-only mechanisms are more diverse than degradation mechanisms. Occupancy-only mechanisms can bidirectionally regulate the target expression levels, while degradation mechanisms can only downregulate RNA levels [144] (**Fig. 1.7 B**).

Non-sense-mediated decay (NMD) is a conserved surveillance pathway in all eukaryotes [155]. This pathway degrades mRNAs that contain premature stop codons. In some cases, the translation of these mRNA could result in production of deleterious overactive or suppressive proteins [156]. The human NMD pathway employs exon–exon junction complex (EJCs) to recognize and degrade aberrant mRNA during the translation steps [156, 157]. In this pathway, premature termination codon (PTC) is an essential recognition feature for triggering degradation. Therefore, by utilizing steric blocking ASOs, pre-mRNA splicing can be modulated to introduce premature codons to trigger the NMD pathway [158]. In this way, manipulation of the splicing site by ASOs will lead to degradation of the target mRNA. ASOs can also inactivate the NMD pathway by modulating pre-mRNA splicing to skip exons that contain PTC and produce a new stabilized mRNA without any PTCs (**Fig. 1.7 B**). EJCs are essential for NMD pathway, therefore, using ASOs to block the EJC binding site is another way to upregulate PTC-contained mRNA levels [144] (**Fig. 1.7 B**).

Direct targeting of mRNA elements or post-transcriptional regulators is another way to employ an occupancy-only mechanism for ASOs. Protein translation can be inhibited by an ASO masking the 5' cap of target mRNA. Alternatively, ASO binding within the coding sequence can cause ribosomes to stall, which triggers mRNA no-go decay [144] (**Fig. 1.7 B**). Sequence and/or structural features near the 5' end of the mRNA can inhibit translation of the downstream gene. For example, the upstream open reading frame is a sequence found upstream of the proper start codon which can be recognized by the ribosome and modulates the translation efficiency of the main coding region. Additionally, translation inhibition elements are stem loop structures found in 5' UTR of mRNA that prevent canonical cap-dependent translation. Targeting of these elements with ASOs can serve to enhance translation by disrupting the repressive elements [144, 148, 159-161]. As mentioned above, RISC can modulate mRNA translation and stability. Thus, ASOs disrupt RISC function by blocking RICS binding to the mRNA or by sequestering the RISC. In these two ways, the base pair interaction between mRNA and RISC is masked and protein translation is enhanced [144].

ASOs have applications outside of the research lab, many ASOs have been approved by the Food and Drug Administration for clinical application in the US. Fomivirsen was the first approved ASO therapeutic and is used for the treatment of cytomegalovirus retinitis (CMV). Fomivirsen binds the mRNA of the IE2 viral protein to recruit RNase H1 for mRNA cleavage, which therefore, stops viral replication [162]. Other approved ASO that use the same RNase H1 mechanism include Mipomersen and Inotersen. Mipomersen targets Apolipoprotein B (Apo B) mRNA, an important therapeutic target for familial hypercholesterolemia treatment [163]. Meanwhile, Inotersen is used for treating polyneuropathy of amyloidosis by targeting pathogenic mRNA encoding the Transthyretin (TTR) protein [164, 165]. Duchenne muscular dystrophy



(DMD) is a severe neuromuscular disease. Symptoms of DMD begin with muscle weakness and finally will result in trouble standing up. DMD is caused by the mutation in the dystrophin gene. Therefore, all the approved ASOs that are used for DMD treatment direct alternate splicing of the dystrophin mRNA to skip the mutation and lead to the production of partially functional proteins. Four ASOs were approved for DMD treatment, including Eteplirsen, Golodirsen, Vitolarsen and Casimersen. Golodirsen and Vitolarsen help pre-mRNA to skip the dystrophin exon 53, while Eteplirsen and Casimersen assist the skipping of exon 51 and 45 respectively [162]. Nusinersen was designed for exon including for the treatment of spinal muscular atrophy, a rare neuromuscular disorder [166]. Nusinersen induces the splicing of survival of motor neuron 2 (SMN2) pre-mRNA to include exon 7. The new splicing SMN2 mRNA can be translated into a functional SMN protein, which helps spinal muscular atrophy treatment [162, 166]. As more and more research are conducting on ASO development, more than 100 ASOs have been enrolled in clinical trials and some of them are close to be approved [147]. The annual market sale of ASO drugs is expected to increase to \$1.62 billion by 2025 from approximately \$1 billion sales in 2021 [162].

This thesis aims to expand our understanding of how pre-miRNA structural features function as key regulators for processing by Dicer/TRBP and how these structures can be targeted for inhibition by ASOs. In Chapter II, I characterized interesting dynamic features of single nucleotide mismatches within the pre-miR-31 stem region. In Chapter III, a high-resolution 3D structure of pre-miR-31 was solved, and I determined how different structural features contribute to the regulation of pre-miR-31 maturation. In Chapter IV, I developed a new type of antisense oligo that targets junction-containing pre-miRNA for Dicer/TRBP cleavage inhibition. Furthermore, I revealed the potential for these ASOs to modulate processing of junction-containing pre-miRNAs *in cell*. Finally, I summarize the overall findings of my thesis work and present

promising future directions for both pre-miR-31 dynamics studies and ASO development.

## 1.5 Reference

1. Lee, R.C., R.L. Feinbaum, and V. Ambros, The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *cell*, 1993. **75**(5): p. 843-854.
2. Bartel, D.P., MicroRNAs: genomics, biogenesis, mechanism, and function. *cell*, 2004. **116**(2): p. 281-297.
3. Almeida, M.I., R.M. Reis, and G.A. Calin, MicroRNA history: discovery, recent applications, and next frontiers. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 2011. **717**(1-2): p. 1-8.
4. Wightman, B., et al., Negative regulatory sequences in the *lin-14* 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. *Genes & development*, 1991. **5**(10): p. 1813-1824.
5. Wightman, B., I. Ha, and G. Ruvkun, Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, 1993. **75**(5): p. 855-862.
6. Reinhart, B.J., et al., The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *nature*, 2000. **403**(6772): p. 901-906.
7. Vella, M.C., et al., The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3' UTR. *Genes & development*, 2004. **18**(2): p. 132-137.
8. Vella, M.C. and F.J. Slack, *C. elegans* microRNAs. *WormBook: The Online Review of C. elegans Biology* [Internet], 2005.
9. Pasquinelli, A.E., et al., Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature*, 2000. **408**(6808): p. 86-89.
10. Abbott, A.L., et al., The *let-7* MicroRNA family members *mir-48*, *mir-84*, and *mir-241*

function together to regulate developmental timing in *Caenorhabditis elegans*. *Developmental cell*, 2005. **9**(3): p. 403-414.

11. Lau, N.C., et al., An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 2001. **294**(5543): p. 858-862.

12. Lim, L.P., et al., The microRNAs of *Caenorhabditis elegans*. *Genes & development*, 2003. **17**(8): p. 991-1008.

13. Kozomara, A. and S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, 2014. **42**(D1): p. D68-D73.

14. Kozomara, A., M. Birgaoanu, and S. Griffiths-Jones, miRBase: from microRNA sequences to function. *Nucleic acids research*, 2019. **47**(D1): p. D155-D162.

15. Kozomara, A. and S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research*, 2010. **39**(suppl\_1): p. D152-D157.

16. Griffiths-Jones, S., et al., miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research*, 2006. **34**(suppl\_1): p. D140-D144.

17. Griffiths-Jones, S., et al., miRBase: tools for microRNA genomics. *Nucleic acids research*, 2007. **36**(suppl\_1): p. D154-D158.

18. Calin, G.A., et al., 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*, 2002. **99**(24): p. 15524-15529.

19. Garzon, R., G.A. Calin, and C.M. Croce, MicroRNAs in cancer. *Annual review of medicine*, 2009. **60**: p. 167-179.

20. Schmittgen, T.D., et al., A high-throughput method to monitor the expression of microRNA precursors. *Nucleic acids research*, 2004. **32**(4): p. e43-e43.

21. Nelson, P.T., et al., Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods*, 2004. **1**(2): p. 155-61.
22. Willenbrock, H., et al., Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. *Rna*, 2009. **15**(11): p. 2028-34.
23. Li, C., et al., Therapeutic microRNA strategies in human cancer. *The AAPS journal*, 2009. **11**: p. 747-757.
24. Acunzo, M., et al., MicroRNA and cancer—a brief overview. *Advances in biological regulation*, 2015. **57**: p. 1-9.
25. Zhao, Y., E. Samal, and D. Srivastava, Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, 2005. **436**(7048): p. 214-220.
26. Kwon, C., et al., MicroRNA1 influences cardiac differentiation in *Drosophila* and regulates Notch signaling. *Proceedings of the National Academy of Sciences*, 2005. **102**(52): p. 18986-18991.
27. Chen, J.-F., et al., The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature genetics*, 2006. **38**(2): p. 228-233.
28. Van Rooij, E., et al., A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. *Proceedings of the National Academy of Sciences*, 2006. **103**(48): p. 18255-18260.
29. Edwards, J.K., et al., MicroRNAs and ultraconserved genes as diagnostic markers and therapeutic targets in cancer and cardiovascular diseases. *Journal of cardiovascular translational research*, 2010. **3**: p. 271-279.
30. Sonkoly, E., et al., MicroRNAs: novel regulators involved in the pathogenesis of psoriasis? *PloS one*, 2007. **2**(7): p. e610.

31. Croker, B.A., et al., SOCS3 negatively regulates IL-6 signaling in vivo. *Nature immunology*, 2003. **4**(6): p. 540-545.
32. Sano, S., et al., Stat3 links activated keratinocytes and immunocytes required for development of psoriasis in a novel transgenic mouse model. *Nature medicine*, 2005. **11**(1): p. 43-49.
33. Sivamani, R.K., et al., Biological therapy of psoriasis. *Indian journal of dermatology*, 2010. **55**(2): p. 161.
34. Li, L., X.P. Chen, and Y.J. Li, MicroRNA-146a and human disease. *Scandinavian journal of immunology*, 2010. **71**(4): p. 227-231.
35. Quinlan, S., et al., MicroRNAs in neurodegenerative diseases. *International review of cell and molecular biology*, 2017. **334**: p. 309-343.
36. Fiore, R., G. Siegel, and G. Schrott, MicroRNA function in neuronal development, plasticity and disease. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2008. **1779**(8): p. 471-478.
37. Lukiw, W.J., Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus. *Neuroreport*, 2007. **18**(3): p. 297-300.
38. Hugon, J. and C. Paquet, Targeting miRNAs in Alzheimer's disease. *Expert Review of Neurotherapeutics*, 2008. **8**(11): p. 1615-1616.
39. Kim, J., et al., A MicroRNA feedback circuit in midbrain dopamine neurons. *Science*, 2007. **317**(5842): p. 1220-1224.
40. Bofill-De Ros, X. and U.A. Vang Ørom, Recent progress in miRNA biogenesis and decay. *RNA biology*, 2024. **21**(1): p. 1-8.
41. Shang, R., et al., microRNAs in action: biogenesis, function and regulation. *Nature*

Reviews Genetics, 2023. **24**(12): p. 816-833.

42. Ha, M. and V.N. Kim, Regulation of microRNA biogenesis. Nature reviews Molecular cell biology, 2014. **15**(8): p. 509-524.

43. Leitão, A.L. and F.J. Enguita, A Structural View of miRNA Biogenesis and Function. Noncoding RNA, 2022. **8**(1).

44. Zeng, Y., R. Yi, and B.R. Cullen, Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. The EMBO journal, 2005. **24**(1): p. 138-148.

45. Gregory, R.I., et al., The Microprocessor complex mediates the genesis of microRNAs. Nature, 2004. **432**(7014): p. 235-240.

46. Han, J., et al., The Drosha-DGCR8 complex in primary microRNA processing. Genes & development, 2004. **18**(24): p. 3016-3027.

47. Denli, A.M., et al., Processing of primary microRNAs by the Microprocessor complex. Nature, 2004. **432**(7014): p. 231-235.

48. Bohnsack, M.T., K. Czaplinski, and D. Görlich, Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. Rna, 2004. **10**(2): p. 185-191.

49. Lund, E., et al., Nuclear export of microRNA precursors. science, 2004. **303**(5654): p. 95-98.

50. Yi, R., et al., Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes & development, 2003. **17**(24): p. 3011-3016.

51. Hutvagner, G., et al., A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science, 2001. **293**(5531): p. 834-838.

52. Ketting, R.F., et al., Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & development*, 2001. **15**(20): p. 2654-2659.
53. Grishok, A., et al., Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell*, 2001. **106**(1): p. 23-34.
54. Bernstein, E., et al., Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 2001. **409**(6818): p. 363-366.
55. Knight, S.W. and B.L. Bass, A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, 2001. **293**(5538): p. 2269-2271.
56. Medley, J.C., G. Panzade, and A.Y. Zinovyeva, microRNA strand selection: Unwinding the rules. *Wiley Interdisciplinary Reviews: RNA*, 2021. **12**(3): p. e1627.
57. Czech, B. and G.J. Hannon, Small RNA sorting: matchmaking for Argonautes. *Nature Reviews Genetics*, 2011. **12**(1): p. 19-31.
58. Müller, M., F. Fazi, and C. Ciaudo, Argonaute proteins: from structure to function in development and pathological cell fate determination. *Frontiers in Cell and Developmental Biology*, 2020. **7**: p. 360.
59. Schirle, N.T., J. Sheu-Gruttadauria, and I.J. MacRae, Structural basis for microRNA targeting. *Science*, 2014. **346**(6209): p. 608-613.
60. Kwon, S.C., et al., Structure of Human DROSHA. *Cell*, 2016. **164**(1): p. 81-90.
61. Partin, A.C., et al., Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. *Molecular Cell*, 2020. **78**(3): p. 411-422.e4.
62. Auyeung, V.C., et al., Beyond secondary structure: primary-sequence determinants license



pri-miRNA hairpins for processing. *Cell*, 2013. **152**(4): p. 844-858.

63. Nguyen, T.A., et al., Functional anatomy of the human microprocessor. *Cell*, 2015. **161**(6): p. 1374-1387.

64. Jin, W., et al., Structural basis for pri-miRNA recognition by Drosha. *Molecular cell*, 2020. **78**(3): p. 423-433. e5.

65. Dang, T.L., et al. Select amino acids in DGCR8 are essential for the UGU-pri-miRNA interaction and processing. *Communications biology*, 2020. **3**, 344 DOI: 10.1038/s42003-020-1071-5.

66. Shang, R. and E.C. Lai, Parameters of clustered suboptimal miRNA biogenesis. *Proceedings of the National Academy of Sciences*, 2023. **120**(41): p. e2306727120.

67. Kwon, S.C., et al., ERH facilitates microRNA maturation through the interaction with the N-terminus of DGCR8. *Nucleic acids research*, 2020. **48**(19): p. 11097-11112.

68. Fang, W. and D.P. Bartel, MicroRNA clustering assists processing of suboptimal microRNA hairpins through the action of the ERH protein. *Molecular cell*, 2020. **78**(2): p. 289-302. e6.

69. Shang, R., et al., Genomic clustering facilitates nuclear processing of suboptimal pri-miRNA loci. *Molecular cell*, 2020. **78**(2): p. 303-316. e4.

70. Hutter, K., et al., SAFB2 enables the processing of suboptimal stem-loop structures in clustered primary miRNA transcripts. *Molecular cell*, 2020. **78**(5): p. 876-889. e6.

71. Wang, J., et al., XPO5 promotes primary miRNA processing independently of RanGTP. *Nature communications*, 2020. **11**(1): p. 1845.

72. Guil, S. and J.F. Cáceres, The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature structural & molecular biology*, 2007. **14**(7): p. 591-596.

73. Wu, K., et al., The role of exportin-5 in microRNA biogenesis and cancer. *Genomics, proteomics & bioinformatics*, 2018. **16**(2): p. 120-126.
74. Leisegang, M.S., et al., Exportin T and Exportin 5: tRNA and miRNA biogenesis—and beyond. *Biological chemistry*, 2012. **393**(7): p. 599-604.
75. Macchi, P., et al., The brain-specific double-stranded RNA-binding protein Staufen2: nucleolar accumulation and isoform-specific exportin-5-dependent export. *Journal of Biological Chemistry*, 2004. **279**(30): p. 31440-31444.
76. Gwizdek, C., et al., Minihelix-containing RNAs mediate exportin-5-dependent nuclear export of the double-stranded RNA-binding protein ILF3. *Journal of Biological Chemistry*, 2004. **279**(2): p. 884-891.
77. Iwasaki, Y.W., et al., Global microRNA elevation by inducible Exportin 5 regulates cell cycle entry. *Rna*, 2013. **19**(4): p. 490-497.
78. Yi, R., et al., Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *Rna*, 2005. **11**(2): p. 220-226.
79. Yamazawa, R., et al., Structural basis for selective binding of export cargoes by exportin-5. *Structure*, 2018. **26**(10): p. 1393-1398. e2.
80. Okada, C., et al., A high-resolution structure of the pre-microRNA nuclear export machinery. *Science*, 2009. **326**(5957): p. 1275-1279.
81. Fareh, M., et al., TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments. *Nature communications*, 2016. **7**(1): p. 13694.
82. Chendrimada, T.P., et al., TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, 2005. **436**(7051): p. 740-744.
83. Chakravarthy, S., et al., Substrate-specific kinetics of Dicer-catalyzed RNA processing.

Journal of molecular biology, 2010. **404**(3): p. 392-402.

84. Daniels, S.M., et al., Characterization of the TRBP domain required for Dicer interaction and function in RNA interference. *BMC Molecular Biology*, 2009. **10**(1): p. 38.
85. Weng, Q., Q. Wu, and Q. Zheng, New discoveries on how DICER efficiently processes pre-miRNA. *MedComm*, 2023. **4**(6): p. e430.
86. Park, J.-E., et al., Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 2011. **475**(7355): p. 201-205.
87. MacRae, I.J., et al., Structural basis for double-stranded RNA processing by Dicer. *Science*, 2006. **311**(5758): p. 195-198.
88. Liu, Z., et al., Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate. *Cell*, 2018. **173**(5): p. 1191-1203. e12.
89. Lee, Y.-Y., et al., Structure of the human DICER–pre-miRNA complex in a dicing state. *Nature*, 2023. **615**(7951): p. 331-338.
90. Torrez, R.M., M.D. Ohi, and A.L. Garner, Structural Insights into the Advances and Mechanistic Understanding of Human Dicer. *Biochemistry*, 2023. **62**(1): p. 1-16.
91. Gu, S., et al., The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell*, 2012. **151**(4): p. 900-911.
92. Luo, Q.-J., et al., RNA structure probing reveals the structural basis of Dicer binding and cleavage. *Nature communications*, 2021. **12**(1): p. 1-12.
93. Lee, Y.-Y., H. Kim, and V.N. Kim, Sequence determinant of small RNA production by DICER. *Nature*, 2023. **615**(7951): p. 323-330.
94. Zhang, H., et al., Single processing center models for human Dicer and bacterial RNase III. *Cell*, 2004. **118**(1): p. 57-68.

95. Jouravleva, K., et al., Structural basis of microRNA biogenesis by Dicer-1 and its partner protein Loqs-PB. *Mol Cell*, 2022. **82**(21): p. 4049-4063 e6.
96. Ma, S., et al., Structure of pre-miR-31 reveals an active role in Dicer-TRBP complex processing. *Proceedings of the National Academy of Sciences*, 2023. **120**(39): p. e2300527120.
97. Baisden, J.T., et al., Visualizing a protonated RNA state that modulates microRNA-21 maturation. *Nature chemical biology*, 2021. **17**(1): p. 80-88.
98. Shortridge, M.D., et al., A slow dynamic RNA switch regulates processing of microRNA-21. *Journal of Molecular Biology*, 2022. **434**(16): p. 167694.
99. Balzeau, J., et al., The LIN28/let-7 pathway in cancer. *Frontiers in genetics*, 2017. **8**: p. 31.
100. Heo, I., et al., TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell*, 2009. **138**(4): p. 696-708.
101. Ustianenko, D., et al., LIN28 selectively modulates a subclass of let-7 microRNAs. *Molecular cell*, 2018. **71**(2): p. 271-283. e5.
102. Shang, R., et al., Regulated dicing of pre-mir-144 via reshaping of its terminal loop. *Nucleic acids research*, 2022. **50**(13): p. 7637-7654.
103. Shortridge, M.D., et al., Drug-like small molecules that inhibit expression of the oncogenic microRNA-21. *ACS Chemical Biology*, 2023. **18**(2): p. 237-250.
104. Nakanishi, K., Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *Wiley Interdisciplinary Reviews: RNA*, 2016. **7**(5): p. 637-660.
105. Nakanishi, K., Anatomy of four human Argonaute proteins. *Nucleic Acids Research*, 2022. **50**(12): p. 6618-6638.
106. Suzuki, H.I., et al., Small-RNA asymmetry is directly driven by mammalian Argonautes. *Nature structural & molecular biology*, 2015. **22**(7): p. 512-521.

107. Frank, F., N. Sonenberg, and B. Nagar, Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 2010. **465**(7299): p. 818-822.
108. Ghildiyal, M., et al., Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, 2008. **320**(5879): p. 1077-1081.
109. Elkayam, E., et al., Multivalent recruitment of human argonaute by GW182. *Molecular cell*, 2017. **67**(4): p. 646-658. e3.
110. Boland, A., et al., Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proceedings of the National Academy of Sciences*, 2011. **108**(26): p. 10466-10471.
111. Eulalio, A., E. Huntzinger, and E. Izaurralde, GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nature structural & molecular biology*, 2008. **15**(4): p. 346-353.
112. Pfaff, J., et al., Structural features of Argonaute–GW182 protein interactions. *Proceedings of the National Academy of Sciences*, 2013. **110**(40): p. E3770-E3779.
113. Sheu-Gruttadauria, J. and I.J. MacRae, Phase transitions in the assembly and function of human miRISC. *Cell*, 2018. **173**(4): p. 946-957. e16.
114. Braun, J.E., et al., GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Molecular cell*, 2011. **44**(1): p. 120-133.
115. Huntzinger, E. and E. Izaurralde, Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, 2011. **12**(2): p. 99-110.
116. Huntzinger, E., et al., The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. *Nucleic acids research*, 2013. **41**(2): p. 978-994.
117. Liu, J., et al., Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, 2004.

**305**(5689): p. 1437-1441.

118. Meister, G., et al., Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell*, 2004. **15**(2): p. 185-197.

119. Park, M.S., et al., Human Argonaute3 has slicer activity. *Nucleic acids research*, 2017. **45**(20): p. 11867-11877.

120. Qi, H.H., et al., Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature*, 2008. **455**(7211): p. 421-424.

121. Horman, S.R., et al., Akt-mediated phosphorylation of argonaute 2 downregulates cleavage and upregulates translational repression of MicroRNA targets. *Molecular cell*, 2013. **50**(3): p. 356-367.

122. Bronevetsky, Y., et al., T cell activation induces proteasomal degradation of Argonaute and rapid remodeling of the microRNA repertoire. *Journal of Experimental Medicine*, 2013. **210**(2): p. 417-432.

123. Rybak, A., et al., The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. *Nature cell biology*, 2009. **11**(12): p. 1411-1420.

124. Johnston, M., et al., HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Molecular biology of the cell*, 2010. **21**(9): p. 1462-1469.

125. Smibert, P., et al., Homeostatic control of Argonaute stability by microRNA availability. *Nature structural & molecular biology*, 2013. **20**(7): p. 789-795.

126. Stepicheva, N.A. and J.L. Song, Function and regulation of microRNA-31 in development and disease. *Molecular reproduction and development*, 2016. **83**(8): p. 654-674.

127. Yu, T., et al., Functions and mechanisms of microRNA-31 in human cancers. *Biomedicine & Pharmacotherapy*, 2018. **108**: p. 1162-1169.

128. Mitamura, T., et al., Downregulation of miRNA-31 induces taxane resistance in ovarian cancer cells through increase of receptor tyrosine kinase MET. *Oncogenesis*, 2013. **2**(3): p. e40-e40.
129. Creighton, C.J., et al., Molecular profiling uncovers a p53-associated role for microRNA-31 in inhibiting the proliferation of serous ovarian carcinomas and other cancers. *Cancer research*, 2010. **70**(5): p. 1906-1915.
130. Noh, J.H., et al., Aberrant regulation of HDAC2 mediates proliferation of hepatocellular carcinoma cells by deregulating expression of G1/S cell cycle proteins. *PloS one*, 2011. **6**(11): p. e28103.
131. Lin, P.-C., et al., Epigenetic repression of miR-31 disrupts androgen receptor homeostasis and contributes to prostate cancer progression. *Cancer research*, 2013. **73**(3): p. 1232-1244.
132. Luo, L.-j., et al., MiR-31 inhibits migration and invasion by targeting SATB2 in triple negative breast cancer. *Gene*, 2016. **594**(1): p. 47-58.
133. Körner, C., et al., MicroRNA-31 sensitizes human breast cells to apoptosis by direct targeting of protein kinase C  $\epsilon$  (PKC $\epsilon$ ). *Journal of Biological Chemistry*, 2013. **288**(12): p. 8750-8761.
134. Yang, M.-H., et al., Elevated microRNA-31 expression regulates colorectal cancer progression by repressing its target gene SATB2. *PloS one*, 2013. **8**(12): p. e85353.
135. Mansour, M.A., et al., SATB1 and SATB2 play opposing roles in c-Myc expression and progression of colorectal cancer. *Oncotarget*, 2016. **7**(4): p. 4993.
136. Mansour, M.A., et al., SATB 2 suppresses the progression of colorectal cancer cells via inactivation of MEK 5/ERK 5 signaling. *The FEBS journal*, 2015. **282**(8): p. 1394-1405.
137. Gurung, A.B. and A. Bhattacharjee, Significance of Ras signaling in cancer and strategies

for its control. Journal-Significance of Ras Signaling in Cancer and Strategies for its Control, 2015.

138. Sanchez-Vega, F., et al., Oncogenic signaling pathways in the cancer genome atlas. Cell, 2018. **173**(2): p. 321-337. e10.

139. Sun, D., et al., MicroRNA-31 activates the RAS pathway and functions as an oncogenic MicroRNA in human colorectal cancer by repressing RAS p21 GTPase activating protein 1 (RASA1). Journal of Biological Chemistry, 2013. **288**(13): p. 9508-9518.

140. Gao, Y., et al., Downregulation of MiR-31 stimulates expression of LATS2 via the hippo pathway and promotes epithelial-mesenchymal transition in esophageal squamous cell carcinoma. Journal of Experimental & Clinical Cancer Research, 2017. **36**: p. 1-20.

141. Edmonds, M.D., et al., MicroRNA-31 initiates lung tumorigenesis and promotes mutant KRAS-driven lung cancer. The Journal of clinical investigation, 2016. **126**(1): p. 349-364.

142. Liu, X., et al., MicroRNA-31 regulated by the extracellular regulated kinase is involved in vascular smooth muscle cell growth via large tumor suppressor homolog 2. Journal of Biological Chemistry, 2011. **286**(49): p. 42371-42380.

143. Liu, X., et al., MicroRNA-31 functions as an oncogenic microRNA in mouse and human lung cancer cells by repressing specific tumor suppressors. The Journal of clinical investigation, 2010. **120**(4): p. 1298-1309.

144. Crooke, S.T., et al., Antisense technology: A review. Journal of Biological Chemistry, 2021. **296**.

145. Zamecnik, P.C. and M.L. Stephenson, Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. Proceedings of the National Academy of Sciences, 1978. **75**(1): p. 280-284.

146. Chernikov, I.V., V.V. Vlassov, and E.L. Chernolovskaya, Current development of siRNA



- bioconjugates: from research to the clinic. *Frontiers in Pharmacology*, 2019. **10**: p. 444.
147. Crooke, S.T., et al., Antisense drug discovery and development technology considered in a pharmacological context. *Biochemical pharmacology*, 2021. **189**: p. 114196.
148. Bennett, C.F., Therapeutic antisense oligonucleotides are coming of age. *Annual review of medicine*, 2019. **70**: p. 307-321.
149. Crooke, S.T., et al., RNA-targeted therapeutics. *Cell metabolism*, 2018. **27**(4): p. 714-739.
150. Cerritelli, S.M. and R.J. Crouch, Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal*, 2009. **276**(6): p. 1494-1505.
151. Tuschl, T., et al., Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & development*, 1999. **13**(24): p. 3191-3197.
152. Bumcrot, D., et al., RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nature chemical biology*, 2006. **2**(12): p. 711-719.
153. Kawamata, T. and Y. Tomari, Making risc. *Trends in biochemical sciences*, 2010. **35**(7): p. 368-376.
154. Yoda, M., et al., ATP-dependent human RISC assembly pathways. *Nature structural & molecular biology*, 2010. **17**(1): p. 17-23.
155. Baker, K.E. and R. Parker, Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current opinion in cell biology*, 2004. **16**(3): p. 293-299.
156. Chang, Y.-F., J.S. Imam, and M.F. Wilkinson, The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.*, 2007. **76**: p. 51-74.
157. Yin, S., et al., Evidence that the nonsense-mediated mRNA decay pathway participates in X chromosome dosage compensation in mammals. *Biochemical and biophysical research communications*, 2009. **383**(3): p. 378-382.

158. Raguraman, P., et al., Antisense oligonucleotide-mediated splice switching: Potential therapeutic approach for cancer mitigation. *Cancers*, 2021. **13**(21): p. 5555.
159. Alghoul, F., et al., Translation inhibitory elements from Hoxa3 and Hoxa11 mRNAs use uORFs for translation inhibition. *Elife*, 2021. **10**: p. e66369.
160. Bao, C., et al., mRNA stem-loops can pause the ribosome by hindering A-site tRNA binding. *Elife*, 2020. **9**: p. e55799.
161. Bao, C., et al., Specific length and structure rather than high thermodynamic stability enable regulatory mRNA stem-loops to pause translation. *Nature Communications*, 2022. **13**(1): p. 988.
162. Migliorati, J.M., et al., Absorption, Distribution, Metabolism, and Excretion of US Food and Drug Administration–Approved Antisense Oligonucleotide Drugs. *Drug Metabolism and Disposition*, 2022. **50**(6): p. 888-897.
163. Parham, J.S. and A.C. Goldberg, Mipomersen and its use in familial hypercholesterolemia. *Expert opinion on pharmacotherapy*, 2019. **20**(2): p. 127-131.
164. Keam, S.J., Inotersen: first global approval. *Drugs*, 2018. **78**: p. 1371-1376.
165. Benson, M.D., et al., Inotersen treatment for patients with hereditary transthyretin amyloidosis. *New England Journal of Medicine*, 2018. **379**(1): p. 22-31.
166. Goodkey, K., et al., Nusinersen in the treatment of spinal muscular atrophy. Exon skipping and inclusion therapies: *Methods and protocols*, 2018: p. 69-76.

## CHAPTER II pH Dependence of C•A, G•A and A•A Mismatches in the Stem of Precursor MicroRNA-31

This chapter is a modified version of a published article with authors listed as follows:

Kotar A, Ma S, Keane SC. pH dependence of C•A, G•A and A•A mismatches in the stem of precursor microRNA-31. *Biophysical Chemistry* 283, 106763 (2022)

**Author Contribution:** Anita Kotar: Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. Sicong Ma: Data curation, Investigation, Writing – review & editing. Sarah C. Keane: Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition, Resources.

NMR, Circular Dichroism experiments and analyses in Fig 2.2-Fig 2.12 and Table 2.1-Table 2.4 were performed by Anita Kotar. Initial methods development was carried out by Anita Kotar. All other works were done by Sicong Ma.

**Abstract:** MicroRNAs (miRNAs) are important regulators of post-transcriptional gene expression. Mature miRNAs are generated from longer transcripts (primary, pri- and precursor, pre-miRNAs) through a series of highly coordinated enzymatic processing steps. The sequence and structure of these pri- and pre-miRNAs play important roles in controlling their processing.

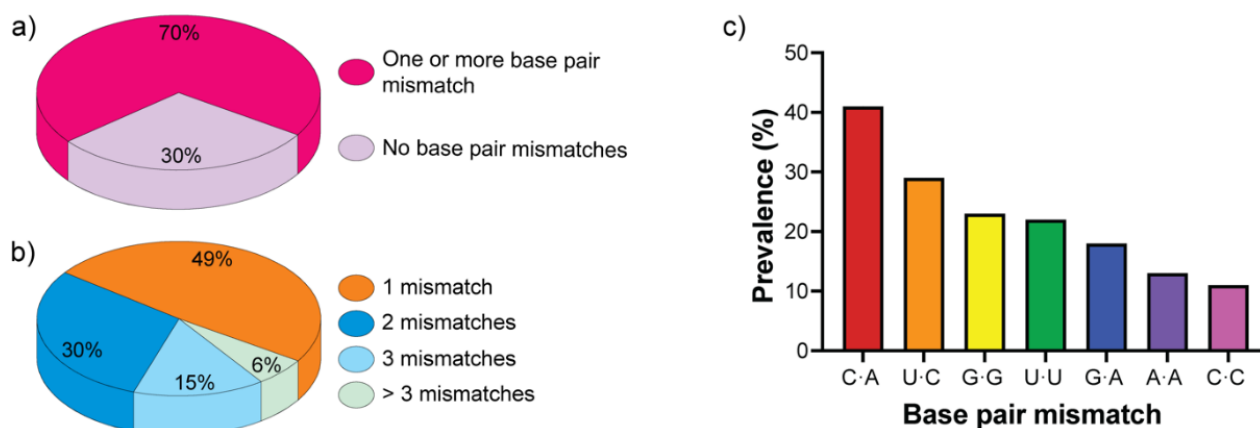
Both pri- and pre-miRNAs adopt hairpin structures with imperfect base pairing in the helical stem. Here, we investigated the role of three base pair mismatches (A·A, G·A, and C·A) present in pre-miRNA-31. Using a combination of NMR spectroscopy and thermal denaturation, we found that nucleotides within the three base pair mismatches displayed unique structural properties, including varying dynamics and sensitivity to solution pH. These studies deepen our understanding of how the physical and chemical properties of base pair mismatches influence RNA structural stability.

## 2.1 Introduction

MicroRNAs (miRNAs) are short (~22 nucleotide) non-coding RNAs that play an important role in the regulation of gene expression through targeting messenger (m) RNAs for post-transcriptional gene silencing [1]. Post-transcriptional regulation of miRNA biogenesis is a multistep process that starts when a long primary (pri-) miRNA is enzymatically processed by Drosha/DiGeorge syndrome critical region 8 (DGCR8) to generate the precursor (pre-) miRNA. The pre-miRNA is exported from the nucleus where it is further processed by the cytoplasmic Dicer enzyme into the mature miRNA duplex [1]. One strand of the mature miRNA duplex is loaded into the RNA-induced silencing complex (RISC), where it serves as a template for complementary recognition of the target mRNA. It has been suggested that one third of genes in the genome are regulated by miRNAs [2]–[4], impacting diverse biological processes [5]–[7]. miRNA processing is tightly regulated to ensure accurate gene expression. The sequence and structure of both pri- and pre-miRNAs play important roles in regulation of miRNA processing as both pri- and pre-miRNAs are differently recognized by the processing enzymes [8]–[13]. The presence of stable basal stems in pri-miRNAs and flexible apical loops in pri-/pre-miRNAs lead to more efficient processing by the biogenesis machinery [14]. The effects of other structural

elements like mismatches within stems of pri-/pre-miRNAs on processing are not as well characterized, but are emerging as important regulators of miRNA maturation [15]. This is especially important given that more than two-thirds of human pre-miRNAs contain at least one base pair mismatch (**Fig. 2.1**).

Mismatches and wobble base pairs in the upper stem of pri-miRNAs can impact the efficiency and accuracy of miRNA processing [16]. Importantly, some non-canonical base pair geometries in pri-/pre-miRNAs can be induced by altering solution conditions, such as pH. An A<sup>+</sup>-G mismatch was found in the stem-loop region of pre-miRNA-21 excited states, which enhanced Dicer processing over its ground conformational state [17]. While in single stranded regions of RNA, the nucleobases are typically uncharged, with reported pK<sub>a</sub> values of 3.5, 4.2, and 9.2 for A, C and G/U, respectively [18]. In higher order structural environments like mismatches, the pK<sub>a</sub> values can be shifted towards neutrality which leads nucleobases to adopt uncommon protonated states that can facilitate different non-canonical interactions [19].



**Figure 2.1. Statistics of base pair mismatches in human pre-miRNA stems.** a) Fraction of RNAs with and without mismatches, b) breakdown of the number of mismatches, and c) the prevalence of various base pair mismatches in pre-miRNA stems.

We examined the base pair mismatches in the stem of the evolutionarily conserved pre-miRNA-31, which is involved in maintaining fertility, embryonic development, bone formation, and myogenesis [20]. miRNA-31 is known to interact with a series of target genes and pathways; thus, its mis-regulation has been connected to various diseases like cancer, autoimmune diseases, and heart conditions such as atrial fibrillation [21]–[24]. Notably, in tumorigenesis miRNA-31 can act as an enhancer of tumor development and progression (in lung, colorectal, non-small-cell lung, head and neck squamous cell, and esophageal squamous cell cancers) as well as a tumor suppressor (in breast, ovarian, and prostate cancers and in hepatocellular and gastric carcinoma) [23]. Additionally, miRNA-31 regulates diverse processes during embryonic implantation and development as well as promotes early sperm development [20]. Moreover, miRNA-31 can positively regulate the proliferation, differentiation, and cell activity of keratinocytes, which are important functions connected to various skin diseases as well as wound healing [22], [24]. Pre-miRNA-31 contains three base pair mismatches (C·A, G·A, and A·A) in its helical stem. Using NMR and UV spectroscopic methods, we evaluated the structural characteristics of these base pair mismatches and analyzed how their properties change when the pH of the buffering solution was altered.

## **2.2 Material and methods**

### **2.2.1 Cataloging of mismatches in human pre-miRNAs**

All 1,917 pre-miRNA sequences were downloaded from the miRBase human microRNA database [25] and secondary structures were predicted using the RNAstructure web server [26] with default parameters. We defined the stem as regions between the 5'-most base paired nucleotides (distal from the apical loop) and the last set of base paired nucleotides which are closest

to the Dicer cleavage site (proximal to the apical loop). Almost all (1,883/1,917) of the extracted pre-miRNA structural information was based on the predicted lowest energy structure. However, in a few cases (34/1,917) the data was based on the lowest energy structure which maintained a canonical stem-loop structure. All relevant information, including the identity of base pairs (A-U, G-C, G-U), single nucleotide mismatches (C·A, A·G, A·A, C·C, U·C, U·U, G·G), and unpaired nucleotides (including internal loops and single nucleotide bulges), was recorded and analyzed.

### 2.2.2 Construct design and template preparation

RNAs examined in this study are listed in **Table 2.1**. DNA oligonucleotides were purchased from Integrated DNA Technologies. The DNA templates were ordered with 2'-O-methoxy modifications at the two 5'-most positions to reduce non-templated transcription [27]. The DNA templates for *in vitro* transcription were created by annealing the DNA oligonucleotides (BottomA: 5'-mGmGAAAGATGGCAATCTCTTGCCTCCTCTCCT*TATAGTGAGTCGTATTA*-3', BottomB: 5'-mGmGCAATATGTTGGTCTCCCAGCATCTTGCCT*TATAGTGAGTCGTATTA*-3'; where m denotes 2'-O-Me modification of the oligonucleotide and italicized nucleotides correspond to the sequence complementary to the T7 promoter) with an oligonucleotide corresponding to the T7 promoter sequence (5'-TAATACGACTCACTATA-3'). Templates were prepared by mixing the desired DNA oligonucleotide (40  $\mu$ L, 200  $\mu$ M) with the complementary oligonucleotide to T7 promoter sequence (20  $\mu$ L, 600  $\mu$ M) together, boiling for 3 min, and then slowly cooling to room temperature. The annealed template was diluted with H<sub>2</sub>O prior to use to produce the partially double-stranded DNA templates at a final concentration approximately 8  $\mu$ M.

**Table 2.1. RNA constructs.**

Construct	RNA sequence 5'→3' <sup>a</sup>
pre-miRNA-31	GGAGAGGAGGCAAGAUGCUGGCAUAGCUGUUGAACUGGG AACCUGCUAUGCCAACAUAUUGCCAUCUUUCC
BottomA	GGAGAGGAGGCAAGAGAUUGCCAUCUUUCC
BottomB	GGCAAGAUGCUGGGAGACCAACAUAUUGCC
BottomA A8U	GGAGAGGUGGCAAGAGAUUGCCAUCUUUCC
BottomB G14U	GGCAAUAUGCUGGGAGACCAACAUAUUGCC
BottomB C18U	GGCAAGAUGUUGGGAGACCAACAUAUUGCC
BottomB 14/18U	GGCAAUAUGUUGGGAGACCAACAUAUUGCC

<sup>a</sup> Red nucleotides indicate non-native nucleotides corresponding to the GAGA tetraloop sequence.

### 2.2.3 RNA Preparation

pre-miRNA-31, BottomA and BottomB RNAs (**Table 2.1**) were prepared by *in vitro* transcription in 1X transcription buffer [40 mM Tris base, 5 mM dithiothreitol (DTT), 1 mM spermidine, and 0.01% Triton-X (pH=8.5)] with addition of 3–6 mM ribonucleoside triphosphates (NTPs), 10–20 mM magnesium chloride (MgCl<sub>2</sub>), 30–40 ng/μL DNA template, 0.2 unit/mL yeast inorganic pyrophosphatase (New England Biolabs) [28], ~15 μM T7 RNA polymerase, and 10–20% (v/v) dimethyl sulfoxide (DMSO). Reaction mixtures were incubated at 37 °C for 3–4 h, with shaking at 70 rpm, and then quenched using a solution of 7 M urea and 500 mM ethylenediaminetetraacetic acid (EDTA), pH=8.5. Reactions were boiled for 3 min and then snap cooled in ice water for 3 min. The transcription mixture was loaded onto preparative-scale 14% (pre-miR-31) or 18% (BottomA, BottomB) denaturing polyacrylamide gels for purification. Gel slices containing the target RNA were crushed and soaked in 1X Tris/Borate/EDTA (TBE) buffer for 24–48 h to extract the RNA. Eluted RNA was filtered and spin concentrated, washed with 2 M high-purity sodium chloride, and exchanged into water using Amicon-15 Centrifugal Filter Units (Millipore, Sigma). RNA quality was verified by running the purified RNA on an analytical 14% (pre-miR-31) or 18% (BottomA, BottomB) denaturing polyacrylamide gel.



The modified RNAs with stabilizing mutations (BottomA\_A8U, BottomB\_G14U, BottomB\_C18U and BottomB\_14/18U) and site-specific  $^{13}\text{C}$ - or  $^{15}\text{N}$ -labeled RNAs were chemically synthesized using 4,4'-dimethoxytrityl (DMT) phosphoramidite chemistry on K&A Laborgeraete GbR DNA/RNA Synthesizer H-8. Samples were purified using the Glen-Pak RNA Cartridge Purification (DMT-ON) protocol [29]. RNAs were washed with 2 M high-purity sodium chloride and exchanged into water using Amicon-15 Centrifugal Filter Units (Millipore, Sigma).

Purified and desalted RNAs were refolded by heating in boiling water for 3 min, followed by incubation on ice for 3 min. For NMR experiments in 100%  $\text{D}_2\text{O}$ , the RNA samples were lyophilized and dissolved in 100%  $\text{D}_2\text{O}$  (99.8%, Cambridge Isotope Laboratories, Inc.) and transferred to an Eppendorf tube containing lyophilized buffer (50 mM K-phosphate buffer (pH=7.5 or pH=5.8) and 1 mM  $\text{MgCl}_2$ ) with RNA concentration of 0.4 mM. Other NMR samples were prepared in 90%/10%  $\text{H}_2\text{O}/\text{D}_2\text{O}$  with 50 mM K-phosphate buffer (pH=7.5 or pH=5.8), 1 mM  $\text{MgCl}_2$ , and 0.2-0.5 mM RNA concentrations.

#### **2.2.4 NMR experiments**

2D  $^1\text{H}$ - $^1\text{H}$  nuclear Overhauser effect spectroscopy (NOESY),  $^1\text{H}$ - $^1\text{H}$  total correlation spectroscopy (TOCSY), and  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear multiple quantum correlation (HMQC) spectra for NMR assignment were recorded at 30 °C for BottomA, and at 37 °C for BottomB. NMR spectra were collected on 600 and 800 MHz Bruker AVANCE NEO spectrometers equipped with a 5 mm three channel inverse (TCI) cryogenic probe and on a 600 MHz Bruker AVANCE DRX NMR spectrometer equipped with a 5 mm pulsed field gradient (PFG) cryoprobe (University of Michigan BioNMR Core). NMR data were processed with NMRFX [30] analyzed with NMRViewJ [31] and MestReNova 12.0.0-20080 [32]. Details related to specific experimental

parameters can be found in the supporting information (**Table 2.2**).  $^1\text{H}$  chemical shifts were referenced to water and  $^{13}\text{C}$  chemical shifts were indirectly referenced from the  $^1\text{H}$  chemical shift [33].

**Table 2.2. NMR experimental parameters.**

	$^1\text{H}$ - $^1\text{H}$ NOESY	$^1\text{H}$ - $^1\text{H}$ TOCSY	$^1\text{H}$ - $^{13}\text{C}$ HMQC
Pulse sequence	noesyphpr	mlevgpph19	hmqcphpr
ds	16	32	16
ns	32-128	128-256	304
sw(F2)	10.0138-10.7724	10.0138-10.2427	8.778
sw(F1)	10.0138-10.7724	10.0138-10.2427	50.0
TD(F2)	8129	2048	1058
TD(F1)	68-800	256-360	80
O1	4.708	4.708	4.705
O2	-	-	148.00
D1	2-5 s	2 s	1.5 s
$\tau_m$	400-500 ms	$\tau_m=80$ ms	-

### 2.2.5 pH titrations and pKa measurements

BottomA and BottomB samples for the pH titration NMR experiments were prepared with 0.2 mM concentration of RNAs in 10%  $\text{D}_2\text{O}$ , and 1 mM  $\text{MgCl}_2$ . Initial pH values were set to 7.5 with 100 mM NaOH using SevenEasy Mettler Toledo pH meter. For the pH titration series, the samples were titrated with 50 mM HCl and after each addition of acid the pH was checked using the pH meter. All solutions were freshly prepared the same day NMR spectra were acquired.

The changes in  $^1\text{H}$  chemical shifts were followed during the pH titration. The graphs of chemical shift changes versus pH values were used to determine pKa values using **Equation 1** [34]:

$$\delta_{obs} = \delta_A + \frac{\delta_{AH} - \delta_A}{1 + 10^{n(\text{pH} - \text{pKa})}} \quad (1)$$

where  $\delta_A$  is the chemical shift at high pH,  $\delta_{AH}$  is the chemical shift at low pH, and  $\delta_{obs}$  is the

observed chemical shift at a given pH.

## 2.2.6 Thermal denaturation of RNA and data analysis

UV-thermal denaturation experiments were performed using an Agilent Cary UV-Vis Multicell Peltier spectrometer with a heating rate of 0.5 °C per min between 10 to 95 °C and 95 to 10 °C. Data points were collected every 1 °C with absorbance detection at 260, 295 and 330 nm. RNA samples (20 μM) were prepared in 50 mM K-phosphate buffer (pH=7.5 or pH=5.8). The melting profiles (at 260 nm) revealed reversible single-transition unfolding and were analyzed using a two-state model with sloping baselines (**Equation 2**) [35]:

$$f(T) = \frac{(m_u T + b_u) + (m_f T + b_f) e^{\left[\frac{\Delta H}{R}\right]\left[\frac{1}{(T_m + 273.15)} - \frac{1}{(T + 273.15)}\right]}}{1 + e^{\left[\frac{\Delta H}{R}\right]\left[\frac{1}{(T_m + 273.15)} - \frac{1}{(T + 273.15)}\right]}} \quad (2)$$

where  $m_u$  and  $m_f$  are the slopes of the lower (unfolded) and upper (folded) baselines,  $b_u$  and  $b_f$  are the y-intercepts of the lower and upper baselines, respectively.  $\Delta H$  (in kcal/mol) is the enthalpy of unfolding,  $T_m$  (in °C) is the melting temperature, and  $R$  is the gas constant (0.001987 kcal/(Kmol)). The entropy change ( $\Delta S$ ) was determined by **equation 3**:

$$\Delta S = \frac{\Delta H}{T_m} \quad (3)$$

The standard free energy change ( $\Delta G^\circ$ ) was calculated at 25 °C using  $\Delta H$  and  $\Delta S$ . Experiments were performed in duplicate and were in good agreement.

## 2.3 Results

### 2.3.1 70% of human pre-miRNAs contain at least one base pair mismatch in the stem

We predicted the secondary structures of all pre-miRNAs using the RNAstructure web

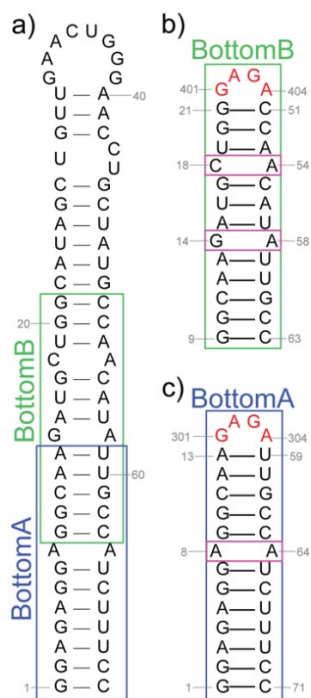
server [26] and analyzed the occurrence of base pair mismatches. We found that approximately 70% (1,351/1,917) of human pre-miRNAs contain at least one base pair mismatch in the stem (**Fig. 2.1a**). Most pre-miRNAs were predicted to contain a single base pair mismatch (49 %), followed by those predicted to contain two (30%). Pre-miRNAs containing three base pair mismatches accounted for 15% of those surveyed, while relatively few (6%) of pre-miRNAs were predicted to contain more than three base pair mismatches (**Fig. 2.1b**). C·A mismatches are found in 41% (559/1,351) of pre-miRNAs containing at least one mismatch. A summary of the presence of other mismatches is as follows: U·C mismatches (29%, 396/1,351); G·G mismatches (23%, 307/1,351); U·U mismatches (22%, 302/1,351); G·A mismatches (18%, 247/1,351); A·A mismatches (13%, 174/1,351), and C·C mismatches (11%, 142/1,351) (**Fig. 2.1c**). Pre-miRNA-31 is predicted to contain three mismatches, C·A, G·A and A·A, in its stem region (**Fig. 2.2a**).

### **2.3.2 NMR data reveal the pre-miRNA-31 base pair mismatches have different conformations**

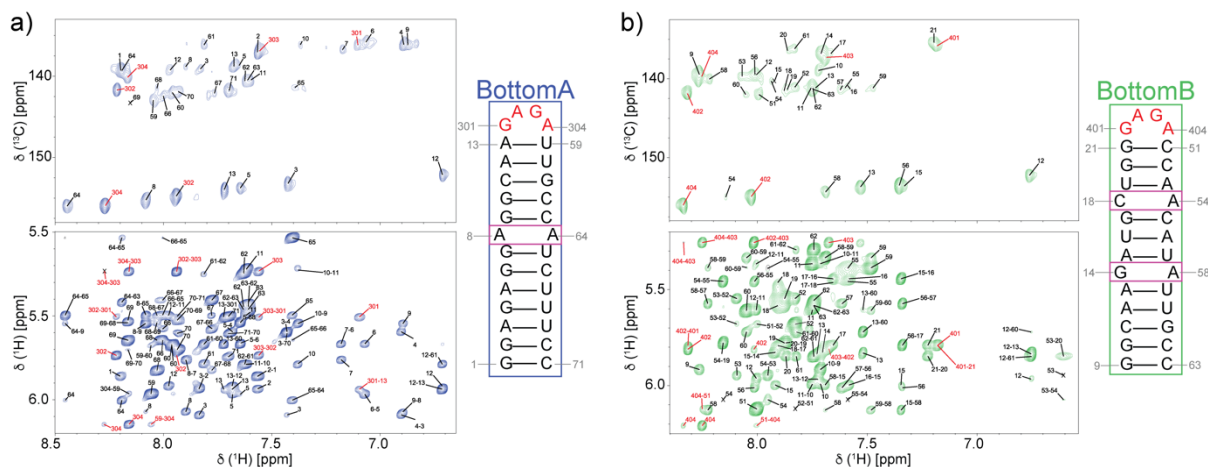
In the predicted pre-miRNA-31 secondary structure (**Fig. 2.2a**), we focused on the helical stem region containing three base pair mismatches. We designed two RNAs, BottomA and BottomB, to examine the structure and properties of these base pair mismatches. The stem of BottomA is composed of four A·U, six G·C base pairs, and two G·U base pairs. BottomA also contains one A·A mismatch. The stem of BottomB consists of five A·U and six G·C base pairs. Additionally, BottomB has one C·A and one G·A mismatch. Both BottomA and BottomB RNAs contain a non-native GAGA tetraloop designed to cap the helical stem (**Fig. 2.2b,c**). The GAGA-tetraloop is structurally stable and yields unambiguous and characteristic patterns of signals in 2D <sup>1</sup>H-<sup>1</sup>H NOESY spectra [36], [37], which facilitated resonance assignments. The observed NOE

signals of BottomA and BottomB revealed that both RNAs adopted an A-form helical stem with a properly folded GAGA tetraloop (**Fig. 2.3**). Both RNAs also adopted the same conformations as observed in full length pre-miRNA-31 as is evidenced by the matching signals in NOESY spectra of both BottomA and BottomB when compared to the full-length pre-miRNA-31 (**Fig. 2.4**).

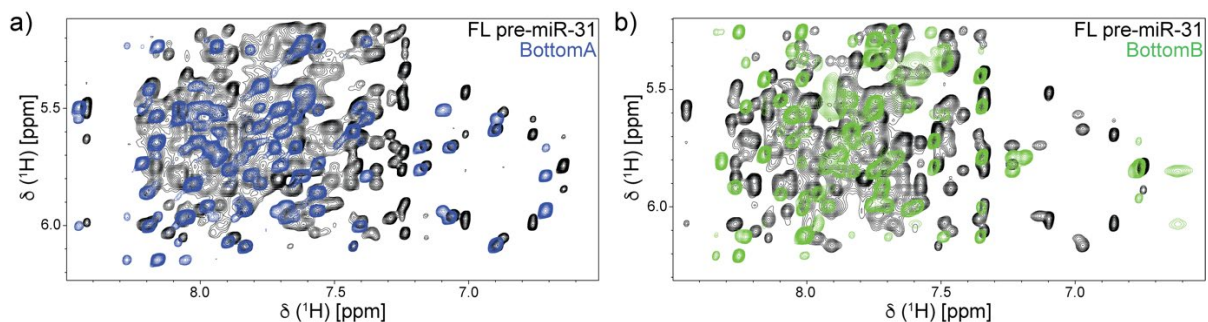
The signals of nonexchangeable protons of BottomA and BottomB were assigned based on analysis of 2D  $^1\text{H}$ - $^1\text{H}$  NOESY, 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY, and  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectra, collected at pH=7.5 (**Fig. 2.3**). All assignments are in good agreement with predicted chemical shifts (**Fig. 2.5**). The largest deviations from predicted values occur at or near mismatched regions, where no chemical shift data of such sequence and structural motifs are reported in the biological magnetic resonance data bank [38] (**Fig. 2.5**). Importantly, this study contributes a number of previously uncharacterized chemical shifts to the RNA chemical shift database (see open circles in **Fig. 2.5**).



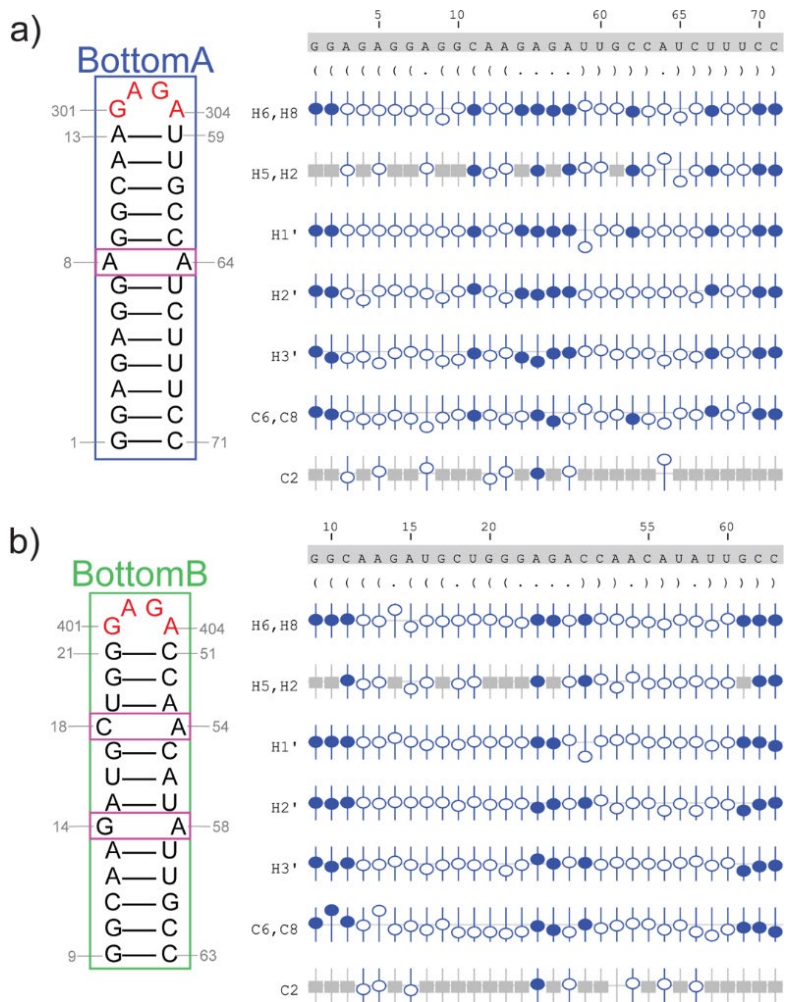
**Figure 2.2. Predicted secondary structure of pre-miRNA-31.** Secondary structure of a) pre-miR-31, b) BottomB, and c) BottomA RNA constructs. Colored rectangles denote the location of BottomA and BottomB constructs relative to pre-miRNA-31. The mismatches of interest are boxed with magenta. The secondary structures were predicted using the RNAstructure webserver [26] and were rendered using RNA2Drawer [74].



**Figure 2.3. Assigned chemical shifts of BottomA and BottomB RNAs.**  $^1\text{H}$ - $^{13}\text{C}$  HMQC (top) and  $^1\text{H}$ - $^1\text{H}$  NOESY (bottom) spectra of a) BottomA and b) BottomB RNAs. The secondary structure of each oligo is shown to the right of the spectra. The signals assigned to GAGA tetraloop are colored red. NMR spectra were recorded at 0.4 mM RNA concentration, 50 mM K-phosphate buffer, pH=7.5, 1 mM  $\text{MgCl}_2$  and 100%  $\text{D}_2\text{O}$ .



**Figure 2.4. Overlay of the aromatic-anomeric region of the  $^1\text{H}$ - $^1\text{H}$  NOESY spectra for a) BottomA (blue) and pre-miRNA-31 (black) and b) BottomB (green) and pre-miRNA-31 (black).** NMR spectra were recorded at 0.4 mM RNA concentration, 50 mM K-phosphate buffer, 1 mM  $\text{MgCl}_2$ , 100%  $\text{D}_2\text{O}$  at 30 °C (BottomA and pre-miR-31) or 37 °C (BottomB and pre-miRNA-31).



