

https://dspace.org/

Optimisation of expressed RNA interference mimics using predicted stem length

Van den Berg, Fiona Taylor

2016

http://hdl.handle.net/10539/21678

Van den Berg, Fiona Taylor (2016) Optimisation of expressed RNA interference mimics using predicted stem length, University of Witwatersrand, Johannesburg, <http://wiredspace.wits.ac.za/handle/10539/21678>

Downloaded from a DSpace repository. https://dspace.org/

OPTIMISATION OF EXPRESSED RNA INTERFERENCE MIMICS USING PREDICTED STEM LENGTH

Fiona Taylor van den Berg

A Thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of
Doctor of Philosophy.

September 2016, Johannesburg

DECLARATION

I, Fiona Taylor van den Berg, declare that this Thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



06 September 2016, Johannesburg.

ABSTRACT

Primary microRNA (pri-miRNA) mimics have been shown to mediate effective gene silencing and are well-suited for therapeutic applications. Pri-miRNA mimics, like natural pri-miRNAs, are processed in the endogenous microRNA (miRNA) biogenesis pathway. Elements of the secondary RNA structure are crucial for processing by the Drosha-DGCR8 microprocessor, including a basal stem of ~ 11 bp. However, structural variation is common and the exact determinants of pri-miRNA processing have been elusive. The aims of this project were to explore the use of natural pri-miRNAs with exceptional basal stems in the design of corresponding mimics and to identify optimal stem features. Natural pri-miRNAs with various basal stem lengths (5 - 15 bp) were exogenously expressed and characterised. Secondary RNA structure was assessed using predictive software (mfold) and pri-miRNA function was assessed in cell culture through target silencing assays and northern blot analyses. The results indicated that miRNAs could be effectively derived from pri-miRNAs with sub-optimal basal stems (9 - 13 bp), but not from those with shorter basal stems (5 - 7 bp) containing large internal bulges. Potentially disruptive conformations predicted for the expressed pri-miRNA sequence were not restrictive. Functional pri-miRNAs with various basal stem conformations were therefore used in the generation of therapeutic mimics targeting HIV-1 sequences. In contrast to the natural pri-miRNAs, not all corresponding mimics were functional and certain pri-miRNA scaffolds and exogenous guides were generally more effective. Analysis of the guide region suggested the importance of thermodynamic asymmetry and indicated 5' sequence and stability preferences in agreement with previous reports. A more comprehensive, updated set of rules was proposed for the design of effective mimics based on natural pri-miRNAs. Overall, sub-optimal basal stem lengths appear to be tolerated by the microprocessor, but total stem length may be a more important parameter. In the design of pri-miRNA mimics, structural, stability and sequence features should be considered and may affect the efficiency of several downstream processing events. Several novel pri-miRNA scaffolds and effective pri-miRNA mimics were characterised here. This work has contributed to our current understanding of pri-miRNA processing and the optimisation of effective pri-miRNA mimics for future therapeutic applications.

PUBLICATIONS & PRESENTATIONS

JOURNAL ARTICLE

Fiona T. van den Berg, John J. Rossi, Patrick Arbuthnot and Marc S. Weinberg. (2016) Design of effective primary microRNA mimics with different basal stem conformations. Molecular Therapy - Nucleic Acids, 5, e278; doi:10.1038/mtna.2015.53; published online 12 January 2016. (PMID: 26756196).

BOOK CHAPTER

Fiona T. van den Berg and Marc S. Weinberg (2011). RNAi-Based Gene Expression Strategies to Combat HIV, Understanding HIV/AIDS Management and Care - Pandemic Approaches in the 21st Century, Fyson Hanania Kasenga (Ed.), ISBN: 978-953-307-603-4, InTech, doi: 10.5772/21340.

CONFERENCE PRESENTATIONS

1) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. Optimisation of primary microRNA secondary structures for effective production of anti-HIV guides. Faculty of Health Sciences Biennial Research Day, 17 September 2014, University of the Witwatersrand, Johannesburg, South Africa.

2) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. Optimisation of primary microRNA secondary structures for effective production of anti-HIV guides. South African Society of Biochemistry & Molecular Biology (SASBMB) Biennial Congress, 6 - 9 July 2014, Rawsonville, South Africa.

3) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. Secondary RNA structure optimisation as a tool in processing of artificial ncRNAs from pri-miRNA precursors. Molecular Biosciences Research Thrust (MBRT) Annual Meeting, 5 December 2012, University of the Witwatersrand, Johannesburg, South Africa.

4) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. Secondary RNA structure optimisation as a tool in processing of artificial ncRNAs from pri-miRNA precursors.

Keystone Symposia 2012: Non-coding RNAs Meeting, 31 March – 5 April 2012, Snowbird, Utah, USA.

5) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. Deriving effective anti-HIV guide sequences from optimised primary microRNA secondary structures. South African Society of Biochemistry & Molecular Biology (SASBMB) Biennial Congress, 29 January – 1 February 2012; Drakensburg, Kwa-Zulu Natal.

6) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. The development of effective pri-miRNA mimics against HIV. 3rd Postgraduate Cross Faculty Symposium, 25 - 29 October 2010, University of the Witwatersrand, Johannesburg, South Africa.

7) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. The development of effective primary-microRNA mimics against HIV. Faculty of Health Sciences Biennial Research Day, 22 September 2010, University of the Witwatersrand, Johannesburg, South Africa

8) Fiona van den Berg, Patrick Arbuthnot, Marc Weinberg. Deriving effective anti-HIV guide sequences from enhanced primary microRNA secondary structures. ESF-EMBO Research Symposium on Antiviral Applications of RNA interference, 30 May – 04 June 2010; Sant Feliu de Guixols, Spain.

9) Fiona T. van den Berg, Patrick Arbuthnot, Marc Weinberg. Deriving effective anti-HIV guide sequences from enhanced primary microRNA secondary structures. American Society of Gene & Cell Therapy: 13th Annual Meeting, 19 – 22 May 2010; Washington D.C., USA.

10) Fiona T. van den Berg, Patrick Arbuthnot & Marc Weinberg. Lecture: Effective target knockdown and processing can be achieved by primary-microRNAs spanning a variety of predicted stem lengths. South African Society of Biochemistry & Molecular Biology (SASBMB) Biennial Congress, 18 – 20 January 2010; Ilanga Estate, Bloemfontein, South Africa.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Professor Marco Weinberg for his support and enthusiasm. Thank you for encouraging me to strive for academic excellence and introducing me to the world of scientific research. I am truly grateful for all of the amazing opportunities that I have had under your supervision, including the chance to present my work at local and international conferences, visit other laboratories and publish my findings.

I wish to thank the head of the Anti-viral Gene Therapy Research Unit (AGTRU) Professor Patrick Arbuthnot for his support, advice and words of encouragement. It has been a privilege to work in your laboratory. I am truly grateful for all of the amazing opportunities that I have had as a member of AGTRU.

I wish to thank all of my funding bodies. This PhD research project would not have been possible without their financial support. I gratefully acknowledge support from the University of the Witwatersrand, the Andrew W. Mellon Foundation, the Poliomyelitis Research Foundation (PRF), the National Research Foundation (NRF), the Department of Science and Technology (DST), the Health Sciences Faculty Research Committee (FRC), the RNA Therapeutics Research Group, the Professional Provident Society (PPS) and the Freda Lawenski Foundation.

Thank you to Professor John Rossi, Dr. Jiehua Zhou and Dr. Xiwei Wu of the Beckman Research Institute of the City of Hope in Duarte, California; and Dr. Jonathan Hart of the Scripps Research Institute in San Diego, California for technical assistance with the generation and analysis of deep sequencing data.

To Dr. Kevin Morris of the Scripps Research Institute in San Diego, California, thank you for hosting me in your laboratory as part of the Fogarty academic exchange programme. It was such a valuable and inspirational experience.

To all members of the AGTRU lab past and present, thank you for keeping me company in this journey and dispensing pearls of wisdom along the way. A special thanks to Sheena Saayman, Neli Nhlabatsi and Samantha Barichievy for anti-viral control constructs and viral reporter plasmids. Thank you to Mrs. Gladys Gagliardi for all of the administrative assistance and the AGTRU cleaning staff.

To my parents, Judith and Dammes van den Berg, thank you for all of your love and boundless support. Your efforts and sacrifices towards my education have gone a long way and will always be remembered. I would not be at this juncture without you.

Last, but not least, to Louis du Pisani, thank you for your understanding and patience at the many twists and turns of this process. I am grateful for the support of you and your family.

In the words of Chuck Palahniuk,
“the future you have tomorrow won’t be the same future you had yesterday”

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
PUBLICATIONS & PRESENTATIONS iv
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS viii
LIST OF FIGURES xii
LIST OF TABLES xiv
LIST OF COMMON ABBREVIATIONS xv

CHAPTER 1

1.1 OVERVIEW 1
1.2 RNA INTERFERENCE (RNAi) 1
1.2.1 Discovery of miRNAs 2
1.2.2 MiRNA nomenclature 3
1.3 THE MAMMALIAN MIRNA BIOGENESIS PATHWAY 4
1.3.1 Organisation and expression of miRNA genes 4
1.3.2 Pri-miRNA processing 6
1.3.3 Pri-miRNA structural determinants 6
1.3.4 Pri-miRNA sequence determinants 10
1.3.5 A model of the Drosha-DGCR8 complex 11
1.3.6 Pre-miRNA export 17
1.3.7 Pre-miRNA processing 18
1.3.8 The RNA-induced silencing complex (RISC) 19
1.3.9 Target recognition & specificity 22
1.3.10 IsomiRs 23
1.3.11 Structure of miRNA processing enzymes 24
1.3.12 Regulation of miRNA biogenesis 32

1.3.13 Relationship of pri-miRNA processing to transcription and splicing 36
1.3.14 Alternative miRNA biogenesis 37
1.3.15 Other small regulatory RNAs 40
1.3.16 Transcriptional Gene Silencing (TGS) 43
1.3.17 The role of mature miRNAs 44
1.4 RNAi AS A THERAPEUTIC TOOL 45
1.4.1 RNAi to target HIV 46
1.4.2 Small interfering RNAs (siRNAs) and artificial miRNAs 47
1.4.3 Short hairpin RNAs (shRNAs) 50
1.4.4 Pri-miRNA mimics 51
1.4.5 Combinatorial RNAi 53
1.4.6 Safety concerns of an RNAi modality 55
1.5 THESIS OBJECTIVES 61

CHAPTER 2

2.1 INTRODUCTION 62
2.2 MATERIALS & METHODS 64
2.2.1 Sequence and secondary RNA structure of selected pri-miRNAs 64
2.2.2 Generation of pri-miRNA expression plasmids 65
2.2.3 Generation of shRNA expression plasmids 72
2.2.4 Generation of target reporter plasmids 75
2.2.5 Assessment of pri-miRNA function in cultured mammalian cells 78
2.2.6 Assessment of pri-miRNA processing in cultured mammalian cells 81
2.2.7 Analysis of pri-miRNA structures *in vitro* 85
2.2.8 Statistical analyses 91
2.3 RESULTS 92
2.3.1 Selection of naturally occurring pri-miRNAs with exceptional basal stems 92

2.3.2 Predicted secondary RNA structures of selected pri-miRNAs..... 92
2.3.3 Functional characterisation of pri-miRNAs with different basal stem conformations 97
2.3.4 Processing of miRNA guides from pri-miRNAs with different basal stem conformations..... 103
2.3.5 Potential pri-miRNA sequence motifs..... 106
2.3.6 RNase mapping of functional and non-functional pri-miRNA structures *in vitro*..... 109
2.4 DISCUSSION..... 114

CHAPTER 3

3.1 INTRODUCTION..... 118
3.2 MATERIALS & METHODS..... 120
3.2.1 Generation of anti-HIV pri-miRNA mimic expression plasmids..... 120
3.2.2 Short hairpin RNA expression plasmids..... 123
3.2.3 Target reporter plasmids..... 124
3.2.4 Assessment of pri-miRNA mimic function in cultured mammalian cells..... 125
3.2.5 Assessment of pri-miRNA mimic processing in cultured mammalian cells..... 125
3.2.6 Analysis of thermodynamic properties of predicted pri-miRNA mimic structures... 126
3.2.7 Analysis of thermodynamic properties of predicted miRNA duplexes..... 127
3.2.8 Statistical analyses..... 127
3.3. RESULTS..... 128
3.3.1 Design of pri-miRNA mimics..... 128
3.3.2 Functional characterisation of pri-miRNA mimics..... 134
3.3.3 Processing of exogenous RNA guides from pri-miRNA mimics..... 138
3.3.4 Deep sequencing of processed RNA guides..... 140
3.3.5 Guide region stability of pri-miRNA mimics..... 143
3.3.6 Stability of expected miRNA duplexes..... 153
3.4 DISCUSSION..... 156

CHAPTER 4

4.1 Summary of results 164

4.2 The current model of pri-miRNA processing by Drosha-DGCR8 165

4.3 Modular processing of pri-miRNAs 169

4.4 Alternative approaches to pri-miRNA mimic design..... 170

4.5 Alternative RNA and combinatorial constructs..... 173

4.6 Delivery of expressed RNA constructs..... 175

4.7 Pre-clinical and clinical studies of expressed RNA constructs 177

4.8 Genome-editing: a new approach to gene therapy..... 179

4.9 Concluding remarks 180

REFERENCES..... 182

APPENDIX A..... 212

APPENDIX B..... 225

APPENDIX C..... 233

LIST OF FIGURES

Figure 1.1: The mammalian miRNA biogenesis pathway 5

Figure 1.2: Pri-miRNA secondary RNA structure and sequence motifs. 7

Figure 1.3: Proposed models for pri-miRNA recognition and the microprocessor. 14

Figure 1.4: Domain organisation of Drosha-DGCR8, Dicer and Argonaute proteins 26

Figure 1.5: RNAi to target the HIV genome. 47

Figure 1.6: Non-expressed and expressed therapeutic constructs. 48

Figure 2.1: PCR and cloning scheme for the generation of pCI-neo-pri-miRNA expression vectors 66

Figure 2.2: PCR and cloning scheme for the generation of shRNA expression vectors 73

Figure 2.3: Cloning scheme for the generation of dual luciferase target vectors 76

Figure 2.4: Dual luciferase expression system for the detection of specific RNAi activity 80

Figure 2.5: The identification of pri-miRNAs with exceptional stem conformations. 93

Figure 2.6: Predicted secondary RNA structures of pri-miRNAs with exceptional basal stem conformations. 95

Figure 2.7: Silencing efficacy of pri-miRNAs. 98

Figure 2.8: Silencing efficacy of pri-miRNAs in other cell lines. 99

Figure 2.9: Silencing efficacy of pri-miRNAs generated from an SV40 expression cassette. .. 101

Figure 2.10: Potential alternative processing of a pri-miRNA with a longer stem. 102

Figure 2.11: Processing of major and minor miRNA guides. 104

Figure 2.12: Potential primary sequence motifs. 107

Figure 2.13: RNase structure map of pri-miR-198. 111

Figure 2.14: RNase structure map of pri-miR-34a. 113

Figure 3.1: ShRNAs used as positive controls for exogenous RNA guides. 123

Figure 3.2: Design and predicted secondary RNA structures of pri-miRNA mimics. 129

Figure 3.3: The proportion of canonical and non-canonical predicted structures. 130

Figure 3.4: Silencing efficacy of pri-miRNA mimics. 135

Figure 3.5: Expected guide-target interactions. 137

Figure 3.6: Processing of exogenous RNA guides from pri-miRNA mimics. 139

Figure 3.7: Precise sequences of processed miRNAs and exogenous RNA guides. 142

Figure 3.8: Predicted stability of the guide region. 144

Figure 3.9: Average thermodynamic profiles of the guide region. 146

Figure 3.10: Average thermodynamic profiles of the guide region. 147

Figure 3.11: Average thermodynamic profiles of the guide region. 148

Figure 3.12: Base pairing profiles of the guide region. 150

Figure 3.13: Guide region stability of individual pri-miRNAs and mimics. 152

Figure 3.14: Terminus stability of expected miRNA duplexes. 154

Figure 3.15: Desirable features for pri-miRNA mimics and therapeutic constructs. 162

Figure 4.1: A comprehensive model of pri-miRNA recognition and cleavage. 166

LIST OF TABLES

Table 2.1: Genomic co-ordinates used to extend miRBase precursor sequences. 64

Table 2.2: Primers for the amplification of pri-miRNAs from genomic DNA 67

Table 2.3: Extension primers for the generation of pri-miRNAs by a two-step PCR 68

Table 2.4: Primers used for DNA sequencing 70

Table 2.5: Primers for the generation of U6-driven shRNAs by a two-step PCR 74

Table 2.6: Oligonucleotides for generation of a target duplex 77

Table 2.7: Oligonucleotide probes for detection of guides in northern blot analysis. 83

Table 2.8: The basal stem length (BSL) and free energy (dG) of predicted secondary RNA structures of pri-miRNAs. 96

Table 2.9: The basal stem length (BSL) and free energy (dG) of predicted secondary RNA structures of expected transcripts. 110

Table 3.1: Anti-HIV guide sequences 120

Table 3.2: Extension primers for the generation of pri-miRNA mimics by a two-step PCR 121

Table 3.3: HIV target sequences 124

Table 3.4: LNA^{TM}-modified probes for detection of guides in northern blot analysis. 125

Table 3.5: The basal stem length (BSL) and free energy (dG) of predicted secondary RNA structures of pri-miRNA mimics. 131

LIST OF COMMON ABBREVIATIONS

(HA)ART	(highly active) antiretroviral therapy
AAV	adeno-associated virus
Ad	adenovirus
Ago	Argonaute
AIDS	acquired immune deficiency syndrome
ATP	adenosine triphosphate
bp	base pair
CCR5	C-C motif receptor 5
CD4	cluster of differentiation 4
CXCR4	C-X-C motif receptor 4
Da	Dalton
DGCR8	DiGeorge syndrome chromosomal / critical region 8
DNA	deoxyribonucleic acid
Exp-5	Exportin-5
Gag	group specific antigen
GTP	guanosine triphosphate
HIV-1	human immunodeficiency virus
HSC	haematopoietic stem cell
Int	integrase
kb	kilobases
lhRNA	long hairpin RNA
LTR	long terminal repeat
METTL3	methyltransferase-like 3
miRNA	microRNA
mRNA	messanger RNA
nt	nucleotide
PACT	protein activator of PKR
pDNA	plasmid DNA
piRNA	PIWI interacting RNA
PKR	protein kinase R
pol II / III	RNA polymerase II / III
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA
PTGS	post-transcriptional gene silencing
RBP	RNA-binding protein
Rev	regulator of expression of virion proteins
Rhed	RNA-binding haem domain
RISC	RNA-induced silencing complex
RLC	RISC loading complex
RNA	ribonucleic acid
RNAi	RNA interference
RNAse	ribonuclease
RPM	Reads Per Million
rpm	revolutions per minute
RRE	Rev response element
rRNA	ribosomal RNA
shRNA	short hairpin RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
ss / dsRNA	single / double-stranded RNA
TAR	trans-activation response element
Tat	trans-activator of transcription
TGS	transcriptional gene silencing
TRBP	TAR RNA-binding protein
tRNA	transfer RNA
UTR	untranslated region

CHAPTER ONE

INTRODUCTION:
RNA interference (RNAi) as a tool in post-transcriptional gene silencing

1.1 OVERVIEW

The discovery of RNA interference (RNAi) provided investigators with a tool for the potent and specific silencing of target genes. This enabled the development of novel therapeutic strategies against various diseases and pathogens. A promising approach for therapeutic gene silencing in humans relies on the endogenous RNAi pathway for microRNA (miRNA) biogenesis. This chapter discusses details of miRNA biogenesis and introduces various artificial mimics that are processed in this pathway to mediate therapeutic gene silencing. In particular, there is a focus on the sequence and structural determinants of primary miRNA (pri-miRNA) processing and the impact of these factors on pri-miRNA mimic design. An emerging model for the processing of typical pri-miRNAs is discussed, highlighting the need for further investigation into the processing of exceptional pri-miRNAs, with a potential benefit for mimic design. The need for a novel and improved therapeutic approach for the treatment of HIV-1 infection is also presented, with a consideration of the efficacy and safety of various RNAi activators for this purpose. In this review, pri-miRNA mimics emerge as a promising option for an effective, safe and flexible strategy for long-term, RNAi-based gene silencing applications. Accordingly, this chapter closes with specific objectives of the thesis, which are centred on the development of novel pri-miRNA mimics and the optimisation of mimic design using predicted stem length.