**Figure 2.5. Sequence analysis and validation of BottomA and BottomB chemical shift assignments.** Secondary structure, nucleotide numbering, and secondary structure in Vienna format for a) BottomA and b) BottomB RNAs. NMRViewJ chemical shift prediction software was used to validate proton (H6/H8, H5/H2, H1', H2', H3') and carbon (C6/C8, C2) assignments. Assigned atoms are represented with blue circles (open and closed), while grey boxes denote atoms that are not present in a given base. Deviation from the predicted chemical shift is shown with the offset from the center. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. All assigned chemical shifts are within 0.2 ppm of the predicted value.

2D  $^1\text{H}$ - $^1\text{H}$  NOESY spectra were used to assign through space correlations and to obtain sequential connectivities in the A-helical stem regions of BottomA and BottomB. The characteristic NOEs of GAGA tetraloop residues were used as reference points which helped to facilitate the resonance assignments of BottomA and BottomB via the established principles of a

sequential walk (**Fig. 2.3**). The NOE signals from C2 protons of adenosine residues were crucial in identifying cross-strand NOE connectivities and consequently identifying mismatch conformations. In the case of the A8·A64 mismatch in BottomA, we observed cross-strand connectivities between A8.H2 and U65.H1', A8.H2 and U65.H6, A64.H2 and G9.H1', as well as A64.H2 and G9.H8 (**Table 2.3**). We also observed strong sequential NOE connectivities with neighboring residues for both A8 and A64. These results indicate that A8 and A64 are oriented inside the stem, stacked relative to their neighboring residues.

Similarly, for the G14·A58 mismatch in BottomB we observe NOE cross-strands connectivities between A58.H2 and A15.H2 (**Table 2.3**). Interestingly, for G14 we observe clear sequential NOE connectivities with A15, while the sequential NOE connectivities between G14 and A13 could not be confirmed due to either missing signals or signal overlap in the aromatic-anomeric region. Taken together, these findings indicate that G14 is not only oriented inside the stem and engaged in common A-form RNA stacking interactions with sequential neighbors but also can adopt conformations that disturb the ideal stacking arrangements or can position itself outside the stem. For the C18·A54 mismatch, we observed cross-strand NOE signals between A54.H2 and U19.H1' (**Table 2.3**). We noted that signals corresponding to C18 as well as neighboring A53 and C55 nucleotides exhibit very broad cross-peaks in aromatic-aromatic and aromatic-anomeric regions of NOESY spectrum. We did not detect similar broadening for proton signals of nucleotides in the other two mismatches or for the other protons in BottomA and BottomB. These observations indicate chemical exchange between different structural conformations happening in the region of the C18·A54 mismatch.



**Table 2.3. Intermolecular NOE cross-peaks defining the A·A, G·A, and C·A mismatches.**

	A·A mismatch	G·A mismatch	C·A mismatch
NOE map			
NOE cross-peaks	A8.H2-G9.H1' A8.H2-U65.H1' A8.H8-G7.H8 A8.H8-G7.H1' A8.H8-G9.H8 A8.H1'-G9.H8 A64.H8-C63.H1' A64.H8-C63.H6 A64.H8-U65.H6 A64.H8-U65.H5 A64.H1'-U65.H6 A64.H2-G9.H8 A64.H2-A8.H2 A64.H2-U65.H1' A64.H2-G9.H1' A64.H2-U65.H6	G14.H8-A15.H8 G14.H1'-A15.H8 A15.H2-A58.H1' A58.H2-A15.H2 A58.H2-A15.H1' A58.H2-U59.H1' A58.H8-U57.H1' A58.H8-U57.H6 A58.H8-U59.H6 A58.H1'-U59.H6 A58.H8-U59.H5	C18.H6-G17.H8 A54.H2-U19.H1' A54.H2-U19.H6 A54.H2-A53.H2 A54.H2-C55.H1' A54.H8-A53.H8 A54.H8-A53.H1' A54.H8-C55.H6 A54.H8-C55.H5 A54.H1'-A53.H2 A54.H1'-C55.H6

### 2.3.3 Stabilization of the structure in the C·A mismatch at lower pH

Protonation of nucleobases can promote formation of non-canonical base pairs in mismatches and other structural elements within RNA molecules and has been connected to many important biological functions [19]. Indeed, there are several examples showing that protonated adenosines may play important mechanistic roles in adenosine deamination, regulation of translational recording, ligand binding in RNA aptamers and riboswitches, and processing of microRNAs, among others [17], [39]–[42]. To explore in more detail how pH influences the

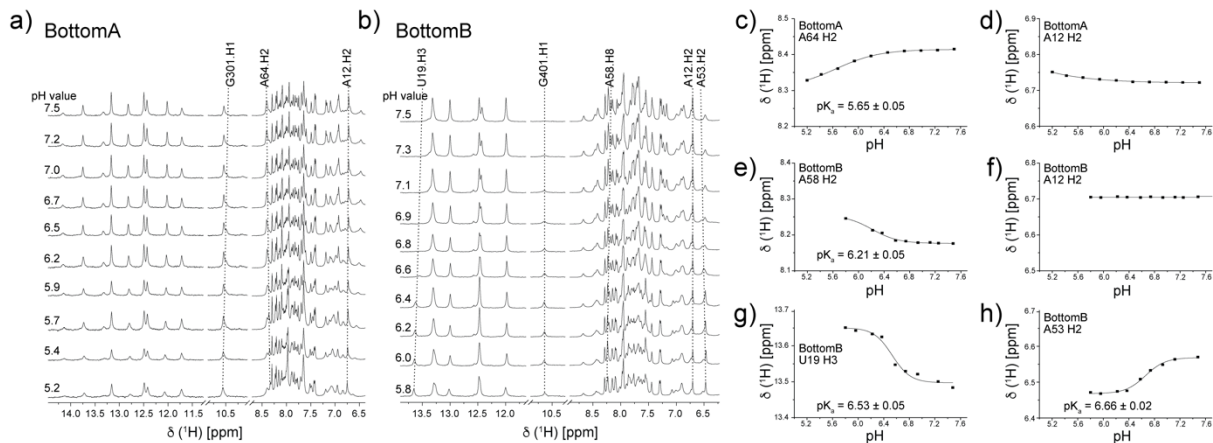
structures and the dynamic behaviors of mismatches in pre-miRNA-31, we analyzed 1D  $^1\text{H}$  NMR spectra of BottomA and BottomB collected at different pH values (**Fig. 2.6**). During the stepwise titration of BottomA with HCl, from pH=7.5 to pH=5.2, we did not observe significant changes in chemical shifts in the imino region, only slight broadening as the pH was lowered (**Fig. 2.6a**). We believe this unexpected line broadening at low pH is due to acid denaturation of the BottomA RNA [34]. In BottomA, we were interested in characterizing the pH-dependent changes of residues in the A8·A64 mismatch. Unfortunately, A8.H8 and A8.H2 signals were too overlapped for a conclusive analysis, however, the A64.H2 proton signal was nicely resolved at 8.41 ppm (pH=7.5). The chemical shift of the A64.H2 signal changes to 8.39 ppm at lower pH values and becomes slightly broader. These results were confirmed in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra where the A64.H2-A64.N1 cross-peak is much broader at pH=5.8 compared to pH=7.5 (**Fig. 2.7**). At 10.5 ppm we detected a signal from G301.H1 (partially overlapped with the signal of the GU base pair) which forms a G·A base pair within the GAGA tetraloop. The G301.H1 signal became sharper at lower pH. This is consistent with previous observations and indicates that the G301.H1 proton exchange with solvent is even slower at low pH [36], [43].

Analysis of the pH titration of BottomB revealed more significant changes in both the imino and aromatic regions of 1D  $^1\text{H}$  NMR spectra compared to BottomA (**Fig. 2.6b**). Consistent with our observations in BottomA, the chemical shift of the G401.H1 signal is at 10.64 ppm, which indicates the stabilization of the G·A base pair from the GAGA tetraloop. The most pronounced change in the NMR spectra during the titration was a sharpening of a signal at 13.65 ppm (pH=5.8) assigned to U19.H3. This signal was very broad at higher pH (barely detectable above the baseline) and became narrow and well-resolved at lower pH. We observed a comparable effect for the A53.H2 signal at 6.47 ppm (pH=5.8) which became sharp and well-resolved as the pH was

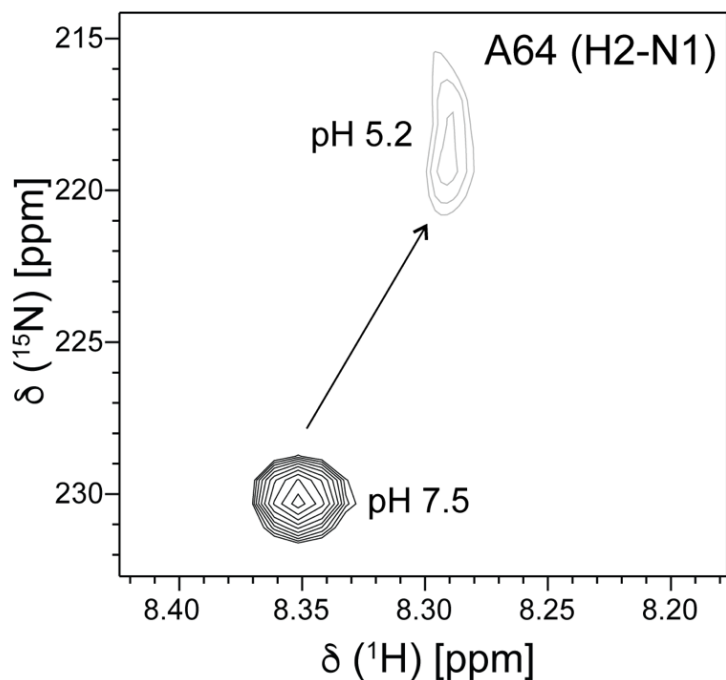
lowered. The signals assigned to C18 from the C·A mismatch were too overlapped to analyze their pH dependence. Interestingly, in the NOESY spectrum recorded at pH=5.8 (**Fig. 2.8**), cross-peaks for the nucleotides in the C·A mismatch and the neighboring residues (in particular A53.H2, C55.H6, C18.H6, and U19.H6) were narrower and more well-defined compared to their broad cross-peaks observed at pH=7.5 (**Fig. 2.3b**). This indicates that at lower pH the chemical exchange between different conformations around the C·A mismatch is not present and a single conformation is stabilized.

For the G14·A58 mismatch, we observed a broadening of the A58.H8 signal when the pH was lowered (**Fig. 2.6b**). Interestingly, in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of site-specifically  $^{15}\text{N}$ -labeled BottomB RNAs, we observed much broader H2-N1 cross-peaks for A54 and A58 compared to A64, even at pH=7.5 (**Fig. 2.9**). Additionally, in the  $^{15}\text{N}$ -A54 sample (pH=7.5), we observed a second broad cross-peak in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum with a  $^{15}\text{N}$  chemical shift of 156.6 ppm, characteristic of protonated N1 atoms (**Fig. 2.10a**). When the pH was lowered to 5.8, only a single  $^1\text{H}$ - $^{15}\text{N}$  HSQC cross-peak ( $\delta^{15}\text{N}=156.6$  ppm) was observed (**Fig. 2.10b**). The A54.N1-A54.H1 cross-peak at low pH is narrower and more defined in both the  $^1\text{H}$  and  $^{15}\text{N}$  dimensions compared to the cross-peak at pH=7.5. Additionally, comparison of the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of BottomB at low and high pH revealed an upfield shift of A54.C2 from 155 ppm (pH=7.5) to 147 ppm (pH=5.8), consistent with protonation at N1 (**Fig. 2.10c**). We did not observe similar upfield  $^{13}\text{C}$  chemical shift changes for other C2 atoms in BottomB at low pH. However, the A58 H2-C2 cross-peak is broader at pH=5.8 compared to pH=7.5, with a relatively small chemical shift change (**Fig. 2.10c**). It is also important to note that the H2-N1 cross-peak in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of A58 was too broad to detect at pH=5.8 which indicates that A58 is engaged in intermediate chemical exchange between its protonated and non-protonated form. Furthermore, our NMR data show that

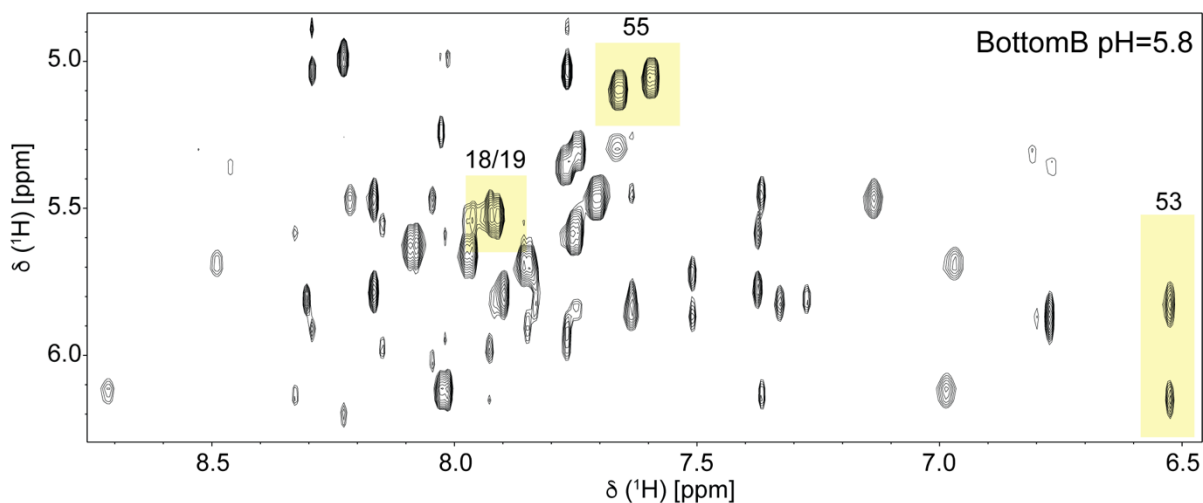
the sharpening of the A53.H2 signal at lower pH is due to the dynamic behavior in the C-A mismatch rather than protonation of A53, since the  $^{13}\text{C}$  chemical shift of the A53 H2-C2 cross-peak is 151.7 ppm, consistent with a non-protonated adenine residue (**Fig. 2.10c**). Moreover, as we can clearly observe from  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra recorded at pH values of 7.5 and 5.8, the C8  $^{13}\text{C}$  and H8  $^1\text{H}$  chemical shifts of A53 ( $\Delta\delta^{13}\text{C}$  0.1 ppm,  $\Delta\delta^1\text{H}$  0.03 ppm) are resistant to changes in pH compared to those for A54 where we observe a clear pH dependence ( $\Delta\delta^{13}\text{C}$  1.9 ppm,  $\Delta\delta^1\text{H}$  0.25 ppm) (**Fig. 2.11**). Collectively, these results strongly suggest that a population of A54 is already protonated at pH=7.5 and virtually completely protonated at pH=5.8 (**Fig. 2.10, 2.11**).



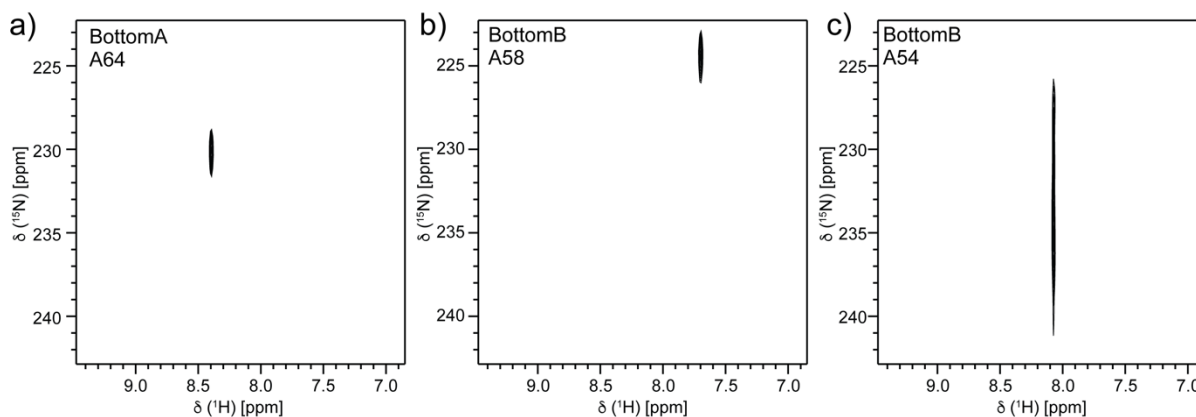
**Figure 2.6. pH dependence of chemical shifts.** The imino and aromatic regions of  $^1\text{H}$  NMR spectra of a) BottomA and b) BottomB RNAs at different pH. The pH values are indicated on the left side of the spectra. The assignments of selected signals are shown above the spectra and the dotted lines follow the changes in the chemical shifts. Effect of pH on chemical shift for c) BottomA A64 H2, d) BottomA A12 H2, e) BottomB A58 H8, f) BottomB A12 H2, g) BottomB U19 H3 and h) BottomB A53 H2.  $pK_a$  values derived from these titrations are indicated within each panel. NMR spectra were recorded at 0.2 mM of RNAs, 1 mM  $\text{MgCl}_2$ , 90%/10%  $\text{H}_2\text{O}/\text{D}_2\text{O}$  at 37 °C.



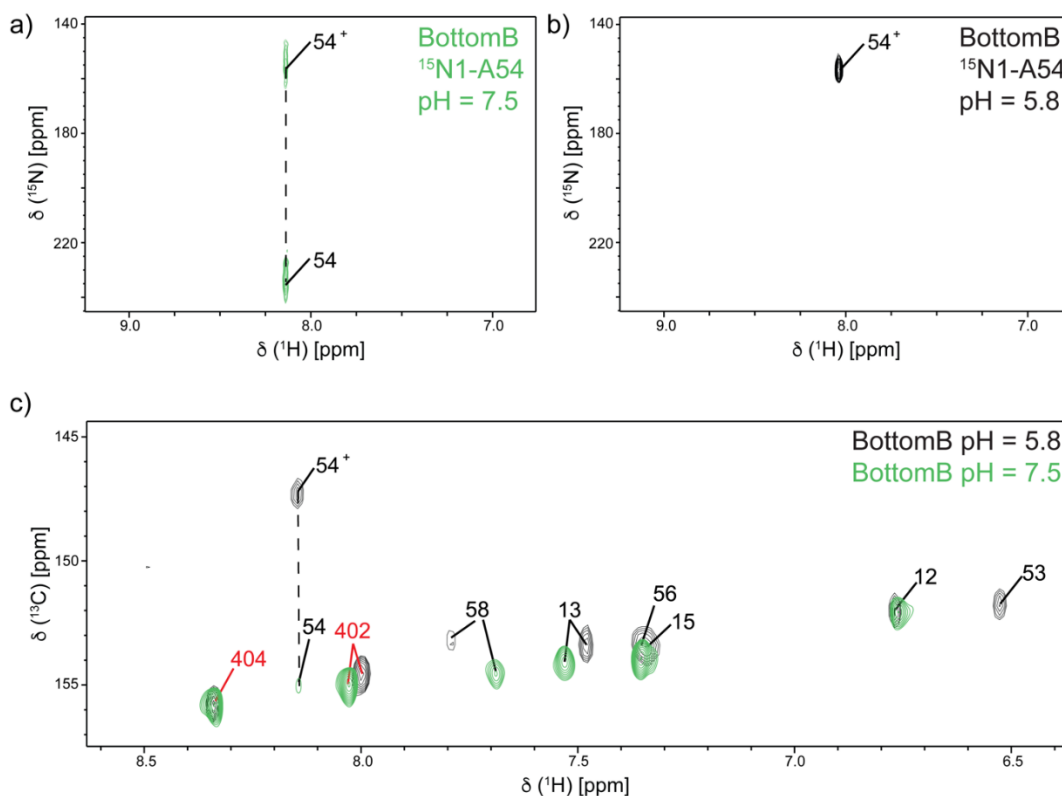
**Figure 2.7. H2-N1 cross-peak of 100%  $^{15}\text{N}$  labeled A64 in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra recorded at pH=7.5 and pH=5.2.** NMR spectra were recorded at 0.3 mM RNA concentration, 50 mM K-phosphate buffer, 1 mM  $\text{MgCl}_2$ , 10%/90%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  at 37 °C. All experimental parameters were identical for both samples.



**Figure 2.8. Aromatic-anomeric region of BottomB RNA  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum.** NMR spectrum was recorded at 0.3 mM RNA concentration, 50 mM K-phosphate buffer, pH 5.8, 1 mM  $\text{MgCl}_2$ , 10%/90%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  and at 37 °C. Yellow boxes highlight regions of the spectrum with noticeably sharper resonances, which are at or near the C-A mismatch. Residue numbers are based on the assignments made at pH=7.5. Complete resonance assignments were not made at pH=5.8.



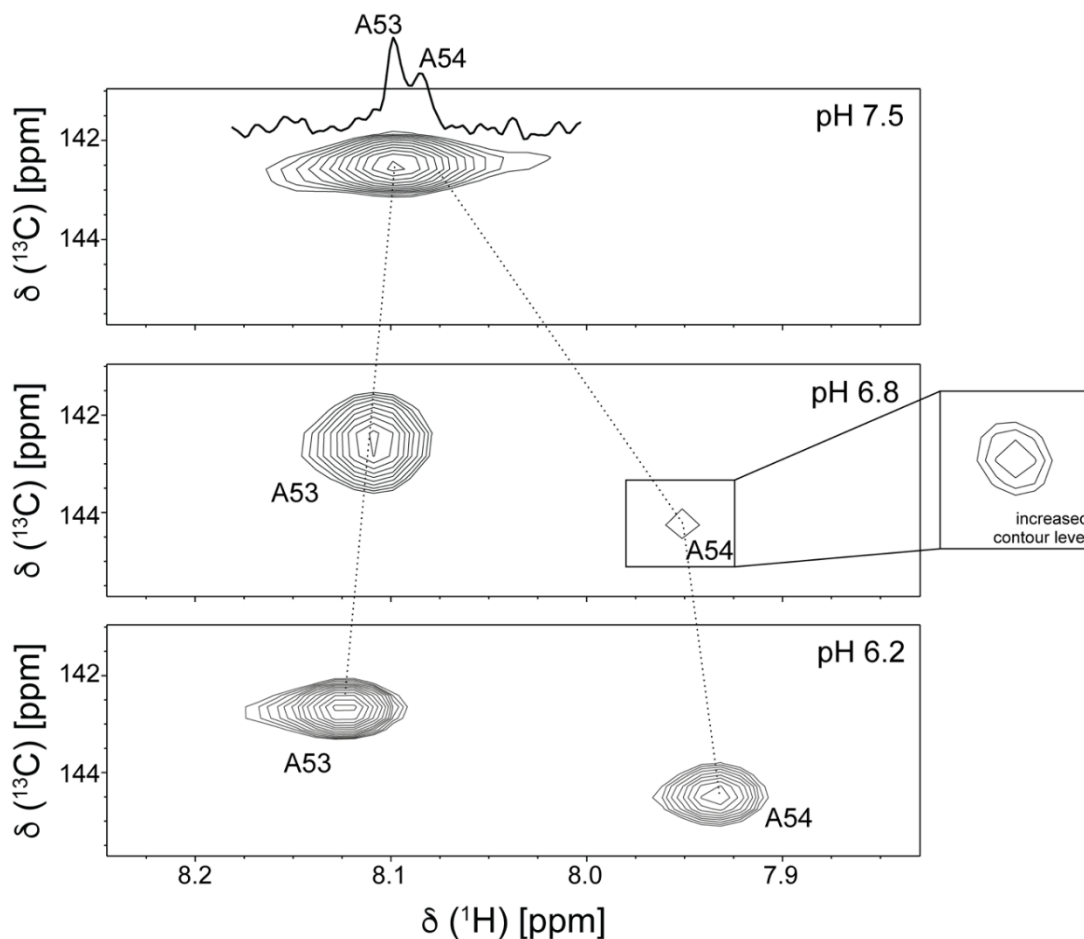
**Figure 2.9. H2-N1 cross-peaks of A58 and A54 are much broader compared to A64. H2-N1 cross-peaks of 100%  $^{15}\text{N}$  labeled a) A64, b) A58 and c) A54 in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra.** NMR spectra were recorded at 0.3 mM RNA concentration, 50 mM K-phosphate buffer, pH 7.5, 1 mM  $\text{MgCl}_2$ , 10%/90%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  and at 37 °C. To obtain NMR spectra of sufficient quality, datasets were recorded with different ns and TD(F1) values as indicated: a) ns=24, TD(F1)=60, b) ns=72, TD(F1)=60, c) ns=272, TD(F1)=76.



**Figure 2.10. Protonation of A54 is favored at low pH.** H2-N1 cross-peak of 100%  $^{15}\text{N}$  labeled A54 in  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra recorded at a) pH=7.5 and b) pH=5.8. H2-C2 cross-peaks of BottomB in  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra c) at pH=7.5 (green) and pH=5.8 (black). NMR spectra were recorded at 0.3 mM RNA concentration, 50 mM K-phosphate buffer, 1 mM  $\text{MgCl}_2$ , 90%/10%  $\text{H}_2\text{O}/\text{D}_2\text{O}$  and at 37 °C (a and c) and 25 °C (b).

### 2.3.4 pK<sub>a</sub> values of adenosine nucleotides in mismatches differ

We followed the changes in 1D <sup>1</sup>H NMR spectra during pH titrations for BottomA and BottomB to determine the pK<sub>a</sub> values of adenine residues (**Fig. 2.6**). pH values lower than 5.2 (BottomA) and 5.8 (BottomB) were not used in the analysis due to significant broadening of the signals, indicative of pH denaturation of the RNA structures. We followed the proton chemical shifts of the well-defined and resolved signals of A64.H2 (A·A mismatch) and A58.H8 (G·A mismatch). While we were not able to directly probe the pK<sub>a</sub> of A54 (C·A mismatch) due to signal overlap, we were able to monitor the pH-dependent chemical shift of the neighboring signals, U19.H3 and A53.H2. U19 and A53 form the A-U base pair above the C18·A54 mismatch and are consequently sensitive to changes in the chemical environment of the C·A mismatch. As a control we followed the signal of A12.H2 which is present in both BottomA and BottomB and whose chemical shift is not pH-dependent. The observed changes in proton chemical shifts as a function of pH were fitted to **Eq. 1** to obtain the pK<sub>a</sub> values of the residues. The lowest pK<sub>a</sub> value was observed for A64,  $5.65 \pm 0.05$  (A·A mismatch, **Fig. 2.6c**). A58 has a pK<sub>a</sub> value of  $6.21 \pm 0.05$  (G·A mismatch, **Fig. 2.6e**). While we could not directly measure the pK<sub>a</sub> value of A54 (in the C·A mismatch) due to spectral overlap, we were able to determine pK<sub>a</sub> values for the neighboring residues U19 (pK<sub>a</sub> =  $6.53 \pm 0.05$ , **Fig. 2.6g**) and A53 (pK<sub>a</sub> =  $6.66 \pm 0.02$ , **Fig. 2.6h**). These findings, coupled with the observation that A54 is protonated at pH=7.5, suggest that the pK<sub>a</sub> of A54 is the highest among the adenines in the base pair mismatches. pH sensitivity is specific for the mismatches in BottomA and BottomB as we observed practically no pH dependence for the A12.H2 signal, which is base paired in the stable region of the stem (**Fig. 2.6d,f**). Slight changes were observed for A12.H2 in the BottomA RNA due to its proximity to the GAGA tetraloop in comparison to the positioning of A12.H2 in the BottomB RNA.



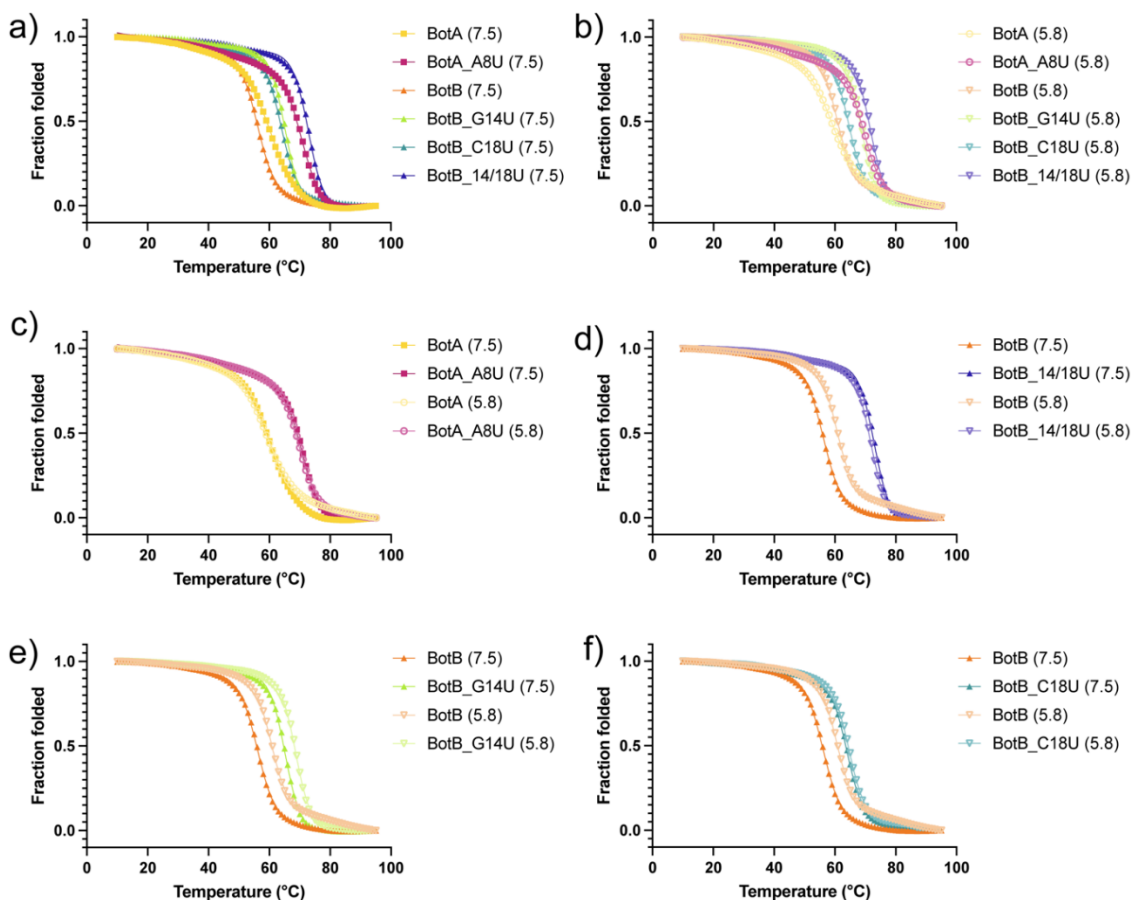
**Figure 2.11. H8-C8 correlations in  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of 100%  $^{13}\text{C}8$  labeled A53 and A54 at different pH values.** NMR spectra were recorded at 0.3 mM RNA concentration, 50 mM K-phosphate buffer, 1 mM  $\text{MgCl}_2$ , 10%/90%  $\text{D}_2\text{O}/\text{H}_2\text{O}$  at 37 °C.

### 2.3.5 Thermal stabilization of the structure in C·A mismatch but not in others

To evaluate if structure specific differences in  $\text{pK}_a$  values of adenosine residues in base pair mismatches are also connected to differences in the thermal stability of the RNAs, we performed UV-melting experiments on BottomA RNA, BottomB RNA, and a series of RNAs harboring single or multiple nucleotide substitutions (**Fig. 2.12**, **Table 2.4**). We designed one mutant for the BottomA RNA, BottomA\_A8U, that generates a canonical AU base pair at the A·A mismatch site. For BottomB, we prepared three mutants which either individually stabilized a single mismatch (BottomB\_G14U, BottomB\_C18U) or stabilized both mismatches (BottomB\_14/18U). The



BottomB RNA, which contains two mismatches, is the least stable at pH=7.5 with a  $T_m$  value of  $55.8 \pm 0.1$  °C. The stabilization of either one of the mismatches increases the thermal stability of BottomB by almost 10 °C ( $T_m$  values for BottomB\_G14U and BottomB\_C18U are  $65.2 \pm 0.2$  °C and  $64.0 \pm 0.2$  °C, respectively). A completely closed stem with both mismatches stabilized in BottomB\_14/18U resulted in the highest thermal stabilization ( $T_m=73.0 \pm 0.1$  °C). Similarly, the stabilization of the A·A mismatch with a U-A base pair in BottomA\_A8U shows thermal stabilization of 10 °C ( $T_m=71.0 \pm 0.1$  °C).



**Figure 2.12. Thermal stability of BottomA and BottomB RNAs.** Normalized UV-melting curves (260 nm) of BottomA, Bottom\_A8U, BottomB, BottomB\_G14U, BottomB\_C18U and BottomB\_G14C18U RNAs at a) pH=7.5 and b) pH=5.8. Comparison of various mutations at different pH values (c-f). Measurements were performed between 10 and 95 °C on 20  $\mu$ M RNA in 50 mM K-phosphate buffer (pH=5.8 or pH=7.5). pH values for each sample are indicated in parentheses. Fitted parameters are reported in **Table 2.4**.

**Table 2.4. Thermal stability of BottomA and BottomB RNAs.**

Oligonucleotide	$T_m^a$ (°C)		$\Delta T_m$ (°C)	$\Delta H^a$ (kcal/mol)		$\Delta G^\circ$ (kcal/mol)		$\Delta\Delta G^\circ$
	pH=7.5 <sup>b</sup>	pH=5.8 <sup>b</sup>		pH=7.5 <sup>b</sup>	pH=5.8 <sup>b</sup>	pH=7.5 <sup>b</sup>	pH=5.8 <sup>b</sup>	
BottomA	61.1 ± 0.3	58.9 ± 0.4	2.2	-47 ± 3	-50 ± 1	-5.0 ± 0.3	-5.0 ± 0.2	0
BottomA_A8U	71.0 ± 0.1	69.7 ± 0.1	1.3	-69 ± 4	-70 ± 2	-9.2 ± 0.6	-9.1 ± 0.3	-0.1
BottomB	55.8 ± 0.1	60.3 ± 0.1	-4.5	-69 ± 2	-77 ± 4	-6.5 ± 0.2	-8.2 ± 0.5	1.7
BottomB_G14U	65.2 ± 0.2	68.8 ± 0.3	-3.6	-85 ± 7	-91 ± 4	-10.1 ± 0.8	-11.7 ± 0.4	1.6
BottomB_C18U	64.0 ± 0.2	64.6 ± 0.1	-0.6	-81 ± 2	-82 ± 3	-9.3 ± 0.3	-9.6 ± 0.5	0.3
BottomB_14/18U	73.0 ± 0.1	71.8 ± 0.1	1.2	-90 ± 5	-88 ± 6	-12.5 ± 0.7	-12.0 ± 0.8	-0.5

<sup>a</sup>  $T_m$  and  $\Delta H$  values are obtained by fitting UV-melting profiles to two-state model using sloping baselines (Equation 2).

<sup>b</sup> Errors report the standard deviation from duplicate experiments.

At low pH, we observed slight destabilization of BottomA and BottomA\_A8U structures compared to pH 7.5 ( $\Delta T_m=2.2$  °C and 1.3 °C, respectively). Interestingly, at pH=5.8 the thermal stability of BottomB ( $T_m=60.3 \pm 0.1$ °C) increased by 4.5 °C relative to that at pH=7.5. This pH-dependent stabilization is connected to the base pair mismatches, since the thermal stability decreased slightly ( $\Delta T_m=1.2$  °C) in the BottomB\_14/18U RNA, which has a fully base-paired stem. Interestingly, when the C·A mismatch is stabilized in BottomB\_C18U there is only 0.6 °C increase in thermal stability at lower pH values. The stabilization of the G·A mismatch in BottomB\_G14U resulted in stabilization of the structure by 3.6 °C. Examination of the difference in the change in free energy ( $\Delta\Delta G^\circ$ ) upon lowering the solution pH further confirms these findings (Table 2.4). BottomB is stabilized by 1.7 kcal/mol when the pH is reduced from 7.5 to 5.8,

consistent with formation of an additional base pair [44]. Furthermore, when the G·A mismatch was stabilized (BottomB\_G14U), we observed a similar stabilization (1.6 kcal/mol) at low pH. However, we did not observe a similar extent of stabilization in the BottomB\_C18U RNA, which substituted the C·A mismatch with a canonical U-A base pair while preserving the G·A mismatch. At low pH, the BottomB\_C18U RNA had a marginal stabilization (0.3 kcal/mol) relative to the BottomB RNA, indicating that the G·A mismatch does not form a base pair at low pH. These results indicate that the C·A mismatch is rearranged at lower pH values in such a manner that it contributes to the stabilization of the structure, in which the protonation of A54.N1 is crucial as indicated by NMR data.

## 2.4. Discussion

Most RNAs contain helical stems with imperfect base pairing, where symmetric (base pair mismatches) and/or asymmetric internal loops disrupt the canonical base pairing. These mismatched regions are thought to mediate important intermolecular interactions, including protein recognition [45]–[51] and define the global RNA structure [52]–[55]. Internal loops have also been shown to be important for RNA-mediated processes like ribozyme cleavage [56]–[58], RNA conformational changes [59]–[61], and RNA processing [62]–[65].

G·A mismatches are the most frequently occurring type of base pair mismatch found in the RNA secondary structure database [66], [67] and our analysis revealed that C·A mismatches are the most common base pair mismatch in the stem of pre-miRNAs. Herein, we investigated the three base pair mismatches, A·A, G·A and C·A, present in the stem of the pre-miRNA-31 (**Fig. 2.2**), which is involved in the regulation of a number of physiological and pathological processes. Our NMR-based analysis enabled us to obtain detailed structural information on these base pair

mismatches and revealed that each of the three base pair mismatches differ in their structures as well as in their dynamics (**Fig. 2.3**). We found the A·A mismatch to be the most structurally restrained, with both adenine residues oriented within the stem and stacked with neighboring nucleotides. In the case of the G·A mismatch, we observed a break in sequential NOE connectivity for G14. This is consistent with a structure in which G14 may adopt different conformations, where it is not only orientated inside the stem but can also adopt conformations that disturb the ideal stacking arrangements or can be flipped outside of the stem. The most flexible and dynamic among the three mismatches is the C·A mismatch as evident by broadening of the many cross-peaks in the NOESY and  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectra that define the C·A mismatch (**Fig. 2.3**). Interestingly, similar broadening is also observed in the sequential and intra-residue NOESY as well as  $^1\text{H}$ - $^{13}\text{C}$  HMQC cross-peaks of U19, A53 and C55 residues which are sequential neighbors of residues A54 and C18 in the C·A mismatch. The signal broadening observed for the C·A mismatch and the neighboring residues suggests that the C·A mismatch is highly dynamic.

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra clearly show that A54.N1 ( $^{15}\text{N}$ 1 labeled A54) is in exchange with its protonated form even at near neutral pH (pH=7.5) (**Fig. 2.10**). Indeed, A54 in the C·A mismatch has a higher  $\text{pK}_a$  value compared to the other mismatched adenosines (A58 and A64 in the G·A and A·A mismatches, respectively) (**Fig. 2.6**). The adenosine residues in all three mismatches (A54, A58 and A64) have  $\text{pK}_a$  values above  $5.65 \pm 0.05$ , which is higher than the value expected for adenines in single stranded RNAs ( $\text{pK}_a \sim 3.5$ ) [18]. As a control, the  $\text{pK}_a$  values of A54, A58 and A64 residues were also compared to A12, which is positioned in a well-defined region of the stem of the BottomB RNA. In contrast to the three adenine residues found in base pair mismatches, we detected virtually no changes in A12.H2 chemical shifts during the pH titrations.

In an unprotonated state, A·C mismatches are not very stable. These mismatches are often stabilized by a single hydrogen bond between the adenine and cytosine bases (C.H4---A.N1 or A.H6---C.N3). However, protonation of the adenosine (A<sup>+</sup>) at the N1 position further stabilizes the mismatch, with the contribution of an A.N1<sup>+</sup>---C.O2 hydrogen bond (in addition to the C.H4---A.N1 hydrogen bond). In BottomB, we observe protonation of the N1 of A54 (**Fig. 2.10**), which can strengthen base pairing in the C·A<sup>+</sup> mismatch, as described above. Consistent with these findings, the BottomB RNA, containing the C·A mismatch, had a higher T<sub>m</sub> (4.5 °C) and more negative free energy ( $\Delta\Delta G^\circ = 1.7$  kcal/mol, **Table 2.4**) when the pH of the solution was reduced. Isolation of these effects to the C·A mismatch was carried out by mutation, stabilizing the G·A mismatch (BottomB\_G14U).

We observe a clear correlation between the pK<sub>a</sub> values of the three mismatched adenosines and the differences in their conformations. In the case of pre-miRNA-31, it seems that the higher pK<sub>a</sub> can be connected to more dynamic behavior of the nucleotide in the stem. pK<sub>a</sub> values are influenced by neighboring base pairs and can, in the case of adenosines in C·A mismatches, vary between 6.5 and 8.1 due only to effects of nearest neighbors and nearby bulges [68]. For a C·A mismatch in a 20-nucleotide long hairpin with neighboring 5'-GC and 3'-UA base pairs, similar to the mismatch in pre-miRNA-31, the mismatched adenosine had a reported pK<sub>a</sub> value of 7.84 and is stabilized by 5.1 °C when the pH value is lowered from 8.79 to 6.13 [68]. These results are in good agreement with the high pK<sub>a</sub> value of A54 in the C·A mismatch and the 4.5 °C stabilization at low pH for the BottomB RNA (**Fig. 2.12, Table 2.4**). In the case of pre-miRNA-31, the highest stabilization of the structure when lowering the pH was observed for the C·A mismatch, with even less stabilization for the G·A mismatch. In the case of the A·A mismatch, we observed a slight destabilization of the structure.

Protonation of adenine bases is known to play an important role in many biological processes [17], [19], [39]–[42]. We found that A54 in the C·A mismatch is protonated in BottomB, even at physiological pH and temperature (**Fig. 2.10**). This base pair mismatch could impact the stability of the stem and thus could affect the regulation of the Dicer processing. Our studies show that adenine bases can adopt various conformations and have different protonation levels, which could have important biological implications. For example, in RNA editing by adenosine deaminases acting on RNA (ADAR) enzymes, an unpaired adenosine that is “flipped out” of the stem increases editing efficiency [69], [70]. Protein-RNA recognition can be modulated by RNA shape and formation of a C·A<sup>+</sup> wobble base pair can cause the helical axis of RNA to bend [71]. In the case of pre-miRNA-21, a dynamic equilibrium exists even at physiological conditions where the A<sup>+</sup>·G mismatch is formed, which enhanced the Dicer processing [17]. It was also strongly implied that the regulation of miRNA biogenesis can be modulated in response to environmental and cellular stimuli [17]. Furthermore, perturbation of the RNA duplex structure by mismatches and wobble base pairs has a negative effect on Drosha/DGCR8 processing [16]. Interestingly, these mismatches and wobble base pairs were not located near cleavage sites but were positioned 10–12 nucleotides downstream of the cleavage sites or 23–26 nt from the basal junction [16]. For transactivation response element RNA-binding protein (TRBP), the protein which recruits pre-miRNAs to Dicer for processing into mature miRNA, it was shown to discriminate between miRNAs based on their secondary structures. TRBP had a strong binding preference for pre-miRNAs, whose stem region had tight base-pairing except for the center region [72], [73].

Our studies contribute to our understanding of the chemical and conformational differences between A·A, G·A and C·A in the stem of the pre-miRNA-31. A54 in the C·A mismatch displayed the highest pH dependence as well as a population of protonated adenine bases even at pH=7.5.

Lowering the pH value resulted in the stabilization of the C·A mismatch, had only a slight effect on the G·A mismatch, and destabilized the A·A mismatch. Our study suggests that the dynamics of mismatches is connected to their pH sensitivity and might play a role in the regulation of pre-miRNA processing.

## 2.5 Reference

- [1] A. Zampetaki, A. Albrecht, and K. Steinhofel, “Long non-coding RNA structure and function: is there a link?,” *Front. Physiol.*, vol. 9, p. 1201, 2018, doi: 10.3389/fphys.2018.01201.
- [2] S. M. Hammond, “An overview of microRNAs,” *Adv. Drug Deliv. Rev.*, vol. 87, pp. 3–14, Jun. 2015, doi: 10.1016/j.addr.2015.05.001.
- [3] R. C. Friedman, K. K. Farh, C. B. Burge, and D. P. Bartel, “Most mammalian mRNAs are conserved targets of microRNAs,” *Genome Res*, vol. 19, no. 1, pp. 92–105, Jan. 2009, doi: 10.1101/gr.082701.108.
- [4] J. Shu, B. V. R. e Silva, T. Gao, Z. Xu, and J. Cui, “Dynamic and modularized microRNA regulation and its implication in human cancers,” *Sci. Rep.*, vol. 7, no. 1, p. 13356, Dec. 2017, doi: 10.1038/s41598-017-13470-5.
- [5] A. M. Ardekani and M. M. Naeini, “The role of microRNAs in human diseases,” *Avicenna J Med Biotechnol*, vol. 2, no. 4, pp. 161–79, Oct. 2010.
- [6] N. Bushati and S. M. Cohen, “microRNA functions,” *Annu Rev Cell Dev Biol*, vol. 23, pp. 175–205, 2007, doi: 10.1146/annurev.cellbio.23.090506.123406.
- [7] S. Lin and R. I. Gregory, “MicroRNA biogenesis pathways in cancer,” *Nat Rev Cancer*, vol. 15, no. 6, pp. 321–33, Jun. 2015, doi: 10.1038/nrc3932.
- [8] V. C. Auyeung, I. Ulitsky, S. E. McGeary, and D. P. Bartel, “Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing,” *Cell*, vol. 152, no. 4, pp. 844–858, Feb. 2013, doi: 10.1016/j.cell.2013.01.031.
- [9] S. Gu et al., “The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo,” *Cell*, vol. 151, no. 4, pp. 900–911, Nov. 2012, doi: 10.1016/j.cell.2012.09.042.



- [10] G. Mirihana Arachchilage, A. C. Dassanayake, and S. Basu, “A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation,” *Chem. Biol.*, vol. 22, no. 2, pp. 262–272, Feb. 2015, doi: 10.1016/j.chembiol.2014.12.013.
- [11] L. Pandolfini et al., “METTL1 Promotes let-7 MicroRNA processing via m7G methylation,” *Mol Cell*, vol. 74, no. 6, pp. 1278-1290 e9, Jun. 2019, doi: 10.1016/j.molcel.2019.03.040.
- [12] K. Nishikura, “Functions and regulation of RNA editing by ADAR deaminases,” *Annu. Rev. Biochem.*, vol. 79, no. 1, pp. 321–349, Jun. 2010, doi: 10.1146/annurev-biochem-060208-105251.
- [13] N. Fernandez, R. A. Cordiner, R. S. Young, N. Hug, S. Macias, and J. F. Cáceres, “Genetic variation and RNA structure regulate microRNA biogenesis,” *Nat. Commun.*, vol. 8, no. 1, p. 15114, Aug. 2017, doi: 10.1038/ncomms15114.
- [14] V. Castilla-Llorente, G. Nicastro, and A. Ramos, “Terminal loop-mediated regulation of miRNA biogenesis: selectivity and mechanisms,” *Biochem. Soc. Trans.*, vol. 41, no. 4, pp. 861–865, Aug. 2013, doi: 10.1042/BST20130058.
- [15] L. Zhu, S. K. Kandasamy, and R. Fukunaga, “Dicer partner protein tunes the length of miRNAs using base-mismatch in the pre-miRNA stem,” *Nucleic Acids Res.*, vol. 46, no. 7, pp. 3726–3741, Apr. 2018, doi: 10.1093/nar/gky043.
- [16] S. Li, T. D. Nguyen, T. L. Nguyen, and T. A. Nguyen, “Mismatched and wobble base pairs govern primary microRNA processing by human Microprocessor,” *Nat. Commun.*, vol. 11, no. 1, p. 1926, Dec. 2020, doi: 10.1038/s41467-020-15674-2.
- [17] J. T. Baisden, J. A. Boyer, B. Zhao, S. M. Hammond, and Q. Zhang, “Visualizing a protonated RNA state that modulates microRNA-21 maturation,” *Nat Chem Biol*, vol. 17, no. 1,

pp. 80–88, Jan. 2021, doi: 10.1038/s41589-020-00667-5.

[18] W. Saenger, *Principles of nucleic acid structure*. New York, NY: Springer New York, 1984. doi: 10.1007/978-1-4612-5190-3.

[19] J. L. Wilcox, A. K. Ahluwalia, and P. C. Bevilacqua, “Charged nucleobases and their potential for RNA catalysis,” *Acc. Chem. Res.*, vol. 44, no. 12, pp. 1270–1279, Dec. 2011, doi: 10.1021/ar2000452.

[20] N. A. Stepicheva and J. L. Song, “Function and regulation of microRNA-31 in development and disease,” *Mol Reprod Dev*, vol. 83, no. 8, pp. 654–74, Aug. 2016, doi: 10.1002/mrd.22678.

[21] S. N. Reilly et al., “Up-regulation of miR-31 in human atrial fibrillation begets the arrhythmia by depleting dystrophin and neuronal nitric oxide synthase,” *Sci Transl Med*, vol. 8, no. 340, p. 340ra74, May 2016, doi: 10.1126/scitranslmed.aac4296.

[22] D. Li et al., “MicroRNA-31 promotes skin wound healing by enhancing keratinocyte proliferation and migration,” *J. Invest. Dermatol.*, vol. 135, no. 6, pp. 1676–1685, Jun. 2015, doi: 10.1038/jid.2015.48.

[23] T. Yu, P. Ma, D. Wu, Y. Shu, and W. Gao, “Functions and mechanisms of microRNA-31 in human cancers,” *Biomed. Pharmacother.*, vol. 108, pp. 1162–1169, Dec. 2018, doi: 10.1016/j.biopha.2018.09.132.

[24] J. Shi et al., “miR-31 mediates inflammatory signaling to promote re-epithelialization during skin wound healing,” *J. Invest. Dermatol.*, vol. 138, no. 10, pp. 2253–2263, Oct. 2018, doi: 10.1016/j.jid.2018.03.1521.

[25] S. Griffiths-Jones, “miRBase: microRNA sequences, targets and gene nomenclature,” *Nucleic Acids Res.*, vol. 34, no. 90001, pp. D140–D144, Jan. 2006, doi: 10.1093/nar/gkj112.

- [26] S. Bellaousov, J. S. Reuter, M. G. Seetin, and D. H. Mathews, "RNAstructure: web servers for RNA secondary structure prediction and analysis," *Nucleic Acids Res.*, vol. 41, no. W1, pp. W471–W474, Jul. 2013, doi: 10.1093/nar/gkt290.
- [27] C. Kao, M. Zheng, and S. Rudisser, "A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase," *RNA*, vol. 5, no. 9, pp. 1268–72, Sep. 1999, doi: 10.1017/s1355838299991033.
- [28] P. R. Cunningham and J. Ofengand, "Use of inorganic pyrophosphatase to improve the yield of in vitro transcription reactions catalyzed by T7 RNA polymerase," *Biotechniques*, vol. 9, no. 6, pp. 713–4, Dec. 1990.
- [29] Glen Research, LLC, "Products for DNA research user guide to Glen-Pak™ purification."
- [30] M. Norris, B. Fetler, J. Marchant, and B. A. Johnson, "NMRFx Processor: a cross-platform NMR data processing program," *J Biomol NMR*, vol. 65, no. 3–4, pp. 205–216, Aug. 2016, doi: 10.1007/s10858-016-0049-6.
- [31] B. A. Johnson and R. A. Blevins, "NMR View: A computer program for the visualization and analysis of NMR data," *J Biomol NMR*, vol. 4, no. 5, pp. 603–14, Sep. 1994, doi: 10.1007/BF00404272.
- [32] M. R. Willcott, "MestRe Nova," *J. Am. Chem. Soc.*, vol. 131, no. 36, pp. 13180–13180, Sep. 2009, doi: 10.1021/ja906709t.
- [33] D. S. Wishart et al., "<sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR," *J Biomol NMR*, vol. 6, no. 2, pp. 135–40, Sep. 1995, doi: 10.1007/BF00211777.
- [34] P. Thaplyal and P. C. Bevilacqua, "Experimental approaches for measuring pK<sub>a</sub>'s in RNA and DNA," in *Methods in Enzymology*, vol. 549, Elsevier, 2014, pp. 189–219. doi: 10.1016/B978-0-12-801122-5.00009-X.

- [35] K. A. Leamy, N. H. Yennawar, and P. C. Bevilacqua, “Cooperative RNA folding under cellular conditions arises from both tertiary structure stabilization and secondary structure destabilization,” *Biochemistry*, vol. 56, no. 27, pp. 3422–3433, Jul. 2017, doi: 10.1021/acs.biochem.7b00325.
- [36] F. M. Jucker, H. A. Heus, P. F. Yip, E. H. Moors, and A. Pardi, “A network of heterogeneous hydrogen bonds in GNRA tetraloops,” *J Mol Biol*, vol. 264, no. 5, pp. 968–80, Dec. 1996, doi: 10.1006/jmbi.1996.0690.
- [37] Y. Liu et al., “NMR chemical shift assignments of RNA oligonucleotides to expand the RNA chemical shift database,” *Biomol. NMR Assign.*, vol. 15, no. 2, pp. 479–490, Oct. 2021, doi: 10.1007/s12104-021-10049-0.
- [38] E. L. Ulrich et al., “BioMagResBank,” *Nucleic Acids Res*, vol. 36, no. Database issue, pp. D402-8, Jan. 2008, doi: 10.1093/nar/gkm957.
- [39] A. C. Wolter et al., “A stably protonated adenine nucleotide with a highly shifted pK a value stabilizes the tertiary structure of a GTP-binding RNA aptamer,” *Angew. Chem. Int. Ed.*, vol. 56, no. 1, pp. 401–404, Jan. 2017, doi: 10.1002/anie.201609184.
- [40] T. N. Malik, E. E. Doherty, V. M. Gaded, T. M. Hill, P. A. Beal, and R. B. Emeson, “Regulation of RNA editing by intracellular acidification,” *Nucleic Acids Res.*, vol. 49, no. 7, pp. 4020–4036, Apr. 2021, doi: 10.1093/nar/gkab157.
- [41] E. Fuchs, C. Falschlunger, R. Micura, and K. Breuker, “The effect of adenine protonation on RNA phosphodiester backbone bond cleavage elucidated by deaza-nucleobase modifications and mass spectrometry,” *Nucleic Acids Res.*, vol. 47, no. 14, pp. 7223–7234, Aug. 2019, doi: 10.1093/nar/gkz574.
- [42] B. Houck-Loomis et al., “An equilibrium-dependent retroviral mRNA switch regulates

translational recoding,” *Nature*, vol. 480, no. 7378, pp. 561–4, Nov. 2011, doi: 10.1038/nature10657.

[43] J. P. Rife, C. S. Cheng, P. B. Moore, and S. A. Strobel, “N<sup>2</sup>-Methylguanosine is iso-energetic with guanosine in RNA duplexes and GNRA tetraloops,” *Nucleic Acids Res.*, vol. 26, no. 16, pp. 3640–3644, Aug. 1998, doi: 10.1093/nar/26.16.3640.

[44] T. Xia et al., “Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs,” *Biochemistry*, vol. 37, no. 42, pp. 14719–14735, Oct. 1998, doi: 10.1021/bi9809425.

[45] Y. Yuan, D. J. Kerwood, A. C. Paoletti, M. F. Shubsda, and P. N. Borer, “Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 × 3 internal loop ,” *Biochemistry*, vol. 42, no. 18, pp. 5259–5269, May 2003, doi: 10.1021/bi034084a.

[46] J. L. Battiste et al., “ $\alpha$  Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex,” *Science*, vol. 273, no. 5281, pp. 1547–1551, Sep. 1996, doi: 10.1126/science.273.5281.1547.

[47] C. Oubridge, N. Ito, P. R. Evans, C.-H. Teo, and K. Nagai, “Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin,” *Nature*, vol. 372, no. 6505, pp. 432–438, Dec. 1994, doi: 10.1038/372432a0.

[48] F. H.-T. Allain, C. C. Gubser, P. W. A. Howe, K. Nagai, D. Neuhaus, and G. Varani, “Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation,” *Nature*, vol. 380, no. 6575, pp. 646–650, Apr. 1996, doi: 10.1038/380646a0.

[49] H. Mao, S. A. White, and J. R. Williamson, “A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex,” *Nat. Struct. Biol.*, vol. 6, no. 12, pp. 1139–1147, Dec. 1999, doi: 10.1038/70081.

- [50] T. Hori, Y. Taguchi, S. Uesugi, and Y. Kurihara, "The RNA ligands for mouse proline-rich RNA-binding protein (mouse Prpp) contain two consensus sequences in separate loop structure," *Nucleic Acids Res.*, vol. 33, no. 1, pp. 190–200, Jan. 2005, doi: 10.1093/nar/gki153.
- [51] V. V. Pham et al., "HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry," *Nat Commun*, vol. 9, no. 1, p. 4266, Oct. 2018, doi: 10.1038/s41467-018-06591-6.
- [52] K. J. Baeyens, H. L. De Bondt, A. Pardi, and S. R. Holbrook, "A curved RNA helix incorporating an internal loop with G•A and A•A non-Watson-Crick base pairing," *Proc. Natl. Acad. Sci.*, vol. 93, no. 23, pp. 12851–12855, Nov. 1996, doi: 10.1073/pnas.93.23.12851.
- [53] L. X. Shen, Z. Cai, and I. Tinoco, "RNA structure at high resolution," *FASEB J.*, vol. 9, no. 11, pp. 1023–1033, Aug. 1995, doi: 10.1096/fasebj.9.11.7544309.
- [54] J. Cavarelli, B. Rees, M. Ruff, J.-C. Thierry, and D. Moras, "Yeast tRNA Asp recognition by its cognate class II aminoacyl-tRNA synthetase," *Nature*, vol. 362, no. 6416, pp. 181–184, Mar. 1993, doi: 10.1038/362181a0.
- [55] M. E. Burkard, D. H. Turner, and I. Tinoco, "10 The interactions that shape RNA structure," *Cold Spring Harb. Monogr. Arch.*, vol. 37, pp. 233–264, 1999.
- [56] B. Hoffmann et al., "NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site," *Proc. Natl. Acad. Sci.*, vol. 100, no. 12, pp. 7003–7008, Jun. 2003, doi: 10.1073/pnas.0832440100.
- [57] A. A. Szewczak and T. R. Cech, "An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis," *RNA N. Y. N.*, vol. 3, no. 8, pp. 838–849, Aug. 1997.
- [58] M. D. Canny, F. M. Jucker, E. Kellogg, A. Khvorova, S. D. Jayasena, and A. Pardi, "Fast cleavage kinetics of a natural hammerhead ribozyme," *J. Am. Chem. Soc.*, vol. 126, no. 35, pp.

10848–10849, Sep. 2004, doi: 10.1021/ja046848v.