1.2 RNA INTERFERENCE (RNAi)

RNAi is a conserved eukaryotic post-transcriptional gene-silencing (PTGS) phenomenon that has been described in numerous organisms. This phenomenon was first described in plants, where the introduction of an exogenous transgene was found to mediate silencing of homologous genes and termed “co-suppression” (Napoli et al., 1990). Similar gene silencing in response to exogenous sequences was observed in the *Neurospora crassa* fungus and referred to as “quelling” (Romano and Macino, 1992). Gene silencing was also observed in *Caenorhabditis elegans* and attributed to anti-sense inhibition of endogenous messenger RNA

(mRNA) transcripts (Fire et al., 1991). However, silencing was also observed in response to sense RNA (Fire et al., 1991; Guo and Kemphues, 1995) and genetic interference was found to be persistent (Fire et al., 1991) and systemic (Voinnet and Baulcombe, 1997). Conflicting ideas surrounding the cause of gene silencing were resolved in a seminal paper by Fire and Mello (Fire et al., 1998), in which dsRNA was found to produce potent and specific RNAi in *C. elegans* and a catalytic or amplification mechanism was proposed. RNAi was subsequently described in other animals, from *Drosophila melanogaster* (Kennerdell and Carthew, 1998) to mice (Svoboda et al., 2000) and humans (Elbashir et al., 2001a). Gene silencing was found to be mediated by short interfering RNAs (siRNAs) of ~ 25 nt in plants (Hamilton and Baulcombe, 1999) and siRNAs of 21 - 23 nt in animals, derived from dsRNA by the action of RNAse III-type processing (Elbashir et al., 2001b; Zamore et al., 2000). SiRNA guide sequences direct cleavage of complementary mRNAs (Martinez et al., 2002a; Schwarz et al., 2002; Zamore et al., 2000) by associating with an RNA-induced silencing complex (Hammond et al., 2000). This mechanism has since been described as central to several endogenous pathways involving small RNAs, most notably, microRNAs (miRNAs) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001).

1.2.1 Discovery of miRNAs

MiRNAs were first described in the control of developmental timing in *C. elegans* and initially described as small temporal RNAs (stRNAs) (Pasquinelli et al., 2000). The Ambros and Ruvkun laboratories proposed that a 22 nt *lin-4* RNA regulated translation of the LIN-14 protein through antisense RNA-RNA interactions at complementary sites in the LIN-14 3'UTR (Lee et al., 1993; Wightman et al., 1993). It was also suggested that the 22 nt *lin-4* RNA product may be derived from a larger 61 nt transcript predicted to form a stem-loop or hairpin structure (Lee et al., 1993). A similar mode of PTGS was described for the 21 nt *let-7* RNA, which has complementary sites in the 3' UTR of LIN-41 (Reinhart et al., 2000; Slack et al., 2000). These early proposals of miRNA derivation and function are now known to be correct. Homologous let-7 RNA sequences and hairpin RNA precursors were identified in a number of other species, including vertebrates, ascidians, hemichordates, molluscs, annelids and arthropods and a link between stRNAs and RNAi was suggested (Pasquinelli et al., 2000). Other 21 - 24 nt small RNAs were also identified in *C. elegans, D. melanogaster* and humans, pointing to the existence of a larger class of small regulatory RNAs termed "microRNAs" (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). The subsequent discovery of other new miRNAs was rapid and miRNAs have since been shown to be ubiquitously expressed in numerous cell types and a range of

organisms from unicellular algae to more complex flowering plants and animals, as well as viruses. An early release of the miRBase database, a repository of published miRNA sequences, listed 506 miRNAs from 6 organisms (Griffiths-Jones, 2004), while the latest release now lists 24 521 miRNA loci from 206 species (Kozomara and Griffiths-Jones, 2014), with new entries from about 2007 onwards greatly facilitated by deep sequencing technologies. There are currently 427 mature miRNA sequences listed in *Arabidopsis thaliana*, 435 in *C. elegans*, 462 in *D. melanogaster*, 1 899 in *Mus musculus* and 2 585 in humans (Release 21).

1.2.2 MiRNA nomenclature

New, distinct miRNAs were named numerically according to the order in which they were described (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001), although the names of a few early miRNAs and homologues persist with prefixes that describe a mutant phenotype from genetic studies, such as “let-7” (lethal) in *C. elegans* (Horvitz et al., 1979). General naming conventions for miRNAs were outlined in the first release of the miRNA registry (Griffiths-Jones, 2004). The use of “miR” to describe both precursors and mature miRNAs in animals is currently preferred (Fang and Bartel, 2015; Nguyen et al., 2015) and used here. Typically, an additional three letter prefix denotes the relevant organism, where the first letter is that of the genus and the second and third letters are derived from the species; for example, miRNA-21 from *Homo sapiens* is “hsa-miR-21” and miRNA-1 from *C. elegans* is “cel-miR-1”. A numerical suffix is used to denote identical copies of a metazoan miRNA originating from different loci, for example, hsa-miR-124-1, hsa-miR-124-2 and hsa-miR-124-3. A letter suffix is used to denote very similar, related metazoan miRNAs that may differ by one or two bases, such as hsa-miR-147a and hsa-miR-147b. MiRNAs may be derived from either arm of a hairpin precursor, but the expression of one miRNA is typically more robust. Initially, an asterisk was used to denote the unexpressed or minor miRNA, but as both miRNAs may be functional the use of a “5p” or “3p” suffix to denote the origin of the miRNA from the respective 5’ or 3’ arm of the hairpin precursor is now preferred, for example, hsa-miR-21-5p and hsa-miR-21-3p (Kozomara and Griffiths-Jones, 2014). However, an asterisk may still be used in more general terms, such as “miRNA*”. As is the convention, the names of miRNA genes are italicised, while the names of mature miRNAs and precursors are not italicised.

1.3 THE MAMMALIAN MIRNA BIOGENESIS PATHWAY

Post-transcriptional regulation of gene expression by miRNAs is now acknowledged as a widespread phenomenon. Research over the past 15 years in *D. melanogaster* and mammalian cell culture has provided much of the mechanistic detail of miRNA biogenesis and gene silencing. An understanding of the mammalian miRNA biogenesis pathway (Figure 1.1) is therefore important and this pathway is discussed here in more detail. Similar biogenesis pathways in plants (Voinnet, 2009) and insects (Lucas and Raikhel, 2013) are reviewed elsewhere.

1.3.1 Organisation and expression of miRNA genes

MiRNAs are typically transcribed from corresponding miRNA genes by RNA pol II to produce transcripts with a characteristic 5' 7-methylguanylate (m7G) cap and 3' polyA tail (Cai et al., 2004; Lee et al., 2004). The location of miRNA genes can affect their mode of transcription and downstream processing. MiRNA genes are located in intergenic, intronic or, less frequently, exonic loci (Rodriguez et al., 2004). A survey of the genomic loci of human miRNAs available on miRBase (Griffiths-Jones et al., 2008) indicated that 62 % were located in introns (43 % coding, 19 % non-coding), 27 % in intergenic regions and 11 % in exons (8 % coding / UTRs, 3 % non-coding) (Morlando et al., 2008). The majority of miRNA genes are therefore located in introns and are frequently co-expressed with their host genes (Baskerville and Bartel, 2005; Liang et al., 2007). However, approximately one third of mammalian and human intronic miRNAs have associated promoter elements that drive transcription independently of the host gene (Corcoran et al., 2009; Monteys et al., 2010; Ozsolak et al., 2008). Intergenic miRNA genes are also expressed from independent transcription units, defined by 5' transcriptional start sites (TSS) and 3' polyadenylation (polyA) signals, as primary transcripts of 1 - 10 kb (Saini et al., 2008). Some miRNAs may be transcribed by RNA pol III, such as those associated with RNA pol III promoter elements in ~ 5 % of intronic miRNA genes (Monteys et al., 2010) or Alu repeat sequences of the human chromosome 19 miRNA cluster (C19MC) (Borchert et al., 2006). Although, for the latter example, more recent evidence has suggested that C19MC miRNAs are in fact derived from pol II-dependent, intron-encoded transcripts expressed in the placenta and pol III-dependent transcription is unlikely to be significant (Bortolin-Cavaille et al., 2009). Almost half of human miRNA genes are organised in clusters (Altuvia et al., 2005), composed of at least two miRNA genes, that are expected to be co-expressed as a single polycistronic transcript (Baskerville and Bartel, 2005; Liang et al., 2007). A polycistronic miRNA transcript contains

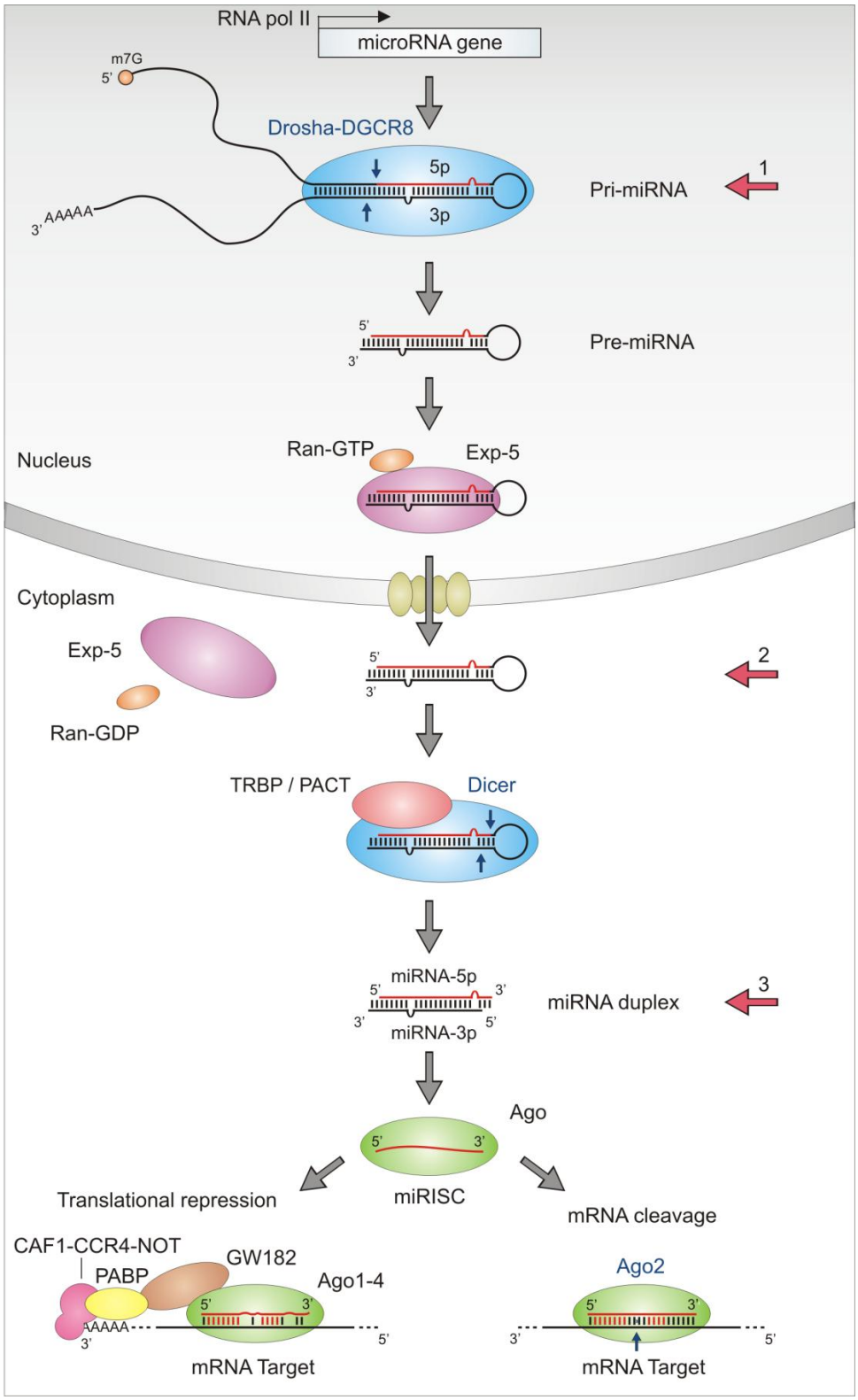


Figure 1.1: The mammalian miRNA biogenesis pathway

Figure 1.1: The mammalian miRNA biogenesis pathway
A schematic diagram of the basic steps in the miRNA biogenesis pathway. Briefly, miRNA genes are transcribed to form primary miRNAs (pri-miRNAs) which are cleaved by Drosha-DGCR8 in the nucleus. The resulting precursor miRNAs (pre-miRNAs) are exported to the cytoplasm by Exp-5 in a Ran-GTP-dependent manner through nuclear pore complexes. Dicer-TRBP cleaves the hairpin to produce a miRNA duplex, which is loaded onto an Ago protein and unwound to form the RNA induced silencing complex (RISC). RISC mediates translational repression or cleavage of a target mRNA depending on the level of complementary miRNA-target pairing. Only Ago2 retains "slicer" activity, while translational repression is mediated through GW182, the polyA binding protein (PABP) and the action of polyadenylase complexes (CAF1-CCR4-NOT). Therapeutic constructs may be introduced into the cell at various points in the pathway and processed accordingly (Red arrows; 1: pri-miRNA mimics; 2: shRNAs; 3: siRNAs, artificial miRNAs).

several adjacent hairpin precursors (Lau et al., 2001) and well-characterised examples include the miR-23a~27a~24-2, miR-106b~93~25 and miR-17~18~19a~20~19b-1~92-1 clusters (Lagos-Quintana et al., 2001; Mourelatos et al., 2002).

1.3.2 Pri-miRNA processing

In the conventional miRNA biogenesis pathway, mature miRNAs are derived from pri-miRNAs in two sequential enzymatic steps (Lee et al., 2002b) (Figure 1.1). In the first processing step in the nucleus, the pri-miRNA is cleaved by the Drosha-DGCR8 RNase III "microprocessor" complex (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004). Pri-miRNA transcripts fold into characteristic hairpin structures with a terminal loop, imperfect duplex stem and unstructured flanking sequences (Han et al., 2006) (Figure 1.2). In a typical pri-miRNA with a total stem length of ~ 3 helical RNA turns (~ 33 bp), the Drosha-DGCR8 cleavage site occurs at ~ 22 bp from the terminal loop and ~ 11 bp from the unstructured flanking sequences, thus defining the upper and lower (basal) stem regions (Han et al., 2006; Zeng et al., 2005b). This cleavage event produces a shorter ~ 65 nt precursor miRNA (pre-miRNA) hairpin with a 3' dinucleotide overhang and creates one end of the mature miRNA/s. MiRNAs can be embedded in the 5' (5p) and / or 3' (3p) arm of the upper stem, although usually only one miRNA species accumulates (Lau et al., 2001).

1.3.3 Pri-miRNA structural determinants

The characteristic secondary RNA structure of pri-miRNAs is important for recognition and cleavage by the microprocessor (Figure 1.2). However, endogenous pri-miRNAs have vast sequence heterogeneity and a great variety of secondary RNA structures have been predicted using RNA folding software (mfold) (Zuker, 2003). Several studies by the Cullen and Kim

research groups have investigated the effects of structural variation for model pri-miRNA hairpins, such as pri-miR-30a and pri-miR-16-1, in an effort to define general structural determinants for efficient processing. However, these studies have yielded some conflicting results, which are discussed in this section. The typical experimental approach in these studies was to introduce strategic mutations to the terminal loop, stem or flanking sequences and assess miRNA production and function through *in vitro* or cell culture assays. In a typical *in vitro* assay, pri-miRNAs were transcribed from an expression vector, radio-labelled and incubated with an immuno-precipitated, FLAG-tagged Drosha complex or a total extract from cells expressing FLAG-tagged Drosha (Lee et al., 2003b). In cell culture, pri-miRNA mutants were expressed in transfected cells and the function of relevant miRNA guides was assessed by target reporter assays, while miRNA levels were assessed using primer extension or northern blot analyses (Zeng and Cullen, 2003).

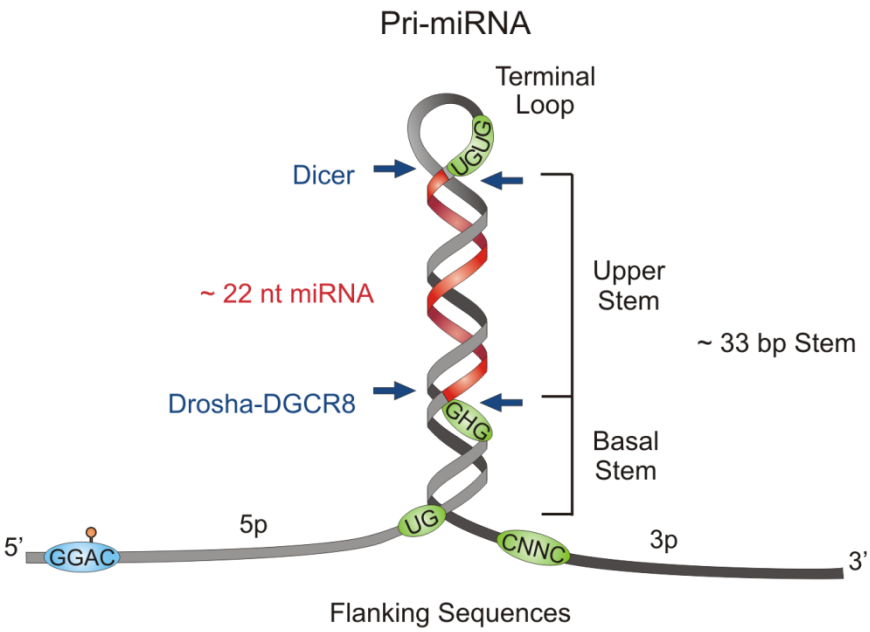


Figure 1.2: Pri-miRNA secondary RNA structure and sequence motifs.
The basic secondary RNA structure of a pri-miRNA depicting the terminal loop, ~ 33 bp imperfect stem and unstructured flanking regions, as described by (Han et al., 2006). Drosha-DGCR8 and Dicer cleavage sites are indicated (arrows) and define the ~ 22nt miRNA sequence (red), which may be located in the 5' (5p) or 3' (3p) arm. The stem extends over ~ 3 helical turns of the RNA duplex and is divided into an upper and lower (basal) region by the Drosha-DGCR8 cleavage site. Sequence motifs (green) that enhance Drosha-DGCR8 processing have recently been described, including 5p UG and 3p CNNC basal motifs, an apical UGUG motif and a mismatched GHG stem motif. Motif positions in the pri-miRNA helix are from the model by (Auyeung et al., 2013). A METTL3 motif (blue) may also occur within 100 nt of the pre-miRNA and enhance processing through N^6-methyladenosine marks (orange).