[59] L. R. Ganser, C.-C. Chu, H. P. Bogerd, M. L. Kelly, B. R. Cullen, and H. M. Al-Hashimi, “Probing RNA conformational equilibria within the functional cellular context,” *Cell Rep.*, vol. 30, no. 8, pp. 2472–2480.e4, Feb. 2020, doi: 10.1016/j.celrep.2020.02.004.

[60] X. Sun, Q. Zhang, and H. M. Al-Hashimi, “Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg<sup>2+</sup> binding: role in the kissing–duplex structural transition,” *Nucleic Acids Res.*, vol. 35, no. 5, pp. 1698–1713, Mar. 2007, doi: 10.1093/nar/gkm020.

[61] S.-F. Torabi et al., “Structural analyses of an RNA stability element interacting with poly(A),” *Proc. Natl. Acad. Sci.*, vol. 118, no. 14, p. e2026656118, Apr. 2021, doi: 10.1073/pnas.2026656118.

[62] W. Ritchie, M. Legendre, and D. Gautheret, “RNA stem–loops: To be or not to be cleaved by RNase III,” *RNA*, vol. 13, no. 4, pp. 457–462, Apr. 2007, doi: 10.1261/rna.366507.

[63] I. Calin-Jageman and A. W. Nicholson, “Mutational analysis of an RNA internal loop as a reactivity epitope for *Escherichia coli* ribonuclease III substrates,” *Biochemistry*, vol. 42, no. 17, pp. 5025–5034, May 2003, doi: 10.1021/bi030004r.

[64] T. L. Nguyen, T. D. Nguyen, S. Bao, S. Li, and T. A. Nguyen, “The internal loops in the lower stem of primary microRNA transcripts facilitate single cleavage of human Microprocessor,” *Nucleic Acids Res.*, vol. 48, no. 5, pp. 2579–2593, Mar. 2020, doi: 10.1093/nar/gkaa018.

[65] W.-C. Lee, S.-H. Lu, M.-H. Lu, C.-J. Yang, S.-H. Wu, and H.-M. Chen, “Asymmetric bulges and mismatches determine 20-nt microRNA formation in plants,” *RNA Biol.*, vol. 12, no. 9, pp. 1054–1066, Sep. 2015, doi: 10.1080/15476286.2015.1079682.

[66] A. R. Davis, C. C. Kirkpatrick, and B. M. Znosko, “Structural characterization of naturally

occurring RNA single mismatches,” *Nucleic Acids Res.*, vol. 39, no. 3, pp. 1081–1094, Feb. 2011, doi: 10.1093/nar/gkq793.

[67] A. R. Davis and B. M. Znosko, “Thermodynamic characterization of single mismatches found in naturally occurring RNA,” *Biochemistry*, vol. 46, no. 46, pp. 13425–13436, Nov. 2007, doi: 10.1021/bi701311c.

[68] J. L. Wilcox and P. C. Bevilacqua, “pK<sub>a</sub> Shifting in double-stranded RNA is highly dependent upon nearest neighbors and bulge positioning,” *Biochemistry*, vol. 52, no. 42, pp. 7470–7476, Oct. 2013, doi: 10.1021/bi400768q.

[69] G. Shevchenko and K. V. Morris, “All I’s on the RADAR: role of ADAR in gene regulation,” *FEBS Lett.*, vol. 592, no. 17, pp. 2860–2873, Sep. 2018, doi: 10.1002/1873-3468.13093.

[70] S. K. Wong, S. Sato, and D. W. Lazinski, “Substrate recognition by ADAR1 and ADAR2,” *RNA*, vol. 7, no. 6, pp. 846–858, Jun. 2001, doi: 10.1017/S135583820101007X.

[71] B. Pan, S. N. Mitra, and M. Sundaralingam, “Structure of a 16-mer RNA duplex r(GCAGACUUAAAUCUGC)<sub>2</sub> with wobble C·A<sup>+</sup> mismatches,” *J. Mol. Biol.*, vol. 283, no. 5, pp. 977–984, Nov. 1998, doi: 10.1006/jmbi.1998.2140.

[72] T. Takahashi et al., “LGP2 virus sensor regulates gene expression network mediated by TRBP-bound microRNAs,” *Nucleic Acids Res.*, vol. 46, no. 17, pp. 9134–9147, Sep. 2018, doi: 10.1093/nar/gky575.

[73] T. Yoshida, Y. Asano, and K. Ui-Tei, “Modulation of microRNA processing by Dicer via its associated dsRNA binding proteins,” *Non-Coding RNA*, vol. 7, no. 3, p. 57, Sep. 2021, doi: 10.3390/ncrna7030057.

[74] P. Z. Johnson, W. K. Kasprzak, B. A. Shapiro, and A. E. Simon, “RNA2Drawer:



geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences,” *RNA Biol.*, vol. 16, no. 12, pp. 1667–1671, Dec. 2019, doi: 10.1080/15476286.2019.1659081.

## CHAPTER III Structure of Pre-miR-31 Reveals an Active Role in Dicer/TRBP Complex Processing

This chapter is a modified version of a published article with authors listed as follows:

Ma, S., Kotar, A., Hall, I., Grote, S., Rouskin, S., & Keane, S. C. (2023). Structure of pre-miR-31 reveals an active role in Dicer–TRBP complex processing. *Proceedings of the National Academy of Sciences*, 120(39), e2300527120.

**Author contributions:** S.M., A.K., and S.C.K. designed research; S.M., A.K., and I.H. performed research; S.M., A.K., I.H., S.G., S.R., and S.C.K. analyzed data; and S.M., A.K., I.H., and S.C.K. wrote the paper.

NMR experiments and analyses in Figure 3.1, Figure 3.2, Appendix A Figure S2-S10, Appendix A Figure S12-S17, Appendix A Table S1, Appendix A Table S2 were performed by A.K.

SAXS experiments and analyses in Appendix A Figure S15 and Appendix A Table S3 were performed by I.H.

DMS-Mapseq data analysis for Figure S1 and Appendix A Figure S1 was performed by S.G.

All other work was done by S.M.

**Abstract:** As an essential post-transcriptional regulator of gene expression, microRNA (miRNA) levels must be strictly maintained. The biogenesis of many miRNAs is mediated by trans-acting protein partners through a variety of mechanisms, including remodeling of the RNA structure. miR-31 functions as an oncogene in numerous cancers and interestingly, its biogenesis is not known to be regulated by protein binding partners. Therefore, the intrinsic structural properties of the precursor element of miR-31 (pre-miR-31) can provide a mechanism by which its biogenesis is regulated. We determined the solution structure of pre-miR-31 to investigate the role of distinct structural elements in regulating processing by the Dicer-TRBP complex. We found that the presence or absence of mismatches within the helical stem do not strongly influence Dicer-TRBP processing of the pre-miRNAs. However, both the apical loop size and structure at the Dicing site are key elements for discrimination by the Dicer-TRBP complex. Interestingly, our NMR-derived structure reveals the presence of a triplet of base pairs that link the Dicer cleavage site and the apical loop. Mutational analysis in this region suggests that the stability of the junction region strongly influences processing by the Dicer-TRBP complex. Our results enrich our understanding of the active role that RNA structure plays in regulating miRNA biogenesis, which has direct implications for the control of gene expression.

### **3.1 Introduction**

MicroRNAs (miRNAs) regulate protein gene expression post-transcriptionally through base pairing with target messenger (m) RNAs to trigger mRNA degradation or translational suppression (1-4). In the nucleus, RNA polymerase II transcribes primary microRNAs (pri-miRNAs) which are subsequently processed by Microprocessor, a protein complex of Drosha and DiGeorge syndrome critical region 8 (DGCR8). Pri-miRNA processing generates precursor

microRNAs (pre-miRNAs) that are exported from the nucleus to the cytoplasm for further processing by Dicer, which functions in complex with transactivation response element RNA-binding protein (TRBP) (5-7). Dicer-TRBP processing results in the production of 21-22 nucleotide (nt) mature miRNA duplexes (4, 8). Mature miRNAs function in concert with Argonaute (Ago) protein to form the miRNA-induced silencing complex, that is responsible for mRNA degradation or translational suppression (2, 4).

Distinctive regulatory elements for pri-miRNAs and pre-miRNAs have been discovered over the past several decades. These elements include specific sequences within the pri-miRNAs and pre-miRNAs that recruit regulatory proteins (9-12). While protein-mediated regulation is indeed important for many pre-miRNAs (13, 14), structural features of pri- and pre-miRNAs can also mediate enzymatic processing in the absence of protein binding (15-21). Therefore, the intrinsic structural properties of pri- and pre-miRNAs may serve as an alternative mechanism for regulation of their biogenesis, suggesting that the RNA is not a passive element in the miRNA biogenesis pathway.

MicroRNA-31 (miR-31) acts as oncogene in multiple cancers. Upregulation of miR-31 in cells is associated with cancer proliferation, anti-apoptosis and migration in multiple cancers by targeting different pathways (22). Interestingly, no protein binding partners have been identified for pre-miR-31 (23), suggesting that the mechanisms for regulating miR-31 biogenesis may be encoded at the RNA level. We sought to examine the RNA structural features that may contribute to the post-transcriptional regulation of pre-miR-31.

Here, we describe the three-dimensional structure of pre-miR-31 and characterized how secondary structural elements throughout pre-miR-31 affect Dicer-TRBP processing. The structure presented in this work is the first full-length pre-miRNA structure determined and

significantly adds to the limited known structures of pre-miRNAs (24, 25). We found that modulating the structure of pre-miR-31 at the dicing site by enlarging the internal loop reduced the rate of Dicer-TRBP processing. Furthermore, we demonstrate that pre-miR-31 RNAs with extended junction regions, which restrict the apical loop size, displayed significantly reduced processing. Whereas pre-miR-31 constructs with large apical loops had near wild type (WT)-like levels of processing. Interestingly, in the absence of TRBP, pre-miR-31 RNAs with large apical loops were poor substrates for Dicer processing. These results suggest that the loop size must be tightly controlled, as too small of an apical loop can inhibit pre-miR-31 maturation, a restriction that can be overcome to some extent by addition of TRBP.

Finally, we found that the junction region functions exquisitely to maximize efficient processing. We note differences in the secondary structure models derived from nuclear magnetic resonance (NMR) spectroscopy and chemical probing in the junction region. Rather than viewing these structures as incompatible, we demonstrate that both structures likely exist in a dynamic equilibrium where the base paired junction transiently samples the open conformation. Our data are consistent with a model in which RNAs can self-regulate their processing in the absence of trans-acting RNA-binding proteins. Recent studies demonstrate the importance of pre-miRNA structural plasticity in regulating their enzymatic processing (20, 21). Our research cements the hypothesis that pre-miRNA structure can regulate its maturation process and further informs on structural features necessary for effective short hairpin (sh) RNA design.

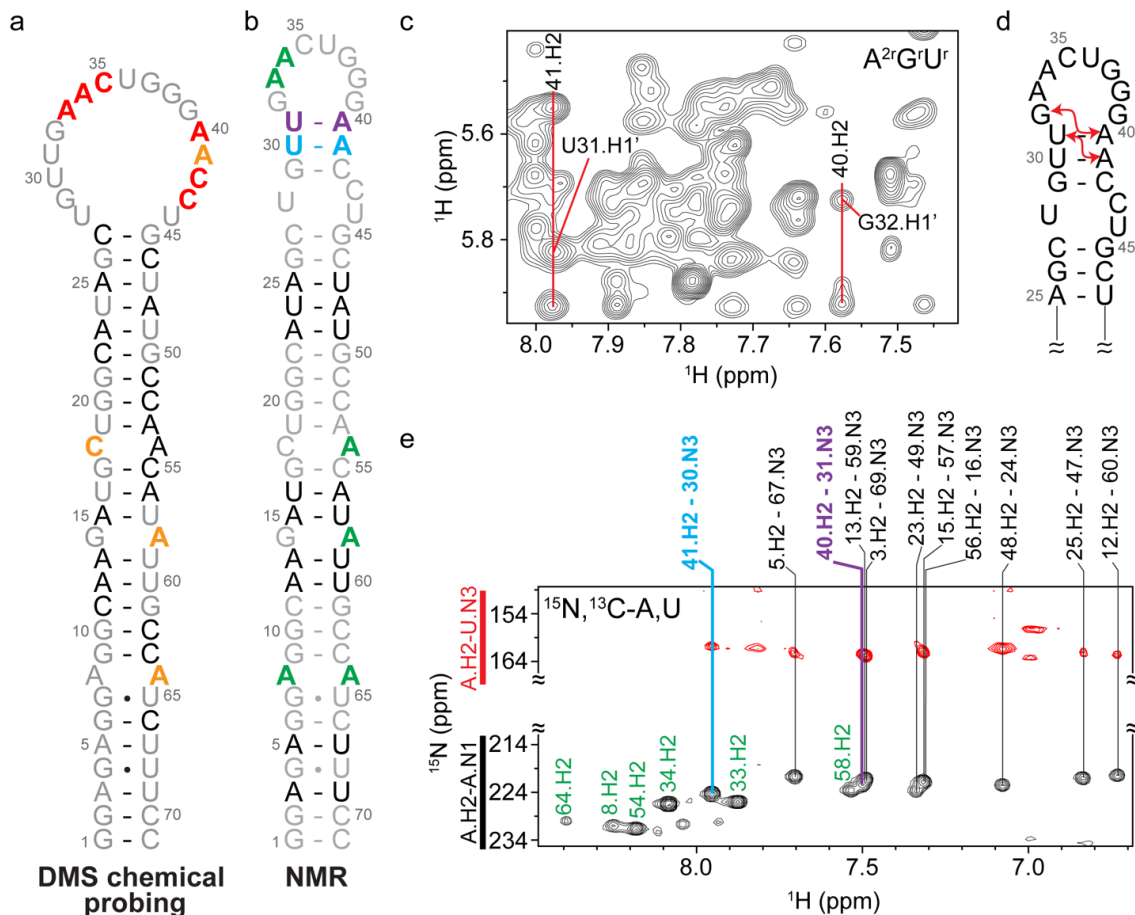
## 3.2 Results

### 3.2.1 The secondary structure of FL-pre-miR-31 contains three mismatches in the helical stem and three base pairs in the apical loop.

The lowest free energy secondary structure of the 71-nt long full length (FL) pre-miR-31 predicted by the RNAstructure webserver (26) is a hairpin composed of three mismatches (A•A, G•A and C•A) in the stem region, a 1x2 internal loop (dicing site), and three base pairs forming the junction region between the internal and apical loops. However, recent *in cell* selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) chemical probing studies (27) revealed that the apical loops of pre-miRNAs are less structured than predicted in the miRbase (28-33). To evaluate the secondary structure of FL pre-miR-31, we performed *in vitro* dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq) (34, 35). The chemical probing derived topology of the entire stem region including the three mismatches is in complete agreement with prediction (**Fig. 3.1a, Appendix A Fig. S1**). However, our *in vitro* chemical probing data suggests that residues within and near the predicted apical loop (A33, A34, C35, A40, A41, C42, and C43) are highly reactive, consistent with these residues being unpaired and forming a large, open apical loop structure (**Fig. 3.1a, Appendix A Fig. S1**).

To better understand the molecular details of the pre-miR-31 hairpin, we determined the solution structure of FL pre-miR-31 using NMR spectroscopy. We used a divide-and-conquer approach to facilitate resonance assignments of FL pre-miR-31 (**Fig. Appendix A S2**). To guide assignments of the FL pre-miR-31 RNA, we combined our previously reported chemical shift assignments for fragments BottomA and BottomB (36) with chemical shift assignments for two additional oligo fragments, TopA (**Appendix A Fig. S3**) and Top (**Appendix A Fig. S4**). However, the large molecular size of FL pre-miR-31 resulted in a severely crowded spectrum, preventing

direct assignments based only on the oligo controls. To better resolve the complex two-dimensional (2D)  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of FL pre-miR-31, we employed a deuterium-edited approach (Appendix A Fig. S5) (37-39). The combination of methods allowed for the nearly complete assignment of non-exchangeable aromatic and ribose (C1', C2', and C3') protons in FL pre-miR-31 (Appendix A Fig. S6, Appendix A Table S1).



**Figure 3.1. Conflicting secondary structure models for pre-miR-31 apical loop. a)** Secondary structure derived from *in vitro* DMS-MapSeq where coloring denotes reactivity of given bases. Red=high reactivity, orange=medium reactivity, black=low reactivity, gray=no data available. **b)** Secondary structure derived from NMR characterization. Coloring is based on identification of A-U base pairs (see panel e). **c)** Portion of a 2D  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of an  $\text{A}^{2r}\text{G}^r\text{U}^r$ -labeled FL pre-miR-31. Adenosine cross-strand NOEs consistent with helical stacking in the junction region are indicated. **d)** Secondary structure of the apical loop region highlighting NOEs noted in c with red arrows. **e)** Best-selective long-range HNN-COSY spectrum identifying A-U base pairs within FL pre-miR-31. Black peaks are adenosine H2-N1 correlations, red peaks are adenosine H2-uracil N3 correlations. Vertical lines indicate the detection of A-U base pairs. Unpaired adenosines are denoted in green, A-U base pairs in the stem region are denoted in black, junction A-U base pairs are denoted in cyan and purple.

The NMR-derived secondary structure of FL pre-miR-31 (**Fig. 3.1b**) is consistent with the predicted lowest free energy structure. We were particularly interested in the structural features of the apical loop of FL pre-miR-31. Analysis of the  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of an  $\text{A}^{2r}\text{G}^r\text{U}^r$ -labeled (adenosine C2 and ribose of adenosine, guanosine and uridine residues are protonated, all other sites deuterated) FL pre-miR-31, revealed strong cross-strand NOEs between A41.H2-U31.H1' and A40.H2-G32.H1' (**Fig. 3.1 c,d**), consistent with a typical A-helical structure in this region. To further explore the base pairing within FL pre-miR-31 we acquired a best selective long-range HNN-COSY (40), which allows for identification of A-U base pairs via detection on the non-exchangeable adenosine C-2 proton rather than detection of the labile imino proton (**Appendix A Fig. S7, S8**). Here, we see clear evidence for 9 of the 10 expected A-U base pairs within the stem on pre-miR-31 (**Fig. 3.1e**). The resonance for A53 is broadened beyond detection at pH = 7.5, likely due to the dynamics of the neighboring C18•A54 mismatch. Furthermore, we observe two additional A.H2-U.N3 signals, which correspond to A41-U30 and A40-U31 base pairs (**Fig. 3.1e**). While A40 and A41 were highly reactive to DMS, and therefore predicted to be unpaired, we provide direct spectroscopic evidence of base pairing within the apical loop.

Consistent with the NMR-derived secondary structure, a pH titration reveals that unpaired residues A8, A54, A64 (mismatches in the helical stem), and A34 (apical loop) are sensitive to the changes in the pH of the solution (**Appendix A Fig. S9**). In contrast, the changes of chemical shifts of A40 and A41 (junction) are notably smaller and resemble those measured for base-paired residues from the stem. Additionally, solvent paramagnetic relaxation enhancement (sPRE) data, which reports on the solvent accessibility of FL pre-miR-31, revealed that G29 and A41 do not show large sPRE values (**Appendix A Fig. S10**) compared to A33, A34, G37 and G38, which are unpaired in the apical loop. Interestingly, for A40 we observe much higher sPRE value indicating

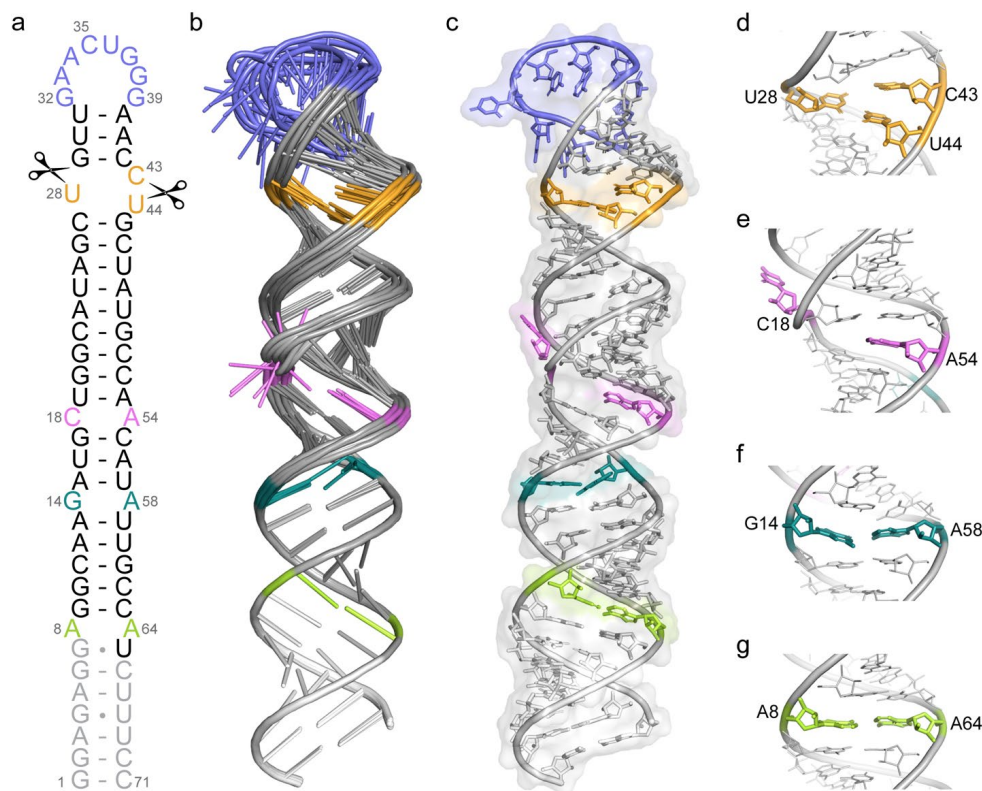


high solvent accessibility of the U31-A40 base pair. These observations suggest that U31-A40 may be a nucleation point for opening the junction based on environmental changes. The sequence of pre-miR-31 is highly conserved in mammals, with mutations or deletions present only in the apical loop region (**Appendix A Fig. S11**). Collectively, our results support the presence of a short base paired element in the junction below the apical loop.

### 3.2.2 Tertiary structure of pre-miR-31

To further our structure-based studies, we determined the three-dimensional structure of FL pre-miR-31 (**Fig. 3.2, Appendix A Table S2**). FL pre-miR-31 adopts a largely elongated hairpin structure, with three base pair mismatches within the helical stem. Nuclear Overhauser effect (NOE) data are consistent with A-helical stacking of 29-GUU-31 and 40-AAC-42, with strong NOEs between A41.H2-U31.H1' and A40.H2-G32.H1' (**Fig. 3.1c, Appendix A Fig. S12**). The HNN-COSY (**Fig. 3.1e**) further defines the base pairing within this region, cinching the apical loop structure and limiting the size of the apical loop to 8 nucleotides. The Dicer-TRBP processing site resides within a 1x2 internal loop containing U28, C43, and U44 (**Fig. 3.2d, Appendix A Fig. S13**). U28 and U44 are co-planar and adopt a *cis* Watson-Crick/Watson-Crick wobble geometry with C43 positioned above U44. We observed a strong NOE between A54.H2 and U19.H1', which positions A54 stacked in an A-helical geometry (**Fig. 3.2e, Appendix A Fig. S13**). No NOEs were observed linking C18 with neighboring residues, therefore C18 was unrestrained in structure calculations and can sample many conformations (**Fig. 3.2b**). No defined NOEs were observed connecting A13 with G14. However, aromatic-aromatic and aromatic-anomeric NOEs position G14 stacked under A15. G14 and A58 have the potential to form a *cis* Watson-Crick/Watson-Crick base pair (**Fig. 3.2f, Appendix A Fig. S13**). The A8•A64 mismatch is well-defined with sequential

and cross-strand NOEs (Fig. 3.2g, Appendix A Fig. S13). The structure was refined using global residual dipolar coupling (RDC) restraints and we observed a strong correlation between experimentally determined and back-calculated residual dipolar couplings (Appendix A Fig. S14). Furthermore, we observe strong agreement between the refined NMR solution structure and scattering data obtained using small angle X-ray scattering (SAXS) (Appendix A Fig. S15, Appendix A Table S3). Importantly, the pre-miR-31 structure is not sensitive to the nature of the monovalent (Appendix A Fig. S16) or divalent (Appendix A Fig. S17) cation in solution (*vide infra*).



**Figure 3.2. Tertiary structure of pre-miR-31.** **a)** NMR-derived secondary structure of FL-pre-miR-31. Dicer cleavage sites are indicated with scissors. Gray nucleotides were included in structural studies but are not present in a Dicing-competent WT pre-miR-31. **b)** Ensemble of 10 lowest energy structures after RDC refinement superimposed over residues 1-13 and 59-71. **c)** Lowest energy structure of pre-miR-31 with a transparent surface rendering. **d)** Enlarged view of the dicing site, colored orange. **e)** Enlarged view of the C•A mismatch, colored pink. **f)** enlarged view of the G•A mismatch, colored teal. **g)** Enlarged view of the A•A mismatch, colored green.

### 3.2.3 TRBP inhibits Dicer processing of pre-miR-31.

In the cell, TRBP is an important co-factor of Dicer that promotes substrate recognition and processing (5, 7, 41, 42), and also contributes to the accuracy of Dicer processing (5, 43). Interestingly, while addition of TRBP generally stimulated the rate of pre-miRNA processing (6), for pre-miR-31, the addition of TRBP reduced the rate of Dicing (42). We observed a similar ~3-fold reduction in the apparent rate constant ( $k_{\text{obs}}$ ) for pre-miR-31 Dicing in the presence of TRBP (**Appendix A Fig. S18, Appendix A Tables S4, S5**).

### 3.2.4 Mismatches within the helical stem region have no impact on Dicer-TRBP cleavage.

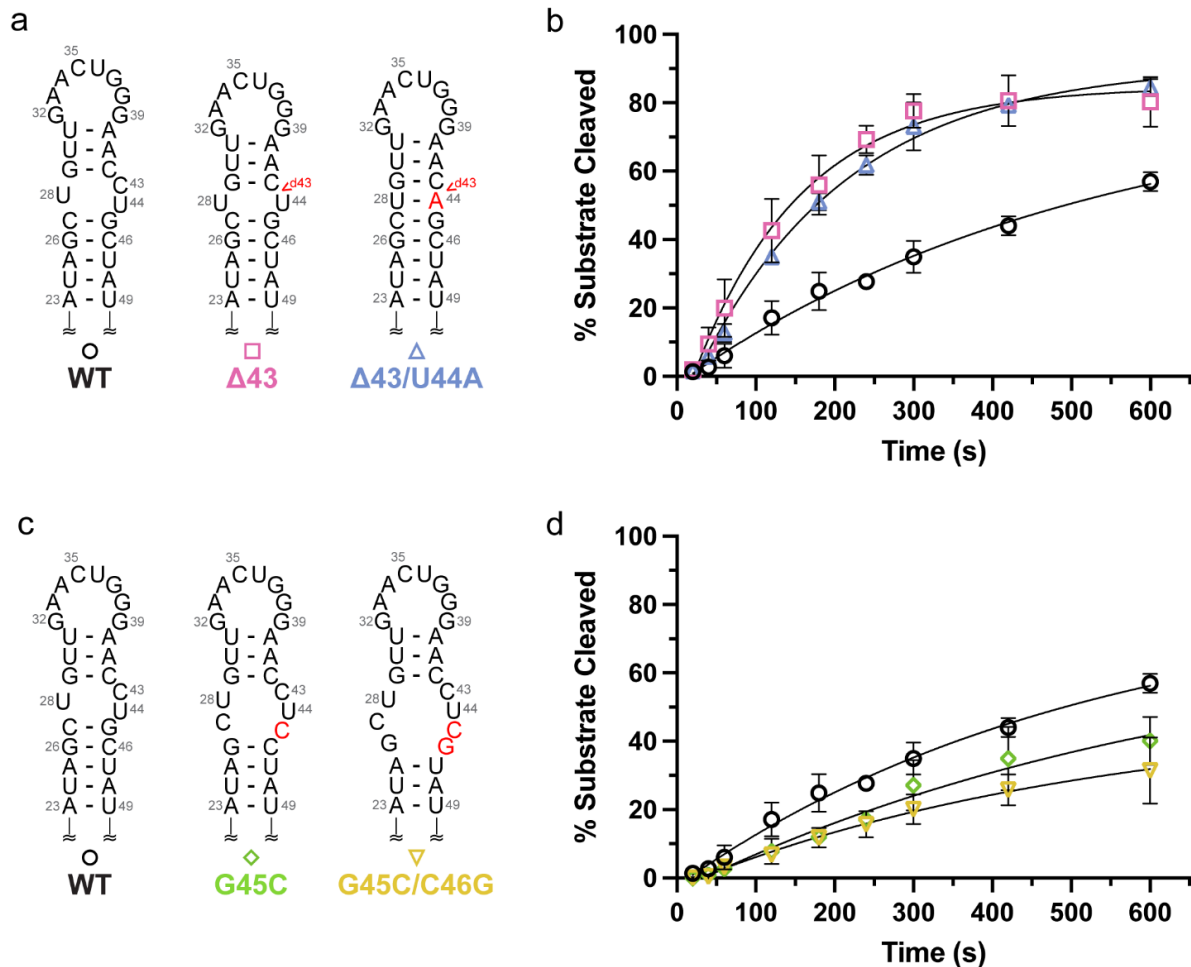
Base pair mismatches are a common feature within the helical stem of pre-miRNAs (36, 44). Studies on fly Dicer-1 suggest that while the length of the pre-miRNA helical stem is important, the presence of mismatches does not significantly affect Dicer processing (16). However, because pre-miR-31 biogenesis does not appear to be regulated by protein binding partners, we considered all aspects of pre-miR-31 structure that could be involved in regulating processing, including individual base pair mismatches.

We first designed a series of mutations that stabilized each mismatch individually and assessed their processing by the Dicer-TRBP complex. Conversion of the G14•A58 mismatch into a canonical U-A base pair (G14U, **Appendix A Fig. S19**) exhibits WT rates and levels of processing (**Appendix A Table S4, Appendix A Fig. S19**). We previously investigated the pH-dependence of the C18•A54 mismatch and found that A54 is partially protonated at physiological pH, suggesting the potential formation of a C•A<sup>+</sup> base pair near neutral pH (36). We were therefore interested in testing if mutations that replaced the mismatch with a canonical U-A or C-G base pair (C18U and A54G, respectively) affected the processing by Dicer-TRBP (**Appendix A Fig. S19**).

As with stabilization of the G•A mismatch, stabilization of the C•A mismatch did not affect the apparent rate of Dicer-TRBP processing or production of mature product (**Appendix A Table S4, Appendix A Fig. S19**). To fully disrupt the C•A mismatch, we engineered an A•A mismatch at this position (C18A, **Appendix A Fig. S19**). This mutant behaved similarly to other stem-stabilized mutations (**Appendix A Table S4, Appendix A Fig. S19**).

All pre-miR-31 stem mutant RNAs were cleaved to a similar extent (approximately 55%) and with a similar apparent rate constant (**Appendix A Table S4**), consistent with studies on fly Dicer-1(16). As described above, the addition of TRBP reduces the pre-miR-31 processing by Dicer (**Appendix A Fig. S18**). Therefore, we examined the processing of stem-mutations with Dicer alone, to discern the role of TRBP for substrate recognition. We found that the four individual stem mutants (C18A, C18U, A54G, G14U) were processed similarly to WT-pre-miR-31 (**Appendix A Fig. S20, Appendix A Table S5**), consistent with our findings with the Dicer-TRBP complex. Additionally, we examined the Dicer processing of a fully base paired double mutant (G14U/A54G) that stabilized both mismatches with canonical base pairs. This fully-base paired pre-miRNA was processed similarly to WT-pre-miR-31 (**Appendix A Fig. S20, Appendix A Table S5**). To determine if the context of the C•A mismatch was important for Dicing, we swapped the bases (18ACsw) to create an A•C mismatch. Again, we observed no significant change in Dicer processing (**Appendix A Fig. S20**).

Collectively, the processing of pre-miR-31 stem mutants was reduced by addition of TRBP, consistent with our observations with WT-pre-miR-31 (**Appendix A Fig. S18**). However, the stem mutations have no effect on Dicer (**Appendix A Fig. S20, Appendix A Table S5**) or Dicer-TRBP (**Appendix A Fig. S19, Appendix A Table S4**) processing, relative to WT.



**Figure 3.3. Structure at the dicing site serves as an important feature for Dicer-TRBP processing.** **a)** Predicted secondary structures of constructs designed to minimize the internal loop at the dicing site. Mutations are indicated with red lettering. **b)** Minimization of the internal loop at the Dicing site enhances the processing by the Dicer-TRBP complex. **c)** Secondary structures of dicing site mutants with expanded internal loop structures. Mutations are indicated with red lettering. **d)** Pre-miR-31 RNAs with larger internal loops at the Dicer cleavage site have reduced Dicer-TRBP processing efficiencies, relative to WT. For all processing assays, average and standard deviation from  $n=3$  independent assays are presented.

### 3.2.5 Structure at the cleavage site affects Dicer-TRBP processing.

The RNase III and helicase domains of Dicer interact with the upper stem loop region (which includes the apical loop and the dicing site) and studies indicate that the structure in this region may regulate Dicer processing (15, 16, 18, 45, 46). To differentiate the importance of structure at distinct elements within the upper stem loop region, we employed a mutational

approach which reshaped the apical loop and the dicing site, independently.

We generated four different Dicer processing site mutants and examined the impact of structure at this site on Dicer-TRBP processing. Mutations that either minimized ( $\Delta 43$ ) or eliminated ( $\Delta 43/U44A$ ) the internal loop at the pre-miR-31 Dicer processing site (**Fig. 3.3a**) displayed significantly enhanced processing (**Fig. 3.3b, Appendix A Table S4**). Additionally, mutations that destabilized base pairs proximal to the dicing site (**Fig. 3.3c**) show a significant reduction in the rate of Dicer-TRBP processing relative to WT (**Fig. 3.3d, Appendix A Table S4**).

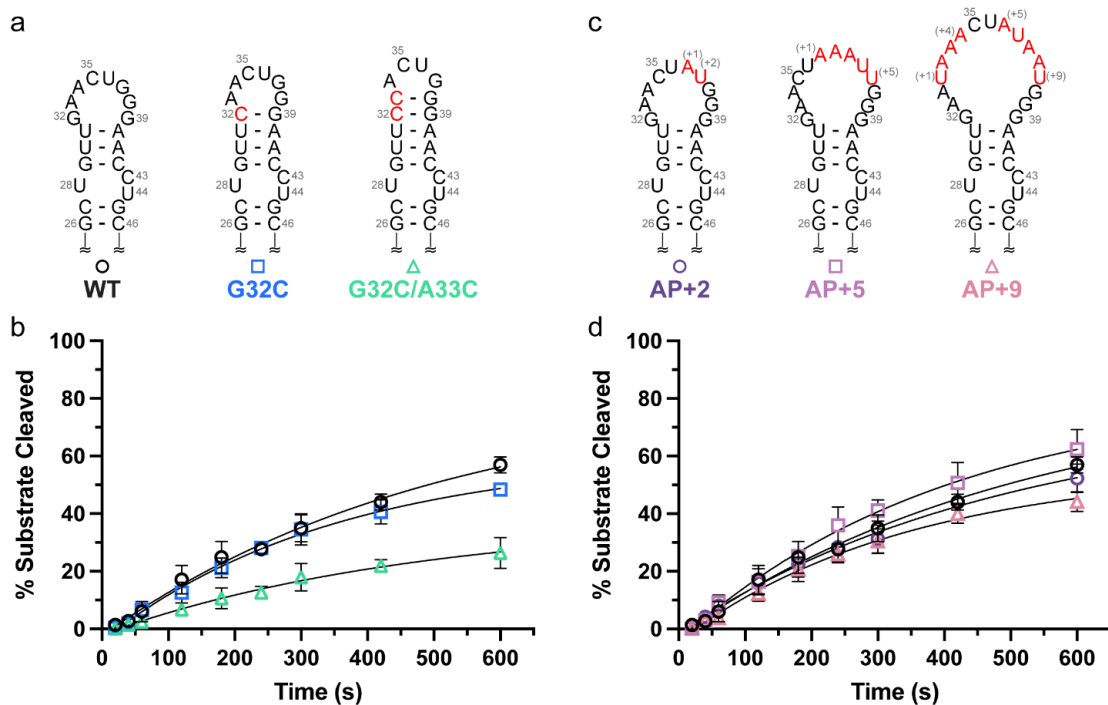
While the addition of TRBP reduces WT-pre-miR-31 processing, the  $\Delta 43C$  and  $\Delta 43C/U44A$  RNAs exhibited similar processing behavior in both Dicer-TRBP (**Fig. 3.3 a,b, Appendix A Table S4**) and Dicer only conditions (**Appendix A Fig. S21 a,b, Appendix A Table S5**). These results indicate that the Dicer-TRBP complex prefers a smaller (or lack of) internal loop structure at the dicing site. Although constructs with a relatively larger internal loop structure at the dicing site are significantly inhibited in a Dicer only assay (**Appendix A Fig. S21 c,d, Appendix A Table S5**), we observe some recovery of activity in the presence of TRBP (**Fig. 3.3 c,d, Appendix A Table S4**).

### **3.2.6 Size and relative position of the apical loop regulates Dicer-TRBP processing**

Apical loop flexibility serves as a control mechanism in many pri/pre-miRNA elements (47, 48) and the apical loop has been identified as a target for regulation by proteins, peptides, and small molecules (13, 49-52). Studies examining the role of apical loop size in Fly Dicer-1 processing revealed that a pre-let-7 RNA with a 4-nt loop were processed less efficiently than a pre-let-7 RNA with a 14-nt loop (16). To build on these findings, we designed two constructs, G32C and G32C/A33C, which minimize the apical loop size by extending the base paired junction

region (**Fig. 3.4a**). We find that the rate of Dicer-TRBP processing of a G32C/A33C mutant, which restricts the pre-miR-31 apical loop to 4-nt, is reduced by  $\sim 2.5$ -fold (**Fig. 3.4b, Appendix A Table S4**). However, the G32C mutant (6-nt apical loop) exhibited WT-like Dicer-TRBP processing rates (**Fig. 3.4b, Appendix A Table S4**).

Pre-miRNAs with small apical loops (3-9 nt long) were identified as poor substrates for human Dicer processing while RNAs with larger apical loops were preferred by Dicer and Drosha (53). We incorporated non-native nucleotides to the apical loop regions of pre-miR-31 to generate AP+2 (10-nt loop), AP+5 (13-nt loop) and AP+9 constructs (17-nt loop) (**Fig. 3.4c**). These larger loop mutants do not influence Dicer-TRBP processing (**Fig. 3.4d, Appendix A Table S4**).



**Figure 3.4. Pre-miR-31 requires a greater than 4-nt apical loop for efficient processing. a)** Secondary structures of pre-miR-31 RNAs engineered to contain smaller apical loops. Sites of mutation are denoted with red lettering. **b)** *In vitro* processing assays with the Dicer-TRBP complex reveal a significant reduction in substrate cleavage for the G32C/A33C RNA. **c)** Secondary structures of mutants designed to extend the pre-miR-31 apical loop size. Non-native nucleotide insertions are indicated with red lettering. **d)** Dicer-TRBP processing of pre-miR-31 RNAs with larger apical loops was largely unchanged relative to WT. For all processing assays, average and standard deviation from  $n=3$  independent assays are presented.

However, Dicer alone is more sensitive to the apical loop size (**Appendix A Fig. S22, Appendix A Table S5**). In the absence of TRBP, G32C has a  $k_{\text{obs}}$  that is 4-fold lower than that of WT (**Appendix A Fig. S22 a,b, Appendix A Table S5**). Additionally, we found that the processing of AP+2, AP+5, and AP+9 RNAs was inhibited in the Dicer only processing assays (**Appendix A Fig. S22 c,d, Appendix A Table S5**).

The reduction in Dicer processing caused by the presence of a larger apical loop can be offset by other factors, including loop position. Previous studies showed that the apical loop or an internal loop 2-nt from cleavage sites could enhance cleavage efficiency of shRNAs (18, 54). Consistent with previous studies, a pre-miR-31 construct containing an 11-nt loop positioned 2-nt from the cleavage site displayed WT-rates of Dicer processing (40UUG, **Appendix A Fig. S23**). Collectively, our findings suggest that TRBP plays an important role in tolerating pre-miRNAs with diverse apical loop sizes and that the presence of TRBP can enlarge the selectivity window for Dicer and promote cleavage for moderately small or large apical loop constructs.

### **3.2.7 Junction residues function as critical control elements for Dicer processing**

Our NMR-derived structure of FL pre-miR-31 revealed the presence of three base pairs in a junction region between the apical loop and the dicer cleavage site (**Fig. 3.1b**). However, *in cell* chemical probing studies with a catalytically inactive Dicer revealed that junction residues were highly reactive, suggesting that these base pairs are absent in the presence of Dicer (27). The high reactivity of these nucleotides *in cell* is consistent with our *in vitro* chemical probing studies (**Fig. 3.1a**) which suggest that pre-miR-31 has an unpaired junction region. To assess the functional importance of these alternative structures, we designed constructs which stabilized or destabilized the junction residues and examined the impact on processing. The junction stability and structure



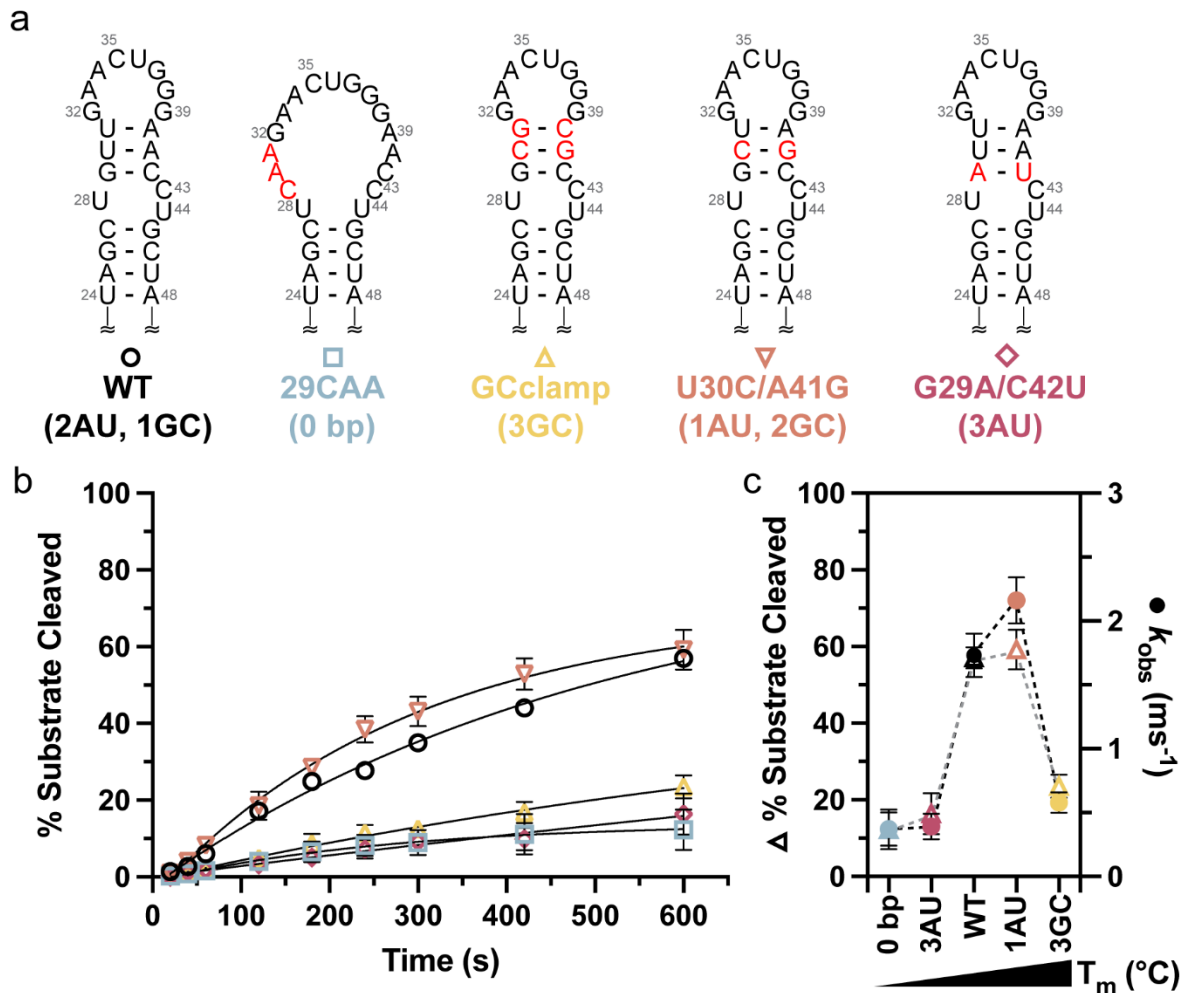
were assessed by thermal denaturation of these constructs (**Appendix A Fig. S24, Appendix A Table S6**).

To mimic the large open loop structure detected by chemical probing, we mutated residues G29, U30, and U31 to prevent base pairing in the junction region (29CAA) (**Fig. 3.5a**). The Dicer-TRBP processing data for 29CAA reveals that it is a poor substrate for Dicer-TRBP processing, with a 5-fold reduction in  $k_{obs}$  for 29CAA relative to WT (**Fig. 3.5b, Appendix A Table S4**). We also designed a construct to stabilize the junction region, where the junction A-U base pairs were replaced with G-C base pairs (GCclamp, **Fig. 3.5a**). Interestingly, the Dicer-TRBP processing of the GCclamp construct was reduced 3-fold, relative to WT (**Fig. 3.5b, Appendix A Table S4**). These data suggest that the stability of the base pairs within the junction region is an important determinant of Dicer-TRBP processing.

To further elucidate how the junction stability of pre-miR-31 regulates Dicer-TRBP processing, we designed two additional junction mutants with different base pairing compositions. The U30C/A41G construct (1AU base pair and 2 GC base pairs, **Fig. 3.5a**) has a  $k_{obs}$  that is 1.25-fold faster than WT (2AU base pairs, 1 GC base pair) (**Fig 3.5b, Appendix A Table S4**). Whereas the  $k_{obs}$  for the G29A/C42U mutant (3 AU base pairs, **Fig. 3.5a**) is reduced 5-fold relative to WT, similar to what we observed for 29CAA (**Fig. 3.5b, Appendix A Table S4**). Similar trends hold for the junction mutants in processing assays without TRBP (**Appendix A Table S5**). These data suggest that the stability of the junction region is finely tuned to maximize Dicer processing.

Collectively, we observe a trend in which pre-miR-31 RNAs with a moderately stable junction (WT, U30C/A41G) are efficiently processed by Dicer-TRBP, suggesting these structures are ideal substrates. However, pre-miR-31 RNAs with either an over-stabilized junction (GCclamp) or a highly destabilized junction (G29A/C42U and 29CAA) are poor substrates,

reflected in their reduced  $k_{obs}$  (Fig. 3.5c). This delicate balance of structural stability within the junction must be optimized to maximize efficient processing.



**Figure 3.5. The junction region is a regulatory element within pre-miR-31. a)** Design of pre-miR-31 RNAs with varying junction stabilities. Mutations are indicated with red lettering. **b)** Time-dependent Dicer-TRBP processing of pre-miR-31 junction RNAs reveals the importance of junction stability. **c)** Correlation between Dicer-TRBP processing (% substrate cleaved and  $k_{obs}$ ) and measured thermal stability (melting temperature,  $T_m$ ) for WT and junction region mutations reveals need for a moderately stable junction for efficient processing. For all thermal denaturation experiments and processing assays, the average and standard deviation from  $n=3$  independent assays are presented.

### 3.3 Discussion

miRNAs play important roles in the post-transcriptional regulation of gene expression in eukaryotes. Due to this important regulatory function, miRNAs are themselves subject to post-transcriptional regulation to ensure appropriate levels of the mature products are produced. Many proteins are known to post-transcriptionally regulate miRNA biogenesis at either the Drosha and/or Dicer processing steps (4, 12, 13). While protein-mediated regulation of miRNA biogenesis can be an important mechanism of control, the intrinsic structural features of pri/pre-miRNAs can also regulate their enzymatic processing (4, 20, 21). Pre-miR-31 is a pre-miRNA with no identified protein binding partners (23), and was therefore an attractive target for uncovering RNA-mediated mechanisms underlying miRNA biogenesis.

To identify the structural basis for Dicer-TRBP processing of pre-miR-31, we solved the high-resolution tertiary structure of the FL pre-miR-31 RNA. Our structural and biochemical studies provide a framework for optimized design of shRNAs and elucidate distinct mechanisms by which RNA structure helps to regulate Dicer-TRBP mediated processing of pre-miR-31 (**Fig. 3.6**).

We found that the presence of mismatches within the pre-miR-31 stem, while a nearly ubiquitous feature of pre-miRNAs, did not significantly influence the processing of pre-miR-31. Furthermore, excluding TRBP did not influence Dicer processing of these mutants. This finding is in contrast to studies with Fly Dicer-1 which demonstrated that mismatches direct the production of different length products in the presence of the Loqs-PB co-protein (44). Previous studies showed the importance of secondary structure at the dicing site for Dicer cleavage of shRNA and some pre-miRNAs (54). Here, we show that RNAs with reduced or eliminated internal loops at the Dicing site are good substrates for Dicer-TRBP processing. However, increasing the internal

loop size negatively impacted Dicer-TRBP processing.

Both apical loop size and position contribute to the regulation of Dicer and Drosha processing (15, 16, 18, 53-55). Our findings re-emphasized the role of apical loop size on Dicer-TRBP processing and provided new insights. Previous studies demonstrate that the presence of a small apical loop inhibits Dicer cleavage (16, 53). We showed not only that a small apical loop inhibits Dicer processing, but also that large apical loops inhibit Dicer processing. Interestingly, the addition of TRBP appears to widen the apical loop size window of RNAs that are efficiently processed. Our study reveals that loop size is one property that should be optimized when designing shRNAs.

Importantly, we found the junction region of pre-miR-31 to be an inherent regulatory site. Our NMR-derived secondary structure stands in contrast to one revealed by *in cell* chemical probing (27). Secondary structures reported based on chemical probing adopt a large apical loop region, where the junction residues are unpaired. We believe that the differences in the NMR and chemical probing derived structures reflect the likely dynamic nature of the base pairs in the junction region, information which can be obstructed in the chemical probing studies. Our structural data are consistent with a model in which some junction base pairs are accessible to solvent and thus more prone to open.

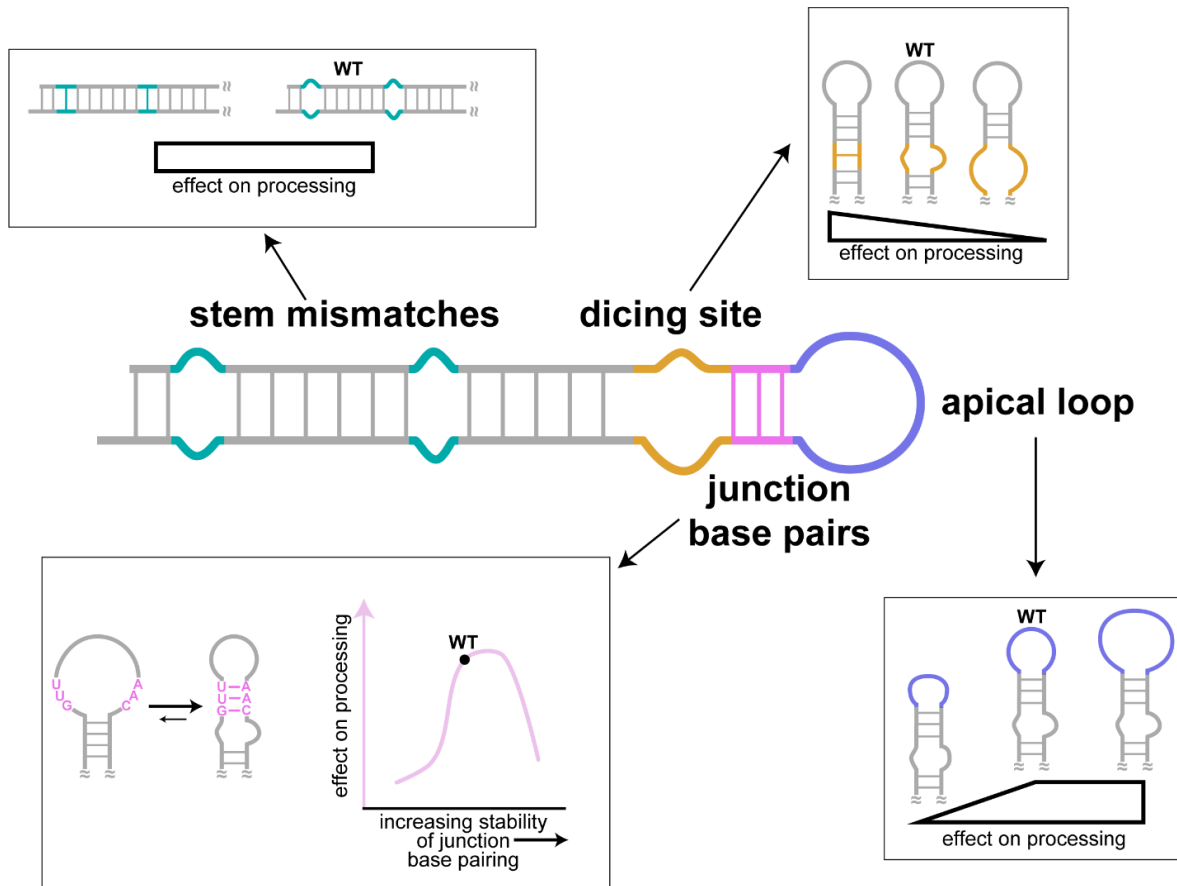
We imagine that the junction region of pre-miR-31 exists in a dynamic equilibrium promoting distinct favorable interactions with the Dicer-TRBP complex. We found that mutations which either stabilized or destabilized the junction region reduced Dicer-TRBP processing. The open apical loop structure sequesters the Dicer cleavage sites in the loop, which may account for the reduced processing levels. Collectively, we found that the stability of the pre-miR-31 junction region is optimized to sample both open and cinched conformations to promote efficient

processing. These findings enrich the understanding of how distinct conformations of pre-miR-31 contribute to Dicer-TRBP processing.

Importantly, we note that our structural and enzymatic studies were conducted *in vitro* under conditions that may not replicate the complex cellular environment. It is indeed possible that regulation in the cell is driven by factors other than pre-miRNA structure and dynamics, for example competition for other pre-miRNAs, varying pre-miRNA expression levels, or crowding factors (56).

Our newly resolved 3D structure of pre-miR-31 in its processing-competent conformation and elucidation of its intrinsic regulatory mechanism informs on the important role that pre-miRNA structural plasticity plays in controlling Dicer processing. Our structural and biochemical studies are consistent with proposed models of pre-miRNA processing based on cryo-EM structures of human Dicer (57) and fly Dicer-1 (55) bound with pre-miRNAs. The pre-let-7 bound human Dicer structure revealed that the pre-let-7 RNA adopts multiple conformations (57). In the “pre-dicing state,” Wang and co-workers posit that the pre-let-7 RNA first binds before the structure is adjusted to form a more stable stem (57). This hypothesis is consistent with our findings that the pre-miR-31 apical loop exists in a conformational equilibrium where a large apical loop structure may be the preferred substrate for Dicer-TRBP binding, but that the structure with a cinched junction region is a “dicing-competent” structure.

The structure of human Dicer bound with let-7a-1<sup>GYM</sup> was recently reported (46). This structure revealed interactions between the Dicer RNase III a domain and pre-miRNA residues above the dicing site, at the junction region identified in pre-miR-31. This structure further revealed that the pre-miRNA structure was significantly distorted from an A-form helical structure near the cleavage site revealing a need for “flexibility” in this region to be reshaped by Dicer.



**Figure 3.6. Secondary structure elements and their contribution to the regulation of pre-miR-31 processing.** The presence or absence of mismatches within the stem of pre-miR-31 had no impact on Dicer-TRBP processing. RNAs with more stabilized Dicing sites were processed more efficiently than the WT sequence, while pre-miRNAs with larger internal loops were poorly processed. Similarly, pre-miRNAs with too small of an apical loop were processed less efficiently than WT pre-miR-31. Interestingly, the WT pre-miR-31 has an inherently encoded structural switch at the junction region. Pre-miR-31 appears to sample both an open loop structure and a closed loop structure. Only RNAs with marginally stable junction regions were maximally processed by Dicer-TRBP. The ability of pre-miR-31 to sample both states promotes processing of pre-miR-31 by the Dicer-TRBP complex.

The recent cryo-EM structures of fly Dicer-1 reveal further details of the Dicer-1-pre-miRNA structure in the “Dicing” state (55). In the “Dicing” state, the structure reveals that the pre-miRNA is highly structured, with the Dicing site sequestered in an A-form helical structure and several base pairs present above the Dicing site. This “Dicing” structure is consistent with our NMR-derived structure, where the stabilization of additional base pairs in the junction promotes

formation of an extended A-helical structure above the dicing site.

Our data suggest that pre-miR-31 is “pre-structured” for Dicer processing. The presence of the junction base pairs in pre-miR-31 supports the formation of an A-form helical structure. However, the stability of these base pairs is critical. The base pairing cannot be too tight, or it may inhibit the restructuring of the RNA for binding. Additionally, the structure cannot be too flexible, as the presence of the A-form geometry is important for Dicer-TRBP processing. Further structural studies are necessary to fully-characterize the structural changes in both the pre-miRNA and Dicer throughout the catalytic cycle.

We also including some information under the Dicer only condition, these data leads to some different conclusions, these data are included in the Appendix B.

### **3.4 Methods**

#### **3.4.1 Preparation of recombinant human Dicer**

Human Dicer protein was purified as previously described (58, 59) with modifications. Sf9 cells with infected His-tagged Dicer baculovirus is purchased from University of Michigan protein core. The cell pellet was lysed in ice-cold lysis buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH = 8.0, 300 mM NaCl, 0.5% Triton X-100, 5% glycerol, 0.5 mM tris(2-carboxyethyl) phosphine (TCEP) and 10 mM imidazole) by sonication. The lysate was pelleted by centrifugation at 30,000 x g for 30 min and the supernatant was mixed with 5 mL pre-equilibrated Ni-NTA resin (Qiagen) in a 50 mL falcon tube. After gently rocking for 1 h at 4 °C, the resin was pelleted by centrifugation at 183 x g for 10 min. The resin was washed with 45 mL wash buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH = 8.0, 300 mM NaCl, 5% glycerol, 0.5 mM TCEP and 20 mM imidazole) 5 times and eluted with elution buffer (50 mM

Na<sub>2</sub>HPO<sub>4</sub> pH = 8.0, 300 mM NaCl, 5% glycerol, 0.5 mM TCEP and 300 mM imidazole). The elutions were dialyzed against dialysis buffer (20 mM Tris pH = 7.5, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 50% glycerol). Purified protein was stored at -80 °C and total protein concentration was determined by Bradford assay (Thermo Fisher Scientific) and the concentration of Dicer was quantified using ImageJ.