Mutations to the terminal loop yielded conflicting results in different studies. Mutations that reduced the terminal loop of pri-miR-30a from 15 nt to between 4 and 8 nt severely reduced miRNA production and function; while a terminal loop of 9 nt permitted moderate function and an unstructured loop of 12 nt with no predicted pairing interactions even appeared to improve miRNA production (Zeng and Cullen, 2003; Zeng et al., 2005b). Similar results were obtained for terminal loop mutations of pri-miR-21, pri-miR-31 and pri-miR-27a, indicating that a reduction in the size of the terminal loop was detrimental to pri-miRNA processing (Zeng et al., 2005b). Loop mutants created with unrelated sequences also generated similar results, suggesting that the size of the terminal loop was more important than a particular sequence. In contrast, mutations that prevented closure of the terminal loop of pri-miR-16-1 did not prevent Drosha cleavage at the intended site, suggesting that a terminal loop was dispensable for efficient processing (Han et al., 2006). The Drosha processing site was not even shifted by inverting the pri-miR-16-1 hairpin, although processing efficiency was reduced. Together, these results suggested varied importance of the terminal loop structure. In one case, the terminal loop was dispensable for processing, while in others, a large unstructured loop of ≥ 10 nt was required for efficient Drosha-DGCR8 processing (Zeng et al., 2005b). However, the latter finding appeared to be more generally applicable, as reducing the 10 nt terminal loop of pri-miR-16-1 to 4 nt or 6 nt also reduced the efficiency of Drosha processing (Han et al., 2006; Zhang and Zeng, 2010).

Studies investigating mutations to the basal stem and flanking sequences also yielded some conflicting results. At least 40 nt of flanking sequence on either side of the pre-miRNA was required for processing of pri-miRNAs from expression constructs in cell culture (Chen et al., 2004); but different flanking sequence requirements were identified for *in vitro* Drosha processing assays of pri-miR-30a, pri-miR-31 and pri-miR-223 mutants (Zeng and Cullen, 2005). Minimal 5p and 3p flanking sequences were indicated as follows: pri-miR-30a (22nt 5p, 15 nt 3p), pri-miR-31 (13 nt 5p, 16 nt 3p), pri-miR-223 (24 nt 5p, 21 nt 3p). A portion of these flanking sequences was expected to pair as the basal stem and corresponding minimum basal stem lengths were therefore 11 bp, 5 bp and 14 bp for pri-miR-30a, -31 and -223, respectively. The basal stem length preference of pri-miR-30a was investigated further in a second study (Zeng et al., 2005b), in which mature miRNAs were effectively derived from mutants with basal stems of 8 bp, 10 bp and 14 bp, but not from those of 3 bp or 18 bp. This suggested that exceptionally short or long basal stem lengths were detrimental to pri-miRNA function. Similarly, short basal stems (0 - 3 bp) were found to abolish pri-miRNA processing for pri-miR-30a (Lee et al., 2003b; Zeng and Cullen, 2003), pri-miR-21 (Zeng and Cullen, 2003) and pri-mR-31 (Zeng and Cullen, 2005) mutants in other studies; while a long basal stem of 23 bp drastically reduced the efficiency of Drosha processing for a pri-miR-16-1 mutant (Han et al., 2006). The absence of unstructured

sequences beyond the basal stem also abolished processing of pri-miR-16-1 (Han et al., 2006), pri-miR-30a and pri-miR-223 (Zeng and Cullen, 2005). However, the absence of unstructured flanking sequences on one side of the basal stem was tolerated for pri-miR-16-1 if at least 10 nt was present on the other side (Han et al., 2006); as was the case for pri-miR-31, but not pri-miR-223 (Zeng and Cullen, 2005).

Together, in spite of conflicting results, generally desirable structural features for efficient Drosha processing were derived from the aforementioned mutation studies. Desirable features included a large, unstructured loop of ≥ 10 nt, a basal stem of ~ 1 helical RNA turn (~ 11 bp) (Zeng et al., 2005b) and at least 10 nt of adjacent flanking sequences that lack a strong secondary RNA structure (Han et al., 2006; Zeng and Cullen, 2005). An average pri-miRNA structure generated from thermodynamic profiling of 280 human pri-miRNAs also supported these general pri-miRNA features, including a basal stem of ~ 11 bp (Han et al., 2006). In a recent study using high-throughput sequencing to analyse pri-miRNA variants for pri-miR-125a, -16-1, -30a and -223, a basal stem length of ~ 11 bp flanked by at least 9 nt of unstructured flanking sequences was also described as optimal (Auyeung et al., 2013).

General pri-miRNA features can not necessarily be applied to all pri-miRNAs and two main issues surrounding the processing of various pri-miRNA structures persisted. Firstly, reports on the structural determinants of the cleavage site were still conflicting. Mutations in pri-miR-30a that increased or decreased the distance between the cleavage site and the terminal loop by 1 bp shifted the Drosha processing site accordingly. This suggested that the cleavage site was measured from the junction between the stem and the terminal loop (Zeng et al., 2005b). In contrast, the cleavage site of pri-miR-16-1 was not shifted by similar 2 bp mutations in the upper stem, but was shifted by a deletion of 4 bp in the basal stem. This suggested that the cleavage site was instead measured from the junction between the basal stem and unstructured flanking sequences (Han et al., 2006). These results were used to propose opposing models for pri-miRNA recognition and cleavage by the microprocessor (section 1.3.5). Secondly, sequencing evidence suggests that functional miRNAs are derived from numerous pri-miRNAs with sub-optimal basal stem lengths and conformations (miRBase) (Kozomara and Griffiths-Jones, 2014). This raises the question of how these pri-miRNAs are processed effectively in the conventional miRNA biogenesis pathway. Processing of exceptional pri-miRNAs may reveal other desirable, but as yet uncharacterised, structural features important for conventional processing. Processing could also involve alternative, non-conventional mechanisms (section 1.3.14). These possibilities are of particular interest in the design of novel and effective pri-miRNA mimics for therapeutic applications. The processing of pri-miRNAs and mimics with exceptional basal stem lengths and

conformations was therefore explored further in this thesis and informed specific research objectives (section 1.5).

1.3.4 Pri-miRNA sequence determinants

Studies using systematic mutations of the pri-miRNA structure identified basic elements required for efficient processing, but also emphasised the lack of dependence on specific sequences. Random sequences used in the generation of structural mutants and artificial pri-miRNAs yielded results similar to those of the original sequences in processing assays (Han et al., 2006; Zeng et al., 2005b). However, these studies relied on a few well-characterised pri-miRNAs and the influence of individual nucleotides was not thoroughly assessed. In two more recent studies, high throughput mutation and sequencing strategies enabled the identification of sequence motifs that enhance pri-miRNA processing. In the first study (Auyeung et al., 2013), more than 10^{11} randomised variants of four human pri-miRNAs (pri-miR-125a, -16-1, -30a and -223) were generated for residues in unpaired regions of the terminal loop and flanking regions. Variants were assessed for Drosha-DGCR8 cleavage in *in vitro* processing assays, followed by sequencing and comparison with the original, wild-type pri-miRNAs to identify beneficial nucleotides. Three primary sequence motifs, namely, the basal UG motif, basal CNNC motif and terminal loop UGUG motif were identified in functional pri-miRNA variants and found to enhance pri-miRNA processing (Figure 1.2). These sequence motifs are conserved in vertebrates and ~ 79 % of human pri-miRNAs have at least one basal or loop motif. The basal UG motif is located in the 5p arm of the pri-miRNA starting at 14 nucleotides before the Drosha-DGCR8 cleavage site. Functional pri-miRNAs with either a U or a G were enriched, while variants with both were further enriched, suggesting that both residues contribute independently to processing efficiency. The basal CNNC motif is located in the 3p arm of the pri-miRNA starting at 17 or 18 nt after the Drosha-DGCR8 cleavage site and enhances processing by recruiting SRp20 (Serine / Arginine-rich protein), which has been implicated in several RNA processes, including splicing regulation, transcription and translation (Corbo et al., 2013). The apical or terminal loop UGUG motif is located at ~ 24 nt after the 5p Drosha-DGCR8 cleavage site and trinucleotide UGU or GUG derivatives of the full motif are common.

In the second study (Fang and Bartel, 2015), 50 000 to 80 000 variants were created for each of three human pri-miRNAs (pri-miR-125a, -16-1 and -30a) with random nucleotides incorporated at paired positions within the stem. A broadly conserved GHG stem motif (H: not G) with a central mismatch was identified in the 3p arm at position 7 - 9 from the basal junction (Figure

1.2). A true GHG-type motif, such as 5p-CUC / 3p-GCG can significantly enhance processing; but weaker variations of the motif, such as those with inverted G-C base pairs (5p-GUC / 3p-CCG; 5p-CUG / 3p-GCC) or a weak central base pair (5p-CUC / 3p-GAG) can also contribute to processing efficiency. Drosha-DGCR8 processing is most enhanced in the presence of all four basal, loop and stem motifs, but only ~ 4.5 % of human pri-miRNAs have all three primary sequence motifs (Auyeung et al., 2013). Analysis of functional pri-miRNA variants suggested that each motif contributed modestly to pri-miRNA function and that pri-miRNAs relied on different motifs and pairing to varying degrees; further suggesting that pri-miRNA processing is a modular phenomenon (Auyeung et al., 2013).

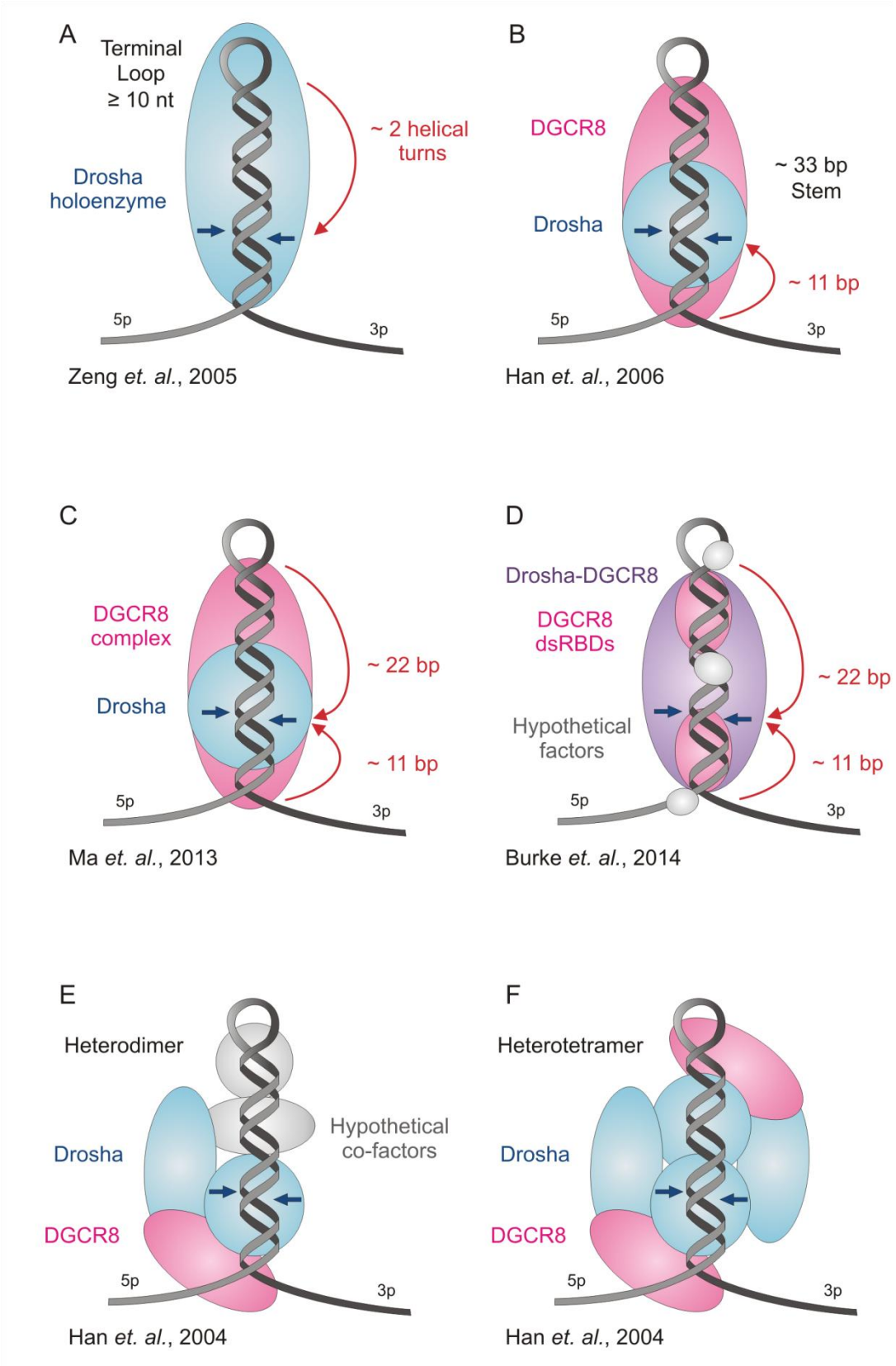
In addition to motifs within or immediately adjacent to the pri-miRNA hairpin, RNA methyltransferase-like 3 (METTL3) recognition motifs enriched within 100 nt of the adjacent flanking sequence (Figure 1.2) were also identified in a systematic search of intergenic and intragenic miRNA-containing regions (Alarcon et al., 2015b). Methylation of pri-miRNAs by METTL3 creates m^6A (N^6-methyladenosine) marks that recruit and enhance recognition by DGCR8 in the nucleus. HnRNPA2B1, a member of the large heterogeneous nuclear ribonucleoprotein (hnRNP) family of proteins involved in numerous aspects of RNA biology (He and Smith, 2009), binds to m^6A marks (RGm6AC; R: purine) for a sub-set of pri-miRNAs, including pri-miR-106b, -93 and -17. HnRNPA2B1 functions as a nuclear m^6A “reader” protein and interacts with DGCR8 to promote pri-miRNA processing (Alarcon et al., 2015a). Additional m^6A reader proteins are likely to exist, as only about half of m^6A-regulated miRNAs are dependent on HnRNPA2B1.

1.3.5 A model of the Drosha-DGCR8 complex

The basic role of the Drosha-DGCR8 complex in pri-miRNA processing has been well established in spite of sequence and structural variation of the pri-miRNA substrates (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004). However, there has been much controversy and seemingly conflicting evidence surrounding the mechanism of pri-miRNA recognition and the form of the microprocessor. Two key issues have been difficult to address: firstly, how the microprocessor distinguishes pri-miRNAs from numerous other RNA hairpins present in cellular transcripts; and secondly, how the microprocessor complex assembles on the pri-miRNA to mediate asymmetric cleavage in the correct orientation. A proper functional and structural analysis of the microprocessor complex has been elusive until recently because a soluble, recombinant protein was difficult to express and purify for Drosha, but less so for DGCR8

(Nguyen et al., 2015). A number of studies therefore focused on DGCR8 alone or on strategic mutations of pri-miRNA substrates (section 1.3.3) to infer processing determinants. Several models for pri-miRNA recognition by Drosha-DGCR8 have been proposed (Figure 1.3).

Strategic mutations of the pri-miR-30a structure and subsequent enzymatic processing assays suggested that the Drosha cleavage site is primarily determined by a distance of ~ 2 helical turns from the ssRNA-dsRNA junction with the terminal loop (Zeng et al., 2005b) (Figure 1.3 A). In contrast, similar mutation experiments using pri-miR-16-1 suggested that the Drosha cleavage site is instead measured at ~ 11 bp from the basal ssRNA-dsRNA junction between the lower stem and flanking sequences (Han et al., 2006). A "ssRNA-dsRNA Junction Anchoring" model was therefore proposed in which DGCR8 acts as a molecular anchor to direct Drosha cleavage (Figure 1.3 B). This model also suggested a sequential assembly of the microprocessor, in which DGCR8 recognises the pri-miRNA and recruits Drosha. However, both the upper and basal stem lengths were later found to affect the precision of Drosha-DGCR8 processing in mutation studies of pri-miR-150, pri-miR-122, pri-miR-142 and pri-miR-16, in which 1 or 2 bp insertions or deletions in the upper or lower stem shifted the cleavage site accordingly, as indicated by deep sequencing of processed miRNA species (Ma et al., 2013). This suggested that the microprocessor cleavage site is co-ordinated by the distance from both upper and lower stem junctions and it was proposed that the DGCR8 complex recognises ~ 3 helical turns of the pri-miRNA stem plus a small part of the ssRNA at basal and apical junctions (Figure 1.3 C). This finding was supported by similar mutation and processing experiments for pri-miR-16, pri-miR-30a and the Simian Virus 40 (SV40) pri-miRNA, which demonstrated that the position and efficiency of Drosha-DGCR8 cleavage are co-operatively co-ordinated by both junctions of the stem, as well as internal structural elements of the stem (Burke et al., 2014). A "ssRNA co-operation model" was therefore proposed in which Drosha-DGCR8 recognises the dsRNA stem and hypothetical factors recognise the apical and basal ssRNA junctions and structural features of the stem like internal loops (Figure 1.3 D). The discovery of basal and apical primary sequence motifs that enhance pri-miRNA processing led to the suggestion that pri-miRNA recognition by Drosha-DGCR8 is a modular phenomenon in which different pri-miRNAs are reliant on individual modules to different extents (Auyeung et al., 2013). Furthermore, the presence of sequence motifs suggested an additional means for distinguishing pri-miRNAs from other RNA hairpins and for distinguishing the basal and apical junctions.