### **3.4.2 Expression and purification of TRBP**

The expression and purification of human TRBP was based on previously described procedures (60, 61). The pET28a-TRBP was purchased from Addgene (Addgene plasmid # 50351). The pET28a-TRBP plasmid was transformed into *E. coil* Rosetta (DE3) pLysS (Novagen), and cells were grown in LB media until cell density reached OD<sub>600</sub>=0.6. Protein expression was induced by addition of 0.2 mM IPTG and cells were incubated at 18 °C for 20 h. The cells were harvested by centrifugation, resuspend in buffer A (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM imidazole, and 5 mM β-mercaptoethanol) and lysed by sonication. The lysate was pelleted by centrifugation at 20,000 x g for 30 min and the supernatant was loaded onto a nickel affinity column and gradient eluted in buffer B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole buffer, and 5 mM β-mercaptoethanol). The protein elution was collected and 10% polyethyleneimine (PEI) was added dropwise (1.5% vol/vol) to remove nucleic acid contaminants. The suspension was stirred at 4 °C for 30 min and the supernatant was collected by centrifugation (30 min at 20,000 rpm). Two sequential ammonium sulfate cuts were performed (with centrifugation for 30 min at 20,000 rpm in between) at 20% and 80% saturation. The pellet from the 80% ammonium sulfate cut was resuspended in buffer C (20 mM Tris-HCl pH 8.0, 2 mM DTT) and dialyzed against buffer D (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM DTT) overnight. The overnight dialyzed sample was



concentrated and loaded into HiLoad 16/600 Superdex 200 (Cytiva) equilibrated with buffer D.

### **3.4.3 Dicer-TRBP complex formation**

The purified Dicer and TRBP proteins were mixed at a 1:3 molar ratio and loaded into HiLoad 16/600 Superdex 200 column (Cytiva) equilibrated with buffer D. The fractions enriched with complex were pooled and concentrated to ~534 nM and flash frozen by liquid nitrogen for use in processing assays.

### **3.4.4 Preparation of DNA templates**

DNA templates for oligo RNAs were purchased from Integrated DNA Technologies (**Appendix A Table S7**). The DNA templates for *in vitro* transcription were created by annealing the DNA oligonucleotides with an oligonucleotide corresponding to the T7 promoter sequence (5'-TAATACGACTCACTATA-3'). Templates were prepared by mixing the desired DNA oligonucleotide (40  $\mu$ L, 200  $\mu$ M) with the complementary oligonucleotide to T7 promoter sequence (20  $\mu$ L, 600  $\mu$ M) together, boiling for 3 min, and then slowly cooling to room temperature. The annealed template was diluted with water prior to use to produce the partially double-stranded DNA templates at a final concentration approximately 8  $\mu$ M.

### **3.4.5 Preparation of plasmid templates for *in vitro* transcription**

The templates for preparation of the extended pre-miR-31 for DMS-MaPseq and FL pre-miR-31 for NMR studies were generated by overlap-extension (OE) polymerase chain reaction (PCR) using EconoTaq PLUS 2x Master Mix (Lucigen) with primers listed in **Appendix A Tables S8 and S9**. The OE PCR template was digested with EcoRI and BamHI restriction enzymes and

inserted into the pUC-19 plasmid. DNA templates for use in *in vitro* transcription reactions were amplified with EconoTaq PLUS 2x Master Mix (Lucigen) using primers UNIV-pUC19\_E105 and miR\_tail\_3buffer\_REV (DMS) or miR31\_4R (NMR, **Appendix A Table S10**).

To ensure the native pre-miR-31 used for processing contained homogeneous 5'-AG sequence, of we included a hammerhead (HH) ribozyme 5' of the pre-miR-31 sequence (62). The native pre-miR-31 template, used to make RNA for processing studies, was generated by OE PCR using EconoTaq PLUS 2x Master Mix (Lucigen) with primers listed in **Appendix A Table S11**. The OE PCR template was digested with EcoRI and BamHI restriction enzymes and inserted into pUC-19 plasmid. The HH-pre-miR-31-HDV plasmid, which was designed to ensure a homogeneous 3' end of the transcript, was generated by inserting HDV ribozyme sequence to 3' end of HH-pre-miR-31 plasmid construct using the Q5 site-directed mutagenesis kit (New England Biolabs) with primers HH-miR-31-HDV-mut-F and HH-miR-31-HDV-mut-R (**Appendix A Table S12**). All subsequent mutations, deletions, and/or insertions were achieved via site-directed mutagenesis (New England Biolabs Q5 site-directed mutagenesis kit) of the HH-pre-miR-31-HDV plasmid with primers listed in **Appendix A Table S12**. Templates prepared from plasmids were amplified with EconoTaq PLUS 2x Master Mix (Lucigen) using primers UNIV-pUC19\_E105 and HDV-AMP-R (**Appendix A Table S10**). All primers were purchased from Integrated DNA Technologies. Plasmid identity was verified by Sanger sequencing (Eurofins Genomics) using the universal M13REV sequencing primer.

### 3.4.6 Preparation of RNA

RNAs were prepared by *in vitro* transcription in 1× transcription buffer [40 mM Tris base, 5 mM dithiothreitol (DTT), 1 mM spermidine and 0.01% Triton-X (pH = 8.5)] with addition of 3–6 mM

ribonucleoside triphosphates (NTPs), 10–20 mM magnesium chloride (MgCl<sub>2</sub>), 30–40 ng/μL DNA template, 0.2 unit/mL yeast inorganic pyrophosphatase (New England Biolabs)(63), ~15 μM T7 RNA polymerase and 10–20% (v/v) dimethyl sulfoxide (DMSO). Reaction mixtures were incubated at 37 °C for 3–4 h, with shaking at 70 rpm, and then quenched using a solution of 7 M urea and 500 mM ethylenediaminetetraacetic acid (EDTA), pH = 8.5. Reactions were boiled for 3 min and then snap cooled in ice water for 3 min. The transcription mixture was loaded onto preparative-scale 10% denaturing polyacrylamide gels for purification. Target RNAs were visualized by UV shadowing and gel slices with RNA were excised. Gel slices were placed into an elutrap electroelution device (The Gel Company) in 1X TBE buffer. RNA was eluted from the gel at constant voltage (120 V) for ~24 h. The eluted RNA was spin concentrated, washed with 2 M high-purity sodium chloride, and exchanged into water using Amicon-15 Centrifugal Filter Units (Millipore, Sigma). RNA purity was confirmed on 10% analytical denaturing gels. RNA concentration was quantified via UV-Vis absorbance. Sequences for all RNAs is provided in **Appendix A Table S13**.

#### **3.4.7 Dimethyl sulfate (DMS) modification of pre-miR-31 RNA**

3 μg of pre-miR-31-tail RNA was denatured at 95 °C for 1 min and incubated on ice for another 3 min. Refolding buffer (300 mM sodium cacodylate and 6 mM MgCl<sub>2</sub>) was added to reach total volume of 97.5 uL (for the 0% control), 97.5 μL (for 2.5% modified sample) or 95 μL (for 5% modified sample). The RNA was incubated in refolding buffer at 37 °C for 40 min. The RNA was treated with either 2.5 μL DMSO (0% DMS), 2.5 μL DMS (2.5% DMS) or 5 μL DMS (5% DMS) followed by incubation at 37 °C while shaking at 250 rpm for 10 min. 60 μL β-mercaptoethanol was added to each reaction to neutralize the residual DMS. The modified RNA was purified using

RNA Clean and Concentrator-5 kit (Zymo) according to manufacturer's instructions.

### **3.4.8 RT-PCR with DMS-modified RNA**

The methylated RNA was reverse transcribed as follows. 0.2  $\mu$ M DMS-modified RNA, 2  $\mu$ l 5 $\times$  first strand buffer (ThermoFisher Scientific), 1  $\mu$ l 10  $\mu$ M reverse primer (miR\_tail\_RT, **Appendix A Table S8**), 1  $\mu$ l dNTP, 0.5  $\mu$ l 0.1 M DTT, 0.5  $\mu$ l RNaseOUT and 0.5  $\mu$ l thermostable group II intron reverse transcriptase, 3<sup>rd</sup> generation (TGIRT-III, Ingex) were mixed. The mixture was incubated at 57  $^{\circ}$ C for 30 min. After the 30 min incubation, the temperature was increased to 85  $^{\circ}$ C for 5 min. 1  $\mu$ L RNase H (New England Biolabs) was added to the mixture to digest the RNA. The reverse-transcribed DNA was PCR amplified using Phusion (NEB) for 27 cycles according to the manufacturer's instruction using primers miR31\_buffer\_F and miR\_tail\_RT (**Appendix A Table S8**). The PCR product was purified by GeneJET PCR purification kit (ThermoFisher Scientific).

### **3.4.9 DMS-MaPseq of pre-miR-31 RNA**

Illumina sequencing adapters were added by ligation mediated PCR using the NEBNext UltraII DNA Library Prep Kit (New England BioLabs). The libraries were Bioanalyzed on a high sensitivity DNA chip, size selected and sequenced on Illumina Miseq 600 cycles (300x300 paired end). The resulting sequencing reads were adapter trimmed using Trim Galore and aligned using bowtie2 ("bowtie2 --local --no-unal --no-discordant --no-mixed --phred33 40 -L 12"). Each read was compared to its reference sequence to count how many mutations occurred at each nucleotide. All sequencing reads were combined together to calculate the average mutations per base and create a mutational profile.

### 3.4.10 Isotopic labeling of RNAs for NMR.

Isotopically-labeled RNAs were produced as described above by replacing the rNTP mixture with rNTPs of appropriate isotope labeling.  $^{15}\text{N}/^{13}\text{C}$  rNTPs were obtained from Cambridge Isotope Laboratories (CIL, Andover, MA). The partially- and per-deuterated rNTPs used for *in vitro* transcription were obtained from Cambridge Isotope Laboratories (CIL, Andover, MA) or generated in house, as described below. Protiation at the C8 position of perdeuterated rGTP and rATP was achieved by incubation with triethylamine (TEA, 5 equiv) in  $\text{H}_2\text{O}$  (60 °C for 24 h and for 5 days, respectively). Deuteration of the C8 position of fully protiated GTP and ATP was achieved by analogous treatment with  $\text{D}_2\text{O}$  (99.8% deuteration; CIL). TEA was subsequently removed by lyophilization.

### 3.4.11 NMR experiments.

Samples for NMR experiments of Top, TopA, pre-miR-31 and FL pre-miR-31 were prepared in 300-350  $\mu\text{L}$  100%  $\text{D}_2\text{O}$  (99.8% deuteration; CIL) or 10%  $\text{D}_2\text{O}/90\%$   $\text{H}_2\text{O}$ , 50 mM K-phosphate buffer (pH 7.5), 1 mM  $\text{MgCl}_2$  of 300-600  $\mu\text{M}$  RNA in Shigemi NMR sample tubes. NMR spectra were collected on 600 and 800 MHz Bruker AVANCE NEO spectrometers equipped with a 5 mm three channel inverse (TCI) cryogenic probe (University of Michigan BioNMR Core). NMR spectra of Top and TopA were recorded at 30°C and of pre-miR-31 and FL pre-miR-31 at 37 °C. The isotopic labeling scheme of FL pre-miR-31 used in specific NMR experiment is indicated in the figure legends. NMR data were processed with NMRFX (64) and analyzed with NMRViewJ (65).  $^1\text{H}$  chemical shifts were referenced to water and  $^{13}\text{C}$  chemical shifts were indirectly referenced from the  $^1\text{H}$  chemical shift (66).

The signals of nonexchangeable protons of Top and TopA were assigned based on analysis of 2D  $^1\text{H}$ - $^1\text{H}$  NOESY ( $\tau_m = 400$  ms), 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY ( $\tau_m = 80$  ms), and  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectra. Additionally, the 2D NOESY spectrum ( $\tau_m = 400$  ms) was recorded for  $\text{A}^{\text{HCH}}$ -labeled Top RNA (A and C fully protiated, G and U perdeuterated). Non-exchangeable  $^1\text{H}$  assignments of FL pre-miR-31 were obtained from 2D NOESY data ( $\tau_m = 400$  ms) recorded on fully protiated FL pre-miR-31 and  $\text{A}^{2\text{rG}^{\text{r}}}$ -,  $\text{A}^{2\text{rG}^{\text{r}}\text{U}^{\text{r}}}$ -,  $\text{A}^{\text{HCH}}$ - and  $\text{G}^{\text{HU}^{\text{6r}}}$ -labeled FL pre-miR-31 (superscripts denote sites of protiation on a given nucleoside, all other sites deuterated).  $^1\text{H}$ - $^1\text{H}$  TOCSY and  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of  $^{15}\text{N}/^{13}\text{C}$  AG-labeled FL pre-miR-31 were analyzed to facilitate the assignment. The NMR samples for pH titration were prepared with 300  $\mu\text{M}$   $^{15}\text{N}$  AU-labeled FL pre-miR-31 in 10%  $\text{D}_2\text{O}/90\%$   $\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$  and 10 mM K-phosphate buffer with pH values 5.8, 6.2, 6.5, 7.0, 7.5 and 8.0.

A best-selective long-range HNN-COSY (40) was recorded to identify AU base pairing in FL pre-miR-31. The spectrum was recorded on 560  $\mu\text{M}$   $^{15}\text{N}$  AU-labeled FL pre-miR-31 in 10%  $\text{D}_2\text{O}/90\%$   $\text{H}_2\text{O}$ , 50 mM K-phosphate buffer (pH=7.5) and 1 mM  $\text{MgCl}_2$ . 64 complex points were recorded with a sweep width of 7.4 kHz for  $^{15}\text{N}$ , and 2048 complex points with a sweep width of 16.6 kHz for  $^1\text{H}$ , 1368 scans per complex increment at 37 °C and 800 MHz.

NMR sPRE (67) data of FL pre-miR-31 were obtained by measuring R1 relaxation rates (68) as a function of the concentration of paramagnetic compound Gd(DTPA-BMA) (69). We acquired  $^1\text{H}$ - $^{13}\text{C}$  HSQC-based pseudo-3D experiments at 0.0, 0.8, 1.6, 2.4, 3.2 and 4.8 mM concentration of the paramagnetic compound. Data were acquired on sample containing 480  $\mu\text{M}$   $^{15}\text{N}/^{13}\text{C}$  A,G-labeled FL pre-miR-31 in 100%  $\text{D}_2\text{O}$  (99.8% deuteration; CIL), 50 mM K-phosphate buffer (pD=7.5) and 1 mM  $\text{MgCl}_2$  at 800 MHz using nine delays (0.02-2s) with two repetitions at every titration point. The data were processed and analyzed using NMRFX (64). The sPRE values

were obtained from the peak intensities of well-resolved peaks in the  $^1\text{H}$ - $^{13}\text{C}$  HSQC-based pseudo-3D experiments. These intensities were fitted to an exponential function (equation 1) (68)

$$I = Ae^{-xR_1} \quad (1)$$

where  $I$  is the intensity of the peak,  $A$  is the amplitude of the relaxation and  $R_1$  is the longitudinal proton relaxation rate. The sPRE values were obtained from the  $R_1$  rates determined in the presence of different concentrations of paramagnetic compound Gd(DTPA-BMA) (equation 2) (67)

$$R_1(c_{Gd}) = m_{sPRE} + R_1^0 \quad (2)$$

where  $R_1(c_{Gd})$  is the  $R_1$  measured at the concentration of the paramagnetic compound ( $c_{Gd}$ ), the slope  $m_{sPRE}$  corresponds to the sPRE and  $R_1^0$  is the fitted  $R_1$  in the absence of the paramagnetic compound. The error of the sPRE value  $\Delta m_{sPRE}$  were obtained from the linear regression as described previously (67).

$^1\text{H}$ - $^{13}\text{C}$  RDCs were recorded using IPAP-HSQC experiments (70) for  $^{15}\text{N}/^{13}\text{C}$  AG-labeled FL pre-miR-31. Two samples were prepared, an isotropic sample containing 400  $\mu\text{M}$  RNA in 90%  $\text{H}_2\text{O}/10\% \text{D}_2\text{O}$ , 50 mM K-phosphate buffer (pH=7.5) and 1 mM  $\text{MgCl}_2$ , and an anisotropic sample containing 600  $\mu\text{M}$  FL pre-miR-31 in the same solvent but also including 10 mg/mL Pfl phage, yielding a solvent  $^2\text{H}$  quadrupole splitting of 11 Hz. 110 complex points were recorded with a sweep width of 8 kHz for  $^{13}\text{C}$ , and 32768 complex points with a sweep width of 14.7 kHz for  $^1\text{H}$ , 200 scans per complex increment at 800 MHz. Spectra were processed and analyzed with Bruker Topspin.

### 3.4.12 Structure calculations.

CYANA was used to generate 640 initial structures via simulated annealing molecular dynamics calculations over 128,000 steps. Upper limits for the NOE distance restraints generally

set at 5.0 Å for weak, 3.3 Å for medium, and 2.7 Å for strong signals, based on peak intensity. Notable exceptions included intraresidue NOEs between H6/H8 and H2' (4.0 Å) and H3' (3.0 Å). For very weak signals, 6.0 Å upper limit restraints were used, including for sequential H1'-H1' NOEs and intraresidue H5-H1' NOEs. Standard torsion angle restraints were included for regions with A-helical geometry, allowing for  $\pm 25^\circ$  deviations from ideality ( $\zeta=-73^\circ$ ,  $\alpha=-62^\circ$ ,  $\beta=180^\circ$ ,  $\gamma=48^\circ$ ,  $\delta=83^\circ$ ,  $\epsilon=-152^\circ$ ). Torsion angles for mismatches were further relaxed to allow for  $\pm 75^\circ$  deviation from ideality. Hydrogen bonding restraints were included for experimentally validated base pairs as were standard planarity restraints. Cross-strand P-P distance restraints were employed for A-form helical regions to prevent the generation of structures with collapsed major grooves (71). A grid search was performed over a broad range of tensor magnitude and rhombicity with weighting of the experimentally determined  $^1\text{H}$ - $^{13}\text{C}$  residual dipolar couplings (RDCs) constraints. 40 input structures were further minimized after singular value decomposition fits of the RDC weights.

The top 20 CYANA-derived structures were then subjected to molecular dynamics simulations and energy minimization with AMBER (72). Only upper limit NOE, hydrogen bond, and dipolar coupling restraints were used, along with restraints to enforce planarity of aromatic residues and standard atomic covalent geometries and chiralities (71, 73). Backbone torsion angle and phosphate-phosphate restraints were excluded during AMBER refinement. Calculations were performed using the RNA.OL3 (74) and generalized Born (75) force fields. NMR restraints and structure statistics are presented in **Appendix A Table S2**.

### **3.4.13 Small angle X-ray scattering (SAXS)**

SAXS with in-line size exclusion chromatography (SEC) and multiangle light scattering (MALS)



was performed at BioCAT (beamline 18ID at the Advanced Photon Source, Chicago). SAXS buffer contained 50 mM potassium phosphate buffer, pH 7.5, 50 mM NaCl, 1 mM MgCl<sub>2</sub> and all data was collected at 20 °C. Full details of SAXS data collection and analysis are presented in the Supporting Information and in **Appendix A Table S3**.

#### **3.4.14 <sup>32</sup>P labeling of RNA**

The 5'-end labeling of RNA was performed using 5 pmol of RNA, 1 μL γ-<sup>32</sup>P-ATP (PerkinElmer) and 10 U T4 polynucleotide kinase (New England Biolabs) in a final volume of 10 μL. Before labeling, RNA was boiled for 3 minutes, and snap cooled by placing on ice for another 3 minutes. The radiolabeled RNA was purified on a G-25 column (Cytiva) according to the manufacturer's instructions. The radiolabeled RNA concentration was determined based on a standard curve which was obtained from the counts per minute of the γ-<sup>32</sup>P-ATP source.

#### **3.4.15 Dicer and Dicer-TRBP processing assay**

Human Dicer protein processing assay was performed as previously described with minimal modifications (20). Concentrated recombinant human Dicer protein was diluted in 1X Dicing buffer (24 mM HEPES or 24 mM Bis-Tris, pH 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 4 μM EDTA). Dicer enzyme was pre-mixed with 80 U RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific) and 5X dicing buffer (120 mM HEPES or 120 mM Bis-Tris, pH 7.5, 0.5 M NaCl, 25 mM MgCl<sub>2</sub>, 0.02 mM EDTA). The <sup>32</sup>P-labeled RNA was heated to 95°C for 3 min and then placed on ice for another 3 min. The RNA (1 μL) was added to pre-mixed solution (9 μL) and incubated at 37°C. The final RNA and enzyme concentration are 2 nM and 20 nM, respectively. The reaction is quenched by adding 10 μL quench buffer (98% Formamide, 20 mM EDTA, trace

bromophenol blue and xylene cyanol) at 20, 40, 60, 120, 180, 240, 300, 420 and 600 sec respectively. After sample was run on a 12% denaturing polyacrylamide gel, the gel was exposed to a phosphor screen, which was scanned by a Typhoon Phosphor Imager (GE Healthcare). The gel image was quantified analyzed by ImageJ. The Dicer cleavage ratio was calculated as the sum of the intensity of products and partially digested products divided by the sum of the intensity of the products, partially digested products, and remaining substrate. Experiments were performed in triplicate. The average, and standard deviation of the measurements are reported. Processing assays with the reconstituted Dicer-TRBP complex were conducted as described for Dicer alone.

The pre-miRNA cleavage ratio was fitted as function of time to get apparent rate constant ( $k_{obs}$ ) using Graphpad Prism Software. Data are fitted using equation 3:

$$Cleavage\ ratio = Y_0 + A \times (1 - e^{-k_{obs}t}) \quad (3)$$

In the fitting parameter,  $Y_0$  and  $A$  are global non-regressive fitting and shared among data set,  $Y_0$  is set to be less than 5 and  $A$  is set to be less than 100. A Student's t-test was used to compare the means of the calculated  $k_{obs}$  between mutant and WT RNAs.

#### CD-thermal denaturation of RNA and data analysis

CD-thermal denaturing of RNAs were performed on JASCO J1500CD spectrometer with a heating rate of 1 °C per min from the 5 °C to 95 °C. Data points were collected every 0.5 °C with absorbance detection at 260 nm. 20 μM RNA samples were premixed in potassium phosphate buffer (pH=7.5) with 1 mM MgCl<sub>2</sub>. The single transition unfolding melting profiles were analyzed using a two-state model using sloping baselines (equation 4) (76).

$$f(T) = \frac{(m_u T + b_u) + (m_f T + b_f) e^{\left[\frac{\Delta H}{R}\right]\left[\frac{1}{(T_m + 273.15)} - \frac{1}{(T + 273.15)}\right]}}{1 + e^{\left[\frac{\Delta H}{R}\right]\left[\frac{1}{(T_m + 273.15)} - \frac{1}{(T + 273.15)}\right]}} \quad (4)$$

where  $m_u$  and  $m_f$  are the slopes of the lower (unfolded) and upper (folded) baselines,  $b_u$  and  $b_f$  are

the y-intercepts of the lower and upper baselines, respectively.  $\Delta H$  (in kcal/mol) is the enthalpy of folding and  $T_m$  (in °C) is the melting temperature,  $R$  is the gas constant (0.001987 kcal/(Kmol)). Experiments were performed in triplicate. The average, and standard deviation of the measurements are reported.

### 3.5 Reference

1. R. C. Friedman, K. K.-H. Farh, C. B. Burge, D. P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs. *Genome research* **19**, 92-105 (2009).
2. L. F. Gebert, I. J. MacRae, Regulation of microRNA function in animals. *Nature reviews Molecular cell biology* **20**, 21-37 (2019).
3. H. Siomi, M. C. Siomi, Posttranscriptional regulation of microRNA biogenesis in animals. *Molecular cell* **38**, 323-332 (2010).
4. M. Ha, V. N. Kim, Regulation of microRNA biogenesis. *Nature reviews Molecular cell biology* **15**, 509-524 (2014).
5. R. C. Wilson *et al.*, Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Molecular cell* **57**, 397-407 (2015).
6. S. Chakravarthy, S. H. Sternberg, C. A. Kellenberger, J. A. Doudna, Substrate-specific kinetics of Dicer-catalyzed RNA processing. *Journal of molecular biology* **404**, 392-402 (2010).
7. M. Fareh *et al.*, TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments. *Nature communications* **7**, 13694 (2016).
8. D. P. Bartel, Metazoan micornas. *Cell* **173**, 20-51 (2018).
9. M. Van Kouwenhove, M. Kedde, R. Agami, MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Reviews Cancer* **11**, 644-656 (2011).
10. L. Pandolfini *et al.*, METTL1 promotes let-7 MicroRNA processing via m7G methylation. *Molecular cell* **74**, 1278-1290. e1279 (2019).
11. I. Heo *et al.*, TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696-708 (2009).
12. J. Balzeau, M. R. Menezes, S. Cao, J. P. Hagan, The LIN28/let-7 pathway in cancer.

*Frontiers in genetics* **8**, 31 (2017).

13. R. Shang *et al.*, Regulated dicing of pre-mir-144 via reshaping of its terminal loop. *Nucleic acids research* **50**, 7637-7654 (2022).
14. N. J. Lehrbach *et al.*, LIN-28 and the poly (U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nature structural & molecular biology* **16**, 1016-1020 (2009).
15. X. Zhang, Y. Zeng, The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic acids research* **38**, 7689-7697 (2010).
16. A. Tsutsumi, T. Kawamata, N. Izumi, H. Seitz, Y. Tomari, Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nature structural & molecular biology* **18**, 1153-1158 (2011).
17. V. C. Auyeung, I. Ulitsky, S. E. McGeary, D. P. Bartel, Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**, 844-858 (2013).
18. S. Gu *et al.*, The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell* **151**, 900-911 (2012).
19. P. Dallaire *et al.*, Structural dynamics control the MicroRNA maturation pathway. *Nucleic acids research* **44**, 9956-9964 (2016).
20. J. T. Baisden, J. A. Boyer, B. Zhao, S. M. Hammond, Q. Zhang, Visualizing a protonated RNA state that modulates microRNA-21 maturation. *Nature chemical biology* **17**, 80-88 (2021).
21. M. D. Shortridge, G. L. Olsen, W. Yang, M. J. Walker, G. Varani, A slow dynamic RNA switch regulates processing of microRNA-21. *Journal of Molecular Biology* **434**, 167694 (2022).
22. T. Yu, P. Ma, D. Wu, Y. Shu, W. Gao, Functions and mechanisms of microRNA-31 in human cancers. *Biomedicine & Pharmacotherapy* **108**, 1162-1169 (2018).

23. T. Treiber *et al.*, A compendium of RNA-binding proteins that regulate microRNA biogenesis. *Molecular cell* **66**, 270-284. e213 (2017).
24. Y. Chen *et al.*, Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic acids research* **44**, 4381-4395 (2016).
25. M. D. Shortridge *et al.*, A macrocyclic peptide ligand binds the oncogenic microRNA-21 precursor and suppresses dicer processing. *ACS chemical biology* **12**, 1611-1620 (2017).
26. S. Bellaousov, J. S. Reuter, M. G. Seetin, D. H. Mathews, RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic acids research* **41**, W471-W474 (2013).
27. Q.-J. Luo *et al.*, RNA structure probing reveals the structural basis of Dicer binding and cleavage. *Nature communications* **12**, 1-12 (2021).
28. A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function. *Nucleic acids research* **47**, D155-D162 (2019).
29. A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* **42**, D68-D73 (2014).
30. A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research* **39**, D152-D157 (2010).
31. S. Griffiths-Jones, H. K. Saini, S. Van Dongen, A. J. Enright, miRBase: tools for microRNA genomics. *Nucleic acids research* **36**, D154-D158 (2007).
32. S. Griffiths-Jones, R. J. Grocock, S. Van Dongen, A. Bateman, A. J. Enright, miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research* **34**, D140-D144 (2006).
33. S. Griffiths-Jones, The microRNA registry. *Nucleic acids research* **32**, D109-D111 (2004).
34. M. Zubradt *et al.*, DMS-MaPseq for genome-wide or targeted RNA structure probing in

vivo. *Nature methods* **14**, 75-82 (2017).

35. P. Tomezsko, H. Swaminathan, S. Rouskin, Viral RNA structure analysis using DMS-MaPseq. *Methods* **183**, 68-75 (2020).
36. A. Kotar, S. Ma, S. C. Keane, pH dependence of C•A, G•A and A•A mismatches in the stem of precursor microRNA-31. *Biophysical Chemistry*, 106763 (2022).
37. K. Lu *et al.*, NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **334**, 242-245 (2011).
38. S. C. Keane *et al.*, Structure of the HIV-1 RNA packaging signal. *Science* **348**, 917-921 (2015).
39. A. Kotar, H. N. Foley, K. M. Baughman, S. C. Keane, Advanced approaches for elucidating structures of large RNAs using NMR spectroscopy and complementary methods. *Methods* **183**, 93-107 (2020).
40. A. Dallmann *et al.*, Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences. *Angewandte Chemie International Edition* **52**, 10487-10490 (2013).
41. H. Y. Lee, K. Zhou, A. M. Smith, C. L. Noland, J. A. Doudna, Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic acids research* **41**, 6568-6576 (2013).
42. H. Y. Lee, J. A. Doudna, TRBP alters human precursor microRNA processing in vitro. *Rna* **18** (2012).
43. R. Fukunaga *et al.*, Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell* **151**, 533-546 (2012).
44. L. Zhu, S. K. Kandasamy, R. Fukunaga, Dicer partner protein tunes the length of miRNAs

- using base-mismatch in the pre-miRNA stem. *Nucleic acids research* **46**, 3726-3741 (2018).
45. Z. Liu, J. Wang, G. Li, H.-W. Wang, Structure of precursor microRNA's terminal loop regulates human Dicer's dicing activity by switching DExH/D domain. *Protein & Cell* **6**, 185-193 (2015).
  46. Y.-Y. Lee, H. Lee, H. Kim, V. N. Kim, S.-H. Roh, Structure of the human DICER-pre-miRNA complex in a dicing state. *Nature*, 1-8 (2023).
  47. Y. Zeng, R. Yi, B. R. Cullen, Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *The EMBO journal* **24**, 138-148 (2005).
  48. M. Trabucchi *et al.*, The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010-1014 (2009).
  49. M. G. Costales *et al.*, Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit. *Journal of the American Chemical Society* **139**, 3446-3455 (2017).
  50. D. Bose *et al.*, Selective inhibition of miR-21 by phage display screened peptide. *Nucleic acids research* **43**, 4342-4352 (2015).
  51. A. Murata, T. Otabe, J. Zhang, K. Nakatani, BzDANP, a small-molecule modulator of pre-miR-29a maturation by Dicer. *ACS Chemical Biology* **11**, 2790-2796 (2016).
  52. D. Ustianenko *et al.*, LIN28 selectively modulates a subclass of let-7 microRNAs. *Molecular cell* **71**, 271-283. e275 (2018).
  53. Y. Feng, X. Zhang, P. Graves, Y. Zeng, A comprehensive analysis of precursor microRNA cleavage by human Dicer. *Rna* **18**, 2083-2092 (2012).
  54. T. D. Nguyen, T. A. Trinh, S. Bao, T. A. Nguyen, Secondary structure RNA elements control the cleavage activity of DICER. *Nature communications* **13**, 1-16 (2022).
  55. K. Jouravleva *et al.*, Structural basis of microRNA biogenesis by Dicer-1 and its partner



protein Loqs-PB. *Mol Cell* **82**, 4049-4063 e4046 (2022).

56. J. P. Sieg, L. N. McKinley, M. J. Huot, N. H. Yennawar, P. C. Bevilacqua, The metabolome weakens RNA thermodynamic stability and strengthens RNA chemical stability. *Biochemistry* **61**, 2579-2591 (2022).

57. Z. Liu *et al.*, Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate. *Cell* **173**, 1191-1203. e1112 (2018).

58. I. J. MacRae, E. Ma, M. Zhou, C. V. Robinson, J. A. Doudna, In vitro reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences* **105**, 512-517 (2008).

59. D. A. Lorenz, A. L. Garner, A click chemistry-based microRNA maturation assay optimized for high-throughput screening. *Chemical communications* **52**, 8267-8270 (2016).

60. T. Takahashi *et al.*, Distinguishable in vitro binding mode of monomeric TRBP and dimeric PACT with siRNA. *PloS one* **8**, e63434 (2013).

61. M. D. Shortridge *et al.*, Drug-like small molecules that inhibit expression of the oncogenic microRNA-21. *ACS Chemical Biology* **18**, 237-250 (2023).

62. E. Ma, I. J. MacRae, J. F. Kirsch, J. A. Doudna, Autoinhibition of human dicer by its internal helicase domain. *Journal of molecular biology* **380**, 237-243 (2008).

63. P. R. Cunningham, J. Ofengand, Use of inorganic pyrophosphatase to improve the yield of in vitro transcription reactions catalyzed by T7 RNA polymerase. *Biotechniques* **9**, 713-714 (1990).

64. M. Norris, B. Fetler, J. Marchant, B. A. Johnson, NMRFX Processor: a cross-platform NMR data processing program. *Journal of biomolecular NMR* **65**, 205-216 (2016).

65. B. A. Johnson, R. A. Blevins, NMR View: A computer program for the visualization and

analysis of NMR data. *Journal of biomolecular NMR* **4**, 603-614 (1994).

66. D. S. Wishart *et al.*,  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift referencing in biomolecular NMR. *Journal of biomolecular NMR* **6**, 135-140 (1995).

67. C. Hartlmueller *et al.*, RNA structure refinement using NMR solvent accessibility data. *Scientific reports* **7**, 1-10 (2017).

68. A. L. Hansen, H. M. Al-Hashimi, Dynamics of large elongated RNA by NMR carbon relaxation. *Journal of the American Chemical Society* **129**, 16072-16082 (2007).

69. P. Caravan, J. J. Ellison, T. J. McMurry, R. B. Lauffer, Gadolinium (III) chelates as MRI contrast agents: structure, dynamics, and applications. *Chemical reviews* **99**, 2293-2352 (1999).

70. A. Enthart, J. C. Freudenberger, J. Furrer, H. Kessler, B. Luy, The CLIP/CLAP-HSQC: pure absorptive spectra for the measurement of one-bond couplings. *Journal of Magnetic Resonance* **192**, 314-322 (2008).

71. B. S. Tolbert *et al.*, Major groove width variations in RNA structures determined by NMR and impact of  $^{13}\text{C}$  residual chemical shift anisotropy and  $^1\text{H}$ - $^{13}\text{C}$  residual dipolar coupling on refinement. *Journal of biomolecular NMR* **47**, 205-219 (2010).

72. D. A. Case *et al.*, The Amber biomolecular simulation programs. *Journal of computational chemistry* **26**, 1668-1688 (2005).

73. I. Yildirim, H. A. Stern, J. D. Tubbs, S. D. Kennedy, D. H. Turner, Benchmarking AMBER force fields for RNA: Comparisons to NMR spectra for single-stranded r (GACC) are improved by revised  $\chi$  torsions. *The journal of physical chemistry B* **115**, 9261-9270 (2011).

74. M. Zgarbová *et al.*, Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of chemical theory and computation* **7**, 2886-2902 (2011).

75. J. Mongan, C. Simmerling, J. A. McCammon, D. A. Case, A. Onufriev, Generalized Born model with a simple, robust molecular volume correction. *Journal of chemical theory and computation* **3**, 156-169 (2007).
76. K. A. Leamy, N. H. Yennawar, P. C. Bevilacqua, Cooperative RNA folding under cellular conditions arises from both tertiary structure stabilization and secondary structure destabilization. *Biochemistry* **56**, 3422-3433 (2017).

## CHAPTER IV Use of Steric Blocking Antisense Oligonucleotides for the Targeted

### Inhibition of Junction Containing Precursor MicroRNAs

Sicong Ma<sup>1</sup>, Samantha Howden<sup>3</sup>, and Sarah C. Keane<sup>1,2\*</sup>

<sup>1</sup>Program in Biophysics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>2</sup>Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

<sup>3</sup>College of LSA, University of Michigan, Ann Arbor, MI 48109, USA.

\*Corresponding author: sckeane@umich.edu

**Author contributions:** Samantha Howden carried out initial cell methods development. All other work was done by Sicong Ma.

#### 4.1 Introduction

Antisense oligonucleotides (ASOs) are a type of chemically modified, short, single-stranded nucleic acid that bind to target RNAs via Watson-Crick base pairing. ASOs are typically 12-30 nucleotides (nts) long and increasing the ASO length results in increased specificity for the target RNA. The mechanisms through which ASOs regulate RNA function can be broadly categorized as either promoting RNA turnover or sterically blocking the targeted RNA which can affect a multitude of processes [1-3].

One common use of ASOs is to promote target RNA cleavage by recruitment of either RNase H1 or Argonaute 2 (Ago2) proteins [4]. RNase H1 is an endogenous endonuclease that is present in the nucleus, cytoplasm, and mitochondria of mammalian cells [5]. RNase H1 plays an important role in mitochondrial genome replication and also functions in genome maintenance by removing R-loop [5, 6]. RNase H1 recognizes RNA-DNA heteroduplexes and catalyzes a phosphoryl transfer reaction that results in a break in the phosphodiester backbone of the RNA [7].

Exogenous small interfering RNAs (siRNAs) can also be used for nucleic acid therapy [4]. Similar to a mature microRNA duplex, siRNA duplexes, which may contain a variety of chemical modifications [8], can be loaded into Ago2 to form the RNA induced silencing complex (RISC). The functional RISC binds to the 3'-untranslated region (UTR) of a specific messenger RNA (mRNA) which induces cleavage of that mRNA. [9, 10].

Steric blocking is another approach for ASO-targeted inhibition of RNA function. Mammalian mRNAs undergo a number of post-transcriptional modifications, including the splicing of one or more intron sequences out of a precursor mRNA [11]. Splicing is a fundamental process in RNA biology and the dysregulation of splicing is associated with numerous diseases and pathologies [12]. ASOs can be used to target the splicing process, broadly regulating gene expression. Duchenne muscular dystrophy (DMD) is a severe neuromuscular disease that is caused by loss-of-function mutations in the dystrophin gene [13]. Eteplirsen, a morpholino ASO, is designed to skip exon 51, which leads to the production of a partially functional dystrophin protein [14]. Conversely, Nusinersen, an ASO containing modifications at both the 2'-hydroxyl and the phosphate backbone, functions to reduce exon skipping and has been shown to be an effective treatment of spinal muscular atrophy, a rare neuromuscular disorder [15]. Nusinersen increases levels of survival of motor neuron 2 (SMN2) mRNAs that include exon 7 which can be translated into a functional SMN protein [15, 16]. Steric blocking ASOs that target RNA elements in the 5'-UTR of mRNAs can function to up- or down-regulate protein translation. For example, ASOs that overlap with or are close to the mRNA start codon can prevent ribosome binding or assembly on the mRNA, reducing translation [4, 17]. Alternatively, ASOs that target repressive elements, like an upstream open reading frame (uORF) or stem-loop structure within the 5'-UTR, can enhance the translation of the downstream gene [18, 19].

Finally, ASOs can be used to sequester mature microRNAs (miRNAs) to enhance protein expression [1, 4, 20]. In the nucleus, RNA polymerase II transcribes primary microRNAs (pri-miRNAs) which are enzymatically cleaved by the Microprocessor complex, which is composed of Drosha and DiGeorge syndrome critical region 8 (DGCR8) proteins. Processing of the pri-miRNA generates a precursor microRNA (pre-miRNA) that is exported from the nucleus to the cytoplasm. In the cytoplasm, the pre-miRNA is further processed by Dicer, which functions in complex with transactivation response element RNA-binding protein (TRBP) [21-23]. Processing of the pre-miRNA by Dicer/TRBP results in the production of a 21-22 nt mature miRNA duplex [24, 25]. Mature miRNAs function in concert with Argonaute (Ago) proteins to form the RNA-induced silencing complex (RISC) which induces mRNA degradation or translational repression [26, 27].

Mature miRNAs are a common ASO target. ASOs that are complementary to the mature miRNA effectively sequester that miRNA within the RISC into a duplex. This sequestration prevents RISC binding to the targeted mRNA, which increases the mRNA lifetime and protein expression. An alternative approach for ASO-based therapeutics involves targeting pre-miRNAs rather than mature miRNAs. ASOs that bind to the apical loop region and block the dicing sites of pre-miR-16, pre-miR-15a, and pre-miR-125b, are effective inhibitors of Dicer processing [28]. However, not all pre-miRNAs apical loops appear to be ideal targets for ASO inhibition as efforts to target the apical loop of pre-miR-21 showed no effect on Dicer/TRBP processing [29]. We therefore sought to examine the common features of pre-miRNAs that make them suitable targets for ASO steric blocking and cleavage inhibition.

Previously, we found that the junction region of pre-miR-31 is essential for efficient Dicer/TRBP processing and that disrupting junction base pairing inhibits Dicer processing [30].

Here, we examined the effect of ASOs that can disrupt the pre-miR-31 junction, promoting an open structure. We found that these junction-disrupting ASOs strongly inhibited pre-miR-31 processing by Dicer/TRBP. Furthermore, we demonstrated that pre-miR-144, which is predicted to contain a similar junction region, is also targetable by ASOs for Dicer/TRBP cleavage inhibition. Examination of the predicted pre-miRNA secondary structures from miRbase revealed that junction regions are a common structural feature among human pre-miRNAs (~18%). Of the identified junction-containing pre-miRNAs, we tested our steric blocking ASO approach on pre-miR-19a and pre-miR-143 and found these ASOs were effective inhibitors of Dicer processing. These results suggest that steric blocking ASOs can be broadly used to inhibit Dicer processing of junction-containing pre-miRNAs. Finally, we demonstrate the effectiveness of these steric-blocking ASOs *in cell* using a dual luciferase reporter assay. These findings provide new insights into our understanding of the role of junction regions in regulating pre-miRNA processing and identifies a new approach for modulating pre-miRNA maturation and translational control.

## **4.2 Materials and Methods**

### **4.2.1 Preparation of recombinant human Dicer**

Human Dicer protein was purified as previously described [31, 32] with modifications. Sf9 cells with infected His-tagged Dicer baculovirus was purchased from the University of Michigan protein core. The cell pellet was lysed in ice-cold lysis buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH = 8.0, 300 mM NaCl, 0.5% Triton X-100, 5% glycerol, 0.5 mM tris(2-carboxyethyl) phosphine (TCEP) and 10 mM imidazole) by sonication. The lysate was pelleted by centrifugation at 30,000 x g for 30 min and the supernatant was mixed with 5 mL Ni-NTA resin (Qiagen) pre-equilibrated with lysis buffer in a 50 mL falcon tube. After gently rocking for 1 h at 4 °C, the resin was pelleted by

centrifugation at 183 x g for 10 min. The resin was washed with 45 mL wash buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH = 8.0, 300 mM NaCl, 5% glycerol, 0.5 mM TCEP and 20 mM imidazole) 5 times and eluted with elution buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub> pH = 8.0, 300 mM NaCl, 5% glycerol, 0.5 mM TCEP and 300 mM imidazole). The elution was dialyzed against dialysis buffer (20 mM Tris pH = 7.5, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 50% glycerol) and the purified protein was stored at -80 °C. Total protein concentration was determined by Bradford assay (Thermo Fisher Scientific) and the concentration of Dicer was quantified using ImageJ [33].

#### **4.2.2 Expression and purification of TRBP**

The expression and purification of human TRBP was based on previously described procedures [34, 35]. The pET28a-TRBP was purchased from Addgene (Addgene plasmid # 50351). The pET28a-TRBP plasmid was transformed into *E. coil* Rosetta (DE3) pLysS (Novagen), and cells were grown in LB media until the cell density reached an OD<sub>600</sub>=0.6. Protein expression was induced by addition of 0.2 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) and cells were incubated at 18 °C for 20 h. The cells were harvested by centrifugation, resuspended in buffer A (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM imidazole, and 5 mM β-mercaptoethanol) and lysed by sonication. The lysate was pelleted by centrifugation at 20,000 x g for 30 min and the supernatant was loaded onto a nickel affinity column (Cytiva) and gradient eluted in buffer B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole buffer, and 5 mM β-mercaptoethanol). The protein elution was collected and 10% polyethyleneimine (PEI) was added dropwise (1.5% vol/vol) to remove nucleic acid contaminants. The suspension was stirred at 4 °C for 30 min and the supernatant was collected by centrifugation (30 min at 20,000 rpm). Two sequential ammonium sulfate cuts were performed (with centrifugation for 30 min at 20,000 rpm in between)



at 20% and 80% saturation. The pellet from the 80% ammonium sulfate cut was resuspended in buffer C (20 mM Tris-HCl pH 8.0, 2 mM Dithiothreitol (DTT)) and dialyzed against buffer D (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM DTT) overnight. The dialyzed sample was concentrated and loaded onto a HiLoad 16/600 Superdex 200 column (Cytiva) equilibrated with buffer D.

#### **4.2.3 Dicer/TRBP complex formation**

The purified Dicer and TRBP proteins were mixed at a 1:3 molar ratio and loaded onto a HiLoad 16/600 Superdex 200 column (Cytiva) equilibrated with buffer D. The fractions enriched with complex were pooled, concentrated, and flash frozen in liquid nitrogen for use in processing assays.

#### **4.2.4 Preparation of DNA templates for *in vitro* transcription**

The DNA template for preparation of the HH-pre-miR-31-HDV was generated as previously described [30]. The DNA template for preparation of the pre-miR-144, pre-miR-21, pre-let-7c, GG-pre-miR-19a and pre-miR-143 were generated by overlap-extension (OE) polymerase chain reaction (PCR) using EconoTaq PLUS 2x Master Mix (Lucigen) with primers listed in **Table 4.1**. The OE PCR template for HH-pre-let-7c was digested with EcoRI and BamHI restriction enzymes and inserted into the pUC-19 plasmid. DNA templates for use in *in vitro* transcription reactions were amplified with EconoTaq PLUS 2x Master Mix (Lucigen) using primers UNIV-pUC19\_E105 and HH-pre-let7c-AMP-R (**Table 4.2**).

#### 4.2.5 Preparation of RNA

RNAs were prepared by *in vitro* transcription in 1× transcription buffer [40 mM Tris base, 5 mM dithiothreitol, 1 mM spermidine and 0.01% Triton-X (pH = 8.5)] with addition of 3–6 mM ribonucleoside triphosphates (NTPs), 10–20 mM magnesium chloride (MgCl<sub>2</sub>), 30–40 ng/μL DNA template, 0.2 unit/mL yeast inorganic pyrophosphatase (New England Biolabs)[36], ~15 μM T7 RNA polymerase and 10–20% (v/v) dimethyl sulfoxide (DMSO). Reaction mixtures were incubated at 37 °C for 3–4 h, with shaking at 70 rpm, and then quenched using a solution of 7 M urea and 500 mM ethylenediaminetetraacetic acid (EDTA), pH = 8.5. Transcription reactions were boiled for 3 min and then snap cooled in ice water for 3 min. The transcription mixture was loaded onto preparative-scale 10% denaturing polyacrylamide gels for purification. Target RNAs were visualized by UV shadowing and gel bands containing RNA were excised. Gel slices were placed into an elutrap electroelution device (The Gel Company) in 1× TBE buffer. RNA was eluted from the gel at constant voltage (120 V) for ~24 h. The eluted RNA was spin concentrated, washed with 2 M high-purity sodium chloride, and exchanged into water using Amicon-15 Centrifugal Filter Units (Millipore, Sigma). RNA purity was confirmed on 12% analytical denaturing gels. RNA concentration was quantified via UV-Vis absorbance. Sequences for all RNAs is provided in **Table 4.3**.

#### 4.2.6 <sup>32</sup>P labeling of RNA

The 5'-end labeling of RNA was performed using 5 pmol of RNA, 1 μL γ-<sup>32</sup>P-ATP (PerkinElmer) and 10 U T4 polynucleotide kinase (New England Biolabs) in a final volume of 10 μL. Before labeling, RNA was boiled for 3 minutes, and snap cooled by placing on ice for another 3 minutes. The radiolabeled RNA was purified on a G-25 column (Cytiva) according to the

manufacturer's instructions. The radiolabeled RNA concentration was determined based on a standard curve which was obtained from the counts per minute of the  $\gamma$ - $^{32}\text{P}$ -ATP source.

#### 4.2.7 Dicer/TRBP antisense oligo cleavage assays

Human Dicer/TRBP processing assays were performed as previously described with minimal modifications [30]. Concentrated recombinant human Dicer/TRBP protein was diluted in 1 $\times$  Dicing buffer (24 mM HEPES pH = 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 4  $\mu$ M EDTA). The Dicer/TRBP enzyme was mixed with 80 U RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific), 5 $\times$  dicing buffer (120 mM HEPES pH = 7.5, 0.5 M NaCl, 25 mM MgCl<sub>2</sub>, 0.02 mM EDTA) and relevant antisense oligo (**Table 4.4**). The  $^{32}\text{P}$ -labeled RNA was heated to 95  $^{\circ}\text{C}$  for 3 min and then placed on ice for another 3 min. The  $^{32}\text{P}$ -labeled RNA (1  $\mu$ L) was added to the Dicer/TRBP mixture (9  $\mu$ L) and incubated at 37  $^{\circ}\text{C}$  for a set period of time. The final RNA and enzyme concentration were 50 nM and 20 nM, respectively. The final antisense oligo concentration was either 500 nM or 5  $\mu$ M, as indicated in the figures. The reaction was quenched by addition of 10  $\mu$ L quench buffer (98% Formamide, 20 mM EDTA, trace bromophenol blue and xylene cyanol) at the end time point. The end time points were chosen for each substrate based on a time course study where a substantial fraction of substrate was cleaved (pre-miR-31 at 30 min and pre-miR-144 at 90 min). Experiments were performed in triplicate. The average and standard deviation of the measurements are reported. Significance was determined using one-way ANOVA test.

For the specificity tests, the end time points were 30 min for both pre-miR-21 and pre-let-7c in the pre-miR-31 related assays and 90 min for pre-miR-31 and pre-let-7c in the pre-miR-144 related assays. After the samples were run on 12% denaturing polyacrylamide gels, the gels were

exposed to a phosphor screen, which was scanned by a Typhoon Phosphor Imager (GE Healthcare). The gel image was quantified by ImageJ [33]. The Dicer/TRBP cleavage ratio was calculated as the sum of the intensity of fully-processed and partially-processed products divided by the sum of the intensity of the fully-processed products, partially-processed products, and remaining substrate (total intensity). Experiments were performed in triplicate. The average and standard deviation of the measurements are reported. Significance was determined using one-way ANOVA test.

#### **4.2.8 Dicer/TRBP pre-miRNA competition assays**

Concentrated recombinant human Dicer/TRBP protein was diluted in 1× Dicing buffer (24 mM HEPES pH = 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 4 μM EDTA). The human Dicer/TRBP enzyme complex was pre-mixed with 80 U RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific), 5× dicing buffer (120 mM HEPES pH = 7.5, 0.5 M NaCl, 25 mM MgCl<sub>2</sub>, 0.02 mM EDTA) and relevant LNA-ASO (**Table 4.4**). The RNA mixtures (defined below) were heated to 95 °C for 3 min and then placed on ice for another 3 min. One RNA mixture was composed of 500 nM <sup>32</sup>P-labeled targeted pre-miRNA and 500 nM unlabeled competitive pre-miRNA. In the second RNA mixture, the labeling strategy was inverted. Here 500 nM <sup>32</sup>P-labeled competitive pre-miRNA was mixed with 500 nM unlabeled targeted pre-miRNA. The RNA mixture (1 μL) was added to a pre-mixed solution of Dicer/TRBP/ASO (9 μL, described above) and incubated at 37°C. The final targeted pre-miRNA, competitive pre-miRNA and enzyme concentration were 50 nM, 50 nM, and 20 nM, respectively. The reactions were conducted in parallel with the two RNA mixtures. The final modified antisense oligo concentration in the reactions was 50, 75, 100, 150, 250, 300, and 400 nM. The reaction was quenched by adding 10

$\mu$ L quench buffer (98% Formamide, 20 mM EDTA, trace bromophenol blue and xylene cyanol) at the end time point. The end time point for the pre-miR-31 competition assay was 30 minutes and was 90 minutes for the pre-miR-144 competition assay. The samples were run on 12% denaturing polyacrylamide gels. The gels were exposed to a phosphor screen and then scanned by a Typhoon Phosphor Imager (GE Healthcare). The gel image was quantified by ImageJ [33]. The Dicer/TRBP cleavage ratio was calculated as the sum of the intensity of fully-processed and partially-processed products divided by the sum of the intensity of the fully-processed products, partially-processed products, and remaining substrate (total intensity). Experiments were performed in triplicate. The average and standard deviation of the measurements are reported. Significance was determined using one-way ANOVA test.

#### **4.2.9 Dicer/TRBP pre-miRNA SYBR Gold cleavage assays**

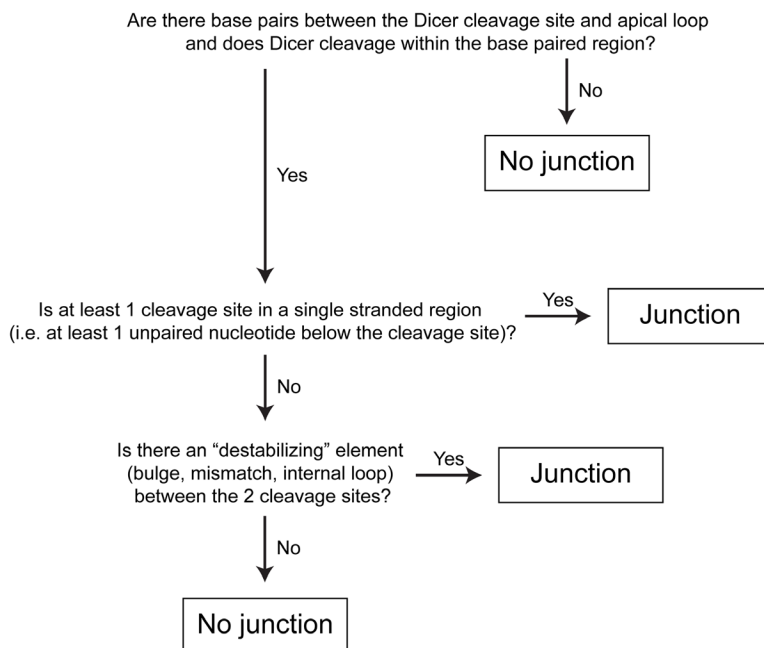
Human Dicer/TRBP protein processing assays were performed as previously described with minimal modifications [30]. Concentrated recombinant human Dicer/TRBP protein was diluted in 1 $\times$  Dicing buffer (24 mM HEPES pH = 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 4  $\mu$ M EDTA). Dicer/TRBP enzyme was pre-mixed with 80 U RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific), 5 $\times$  dicing buffer (120 mM, pH = 7.5, 0.5 M NaCl, 25 mM MgCl<sub>2</sub>, 0.02 mM EDTA) and antisense oligos. The RNA was heated to 95  $^{\circ}$ C for 3 min and then placed on ice for another 3 min. The RNA (1  $\mu$ L) was added to the pre-mixed solution of Dicer:TRBP/ASO (9  $\mu$ L) and incubated at 37  $^{\circ}$ C. The final RNA concentrations for pre-miR-144 and pre-miR-143 were 50 nM and 100 nM for GG-pre-miR-19a. The final enzyme concentration was 20 nM for all reactions. The final antisense oligo concentration was either 500 nM or 5  $\mu$ M for pre-miR-144 and pre-miR-143 and was either 1  $\mu$ M or 10  $\mu$ M for GG-pre-miR-19a. The reaction was quenched by

adding 10  $\mu$ L quench buffer (98% Formamide, 20 mM EDTA, trace bromophenol blue and xylene cyanol) at the end time point. The end time point for pre-miR-144 and pre-miR-143 RNA was 120 min and was 30 min for GG-pre-miR-19a based on kinetic analysis of the substrate processing. After the sample was run on a 20% denaturing polyacrylamide gel, the gel was stained with 1 $\times$  SYBR Gold Nucleic Acid Gel Stain for 30 min (Thermo Fisher Scientific). The gel image was quantified by ImageJ [33]. The Dicer/TRBP cleavage ratio was calculated as the sum of the intensity of fully-processed and partially-processed products divided by the sum of the intensity of the fully-processed products, partially-processed products, and remaining substrate (total intensity). The individual fully-processed 5p strand and 3p strand intensity ratios were calculated as the sum of the intensity of fully-processed 5p or 3p strand intensity divided by the sum of the intensity of the fully-processed products, partially-processed products, and remaining substrate. Experiments were performed in triplicate. The average and standard deviation of the measurements are reported. Significance was determined using one-way ANOVA test.

#### **4.2.10 Cataloging of junction regions of human pre-miRNAs**

All 1,613 pre-miRNA secondary structures were analyzed based on the structure provided in the miRbase human microRNA database [37]. We categorized pre-miRNAs as junction-containing or not based on the following criteria (**Fig. 4.1**). First, we evaluated if there were predicted base pairs between the dicer cleavage site and the apical loop. If one or both cleavage sites were within that base paired region, those pre-miRNAs were classified as not containing a junction region. All other pre-miRNAs would be considered as containing a junction if one of the two additional criteria were met; a) at least one cleavage site is present in a single stranded region with at least one unpaired nt below the cleavage site, and b) if both cleavage sites are within a base

paired region, an unpaired region exists between two cleavage sites. All relevant information, including the cleavage site position (junction region, stem region, apical loop or internal loop), and the size and composition of the base paired junction region, was recorded and analyzed.



**Figure 4.1 Flowchart of junction region-containing pre-miRNAs determination.**

#### 4.2.11 pmirGLO plasmid construction

The pmirGLO plasmid construction was based on the manufacturer manuals. The pmirGLO dual- luciferase miRNA target expression vector was purchased from Promega. The vector was digested with PmeI and XbaI restriction enzymes (New England Biolabs). The inserts for pmirGLO miR144 and pmirGLO miR31 plasmids were purchased from Integrated DNA Technologies (**Table 4.5**) and were also digested with PmeI and XbaI restriction enzyme. The inserts were each ligated with digested pmirGLO plasmids using a quick ligation kit (New England Biolabs) and transformed into JM109 competent cells (Promega). The plasmid was verified by digestion with the NotI restriction enzyme (New England Biolabs), according to

manual.

#### **4.2.12 pCMV plasmid construction**

The pCMV miR144 plasmid was a generous gift from Dr. Eric Lai (Memorial Sloan Kettering Cancer Center). The insert for pCMV miR31 was generated by OE PCR using primers listed in **Table 4.6**. The insert was digested by XhoI and BglIII (New England Biolabs) and ligated with digested pCMV vector. The ligated product was transformed into DH5 $\alpha$  competent cells, and the plasmid integrity was verified by sequencing.

#### **4.2.13 *In cell* dual luciferase assays**

HEK293 cells (ATCC) were grown in Dulbecco's Modified Eagle Medium (DMEM, Gibco) with 10% fetal bovine serum (FBS, Gibco) and 1% Penicilin/Streptomycin. HEK293 cells were detached by trypsin treatment and seeded into 96-well plates with 40K cells per well. Cells were allowed to recover for about 24 h. Subsequently, these cells were co-transfected with 50 ng pmirGLO plasmid (reporter plasmid), 50 ng pCMV plasmid (overexpression plasmid), and 10 nM final concentration of ASO or positive/negative control. Positive control ASOs were purchased from Horizon Discovery. Post transfection, the plates were incubated for another 24 h. The Promega dual-luciferase reporter assay system was used to detect firefly and *Renilla* luciferase activities and the data was analyzed according to the manufacturer protocols.