A
Terminal
Loop
≥ 10 nt
~ 2 helical
turns
Drosha
holoenzyme
5p
3p
Zeng et. al., 2005
B
DGCR8
~ 33 bp
Stem
Drosha
~ 11 bp
5p
3p
Han et. al., 2006
C
DGCR8
complex
~ 22 bp
Drosha
~ 11 bp
5p
3p
Ma et. al., 2013
D
Drosha-DGCR8
DGCR8
dsRBDs
~ 22 bp
Hypothetical
factors
~ 11 bp
5p
3p
Burke et. al., 2014
E
Heterodimer
Hypothetical
co-factors
Drosha
DGCR8
5p
3p
Han et. al., 2004
F
Heterotetramer
Drosha
DGCR8
5p
3p
Han et. al., 2004

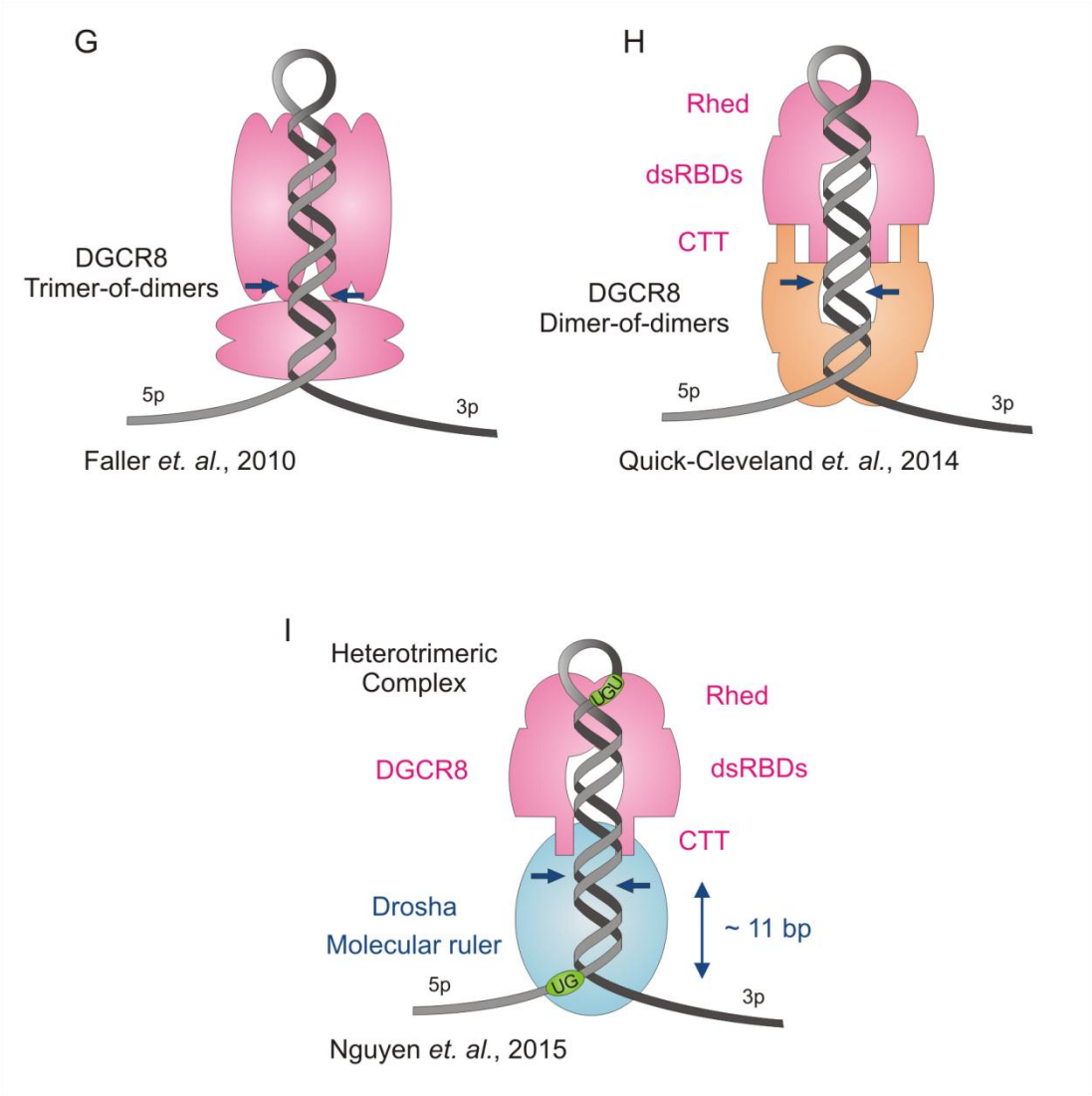


Figure 1.3: Proposed models for pri-miRNA recognition and the microprocessor.

(**A**) Apical junction / loop model. (**B**) ssRNA-dsRNA junction anchoring model. (**C**) Upper and lower stem junction model. (**D**) ssRNA co-operation model. (**E**) Heterodimer model. (**F**) Heterotetramer model. (**G**) Co-operative trimer model. (**H**) Molecular clamp model. DGCR8 domains are indicated. Rhed: RNA-binding haem domain; dsRBD: double-stranded RNA-binding domain; CTT: C terminal tail. (**I**) Current heterotrimeric model in which Drosha functions as a molecular ruler. The figure in each panel was derived from the model in the publication indicated.

Several studies have also generated conflicting results concerning the stoichiometry of the microprocessor. Two possible models supporting a “single processing centre” were initially proposed, in which the two Drosha RNAse III domains form an intramolecular dimer at the cleavage site (Han et al., 2004). In the first “heterodimer model”, it was proposed that the ~ 650 kDa active microprocessor complex may be composed of a single Drosha sub-unit which binds to the stem and to a single DGCR8 sub-unit, which itself binds to the stem and the single stranded flanking sequences to correctly orient the complex (Figure 1.3 E). Additional hypothetical co-factors may also be included in the complex. In the second “heterotetramer model”, it was proposed that the complex may instead include more than one copy of each protein to form a “homodimer of the heterodimer” in which the DGCR8 sub-units may bind to both the basal and apical ssRNA (Figure 1.3 F). In contrast to these models, other data suggested that DGCR8 alone is primarily responsible for pri-miRNA recognition as Drosha lacked any significant RNA-binding activity (Yeom et al., 2006). A number of subsequent studies therefore focused on the interaction between DGCR8 and the pri-miRNA to explain the basis of recognition in the absence of Drosha. In the crystal structure of the human DGCR8 core, two double-stranded RNA-binding domains (dsRBDs) were positioned such that each dsRBD can potentially bind an ~ 11 bp portion of either the upper or lower stem of a single pri-miRNA, provided that the remaining ~ 11 bp of the central stem can adopt a bent conformation (Sohn et al., 2007). An alternative explanation was that the upper and lower stems of a single pri-miRNA might be bound by two separate DGCR8 molecules, leaving each DGCR8 free to interact with two different pri-miRNAs. A haem co-factor (a complex of iron and protoporphyrin IX) was subsequently identified that promotes DGCR8 dimerisation and the oligomerisation state of DGCR8 was estimated using size-exclusion chromatography (Faller et al., 2007). Based on these studies, a “co-operative trimer model” was proposed in which a haem-bound dimer appears to form a catalytically active, co-operative trimer (“trimer of dimers”) (Figure 1.3 G), which may also explain the basis of the ~ 11 bp measuring mechanism in the “ssRNA-dsRNA Junction Anchoring” model (Han et al., 2006). The “co-operative trimer” model was supported by the results of 3D electron tomography and it was further suggested that DGCR8 may be able to distinguish between pri-miRNAs and other non-specific RNAs by virtue of co-operative pri-miRNA binding and subtle differences in binding affinity (Faller et al., 2010). However, it was later suggested that the 3D tomography results might be more consistent with a “dimer of dimers” in agreement with a “molecular clamp” model (Quick-Cleveland et al., 2014) (Figure 1.3 H). In this unifying model, it was proposed that each of the apical and basal junctions is recognised and “clamped” by a DGCR8 dimer through the action of an RNA-binding haem domain (Rhed) and topological differences between the two junctions may account for cleavage asymmetry of the microprocessor. However, again in contrast, DGCR8 was found to bind ssRNA

and dsRNA non-specifically, suggesting that a pre-formed Drosha-DGCR8 heterodimer is required for specific binding (Roth et al., 2013). A solution structure of the Drosha dsRBD also suggested that this domain does in fact participate in RNA interactions with DGCR8 (Mueller et al., 2010).

Issues surrounding pri-miRNA recognition and assembly of the microprocessor have only recently been resolved. A soluble variant of a recombinant Drosha protein was generated through co-expression with a short 23 residue fragment of DGCR8, thus enabling subsequent biochemical and structural studies (Kwon et al., 2015; Nguyen et al., 2015). In the current model of the functional microprocessor (Figure 1.3 I), it was proposed that Drosha-DGCR8 is a heterotrimeric complex containing one Drosha sub-unit and two DGCR8 dimer sub-units and spans the full length of the hairpin (Nguyen et al., 2015). Drosha recognises the basal junction and basal UG motif and acts as a "molecular ruler" to cleave at a distance of ~ 11 bp from the basal junction, while two parallel DGCR8 dimers recognise the upper stem, apical junction and loop UGU motif to ensure processing fidelity. This model confirms that both the secondary RNA structure and primary sequence motifs contribute to pri-miRNA recognition and orientation of the microprocessor without the need for other recognition factors. This model is supported by a corresponding crystal structure of Drosha (Kwon et al., 2015). A heterotrimeric complex was also described in another recent study using a single-molecule photo-bleaching assay to assess the stoichiometry of the microprocessor complex formed in solution by full-length versions of both proteins purified from human cells (Herbert et al., 2015). The latter study also supported a pre-formed microprocessor complex, instead of one that is assembled sequentially (Han et al., 2006), as the heterotrimeric complex was detected in the absence of a pri-miRNA substrate.

The emerging model of pri-miRNA processing by the Drosha-DGCR8 microprocessor is complex and involves determinants of both the secondary RNA structure and pri-miRNA sequence. It is now clear that Drosha-DGCR8 associates with the full pri-miRNA and the cleavage site is measured at ~ 11 bp from the basal junction by the Drosha "molecular ruler". Processing is also enhanced by specific recognition of basal UG and apical UGU motifs. This model is well suited to explain the processing of a variety of pri-miRNA sequences, provided that a conventional structure can be adopted and specific motifs are present. However, variation in the length and conformation of the basal stem is common and could conceivably affect the accuracy and efficiency of cleavage by the Drosha molecular "ruler"; and only ~ 7 % of human pri-miRNAs have both sequence motifs (Auyeung et al., 2013). Again, this raises the question of how exceptional pri-miRNAs are processed effectively in the conventional miRNA biogenesis

pathway. This highlights the need for further investigation, especially with respect to the design of novel and effective pri-miRNA mimics.

1.3.6 Pre-miRNA export

The determinants and mode of Drosha-DGCR8 processing have been a source of confusion and controversy, but greater clarity has generally been achieved for downstream processing events. These events involve less substrate heterogeneity and protein structures were typically obtained sooner. Following Drosha-DGCR8 cleavage, the ~ 65 nt precursor miRNA (pre-miRNA) is exported from the nucleus to the cytoplasm by Exportin-5 (Exp-5, XPO-5) (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003). Exp-5 (Brownawell and Macara, 2002) is a karyopherin transport protein that interacts with nucleoporins of the nuclear pore complexes and shuttles cargo from the nucleus to the cytoplasm. Exp-5 binds specifically to pre-miRNAs in a sequence independent manner, but only in the presence of a Ran-GTP co-factor (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003). In the cytoplasm, GTP hydrolysis releases Ran (Ran guanosine phosphatase) and the pre-miRNA cargo. In contrast to Drosha, a recombinant Exp-5 was expressed and purified prior to the identification of its role in RNAi (Brownawell and Macara, 2002) and the pre-miRNA structural determinants of Exp-5 binding have been largely elucidated. Assays with various pri-miR-30a mutants indicated that Exp-5 binding is facilitated by a pre-miRNA (upper) stem of ≥ 16 bp and a 3' overhang (Zeng and Cullen, 2004). Efficient binding was observed for a pre-miRNA stem of 25 nt, 21 nt and 18 nt, while a shorter stem of 15 or 12 nt blocked Exp-5 binding. Reduction of the terminal loop from 15 nt to 6 nt had no effect on Exp-5 binding; although reduced binding was observed for a terminal loop of 4 nt. Structural data for Exp-5 in complex with pre-miR-30a was provided by a crystal structure of the complex solved at 2.9 Å (Okada et al., 2009) and all-atom multi-nanosecond molecular dynamic (MD) simulations of Exp-5 in aqueous solution (Wang et al., 2011). These structures revealed that Exp-5 recognises the 2 nt 3' overhang and the dsRNA stem of the pre-miRNA. However, some tolerance of the overhangs has been indicated (Zeng and Cullen, 2004), as pre-miR-30a mutants with a 1 nt 3' overhang, 5' and 3' 2nt overhangs or a blunt end were bound as well as the standard pre-miRNA. Longer 3' overhangs did not reduce Exp-5 binding, but longer 5' overhangs blocked binding. In addition to its function as an exporter, Exp-5 protects the pre-miRNA from nuclear degradation (Zeng and Cullen, 2004) by surrounding the pre-miRNA in a baseball mitt-like conformation (Okada et al., 2009).

1.3.7 Pre-miRNA processing

Once in the cytoplasm, pre-miRNAs are subject to a second processing step in the miRNA biogenesis pathway (Lee et al., 2002b) (Figure 1.1). The dsRNA stem of the pre-miRNA is cleaved near the terminal loop by the RNase III enzyme Dicer (Bernstein et al., 2001), thus producing a staggered miRNA duplex with 3' dinucleotide overhangs. The conserved role of Dicer in RNAi and the generation of small RNAs was first described in *C. elegans, D. melanogaster* and humans (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). Dicer is bound by two double stranded RNA-binding proteins (dsRBPs), namely the TAR (trans-activation response) RNA-binding protein (TRBP) (Chendrimada et al., 2005; Gregory et al., 2005) and the protein activator of PKR (PACT) (Lee et al., 2006). In association with TRBP, Dicer is recruited to Argonaute 2 (Ago2) to form the RISC-loading complex (RLC) (MacRae et al., 2008). Both TRBP and PACT facilitate miRNA production (Kok et al., 2007) and appear to act in a non-redundant manner to alter the behaviour of Dicer cleavage, thus contributing to miRNA length determination and guide strand selection for a subset of mammalian miRNAs (Lee et al., 2013; Wilson et al., 2015). These proteins also assist in distinguishing between pre-siRNAs and pre-miRNAs, as PACT inhibits the processing of pre-siRNAs (Lee et al., 2013). Only TRBP or PACT can bind to directly to Dicer at one time through a common binding site (Wilson et al., 2015).

Again in contrast to Drosha, a recombinant form of Dicer was purified shortly after its role in RNAi was discovered (Zhang et al., 2002) and the basis of pre-miRNA recognition by Dicer has been largely elucidated. Human Dicer specifically cleaves dsRNA at a distance of ~ 22 nt from the termini to produce a 2 nt 3' overhang (Zhang et al., 2002; Zhang et al., 2004). Initially it was proposed that Dicer measures this distance specifically from the 3' overhang in a "3' end counting rule" (Vermeulen et al., 2005), but an additional "5' end counting rule" (Park et al., 2011) and "loop counting rule" (Gu et al., 2012) have since been proposed. Both termini and the terminal loop can therefore influence Dicer cleavage and recognition of the whole pre-miRNA was supported by the structure of human Dicer (Lau et al., 2012). Dicer has a preference for pre-miRNA substrates with a 2 nt 3' overhang, such as those generated by RNase III Drosha cleavage, and cleaves these substrates uniformly and efficiently (Park et al., 2011). Pre-miRNA substrates with shorter 1 nt 3' overhangs or blunt termini can also be cleaved, but less efficiently; while a longer 3' overhang can shift the cleavage site for certain pre-miRNAs (Vermeulen et al., 2005; Zhang et al., 2004). If the terminal structure is not optimal, certain pre-miRNAs appear to rely more on cleavage measured from a particular terminus (Park et al., 2011). For example, when 1 - 3 extra uridine residues were added to the 3' end of the pre-miRNA, cleavage of pre-

miR-143 was largely measured from the 5' end, while cleavage of pre-miR-142 was largely measured from the 3' end and cleavage of pre-miR-200c appeared to be measured from either end. Therefore, the Dicer cleavage site does not necessarily shift in response to 3' heterogeneity of the pre-miRNA (section 1.3.10), but this depends on the pre-miRNA identity (Park et al., 2011). The sequence of the 3' overhang can also affect the efficiency of Dicer cleavage for certain pre-miRNAs (Vermeulen et al., 2005). Optimal processing was observed when the 2 nt 3' overhang had a C in the second last position and an A in the last position; while less optimal processing was observed with an A in the second last position and a C or U in the last position. Dicer may also have sequence preferences that influence cleavage site selection near the terminal loop and specifically avoid NpG sequences (Starega-Roslan et al., 2015a). The size of the terminal loop can also affect Dicer cleavage activity, as observed for Drosha (Zhang and Zeng, 2010). Pri-miR-30a, -16-1 and -31 mutants with smaller terminal loops of 4 - 6 nt had reduced Dicer cleavage activity *in vitro* (Zhang and Zeng, 2010), which may be a consequence of reduced pre-miRNA flexibility and enzyme turn-over. RNA editing of the terminal loop by adenosine deaminases acting on RNA (ADARs) can also affect Dicer activity (Kawahara et al., 2007a; Liu et al., 2015). The position of the loop is important for cleavage consistency and the loop counting rule dictates that Dicer will cleave precisely if the cleavage site is positioned at exactly 2 nt from the terminal loop or an internal bulge (Gu et al., 2012).

1.3.8 The RNA-induced silencing complex (RISC)

The RISC (Hammond et al., 2000) is composed of an Argonaute (Ago) protein and an RNA guide sequence (siRNA / miRNA) that directs the silencing of target genes (Hammond et al., 2001). The assembly of RISC involves two successive steps: loading of the miRNA duplex to form a pre-RISC and duplex unwinding to form a mature RISC (Kawamata and Tomari, 2010). Details of RISC assembly have been primarily described in *Drosophila* and there is still some uncertainty about precise mechanisms in humans. In the first step, the miRNA duplex produced by Dicer cleavage is loaded into an Argonaute (Ago) protein. An important aspect of this step is the asymmetric loading of the miRNA duplex into Ago, as the strand of the duplex with lower 5' stability is typically selected as the mature RNA guide (Khvorova et al., 2003; Schwarz et al., 2003). Sequence and structural preferences of the Ago proteins can affect RISC loading. There are four closely related human Argonaute proteins (Ago1 - 4), each of which can associate with miRNAs in a ribonucleoprotein complex (Meister et al., 2004). These proteins have similar preferences and single mismatches in the central region (position 8 - 11) of the RNA duplex promote RISC loading (Yoda et al., 2010). Ago2 also has a preference for miRNAs with a 5' U or

A (Frank et al., 2010). Dicer-TRBP and Dicer-PACT complexes can enhance strand selection abilities of Ago2 (Noland and Doudna, 2013). The RISC-loading complex (RLC), composed of Dicer, TRBP and Ago2 (Gregory et al., 2005; MacRae et al., 2008) may therefore facilitate formation of pre-RISC and asymmetric loading. After cleavage, dsRNA is re-positioned within the Dicer complex to enable sensing of thermodynamic asymmetry, suggesting that Dicer itself could facilitate strand-selective loading (Noland and Doudna, 2013). However, there is some debate over whether Dicer processing and RISC assembly are coupled. Early reports indicated that Dicer is important for RISC assembly (Gregory et al., 2005; MacRae et al., 2008; Maniataki and Mourelatos, 2005) and a RLC capable of both pre-miRNA processing and Ago loading was recently confirmed in mammalian cells (Liu et al., 2012c) and supported by structural data (Wang et al., 2009b). Other work has instead suggested that Dicer cleavage and RISC assembly are uncoupled (Yoda et al., 2010) and that Dicer is dispensable for asymmetric loading of RISC in mammals (Betancur and Tomari, 2012). Earlier reports also indicated that RISC assembly does not require ATP (Gregory et al., 2005; Maniataki and Mourelatos, 2005), but more recent work shows that human RISC assembly is greatly facilitated by ATP (Yoda et al., 2010).

In the second step of RISC assembly, the duplex is unwound and the guide strand remains associated with the Ago protein, thus creating a mature minimal RISC (Figure 1.1). Unwinding of the RNA duplex can be slicer-dependent or slicer-independent, depending on the associated Ago protein. Ago2 is the only human Ago protein to retain endonuclease “slicer” activity (Liu et al., 2004; Meister et al., 2004) and can cleave the passenger strand to facilitate unwinding (Leuschner et al., 2006; Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). Cleavage occurs at a position of the duplex corresponding to between position 10 and 11 of the guide (Elbashir et al., 2001b 2001; Elbashir et al., 2001c), but central mismatches are common in the miRNA duplex and prevent passenger strand cleavage. Slicer-independent unwinding therefore occurs for complexes containing Ago 1 - 4 and is facilitated by single mismatches in the seed region (position 2 - 8) and to a lesser degree, the 3’ mid-region (position 12 - 15) (Yoda et al., 2010). Slicer-independent unwinding has recently been described as the general mechanism for RISC maturation in humans, as unwinding of both near-perfect siRNA duplexes and imperfect miRNA duplexes is facilitated by physiological temperatures (Park and Shin, 2015).

The ~ 22 nt mature guide strand directs RISC to mRNAs with complementary target sequences, while the passenger strand is removed and degraded by the C3PO endoribonuclease (Liu et al., 2009; Ye et al., 2011). RISC mediates cleavage or translational repression of the target mRNA, depending on the level of complementary pairing with the miRNA (Hutvagner and Zamore, 2002; Zeng et al., 2003) (Figure 1.1). Near-perfect miRNA-target interactions can induce mRNA

cleavage by complexes containing Ago2 (Liu et al., 2004; Meister et al., 2004), while imperfect miRNA-target interactions can induce translational repression by complexes containing Ago 1 - 4 (Pillai et al., 2004; Wu et al., 2008). RISC typically binds to several imperfect sites in the 3' UTR of a mRNA, which results in translational repression, deadenylation and subsequent degradation of the target mRNA with co-localisation in cytoplasmic processing bodies (P bodies) (Eulalio et al., 2007a). GW182 proteins associate with RISC at the 3' UTR and recruit the polyA-binding protein (PABP) (Fabian et al., 2009) and the PAN2-PAN3 and CAF1-CCR4-NOT deadenylase complexes (Behm-Ansmant et al., 2006; Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). Deadenylation of the mRNA results in decapping by the enzyme DCP2 and degradation by the major cytoplasmic 5' - 3' exonuclease XRN1 (Behm-Ansmant et al., 2006; Eulalio et al., 2007b). Further details concerning the molecular stages of translational repression are reviewed by Huntzinger *et. al.* (Huntzinger and Izaurralde, 2011). Functional assays in human cells indicate that translational repression and degradation are mechanistically linked through the interactions of GW182 proteins with PABP and deadenylase complexes (Huntzinger et al., 2013). General sorting of miRNAs and miRNAs* between different Ago proteins has been described in *Drosophila* (Czech et al., 2009; Okamura et al., 2009) and *C. elegans* (Steiner et al., 2007), but a general sorting mechanism has not been demonstrated in humans (Azuma-Mukai et al., 2008; Burroughs et al., 2011; Dueck et al., 2012). The identity of the Ago protein in RISC may be linked to the length of certain miRNAs (Dueck et al., 2012; Juvvuna et al., 2012) or particular isomiRs (Burroughs et al., 2011) by virtue of varying degrees of RNA protection, but the significance of this association is not yet clear.

RISC is typically a very stable complex and can direct multiple rounds of mRNA silencing (Hutvagner and Zamore, 2002). RISC loading was initially assumed to be an irreversible process which prevents miRNA release, thus enhancing RISC turnover and silencing of corresponding targets (Okamura, 2013). However, loading and unloading of RISC has recently been described as a reversible process affected by the nature of pairing between the miRNA and target sequences (De et al., 2013). A fully complementary target sequence can enhance unloading of a 21 nt RNA guide, thus reducing RISC stability and enabling reprogramming of Ago with a different guide. This unloading effect is less pronounced for RNA guides with a higher affinity for Ago (5' U or A) and is attenuated by 5' pairing and 3' mismatches in the guide-target interaction. These features are common for natural miRNAs and their interactions with endogenous mRNA targets, suggesting that miRNA-mediated silencing is not typically inhibited by RISC unloading. In fact, a single miRNA is expected to participate in at least two rounds of silencing for an imperfect target in mammalian cells (Baccarini et al., 2011). Reversible loading and unloading of RISC could suggest a sampling mechanism through which Ago identifies and retains preferred

miRNAs (De et al., 2013). Once released from RISC, miRNAs are susceptible to degradation by various ribonucleases (section 1.3.12).

1.3.9 Target recognition & specificity

MiRNA recognition of a target sequence is primarily mediated through complementary pairing with the 5' region of the miRNA. Earlier observations indicated higher 5' conservation for related miRNA sequences in *C. elegans* (Lim et al., 2003) and perfect pairing between the 5' miRNA sequence and expected targets in *Drosophila* (Lai, 2002). A specific 5' "seed" region (nucleotides 2 - 8) was defined and used to successfully predict mammalian target genes using the TargetScan algorithm (Lewis et al., 2003). This implied a functional significance of the miRNA seed region in target recognition, which has since been confirmed experimentally. Effective target silencing was largely dictated by perfect complementary pairing with the seed region (Doench and Sharp, 2004; Kloosterman et al., 2004; Lai et al., 2005; Lim et al., 2005) and the introduction of mismatches or G - U wobbles to this miRNA - target interaction was detrimental to mRNA repression (Doench and Sharp, 2004). However, a bulge in the mRNA target may be tolerated at a position corresponding to between position 5 and 6 of the seed, as demonstrated for a subset of miR-124 targets in the mouse brain (Chi et al., 2012). The seed is so crucial to target recognition that effective target silencing can be enabled with perfect pairing to the 5' of the miRNA, but little or no complementary pairing to the 3' of the miRNA (Brennecke et al., 2005). Different miRNAs with the same seed sequence can also regulate the same targets (Lai et al., 2005). MiR-147b and miR-210, for example, are from distinct miRNA families, but share the same seed sequence and have common mRNA targets; while miR147a and miR147b are identical except for one base in the seed region and mediate different phenotypic effects (Bertero et al., 2012). Nonetheless, different miRNAs with same seed sequence may have moderately varied silencing as a consequence of different interactions over the 3' region (Bartel, 2009). Supplementary pairing to position 13 - 16 of the miRNA can enhance target silencing (Grimson et al., 2007), while stronger compensatory pairing to the 3' region can enable function for targets with weak seed pairing (Brennecke et al., 2005). The function of some miRNAs is therefore more reliant on a precise seed sequence, while the function of other miRNAs is promoted by non-seed sequences (Zhang et al., 2015). In contrast to 3' compensatory pairing, consecutive pairing with the last 2 - 4 nt of a 21 nt guide RNA can reduce silencing efficacy by enhancing RISC unloading (De et al., 2013).

Other factors can also influence target silencing, such as the number and location of target sites. Several adjacent or closely spaced target sites can enhance miRNA silencing in a synergistic manner (Grimson et al., 2007; Kloosterman et al., 2004) and can even overlap at non-seed regions (Doench and Sharp, 2004). Effective target sites are typically located near the ends of a 3' UTR and are at least 15 nt from a stop codon (Grimson et al., 2007), but functional target sites may also be located in the 5' UTR or coding region (Kloosterman et al., 2004; Saxena et al., 2003). A target site with an A corresponding to position 1 of the miRNA, a U or A corresponding to position 9 and increased AU content of the 3' region can enhance miRNA-mediated repression (Nielsen et al., 2007), in agreement with common miRNA features. Accessibility of the target site is also an important factor in miRNA binding and target sites are preferentially located in regions of the genome with high accessibility (Kertesz et al., 2007). A novel RNA-binding function was also recently described for Ago proteins where they can associate with mRNA targets independently of guide miRNAs (Li et al., 2014a). This feature of Ago proteins is enabled by sequence preferences of Ago upstream of the binding site which may assist miRNAs in target recognition.

1.3.10 IsomiRs

Slightly different variations of miRNAs have been identified from deep sequencing data and are known as "isomiRs", a play on the term isomer. The functional consequences of sequence variation vary, but can affect target recognition, RISC loading and miRNA stability (Neilsen et al., 2012). IsomiRs can result from imprecise RNase III cleavage, general RNA editing or 3' modifications. In the first case, alternative Drosha cleavage can occur when the upper and lower stem lengths of the pri-miRNA are sub-optimal, resulting in multiple cleavage sites and the generation of 5' isomiRs (Ma et al., 2013). Even a 1 or 2 bp change in the stem can affect the precision of Drosha processing, which can in turn affect the register of the miRNA seed region, subsequent Dicer processing and strand selection. Pri-miR-142, for example, has isomiRs generated from three different Drosha cleavage sites (Wu et al., 2009). The position of an internal bulge within a stem of uniform length can also alter the consistency of Drosha cleavage (Burke et al., 2014). In addition, Drosha and Dicer may have inherent sequence preferences (Starega-Roslan et al., 2015b). In the absence of internal bulges Dicer seems to specifically avoid cleavage sites in the 3p arm that generate a 5' G (Starega-Roslan et al., 2015a). Templated isomiRs and 5' isomiRs can therefore be produced by imprecise Drosha and Dicer cleavage.

In the second case, isomiRs can be generated through A - I editing by the adenosine deaminases acting on RNA (ADARs) which convert adenosine residues to inosine residues (Neilsen et al., 2012). An alteration in the mature miRNA sequence can alter pairing capacity (A - U; I - C) and the specificity of target recognition; as was demonstrated for members of the miR-376 cluster where edited miR-376a effectively silenced a different set of genes (Kawahara et al., 2007b). RNA editing sites were only identified in 30 miRNAs in large sequencing data sets, but ~ 21 % of 3' UTR editing sites also affected miRNA seed target sites (Peng et al., 2012). Currently, the physiological relevance of A - I miRNA editing appears to be limited to a small subset of miRNAs (Neilsen et al., 2012).

In the third case, miRNA sequences are commonly subject to 3' RNA editing mechanisms (Chiang et al., 2010; Landgraf et al., 2007; Wyman et al., 2011). Non-templated mono-adenylation or mono-uridyaltion at the 3' ends of miRNAs is most common (Wyman et al., 2011). Nucleotide addition is catalysed by RNA nucleotidyl transferases, including PAPD4 and PAPD5 cytoplasmic polyA polymerases and TUT1 and TUT4 (ZCCHC11) terminal uridyl transferases (Wyman et al., 2011). The 3' ends of miRNAs are also subject to trimming by 3' - 5' exonuclease activity, such as the Nibbler enzyme in *Drosophila* (Han et al., 2011; Liu et al., 2011). A similar exonuclease has not yet been described in humans, but 22 nt miRNAs associated with Ago2 in mammalian cells may undergo specific 3' trimming to 21 nt (Juvvuna et al., 2012). These 3' mechanisms are not expected to significantly affect the targeting function of mature miRNAs, as the 5' miRNA seed region and 3' supplementary sites are unaffected. However, 3' non-templated addition is also common for pre-miRNAs, with mono-uridylation being common and cytidine or adenosine modification less common (Newman et al., 2011). In turn, this can shift the Dicer cleavage site and the seed sequence of the mature miRNA. Longer nucleotide additions are also associated with RNA stability and degradation. Non-templated 3' addition is common but not universal as certain human miRNAs are frequently modified, but most are modified infrequently or not at all (Wyman et al., 2011). Apart from affecting miRNA stability, it is still not clear whether isomiRs have more widespread functional significance (Neilsen et al., 2012).

1.3.11 Structure of miRNA processing enzymes

The key proteins involved in miRNA biogenesis and target silencing have several common features. The Drosha-DGCR8 and Dicer processing enzymes have double-stranded RNA-binding domain (dsRBDs) and tandem catalytic RNase III domains; while the processing and silencing complexes include a PAZ (Piwi / Argonaute / Zwille) or PAZ-like domain. However,

these proteins also have unique features which influence their function. Details of the protein structure and function below are crucial for an understanding of miRNA-mediated silencing and the design of effective RNAi-based therapeutic mimics.

Structure of Drosha

Drosha was identified as the enzyme that cleaves a pri-miRNA to a pre-miRNA in *in vitro* studies with an immunopurified Drosha protein (Lee et al., 2003b). Drosha is classified as an RNAse III enzyme as it is a dsRNA-specific endonuclease that produces a staggered cleavage of the RNA helix and has a requirement for Mg^{2+} ions (Lee et al., 2003b; Zamore, 2001). It is also classified as a class II RNAse III enzyme because it has tandem endonuclease domains, one dsRBD and an extended N terminal domain (Filippov et al., 2000; Lee et al., 2003b) (Figure 1.4 A). The crystal structure of human Drosha was recently solved at a resolution of 3.2 Å and revealed that Drosha is an elongated protein in which the two RNase III domains (RIIIDa, RIIIDb) are positioned above a central domain (CED) (Kwon et al., 2015). The catalytic site of RIIIDa is positioned to cleave the 3p arm of the pri-miRNA, while the catalytic site of RIIIDb is positioned to cleave the 5p arm, thus producing the characteristic 3' 2 nt overhang (Han et al., 2004). The two RNAse III domains form an intramolecular dimer to create a single processing centre which is connected to the central region through a long helix known as the "connector helix". The CED is composed of three distinct sub-domains, including the connector helix, a PAZ-like sub-domain and the N terminal half of the CED known as the "platform".

In support of the heterotrimeric microprocessor complex, a DGCR8-binding site lined by hydrophobic residues was located on each RNase III domain. The interaction of RIIIDb with DGCR8 was found to be crucial for Drosha stability, while the interaction of RIIIDa with DGCR8 was not. Unique features were also identified that support Drosha's role as a "molecular ruler". These include non-canonical zinc-finger motifs which contribute to the interaction between RIIIDa and the CED; a long insertion in the RIIIDa that forms a "bump helix" and "mobile-basic" (MB) helix which may contribute to the RNA interaction; and a kinked link between the connector helix and RIIIDa that produced a stooped conformation important for interaction with the substrate. The bump helix, MB helix, platform and PAZ-like modules of Drosha appear to work together to enable recognition of the basal junction and several positively-charged residues on the surface of the CED may facilitate interactions with the RNA and processing. The R-rich domain of Drosha was found to have an affinity for ssRNA (Zeng and Cullen, 2005). In addition

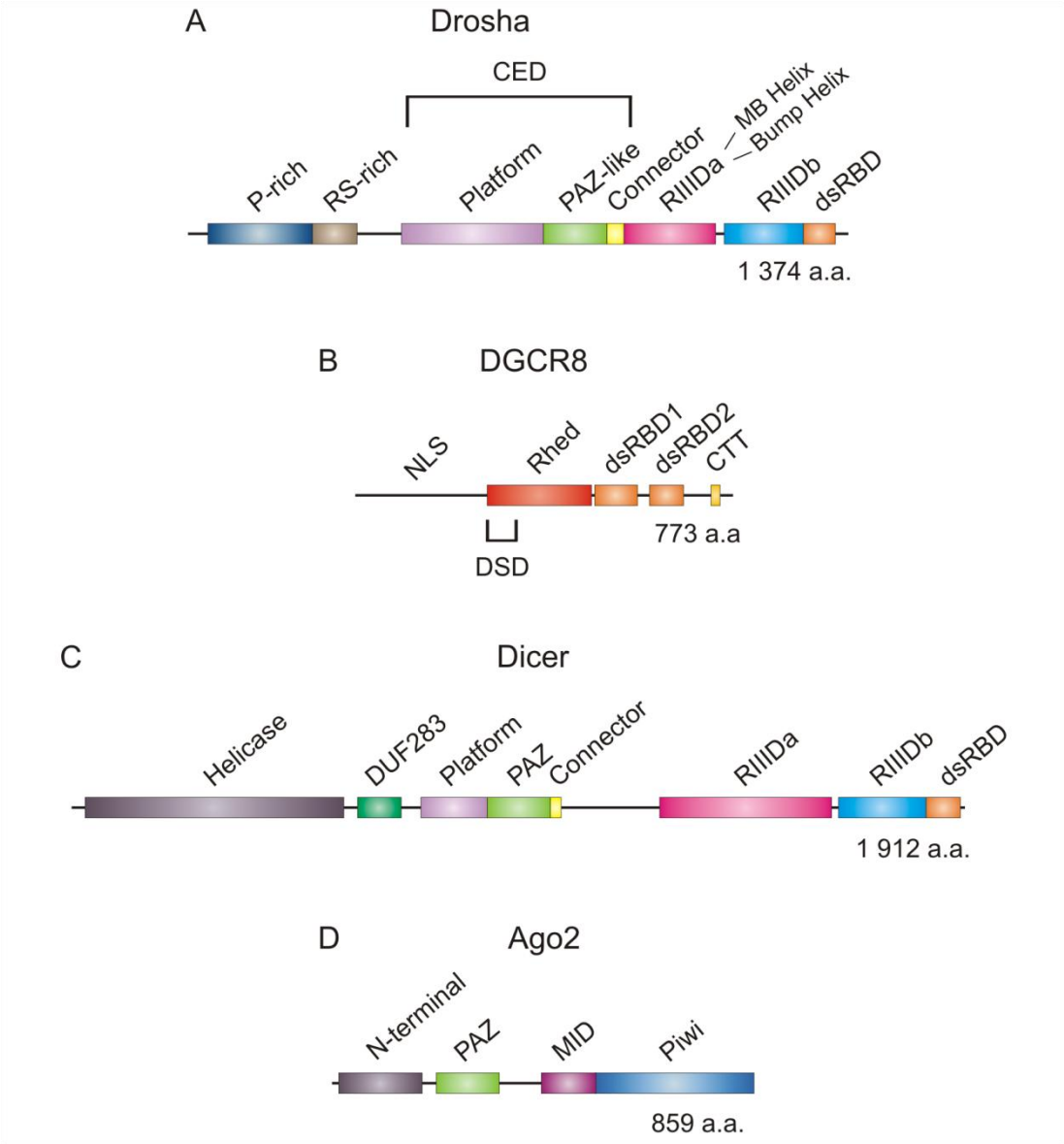


Figure 1.4: Domain organisation of Drosha-DGCR8, Dicer and Argonaute proteins

Protein domain organisation of human (**A**) Drosha, (**B**) DGCR8, (**C**) Dicer and (**D**) Argonaute proteins. These diagrams were adapted from those by (Kwon et al., 2015), (Lee et al., 2003b), (Nguyen et al., 2015) and (Schirle, et al., 2014). P-rich: proline-rich; RS-rich: arginine / serine-rich; CED: central domain; PAZ: Piwi, Argonaute, Zwille proteins; RIIID: RNase III domain; dsRBD: double-stranded RNA-binding domain; Rhed: RNA-binding haem domain; CTT: C-terminal tail; DUF283: Domain of unknown function 283; MID: middle domain; Piwi: Piwi domain.

to the crystal structure, a nuclear magnetic resonance (NMR) solution structure of the Drosha dsRBD suggested that this domain does in fact retain RNA-binding activity and may participate in RNA interactions of the microprocessor complex (Mueller et al., 2010). Unique features of the Drosha dsRBD were also identified, including a long negatively-charged loop that may facilitate protein-protein interactions.