#### **4.2.14 Quantitative Real-Time PCR Analysis**

HEK293 cells (ATCC) were grown in DMEM (Gibco) with 10% FBS (Gibco) and 1% Penicilin/Streptomycin. HEK293 cells were detached by trypsin treatment and seeded into 24-well



plates with 200K cells per well. Cells were allowed to recover for about 24 h, and then co-transfected with 250 ng pCMV plasmid (overexpression plasmid) and 10 nM final concentration of antisense oligo or positive/negative control. Post transfection, the plates were incubated for another 24 h. Cells were detached by trypsin treatment and RNAs were extracted using miRNeasy Mini Kit (Qiagen). RNA concentration was determined by NanoDrop One C (ThermoFisher Scientific). Expression level of all microRNAs was assessed utilizing the Taqman advanced miRNA assays (Applied Biosystems) and expression level of U6 level was assessed utilizing the Taqman microRNA assay (Applied Biosystems). The complementary DNA (cDNA) was reverse transcribed by Taqman Reverse Transcription Kit (Applied Biosystems) for U6 and Taqman advanced miRNA cDNA synthesis Kit (Applied Biosystems) for miRNAs. U6 was used as internal control for all miRNA levels. The Ct was determined using single threshold method. All experiments were performed in triplicate, with all samples normalized to U6 RNA level and relative expression levels were calculated using  $2^{-\Delta\Delta Ct}$  method.

## **4.3 Results**

### **4.3.1 Antisense oligos inhibit pre-miR-31 cleavage by Dicer/TRBP**

RNAs are known to be conformationally dynamic molecules, and the conformational dynamics of pre-miRNAs can be important for regulating Dicer/TRBP binding and processing. Studies on pre-miR-21 revealed the existence of both ground and excited state conformations and demonstrated that Dicer cleaved the excited state conformation more efficiently than the ground state conformation [38]. Pre-miR-21 processing can also be inhibited by drug-like small molecules, which bind to pre-miR-21 and stabilize it into a conformation that is a poor substrate for Dicer processing [35]. We previously identified the junction region is a regulatory element in

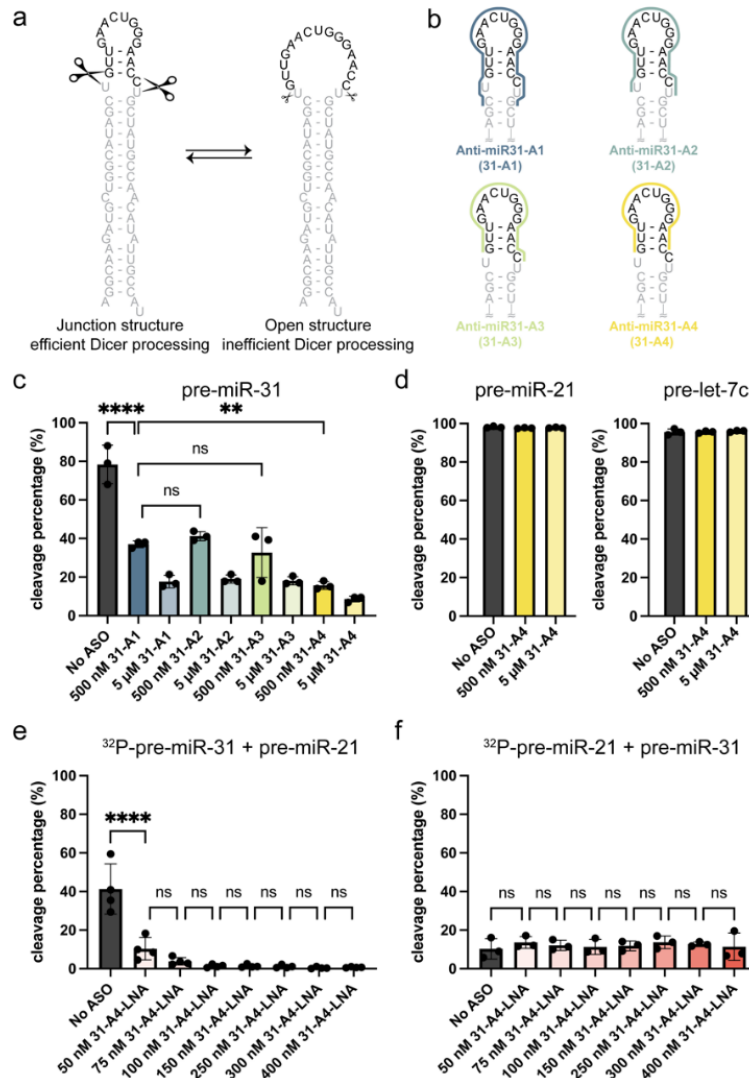
pre-miR-31 and demonstrated that altering the stability of this region affected Dicer/TRBP cleavage (**Fig. 4.2a**) [30]. The open structure of pre-miR-31, which does not contain a pre-formed junction region, is a poor substrate for Dicer/TRBP cleavage [30]. We therefore explored the use of steric-blocking antisense oligos that could compete with pre-miR-31 junction and stabilize the RNA into an “open” conformation to inhibit pre-miR-31 maturation.

We designed a series of antisense oligos, intended to bind to the pre-miR-31 apical loop region and disrupt the junction base pairs (**Fig. 4.2b, Table 4.4**). Anti-miR31-A1 (31-A1) is the longest ASO, spanning the apical loop, junction, dicing site, and disrupting one additional C-G base pair directly beneath the internal loop. Anti-miR31-A2 (31-A2) is complementary to full apical loop sequence of the open conformation for pre-miR-31, which partially sequesters the Dicer cleavage site in the loop. Anti-miR31-A3 (31-A3) binds to the apical loop sequence of the pre-miR-31 without spanning the Dicer cleavage site (i.e. the sequence that would be cleaved by Dicer). Anti-miR31-A4 (31-A4) is the shortest ASO and only disrupts the junction base pairs, annealing across the apical loop. We expected all ASOs to disrupt junction base pairing, promoting intermolecular rather than intramolecular interactions at the junction region. *In vitro* Dicer/TRBP processing assays on pre-miR-31 in the absence or presence of different concentrations of various Anti-miR31 ASOs revealed significant inhibition of processing with 31-A4 providing the highest degree of inhibition (**Fig. 4.2c**). Subsequent studies were therefore conducted only with 31-A4.

We next used the siRNA Target Finder tool (<https://www.genscript.com/tools/sirna-target-finder>) to determine if 31-A4 would target any other pre-miRNA sequence. No pre-miRNAs were identified that had sequences complementary to 31-A4. To examine the specificity of 31-A4 inhibition, we assessed the processing of pre-miR-21 and pre-let-7c incubated with excess amount of 31-A4. We did not observe a detectable reduction in Dicer/TRBP processing for either pre-miR-

21 or pre-let-7c in the presence of 31-A4 ASO (**Fig. 4.2d**). These results suggest that 31-A4 acts specifically to inhibit pre-miR-31 processing and is not a global repressor of Dicing for pre-miRNAs.

While our *in vitro* assays were conducted with DNA oligonucleotides containing a native phosphodiester backbone, in the cell, naked DNAs are readily degraded and are therefore not suitable for antisense oligo design [1]. A variety of sugar, backbone, and base modifications can be incorporated into the ASO to enhance the ASO stability and binding affinity [1, 16, 39, 40]. We therefore ordered a new version of the 31-A4 ASO that contained a modified phosphorothioate backbone and mixed locked nucleic acid (LNA) bases where the 2' oxygen and the 4' carbon of some pentose sugars were linked with a methylene bridge (anti-miR31-A4-LNA, 31-A4-LNA, **Table 4.4**). We employed a competition assay to assess the impact of our modified ASO on the Dicer/TRBP processing of pre-miR-31 in the presence of a competitor pre-miRNA. Here, equal amounts of the target pre-miRNA (pre-miR-31) and a competitor pre-miRNA (pre-miR-21) were mixed; one of the pre-miRNAs was radiolabeled at the 5'-end while the other pre-miRNA was unlabeled. Incubation of the pre-miRNA mixture in the presence and absence of 31-A4-LNA and addition of Dicer permits the analysis of the radiolabeled pre-miRNA. By changing which pre-miRNA was radiolabeled, we were able to detect the cleavage percentage of each pre-miRNA in the competition cleavage assay. We found that pre-miR-31 cleavage was inhibited by 31-A4-LNA in a dose dependent manner (**Fig. 4.2e**) while there was no detectable impact on pre-miR-21 cleavage by Dicer/TRBP (**Fig. 4.2f**). Collectively, we found that all junction-spanning ASOs of pre-miR-31 strongly inhibit Dicer/TRBP cleavage with no off-target effects on other pre-miRNAs examined.

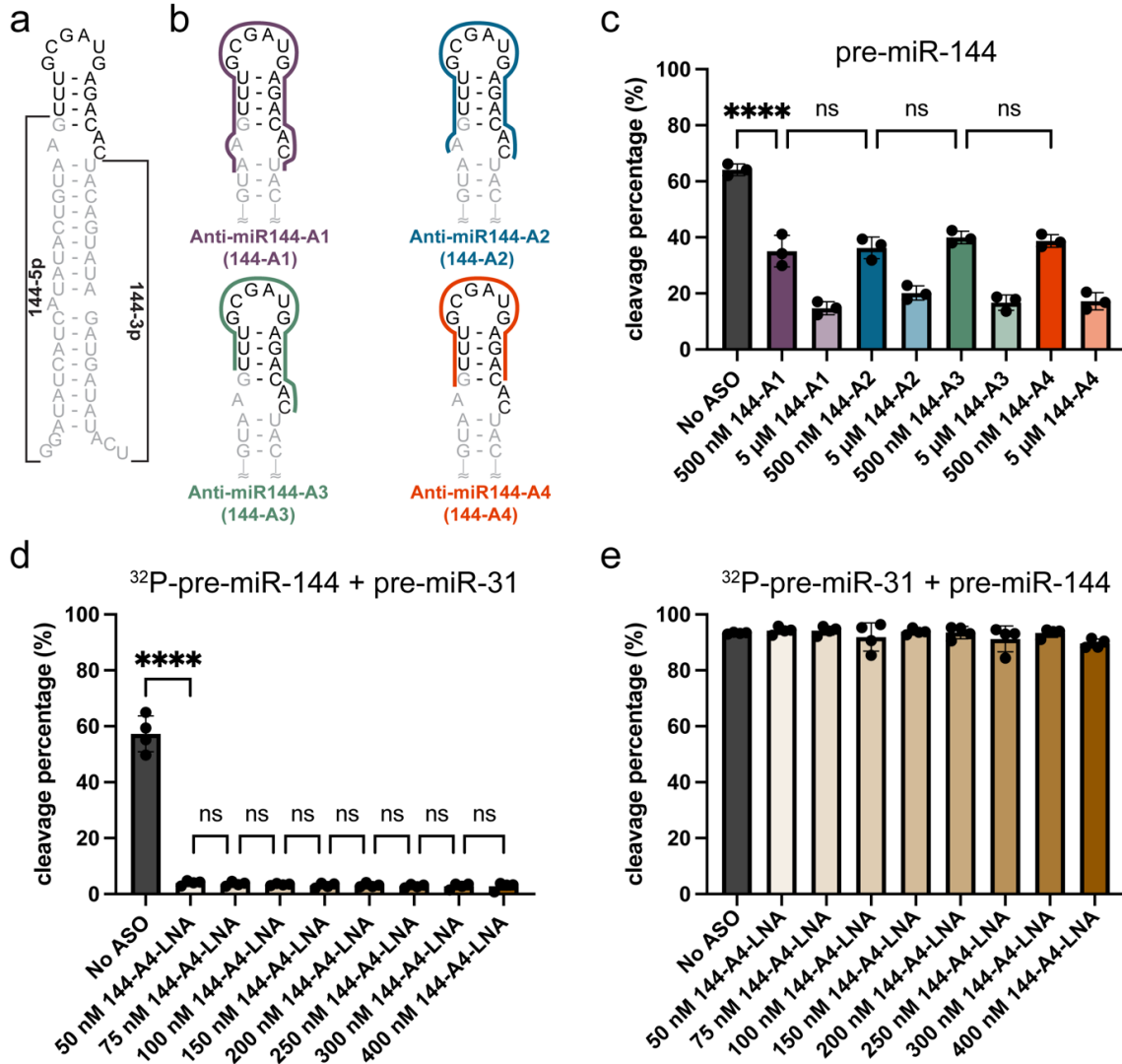


**Figure 4.2. Antisense oligos that target the pre-miR-31 junction structure are potent and specific inhibitors of Dicer/TRBP processing in vitro.** **a)** Pre-miR-31 conformational equilibrium highlighting the junction structure containing structure, a good substrate for processing by the Dicer/TRBP complex, and an open loop structure, which is a poor substrate for Dicer/TRBP processing. The mature miR-31 duplex sequence is gray and the cleaved off apical loop is black. Scissors indicate Dicer cleavage sites with sizes to reflect the extent to which each conformer is processed. **b)** Pre-miR<sub>31</sub> ASO design. Colored lines span the nucleotides within pre-miR-31 that each anti-miR-31 ASO is complimentary to. **c)** Effect of addition of anti-miR-31-A1-A4 to a pre-miR-31 Dicer/TRBP cleavage assay. **d)** Dicer/TRBP processing assays of pre-miR-21 and pre-let-7c in the presence and absence of anti-miR31-A4. **e)** Competition cleavage assay conducted on <sup>32</sup>P-pre-miR-31 and unlabeled pre-miR-21 in the presence of anti-miR-31-A4-LNA. **f)** Competition cleavage assay conducted on <sup>32</sup>P-pre-miR-21 and unlabeled pre-miR-31 in the presence of anti-miR-31-A4-LNA. For all processing assays, the average and standard deviation from n=3 or n=4 independent assays are presented. \*\*\*\* p<0.0001, \*\*p<0.01, and ns indicates no significant difference from an ordinary one-way ANOVA Tukey analysis.

### 4.3.2 Pre-miR-144 is a targetable pre-miRNA containing a junction region

Secondary structure probing of pre-miR-144 revealed a junction-containing structure [41]. Mutants that disrupted the pre-miR-144 junction base pairs inhibited Dicer processing *in cell* [42], similar to our observations in pre-miR-31 [30]. We therefore designed steric blocking ASOs to target the pre-miR-144 junction region and assessed the effect of the ASOs on inhibiting its Dicer-mediated maturation.

Using the secondary structure of pre-miR-144 as a reference (**Fig. 4.3a**), we designed four ASOs based on similar principles described above for pre-miR-31 (**Fig. 4.3b**). Anti-miR144-A1 (144-A1) was designed to fully destabilize the junction, span the dicing site, and destabilize one additional A-U base pair positioned directly beneath the 1x2 internal loop. Anti-miR144-A2 (144-A2) spans the internal loop, junction region, and apical loop while anti-miR144-A3 (144-A3), and anti-miR144-A4 (144-A4) are shorter ASOs that destabilize the junction base pairs, locking the junction region in an “open” conformation. In our *in vitro* processing assays, all pre-miR-144 ASOs inhibited Dicer/TRBP cleavage to a similar extent (**Fig. 4.3c**). The 144 ASOs did not impact the processing of other pre-miRNAs by Dicer/TRBP (**Fig. 4.4a, b**), even for the pre-miR-31, which also contains a junction region.

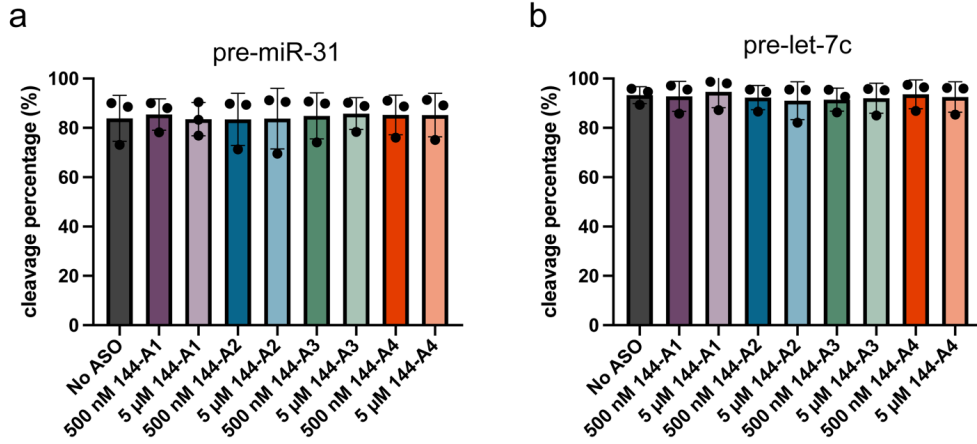


**Figure 4.3. ASOs targeting the pre-miR-144 junction inhibit Dicer/TRBP processing in vitro.** **a)** Predicted secondary structure of pre-miR-144. The mature miR-144 duplex sequence is gray and the cleaved apical loop is black. The 144-5p and 144-3p products are labeled, 144-3p serves as the guide strand. **b)** ASO design. Colored lines span the nucleotides within pre-miR-144 that each anti-miR-144 ASO is complimentary to. **c)** Effect of addition of anti-miR-144-A1-A4 to a pre-miR-144 Dicer/TRBP cleavage assay. **d)** Competition cleavage assay conducted on  $^{32}\text{P}$ -pre-miR-144 and unlabeled pre-miR-31 in the presence of anti-miR-144-A4-LNA. **e)** Competition cleavage assay conducted on  $^{32}\text{P}$ -pre-miR-31 and unlabeled pre-miR-144 in the presence of anti-miR-144-A4-LNA. For all processing assays, the average and standard deviation from  $n=3$  or  $n=4$  independent assays are presented. \*\*\*\*  $p < 0.0001$  and ns indicates no significant difference from an ordinary one-way ANOVA Tukey analysis.

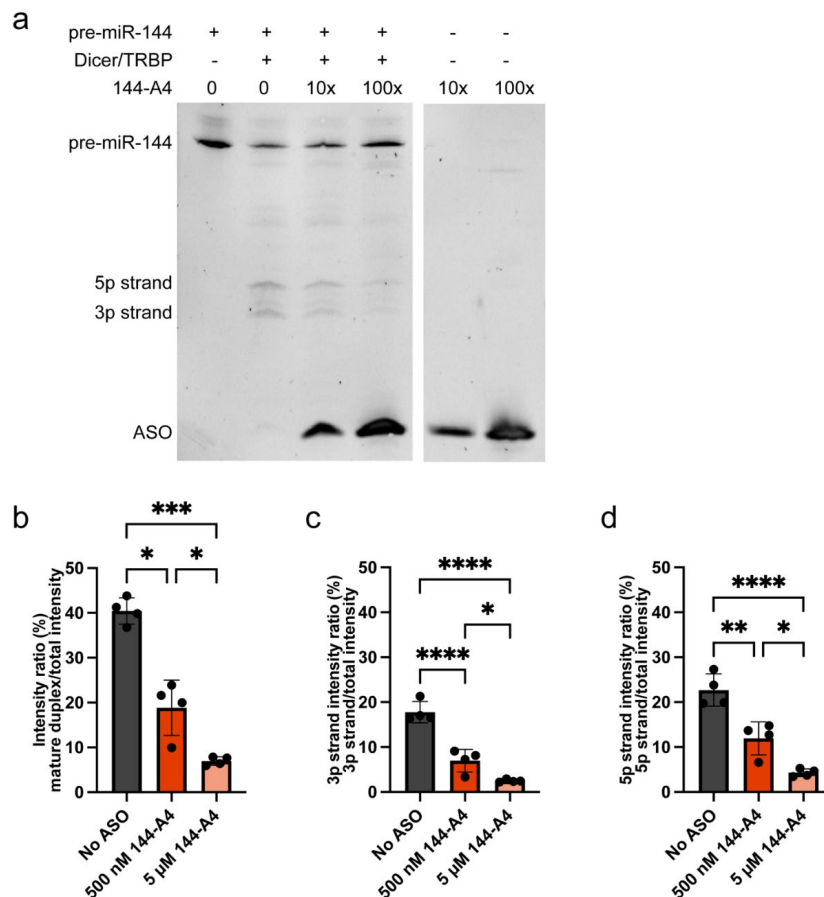
Because all of the anti-pre-miR-144 ASOs that we tested inhibited Dicer processing to a similar extent, we moved forward with the 144-A4 construct, the minimal ASO construct that only

disrupts the junction formation. The guide strand is the functional component of the mature miRNA duplex which forms the RISC and promotes sequence-specific binding to the target mRNA. The production of the guide strand therefore plays an important role in post-transcriptional gene expression. In pre- miR-144, the guide strand is the 3'-end cleavage product (3p, **Fig. 4.3a**), which is masked in our 5'-<sup>32</sup>P radiolabeled experiments. To ensure that the ASO didn't differentially impact generation of the 5' (5p) or 3' (3p) product, we performed Dicer/TRBP cleavage assays and analyzed all processing products using SYBR gold stain (**Fig. 4.5**). Consistent with our observations from the 5'-<sup>32</sup>P radiolabeled processing experiments, we observed that addition of 144-A4 inhibited the production of mature miR-144 duplex when analyzed by SYBR staining (**Fig. 4.5b**). Furthermore, we analyzed the production of the miR-144 3p and 5p strands individually. Here we found a similar dose-dependent inhibition of 5p and 3p products in the presence of 144-A4, indicating that the production of both strands of mature miR-144 was inhibited to a similar extent (**Fig. 4.5c, d**). Unlike using a radiolabeled substrate, the band intensity of SYBR stained gels is dependent on RNA size, therefore we cannot make direct comparisons across samples. Importantly, these findings ensured that both strands of pre-miR-144 were produced and the processing of the 3p strand was also inhibited by addition of 144-A4.

We next examined the effect of a modified anti-miR144-A4 ASO (144-A4-LNA) on Dicer processing. Competition cleavage assays were conducted on pre-miR-144 in the presence of pre-miR-31, which contains a similar junction region. We observed strong cleavage inhibition of pre-miR-144 by 144-LNA (**Fig. 4.3d**) while pre-miR-31 cleavage levels were unaffected (**Fig. 4.3e**), indicating that the pre-miR-144 ASO can strongly inhibit pre-miR-144 maturation without perturbing the processing of other similarly structured pre-miRNA.

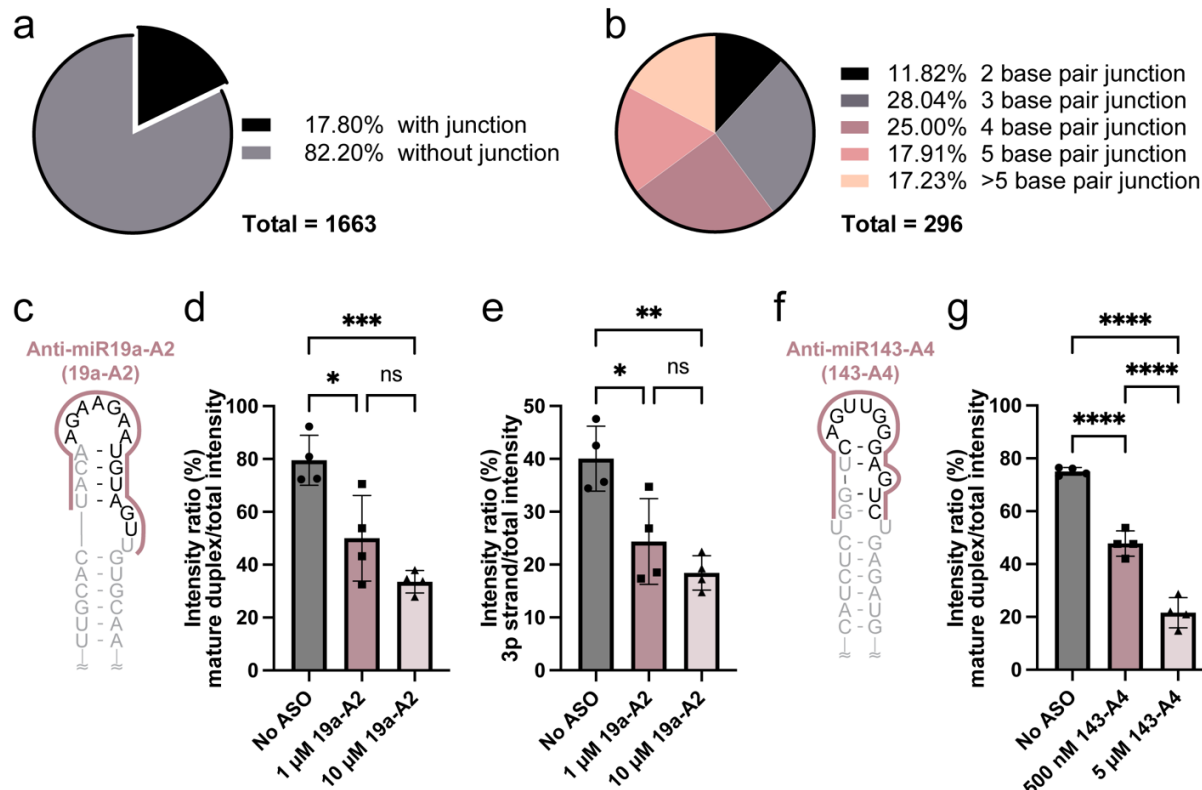


**Figure 4.4. Anti-miR144 ASO are specific and do not affect the processing of pre-miR-31 or pre-let-7c.** **a)** Dicer/TRBP processing of pre-miR-31 in the presence and absence of different anti-miR144 ASOs. **b)** Dicer/TRBP processing of pre-let-7c in the presence and absence of different anti-miR144 ASOs. No significant differences were identified from an ordinary one-way ANOVA Tukey analysis.



**Figure 4.5. ASO 144-A4 inhibits Dicer/TRBP processing of pre-miR-144.** **a)** Representative pre-miR-144 processing assay, visualized by SYBR gold staining. **b-d)** Quantification of pre-miR-144 processing products. \*\*\*\* p<0.0001, \*\*\*p<0.001, \*\*p<0.01, and \*p<0.05 from an ordinary one-way ANOVA Tukey analysis.





**Figure 4.6. Junction regions are a common feature of pre-miRNAs and are effective ASO targets.** **a)** Analysis of secondary structures predicted in the miRbase reveal the prevalence of junction-containing pre-miRNAs. **b)** Breakdown of predicted junction length. **c)** Pre-miR-19a is predicted to contain a four base pair junction. Anti-miR19a-A2 binding site is indicated. The mature miR-19a duplex sequence is gray and the cleaved apical loop is black. **d)** Quantification of pre-miR-19a processing products with increasing concentrations of anti-miR19a-A2. **e)** Quantification of the pre-miR-19a 3p strand product in Dicer/TRBP processing assays that include increasing concentrations of anti-miR19a-A2. **f)** Pre-miR-143 predicted secondary structure with the anti-miR143-A4 binding site indicated. The mature miR-143 duplex sequence is gray and the cleaved off apical loop is black. **g)** Quantification of pre-miR-143 processing products with increasing concentrations of anti-miR143-A4. For all processing assays, average and standard deviation from n=4 independent assays are presented. \*\*\*\* p<0.0001, \*\*\*p<0.001, \*\*p<0.01, \*p<0.05, and ns indicates no significant difference from an ordinary one-way ANOVA Tukey analysis.

### 4.3.3 The junction region is a common structural feature within pre-miRNAs

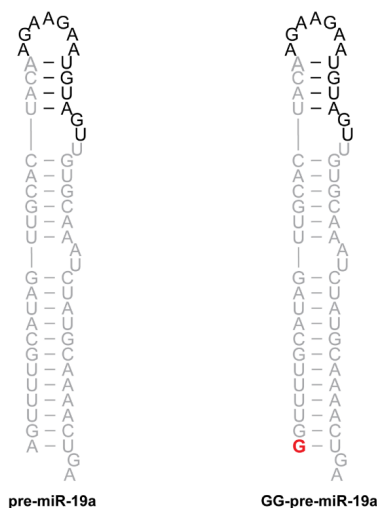
We previously showed that the stability of the junction region is important for pre-miR-31 processing [30]. Furthermore, we demonstrated that the junction region is a targetable element for inhibiting Dicer/TRBP cleavage of pre-miR-31 (Fig. 4.2) and pre-miR-144 (Fig. 4.3) using ASOs. We were therefore interested to know how many pre-miRNAs are predicted to contain a junction

region. We examined the predicted secondary structures of pre-miRNAs provided in the miRbase database [37, 43-46] and classified pre-miRNAs based on whether or not they are predicted to contain a junction (**Fig. 4.1**) Among the 1,663 pre-miRNAs, 296 (~18%) are predicted to contain a junction region (**Fig. 4.6a**). Of these 296 pre-miRNAs, ~70% are predicted to contain junction regions between three and five base pairs in length (**Fig. 4.6b**).

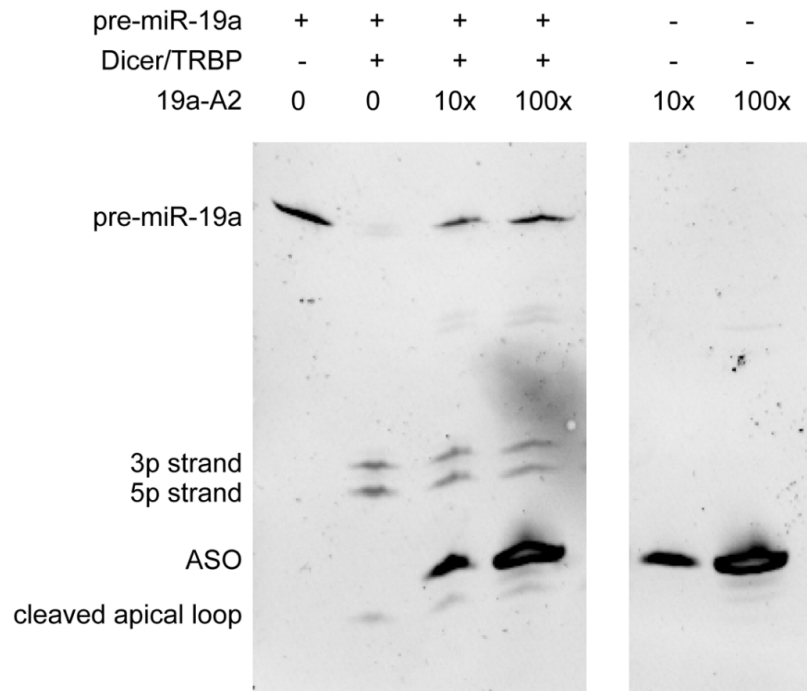
#### **4.3.4 Targeting junction containing pre-miRNAs is a reasonable approach for inhibiting Dicer/TRBP processing**

To expand our study of ASOs targeting junction-containing pre-miRNAs, we conducted additional processing assays using pre-miR-19a and pre-miR-143, two pre-miRNAs predicted to contain four base pair junction based on our analysis. We designed two antisense oligos, anti-miR19a-A2 (19a-A2) and anti-miR143-A4 (143-A4), which were designed to target pre-miR-19a and pre-miR-143, respectively (**Fig 4.6c, f**). To simplify the transcription and purification of pre-miR-19a, the first adenosine was replaced with a guanosine to generate GG-pre-miR-19a, which has the same predicted secondary structure as the native pre-miR-19a (**Fig. 4.7**). Due to the high A U content within the junction and apical loop of pre-miR-19a, we designed a longer A4-like ASO, spanning the dicer cleavage site, junction region, and internal and apical loops, to ensure tight binding under the reaction conditions. Incubation with 10-fold and 100-fold excess ASO resulted in significant inhibition of Dicer processing of both GG-pre-miR19a, and pre-miR-143 (**Fig. 4.6c-g**). The guide strand (3p) intensity ratio for GG-pre-miR-19a was also significantly reduced when treated with high concentrations of 19a-A2 (**Fig 4.6e**), as the intact GG-pre-miR-19a accumulates (**Fig. 4.8**). For pre-miR-143, we detected similar inhibitory effects when treated with a high concentration of 143-A4 (**Fig. 4.6g, 4.9**).

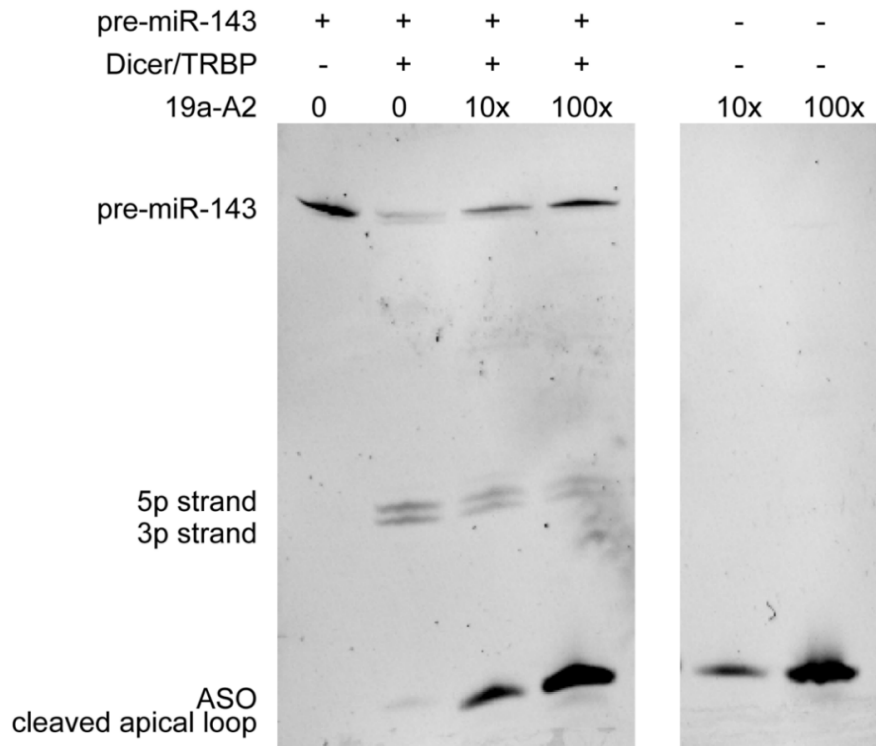
While the molecular details of the interaction between the human Dicer helicase domain and the apical loop of pre-miRNAs is masked in human Dicer-RNA complex structures, this domain provides important contacts that mediate the specific processing of pre-miRNAs [47-49]. We therefore sought to determine if the inhibition of Dicer processing observed in our ASO assays is the result of disruption between the pre-miRNA apical loop and Dicer helicase domain recognition. We selected pre-let-7c from the pool of non-junction containing pre-miRNAs to investigate this possibility. We designed anti-let-7c-A1 (**Fig. 4.10a, Table 4.4**), a 14-nt ASO designed to target the apical loop region of the non-junction-containing pre-let-7c. Interestingly, we found that even 100-fold excess of anti-let-7c-A1 did not significantly inhibit Dicer/TRBP cleavage on pre-let-7c (**Fig. 4.10b**), suggesting that the ASO does not function as a steric block to Dicer/TRBP recognition. Collectively, our findings suggest that, for junction-containing pre-miRNAs, targeting of the junction and apical loop regions is an alternative approach for inhibition of Dicer processing. This success suggests that the junction region is an important domain within a subset of pre-miRNAs, and that junction formation facilitates the efficient cleavage by Dicer/TRBP.



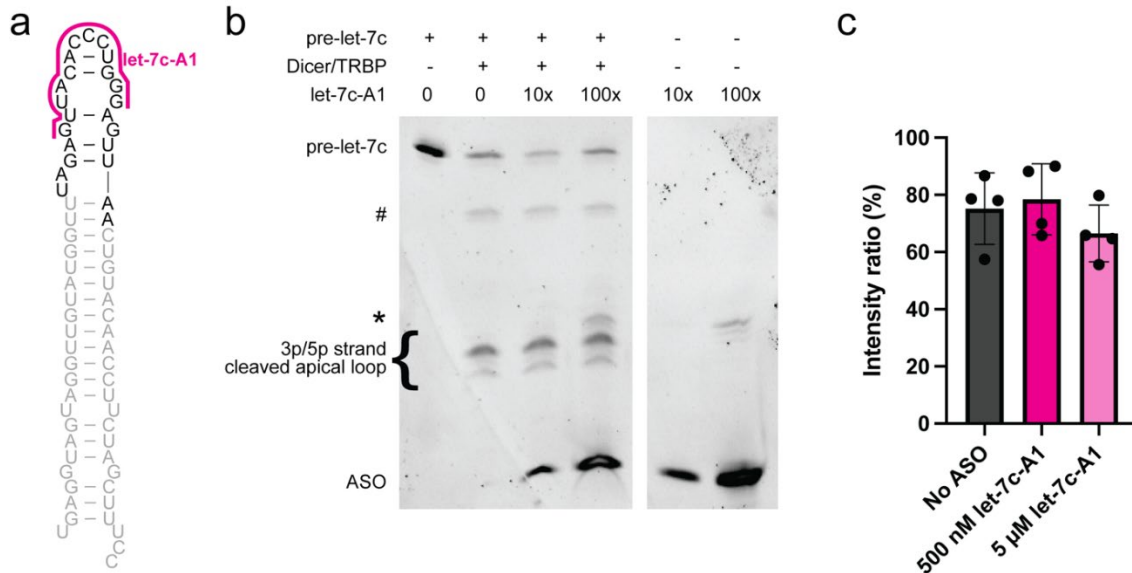
**Figure 4.7. Secondary structure of pre-miR-19a and GG-pre-miR-19a**



**Figure 4.8. Representative GG-pre-miR-19a processing assay, visualized by SYBR gold staining.**



**Figure 4.9. Representative pre-miR-143 processing assay, visualized by SYBR gold staining.**



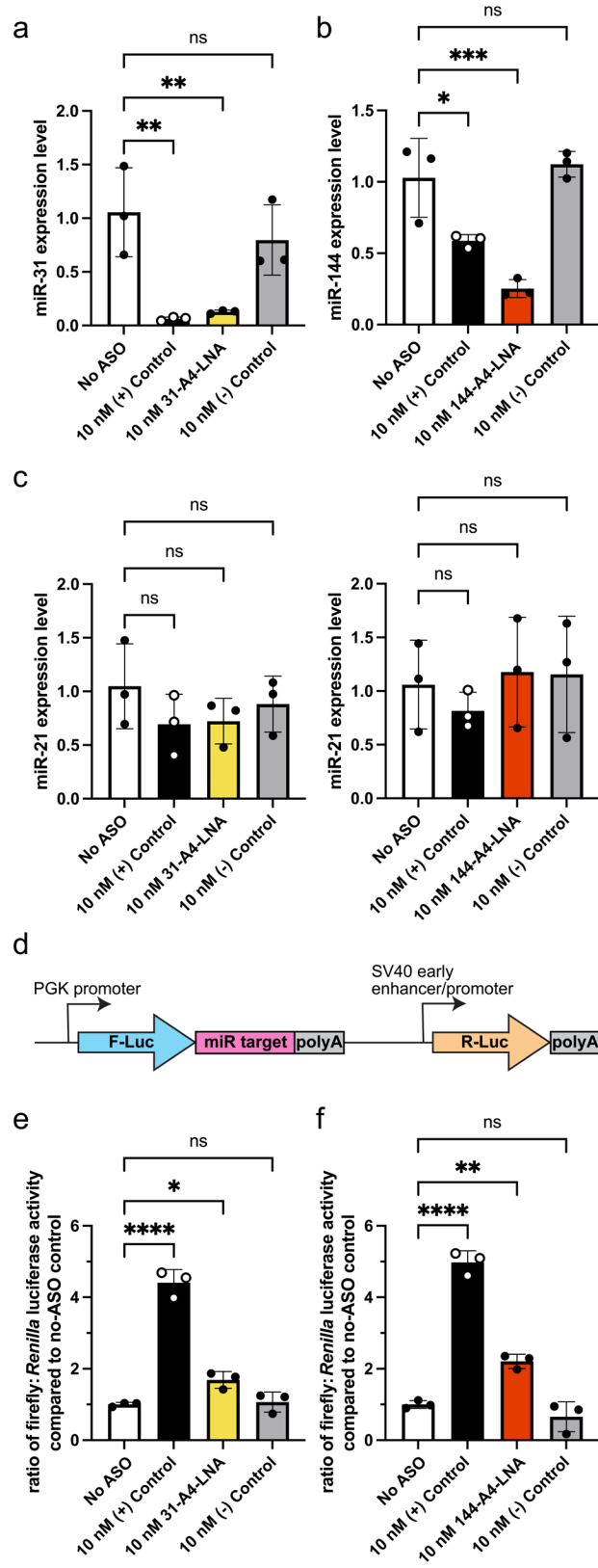
**Figure 4.10. Apical loop region targeted ASO does not affect pre-let-7c processing by Dicer/TRBP** **a)** Predicted secondary structure of pre-let-7c. The mature let-7c duplex sequence is gray and cleaved off apical loop is black. **b)** Representative pre-let-7c processing assay, visualized by SYBR gold staining. # indicates single processed pre-let-7c product, \* indicates an impurity from the anti-let-7c-A1 ASO. **c)** The intensity ratio of mature let-7c duplexes is not affected as let-7c-A1 concentration increases in the reaction.

#### 4.3.5 ASO targeting of junction-containing pre-miRNAs inhibit Dicer cleavage *in cell*

Our *in vitro* processing assays revealed that ASOs targeting the junction region of pre-miRNAs were inhibitors of Dicer/TRBP processing. We next sought to validate the inhibitory effects of 31-A4-LNA and 144-A4-LNA on Dicer/TRBP processing of the pre-miRNA elements *in cell*. First, we quantified mature miRNA levels (miR-31 or miR-144) in HEK293 cells after transfection with the relevant ASO using quantitative reverse transcription polymerase chain reaction (qRT-PCR). We detected significant inhibition of mature miR-31 levels in cells transfected with either commercially available positive controls (complementary to the mature miRNA sequence) or our 31-A4-LNA (**Fig. 4.11a**). We did not detect significant changes in mature miR-31 levels in cells transfected with a negative control (**Table 4.4**). Transfection of HEK293 cells expressing pre-miR-144 with 144-A4-LNA also strongly inhibited miR-144

production (**Fig.4.11b**). Importantly, we verified that ASO transfection did not affect off-target miR-21 levels (**Fig. 4.11c**).

Both 31-A4-LNA and 144-A4-LNA ASOs were effective in reducing mature miRNA levels. We therefore used a dual luciferase assay to determine if the ASOs affected gene expression (**Fig. 4.11d**). We co-transfected an overexpression plasmid (for either pre-miR-31 or pre-miR-144), relevant ASOs, and the respective dual luciferase reporter plasmid into HEK293 cells. After incubation for 24 hours, we quantified the relative luciferase activity. We found that the luciferase signal was elevated in both 31-A4-LNA and 144-A4-LNA treated samples, relative to their No ASO treated controls (**Fig. 4.11e,f**). These results further confirm that our junction region targeted ASOs inhibit miRNA production *in cells*.



Legend expanded into next page

**Figure 4.11. Anti-junction-region ASOs can inhibit pre-miRNA maturation in cell.** **a)** RT-qPCR analysis of mature miR-31 expression levels in HEK293 cells upon treatment with various ASOs. **b)** RT-qPCR analysis of mature miR-144 expression levels in HEK293 cells upon treatment with various ASOs. **c)** RT-qPCR analysis of mature miR-21 expression levels in HEK293 cells within miR-31 (left) and miR-144 (right) *in cell* experiment samples. **d)** Diagram of the pmirGLO vector for the dual luciferase assays. The expression of firefly luciferase (F-Luc, blue) is under the control of the target mature miRNA while *Renilla* luciferase (R-Luc, yellow) expression is not controlled by any miRNA sequence. R-Luc luciferase signals are used as internal normalization control. **e,f)** Luciferase reporter assay for miR-31 (**e**) and miR-144 (**f**) with various ASO treatment in HEK 293 cells. The post-transcriptional inhibition of F-Luc translation by miR-31 and miR-144 are recovered by anti-junction-region ASOs, suggesting that these ASO function to reduce the levels of miR-31 and miR-144. For all assays, average and standard deviation from n=3 independent assays are presented. \*\*\*\* $p < 0.0001$ , \*\* $p < 0.01$ , \* $p < 0.05$ , and ns indicates no significant difference from an ordinary one-way ANOVA analysis.

#### 4.4 Discussion

ASOs play important roles in nucleic therapeutics and there is growing interest in developing novel ASOs that target different aspects of nucleic acid biology [1, 50]. MicroRNAs are one type of nucleic acid that have been targeted by ASOs for therapeutic purposes [51, 52]. However, most ASOs target the mature sequence of microRNAs, with relatively little research aimed at targeting pri/pre-element microRNA using ASOs [28]. Interestingly, a virus-encoded miRNA (HHV-6A miR-aU14) has been shown to inhibit pre-miR-30 processing through RNA-RNA interactions within the pre-miR-30 apical loop [53]. However, the artificial ASOs, specifically designed to inhibit pre-miRNA processing have remained elusive [28, 29]. Our previous research uncovered the importance of the pre-miR-31 junction region in regulating Dicer/TRBP processing [30]. We found that the junction-containing structure was a good substrate for Dicer/TRBP processing while mutations that disrupted the junction base pairs reduced Dicer/TRBP processing. Therefore, pre-miR-31 serves as a model for investigating if ASOs could be used to disrupt the junction structures and inhibit Dicer/TRBP processing.



We found that blocking junction base pair formation within pre-miR-31 and locking the RNA into an open junction conformation strongly inhibited Dicer/TRBP processing. These junction-targeting ASOs show high specificity in inhibiting pre-miR-31 processing while not inhibiting Dicer/TRBP cleavage of other pre-miRNAs. Additionally, a highly modified version of the anti-miR-31-A4 ASO strongly inhibited pre-miR-31 cleavage in the presence of other pre-miRNAs both *in vitro* and *in cell*.

The utility of junction-targeting ASOs was not limited to pre-miR-31. Pre-miR-144 has a similar junction structure and ASOs designed to disrupt that junction structure showed high specificity and high efficiency for inhibiting Dicer/TRBP cleavage. Additionally, our *in cell* further validate these findings. The presence of a junction region in both pre-miR-144 and pre-miR-31 may explain why we were successful in targeting these RNAs with ASOs while previous studies using ASOs to target the pre-miR-21 (which does not contain a junction) apical loop had no impact on Dicer/TRBP cleavage [29].

Our findings reveal that a number of pre-miRNAs are predicted to contain junction regions and may be targetable using this antisense oligo approach. Our findings pave a new way for steric blocking antisense oligos design in inhibiting pre-miRNA cleavage by Dicer/TRBP.

**Table 4.1. Overlap extension PCR primers for generation of pre-miRNA templates.**

Construct	Primer name	5'-sequence-3'
HH-pre-miR-21	HH-pre-miR-21-1F	TAATACGACTCACTATAGGCTCG
	HH-pre-miR-21-2R	ACCCTGATGGTGTCTGAAAAACGTACCCTGATGGTGTACGAGCCTATAGTGAGTCGTAT
	HH-pre-miR-21-3F	TTCAGACACCATCAGGGTCTGCTGATAAGCTACTGATGAGTCCGTGAGG
	HH-pre-miR-21-4R	CTAGACGGTACCGGGTACCGTTTCGTCCTCACGGACTCATCAGTA
	HH-pre-miR-21-5F	ACCCGGTACCGTCTAGCTTATCAGACTGATGTGACTGTTGAATCTCATGGCAACACCAG
	HH-pre-miR-21-6R	mGmACAGCCCATCGACTGGTGTGCCATGAGATT
pre-miR-let7c	pre-miR-let7c-nat-1F	CCGGAATTCTAATACGACTCACTATAGGGCTCGTACACCATCAGGGTACGTTTTTCA
	pre-miR-let7c-nat-2R	GCAGACCCTGATGGTGTCTGAAAAACGTACCCTGATGGTGTACGAGCC
	pre-miR-let7c-nat-3F	ACACCATCAGGGTCTGCTACTACCTCACTGATGAGTCCGTGAGGACGAA
	pre-miR-let7c-nat-4R	CCTCAGACGGTACCGGGTACCGTTTCGTCCTCACGGACTCAT
	pre-miR-let7c-nat-5F	CCGGTACCGTCTGAGGTAGTAGGTTGTATGGTTTAGAGTTACACCCTGGGAGTAACTGT
	pre-miR-let7c-nat-6R	CCGTCGCGGATCCGGAAAGCTAGAAGGTTGTACAGTTAACTCCCAGGGTGTA
HH-pre-miR-144	HH-pre-144-1F	TAATACGACTCACTATAGGCTCG
	HH-pre-144-2R	ACCCTGATGGTGTCTGAAAAACGTACCCTGATGGTGTACGAGCCTATAGTGAGTCGTATT
	HH-pre-144-3F	TTCAGACACCATCAGGGTCTGTGATGATATCCCTGATGAGTCCGTG
	HH-pre-144-4R	CGACGGTACCGGGTACCGTTTCGTCCTCACGGACTCATCAGGGA
	HH-pre-144-5F	ACCCGGTACCGTCGGATATCATCATATACTGTAAAGTTTGCATGAGACACTAC
	HH-pre-144-6R	mAmGTACATCATCTATACTGTAGTGTCTCATCGCAAATT

GG-pre-miR-19a	GG-pre-miR19a-F	TAATACGACTCACTATAGGTTTTGCATAGTTG CACTACAAG
	GG-pre-miR19a-R	mUmCAGTTTTGCATAGATTTGCACA TTCTTCTTGTAGTGCAACTATGCAAAACC
Pre-miR-143	pre-miR143-F	TAATACGACTCACTATAGGTGCAGTGCTGCA TCTCTGGTCAGTT
	pre-miR143-R	mGmAGCTACAGTGCTTCATCTCAGACTCCCA ACTGACCAGAGATGCAGCACTGCACCT

**Table 4.2. Amplification primers for template.**

Amplification primer	5'-sequence-3' <sup>a</sup>	application
UNIV-pUC19_E105	TCTTCGCTATTACGCCAGCTGGCGA AA	Forward primer for amplification of DNA template for HH-pre-let-7c, HH-pre-miR-31 from plasmid
HDV-AMP-R	mUmAATGTGAGAATTGGCTACGTT GAAACAACGCATTACCG	Reverse primer for amplification of DNA template for HH-pre-miR-31-HDV from plasmid
HH-pre-let7c-AMP-R	mGmGAAAGCTAGAAGGTTGTACAG TTAACTCCCAGGG	Reverse primer for amplification of DNA template for HH-pre-let-7c from plasmid

**Table 4.3. RNA sequences**

RNA name	5'-sequence-3' <sup>a</sup>
pre-miR-31	AGGCAAGAUGCUGGCAUAGCUGUUGAACUGGGAACCUGCUAUGC CAACAUAUUGCCA
pre-miR-21	UAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGGCAACAC CAGUCGAUGGGCUGU
pre-let-7c	UGAGGUAGUAGGUUGUAUGGUUUAGAGUUACACCCUGGGAGUU AACUGUACAACCUUCUAGCUUCC
pre-miR-144	GGUAUUAUCAUAUACUGUAAGUUUGCGAUGAGACACUACAGU AUAGAUGAUGUACU
GG-pre-miR-19a	gGUUUUGCAUAGUUGCACUACAAGAAGAAUGUAGUUGUGCAAU CUAUGCAAACUGA
pre-miR-143	GGUGCAGUGCUGCAUCUCUGGUCAGUUGGGAGUCUGAGAUGAA GCACUGUAGCUC

<sup>a</sup>Non-native residues are in lowercase.

**Table 4.4. Antisense oligo sequences.**

Antisense oligo	primer names	5'-sequence-3' <sup>1a</sup>
Anti-miR31-A1	anti-miR31-D3	CAGGTTCCCAGTTCAACAG
Anti-miR31-A2	anti-miR31-loop-D	AGGTTCCCAGTTCAACA
Anti-miR31-A3	anti-miR31-D1	GGTCCCAGTTCAAC
Anti-miR31-A4	anti-miR31-D2	GTTCCCAGTTCAAC
Anti-miR31-A4-LNA	anti-miR31-D2-LNA	+G**T**T*C**C*C*A**G*T**T*C**A**A**C
Anti-miR144-A1	Anti-miR144-D3	AGTGTCTCATCGCAAACCTT
Anti-miR144-A2	Anti-miR144-loop-D	GTGTCTCATCGCAAACCT
Anti-miR144-A3	Anti-miR144-D1	GTGTCTCATCGCAAA
Anti-miR144-A4	Anti-miR144-D2	GTCTCATCGCAAAC
Anti-miR144-A4-LNA	anti-miR144-D2-LNA	+G**T**C*T**C*A*T**C*G**C*A**A**A**C
Anti-miR19a-A2	Anti-miR19a-loop-D	AACTACATTCTTCTTGTA
Anti-miR143-A4	Anti-miR143-D2	GACTCCCAACTGACC
Anti-let-7c-A1	pre-let-7c-AP-ASO	TCCCAGGGTGTAAC
ASO (-) control	anti-miR-31-locked	+C*C**G*T*T**C**T*A**C**G*A**C*C**G**T
Anti-miR-31 (+) control	Horizon Discovery	Catalog ID: IH-300507-06
Anti-miR-144 (+) control	Horizon Discovery	Catalog ID: IH-300612-06

<sup>a\*\*</sup> indicates phosphorothioate backbone, '+' indicates locked nucleic acid.

**Table 4.5. Templates for generating pmirGLO plasmid inserts.**

Construct	Primer name	5'-sequence-3'
pmirGLO miR31	pmirGLO-miR-31-F	AAACTAGCGGCCGCTAGTAGCTATGCCAGC ATCTTGCCTT
	pmirGLO-miR-31-R	CTAGAAGGCAAGATGCTGGCATAGCTACTA GCGGCCGCTAGTTT
pmirGLO miR144	pmirGLO-miR144-N-F	GCGTTTAAACTAGCGGCCGCTAGTAGTACA TCATCTATACTGTATCTAGAGC
	pmirGLO-miR144-N-R	GCTCTAGATACAGTATAGATGATGTACTAC TAGCGGCCGCTAGTTTAAACGC

**Table 4.6. Overlap extension PCR primers for generation of pCMV miR31 plasmid insert.**

Construct	Primer name	5'-sequence-3'
pCMV miR31	cmv-miR31-1F	GCAGATCTAGTCATAGTATTCTCCTGTA ACTTGGAAGTGGAGAGGAGGCAAGATGCT
	cmv-miR31-2R	GGCATAGCAGGTTCCAGTTCAACAGCTATG CCAGCATCTTGCCCTCCTCT
	cmv-miR31-3F	TGGGAACCTGCTATGCCAACATATTGCCATC TTTCTGTCTGACAGCAGCCATGGCCAC
	cmv-miR31-4R	GCCTCGAGGCATGCAGGTGGCCATGGCTGCT GTCAGACAGGAAA

#### 4.5 Reference

1. Bennett, C.F., *Therapeutic antisense oligonucleotides are coming of age*. Annual review of medicine, 2019. **70**: p. 307-321.
2. Baker, K.E. and R. Parker, *Nonsense-mediated mRNA decay: terminating erroneous gene expression*. Current opinion in cell biology, 2004. **16**(3): p. 293-299.
3. Marrosu, E., et al., *Gapmer antisense oligonucleotides suppress the mutant allele of COL6A3 and restore functional protein in ullrich muscular dystrophy*. Molecular Therapy-Nucleic Acids, 2017. **8**: p. 416-427.
4. Crooke, S.T., et al., *RNA-targeted therapeutics*. Cell metabolism, 2018. **27**(4): p. 714-739.
5. Cerritelli, S.M. and R.J. Crouch, *Ribonuclease H: the enzymes in eukaryotes*. The FEBS journal, 2009. **276**(6): p. 1494-1505.
6. Tadokoro, T. and S. Kanaya, *Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes*. The FEBS Journal, 2009. **276**(6): p. 1482-1493.
7. Nowotny, M., et al., *Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis*. Cell, 2005. **121**(7): p. 1005-1016.
8. Dar, S.A., et al., *siRNAmoD: A database of experimentally validated chemically modified siRNAs*. Scientific Reports, 2016. **6**(1): p. 20031.
9. Kawamata, T. and Y. Tomari, *Making risc*. Trends in biochemical sciences, 2010. **35**(7): p. 368-376.
10. Yoda, M., et al., *ATP-dependent human RISC assembly pathways*. Nature structural & molecular biology, 2010. **17**(1): p. 17-23.
11. Sharp, P.A., *The centrality of RNA*. Cell, 2009. **136**(4): p. 577-580.

12. Anna, A. and G. Monika, *Splicing mutations in human genetic disorders: examples, detection, and confirmation*. J Appl Genet, 2018. **59**(3): p. 253-268.
13. Hoffman, E.P., R.H. Brown, and L.M. Kunkel, *Dystrophin: The protein product of the duchenne muscular dystrophy locus*. Cell, 1987. **51**(6): p. 919-928.
14. Kinali, M., et al., *Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study*. The Lancet Neurology, 2009. **8**(10): p. 918-928.
15. Goodkey, K., et al., *Nusinersen in the treatment of spinal muscular atrophy*. Exon skipping and inclusion therapies: Methods and protocols, 2018: p. 69-76.
16. Migliorati, J.M., et al., *Absorption, Distribution, Metabolism, and Excretion of US Food and Drug Administration–Approved Antisense Oligonucleotide Drugs*. Drug Metabolism and Disposition, 2022. **50**(6): p. 888-897.
17. Braasch, D.A., Y. Liu, and D.R. Corey, *Antisense inhibition of gene expression in cells by oligonucleotides incorporating locked nucleic acids: effect of mRNA target sequence and chimera design*. Nucleic Acids Research, 2002. **30**(23): p. 5160-5167.
18. Liang, X.-h., et al., *Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames*. Nature Biotechnology, 2016. **34**(8): p. 875-880.
19. Liang, X.-h., et al., *Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels*. Nucleic Acids Research, 2017. **45**(16): p. 9528-9546.
20. Elmén, J., et al., *LNA-mediated microRNA silencing in non-human primates*. Nature, 2008. **452**(7189): p. 896-899.

21. Wilson, R.C., et al., *Dicer/TRBP complex formation ensures accurate mammalian microRNA biogenesis*. *Molecular cell*, 2015. **57**(3): p. 397-407.
22. Chakravarthy, S., et al., *Substrate-specific kinetics of Dicer-catalyzed RNA processing*. *Journal of molecular biology*, 2010. **404**(3): p. 392-402.
23. Fareh, M., et al., *TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments*. *Nature communications*, 2016. **7**(1): p. 13694.
24. Bartel, D.P., *Metazoan micrornas*. *Cell*, 2018. **173**(1): p. 20-51.
25. Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis*. *Nature reviews Molecular cell biology*, 2014. **15**(8): p. 509-524.
26. Paroo, Z., Q. Liu, and X. Wang, *Biochemical mechanisms of the RNA-induced silencing complex*. *Cell Research*, 2007. **17**(3): p. 187-194.
27. Piao, X., et al., *CCR4-NOT deadenylates mRNA associated with RNA-induced silencing complexes in human cells*. *Molecular and cellular biology*, 2010. **30**(6): p. 1486-1494.
28. Lee, Y.S., et al., *Depletion of human micro-RNA miR-125b reveals that it is critical for the proliferation of differentiated cells but not for the down-regulation of putative targets during differentiation*. *Journal of Biological Chemistry*, 2005. **280**(17): p. 16635-16641.
29. Bhattarai, U., et al., *Bifunctional small molecule-oligonucleotide hybrid as microRNA inhibitor*. *Bioorganic & medicinal chemistry*, 2020. **28**(7): p. 115394.
30. Ma, S., et al., *Structure of pre-miR-31 reveals an active role in Dicer–TRBP complex processing*. *Proceedings of the National Academy of Sciences*, 2023. **120**(39): p. e2300527120.
31. MacRae, I.J., et al., *In vitro reconstitution of the human RISC-loading complex*. *Proceedings of the National Academy of Sciences*, 2008. **105**(2): p. 512-517.
32. Lorenz, D.A. and A.L. Garner, *A click chemistry-based microRNA maturation assay*



*optimized for high-throughput screening*. Chemical communications, 2016. **52**(53): p. 8267-8270.

33. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis*. Nature Methods, 2012. **9**(7): p. 671-675.

34. Takahashi, T., et al., *Distinguishable in vitro binding mode of monomeric TRBP and dimeric PACT with siRNA*. PloS one, 2013. **8**(5): p. e63434.

35. Shortridge, M.D., et al., *Drug-like small molecules that inhibit expression of the oncogenic microRNA-21*. ACS Chemical Biology, 2023. **18**(2): p. 237-250.

36. Cunningham, P.R. and J. Ofengand, *Use of inorganic pyrophosphatase to improve the yield of in vitro transcription reactions catalyzed by T7 RNA polymerase*. Biotechniques, 1990. **9**(6): p. 713-714.

37. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature*. Nucleic acids research, 2006. **34**(suppl\_1): p. D140-D144.

38. Shortridge, M.D., et al., *A slow dynamic RNA switch regulates processing of microRNA-21*. Journal of Molecular Biology, 2022. **434**(16): p. 167694.

39. Crooke, S.T., et al., *Antisense technology: A review*. Journal of Biological Chemistry, 2021. **296**.

40. Raguraman, P., et al., *Antisense oligonucleotide-mediated splice switching: Potential therapeutic approach for cancer mitigation*. Cancers, 2021. **13**(21): p. 5555.

41. Shang, R., et al., *Genomic clustering facilitates nuclear processing of suboptimal pri-miRNA loci*. Molecular cell, 2020. **78**(2): p. 303-316. e4.

42. Shang, R., et al., *Regulated dicing of pre-mir-144 via reshaping of its terminal loop*. Nucleic acids research, 2022. **50**(13): p. 7637-7654.

43. Kozomara, A., M. Birgaoanu, and S. Griffiths-Jones, *miRBase: from microRNA sequences*

*to function*. Nucleic acids research, 2019. **47**(D1): p. D155-D162.

44. Kozomara, A. and S. Griffiths-Jones, *miRBase: annotating high confidence microRNAs using deep sequencing data*. Nucleic acids research, 2014. **42**(D1): p. D68-D73.

45. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data*. Nucleic acids research, 2010. **39**(suppl\_1): p. D152-D157.

46. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics*. Nucleic acids research, 2007. **36**(suppl\_1): p. D154-D158.

47. Lee, Y.-Y., et al., *Structure of the human DICER-pre-miRNA complex in a dicing state*. Nature, 2023. **615**(7951): p. 331-338.

48. Liu, Z., et al., *Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate*. Cell, 2018. **173**(5): p. 1191-1203. e12.

49. Leitão, A.L. and F.J. Enguita, *A Structural View of miRNA Biogenesis and Function*. Noncoding RNA, 2022. **8**(1).

50. Wurster, C.D. and A.C. Ludolph, *Antisense oligonucleotides in neurological disorders*. Therapeutic advances in neurological disorders, 2018. **11**: p. 1756286418776932.

51. Reda El Sayed, S., et al., *MicroRNA therapeutics in cancer: current advances and challenges*. Cancers, 2021. **13**(11): p. 2680.

52. Seo, H.A., et al., *Microrna-based combinatorial cancer therapy: Effects of micrornas on the efficacy of anti-cancer therapies*. Cells, 2019. **9**(1): p. 29.

53. Hennig, T., et al., *Selective inhibition of miRNA processing by a herpesvirus-encoded miRNA*. Nature, 2022. **605**(7910): p. 539-544.

## CHAPTER V Conclusion and Future Directions

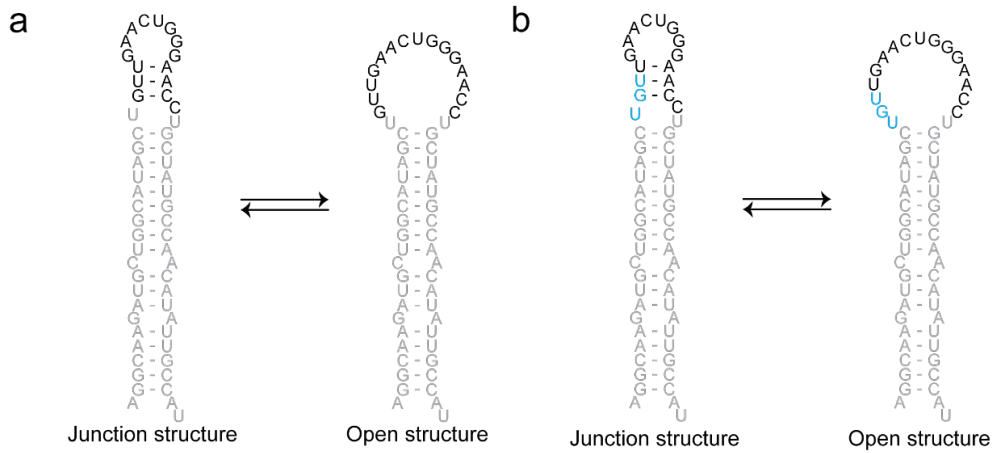
### 5.1 Conclusion

MicroRNAs (miRNA) post-transcriptionally regulate approximately one third of all protein-coding genes in the human genome [1]. Production of the proper levels of specific microRNA products is essential for proper cellular growth and development. The Dicer/TRBP complex processes precursor miRNAs (pre-miRNAs) in the cytoplasm to produce mature microRNAs [2, 3]. Many factors, including RNA sequence, structure, and interactions with protein binding partners, can influence Dicing and thereby the levels of mature miRNAs that are produced. It is therefore important to understand the various effectors of Dicing at the molecular level [4-9]. Pre-miRNA structure and dynamics can be essential for regulation of miRNA biogenesis [7-11]. However, prior to my thesis work, the studies defining how pre-miRNA tertiary structure or dynamics affected Dicer/TRBP processing were limited to a single pre-miRNA (pre-miR-21) [8, 9]. To expand on these structure-function studies, I examined the structural features that are important for pre-miR-31 processing. In chapter II of this thesis, I examined the secondary structure of pre-miR-31, identifying several mismatches in the helical stem [12]. The pre-miR-31 stem region has a dynamic C•A mismatch, which has a strong pH sensitivity. The adenosine in this mismatch has a significantly shifted pKa such that this C•A mismatch can form a stable base pair at near physiological pH, enhancing the stability of the pre-miR-31 stem region. Other mismatches in the stem region do not have similar pH sensitivity [12]. Furthermore, in Chapter

III, I identified how different structural features within pre-miR-31 influence the Dicer/TRBP processing of pre-miR-31. While the mismatches within the stem had no impact on Dicer/TRBP processing, I found that restricting the size of the internal loop near the dicing site promoted efficient dicing by Dicer/TRBP. Furthermore, in this Chapter, I also demonstrated the importance of TRBP in helping Dicer recognize structural features that will hinder Dicer cleavage (i.e. too large of an apical loop). However, the most important discovery in this Chapter was the identification that the junction region within pre-miR-31 serves as an inherent regulatory element for Dicer/TRBP processing [11]. With the aim of exploiting the regulation encoded within the junction, I designed a series of ASOs that specifically inhibit pre-miR-31 cleavage by Dicer/TRBP *in vitro* and *in cells* in Chapter IV. In this chapter, I also showed that the junction region is a common regulatory element inside pre-miRNAs and targeting these junction regions with ASOs can inhibit processing of the target pre-miRNA both *in vitro* and *in cells*. These results provide the framework for a new way for inhibiting miRNA production complementary to traditional anti-miRNA ASO designs [13]. The following Chapter proposes future directions in the research of how long miRNA transcripts regulates its enzymatic steps and novel antisense oligoes development.

## 5.2 Future directions

### 5.2.1 Junction region dynamics of pri-miR-31 and its role in Drosha/DGCR8 processing



**Figure 5.1. Two conformations of pre-miR-31. a)** Pre-miR-31 can adopt two different conformations, either containing (left) or lacking (right) a base paired junction region. **b)** The UGU motif (blue) sequence is sequestered in the junction region of pre-miR-31.

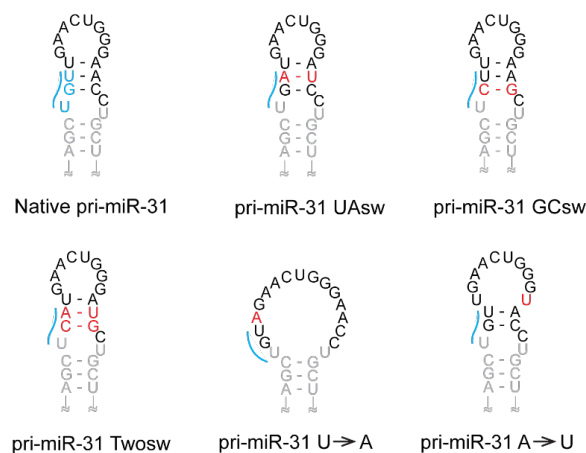
As discussed in Chapter III, the junction stability serves as an important regulatory element for Dicer/TRBP processing of pre-miR-31. The DMS-MaPseq chemical probing and solution nuclear magnetic resonance (NMR) studies revealed two different secondary structures near the apical loop of pre-miR-31. One structure contained a large apical loop while the other contained additional base pairs between the dicing site and the apical loop, defined as the junction region.

These two different conformations were determined using different structure biology methods, however, the dynamic information about the interconversion between states remains elusive. The conformational dynamics of RNAs have been shown to be important as “excited state” structures are often associated with directing biological function [8, 9]. Therefore, using NMR methods to study the inherent dynamics of the pre-miR-31 junction region will allow us to quantify the time scales under which these two conformations are sampled and the relative populations of the two conformations (**Fig. 5.1a**). The dynamics of RNAs can be divided into slow, intermediate,

and fast time scales and different NMR methods can be used to study conformation exchange in these different time scales. In our previous study, we did not observe additional signals in the NOESY spectrum, which is characteristic of slow exchange dynamics. This result suggests that pre-miR-31 conformational exchange does not occur in the slow time scale, although it is still possible that the conformational exchange is slow but the population in the excited state is sufficiently low that it can't be detected in the NOESY. Chemical exchange saturation transfer (CEST) is a useful method for detecting excited states of RNAs, especially when the excited states population is low and the chemical shift is absent in the NMR spectra [14]. This method has been used to detect an “invisible” excited state in the pre-miR-21 [8]. The CEST experiment can be used to detect the exchange which occurs in the milliseconds to seconds timescale. Carr–Purcell–Meiboom–Gill spin echo (CPMG) and  $R_{1\rho}$  are two NMR experiments that can be used to study conformational exchange on the microsecond to millisecond timescale. CPMG has been used to study the structure change of a pseudoknot element caused by miRNA binding [15] and  $R_{1\rho}$  experiments have been used in studying the conformational dynamics that occur in different HIV RNA elements [16, 17]. Solid state NMR can also be used to study the nanoseconds to microseconds level dynamics in RNAs, as has been demonstrated for the HIV TAR element [18-21]. These different NMR methods can be applied to pre-miR-31 to determine the exchange rate between the two discrete conformations.

Interestingly, pre-miR-31 and pri-miR-31 share the same upper stem region and apical loop - in other words, the junction region may also be present in pri-miR-31. As detailed in Chapter I, the UGU motif in the apical loop region of pri-miRNAs can be recognized by the DGCR8 protein and enhance Drosha/DGCR8 cleavage efficiency and accuracy [22, 23]. The UGU motif, when present, is typically located in a single strand terminal loop region (**Fig 1.1**) [22, 24]. Both pri- and

pre-miR-31 contain UGU motifs, however, when the junction structure is formed, the UGU motif is sequestered within the junction whereas the UGU motif would be appropriately positioned in the apical loop in the open structure (**Fig. 5.1b**). In Chapter III, I showed that the open structure was a poor substrate for Dicer/TRBP cleavage while the junction structure was a good Dicer/TRBP substrate [11]. I hypothesize that junction region of pre-miR-31 exists in a dynamic equilibrium. The dynamics of the junction region may allow pre-miR-31 to adopt a cinched junction structure, which is ideal for Dicer/TRBP processing, and the open structure, where the UGU motif is exposed, may be favored by Drosha/DGCR8 for optimal cleavage efficiency and accuracy. To test this hypothesis, several mutations can be introduced for *in vitro* processing studies. The UGU motif can be disrupted by mutating the junction residues such that the secondary structure of the pri-miR-31 junction region was retained (**Fig. 5.2**). The following are potential constructs that could be used; UAsw, GCsw and Twosw, all constructs mutate the UGU motif in the junction region of pri-miR-31. If the UGU motif inside pri-miR-31 is recognized by the Drosha/DGCR8 protein, these mutations should inhibit or affect the accuracy of processing by Drosha/DGCR8. Another set of mutations can be designed to promote the opening of the junction region, which is expected to enhance the recognition of pri-miR-31 by Drosha/DGCR8. The U→A and A→U constructs mutate the top UA base pairs in the junction region (**Fig 5.2**) and promote the release of the UGU motif from a base paired conformation. These two mutations would be expected to enhance the processing efficiency of pri-miR-31 by Drosha/DGCR8.



**Figure 5.2. Proposed mutants of pri-miR-31 junction region.** The UGU motif in the native pri-miR-31 structure is labeled in blue text and the position of the UGU motif in the mutant constructs are indicated using blue lines. The mutated nucleotides are labeled in red. Pri-miR-31 UAsw, GCsw, and Twosw are proposed as poor Drosha/DGCR8 substrates while pri-miR-31 U→A and pri-miR-31 A→U are proposed as good Drosha/DGCR8 substrates.

### 5.2.2 pre-miR-31 structural role in regulating Dicer/TRBP binding affinity

In Chapter III, we found that the pre-miR-31 29CAA RNA had a 6-fold tighter binding affinity for Dicer (in the absence of TRBP), relative to the pre-miR-31 GCclamp mutant (**Appendix B**). Our results suggested that 29CAA adopted a conformation which promoted Dicer binding, yet this RNA was a poor substrate for Dicer cleavage (10-fold reduction in  $k_{obs}$  relative to WT). In our binding and processing experiments using Dicer enzyme alone (without TRBP), we found that WT pre-miR-31 had the tightest binding affinity and near maximal processing kinetics among all the junction region mutants tested. However, how the stability of the junction region affects the Dicer/TRBP binding has not been investigated. Therefore, supplemental experiments for Dicer/TRBP binding affinity with junction region mutants can be carried out to investigate how the stability of junction region affects Dicer/TRBP binding on pre-miR-31. This will help us understand what role of open conformation of pre-miR-31 plays in regulating Dicer/TRBP binding to pre-miR-31.



### 5.2.3 Novel design of shRNA combined with junction region targeted ASOs

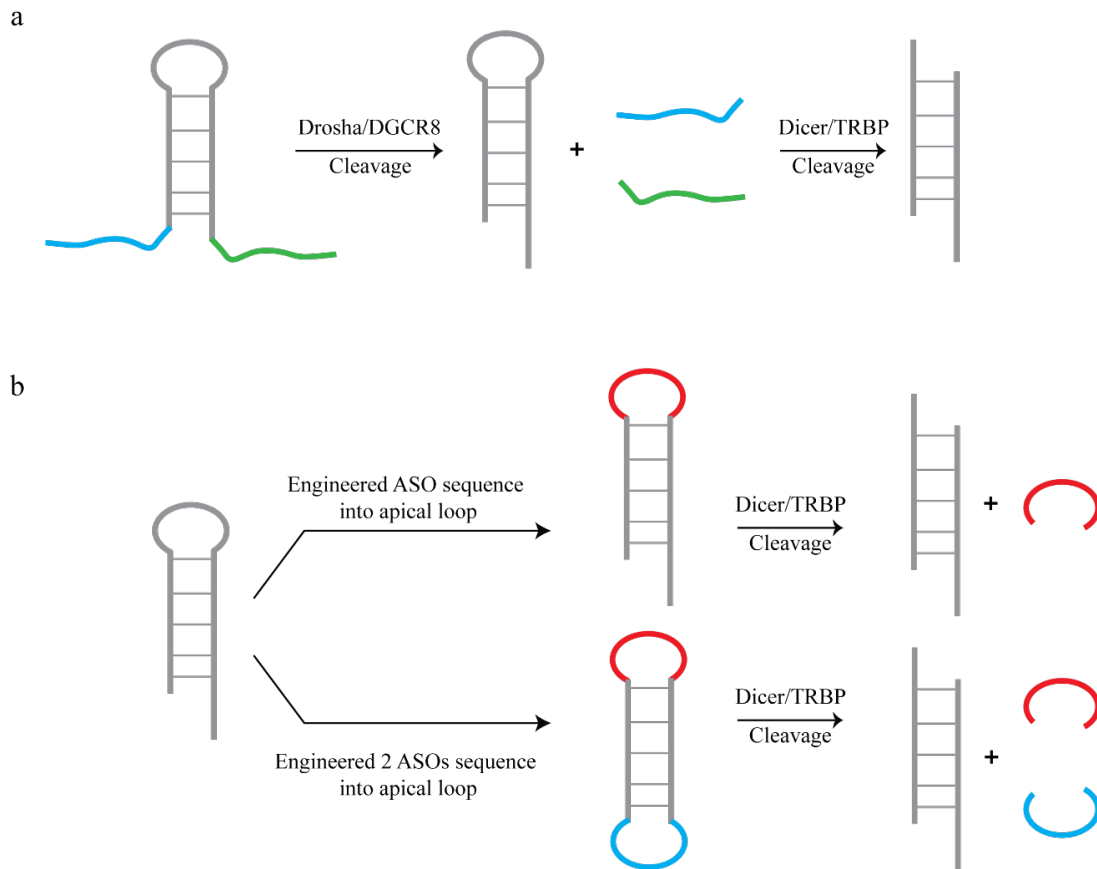
Short hairpin RNAs (shRNAs) are a useful tool for targeting any desired mRNA for repression as a therapeutic method [25-27]. The structure of a shRNA is very similar to that of a pre-miRNA, which contains a base paired stem capped with an apical loop. Many studies have sought to determine the importance of shRNA structure for the shRNA to effectively knock-down a target gene. In general, shRNAs consist of a 19-29 dsRNA stem region connected by an apical loop, which is typically 4-15 nucleotides in length [28]. The stem region of an shRNA is an important region because it contains the seed sequence which is required for sequence-specific targeting on a mRNA. Although the length of the apical loop in a shRNA has been shown to be important for knock-down efficiency [28-30], the possibility of using the apical loop cleavage product of shRNA as a therapeutic ASO has not yet been explored. Recent studies have shown that transfection of artificial pri-miRNA into cells can deliver both siRNAs and ASOs at the same time, allowing researchers to achieve multimodal gene regulation [31]. The approach was simple, requiring the ligation of one ASO to a shRNA as a flanking motif at the 3' end to generate an artificial primary miRNA (**Fig. 5.3a**). This new pri-miRNA can be processed via the canonical miRNA biogenesis pathway to generate new pre-miRNA, where the 3'-flanking sequence can be processed off by Drosha/DGCR8 (**Fig. 5.3a**). The newly generated pre-miRNA encompasses shRNA as the main body and produces siRNA for gene silencing (**Fig. 5.3a**). The cleaved off flanking sequence can also function to sequester a mature miRNA or function as splicing modulator to alter pre-mRNA splicing, as has been demonstrated for miR-21 and survival motor neuron 2 (SMN2), respectively (**Fig. 5.3a**) [31].

These findings support the idea that multiple ASOs can be combined to generate a multifunctional therapeutic. Therefore, it may be possible for ASOs that target a pre-miRNA

junction region to function together with a shRNA (**Fig. 5.3b, top**). In cancer cells, some miRNAs are oncogenic and overexpressed, and some of these pre-miRNAs contain a junction region (i.e. pre-miR-31) [32, 33]. Combining the shRNA with anti-junction-targeting ASOs may enhance the inhibition of cancer cell growth. Circular shRNAs show higher knock down efficiency [34], therefore, forming a circularized shRNA that contains two “apical loops” encoding anti-junction-region ASOs (**Fig. 5.3b, bottom**) may enhance overall efficiency for pre-miRNA knockdown due to the effectively doubled concentration of anti-junction-region ASOs relative to non-circularized anti-junction-region ASO-contained shRNA. Furthermore, the circularized shRNAs can contain two apical loops that encode different anti-junction-region ASO sequences, allowing for the targeting of two different junction-containing pre-miRNAs at the same time. (**Fig. 5.3b, bottom**).

To test the effectiveness of this design, I propose to engineer a shRNA/circularized shRNA pair which uses the pre-miR-21 stem region as the shRNA stem template and uses the anti-miR31-A4 sequences (detailed in chapter IV) as the loop sequence(s). Using a dual luciferase miRNA assay, which involves transfection of a pre-miR-31 overexpression plasmid, a miR-21 luciferase plasmid, a miR-31 luciferase plasmid, and the engineered shRNA or circularized shRNA into HEK 293 cells, the inhibitory effectiveness of the anti-miR31-A4 ASO can be assessed. If the transfected shRNAs are properly cleaved by Dicer/TRBP *in cell*, the generated miR-21 mature sequence will bind to the 3' UTR of the miR-21 luciferase reporter mRNA and decrease the expression of the upstream luciferase protein. The decrease in the observed signal from the miR-21 luciferase control can be used as a measure of the successful transfection of the shRNA/circularized shRNA. The change in the signal from the miR-31 luciferase reporter protein can be used to evaluate the extent of inhibition on the pre-miR-31 by the cleaved-off loops *in cells*. If this principle is validated in cell experiments, the shRNA can serve a multifunctional antisense oligo in the future nucleic

acid therapeutic areas.



**Figure 5.3 New principle for designing shRNA. a)** Approach for artificial pri-miRNA design. An antisense oligonucleotide (ASO, green) is ligated to the 3' end of a shRNA, generating a flanking sequence. The 5' end flanking motif (blue) does not function as a ASO. The artificial pri-miRNA can be processed by the Drosha/DGCR8 complex to generate an shRNA with siRNA sequence embedded in the stem and a functional ASO (green). The shRNA can be further cleaved by the Dicer/TRBP complex to produce the functional siRNA (gray). **b)** Conventional design of a shRNA (left). The apical loop sequence can be engineered to include an anti pre-miRNA junction region ASO sequence (red, top). The Dicer/TRBP complex can cleave this engineered shRNA to generate a functional siRNA (gray) and a function anti pre-miRNA junction region ASO (red). The apical loop of a conventional shRNA can be replaced by an anti pre-miRNA junction region ASO sequence (red, bottom) while the open 5' end and 3' end can be circularized by teathering another anti pre-miRNA junction region ASO sequence (blue). The Dicer/TRBP complex can cleave these circular shRNA to generate a functional siRNA (gray) and two functional anti pre-miRNA junction region ASOs (blue and red).

### 5.3 Reference

1. Jackson, R.J. and N. Standart, *How do microRNAs regulate gene expression?* Sci STKE, 2007. **2007**(367): p. re1.
2. Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis*. Nature reviews Molecular cell biology, 2014. **15**(8): p. 509-524.
3. Shang, R., et al., *microRNAs in action: biogenesis, function and regulation*. Nature Reviews Genetics, 2023. **24**(12): p. 816-833.
4. Balzeau, J., et al., *The LIN28/let-7 pathway in cancer*. Frontiers in genetics, 2017. **8**: p. 31.
5. Abbott, A.L., et al., *The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans*. Developmental cell, 2005. **9**(3): p. 403-414.
6. Ustianenko, D., et al., *LIN28 selectively modulates a subclass of let-7 microRNAs*. Molecular cell, 2018. **71**(2): p. 271-283. e5.
7. Lee, Y.-Y., H. Kim, and V.N. Kim, *Sequence determinant of small RNA production by DICER*. Nature, 2023. **615**(7951): p. 323-330.
8. Baisden, J.T., et al., *Visualizing a protonated RNA state that modulates microRNA-21 maturation*. Nature chemical biology, 2021. **17**(1): p. 80-88.
9. Shortridge, M.D., et al., *A slow dynamic RNA switch regulates processing of microRNA-21*. Journal of Molecular Biology, 2022. **434**(16): p. 167694.
10. Lee, Y.-Y., et al., *Structure of the human DICER-pre-miRNA complex in a dicing state*. Nature, 2023. **615**(7951): p. 331-338.
11. Ma, S., et al., *Structure of pre-miR-31 reveals an active role in Dicer-TRBP complex processing*. Proceedings of the National Academy of Sciences, 2023. **120**(39): p. e2300527120.

12. Kotar, A., S. Ma, and S.C. Keane, *pH dependence of C•A, G•A and A•A mismatches in the stem of precursor microRNA-31*. *Biophysical Chemistry*, 2022: p. 106763.
13. Crooke, S.T., et al., *Antisense technology: A review*. *Journal of Biological Chemistry*, 2021. **296**.
14. Marušič, M., J. Schlagnitweit, and K. Petzold, *RNA dynamics by NMR spectroscopy*. *ChemBioChem*, 2019. **20**(21): p. 2685-2710.
15. Chen, B., et al., *CCR5 RNA Pseudoknots: Residue and Site-Specific Labeling correlate Internal Motions with microRNA Binding*. *Chemistry—A European Journal*, 2018. **24**(21): p. 5462-5468.
16. Dethoff, E.A., et al., *Visualizing transient low-populated structures of RNA*. *Nature*, 2012. **491**(7426): p. 724-728.
17. Dethoff, E.A., et al., *Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis*. *Biophysical journal*, 2008. **95**(8): p. 3906-3915.
18. Huang, W., et al., *Ultraslow domain motions in HIV-1 TAR RNA revealed by solid-state deuterium NMR*. *The Journal of Physical Chemistry B*, 2017. **121**(1): p. 110-117.
19. Olsen, G.L., et al., *Hydration dependent dynamics in RNA*. *Journal of biomolecular NMR*, 2009. **45**: p. 133-142.
20. Olsen, G.L., et al., *Intermediate rate atomic trajectories of RNA by solid-state NMR spectroscopy*. *Journal of the American Chemical Society*, 2010. **132**(1): p. 303-308.
21. Olsen, G.L., et al., *Solid-state deuterium NMR studies reveal  $\mu$ s-ns motions in the HIV-1 transactivation response RNA recognition site*. *Journal of the American Chemical Society*, 2008. **130**(10): p. 2896-2897.

22. Dang, T.L., et al. *Select amino acids in DGCR8 are essential for the UGU-pri-miRNA interaction and processing*. *Communications biology*, 2020. **3**, 344 DOI: 10.1038/s42003-020-1071-5.
23. Bofill-De Ros, X. and U.A. Vang Ørom, *Recent progress in miRNA biogenesis and decay*. *RNA biology*, 2024. **21**(1): p. 1-8.
24. Partin, A.C., et al., *Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs*. *Nature communications*, 2017. **8**(1): p. 1737.
25. Bennett, C.F., *Therapeutic antisense oligonucleotides are coming of age*. *Annual review of medicine*, 2019. **70**: p. 307-321.
26. Crooke, S.T., et al., *Antisense drug discovery and development technology considered in a pharmacological context*. *Biochemical pharmacology*, 2021. **189**: p. 114196.
27. Migliorati, J.M., et al., *Absorption, Distribution, Metabolism, and Excretion of US Food and Drug Administration–Approved Antisense Oligonucleotide Drugs*. *Drug Metabolism and Disposition*, 2022. **50**(6): p. 888-897.
28. Jensen, S.M.R., et al., *Functional selection of shRNA loops from randomized retroviral libraries*. 2012.
29. Brummelkamp, T.R., R. Bernards, and R. Agami, *A system for stable expression of short interfering RNAs in mammalian cells*. *Science*, 2002. **296**(5567): p. 550-3.
30. Paddison, P.J., A.A. Caudy, and G.J. Hannon, *Stable suppression of gene expression by RNAi in mammalian cells*. *Proceedings of the National Academy of Sciences*, 2002. **99**(3): p. 1443-1448.
31. Kim, H., et al., *Artificial primary-miRNAs as a platform for simultaneous delivery of siRNA and antisense oligonucleotide for multimodal gene regulation*. *Journal of Controlled Release*,

2022. **349**: p. 983-991.

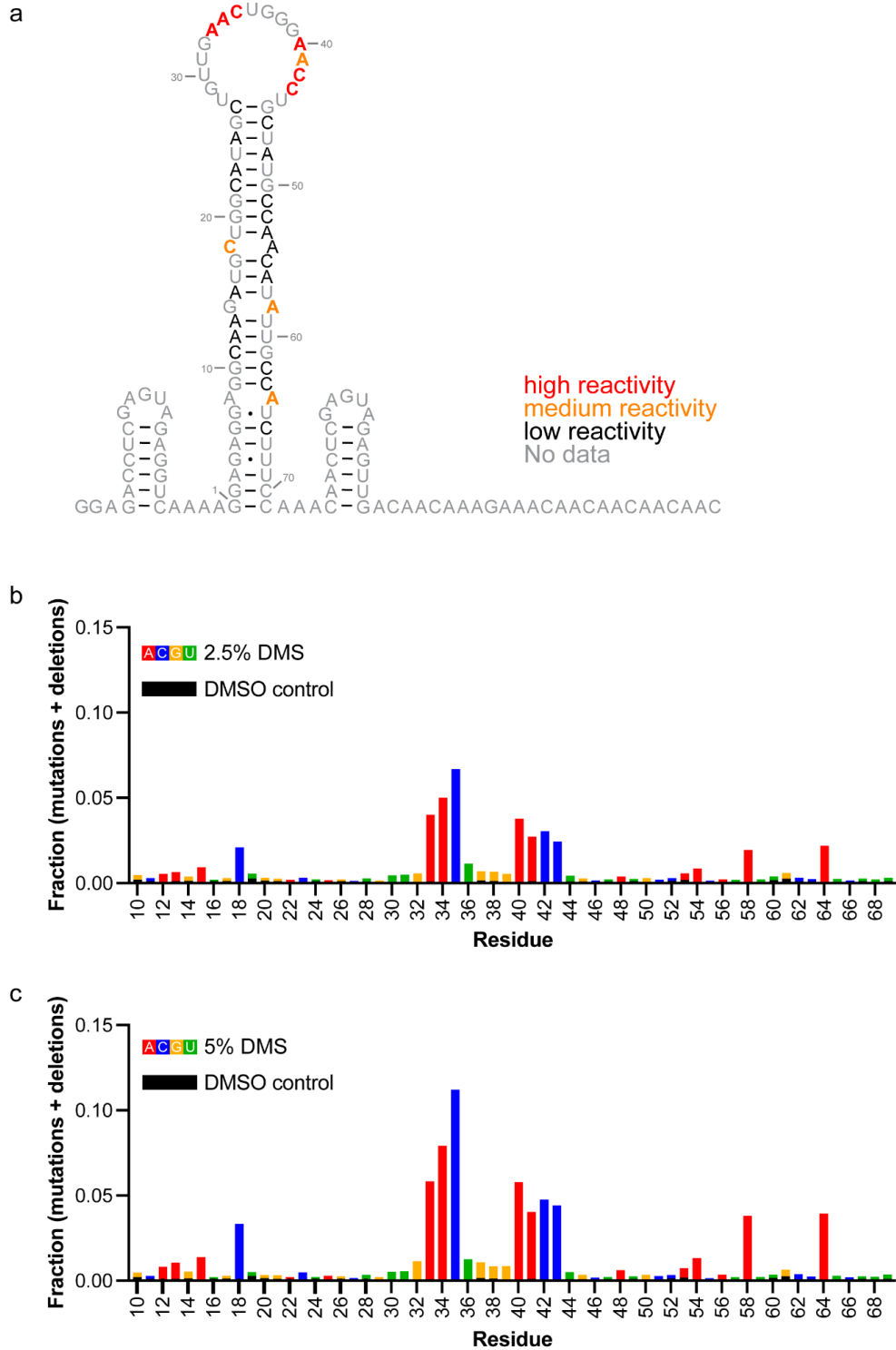
32. Acunzo, M., et al., *MicroRNA and cancer—a brief overview*. *Advances in biological regulation*, 2015. **57**: p. 1-9.

33. Reda El Sayed, S., et al., *MicroRNA therapeutics in cancer: current advances and challenges*. *Cancers*, 2021. **13**(11): p. 2680.

34. Abe, H., et al. *Synthetic nanocircular RNA for controlling of gene expression*. in *Nucleic Acids Symposium Series*. 2009. Oxford University Press.

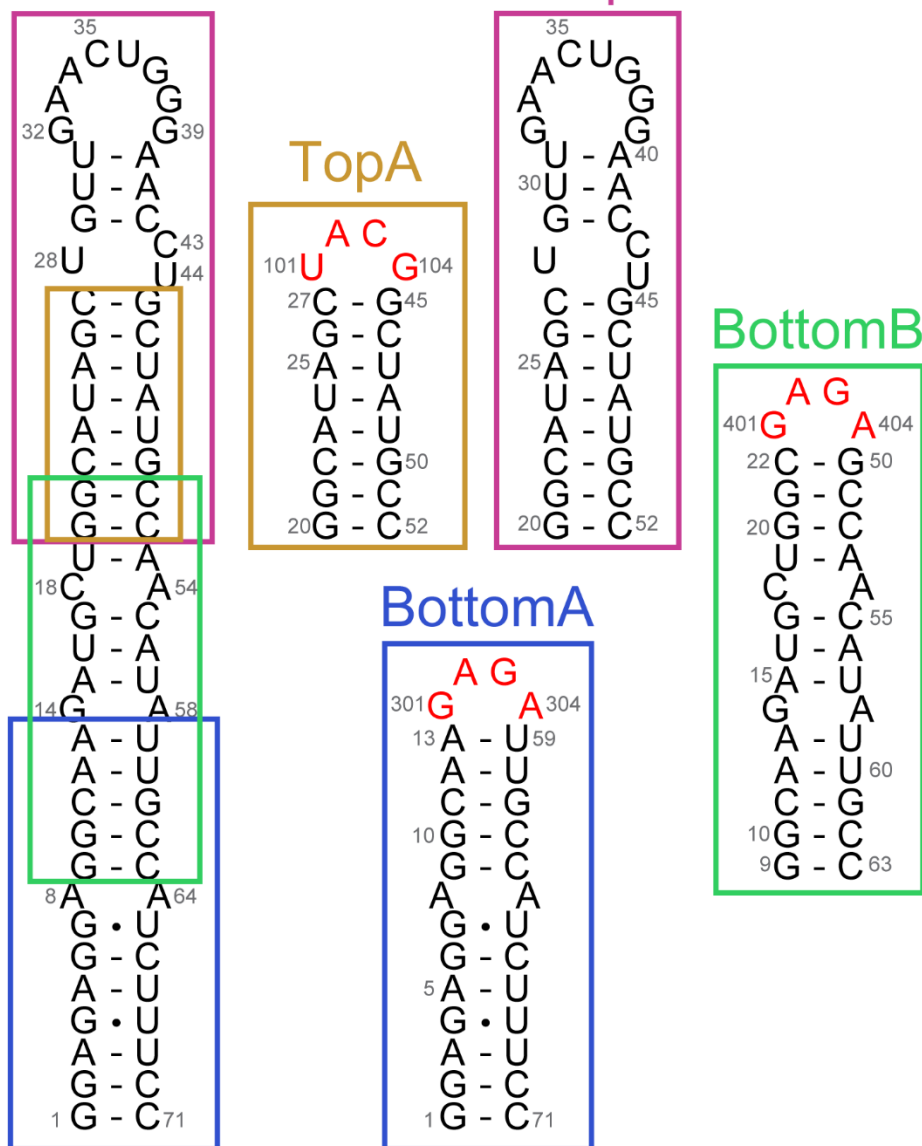
## **Appendix A Supplemental Figure and Tables for Chapter III**



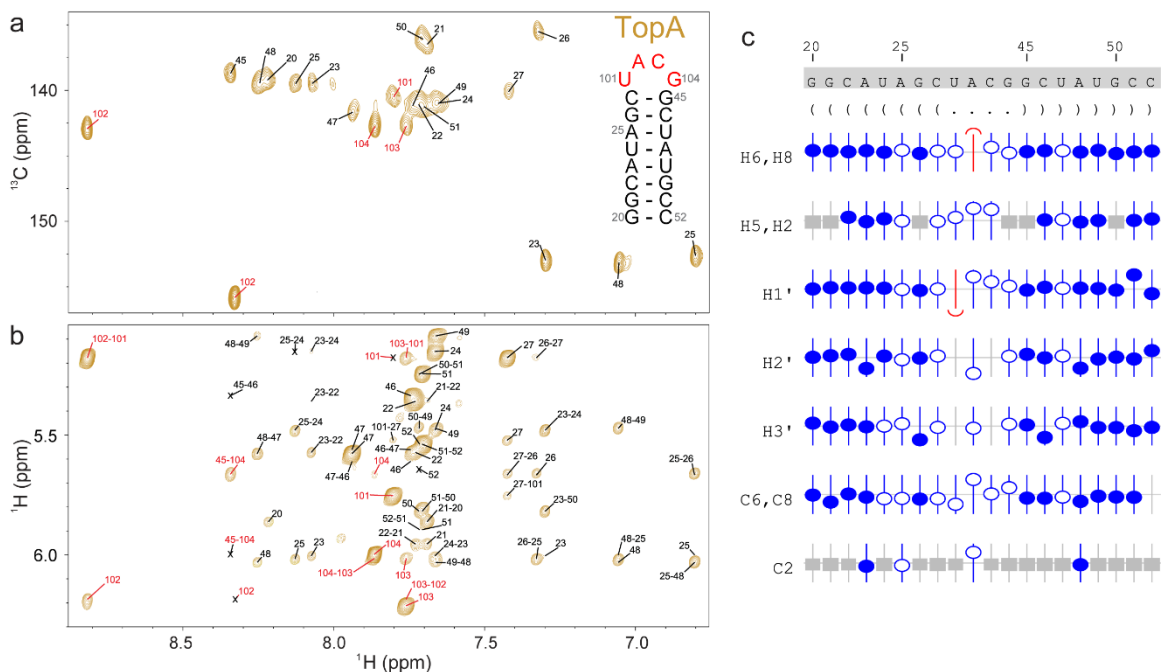


**Appendix A Fig. S1. In vitro DMS-MaPseq of pre-miRNA-31 RNA.** a) Construct design and reactivity scores. 5' and 3' extensions were included to facilitate library preparation and sequencing. These regions were not predicted to disrupt the folding of pre-miR-31. b and c) Fraction of mutations and deletions on a per-residue basis upon reaction with 2.5% DMS (b) or 5% DMS (c). A control where DMSO was included in the reaction rather than DMS indicates minimal background mutations (black bars in b and c). Secondary structure was rendered using

## pre-miR-31



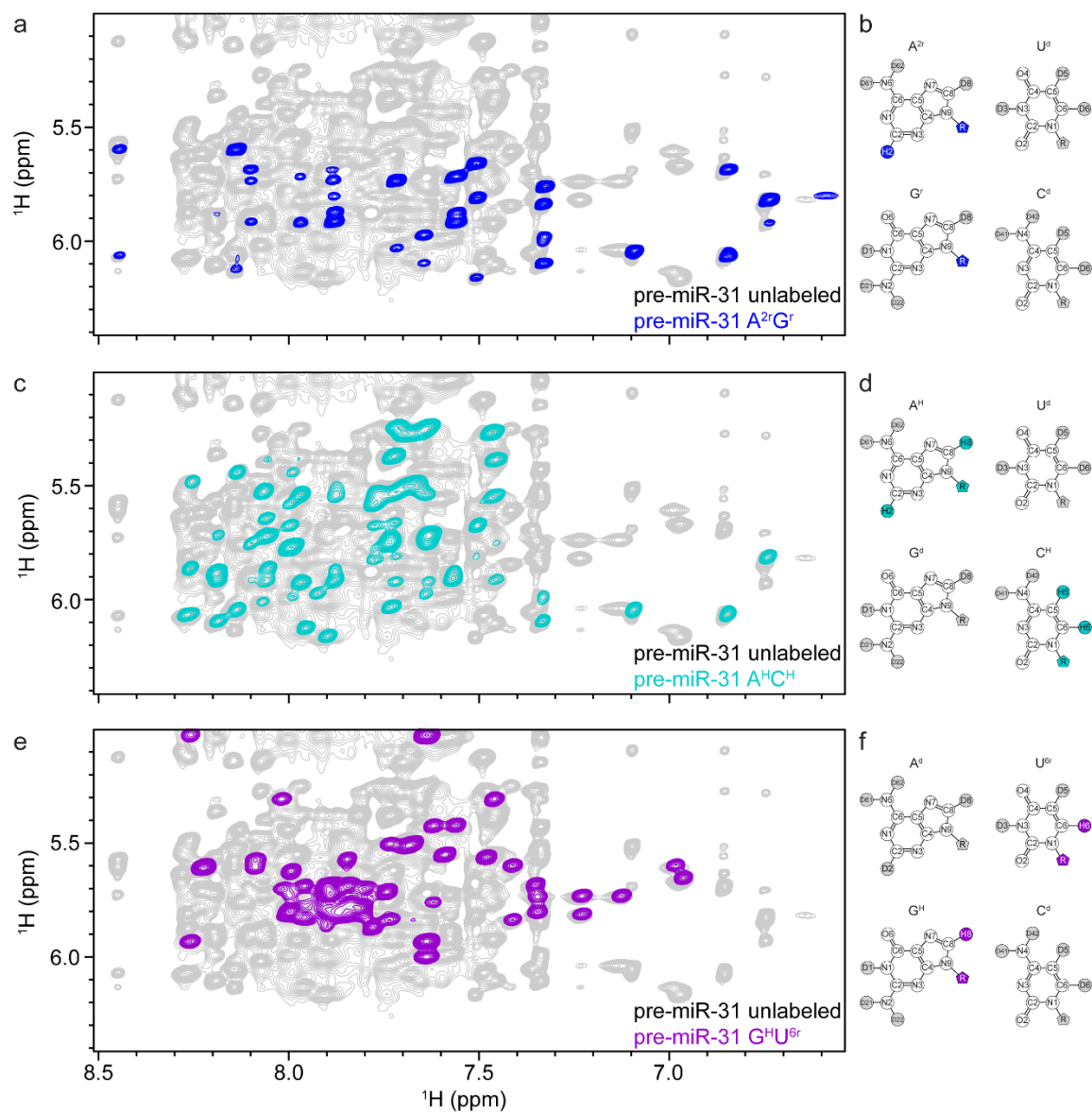
**Appendix A Fig. S2. Oligo controls for NMR chemical shift assignment of FL-pre-miR-31.** Four oligonucleotide controls were designed to cover the entire FL-pre-miR-31 sequence. Non-native tetraloops (red) were included to cap oligos that truncated the apical loop. Secondary structures were rendered using RNA2Drawer (1).



**Appendix A Fig. S3. Assigned chemical shifts of TopA RNA.** **a)**  $^1\text{H}$ - $^{13}\text{C}$  HMQC and **b)**  $^1\text{H}$ - $^1\text{H}$  NOESY of TopA RNA. The signals assigned to GAGA tetraloop are colored red. NMR spectra were recorded at 0.4 mM RNA concentration, 50 mM K-phosphate buffer, pH=7.5, 1 mM  $\text{MgCl}_2$  and 100%  $\text{D}_2\text{O}$ . **c)** Sequence analysis and validation of TopA chemical shift assignments. Nucleotide numbering, sequence, and secondary structure in Vienna format for TopA RNA. NMRViewJ chemical shift prediction software was used to validate proton (H6/H8, H5/H2, H1', H2', H3') and carbon (C6/C8, C2) assignments. Assigned atoms are represented with blue circles (open and closed), while grey boxes denote atoms that are not present in a given base. Deviation from the predicted chemical shift is shown with the offset from the center. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB. Secondary structures were rendered using RNA2Drawer(1).



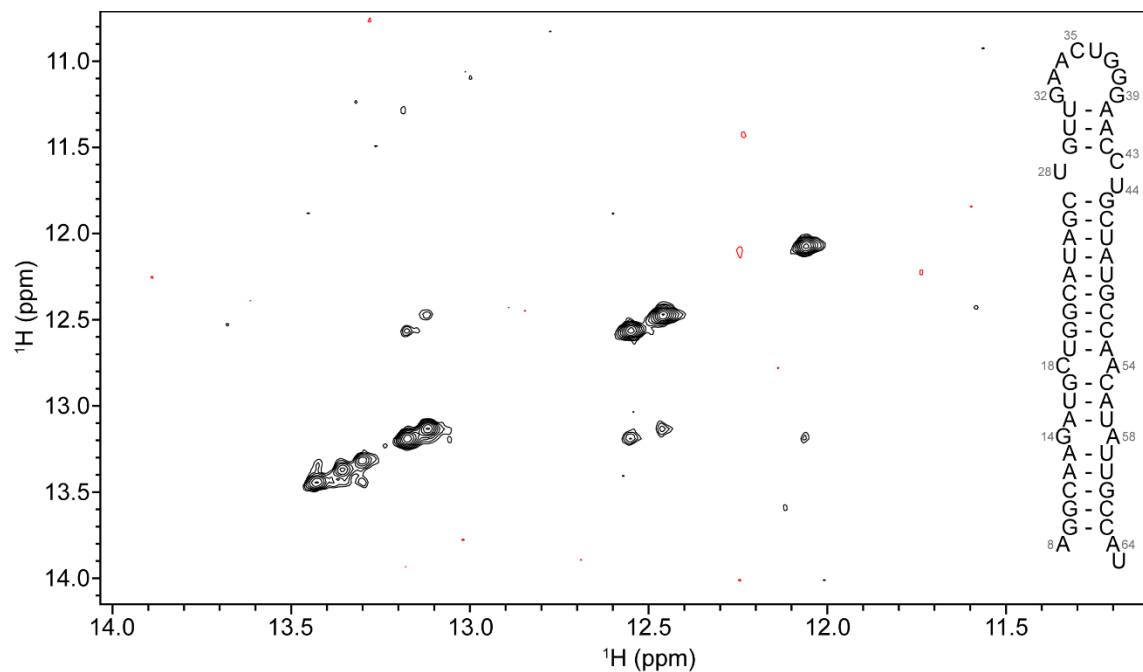
**Appendix A Fig. S4. Assigned chemical shifts of Top RNA.** **a)**  $^1\text{H}$ - $^{13}\text{C}$  HMQC and **b)**  $^1\text{H}$ - $^1\text{H}$  NOESY of Top RNA. **c)**  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum overlay of fully-protonated (pink) and  $\text{A}^{13}\text{C}^{\text{H}}$ -labeled (teal) Top RNA. Secondary structure is colored to indicate the proton position (teal) in the  $\text{A}^{13}\text{C}^{\text{H}}$ -labeled Top RNA sample. All other sites are perdeuterated (black). NMR spectra were recorded at 0.4 mM RNA concentration, 50 mM K-phosphate buffer, pH=7.5, 1 mM  $\text{MgCl}_2$  and 100%  $\text{D}_2\text{O}$ . **d)** Sequence analysis and validation of Top chemical shift assignments. Nucleotide numbering, sequence, and secondary structure in Vienna format for Top RNA. NMRViewJ chemical shift prediction software was used to validate proton (H6/H8, H5/H2, H1', H2', H3') and carbon (C6/C8, C2) assignments. Assigned atoms are represented with blue circles (open and closed), while grey boxes denote atoms that are not present in a given base. Deviation from the predicted chemical shift is shown with the offset from the center. Filled circles indicate that there are chemical shifts for atoms with the same set of attributes in the BMRB. Open circles indicate atoms that have a prediction, but for which no exact matches of the attributes are available in the BMRB.



**Appendix A Fig. S5. Deuterium labeling improves spectral quality by reducing overlap.** **a)** Overlay of the unlabeled (fully protiated, gray) and A<sup>2r</sup>G<sup>r</sup>-labeled (blue) pre-miR-31 <sup>1</sup>H-<sup>1</sup>H NOESY spectra. **b)** Chemical structures of the four nucleosides indicating sites of selective deuteration (gray shade). Sites containing non-exchangeable protons are colored blue. **c)** Overlay of the unlabeled (fully protiated, gray) and A<sup>H</sup>C<sup>H</sup>-labeled (teal) pre-miR-31 <sup>1</sup>H-<sup>1</sup>H NOESY spectra. **d)** Chemical structures of the four nucleosides indicating sites of selective deuteration (gray shade). Sites containing non-exchangeable protons are colored teal. **e)** Overlay of the unlabeled (fully protiated, gray) and G<sup>H</sup>U<sup>6r</sup>-labeled (purple) pre-miR-31 <sup>1</sup>H-<sup>1</sup>H NOESY spectra. **f)** Chemical structures of the four nucleosides indicating sites of selective deuteration (gray shade). Sites containing non-exchangeable protons are colored purple.

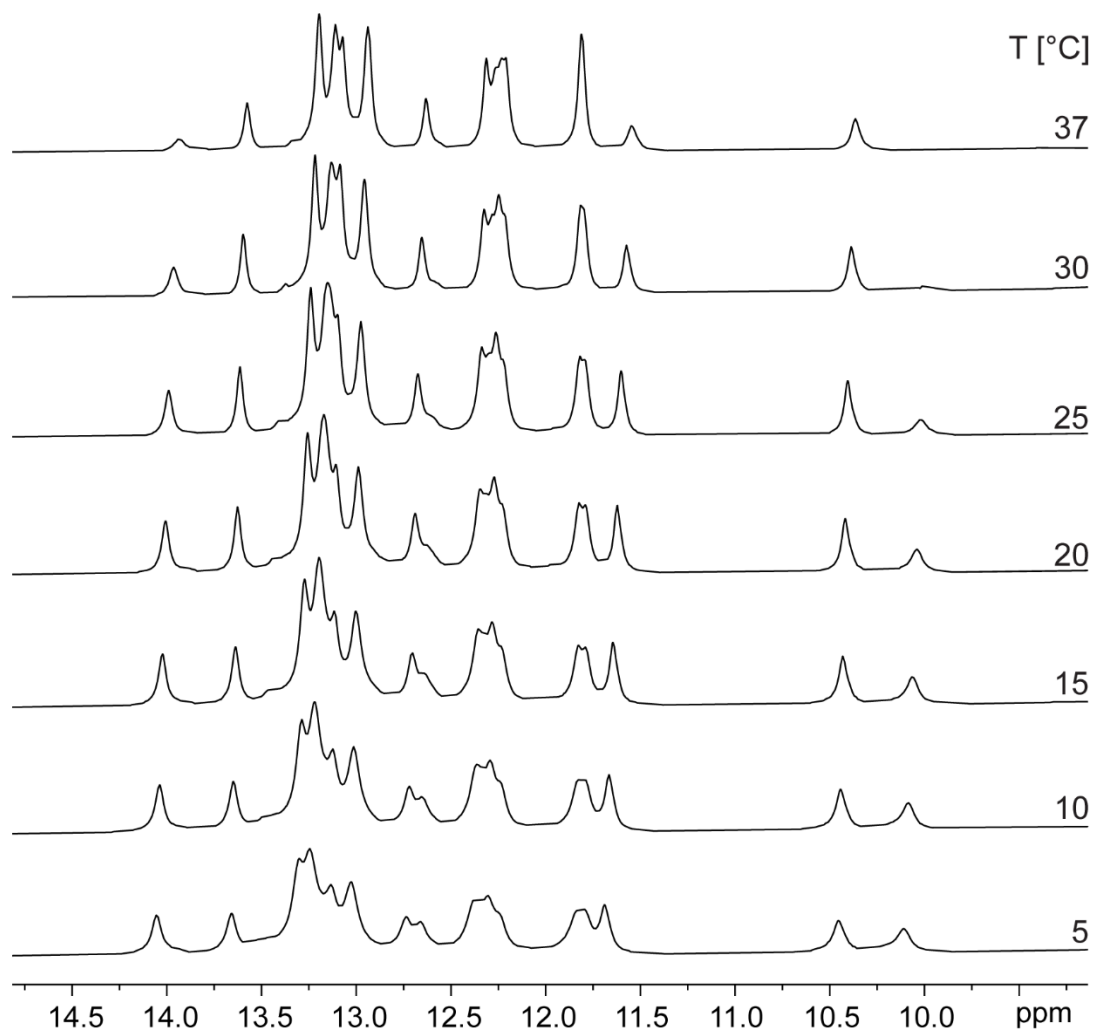


**Appendix A Fig. S6. Summary of the secondary structure, NOE connectivity, and chemical shift assignment validation for FL pre-miR-31.** **a)** The secondary structure is shown beneath the sequence in Vienna format along with arrows to denote helical regions. **b)** NOE upper limit restraints for specified proton pairs used in CYANA and AMBER calculations are drawn as black bars. The thickness of the bar is representative of the strength of the measured NOE. **c)** NMRViewJ chemical shift prediction software was used to validate assignments of H6/H8, H5/H2, H1', H2', H3', C8 and C2. Protons that have been assigned in FL pre-miR-31 are indicated with blue circles (open and closed). Deviation from the predicted chemical shift is represented by deviation from the center. Predictions that are based on examples in the database of chemical shifts are shown as filled circles, predictions without data are shown as open circles. Filled grey squares are present in for nucleotides that do not contain a given proton or carbon. Unassigned resonances have open grey symbols.

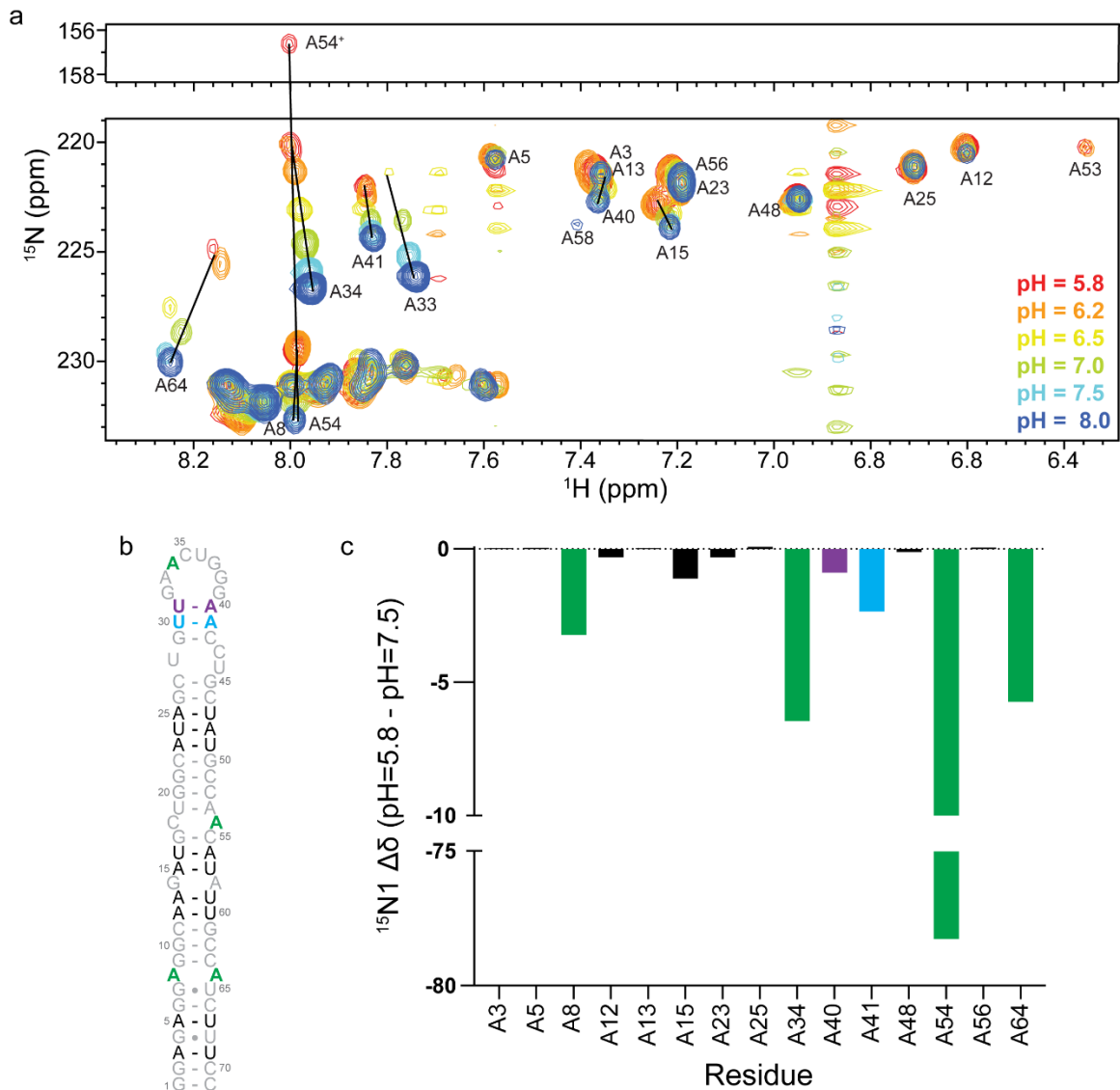


**Appendix A Fig. S7. Imino proton NOESY spectrum of pre-miR-31.** The imino proton NOESY spectrum (mixing time 300 ms) of pre-miR-31 exhibited severe line broadening. The NMR spectrum was recorded at 0.33 mM RNA concentration, 50 mM K-phosphate buffer, pH=7.5, 1 mM MgCl<sub>2</sub> and 90% H<sub>2</sub>O/10% D<sub>2</sub>O at 37°C and 600 MHz.