Structure of DGCR8

The DGCR8 (DiGeorge syndrome chromosomal / critical region 8) protein (Landthaler et al., 2004; Shiohama et al., 2003) was found to be an essential component of a smaller Drosha complex capable of mediating pri-miRNA processing, termed the microprocessor (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004). The DGCR8 protein is composed of an RNA-binding haem domain (Rhed) (Quick-Cleveland et al., 2014), two tandem dsRBDs and a C-terminal region or tail (CTT) (Yeom et al., 2006) (Figure 1.4 B). The Rheds and haem co-factor are important for DGCR8 dimerisation (Faller et al., 2007), while the N terminal region was found to be dispensable for processing activity, but contains a nuclear localisation signal (NLS) (Yeom et al., 2006). The haem-binding domain contains a conserved dimerisation sub-domain (DSD) (residues 298 - 352) (Senturia et al., 2010) and a tryptophan WW motif which plays a role in both dimerisation and association with the haem co-factor. The WW motif may also interact with the proline-rich domain of Drosha (Gregory et al., 2004). A conserved Cys352 residue of the DSD probably functions as a ligand for iron of the haem co-factor (Faller et al., 2007). Ferric Fe(III), but not ferrous Fe(II), activates DGCR8 for efficient processing (Barr et al., 2012). The crystal structure of the DGCR8 core, including two dsRBDS and the CTT, was resolved at 2.6 Å and revealed a “butterfly” conformation composed of an α / β-fold with seven helices and eight strands (Sohn et al., 2007). The two dsRBDs interact with each other and the CTT through hydrophobic interactions and H-bonds. In the current heterotrimeric model of pri-miRNA recognition (Nguyen et al., 2015), the various domains of DGCR8 have different roles: Rhed enhances processing accuracy; the two dsRBDS recognise the upper stem and enhance processing efficiency; and the CTT interacts with Drosha to stabilise the complex and activate Drosha cleavage, in agreement with earlier studies (Yeom et al., 2006). Interactions between the three Drosha and DGCR8 sub-units stabilise the heterotrimeric complex, including interactions of the Rhed-mediated DGCR8 dimer; binding between the Drosha RIIIDb and one DGCR8 CTT; and an interaction between the Drosha RIIIDa and the other DGCR8 CTT (Kwon et al., 2015).

Structure of Dicer

Dicer was identified as the enzyme responsible for generating ~ 22 nt RNAs from dsRNA in RNAi in *Drosophila* (Bernstein et al., 2001) and conserved homologues with a similar role were subsequently identified in *C. elegans, D. melanogaster* and humans (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). Dicer is a class III RNase III enzyme because it contains two RNase III domains (RIIIDa, RIIIDb) and an N terminal DExD helicase domain (Bernstein et al., 2001) (Figure 1.4 C). Human Dicer is a 220 kDa multidomain protein and also includes a Domain of Unknown Function (DUF283), Piwi / Argonaute / Zwille (PAZ) family domain and a dsRBD (Zhang et al., 2002). The structure of the PAZ domain was initially informed by crystal and solution structures generated for the PAZ domain of *Drosophila* Ago1 and Ago2 (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003) and human Ago1 (Ma et al., 2004). In each case, a conserved hydrophobic pocket was identified that could bind a 2 nt 3' overhang. The crystal structure of the human PAZ domain in complex with a siRNA-like duplex indicated that this pocket specifically anchors the 2 nt 3' overhang, making contacts with the phosphodiester backbone of the overhanging 3' strand and the 5' residue of the other strand (Ma et al., 2004). Specific aromatic residues of the 3' pocket make contact with the 3' overhang (Lingel et al., 2004), suggesting that the PAZ domain is an RNA end-binding module (Ma et al., 2004).

Structural information for human Dicer was initially derived from the crystal structure of the protozoan *Giardia intestinalis* Dicer solved at a resolution of 3.3 Å, which contains only the conserved PAZ domain and two RNase III domains (Macrae et al., 2006). This structure revealed an elongated shape with the two RNase III domains positioned next to each other and connected to the PAZ domain by a long α helix known as the "connector helix", which is encircled by a platform domain. The RNase III domains are connected by a large, helical bridging domain and, like Drosha, form an intramolecular dimer to create a single processing centre (Zhang et al., 2004). RIIIDa is positioned to cleave the 3p arm and RIIIDb is positioned to cleave the 5p arm as in Drosha (Zhang et al., 2004), but recent pre-miRNA cleavage assays indicated that RIIIDa specifically avoids cleaving at a position that generates a 5' G (Starega-Roslan et al., 2015a). This suggests that the RIIIDa may be able to move slightly within the intramolecular dimer, possibly implying flexibility of the bridging domain. A distance of 65 Å was identified between the PAZ and RNAse III domains, which corresponded to the length of a typical 25 bp RNA product produced by *Giardi*a Dicer. This suggested that the arrangement of these domains enabled Dicer to recognise the ends of a dsRNA and function as a "molecular ruler" (Macrae et al., 2006). Single-particle electron microscopy analysis indicated that the

complete human Dicer adopts an L-shaped conformation (Wang et al., 2009b), which was supported by similar experiments for the Dicer-TRBP complex (Lau et al., 2012).

A crystal structure of the platform - PAZ - connector helix domains revealed that, in addition to the 3' overhang - binding pocket in the PAZ domain (Ma et al., 2004), a 5' phosphate - binding pocket is also formed by basic residues of these domains (Park et al., 2011; Tian et al., 2014). This indicated that both ends of the staggered dsRNA end are recognised by Dicer. The loop of the pre-miRNA is also recognised by the helicase sub-domains which form a clamp-like structure adjacent to the RNase III catalytic sites (Lau et al., 2012), in agreement with a similar observation for *Drosophila* Dicer-1 where the pre-miRNA loop is recognised and appears to co-ordinate the cleavage site (Tsutsumi et al., 2011). A domain localisation strategy using biotinylation-streptavadin labelling and electron microscopy revealed that the nuclease core of human Dicer is arranged slightly differently to that of *Giardia* to enable the production of 21 - 23 nt dsRNAs (Lau et al., 2012). Notably, the helicase and PAZ domains are positioned at opposite ends of the 3D model and recognise the terminal pre-miRNA loop and 3' overhang, respectively; while the dsRBD and RIIIDs are positioned in between to enable binding and cleavage at an appropriate distance (Lau et al., 2012). Further investigation has shown that structural changes in the helicase domain upon binding a pre-miRNA and dsRBDS (TRBP or PACT) enable productive substrate recognition and explain Dicer's preference for pre-miRNAs over pre-siRNAs (Taylor et al., 2013).

Structure of Argonautes

Argonautes are a highly conserved family of proteins and are characterised by the presence of an N terminal or central PAZ (PIWI, Argonaute, Zwille / Pinhead) domain and C terminal PIWI domain (Carmell et al., 2002; Cerutti et al., 2000). The name of the PIWI protein was originally derived from the "P-element-induced wimpy testis" mutant phenotype in genetic studies in *Drosophila* (Lin and Spradling, 1997). There are eight members of the human Argonaute family, which is subdivided into the Argonaute (Ago) and PIWI sub-families on the basis of sequence homology (Sasaki et al., 2003). There are four members of the Ago sub-family (Ago 1 / EIF2C1, Ago 2 / EIF2C2, Ago 3 / EIF2C3, Ago 4 / EIF2C4) which are expressed in a variety of tissues; and four members of the PIWI sub-family (PIWIL1 / HIWI, PIWIL2 / HILI, PIWIL3, PIWIL4 / HIWI2) which are mainly expressed in the germline tissue. Ago proteins are the central component of RISC (Hammond et al., 2000; Hammond et al., 2001) and associate with miRNAs to mediate cleavage or translational repression of the target mRNAs in a general PTGS

mechanism; while PIWI proteins are the central component of piRISC (piRNA-induced silencing complex) and associate with piRNAs in the germline tissue to silence transposons (section 1.3.15).

Human Ago proteins consist of N terminal, PAZ, middle (MID) and PIWI domains (Song et al., 2004) (Figure 1.4 D). The structure of eukaryotic Argonaute proteins was initially informed by crystal and solution structures of model domains and proteins, including the PAZ domains of *Drosophila* and human Ago proteins, bacterial Ago proteins and archaeal PIWI proteins. The highly conserved PAZ domain is found exclusively in Dicer and Argonaute proteins and functions similarly in both as an end binding module that specifically binds to a 2 nt 3' overhang of dsRNA (Lingel et al., 2003, 2004; Ma et al., 2004; Song et al., 2003; Yan et al., 2003). A highly conserved, 5' phosphate-binding pocket was identified in crystal structures of the *Archaeoglobus fulgidus* (*Af*) PIWI (Ma et al., 2005; Parker et al., 2005) and *Thermus thermophilus* Ago (Wang et al., 2008) proteins bound to a siRNA or siRNA-like duplex. This pocket is located primarily within the MID domain and a divalent cation assists in anchoring the 5' phosphate of an RNA guide. In support of the 5' phosphate-binding pocket, RISC was found to be reliant on Mg^{2+} (Schwarz et al., 2004), implying that this is the co-ordinated ion (Ma et al., 2005); and a 5' phosphate was found to be important for the stability and fidelity of the minimal human RISC (Rivas et al., 2005). The 5' nucleotide of the RNA guide also appears to associate with Ago in a specific manner. In complex with *Af* PIWI, the 5' base pair of the guide is unwound, while the rest of the guide-target duplex is accommodated within a basic channel spanning the equivalent of the MID-PIWI domain interface (Ma et al., 2005; Parker et al., 2005). A crystal structure of the human Ago2 MID domain in complex with various nucleoside monophosphates also showed that specific contacts are made between a rigid loop of the MID domain and U or A bases, while interactions with a C or G base are less favourable. Ago2 has a higher affinity for UMP or AMP, thus explaining Ago binding preferences for RNA guides with a 5' U or A (Frank et al., 2010). The PIWI domain was identified as an RNase H-like domain in the crystal structures of the *Pyrococcus furiosus* (*Pfu*) Argonaute (2.25 Å resolution) (Song et al., 2004) and the *Af* PIWI protein (Domain B; 1.9 Å resolution) (Parker et al., 2004). Conserved aspartate residues and an aspartate-aspartate-glutamate (DDE) motif indicated a potential RNase H -type catalytic site (Parker et al., 2004; Song et al., 2004), suggesting a role for the PIWI domain in cleavage of the target mRNA. A similar but unique aspartate-aspartate-histidine (DDH) motif was identified in human Ago2 and found to confer "slicer" activity for mRNA cleavage *in vitro* (Liu et al., 2004; Rivas et al., 2005).

In terms of Ago as a whole, the structure of *Pfu* Ago indicated that the PAZ domain is positioned above a crescent-shaped base composed of the N-terminal, MID and PIWI domains, creating a groove between the PAZ and PIWI domains (Song et al., 2004). This suggested a model for RNA-guided mRNA cleavage by Argonaute, in which the RNA guide binds to the PAZ domain, pairs with a mRNA target and orients the mRNA in the groove for cleavage by the PIWI domain on the opposite side (Song et al., 2004). The distance between the 5' pocket and the catalytic site of *Af* PIWI corresponded to cleavage of the mRNA target at a fixed distance from the 5' end of the guide between position 10 and 11 (Ma et al., 2005), as observed for functional siRNAs (Elbashir et al., 2001b).

Early observations for model domains and proteins were supported by crystal structures of the full length human Ago2 (hAgo2) protein bound to guide RNAs (2.2 - 2.3 Å resolution) (Elkayam et al., 2012; Schirle and MacRae, 2012). The features of individual domains and overall domain arrangement are very similar and support a close evolutionary link for Ago proteins (Elkayam et al., 2012). However, the precise positions of well-conserved individual domains differ for prokaryotic (Wang et al., 2008) and human Ago proteins; and additional secondary RNA structures and extended loops of the latter may enable the binding of co-factors (Schirle and MacRae, 2012). In particular, tandem tryptophan binding pockets in the human PIWI domain probably form a surface for interaction with GW182 proteins involved in the translational repression / degradation mechanism of RISC or other tryptophan-rich co-factors (Schirle and MacRae, 2012). Like *Pfu* Ago (Song et al., 2004), hAgo2 adopts a bi-lobed structure where the PAZ domain is positioned above a crescent-shaped base formed by the N-terminal, PIWI and MID domains. This conformation has been described as a "duck", in which the PAZ domain is the head, PIWI is the body and the N terminal and MID domains are the wings (Elkayam et al., 2012).

A recent crystal structure of hAgo2 in complex with a representative guide-target RNA duplex has provided clarity for mode of target recognition, suggesting that it is a stepwise mechanism (Schirle et al., 2014). Initial pairing of position 2 - 5 of the guide promotes conformational changes in the protein and the guide to expose the full seed (2 -8) and the supplementary pairing region (13 - 16). The 5' of the guide is anchored by the 5' pocket of the MID domain; positions 2 - 6 of the guide assume a helical conformation and are bound in a minor groove; positions 14 - 18 of the guide are threaded through a narrow channel between the N terminal and PAZ domains (N-PAZ channel); and the end of the guide binds to the 3' pocket of PAZ (Schirle et al., 2014). Initial pairing (2 - 5) appears to cause a shift in helix-7 of the protein, which relaxes a kink between position 6 and 7 of the guide caused by isoleucine 365 and widens the

N-PAZ channel. These changes enable the full seed and supplementary regions of the guide to adopt an A-form or near A-form conformation for optimal pairing with the target. Pairing beyond position 8 of the guide is restricted by narrowing of the central RNA-binding groove, supporting the definition of the seed region (2 - 8). In general, hAgo2 associates with the guide RNA through H bonds and salt linkages with the sugar-phosphate backbone, explaining the basis of sequence-independent recognition. In addition to the preference for an RNA guide with a 5' U, hAgo2 also appears to specifically accommodate the complementary A at the start of the target sequence in a narrow pocket with a 3 times higher affinity than for targets starting with a U, G or C. The recent structure of hAgo2 (Schirle et al., 2014) has therefore advanced and consolidated the current understanding of miRNA binding and target recognition.

1.3.12 Regulation of miRNA biogenesis

MiRNA biogenesis and function is regulated at several steps in the pathway, including transcription, Drosha, Dicer and Ago processing and miRNA decay. These points of regulation are important to consider in the context of a therapeutic strategy and may affect the success of pri-miRNA mimics. Several examples of regulation are summarised below, while other examples may be found in the literature and recent reviews (Ha and Kim, 2014; Kim et al., 2009; Krol et al., 2010).

Regulation of expression

The expression of intronic and exonic miRNAs is subject to transcriptional regulation of the host gene. MiRNAs in independent transcriptional units are typically transcribed by RNA pol II (Cai et al., 2004; Lee et al., 2004) and are therefore subject to conventional transcriptional regulation. Several transcription factors have been identified in the positive and negative regulation of miRNA expression in specific tissues, developmental stages or disease states. The p53 tumour suppressor positively regulates the expression of the miR-34 family (Chang et al., 2007 2007) which functions in cell cycle arrest. In contrast, the proto-oncogenic transcription factor c-MYC activates the expression of the oncogenic miR-17~92 cluster (He et al., 2005; O'Donnell et al., 2005), as well as downregulating the expression of multiple miRNAs (Chang et al., 2008). E2F7 and E2F1-3 transcription factors function in an antagonistic manner to regulate the expression of several miRNAs (miR-25, -26a, 27b, 92a and -7) that promote cellular proliferation (Mitxelena et al., 2016).

Regulation of miRNA networks may also involve feedback loops. E2F1-3 regulates transcription of the miR-17~92 cluster, while miR-20a of this cluster represses the translation of E2F2 and E2F3, suggesting an auto-regulatory loop (Sylvestre et al., 2007). MiRNA expression may also be regulated by epigenetic mechanisms and active promoters are associated with H3K4 (Histone 3, lysine residue 4) and H2A.Z and RNA pol II marks (Nepal et al., 2015). Hyper methylation by DNMT1 (DNA methyltransferase 1) and DNMT3 of the CpG islands associated with the promoters of downregulated miRNAs is common, especially in cancer (Weber et al., 2007). The large chromosome 19 miRNA cluster is expressed from a CpG-rich promoter showing maternal imprinting in the placenta and is therefore only expressed from the paternal allele (Noguer-Dance et al., 2010). Individual miRNAs of a polycistronic cluster, like miR-106b~93~25, may also be differentially expressed as a result of alternative spicing (Ramalingam et al., 2014).

Regulation of Drosha-DGCR8 processing

Activity of the Drosha-DGCR8 complex may be regulated in several ways. Drosha and DGCR8 appear to cross-regulate each other post-transcriptionally to maintain processing homeostasis (Han et al., 2009). Drosha-DGCR8 destabilises the DGCR8 mRNA by cleaving a hairpin within the transcript and DGCR8 stabilises the Drosha protein. The activity and level of Drosha and DGCR8 proteins are altered by protein modifications, including phosphorylation and acetylation. Phosphorylation of DGCR8 increases protein stability (Herbert et al., 2013), while phosphorylation of Drosha at specific serine residues by GSK3β (glycogen synthase kinase 3β) is required for nuclear localisation (Tang et al., 2011). Acetylation of Drosha inhibits ubiquitination and degradation (Tang et al., 2013). Drosha-DGCR8 function can also be altered by associated co-factors, like haem, which promote DGCR8 dimerisation and prevent auto-inhibition of the haem-binding region (Faller et al., 2007). Accessory proteins may play a role in processing of specific miRNAs. The p53 tumour suppressor interacts with the Drosha complex through the DEAD-box RNA helicase p68 to facilitate the processing of specific pri-miRNAs in response to DNA damage (Suzuki et al., 2009). TDP-43 (TAR DNA-binding protein 43) specifically enhances Drosha stability and processing during neuronal differentiation (Di Carlo et al., 2013). Nucleolin, a ribosomal RNA (rRNA) processing protein, is required for effective processing of miR-15a and miR-16 (Pickering et al., 2011). The NF90-NF45 (nuclear factor) complex binds directly to pri-miRNAs like pri-miR-21 and pri-let-7a to inhibit processing (Sakamoto et al., 2009). In contrast, the KH-type splicing regulatory protein (KSRP) binds to the terminal loop of a subset of pri-miRNAs and associates with both Drosha and Dicer complexes

to promote processing (Trabucchi et al., 2009). Similarly, the RNA-binding protein hnRNPA1 binds to the terminal loop of pri-miR-18a to promote Drosha processing (Guil and Caceres, 2007). Drosha-DGCR8 activity is enhanced by sequence motifs of the pri-miRNA substrate (section 1.3.4); including apical and basal sequence motifs which interact with the microprocessor directly and METTL3 motifs which recruit hnRNPA2B1. The organisation of miRNAs within polycistronic clusters can also affect the processing of individual miRNAs. The miR-17~92 cluster forms a globular tertiary structure and miRNAs located on the surface of the structure are processed more efficiently than those within the structure (Chaulk et al., 2011). In the miR-11~998 cluster, processing of miR-998 is reliant on the presence of a second miRNA within the cluster (Truscott et al., 2016).

Regulation of pre-miRNA export and processing

Regulation occurs at several downstream events in pri-miRNA processing, including pre-miRNA export and processing. Exp-5 is regulated by a post-transcriptional mechanism dependent on PI3K (phosphatidylinositol 3-kinase) (Iwasaki et al., 2013). Dicer activity is regulated by TRBP, which stabilises Dicer and affects the rate and consistency of processing (Chendrimada et al., 2005; Lee and Doudna, 2012). TRBP and PACT also affect the efficiency of dsRNA processing by Dicer in different ways (Lee et al., 2013). Phosphorylation of TRBP by the mitogen-activated protein kinase (MAPK) Erk increased stability of the Dicer-TRBP complex and the production of certain growth-promoting miRNAs (Paroo et al., 2009). Dicer is also regulated by an autoregulatory feedback loop, in which mature let-7 directly targets the mRNA coding sequence (Forman et al., 2008). Furthermore, the conformation and activity of Dicer can be regulated in response to ADAR editing of the pre-miRNA terminal loop, where certain terminal loop sequences of pre-miR-151 enhance Dicer activity (Liu et al., 2015).

Argonaute regulation

Argonaute proteins may also be regulated by post-translational modifications which in turn affect RISC activity. Prolyl 4-hydroxylation of Ago2 by the collagen prolyl-4-hydroxylase enhances its stability and enables interaction with heat shock protein 90 (Hsp90) (Qi et al., 2008; Wu et al., 2011a). Hsc70 / Hsp90 chaperones enhance ATP-dependent loading of small RNA duplexes, possibly by inducing a conformational change in Ago (Iwasaki et al., 2010). Phosphorylation of Ago2 at serine 387 through the MAPK pathway facilitates localisation to P bodies (Zeng et al.,

2008); and Akt-mediated phosphorylation of the same residue represses cleavage and upregulates translational repression (Horman et al., 2013). In contrast, EFGR (epidermal growth factor receptor) phosphorylates Ago2 at tyrosine 393 in response to hypoxia, which suppresses the maturation of specific miRNAs (Shen et al., 2013). Phosphorylation of tyrosine 529 in the 5'-binding pocket is also expected to negatively affect miRNA binding (Rudel et al., 2011). In mouse stem cells, lin-41 acts as an E3 ubiquitin ligase to mediate ubiquitylation and lysosomal degradation of Ago2 (Rybak et al., 2009). Ago2 levels also appear to be regulated by miRNA abundance, as unbound Ago2 is unstable, suggesting a regulatory feedback mechanism that finely tunes Ago levels (Martinez and Gregory, 2013; Smibert et al., 2013). In response to DNA damage, TP53 interacts with Ago2 to regulate its association with miRNAs (Krell et al., 2015).

Regulation of miRNA abundance

MiRNA abundance may be affected by the stability of both miRNA precursors and mature miRNAs. Non-templated 3' end modification of pre-miRNAs appears to be a common phenomenon (Newman et al., 2011). Lin28 proteins mediate 3' uridyaltion of specific pre-miRNAs through recognition and binding of a tetra-nucleotide sequence motif in the terminal loop and subsequent recruitment of TUT4; thus blocking Dicer processing and inducing pre-miRNA degradation (Heo et al., 2008; Heo et al., 2009). In contrast, mono-uridylation of pre-miRNAs with a 1 nt 3' overhang by TUT7, TUT4 or TUT2 enables the formation of a dinucleotide overhang to permit Dicer processing (Heo et al., 2012). MiRNA levels may be directly affected by non-templated 3' additions (section 1.3.10). The 3' adenylation by the polyA polymerase GLD-2 has been implicated in the stabilisation of miR-122 in the liver (Katoh et al., 2015); while 3' adenylated miR-122 is subsequently deadenylated and destabilised by the polyA specific ribonuclease (PARN) (Katoh et al., 2015). Similarly, polyadenylation of miR-21 by the PAPD5 cytoplasmic polyA polymerase may signal degradation by recruiting PARN (Boele et al., 2014). The frequency of 3' modification for certain miRNAs changes during differentiation, suggesting that this is a regulated and physiologically relevant process in humans (Wyman et al., 2011). Active miRNA turnover may also be mediated by various exoribonucleases and in certain cellular conditions (reviewed in (Ruegger and Grosshans, 2012)). For example, specific miRNAs were degraded by 3' to 5' exonuclease activity of the interferon-induced human polynucleotide phosphorylase (PNP) in melanoma cells (Das et al., 2010). Binding to highly complementary targets can also influence miRNA stability and initiate target RNA-directed miRNA degradation (TDMD) (de la Mata et al., 2015). TDMD may be a consequence of enhanced 3' modification and trimming (Ameres et al., 2010) or dissociation from RISC (De et al., 2013), both of which

render the miRNA susceptible to degradation. In addition to degradation mechanisms, miRNA abundance may be reduced by the presence of circular RNAs (circRNAs) with multiple target sites that act as “sponges” and compete with endogenous targets for miRNA-binding (Hansen et al., 2013; Memczak et al., 2013). However, it is not yet clear whether the activity of such competitive endogenous RNAs (ceRNAs) constitutes a general regulatory mechanism (reviewed in (Thomson and Dinger, 2016)).

1.3.13 Relationship of pri-miRNA processing to transcription and splicing

MiRNAs are processed from RNA transcripts that may also be involved in other cellular processes, such as transcription, splicing and translation. The relationship between pri-miRNA processing and other processes varies and may be determined by the genomic context of the pri-miRNA (intergenic, intronic, exonic). Transcription and processing of endogenous pri-miRNAs appear to be coupled. The microprocessor is specifically recruited to sites of pol II transcription for intronic and intergenic miRNAs, indicating that pri-miRNA processing is co-transcriptional (Morlando et al., 2008). Pri-miRNAs retained at transcription sites are processed more efficiently, suggesting that processing is enhanced by coupling to transcription (Pawlicki and Steitz, 2008). In addition, Drosha cleavage of intergenic miRNAs appears to be coupled with transcriptional termination by creating a 5’ substrate for the 5’ - 3’ Xrn2 exonuclease (Ballarino et al., 2009).

The relationship between pri-miRNA processing and splicing is more complex. Drosha can cleave intronic pri-miRNAs from an unspliced intron, indicating that splicing is not a prerequisite for processing and that processing occurs before splicing catalysis (Kim and Kim, 2007). As Drosha cleavage would separate two exons, it was suggested that the microprocessor cleaves before splicing is complete, possibly between the commitment and excision steps to allow for exon-tethering (Kim and Kim, 2007). Subsequent experiments supported a physical association between the microprocessor and splicing machinery (Agranat-Tamir et al.; Kataoka et al., 2009), indicating that a combined microprocessor - spliceosome complex may mediate pre-miRNA cleavage prior to the completion of splicing. Spliceosome assembly also appears to positively influence processing of intronic miRNAs, but not intergenic miRNAs, while processing of intronic miRNAs promotes splicing; suggesting a mutually co-operative relationship between the microprocessor and the spliceosome (Janas et al., 2011). However, other evidence suggests that the microprocessor and spliceosome may compete with each other, as the inhibition of splicing increases the expression of mature intronic miRNAs and knockdown of Drosha

increases splicing (Agranat-Tamir et al., 2014). Together, this may suggest co-operative interaction through recruitment, but competition of function, although the functional association between the microprocessor and spliceosome has not yet been fully characterised.

Cleavage of an intronic miRNA hairpin does not impair accumulation of the host mRNA, implying that an intronic miRNA and protein can be derived from a single transcript (Kim and Kim, 2007). In contrast, splicing and miRNA processing are mutually exclusive for rare pre-miRNAs that overlap splice sites at intron - exon boundaries (Melamed et al., 2013). Splicing interrupts the pre-miRNA sequence and prevents processing of mature miRNAs like miR-412 and miR-202, while exon skipping allows Drosha-DGCR8 to successfully cleave a pre-miRNA from the transcript. Exonic miRNAs like miR-155 can be derived from spliced or unspliced nuclear transcripts (Slezak-Prochazka et al., 2013), but expression of an exonic miRNA and translation of the host exon are expected to be mutually exclusive. This phenomenon was observed for miR-198 positioned in the 3' UTR of a protein-coding gene, where TGF-β signalling caused a switch from miRNA to mRNA expression (Sundaram et al., 2013).

1.3.14 Alternative miRNA biogenesis

Most miRNAs are derived from the canonical miRNA biogenesis pathway, but others are instead derived through alternative miRNA biogenesis mechanisms. In alternative biogenesis, miRNA precursors typically resemble other RNA intermediates in the cell and are processed accordingly. Alternative biogenesis can generally be classified as Drosha-independent or Dicer-independent (Curtis et al., 2012). The functional significance of miRNAs with alternative biogenesis has not yet been fully characterised and may only apply to a few select miRNAs. Nonetheless, other alternative modes of processing are of interest in a therapeutic context and may also provide novel strategies.

Mirtrons

The most common type of Drosha-independent biogenesis involves introns / pre-miRNAs known as "mirtrons" (Ruby et al., 2007). Mirtrons were originally identified in *D. melanogaster, C. elegans* (Okamura et al., 2007; Ruby et al., 2007) and mammals (Berezikov et al., 2007). Functional characterisation studies indicated that mirtrons, like miR-877, miR-1224 and

miR-1226, are processed by an alternative splicing-dependent mechanism (Okamura et al., 2007; Ruby et al., 2007; Sibley et al., 2012b). Mirtrons are spliced from the host transcript as introns and fold into a complete pre-miRNA, thus bypassing cleavage by Drosha-DGCR8 (Ruby et al., 2007). The 5' and 3' ends of a canonical mirtron are therefore defined by the 5' splice donor site (5' GU) and 3' splice acceptor site (AG 3'), respectively. In splicing, the 5' donor site is covalently linked to a central branch point (A) to form a lariat structure. Mirtron processing is subsequently dependent on the lariat debranching enzyme to produce a linear sequence capable of folding into a pre-miRNA (Ruby et al., 2007). The characteristic 3' dinucleotide overhang is formed by the unpaired splice acceptor site, although single symmetrical overhangs are also common as a result of pairing between the U of the 5' donor site and the A of the 3' acceptor site (Berezikov et al., 2007). The mirtron hairpin enters the miRNA biogenesis pathway as a pre-miRNA and is exported from the nucleus by Exp-5, followed by Dicer cleavage (Okamura et al., 2007).

A total of almost 500 putative human mirtrons have been identified from the analysis of large-scale sequencing data (Ladewig et al., 2012; Wen et al., 2015). Four mirtron sub-classes have been described, including conventional mirtrons (6.9 %), 5' tail mirtrons (86 %) (Ladewig et al., 2012), 3' tail mirtrons (3.8 %) (Flynt et al., 2010) and two-tail mirtrons (3.6 %) (Castellano and Stebbing, 2013). In the case of non-canonical mirtrons, additional 5' or 3' sequences ("tails") need to be resected or trimmed to generate a functional pre-miRNA. The enzymes responsible for trimming have not yet been identified, although trimming of a 3' tail by the RNA exosome, a multi-sub-unit 3' - 5' exonuclease, was demonstrated for miR-1017 in *Drosophila* (Flynt et al., 2010). The majority of human mirtrons exhibited a bias for 5' U at 5' ends of both 5p and 3p species, as is the case for conventional miRNAs (Wen et al., 2015). A 5' G was expected for conventional and 3' tailed mirtrons as a consequence of the 5' GU splice donor site, but several mirtrons instead initiated with the U of the donor site; implying a 5' decapping or trimming mechanism (Wen et al., 2015). Mirtrons are also subject to modification and non-templated 3' monouridylation is common (Westholm et al., 2012). In *Drosophila*, the uridyltransferase (TUTase) Tailor preferentially uridylates mirtrons and canonical pre-miRNAs with a 3' G to suppress Dicer processing and biogenesis (Bortolamiol-Becet et al., 2015; Reimao-Pinto et al., 2015). The miR-1236 mirtron has been shown to silence the vascular endothelial growth factor receptor (VEGFR3) (Jones et al., 2012), but in general, the targets of mirtron silencing require further experimental validation (Curtis et al., 2012).

MiRNAs derived from shRNAs, lhRNAs, tRNAs and snoRNAs

Several other classes of Drosha-independent miRNAs have been described where a miRNA can be derived from a short or long hairpin RNA (shRNA / lhRNA), transfer RNA (tRNA) or small nucleolar RNA (snoRNA). In mouse ESCs, miR-320 and miR-484 were found to be derived from endogenous short hairpin RNAs (shRNAs) that only require Dicer processing (Babiarz et al., 2008). Four miRNAs were derived from both arms of a long hairpin RNA for mirtron pre-miR-3102, presumably by sequential Dicer processing (Chiang et al., 2010). tRNA-derived small RNAs (tsRNAs; type I) can be generated independently of Drosha from the 3' end of a misfolded tRNA by Dicer (Haussecker et al., 2010). Similarly, pre-miRNAs encoded by the murine γ-herpesvirus 68 (MHV68) bear a 5' tRNA moiety and were processed by tRNase Z instead of Drosha-DGCR8 (Bogerd et al., 2010).

Small RNAs that function like miRNAs are also derived independently of Drosha-DGCR8 from small nucleolar RNAs (snoRNAs). SnoRNA-derived miRNAs (sno-miRNAs) (Brameier et al., 2011) of 21 - 26 nt were identified in *Giardia* from 3 different snoRNAs (Saraiya and Wang, 2008), while similar 20 - 22 nt sno-miRNAs were identified from the human ACA45 snoRNA (Ender et al., 2008). Evidence suggests that sno-miRNAs mediate translational repression of a target and are reliant on Dicer and Argonaute for processing (Ender et al., 2008; Saraiya and Wang, 2008). There are two classes of snoRNAs that act in ribonucleoproteins (RNPs) to direct the enzyme modification of other RNAs by anti-sense interactions; the H/ACA box snoRNAs mediate pseudouridylation, while the C/D box snoRNAs mediate methylation (Matera et al., 2007). Sno-derived RNAs (sdRNAs) have been mapped to both classes of snoRNAs in animals, with 20 - 24 nt sdRNAs derived from the 3' end of H/ACA snoRNAs and ~ 17 - 19 nt or > 27 nt sdRNAs derived from the 5' end of C/D snoRNAs (Taft et al., 2009). However, further functional characterisation of these candidate sno-miRNAs is required. Eleven additional C/D human sno-miRNAs have been validated experimentally, including U3, U78 and HBII-336 (Brameier et al., 2011).

Dicer-independent miRNAs

Dicer-independent processing has been described for splicing-independent mirtron-like miRNAs (simtrons) including miR-1225 and miR-1228 (Havens et al., 2012). Simtron biogenesis is reliant on Drosha, but not DGCR8, Dicer, Exp-5 or Ago2. However, simtrons still associate with all four Ago proteins to mediate silencing of target genes. This suggests that simtrons are functional, but

may also rely on other, as yet identified, processing enzymes in addition to Drosha. Dicer-independent processing has also been described for pre-miRNA-451, which is crucial for normal development in vertebrates (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010a). Instead of Dicer cleavage of both strands of the pre-miRNA, the 3p passenger strand of the pre-miRNA is cleaved by Ago2 at a position 10 nt from the 5' end. Subsequent 3' uridylation and trimming of the cleaved pre-miRNA produce a mature miRNA. The poly(A)-specific ribonuclease (PARN) has been identified as the 3' - 5' exonuclease involved in trimming of pre-miR-451 (Yoda et al., 2013). Ago2 cleavage of pre-miRNAs was previously observed as a general cellular mechanism that may contribute to selection of the guide strand and removal of the passenger strand; but Ago2-cleaved pre-miRNAs (ac-pre-miRNAs) were found to be Dicer substrates (Diederichs and Haber, 2007). This suggests that Ago2 cleavage may be a common occurrence, but only plays a significant role in the biogenesis of rare miRNAs involved in vertebrate development. tRNA-derived small RNAs (tsRNAs; type II) can also be generated independently of Dicer as mature sequences from the 3' end of a tRNA by tRNase Z (Haussecker et al., 2010).

Agotrons

Drosha-independent and Dicer-independent processing has recently been described (Hansen et al., 2016). Short introns of 80 - 100 nt associate directly with Ago2 in the cytoplasm and have been termed "agotrons". Agotrons appear to be spliced like mirtrons, but lack Dicer processing and associate with Ago2 as unprocessed > 30 nt pre-miRNAs capable of gene silencing. Only 87 agotron candidates have been identified in humans, including miR-1225 which was previously classified as a simtron (Havens et al., 2012). This finding suggests that miRNAs may be generated independently of the canonical processing pathway with potential implications for therapeutic development.

1.3.15 Other small regulatory RNAs

Other small RNAs with regulatory roles have been described in germline and somatic cell types, including endogenous siRNAs (endo-siRNAs) and PIWI-interacting RNAs (piRNAs). The processing and function of these small RNAs are reliant on components of the miRNA biogenesis pathway or other similar pathways involving related proteins; but full details are still emerging. However, it is clear that the activity of these RNAs involves the more obscure elements of the genome, including retrotransposons and pseudogenes. This has added another level of complexity to our understanding of these elements and provided a new perspective on

previous observations and anomalies. Both endo-siRNAs and piRNAs have roles in silencing of transposons in the germline, which is crucial for fertility and integrity of the genome. While not the investigated in this research project, an understanding of these closely related small RNAs and processes could reveal novel therapeutic opportunities.

Endo-siRNAs

Small ~ 21 nt RNAs equivalent to endogenous siRNAs (endo-siRNAs or esiRNAs) were identified in *C. elegans* (Ruby et al., 2006), the gonad and somatic cells of *Drosophila* (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008) and oocytes and embryonic stem cells (ESCs) in mice (Babiarz et al., 2008; Tam et al., 2008; Watanabe et al., 2008). Endo-siRNAs were derived from stem-loop structures (Babiarz et al., 2008; Kawamura et al., 2008; Okamura et al., 2008), retrotransposons (Babiarz et al., 2008; Kawamura et al., 2008), complementary transcription products of a gene and pseudogene (*trans*), bi-directional transcription of a single locus (*cis*) or an inverted repeat pseudogene (Tam et al., 2008; Watanabe et al., 2008). Endo-siRNAs are processed independently of Drosha-DGCR8 and associate with Ago2 to direct cleavage of target RNAs. Target transcripts include those of protein-coding genes and mobile elements like retrotransposons (Czech et al., 2008; Ghildiyal et al., 2008). In mouse oocytes, endo-siRNA targets include several genes involved in spindle assembly and chromosome segregation (Tam et al., 2008) and endo-siRNAs appear to play an essential role in meiosis (Stein et al., 2015). Fertility and efficient processing of endo-siRNAs in female mice was also attributed to an oocyte-specific Dicer (Dicer ^{O}) isoform, which can process siRNAs efficiently from dsRNA (Flemr et al., 2013).

piRNAs

PIWI interacting RNAs (piRNAs) are 26 -31 nt RNAs that function to silence transposons in the germline and are important for gametogenesis, embryogenesis and stem cell maintenance in animals. Small RNAs were first implicated in silencing of transposons in the *Drosophila* germline (Aravin et al., 2001). Nearly 80 % of *Drosophila* piRNAs were derived from regions of repeated sequence (Brennecke et al., 2007) and can also be described in more general terms as repeat associated small interfering RNAs (rasiRNAs) (Aravin et al., 2003). Mammalian piRNAs were subsequently described in the testes of mice (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Watanabe et al., 2006) and rats (Lau et al., 2006); and later, in mouse oocytes (Tam et

al., 2008; Watanabe et al., 2008). PiRNAs are transcribed from discrete genomic clusters, comprising transposon sequences and defective transposon remnants (Brennecke et al., 2007). In turn, piRNAs primarily function in the silencing of complementary transposon sequences, suggesting that piRNA clusters serve as a "genetic memory of past transposon challenges" (Czech and Hannon, 2016).