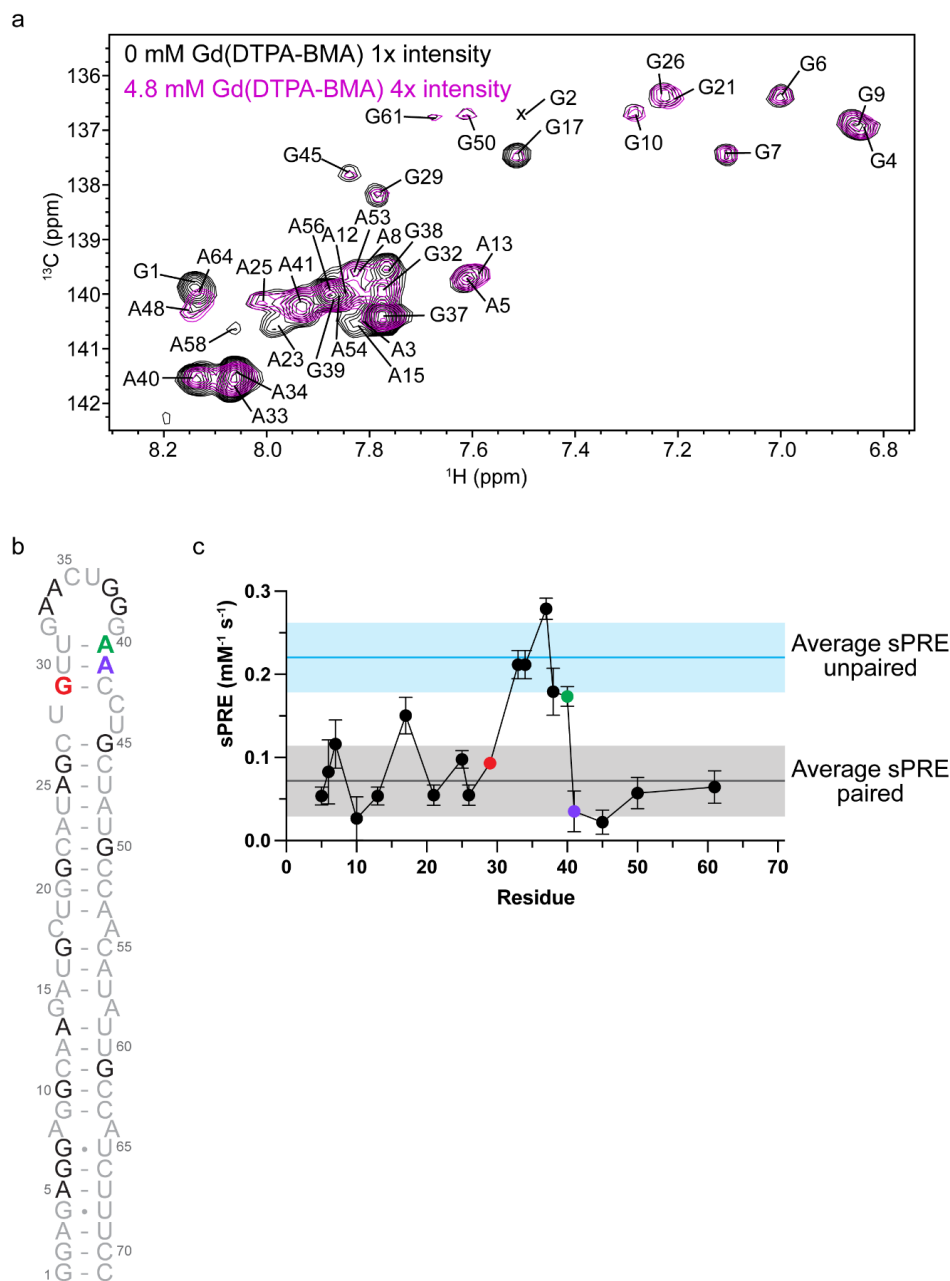




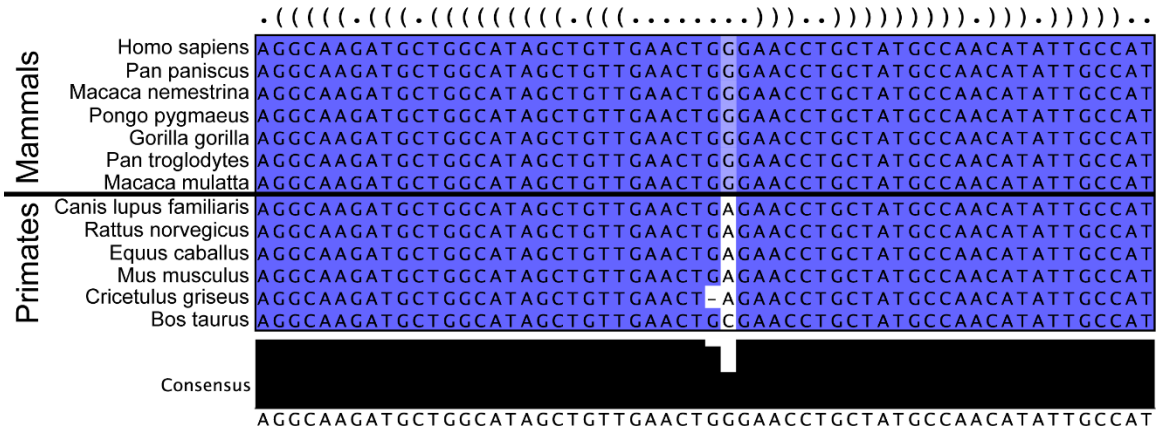
**Appendix A Fig. S8. Imino region of <sup>1</sup>H spectra of pre-miR-31 FL at different temperatures.** The NMR spectra were recorded at 0.3 mM RNA concentration, 50 mM K-phosphate buffer, pH=7.5, 1 mM MgCl<sub>2</sub> and 90% H<sub>2</sub>O/10% D<sub>2</sub>O at 600 MHz and at temperatures between 5 and 37 °C as indicated on the right side of spectra.



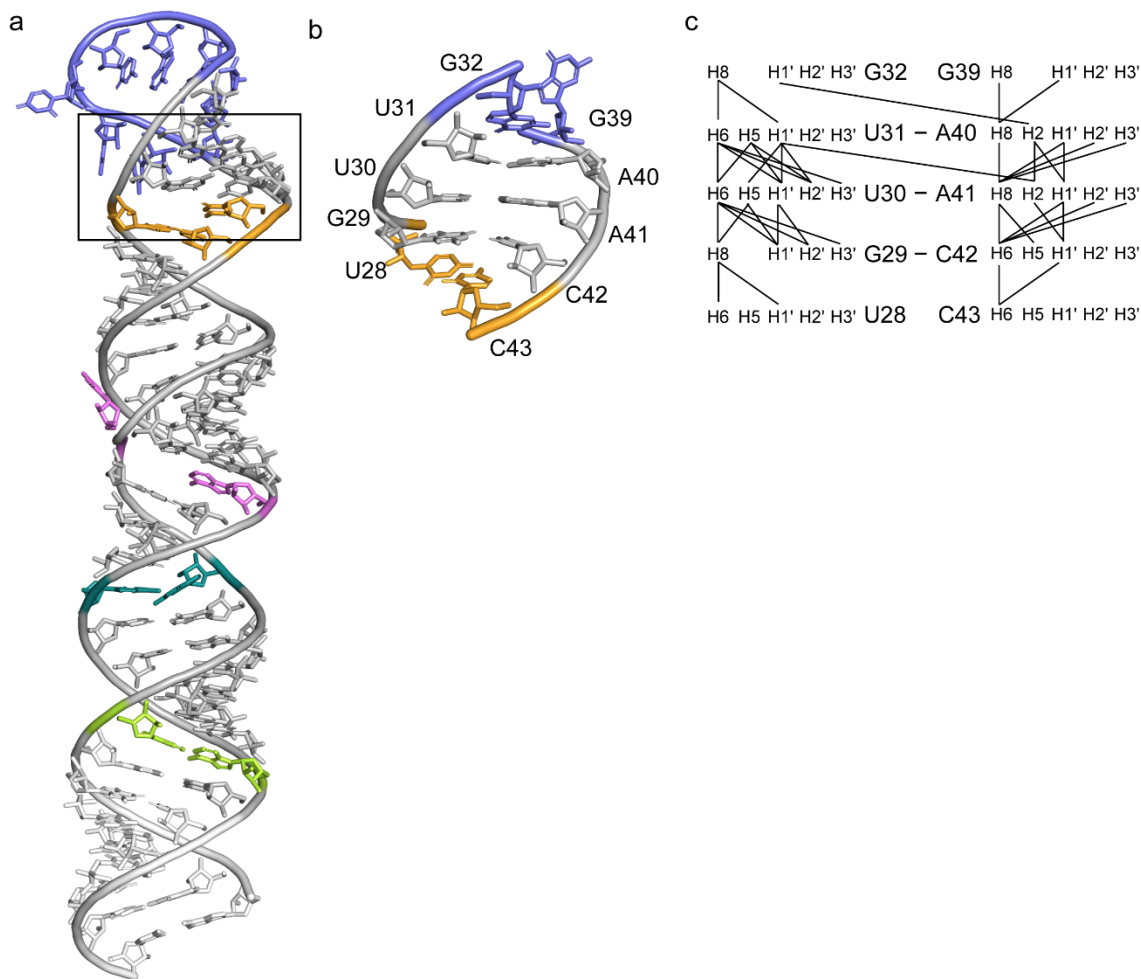
**Appendix A Fig. S9. pH-dependence of unpaired adenosines.** **a)** BEST selective  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -AU labeled FL pre-miR-31, collected at various pH conditions. **b)** Secondary structure of FL pre-miR-31. **c)** Quantification of chemical shift perturbations (pH=5.8 – pH=7.5). A33 and A58 were not included in this analysis. The cross-peak of A33 is too broad to detect at pH=5.8 and cross-peak of A58 is severely overlapped at pH=5.8. Coloring follows the secondary structure in panel b.



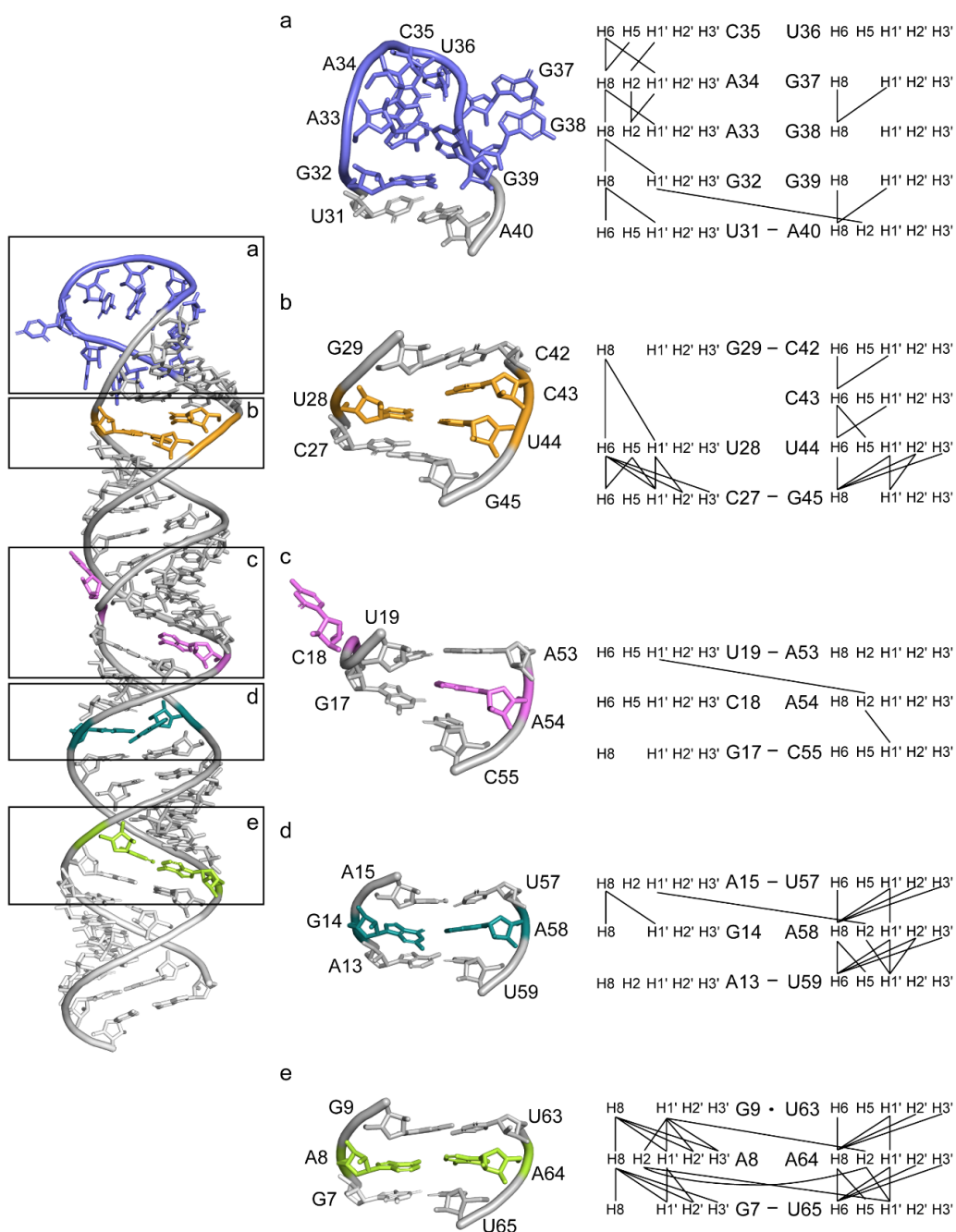
**Appendix A Fig. S10. Solvent paramagnetic relaxation effect analysis of FL pre-miR-31 reveals solvent accessibility in the loop region. a)**  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra of  $^{15}\text{N}/^{13}\text{C}$  A,G-labeled FL pre-miR-31 in the absence (black) and in the presence (magenta) of 4.8 mM paramagnetic compound Gd(DTPA-BMA). NMR spectra were recorded at 0.48 mM RNA concentration, 50 mM K-phosphate buffer (pD=7.5), 1 mM  $\text{MgCl}_2$  and 100%  $\text{D}_2\text{O}$  at 37 °C and 800 MHz. The H8-C8 assignments are labeled on the spectra. **b)** FL pre-miR-31 secondary structure colored to indicate the position of junction residues G29, A40 and A41 (red, green, and purple, respectively). Residues shaded gray were not included in the analysis. **c)** sPRE data for aromatic H8 protons of FL pre-miR-31. The errors of the sPRE values were obtained from the linear regression as described previously (2). The average sPRE values  $\pm$  one standard deviation for unpaired (blue shading) and paired (grey shading) residues are indicated.



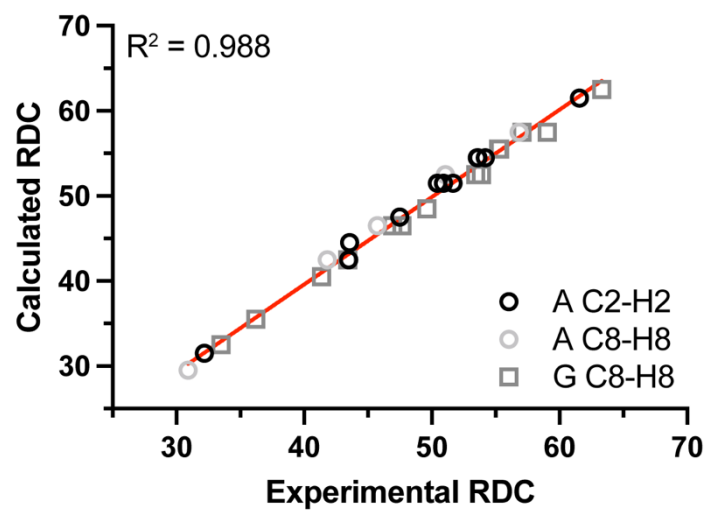
**Appendix A Fig. S11. Sequence conservation of pre-miR-31.**



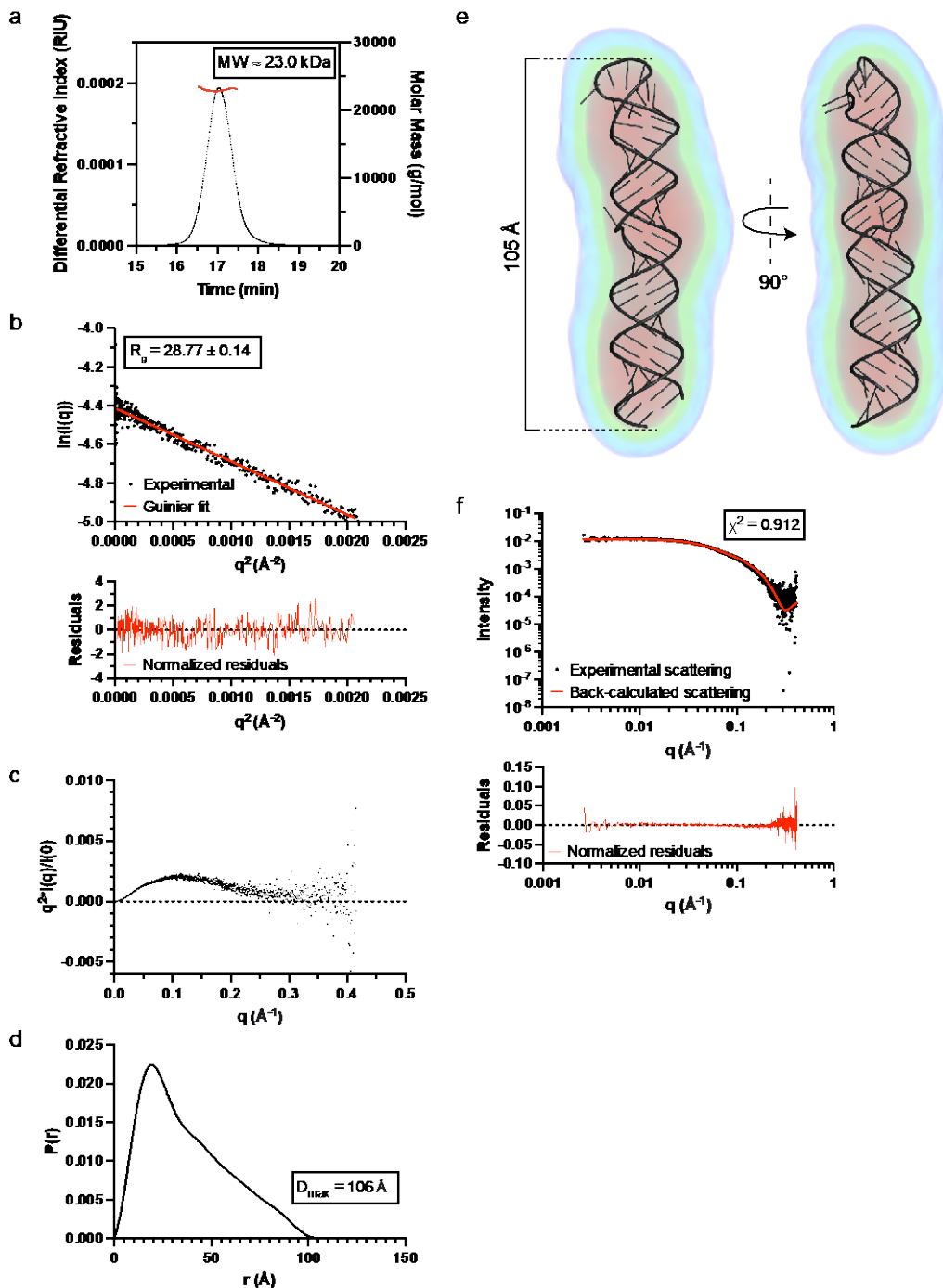
**Appendix A Fig. S12. Schematic presentation of inter-residue NOEs defining the pre-miR-31 junction base pairs. a)** Structure of pre-miR-31 with junction region boxed. **b)** Expanded view of the junction. **c)** Schematic of residues, relevant atoms, and identified inter-residue NOEs used in structure calculations. The black lines represent the inter-residue NOEs.



**Appendix A Fig. S13. Schematic presentation of inter-residue NOEs defines the apical and internal loops in pre-miR-31.** (left) Structure of pre-miR-31 with mismatch/loop regions indicated. (middle) Expanded view of the **a**) apical loop, **b**) dicing site, **c**) C•A mismatch, **d**) A•G mismatch, and **e**) A•A mismatch. (right) Schematic of residues, relevant atoms, and identified inter-residue NOEs used in structure calculations. The black lines represent the inter-residue NOEs.

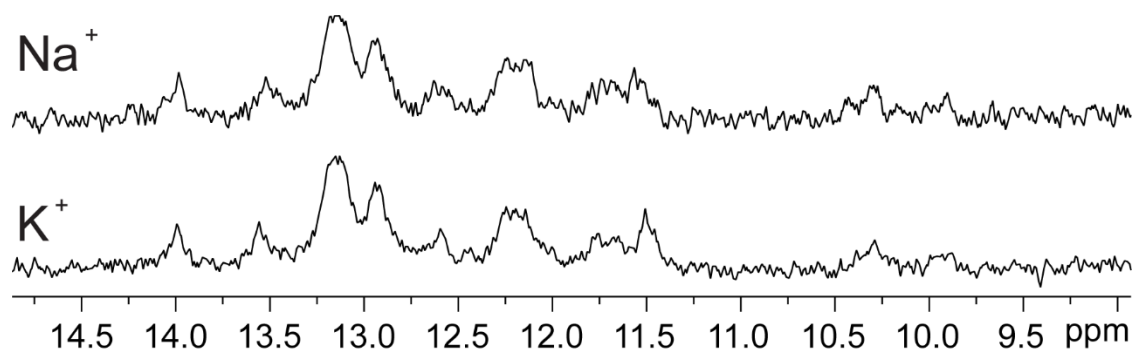


Appendix A Fig. S14. Correlation plot between measured and back-calculated RDCs for the lowest energy pre-miR-31 structure.

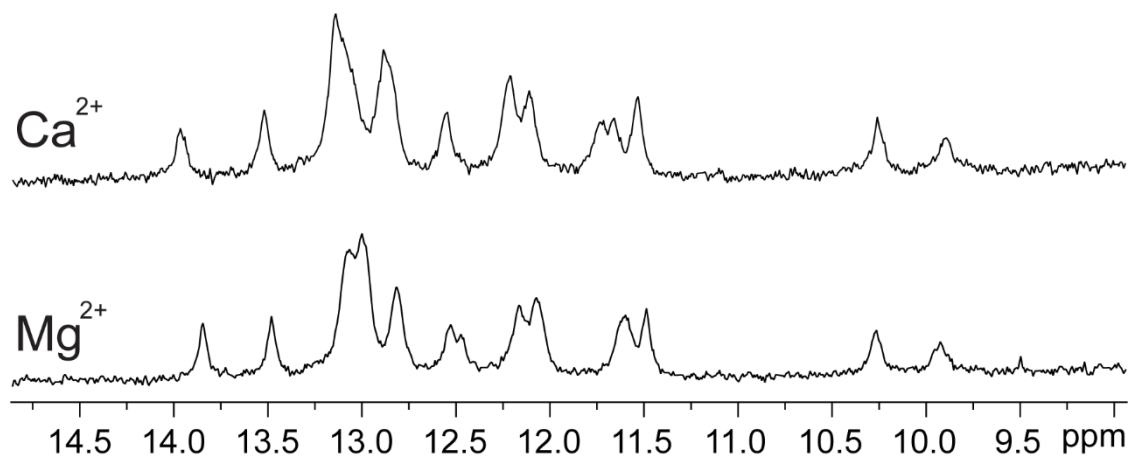


**Appendix A Fig. S15. Validation of pre-miR<sub>31</sub> tertiary structure by SAXS.** **a)** SEC-MALS of pre-miR-31 FL. **b)** Guinier analysis of pre-miR-31 FL used to derive  $R_g$  and  $I(0)$  parameters (top). Normalized residuals are flat and randomly distributed about zero (bottom). **c)** Dimensionless Kratky plot of pre-miR-31 FL. **d)** Pair distance distribution  $[P(r)]$  plot of pre-miR-31 FL. **e)** The lowest energy AMBER-refined pre-miR-31 FL structure aligned to the SAXS electron density map reconstruction. **f)** FoXS back-calculated scattering curve (red) of lowest energy structure fit to experimental SAXS data (black circles) (top). Normalized residuals are flat and randomly distributed about zero (bottom).

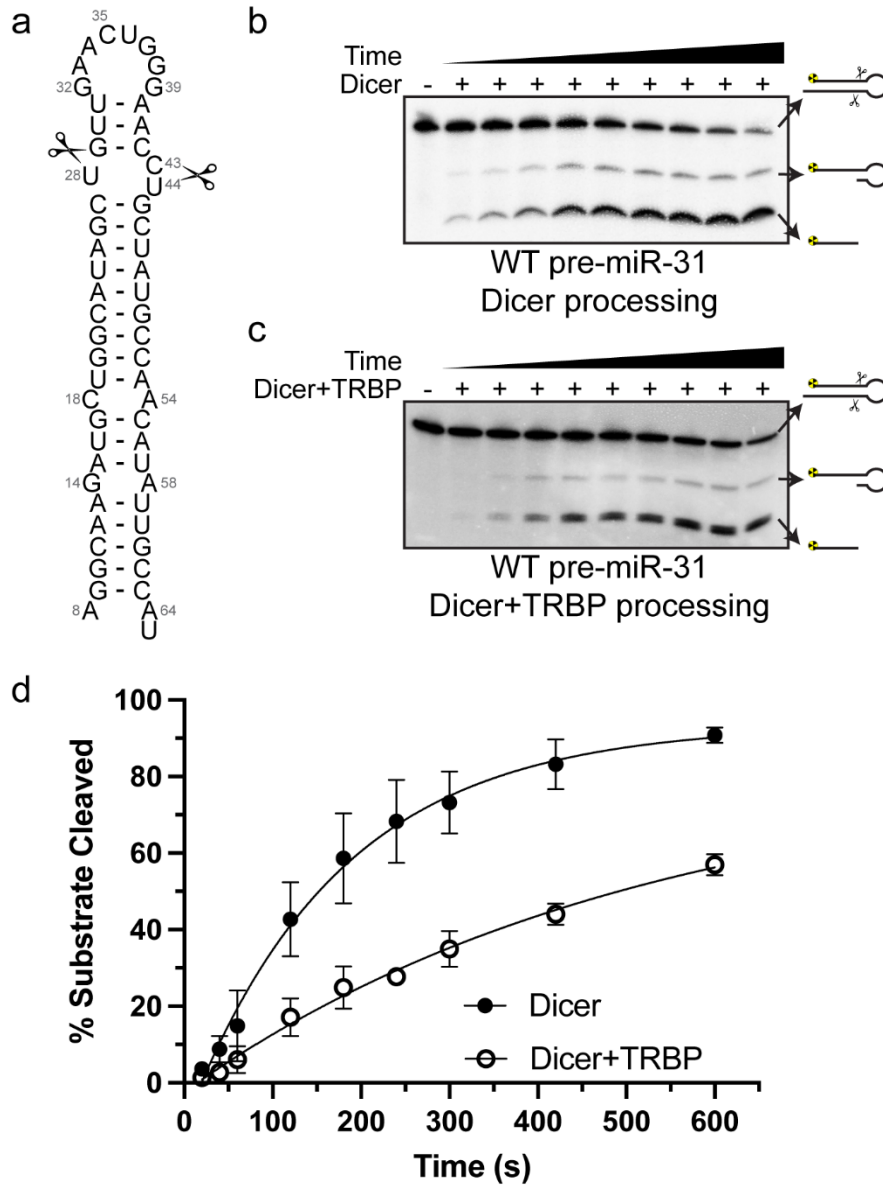




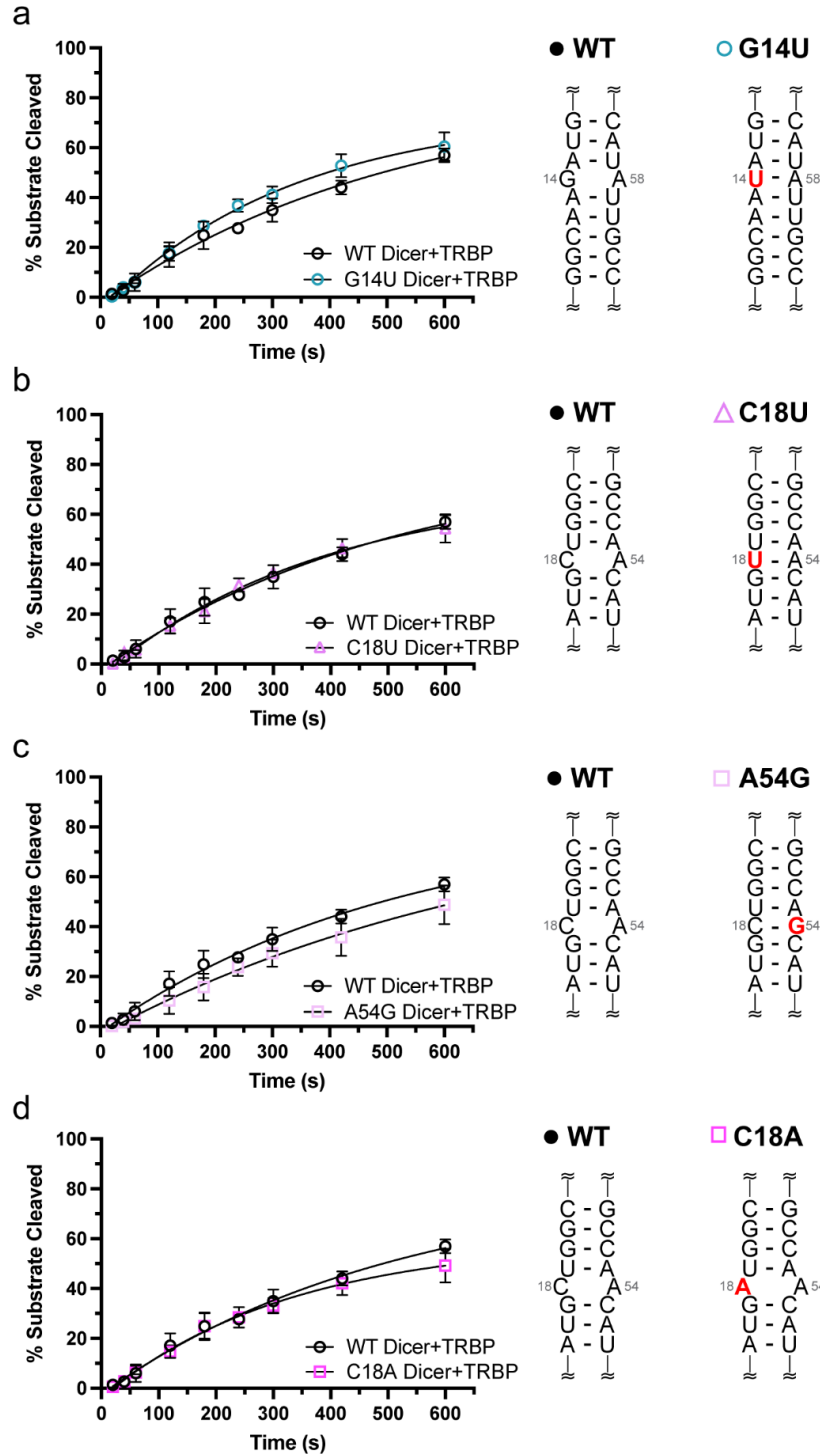
**Appendix A Fig. S16. Structure of pre-miR-31 is not sensitive to the monovalent cation.** The NMR spectra were recorded at 0.01 mM RNA concentration,  $\text{Na}^+$ : 24 mM Na-phosphate buffer, pH=7.5, 100 mM NaCl, 5 mM  $\text{MgCl}_2$ , 4  $\mu\text{M}$  EDTA; and  $\text{K}^+$ : 24 mM K-phosphate buffer, pH=7.5, 100 mM KCl, 5 mM  $\text{MgCl}_2$ , 4  $\mu\text{M}$  EDTA, both at 10 °C and at 600 MHz.



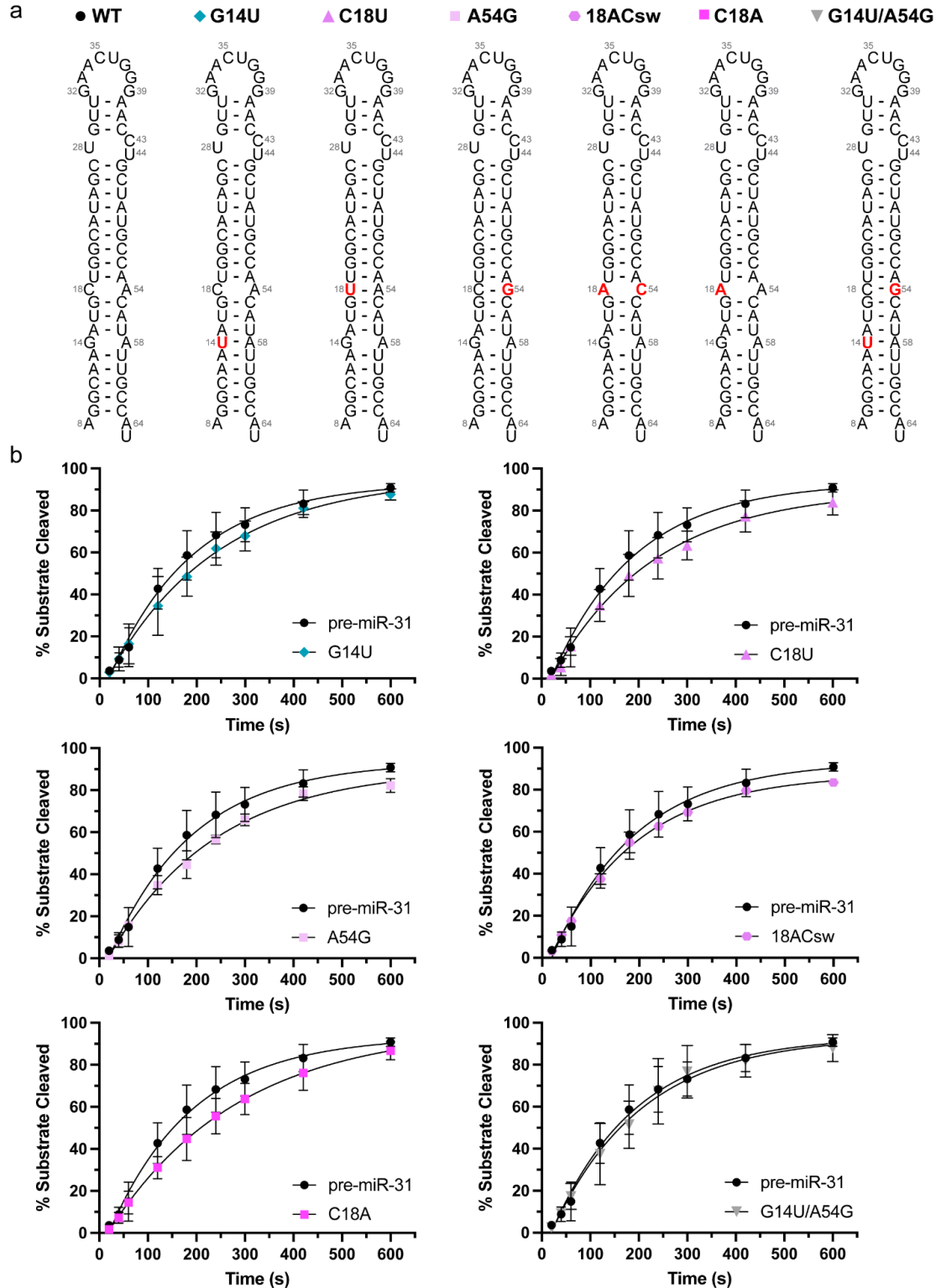
**Appendix A Fig. S17. Structure of pre-miR-31 is not sensitive to the divalent cation.** The NMR spectra were recorded at 0.01 mM RNA concentration, in 24 mM K-phosphate buffer, pH=7.5, 5 mM  $\text{MgCl}_2$  or 5 mM  $\text{CaCl}_2$  (as indicated on the left) and 90%  $\text{H}_2\text{O}/10\%$   $\text{D}_2\text{O}$  at  $10^\circ\text{C}$  and at 600 MHz.

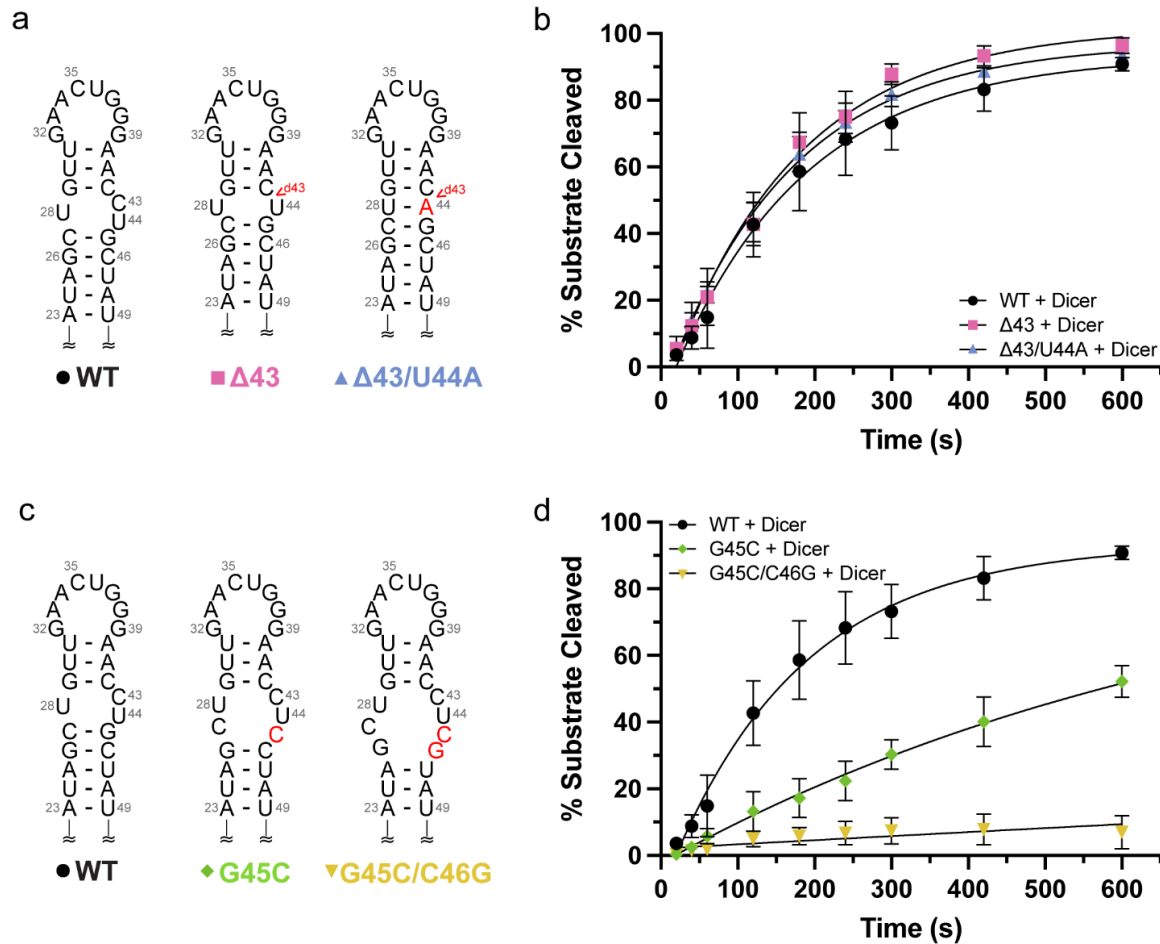


**Appendix A Fig. S18. Dicer processing co-factor, TRBP, inhibits pre-miR-31 processing.** **a)** Secondary structures of pre-miR-31. **b)** Denaturing polyacrylamide gel resolving the RNA products upon incubation with Dicer. **c)** Denaturing polyacrylamide gel resolving the RNA products upon incubation with Dicer-TRBP complex. **d)** Quantification of pre-miR-31 processing with either Dicer (closed circles) or the Dicer-TRBP complex (open circles). Average and standard deviation from n=3 independent assays are presented.

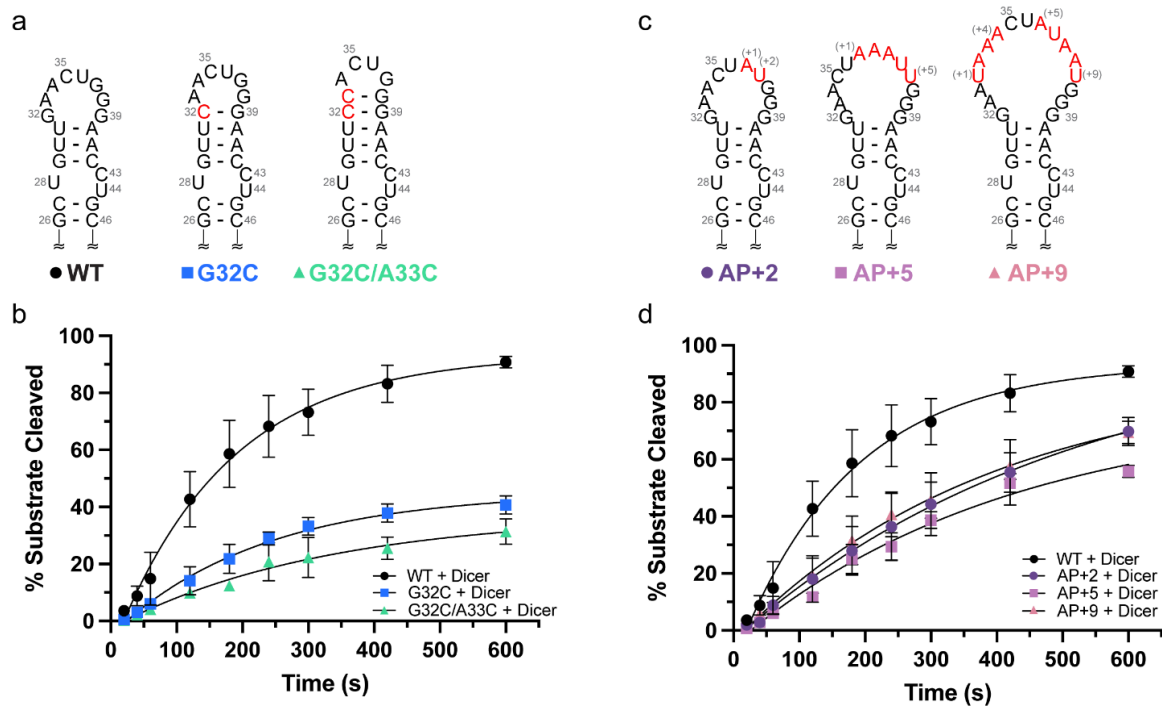


**Appendix A Fig. S19. Mismatches in the stem of pre-miR-31 do not significantly impact Dicer-TRBP processing of the substrate RNA.** Processing assays for WT pre-miR-31 and a) G14U, b) C18U, c) A54G, and d) C18A RNAs. Quantification of pre-miR-31 processing with the Dicer-TRBP complex. Average and standard deviation from n=3 independent assays are presented. Regions of the secondary structures of constructs designed to stabilize or destabilize the stem mismatches are included for clarity. Mutations are indicated with red lettering.

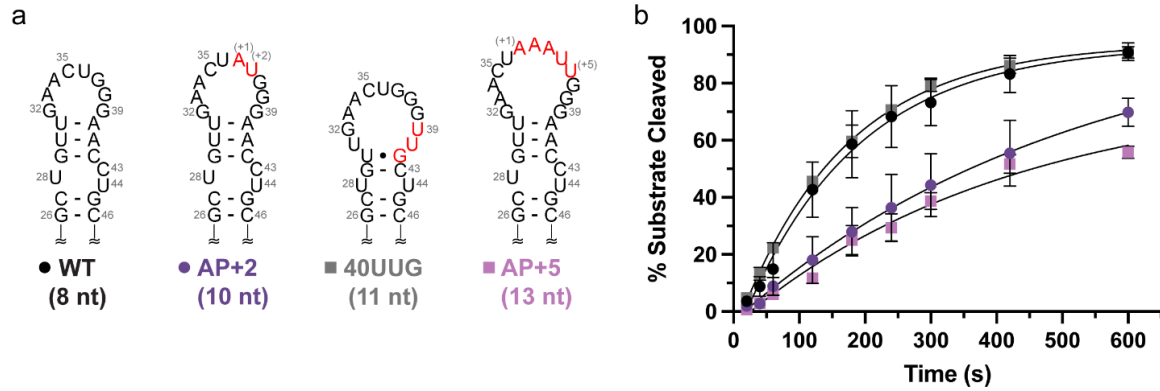




**Appendix A Fig. S21. Structure at the dicing site serves as a control element for Dicer processing.** **a)** Predicted secondary structures of constructs designed to minimize the internal loop at the dicing site. Mutations are indicated with red lettering. **b)** Minimization of the internal loop at the Dicing site does not inhibit Dicer processing. **c)** Dicing site mutants with expanded internal loop structures. Mutations are indicated with red lettering. **d)** Pre-miR-31 RNAs with larger internal loops at the Dicer cleavage site have reduced Dicer processing, relative to WT.

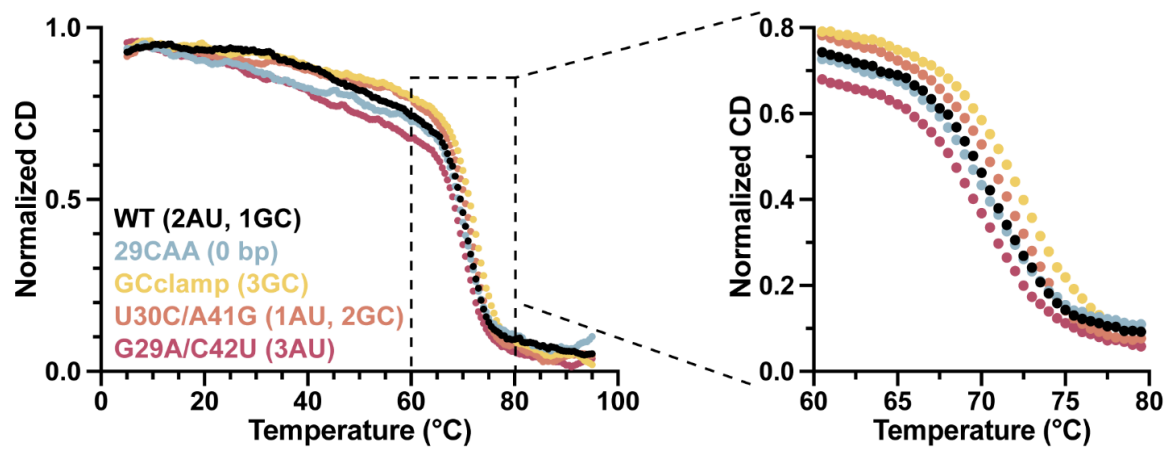


**Appendix A Fig. S22. Dicer processing assays for apical loop mutations.** **a)** Secondary structures of pre-miR-31 RNAs with smaller apical loops. Sites of mutation are denoted with red lettering. **b)** *In vitro* Dicer processing assays reveal a significant reduction in substrate cleavage for G32C and G32C/A33C RNAs. **c)** Secondary structures of mutants designed to extend the pre-miR-31 apical loop. Insertions are indicated with red lettering. **d)** Dicer processing of pre-miR-31 RNAs with larger apical loops was moderately reduced relative to WT. Average and standard deviation from n=3 independent assays are presented.



**Appendix A Fig. S23. A two base-pair junction between the apical loop and dicing site recovers reduced Dicer processing efficiency due to large apical loop size. a)** Secondary structures of WT, AP+2, 40UUG, and AP+5 pre-miR-31 RNAs which have 8, 10, 11, and 13 nucleotide apical loops, respectively. Mutations are indicated with red lettering. **b)** Dicer processing assay of pre-miR-31 RNAs. Average and standard deviation from n=3 independent assays are presented.





**Appendix A Fig. S24. Thermal stability of junction mutants.** RNA thermal denaturation monitored by circular dichroism.

**Appendix A Table S1. Chemical shift completeness.<sup>a</sup>**

<b>TopA RNA</b>	A (% assigned)	C (% assigned)	G (% assigned)	U (% assigned)
H8/H6	100	100	100	100
H2/H5	100	100	/	100
H1'	100	100	100	100
H2'	100	83.3	100	75
H3'	100	83.3	100	75
C6/C8	100	83.3	100	100
C2	100	/	/	/
<b>Top RNA</b>				
<b>Top RNA</b>	A (% assigned)	C (% assigned)	G (% assigned)	U (% assigned)
H8/H6	100	100	90	100
H2/H5	100	100	100	100
H1'	100	100	90	100
H2'	100	100	100	100
H3'	100	100	100	100
C6/C8	100	100	90	100
C2	100	/	/	/
<b>FL pre-miR-31</b>				
<b>FL pre-miR-31</b>	A (% assigned)	C (% assigned)	G (% assigned)	U (% assigned)
H8/H6	100	100	100	100
H2/H5	100	100	100	100
H1'	100	93.8	95	100
H2'	100	93.8	100	100
H3'	94.4	93.8	90	100
C8	100	/	95	/
C2	94.4	/	/	/

<sup>a</sup> “/” indicates a given atom is not present in the nucleoside.

**Appendix A Table S2. NMR restraints and structural statistics for the FL pre-miR-31 structure.<sup>a</sup>**

<b>Cyana<sup>b</sup></b>	
NOE-derived restraints	516
Intraresidue	138
Sequential	359
Long range ( $ i - j  > 1$ )	19
H-bond restraints	134
RDC	30
NOE restraints/residue	7.3
Total restraints/residue	9.6
Target function (Å <sup>2</sup> )	1.99 ± 0.02
Upper distance viol. (Å <sup>2</sup> )	0.0269 ± 0.0002
Lower distance viol. (Å <sup>2</sup> )	0.0412 ± 0.0001
RMSD <sup>c</sup> (Å)	2.68 ± 0.15
Q value	5.4 ± 0.3 %
<b>Amber<sup>d</sup></b>	
Amber energy	-16,427.7
Distance	160.1
Torsion	7.3
RMSD <sup>c</sup> (Å)	2.16 ± 0.48
RMSD lower stem (1-13, 59-71) (Å)	0.22 ± 0.08
RMSD full stem (1-27, 45-71) (Å)	1.30 ± 0.37
RMSD bp in loop (29-31, 40-42) (Å)	0.33 ± 0.11
RMSD loop (32-39) (Å)	3.23 ± 1.04
<b>MolProbity analysis<sup>e</sup></b>	
Clashscore	0.44
Probably wrong sugar pucker (%)	0
Bad backbone conformation (%)	4
Bad bonds (%)	0
Bad angles (%)	0

<sup>a</sup>Statistics are reported for the entire structure unless otherwise specified.

<sup>b</sup>Statistics for the 20 structures with lowest target function.

<sup>c</sup>RMSD: root mean squared deviation.

<sup>d</sup>Statistics for the 20 lowest energy structures.

<sup>e</sup>The 20 amber-refined structures were evaluated using the MolProbity webserver (3, 4)

**Appendix A Table S3. SEC-SAXS data acquisition, sample details, data analysis, model fitting, and software used.**

<b>(a) Sample details</b>	
Organism	Human
Source	<i>In vitro</i> transcribed RNA
<i>Scattering particle composition</i>	
DNA/RNA(s)	Precursor microRNA 31, LM608178
<i>Sample environment/configuration</i>	
Solvent composition	50 mM Potassium Phosphate pH 7.5, 50 mM NaCl, 1 mM MgCl <sub>2</sub>
Sample temperature (°C)	20 °C
In beam sample cell	1.0 mm ID quartz capillary
<i>Size Exclusion Chromatography SEC-SAS</i>	
Sample injection concentration	2.31 mg/mL
Sample injection volume	0.225 mL
SEC column type	Superdex 75 Increase 10/300 GL
SEC flowrate, mL/min	0.6 mL/min
<b>(b) SAS data collection</b>	
Data acquisition/reduction software	SEC-MALS-SAXS. Size separation using a Superdex 200 Increase 10/300 GL column and a 1260 Infinity II HPLC. SAXS data was measured in a sheath-flow cell(5), effective path length 0.542 mm. Data reduced with RAW 2.1.4.
Source/instrument description or reference	BioCAT facility at the Advanced Photon Source beamline 18ID with Eiger2 XE 9M (dectris) detector
Measured $q$ -range ( $q_{min} - q_{max}$ ; Å <sup>-1</sup> , nm <sup>-1</sup> )	0.0027 – 0.42 Å <sup>-1</sup>
Method for scaling intensities	Glassy Carbon, NIST SRM 3600
Exposure time(s), number of exposures. For SEC-SAS, final number of sample frames used for averaging.	0.2 s exposure time with a 1 s total exposure period (0.2 s on, 0.8 s off). 2248 exposures with 20 frames from elution peak used for averaging.
<b>(c) SAS-derived structural parameters</b>	
Methods/Software	Radial averaging; frame comparison, averaging, and subtraction done using BioXTAS RAW 2.1.4 (6)
<i>Guinier Analysis</i>	
$I(0) \pm \sigma$ (cm <sup>-1</sup> ; a.u)	0.01 ± 3.38e-5
$R_g \pm \sigma$ (Å, nm)	28.77 ± 0.14
$min < qR_g < max$ limit (or data point range)	0.077679 < $qR_g$ < 11.959689
Linear fit assessment (definition)	0.98 (r <sup>2</sup> )
<i>PDDF/P(r) analysis</i>	
$I(0) \pm \sigma$ (cm <sup>-1</sup> ; a.u.)	0.01 ± 3.9e-5 (a.u.)

$R_g \pm \sigma$ (Å, nm)	30.46 ± 0.16 (Å)
$d_{\max}$ (Å, nm)	106.0 (Å)
$q$ -range (Å <sup>-1</sup> , nm <sup>-1</sup> )	0.005 – 0.416 (Å <sup>-1</sup> )
$P(r)$ fit assessment (definition)	0.584 ( $\chi^2$ )
<b>(d) Scattering particle size</b>	
Methods/Software	BioXTAS RAW 2.1.4 (6)
<i>Volume estimates</i>	
Porod volume, $V_p$ (Å <sup>3</sup> , nm <sup>3</sup> )	25.9 (Å <sup>3</sup> )
<i>Molecular weight (M) estimates (kDa)</i>	
From chemical composition	23.1
From SAS, concentration independent method	25.5 (Vc)
From SAS-independent measure	23.0 (MALS)
<b>(e) Modelling (a complete sub-panel for each method)</b>	
<i>Shape modelling method(s) (if used)</i>	DENSS
Software	BioXTAS RAW 2.1.4 (6)
$q$ -range for fit ( $q_{\min} - q_{\max}$ ; Å <sup>-1</sup> , nm <sup>-1</sup> )	
Symmetry/anisotropy assumptions	
Number of individual model reconstructions	20
$\chi^2$ , CorMap $P$ -values for fit	3.398e-3 ( $\chi^2$ )
<b>(f) Data and model deposition</b>	
SASBDB IDs	SASDRF9

**Appendix A Table S4. Dicer+TRBP cleavage of pre-miR-31 RNAs.**

Region mutated	RNA constructs	% substrate cleaved (10 min) <sup>a</sup>	$k_{obs}$ (ms <sup>-1</sup> ) <sup>a</sup>	$k_{obs}$ Fold change	P value relative to WT <sup>b</sup>	Difference Level <sup>c</sup>
-	pre-miR-31 WT	57.0 ± 2.8	1.73 ± 0.17	-	-	-
Stem mutations	pre-miR-31 G14U	60.5 ± 5.7	2.09 ± 0.27	1.20	0.128	NS
	pre-miR-31 C18U	54.3 ± 5.7	1.75 ± 0.22	1.01	0.9489	NS
	pre-miR-31 A54G	48.8 ± 7.8	1.33 ± 0.30	0.77	0.1145	NS
	pre-miR-31 C18A	49.2 ± 6.8	1.58 ± 0.22	0.91	0.4004	NS
Dicing site mutations	pre-miR-31 Δ43C	80.3 ± 7.3	5.65 ± 1.25	3.26	0.0057	**
	pre-miR-31 Δ43/U44A	84.5 ± 2.5	4.70 ± 0.16	2.71	<0.0001	****
	pre-miR-31 G45C	40.1 ± 7.0	1.11 ± 0.26	0.64	0.0255	*
	pre-miR-31 G45C/C46G	31.4 ± 9.7	0.84 ± 0.18	0.49	0.0035	**
Apical loop mutations	pre-miR-31 G32C	48.3 ± 2.1	1.52 ± 0.18	0.88	0.2146	NS
	pre-miR-31 G32C/A33C	26.4 ± 5.3	0.71 ± 0.14	0.41	0.0013	**
	pre-miR-31 AP+2	52.3 ± 4.9	1.62 ± 0.13	0.93	0.4079	NS
	pre-miR-31 AP+5	62.3 ± 6.9	2.07 ± 0.33	1.20	0.1917	NS
	pre-miR-31 AP+9	44.2 ± 3.5	1.39 ± 0.17	0.80	0.0703	NS
Junction mutations	pre-miR-31 29CAA	12.3 ± 5.2	0.37 ± 0.13	0.21	0.0004	***
	pre-miR-31 GCclamp	23.5 ± 3.0	0.58 ± 0.08	0.33	0.0005	***
	pre-miR-31 U30C/A41G	59.2 ± 5.2	2.16 ± 0.18	1.25	0.0421	*
	pre-miR-31 G29A/C42U	16.4 ± 5.4	0.39 ± 0.10	0.23	0.0003	***

<sup>a</sup> Average and standard deviation from n=3 independent assays are presented.

<sup>b</sup> P value based on an unpaired parametric t-test.

<sup>c</sup> \* P value <0.05, \*\* P value <0.01, \*\*\* P value <0.001, \*\*\*\* P value <0.0001, NS indicates no significant difference.

**Appendix A Table S5. Dicer cleavage of pre-miR-31 RNAs.**

Region mutated	RNA constructs	% substrate cleaved (10 min) <sup>a</sup>	$k_{obs}$ (ms <sup>-1</sup> ) <sup>a</sup>	$k_{obs}$ fold change	P value relative to WT <sup>b</sup>	Difference Level <sup>c</sup>
-	pre-miR-31 WT	90.8 ± 1.7	4.90 ± 1.07	-	-	-
Stem mutations	pre-miR-31 G14U	87.6 ± 2.7	4.19 ± 0.88	0.85	0.423	NS
	pre-miR-31 C18U	83.8 ± 6.1	3.82 ± 0.88	0.78	0.250	NS
	pre-miR-31 A54G	82.2 ± 3.2	3.72 ± 0.22	0.76	0.135	NS
	pre-miR-31 G14U/A54G	88.0 ± 6.4	4.71 ± 1.38	0.96	0.858	NS
	pre-miR-31 18ACsw	83.4 ± 1.4	4.27 ± 0.16	0.87	0.367	NS
	pre-miR-31 C18A	86.6 ± 4.3	3.66 ± 0.75	0.75	0.175	NS
Dicing site mutations	pre-miR-31 Δ43C	98.0 ± 2.3	6.11 ± 0.80	1.25	0.146	NS
	pre-miR-31 Δ43/U44A	93.5 ± 1.2	5.72 ± 0.61	1.17	0.311	NS
	pre-miR-31 G45C	52.2 ± 4.8	1.34 ± 0.25	0.27	0.0049	**
	pre-miR-31 G45C/C46G	7.0 ± 4.9	0.28 ± 0.12	0.06	0.0017	**
Apical loop mutations	pre-miR-31 G32C	41.8 ± 3.2	1.28 ± 0.15	0.26	0.0044	**
	pre-miR-31 G32C/A33C	31.4 ± 4.4	0.86 ± 0.15	0.17	0.0029	**
	pre-miR-31 AP+2	69.8 ± 5.0	2.14 ± 0.52	0.44	0.0158	*
	pre-miR-31 AP+5	55.8 ± 2.2	1.68 ± 0.15	0.34	0.0067	**
	pre-miR-31 AP+9	69.5 ± 3.9	2.26 ± 0.39	0.46	0.0159	*
	pre-miR-31 40UUG	91.0 ± 3.2	5.39 ± 0.12	1.10	0.473	NS
Junction mutations	pre-miR-31 29CAA	15 ± 4	0.52 ± 0.09	0.11	0.0004	***
	pre-miR-31 GCclamp	58.1 ± 6.7	1.71 ± 0.10	0.35	0.0067	**
	pre-miR-31 U30C/A41G	91.1 ± 3.4	4.12 ± 0.39	0.84	0.300	NS

	pre-miR-31 G29A/C42U	24.8 ± 1.1	0.57 ± 0.06	0.12	0.0022	**
--	-------------------------	------------	----------------	------	--------	----

<sup>a</sup> Average and standard deviation from n=3 independent assays are presented.

<sup>b</sup> P value based on an unpaired parametric t-test.

<sup>c</sup> \* P value <0.05, \*\* P value <0.01, \*\*\* P value <0.001, NS indicates no significant difference.



**Appendix A Table S6. Thermal stability of pre-miR-31 RNAs.**

<b>Region mutated</b>	<b>RNA constructs</b>	<b>T<sub>m</sub> (°C)<sup>a</sup></b>
-	pre-miR-31 WT	70.5 ± 0.1
Junction mutations	pre-miR-31 29CAA	70.4 ± 0.2
	pre-miR-31 3GCclamp	72.0 ± 0.2
	pre-miR-31 U30C A41G	71.4 ± 0.2
	pre-miR-31 G29A C42U	70.3 ± 0.4

<sup>a</sup> T<sub>m</sub> values were obtained by fitting CD thermal denaturation profiles to a two-state unfolding model using sloping baselines. Average and standard deviation from n=3 independent assays are presented.

**Appendix A Table S7. Synthetic DNA templates and associated RNA constructs.**

	5'-sequence-3' <sup>a,b,c</sup>	
	<b>DNA</b>	<b>RNA</b>
Top	mGmGCATAGCAGGTTCCCAGTTCAACAG CTATGCCT <i><u>TATAGTGAGTCGTATTA</u></i>	GGCAUAGCUGUUGAAC UGGGAACCUGCUAUGC C
TopA	mGmGCATAGCCGTAGCTATGCCT <i><u>TATAGTG</u></i> <i><u>AGTCGTATTA</u></i>	GGCAUAGCU <u>ACG</u> GCUA UGCC

<sup>a</sup> m denotes 2'-O-Me modification of the primer.

<sup>b</sup> Italicized nucleotides correspond to the sequence complementary to the T7 promoter.

<sup>c</sup> Red nucleotides indicate non-native tetraloop sequences.

**Appendix A Table S8. DNA primers for pre-miR-31-tail (DMS) experiments.**

OE-PCR primers	5'-sequence-3'
miR31_tail-1F	GCAGCTGAATTCTTCTAATACGACTCACTATAGGAGACCTCGAGTAGAGTCAAAA
miR31_tail-2R	CCAGCATCTTGCCTCCTCTCCTTTTGACCTCTACTCGAGGTCTCCTATAGTG
miR31_tail-3F	GGAGGCAAGATGCTGGCATAGCTGTTGAACTGGGAACCTGCTATGCCAACAT
miR31_tail-4R	AGTTGTTTGGAAAGATGGCAATATGTTGGCATAGCAGGTTCCC
miR31_tail-5F	TTGCCATCTTTCCAAACAACCTCGAGTAGAGTTGACAACAAAGAAACAACAACAACGG
miR31_tail-6R	GCAGGAGGATCCGTTGTTGTTGTTGTTTCTTTGTTGTC
miR_tail_RT	GTTGTTGTTGTTGTTTCTTTGTTGTCAACTCTACTCGAGTTGTTT
miR31_buffer_F	GGAGACCTCGAGTAGAGGTCAAAGGAGAGG

**Appendix A Table S9. DNA primers for generation of the pre-miR-31 (NMR) template.**

OE-PCR primers	5'-sequence-3'
miR-31FL-OE-1F	GTGTCAGAATTCTAATACGACTCACTATAGGAGAGGAGGCAAG
miR-31FL-OE-2R	CCAGTTCAACAGCTATGCCAGCATCTTGCCTCCTCTCCTATAGTGAGT CGTA
miR-31FL-OE-3F	GCTGGCATAGCTGTTGAACTGGGAACCTGCTATGCCAACATATTGCC ATCTTT
miR-31FL-OE-4R	CATAGCGGATCCGGAAAGATGGCAATATGTTGGCATAGCAGGT

**Appendix A Table S10. Amplification primers for template.**

Amplification primers	5'-sequence-3' <sup>a</sup>	application
UNIV-pUC19_E105	TCTTCGCTATTACGCCAGCTGGCGAAA	Forward primer for amplification of DNA template for pre-miR-31 NMR construct and all processing constructs
HDV-AMP-R	mUmAATGTGAGAATTGGCTACGTTGAAACAACGCATTACCG	Reverse primer for amplification of DNA template for all pre-miR-31 processing constructs
miR31_4R	mGmGAAAGATGGCAATATGTTGGCATAGCAGGTT	Reverse primer for amplification of DNA template for pre-miR-31 NMR construct
miR_tail_3buffer_REV	mGmUTGTTGTTGTTGTTTCTTTGTTGTCAACTCTACTCGAGTTG	Reverse primer for amplification of DNA template for pre-miR-31 DMS construct

<sup>a</sup> m denotes 2'-O-Me modification of the primer.

**Appendix A Table S11. DNA primers for HH-pre-miR-31 template.**

OE-PCR primers	5'-sequence-3'
HH_miR31_N at 1F	CCGGAATTCTAATACGACTCACTATAGGGCTC
HH_miR31_N at 2R	ACGTACCCTGATGGTGTACGAGCCCTATAGTGAGTCGTATTA
HH_miR31_N at 3F	ACACCATCAGGGTACGTTTTTCAGACACCATCAGGGTCTGGCATC TTGCCTCT
HH_miR31_N at 4R	CTGACGGTACCGGGTACCGTTTCGTCCTCACGGACTCATCAGAGG CAAGATGCCAGACC
HH_miR31_N at 5F	ACCCGGTACCGTCAGGCAAGATGCTGGCATAGCTGTTGAACTGG GAACCTGCTATGCCAA
HH_miR31_N at 6R	CCGTCGCGGATCCATGGCAATATGTTGGCATAGCAGGTTCCCAGT

**Appendix A Table S12. Mutation DNA primers for processing constructs.**

Mutagenesis primers	5'-sequence-3'	Application
HH-miR-31-HDV-mut-F	TAATGCGTTGTTTCAACGTAGC CAATTCTCACATTAGGATCCTC TAGAGTCGAC	Forward primer to insert HDV-like sequence to 3' end of HH-pre-miR-31
HH-miR-31-HDV-mut-R	CCGACACTACGACGGGGACGTT TCTCACTCAGTGTCATGGCAAT ATGTTGGCATAG	Reverse primer to insert HDV-like sequence to 3' end of HH-pre-miR-31
HH-A54G-HDV-mut-R	CCGACACTACGACGGGGACGTT TCTCACTCAGTGTCATGGCAAT ATGCTGGCATAG	Reverse primer to insert HDV-like sequence to 3' end of HH-pre-miR-31-A54G
miR-31-Nat-G14U-Mut-F	CGTCAGGCAATATGCTGGCATA G	Forward primer for G14U construct mutation
miR-31-Nat-G14U-Mut-R	GTACCGGGTACCGTTTCG	Reverse primer for G14U construct mutation
miR31-G14U-11nt-F	GGTCTGGCATATTGCCTCTGA	Forward primer for G14U hammerhead complementary sequence mutation
miR31-G14U-11nt-R	CTGATGGTGTCTGAAAAACG	Reverse primer for G14U hammerhead complementary sequence mutation
miR-31-Nat-C18U-Mut-F	AGGCAAGATGTTGGCATAGCTG TTG	Forward primer for C18U construct mutation
miR-31-Nat-C18U-Mut-R	GACGGTACCGGGTACCGT	Reverse primer for C18U construct mutation
miR31-C18U-11nt-F	TCAGGGTCTGACATCTTGCCT	Forward primer for C18U hammerhead complementary sequence mutation
miR31-C18U-11nt-R	TGGTGTCTGAAAAACGTAC	Reverse primer for C18U hammerhead complementary sequence mutation
miR-31-Nat-C18A-31-F	AGGCAAGATGATGGCATAGCT G	Forward primer for C18A construct mutation
miR-31-Nat-C18A-31-R	GACGGTACCGGGTACCGT	Reverse primer for C18A construct mutation
miR-31-Nat-C18A-11nt-F	TCAGGGTCTGTCATCTTGCCT	Forward primer for C18A hammerhead complementary sequence mutation
miR-31-Nat-C18A-11nt-R	TGGTGTCTGAAAAACGTAC	Reverse primer for C18A hammerhead complementary sequence mutation
miR31-18Acsw-mut-F	TGCTATGCCACCATATTGCCAT G	Forward primer for 18Acsw construct mutation
miR31-18Acsw-mut-R	GGTTCCCAGTTCAACAGC	Reverse primer for 18Acsw construct mutation

miR-31-Nat-29CAA-Mut-F	TGGCATAGCTCAAGAAGACTGGGAACC	Forward primer for 29CAA construct mutation
miR-31-Nat-29CAA-Mut-R	GCATCTTGCCTGACGGTA	Reverse primer for 29CAA construct mutation
miR31-nat-G32C-F	CATAGCTGTTCAACTGGGAACC	Forward primer for G32C construct mutation
miR31-nat-G32C-R	CCAGCATCTTGCCTGACG	Reverse primer for G32C construct mutation
miR31-Nat-G32C/A33C-F	CATAGCTGTTCCACTGGGAACCTG	Forward primer for G32C/A33C construct mutation
miR31-Nat-G32C/A33C-R	CCAGCATCTTGCCTGACG	Reverse primer for G32C/A33C construct mutation
miR31-nat-40UUG-mut-F	TTGAACTGGGTTGCTGCTATGCCAAC	Forward primer for 40UUG construct mutation
miR31-nat-40UUG-mut-R	CAGCTATGCCAGCATCTTG	Reverse primer for 40UUG construct mutation
miR31-nat-A54G-mut-F	TGCTATGCCAGCATATTGCCAT	Forward primer for A54G construct mutation
miR31-nat-A54G-mut-R	GGTTCACAGTTCAACAGC	Reverse primer for A54G construct mutation
miR31-AP+2-mut-F	ATGGGAACCTGCTATGCCAA	Forward primer for AP+2 construct mutation
miR31-AP+2-mut-R	AGTTCAACAGCTATGCCAG	Reverse primer for AP+2 construct mutation
miR31-AP+5-mut-F	ATAATGGGAACCTGCTATGCCAA	Forward primer for AP+5 construct mutation
miR31-AP+5-mut-R	AGTTCAACAGCTATGCCAG	Reverse primer for AP+5 construct mutation
miR31-AP+9-mut-F	TATAATGGGAACCTGCTATGCCAA	Forward primer for AP+9 construct mutation
miR31-AP+9-mut-R	GTTTATTCAACAGCTATGCCAGCAT	Reverse primer for AP+9 construct mutation
miR31-3Gcclamp-mut-F	TGGGCGCCTGCTATGCCAACATATTG	Forward primer for 3Gcclamp construct mutation
miR31-3Gcclamp-mut-R	GTTCCGCAGCTATGCCAGCATCTTG	Reverse primer for 3Gcclamp construct mutation
miR31-G45C-mut-F	CTGGGAACCTCCTATGCCAAC	Forward primer for G45C construct mutation
miR31-G45C-mut-R	TTCAACAGCTATGCCAGC	Reverse primer for G45C construct mutation
miR31-G45C C46G-mut-F	CTGGGAACCTCGTATGCCAACATATTG	Forward primer for G45C/C46G construct mutation
miR31-G45C C46G-mut-R	TTCAACAGCTATGCCAGC	Reverse primer for G45C/C46G construct mutation
miR31-d43C-	TGCTATGCCAACATATTGC	Forward primer for $\Delta$ 43



mut-F		construct mutation
miR31-d43C-mut-R	GTTCCCAGTTCAACAGCTATG	Reverse primer for $\Delta$ 43 construct mutation
miR31-4344A-mut-F	AACTGGGAACAGCTATGCCAAC	Forward primer for $\Delta$ 43/U44A construct mutation
miR31-4344A-mut-R	CAACAGCTATGCCAGCAT	Reverse primer for $\Delta$ 43/U44A construct mutation
G29A C42U-mut-FWD	TGGGAATCTGCTATGCCAACAT ATTG	Forward primer for G29A/C42U construct mutation
G29A C42U-mut-REV	GTTCAATAGCTATGCCAGCATC TTG	Reverse primer for G29A/C42U construct mutation
U30C A41G-mut-FWD	TGGGAGCCTGCTATGCCAACAT ATTG	Forward primer for U30C/A41G construct mutation
U30C A41G-mut-REV	G TTCAGCAGCTATGCCAGCATC TTG	Reverse primer for U30C/A41G construct mutation

**Appendix A Table S13. RNA sequences used for structural and processing studies.**

construct name	5'-RNA sequence-3'	Application
miR-31_DMS	GGAGACCUCGAGUAGAGGUCAAAG GAGAGGAGGCAAGAUGCUGGCAUA GCUGUUGAACUGGGAACCUGCUAUG CCAACAUAUUGCCAUCUUUCCAAAC AACUCGAGUAGAGUUGACAACAAAG AAACAACAACAACAAC	DMS chemical probing
FL-pre-miR-31	GGAGAGGAGGCAAGAUGCUGGCAU AGCUGUUGAACUGGGAACCUGCUAU GCCAACAUUAUUGCCAUCUUUCC	Structure
WT pre-miR-31	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGAACCUGCUAUGCCAACAU AUUGCCAU	Processing
WT pre-miR-31 -G14U	AGGCAAUAUGCUGGCAUAGCUGUUG AACUGGGAACCUGCUAUGCCAACAU AUUGCCAU	Processing
WT pre-miR-31-C18U	AGGCAAGAUGUUGGCAUAGCUGUU GAACUGGGAACCUGCUAUGCCAACA UAUUGCCAU	Processing
WT pre-miR-31-C18A	AGGCAAGAUGAUGGCAUAGCUGUU GAACUGGGAACCUGCUAUGCCAACA UAUUGCCAU	Processing
WT pre-miR-31-18Acsw	AGGCAAGAUGAUGGCAUAGCUGUU GAACUGGGAACCUGCUAUGCCACCA UAUUGCCAU	Processing
WT pre-miR-31-A54G	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGAACCUGCUAUGCCAGCAU AUUGCCAU	Processing
WT pre-miR-31-40UUG	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGUUGCUGCUAUGCCAACAU AUUGCCAU	Processing
WT pre-miR-31-29CAA	AGGCAAGAUGCUGGCAUAGCUC AAG AACUGGGAACCUGCUAUGCCAACAU AUUGCCAU	Processing
WT pre-miR-31-G32C	AGGCAAGAUGCUGGCAUAGCUGUUC AACUGGGAACCUGCUAUGCCAACAU AUUGCCAU	Processing
WT pre-miR-31-G32C/A33C	AGGCAAGAUGCUGGCAUAGCUGUUC CACUGGGAACCUGCUAUGCCAACAU AUUGCCAU	Processing
WT pre-miR-31-G14U/A54G	AGGCAAUAUGCUGGCAUAGCUGUUG AACUGGGAACCUGCUAUGCCAGCAU AUUGCCAU	Processing
WT pre-miR-31-AP+2	AGGCAAGAUGCUGGCAUAGCUGUUG AACUAUGGGAACCUGCUAUGCCAAC	Processing

	AUAUUGCCAU	
WT pre-miR-31-AP+5	AGGCAAGAUGCUGGCAUAGCUGUUG AACUAUAAUGGGAACCUAGCUAUGCC ACAUAUUGCCAU	Processing
WT pre-miR-31-AP+9	AGGCAAGAUGCUGGCAUAGCUGUUG AAUAAACUAUAAUGGGAACCUAGCUA UGCCAACAUAUUGCCAU	Processing
WT -miR-31-Gcclamp	AGGCAAGAUGCUGGCAUAGCUGCGG AACUGGGCGCCUGCUAUGCCAACA AUUGCCAU	Processing
WT pre-miR-31-G45C	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGAACCUCCUAUGCCAACA AUUGCCAU	Processing
WT pre-miR-31- G45C/C46G	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGAACCUAGCUAUGCCAACA AUUGCCAU	Processing
WT pre-miR-31- $\Delta$ 43	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGAACUGCUAUGCCAACA UAUUGCCAU	Processing
WT pre-miR-31- $\Delta$ 43/U44A	AGGCAAGAUGCUGGCAUAGCUGUUG AACUGGGAACAGCUAUGCCAACA UAUUGCCAU	Processing
WT pre-miR-31- U30C/A41G	AGGCAAGAUGCUGGCAUAGCUGCUG AACUGGGAGCCUGCUAUGCCAACA AUUGCCAU	Processing
WT pre-miR-31- G29A/C42U	AGGCAAGAUGCUGGCAUAGCUAUUG AACUGGGAUCUGCUAUGCCAACA AUUGCCAU	Processing

## References

1. P. Z. Johnson, W. K. Kasprzak, B. A. Shapiro, A. E. Simon, RNA2Drawer: geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. *RNA biology* **16**, 1667-1671 (2019).
2. C. Hartmueller *et al.*, RNA structure refinement using NMR solvent accessibility data. *Scientific reports* **7**, 1-10 (2017).
3. D. I. L.-F. A. Chen, V. B. J. K. G. Wang, X. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-W383 (2007).
4. C. J. Williams *et al.*, MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**, 293-315 (2018).
5. N. Kirby *et al.*, Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallographica Section D: Structural Biology* **72**, 1254-1266 (2016).
6. J. B. Hopkins, R. E. Gillilan, S. Skou, BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of applied crystallography* **50**, 1545-1553 (2017).

## Appendix B Dicer Only Data and Conclusions for Chapter III

### Mismatches within the helical stem region have no impact on Dicer cleavage

Base pair mismatches are a common feature within the helical stem of precursor microRNAs[1]. Increasing the length of the pre-miR helical stem by including additional base paired sequences is detrimental for Dicer processing[2, 3]. Studies on fly Dicer-1 suggest that while the length of the pre-miR helical stem is important, the presence of mismatches does not significantly affect Dicer processing[4]. However, because pre-miR-31 biogenesis does not appear to be regulated by protein binding partners, we wanted to consider all aspects of pre-miR-31 structure that could be involved in regulating processing. To investigate the role of individual base pair mismatches in the Dicer processing of WT pre-miR-31, we sought to stabilize the G14•A58 mismatch. We made a single point mutation (G14U) which converted the mismatch into a canonical U-A base pair (**Appendix A Fig. S20**). Quantification of Dicer processing revealed WT-levels of processing of the G14U mutant pre-miR (**Appendix A Table S5, Appendix A Fig. S20**).

We previously investigated the pH-dependence of the C18•A54 mismatch and found that A54 is partially protonated at physiological pH, suggesting that these bases can form a C•A<sup>+</sup> base pair near neutral pH[1]. We were therefore interested in testing if mutations that replaced the mismatch with a canonical U-A or C-G base pair (C18U and A54G, respectively) affected the processing by Dicer (**Appendix A Fig. S20**). As with stabilization of the G•A mismatch, stabilization of the C•A mismatch did not affect the efficiency of Dicer processing (**Appendix A**

**Table S5, Appendix A Fig. S20**). We next examined the Dicer processing efficiency of mutant (G14U/A54G) that stabilized both mismatches with canonical base pairs. We found that pre-miR-31 G14U/A54G was processed similarly to WT (**Appendix A Table S5, Appendix A Fig. S20**). We next examined the importance of the context of the C•A mismatch by swapping the bases (18ACsw). Again, we observed no significant change in Dicer processing efficiency (**Appendix A Table S5, Appendix A Fig. S20**).

All pre-miR-31 mutant RNAs we examined were cleaved to approximately 90%. Maintaining the same stem length, the absence of one (G14U, C18U, A54G) or two (G14U/A54G) mismatches within the stem of WT pre-miR-31 does not significantly alter the Dicer cleavage efficiency, consistent with studies on fly Dicer-1.[4] However, the measured binding affinity of G14U/A54G for Dicer decreased 2.5-fold relative to WT (**Table 1**). Binding of G14U, C18U, A54G mutants to Dicer were similar to WT while the binding affinity of 18ACsw was slightly enhanced (2-fold). These findings suggest that the mismatches in pre-miR-31 stem are important features for Dicer binding.

### **Structure at the cleavage site affects Dicer processing**

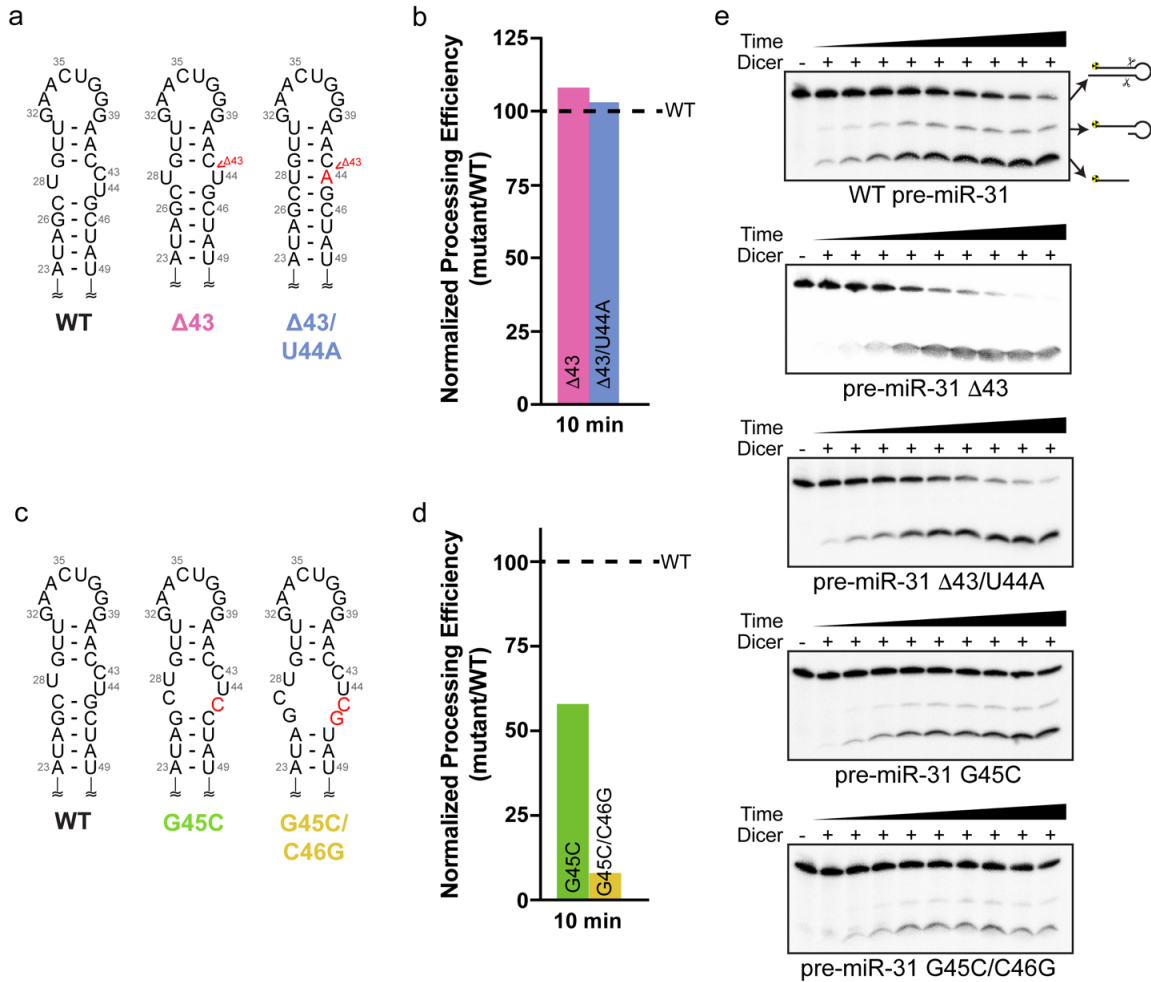
The RNase III and helicase domains of Dicer interact with the upper stem loop region (which includes the apical loop and the dicing site).[2, 5, 6] Studies strongly indicate that the structure in this region may regulate Dicer processing.[2, 4-6] To distinguish between the importance of structure at distinct regions within the upper stem loop regions, we employed a mutational approach which reshaped the apical loop and the dicing site, independently.

First, we generated four different Dicer processing site mutants and examined the impact

of structure at this site on Dicer processing. We examined two mutations that either minimized ( $\Delta 43$ ) or eliminated ( $\Delta 43/U44A$ ) the internal loop at the pre-miR-31 Dicer processing site (**Fig. 1a**). The  $\Delta 43$  construct is processed more efficiently than WT. This is particularly noticeable at timepoints early in the reaction. Interestingly, the  $\Delta 43/U44A$  construct exhibited a slight processing enhancement relative to WT, but was not processed as efficiently as  $\Delta 43$  (**Fig. 1b**). These findings suggest that a small 1x1 internal loop structure serves as a better substrate for Dicer processing.

Conversely, we found that mutations that enlarged the internal loop at the dicing site resulted in RNAs that were inefficiently processed by Dicer (**Fig. 1c**). The G45C mutant, which increases the WT 1x2 internal loop to a 2x3 internal loop, has ~50% reduced processing efficiency while the G45C/C46G mutant (3x4 internal loop) exhibits almost no processing (**Fig. 1d**). Furthermore, we found that  $\Delta 43C$  and  $\Delta 43C/U44A$ , which minimized and eliminated the internal loop, respectively, promoted 5' strand cleavage by Dicer, eliminating the partially processed intermediate, and generating more mature miR (**Fig. 1e**).

Collectively, we found that a 1x1 internal loop at the Dicing site is the best substrate for Dicer processing, while a fully base paired or the native 1x2 internal loop at cleavage site are suboptimal substrates. Pre-miRs with too large of an internal loop around the cleavage site are poor substrates for Dicer to cleave. The binding affinity for Dicer was measured and we found that the  $\Delta 43$  mutant bound Dicer with near WT affinity, while the  $\Delta 43/U44A$  mutant and the G45C mutant both had a slightly weaker affinity. Introduction of a large internal loop (45/46) reduced binding by ~6-fold (**Table 1**). Together, our results suggest that Dicer binding affinity and processing efficiency are not strictly correlated, consistent with previous studies[3].



**Appendix B Figure 1. Structure at the dicing site serves as a control element for Dicer processing.** **a)** Secondary structures of constructs designed to minimize the internal loop at the dicing site. Mutations are indicated with red lettering. **b)** Dicer processing efficiency for  $\Delta 43$  and  $\Delta 43/U44A$  mutants normalized to WT pre-miR-31 at 10 min. **c)** Secondary structures of constructs designed to expand the internal loop at the dicing site. Mutations are indicated with red lettering. **d)** Dicer processing efficiency for G45C and G45C/C46G mutants normalized to WT pre-miR-31 at 10 min. **e)** Processing assay gels of hDicer (20 nM) with WT and dicing site mutant pre-miR-31 RNAs (2 nM) at pH = 7.5.



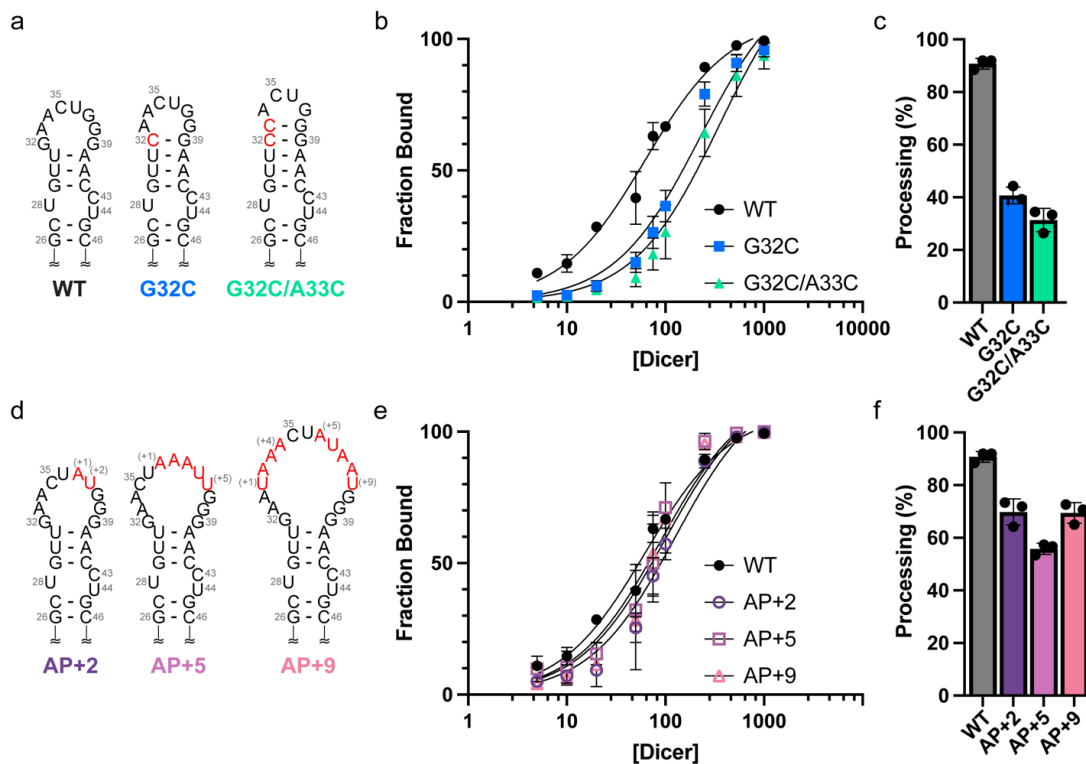
## Size and relative position of the apical loop regulates Dicer processing efficiency and specificity

We next examined the impact of apical loop size on Dicer processing. Apical loop flexibility serves as a control mechanism in many pre-miR/pri-miR elements[7, 8] and the apical loop has been identified as a target for regulation by small molecules or peptides[9-11]. Fly Dicer-1 binds to pre-let-7 with 4-nt loop six times weaker than pre-let-7 with 14-nt loop,[4] and the weaker binding leads to poorer cleavage efficiency. However, another study shows that human Dicer binds similarly with different loop sized pre-miR mutants and has uncoupled Dicing activity[3]. To further elucidate these findings, we designed two constructs, G32C and G32C/A33C, which minimize the apical loop size by forming one or two canonical base pairs within the otherwise unpaired region (**Fig. 2a**). Dicer binds the G32C RNA (6-nt loop) and the G32C/A33C (4-nt loop) about four times and six times weaker than WT pre-miR-31 (8-nt loop), respectively (**Fig. 2b, Table 1**). The reduced binding affinity correlates with reduced cleavage efficiency (**Fig. 2c**). This result is consistent with observations made with fly Dicer-1[4].

Pre-miRs with small apical loops (3-9 nt long) were identified as poor substrates for human Dicer processing, and RNAs with larger apical loops were preferred by Dicer and Drosha[3]. We next examined how increasing the apical loop size impacted Dicer cleavage. We added non-native nucleotides to the apical loop regions of pre-miR-31 to generate AP+2 (10-nt loop), AP+5 (13-nt loop) and AP+9 constructs (17-nt loop) (**Fig. 2d**). These larger loop mutants bound human Dicer ~2-fold weaker than WT (**Fig. 2e, Table 1**). We found that increasing the apical loop size reduced Dicer processing, but not to the same extent as minimizing the apical loop size (**Fig. 2f**).

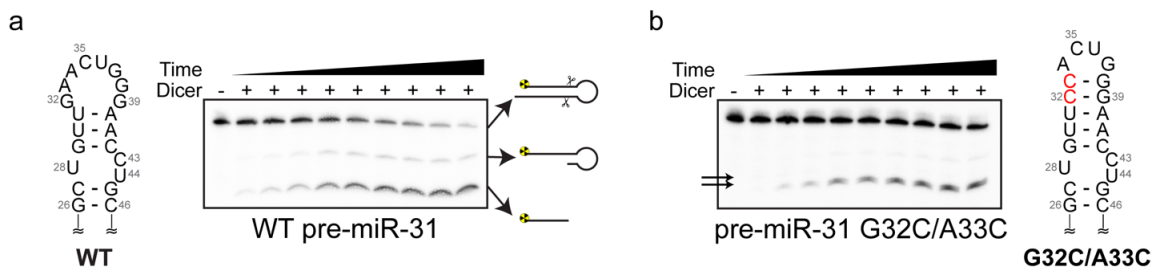
The reduction in processing efficiency caused by the presence of a larger apical loop can be offset by other factors. Previous studies showed that the apical loop or an internal loop 2-nt

from cleavage sites could enhance cleavage efficiency of shRNAs[5, 12]. Consistent with previous studies, we observed WT-level processing for a pre-miR-31 construct which contains an 11-nt loop positioned 2-nt from the cleavage site (40UUG, **Appendix A Fig. S23**). Furthermore, the 40UUG construct generates a U•U mismatch at the dicing site. We demonstrated that dicing site mutants that have 1x1 internal loops at the dicing site are better substrates for Dicer. The restructuring of the Dicing site may further compensate the presence of a larger apical loop.



**Appendix B Figure 2. Apical loop size is optimized for efficient Dicer binding and processing.**

**a)** Secondary structures of constructs designed to minimize the pre-miR-31 apical loop. Mutations are indicated with red lettering. **b)** Quantification of the binding affinity of pre-miR-31 RNAs with Dicer. Solid lines represent best fits to a one site specific binding equation. **c)** Histogram quantifying the Dicer processing efficiencies of pre-miR-31 RNAs at 10 min. **d)** Secondary structures of constructs designed to extend the pre-miR-31 apical loop. Insertions are indicated with red lettering. **e)** Quantification of the binding affinity of pre-miR-31 RNAs with Dicer. Solid lines represent best fits to a one site specific binding equation. **f)** Histogram quantifying the Dicer processing efficiencies of pre-miR-31 RNAs at 10 min. For all binding and processing assays, average and standard deviation from n=3 independent assays are presented. Individual replicates shown with black circles.



**Appendix B Figure. 3. Extending the length of the helical junction between the apical loop and dicing site reduces processing efficiency and accuracy. a)** Secondary structure and Dicer processing gel of WT pre-miR-31. **b)** Secondary structure and Dicer processing gel of pre-miR-31 G32C/A33C. Two mature products are detected (black arrows).

In addition to enhanced cleavage efficiency, cleavage accuracy is also affected by the loop position. Extension of the helical region between the dicing site and the apical loop results in the generation of mature products of varying lengths. In the G32C/A33C mutant, which shifts the loop position 2-nt up relative to WT, we detected two mature product bands, while for WT, only 1 mature product was observed (**Fig. 3**). We conclude that for pre-miRs, loop size can control Dicer processing efficiency in a bidirectional way. Furthermore, we show that the position of the loop relative to the dicer processing site is essential for accurate and efficient cleavage of Dicer, consistent with the previously described loop counting rule[5].

### Junction residues function as critical control elements for Dicer processing

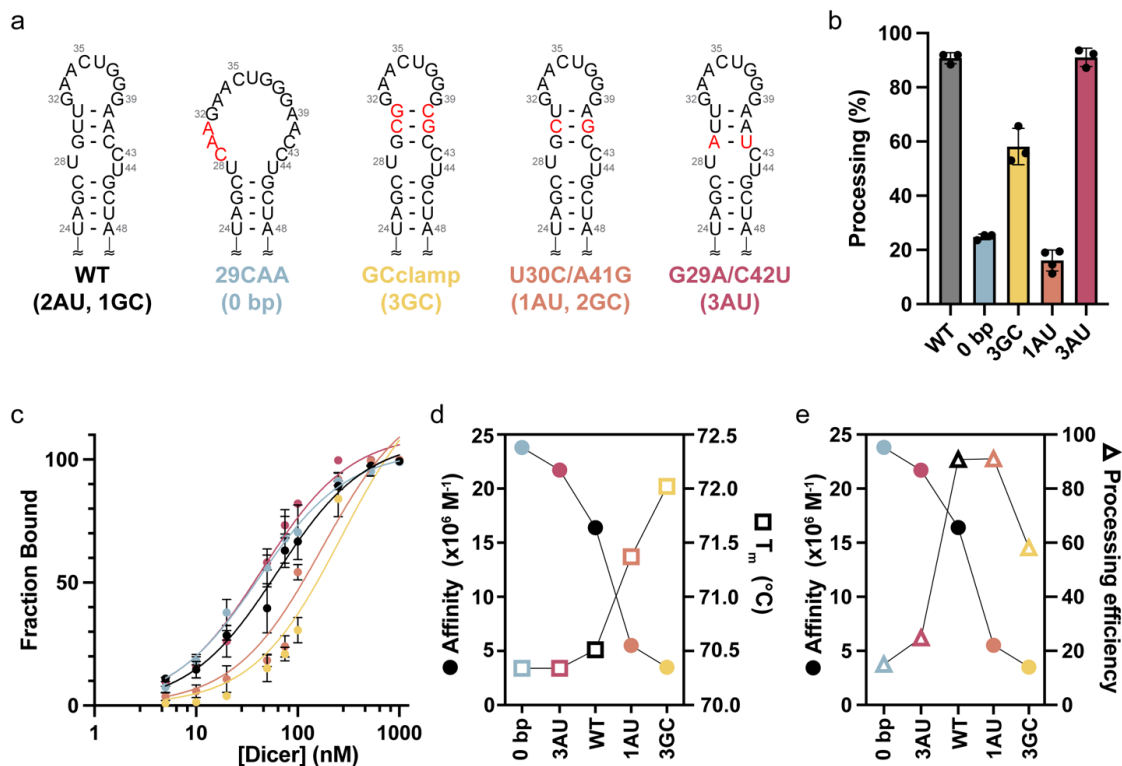
Our NMR-derived structure of FL pre-miR-31 revealed the presence of three base pairs in a junction region between the apical loop and the dicer cleavage site (**Fig. 3.1b**). However, *in cell* chemical probing studies revealed that junction residues were highly reactive, suggesting that these base pairs are absent in the presence of Dicer[13]. The high reactivity of these nucleotides *in cell* is consistent with our *in vitro* chemical probing studies (**Fig. 3.1a**) which suggest that pre-miR-31 has a large apical loop region. To resolve these conflicting models, we designed constructs which stabilized or destabilized the junction residues and examined their Dicer binding affinity and Dicer

cleavage efficiency.

To mimic the large open loop structure detected by chemical probing, we mutated residues G29, U30, and U31 to prevent base pairing in the junction region (29CAA) (**Fig. 4a**). The processing data for 29CAA reveals that it is a poor substrate for Dicer processing, with only 15% of the 29CAA precursor converted to mature product (**Fig. 4b**). We also designed a construct to stabilize the junction region, as defined by NMR data. Here, the junction A-U base pairs were replaced with G-C base pairs (GCclamp, **Fig. 4a**). Interestingly, the GCclamp construct reduced the cleavage efficiency to ~ 60% (**Fig. 4b**). These data suggest that the stability of the base pairs within the junction region is an important determinant of Dicer processing.

To further elucidate how the junction stability of pre-miR-31 regulates Dicer processing, we designed two additional junction mutants with different base pairing compositions. The U30C/A41G construct (1AU base pair and 2 GC base pairs, **Fig. 4a**) is processed as efficiently as WT (2AU base pairs, 1 GC base pair) at 10 minutes (**Fig 4b**). Whereas the G29A/C42U mutant (3 AU base pairs, **Fig. 4a**) is processed with ~20% efficiency (**Fig. 4b**). These data suggest that the stability of the junction region is finely tuned to maximize dicer processing efficiency.

To better characterize the junction stability, we performed thermal denaturation experiments for these constructs. We found that 29CAA, G29A/C42U and WT pre-miR-31 had similar melting temperatures, consistent with a model in which they adopt a similar open loop structure. The observed melting temperature of U30C/A41G and GCclamp increased by 1 °C and 1.5 °C, respectively relative to WT pre-miR-31 (**Appendix A Fig S24, Appendix A Table S6**). The observed increase in melting temperature suggests that the base pairs in the junction region of these RNAs are more stable than WT.



**Appendix B Figure 4. The junction region is a regulatory element within pre-miR-31.** **a)** Secondary structures of constructs designed to perturb the stability of the pre-miR-31 junction region. Mutations are indicated with red lettering. **b)** Histogram quantifying the Dicer processing efficiencies of pre-miR-31 RNAs at 10 min. **c)** Quantification of the binding affinity of pre-miR-31 RNAs with Dicer. Solid lines represent best fits to a one site specific binding equation. **d)** Inverse correlation between calculated binding affinity and measured thermal stability (melting temperature,  $T_m$ ) for WT and junction region mutations. **e)** Correlation between Dicer binding affinity and Dicer processing efficiency for junction region mutations. For all binding and processing assays, average and standard deviation from  $n=3$  independent assays are presented. Individual replicates shown with black circles.

We show that 29CAA is poorly processed (**Fig. 4b**), however, this RNA adopts an open loop structure, consistent with the Dicer-bound structure identified *in cell*[13]. Therefore, we hypothesized that the open loop structure may contribute favorably to dicer binding. We found that 29CAA and G29A/C42U, which both have destabilized junction regions have similar binding affinities, which are slightly tighter than WT (**Fig. 4c, Table 1**). However, mutations that stabilized the junction region (GCclamp, U30C/A41G) exhibited weaker binding relative to wildtype (**Fig. 4c, Table 1**). Collectively, we observe an inverse relationship between junction stability ( $T_m$ ) and

binding affinity (**Fig. 4d**), consistent with a model in which the binding affinity between Dicer and the pre-miR substrate is determined by the structural stability at the junction.

This delicate balance of structural stability within the junction must be optimized to maximize both high affinity binding and efficient processing. WT pre-miR-31 is precisely tuned to maximize both binding affinity and processing efficiency (**Fig. 4e**). While U30C/A41G maintains high efficiency processing, the increased stabilization of the junction leads to an RNA with reduced binding affinity. Similarly, 29CAA, which has an open loop structure that promotes high affinity binding is poorly processed (**Fig. 4e**).

## **Discussion**

miRs play an important role in the post-transcriptional regulation of gene expression in cells. miRs are themselves subject to post-transcriptional regulation to ensure appropriate levels of the mature products are produced. Many proteins are known to post-transcriptionally regulate miR biogenesis at either the Drosha and/or Dicer processing steps [14-16]. While protein-mediated regulation of miR biogenesis can be an important mechanism of control, the intrinsic structural features of pri/pre-miRs can also regulate the enzymatic processing of miRs[16-18]. In fact, in a recent proteomics screen, pre-miR-31 is one of two human miRs (72 miRs examined) with no identified protein binding partners[19]. Therefore, we were interested in uncovering the RNA-mediated mechanisms regulating miR-31 biogenesis. To better understand the structural basis for processing, we solved the high-resolution tertiary structure of pre-miR-31. Our structural and biochemical studies provide a framework for optimized design of shRNAs and elucidate distinct mechanisms by which RNA structure helps to regulate Dicer-mediated processing of pre-miR-31 (**Fig. 5**). We found that the presence of mismatches within the pre-miR-31 stem, while a nearly

ubiquitous feature of pre-miRs, did not significantly influence the processing of pre-miR-31. We also showed that destabilizing the dicing site by introduction of a larger internal loop inhibited processing of pre-miR-31. Furthermore, we show that apical loop size controls Dicer processing in a bidirectional manner. Finally, we provide strong evidence that stability of pre-miR-31 junction region serves as a potent regulatory factor for Dicer binding and processing.

Our structure reveals that pre-miR-31 adopts an elongated A-helical structure with three mismatches within the stem region. Both the A•A and G•A mismatches are stacked with their flanking nucleobases. The C•A mismatch is less well-defined. A54 appears to participate in A-helical stacking while C18 samples many conformations. The dicing site is marked by a highly ordered 1x2 internal loop and is linked to the 8-nt apical loop by a 3 base pair junction region.

We previously showed that A54 has an elevated  $pK_a$  and that a C•A<sup>+</sup> mismatch within pre-miR-31 can form at near neutral pH[1]. A similar pH-regulated conformational switch near the Dicer cleavage site in pre-miR-21 was shown to regulate Dicer processing[17]. However, in pre-miR-31, we found that formation of a base pair at the mismatch does not regulate Dicer processing. In fact, our processing assays show that mutations designed to either stabilize or destabilize the stem mismatches have no effect on Dicer processing. Although the pH-sensitive mismatch within the stem of pre-miR-31 had no effect on Dicer recognition and processing, this and other mismatches may help to regulate Drosha processing[20].

Previous studies show the importance of secondary structure at the dicing site for Dicer cleavage of shRNA and some pre-miRs.[12] Here, we show that substitution to form a 1x1 internal loop at the Dicing site makes itself a slightly better pre-miR substrate for Dicer processing than the native 1x2 internal loop or fully base paired structure at dicing site. However, increasing the internal loop size negatively impacted Dicer processing. Interestingly, we also found that

minimizing or eliminating the internal loop at the dicing site promotes 5' strand cleavage by Dicer and effectively eliminates the partially processed intermediate, converting all processed pre-miR to the mature product.

Both apical loop size and position contribute to the regulation of Dicer and Drosha processing[2-5]. Our findings re-emphasized the efficiency control by loop size and efficiency/accuracy control by loop position and provide new insights. Previous studies demonstrate that the presence of a small apical loop inhibits Dicer cleavage[3, 4]. We showed not only that a small apical loop inhibits Dicer processing, but also that large apical loops negatively regulate Dicer processing efficiency. We attribute at least a portion of the reduced processing to the weaker binding to Dicer of pre-miRs with small apical loops. We show that as the distance between the cleavage site and the apical loop increases, the processing accuracy decreases. Furthermore, we found that inclusion of a two base pair spacer between the dicing site and the apical loop compensates for the cleavage inhibition caused by a larger apical loop. These findings further validated the loop counting rule[5] in which Dicer has a higher processing efficiency and accuracy when the dicing site is positioned two base pairs below the apical or an internal loop. Our study reveals that loop size is one property that should be optimized when designing shRNAs where large apical loops can reduce Dicer cleavage.

Importantly, we found that the stability of junction region of pre-miR-31 is an inherent regulatory mechanism. Our NMR-derived secondary structure stands in contrast to one revealed by both *in cell* chemical probing[13] and our own *in vitro* chemical probing studies. Secondary structures reported based on chemical probing adopt a large apical loop region, where the junction residues are not engaged in base pairing. We believe that the differences in the NMR and chemical probing derived structures reflect the likely dynamic nature of the base pairs in the junction region,



information which can be obstructed in the chemical probing studies. Early chemical probing studies[21, 22] suggest that in the cell, RNAs are generally less folded than *in vitro*. Consistent with this hypothesis, recent *in cell* selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) chemical probing studies revealed that the apical loops of pre-miRs are less structured than predicted in the miRbase.[13, 23-28] Our structural data are consistent with a model in which base pairs in the junction region are very accessible to the solvent and thus more prone to open, so we believe that both an open and cinched junction region exist in a dynamic equilibrium.

We imagine that these two different pre-miR-31 structures both exist and promote distinct favorable interactions with Dicer. We therefore sought to determine the different contributions from the open loop and cinched junction structures. We first examined mutations designed to stabilize the junction region, favoring a cinched junction, consistent with the NMR-derived structure. We found that mutations which stabilized the junction region reduced Dicer binding affinity yet maintained Dicer cleavage. Conversely, we show that mutations which destabilized the junction region, promoting an open apical loop structure, promote binding to Dicer yet inhibit processing. The open apical loop structure sequesters the Dicer cleavage sites in the loop, which may account for the reduced processing efficiency. Collectively, we found that the stability of the pre-miR-31 junction region is optimized to sample both open and cinched conformations to promote both high affinity binding and high efficiency processing. These findings enrich the understanding of how distinct conformations of pre-miR-31 contribute to Dicer binding and processing.

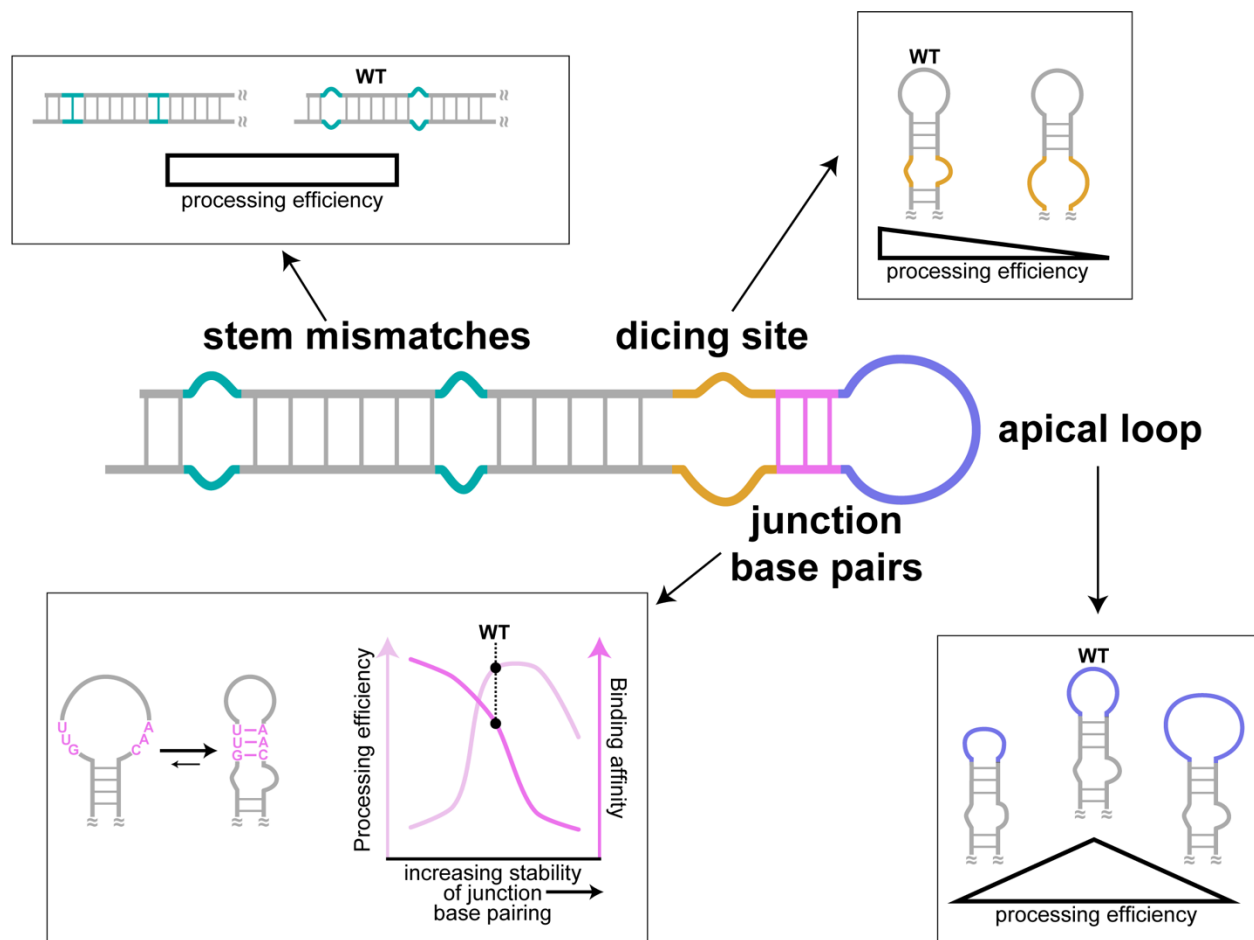
Our newly resolved 3D structure of pre-miR-31 in its processing-competent conformation and elucidation of its intrinsic regulatory mechanism informs on the important role that pre-miR apical loop plasticity plays in controlling Dicer processing. Our structural and biochemical studies

are consistent with proposed models of pre-miR processing based on cryo-EM structures of human Dicer[29] and fly Dicer-1[30] bound with pre-miRs. The pre-let-7 bound human Dicer structure revealed that the pre-let-7 RNA adopts multiple conformations[29]. In the “pre-dicing state,” Wang and co-workers posit that the pre-let-7 RNA first binds before the structure is adjusted to form a more stable stem[29]. This hypothesis is consistent with our findings that the pre-miR-31 large apical loop structure is the preferred substrate for Dicer binding, but that the structure with a cinched junction region is a “dicing-competent” structure. The recent cryo-EM structures of fly Dicer-1 reveal further details of the Dicer-1-pre-miR structure in the “Dicing” state[30]. In the “Dicing” structure, the dicing activity of Dicer-1 is inhibited by replacing  $Mg^{2+}$  with  $Ca^{2+}$ . The structure reveals that the pre-miR is highly structured in the “Dicing” state, with the Dicing site sequestered in an A-form helical structure and several base pairs present above the Dicing site. This “Dicing” structure is consistent with our NMR-derived structure, where the stabilization of additional base pairs in the apical loop promotes formation of an extended A-helical structure above the dicing site. Our data suggest that pre-miR-31 is “pre-structured” for Dicer processing. Further structural studies will be necessary to fully-characterize the structural changes in both the pre-miR and Dicer throughout the catalytic cycle.

**Appendix B Table 1. Dicer binding affinity of pre-miR-31 mutations.**

<b>Region mutated</b>	<b>RNA constructs</b>	<b>Binding affinity (x10<sup>6</sup> M<sup>-1</sup>)<sup>a</sup></b>
-	pre-miR-31 WT	17 ± 2
Stem mutations	pre-miR-31 G14U	19 ± 3
	pre-miR-31 C18U	13 ± 1
	pre-miR-31 A54G	23 ± 4
	pre-miR-31 G14U/A54G	6 ± 1
	pre-miR-31 18ACsw	36 ± 6
Dicing site mutations	pre-miR-31 Δ43C	11 ± 2
	pre-miR-31 Δ43/U44A	7 ± 1
	pre-miR-31 G45C	9 ± 2
	pre-miR-31 G45C/C46G	3 ± 1
Apical loop mutations	pre-miR-31 G32C	5 ± 1
	pre-miR-31 G32C/A33C	3 ± 1
	pre-miR-31 AP+2	8 ± 2
	pre-miR-31 AP+5	11 ± 3
	pre-miR-31 AP+9	11 ± 3
	pre-miR-31 40UUG	18 ± 2
Junction mutations	pre-miR-31 29CAA	25 ± 5
	pre-miR-31 3GCclamp	4 ± 1
	pre-miR-31 U30C A41G	6 ± 2
	pre-miR-31 G29A C42U	22 ± 3

<sup>a</sup> Average and standard deviation from n=3 independent assays are presented.



**Appendix B Figure 5. Secondary structure elements and their contribution to the regulation of pre-miR-31 processing.** The presence or absence of mismatches within the stem of pre-miR-31 had no impact on Dicer processing. More highly stabilized Dicing sites were processed as efficiently as the WT sequence, but pre-miRs with larger internal loops were not processed efficiently. Similarly, pre-miRs with either too small or too large apical loops were processed less efficiently than WT pre-miR-31. Interestingly, the WT pre-miR-31 has an inherently encoded structural switch at the junction region. Pre-miR-31 appears to sample both an open loop structure, which favors binding, and a closed loop structure, which promotes processing. This allows WT pre-miR-31 to maximize both binding with and processing by Dicer.

## References

1. Kotar, A., S. Ma, and S.C. Keane, *pH dependence of C•A, G•A and A•A mismatches in the stem of precursor microRNA-31*. *Biophysical Chemistry*, 2022: p. 106763.
2. Zhang, X. and Y. Zeng, *The terminal loop region controls microRNA processing by Drosha and Dicer*. *Nucleic acids research*, 2010. **38**(21): p. 7689-7697.
3. Feng, Y., et al., *A comprehensive analysis of precursor microRNA cleavage by human Dicer*. *Rna*, 2012. **18**(11): p. 2083-2092.
4. Tsutsumi, A., et al., *Recognition of the pre-miRNA structure by Drosophila Dicer-1*. *Nature structural & molecular biology*, 2011. **18**(10): p. 1153-1158.
5. Gu, S., et al., *The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo*. *Cell*, 2012. **151**(4): p. 900-911.
6. Liu, Z., et al., *Structure of precursor microRNA's terminal loop regulates human Dicer's dicing activity by switching DExH/D domain*. *Protein & Cell*, 2015. **6**(3): p. 185-193.
7. Zeng, Y., R. Yi, and B.R. Cullen, *Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha*. *The EMBO journal*, 2005. **24**(1): p. 138-148.
8. Trabucchi, M., et al., *The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs*. *Nature*, 2009. **459**(7249): p. 1010-1014.
9. Costales, M.G., et al., *Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit*. *Journal of the American Chemical Society*, 2017. **139**(9): p. 3446-3455.
10. Bose, D., et al., *Selective inhibition of miR-21 by phage display screened peptide*. *Nucleic acids research*, 2015. **43**(8): p. 4342-4352.
11. Murata, A., et al., *BzDANP, a small-molecule modulator of pre-miR-29a maturation by*

- Dicer*. ACS Chemical Biology, 2016. **11**(10): p. 2790-2796.
12. Nguyen, T.D., et al., *Secondary structure RNA elements control the cleavage activity of DICER*. Nature communications, 2022. **13**(1): p. 1-16.
  13. Luo, Q.-J., et al., *RNA structure probing reveals the structural basis of Dicer binding and cleavage*. Nature communications, 2021. **12**(1): p. 1-12.
  14. Balzeau, J., et al., *The LIN28/let-7 pathway in cancer*. Frontiers in genetics, 2017. **8**: p. 31.
  15. Shang, R., et al., *Regulated dicing of pre-mir-144 via reshaping of its terminal loop*. Nucleic acids research, 2022. **50**(13): p. 7637-7654.
  16. Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis*. Nature reviews Molecular cell biology, 2014. **15**(8): p. 509-524.
  17. Baisden, J.T., et al., *Visualizing a protonated RNA state that modulates microRNA-21 maturation*. Nature chemical biology, 2021. **17**(1): p. 80-88.
  18. Shortridge, M.D., et al., *A slow dynamic RNA switch regulates processing of microRNA-21*. Journal of Molecular Biology, 2022. **434**(16): p. 167694.
  19. Treiber, T., et al., *A compendium of RNA-binding proteins that regulate microRNA biogenesis*. Molecular cell, 2017. **66**(2): p. 270-284. e13.
  20. Li, S., et al., *Mismatched and wobble base pairs govern primary microRNA processing by human Microprocessor*. Nature communications, 2020. **11**(1): p. 1-17.
  21. Rouskin, S., et al., *Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo*. Nature, 2014. **505**(7485): p. 701-705.
  22. Spitale, R.C., et al., *Structural imprints in vivo decode RNA regulatory mechanisms*. Nature, 2015. **519**(7544): p. 486-490.
  23. Kozomara, A., M. Birgaoanu, and S. Griffiths-Jones, *miRBase: from microRNA sequences*

*to function*. Nucleic acids research, 2019. **47**(D1): p. D155-D162.

24. Kozomara, A. and S. Griffiths-Jones, *miRBase: annotating high confidence microRNAs using deep sequencing data*. Nucleic acids research, 2014. **42**(D1): p. D68-D73.

25. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data*. Nucleic acids research, 2010. **39**(suppl\_1): p. D152-D157.

26. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics*. Nucleic acids research, 2007. **36**(suppl\_1): p. D154-D158.

27. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature*. Nucleic acids research, 2006. **34**(suppl\_1): p. D140-D144.

28. Griffiths-Jones, S., *The microRNA registry*. Nucleic acids research, 2004. **32**(suppl\_1): p. D109-D111.

29. Liu, Z., et al., *Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate*. Cell, 2018. **173**(5): p. 1191-1203. e12.

30. Jouravleva, K., et al., *Structural Basis of MicroRNA Biogenesis by Dicer-1 and Its Partner Protein Loqs-PB*. bioRxiv, 2022.