PiRNA biogenesis has certain similarities with miRNA biogenesis, but processing is independent of both Drosha-DGCR8 and Dicer and full details are still emerging. PiRNAs are initially transcribed as long, precursors (Aravin et al., 2006) which lack characteristic hairpin conformations and are exported to the cytoplasm. PiRNA precursors are processed into shorter intermediates by phospholipase D endonucleases of the outer-mitochondrial membrane, including Zucchini in flies (Ipsaro et al., 2012; Nishimasu et al., 2012) and MITOPLD in mice (Huang et al., 2011; Qi et al., 2011). PiRNA intermediates are loaded into members of the PIWI sub-family of Argonaute proteins in *Drosophila* (PIWI, Aubergine (Aub), Ago3) and mice (MILI, MIWI, MIWI2) to form a piRNA induced silencing complex (piRISC). This complex functions like miRISC to silence complementary transposon sequences. PiRNAs of different lengths are associated with particular PIWI proteins, presumably as a consequence of varying degrees of nucleotide protection (Watanabe and Lin, 2014).

PiRNA-mediated silencing of transposons also involves a unique "ping-pong" amplification cycle which has been well -characterised in *Drosophila* (Brennecke et al., 2007; Gunawardane et al., 2007). Briefly, primary piRNAs direct the cleavage of complementary transposons at a position corresponding to between nucleotide 10 and 11 of the piRNA guide. This generates a 5' and 3' transposon cleavage product, the latter of which then functions as a secondary piRNA. Secondary piRNAs direct cleavage of primary piRNA transcripts, resulting in the generation of additional primary piRNAs. In this way, the original piRNA product is amplified in response to the level of transposon activity to mediate an appropriate PTGS response. Consecutive, phased primary piRNA production has also been reported recently (Han et al., 2015; Mohn et al., 2015), which diversifies the pool of piRNAs targeting a particular transposon. PiRNAs therefore function in an adaptive genetic response against harmful transposons. In addition to PTGS, piRNAs also associate with nuclear PIWI to direct transcriptional gene silencing (TGS) of transposon genes through heterochromatin formation, reviewed in (Czech and Hannon, 2016). Further details of piRNA biogenesis and function have been discussed in several recent reviews (Czech and Hannon, 2016; Siomi et al., 2011; Watanabe and Lin, 2014).

1.3.16 Transcriptional Gene Silencing (TGS)

TGS by small RNAs and components of the RNAi pathway has been demonstrated in plants, yeast and *C. elegans,* but a similar phenomenon has not yet been conclusively demonstrated in mammals. An established TGS mechanism in mammals could enable therapeutic silencing of the transcription of target genes in addition to the current PTGS approach. This would enable even more powerful gene silencing and the ability to target integrated or latent pathogenic sequences. In yeast, a nuclear version of RISC known as the RNA-induced transcriptional silencing complex (RITS) mediates sequence-specific chromatin modification and epigenetic silencing (Verdel et al., 2004). RITS contains Ago1, Chp1 (a heterochromatin-associated chromodomain protein), Tas3 and Dicer-dependent small RNAs. Small RNAs derived from centromeric repeats direct RITS to complementary regions of the chromatin, which signals transcriptional repression through the formation and maintenance of H3K9 (Histone 3, lysine residue 9) methylation (Volpe et al., 2002). RITS associates with the Clr4-Rik-1-Cul4 (CLRC) complex responsible for H3K9 methylation (Hong et al., 2005).

TGS was initially reported in mammals using synthetic promoter-directed siRNAs (Kawasaki and Taira, 2004; Morris et al., 2004). The mechanisms of mammalian TGS are still unclear, but involve RNAi components. Ago1, along with DNMT3a and histone deacetyalse 1 (HDAC1), has been implicated in the initiation of long-term transcriptional silencing in human cells; while DNMT1 was necessary for the maintenance of silencing (Hawkins et al., 2009). SiRNA treatment was found to increase both H3K9 and H3K27 methylation of a targeted promoter, suggesting a therapeutic application for TGS (Weinberg et al., 2006). Ago1 also directly interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cells (Huang et al., 2013). MiRNAs may be involved in transcriptional gene silencing (TGS) following import to the nucleus (Hwang et al., 2007). Ago1 has also been implicated in the regulation of constitutive and alternative splicing through preferential binding to active transcriptional enhancers through the enhancer RNAs (eRNAs) (Allo et al., 2014). A recent and comprehensive review of current data highlighted strong supporting evidence for RNAi proteins in the nucleus and a TGS mechanism in mammals (Kalantari et al., 2016). It is not yet clear whether mammalian miRNAs (or piRNAs) play a general and physiologically relevant role in the nucleus and further investigation of mammalian TGS mechanisms is required.

1.3.17 The role of mature miRNAs

MiRNA-mediated gene silencing is a widespread phenomenon and miRNAs function in a vast array of cellular processes. MiRNA function has been described in development, differentiation, metabolism, the cell cycle, ageing and aberrant gene expression in disease. MiRNA function can be especially complex in the context of established regulatory networks. MiRNA expression is often tissue-specific or stage-specific and a normal phenotype relies on a careful balance of miRNA function. Serum miRNAs can also be used as biomarkers in applications for disease diagnosis. The variety of cellular functions that can be affected by a single miRNA is an important consideration in therapeutic approaches. To illustrate the variety of possible regulatory activities by miRNAs, examples for the miR-34 family are summarised below.

The miR-34 family has been studied extensively and is well-known for its role in tumour suppression. Ectopic expression of miR-34 family members can lead to apoptosis, cell cycle arrest or senescence (Hermeking, 2010). The miR-34 family are direct targets of p53 and downregulate genes involved in cell cycle progression, including cyclin E2, cyclin-dependent kinase 4 and the hepatocyte growth factor receptor, leading to inhibition of the cell cycle and growth (He et al., 2007). MiR-34a inhibits SIRT1 (silent information regulator 1), which regulates cellular senescence and longevity, ultimately leading to apoptosis (Yamakuchi et al., 2008). The miR-34 family features in a variety of pathways and cancers as a consequence of its p53-dependent role, but individual members of this family also have several independent roles. MiR-34a plays a negative role in cardiac ageing, fibrosis and cell death following myocardial infarction, in part, by targeting the *PPP1R10* gene which enhances telomere attrition, the DNA damage response and apoptosis (Boon et al., 2013). In contrast, miR-34 has a positive effect on ageing in the brain, aiding in long-term brain integrity in *Drosophila,* where upregulation of miR-34 extends lifespan and mitigates neurodegeneration by a pathogenic polyglutamine protein (Liu et al., 2012a). MiR-34a, but not miR-34b or -c, plays a role in neuronal development by regulating the expression of synaptic proteins like syntaxin-1A in response to the Tap73 transcription factor (Agostini et al., 2011). MiR-34a also represses the long isoform of the tau protein, required for assembly and stability of microtubules, with possible implications for neurodegenerative disease (Dickson et al., 2013). MiR-34b regulates mutliciliogenesis and kidney morphogenesis through the *cmyb* transcription factor (Wang et al., 2013); while miR-34c plays a critical role in bone development and osteoblastogenesis by regulating Notch signalling and bone homeostasis (Bae et al., 2012). MiR-34c also upregulates germ cell-specific genes and functions in the late steps of spermatogenesis in mammals (Bouhallier et al., 2010); and

enables the first zygotic cleavage division in mice by regulation of the Bcl-2 anti-apoptotic regulator (Liu et al., 2012b).

1.4 RNAi AS A THERAPEUTIC TOOL

The miRNA biogenesis pathway has been extensively characterised and can be used to re-direct gene silencing in therapeutic applications. Therapeutic constructs can be introduced at different points in the pathway and are processed accordingly (Figure 1.1, red arrows). These constructs include small interfering RNAs (siRNAs), short hairpin RNAs (shRNAs), long hairpin RNAs (lhRNAs) and pri-miRNA mimics. RNAi-based constructs have been successfully used to target a number of diseases and have entered clinical trials for age-related macular degeneration (AMD), asthma, various types of cancer and infections caused by the respiratory syncytial virus (RSV), hepatitis C virus (HCV) and the human immunodeficiency virus (HIV) (reviewed in (Burnett and Rossi, 2012; Li et al., 2014b)). Pri-miRNA mimics have emerged as effective, flexible and safe mediators of RNAi in sustained gene silencing applications, with the potential for silencing multiple targets in a combinatorial RNAi approach. Pri-miRNA mimics based on the well-characterised pri-miR-30 backbone have been used in a number of therapeutic studies, including those targeting the hepatitis B virus (HBV) (Ely et al., 2009), HIV (Boden et al., 2004), Huntington's disease (Monteys et al., 2015), spinocerebellar ataxia (Ramachandran et al., 2014) and clinical trials against various types of cancer (Senzer et al., 2012). However, pri-miRNA mimics have thus far relied on only a handful of natural miRNA precursors and the use of hundreds of other precursors remains largely unexplored. The efficacy of mimics is dependent on both structural and sequence features (McManus et al., 2002) (section 1.3.3 & 1.3.4) and it is feasible that novel mimics derived from other exceptional precursors may have enhanced function and therapeutic potential. The characterisation of novel pri-miRNA mimics in this project may provide improved therapeutic efficacy for a number of applications and expand the current repertoire of effective pri-miRNA mimics (section 1.5). RNAi-based therapies have shown much promise in the field of anti-HIV gene therapy (Swamy et al., 2016) and pri-miRNA mimics are especially well-suited for this application (Chung et al., 2014; Chung et al., 2012), which requires long-term, multi-targeting to successfully inhibit viral replication and prevent the rapid emergence of escape mutants. The development of various RNAi constructs and the benefits of pri-miRNA mimics are therefore discussed further in the context of anti-HIV therapeutic applications.

1.4.1 RNAi to target HIV

HIV infection and acquired immune deficiency syndrome (AIDS) remain a global pandemic with about 36.9 million people worldwide living with the disease (UNAIDS 2014 statistics: http://aidsinfo.unaids.org/, accessed 20 February 2016). The burden of this pandemic rests in sub-Saharan Africa with approximately 70 % of HIV-infected individuals residing in this region. In South Africa alone, there were 340 000 new infections and 140 000 AIDS-related deaths reported in 2014. Anti-retroviral drug regimens like HAART (highly active antiretroviral gene therapy) have made a significant contribution to disease progression and life extension. It is estimated that around 300 000 deaths were averted in South Africa during the period of 2010 - 2014 through the use of antiretroviral therapy (ART). However, the side effects of ART drugs can be severe, including cardiovascular disease, hepatotoxicity, anaemia, lactic acidosis, gastrointestinal disorders and toxic epidermal necrosis (HHS, 2016). These side-effects may contribute to a lack of patient adherence to drug regimens with a subsequent loss of virological control and the emergence of drug resistance. Drug resistance is common and a major concern in the success of ART. In the period between 1998 and 2015, resistance to tenofovir was highest in sub-Saharan Africa, occurring in ~ 57 % of patients with virological failure in response to first-line tenofovir-containing ART (TenoRes Study, 2016). Therefore, in spite of the current success of HIV treatments, there is room for improvement.

Anti-viral gene therapies may satisfy the need for continuous treatment without the burden of life-long patient adherence. An anti-HIV gene therapy could be administered as a once-off alternative treatment for long-term viral suppression or as an adjuvant to existing therapies (Scherer et al., 2007a). RNAi-based strategies against HIV are particularly promising and have been used to target the expression of specific HIV-1 genes and regions of the HIV-1 genome (Figure 1.5). Many details on the structure, function and life cycle of HIV-1 have been elucidated since the isolation of HIV (Barre-Sinoussi et al., 1983; Popovic et al., 1984) and are described in numerous other reviews, including (Engelman and Cherepanov, 2012; Freed, 2001, 2015). RNAi-based strategies may also target HIV-dependency factors (HDFs) (Brass et al., 2008) of the host cell that are utilised by the virus in the course of infection. Cellular host receptors required for viral entry are commonly targeted, including the CD4 (cluster of differentiation 4) receptor (Bour et al., 1995) and the CXCR4 (C-X-C motif receptor 4) or CCR5 (C-C motif receptor 5) chemokine co-receptors (Alkhatib et al., 1996; Deng et al., 1996; Dragic et al., 1996). Several different non-expressed and expressed therapeutic constructs (Figure 1.6) have been investigated for gene silencing purposes and are discussed in more detail with respect to HIV inhibition.

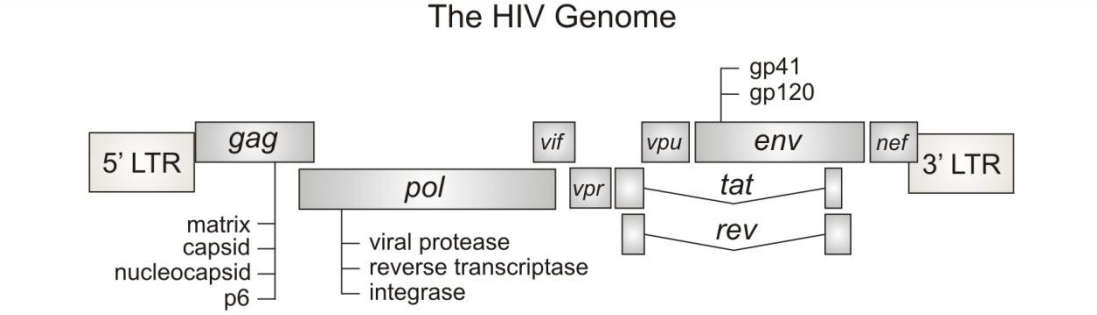


Figure 1.5: RNAi to target the HIV genome.

The ~ 9.2 kb HIV genome (Alizon et al., 1984; Wain-Hobson et al., 1985) encompasses nine genes which code for fifteen viral proteins, including enzymes and structural, regulatory and accessory proteins. HIV produces three polyproteins (Gag, Pol, Env) which are processed into individual functional proteins by the cellular or viral protease. LTR: long terminal repeat; *gag*: group specific antigen; *pol*: polymerase; *vif*: viral infectivity factor; *vpr*: viral protein R; *vpu*: viral protein U; *tat*: trans-activator of transcription; *rev*: regulator of expression of virion proteins; *env*: envelope; gp: glycoprotein; *nef*: negative factor. This diagram was adapted from (Saayman et al., 2008).

1.4.2 Small interfering RNAs (siRNAs) and artificial miRNAs

The potential use of synthetic siRNAs for novel therapeutic applications and gene function studies was identified early on in the field of RNAi. Specific silencing of target genes by synthetic 21 nt siRNA duplexes was demonstrated in *Drosophila* (Elbashir et al., 2001b) and in human cells (Caplen et al., 2001; Elbashir et al., 2001a). This work showed that siRNA function can be uncoupled from dsRNA precursors (Elbashir et al., 2001b), making siRNAs a powerful tool for transient gene silencing. SiRNAs function like miRNAs in the mammalian miRNA biogenesis pathway and associate with RISC to mediate cleavage or translational repression of a target mRNA, depending on the level of complementary guide-target pairing (Doench et al., 2003; Zeng et al., 2003) and the Ago identity. However, gene silencing approaches typically rely on fully complementary siRNAs. Synthetic siRNA or miRNA duplexes can be easily introduced into the cell by transfection with cationic liposomes to mediate transient gene silencing. Benefits of a synthetic siRNA approach include potent and specific gene silencing, ease of synthesis and evasion of non-specific effects commonly observed for longer dsRNA (Caplen et al., 2001; Elbashir et al., 2001a).

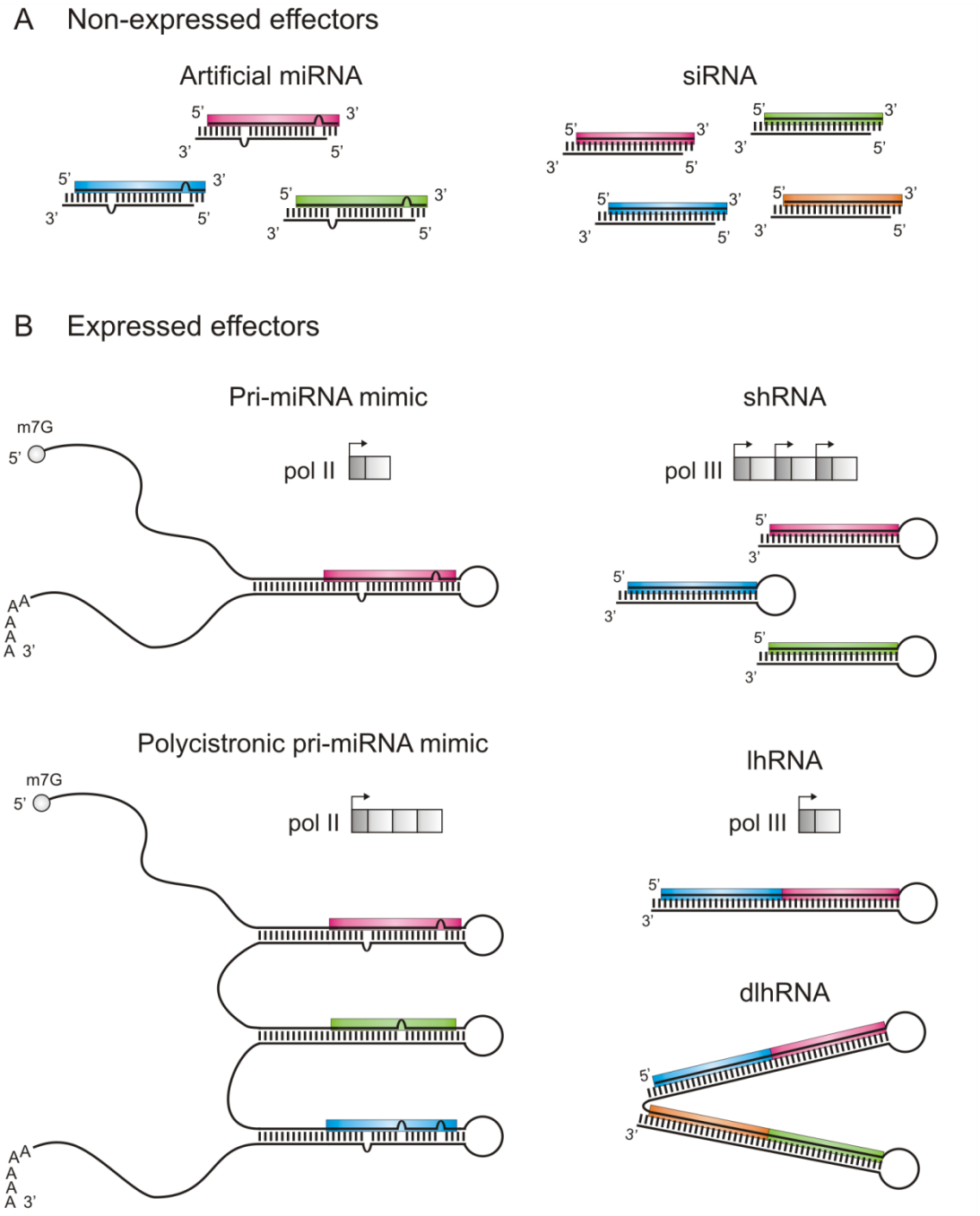


Figure 1.6: Non-expressed and expressed therapeutic constructs.

A schematic diagram indicating various (**A**) non-expressed and (**B**) expressed RNAi activators derived from pol II or pol III cassettes. ShRNA: short hairpin RNA; lhRNA: long hairpin RNA; dhlRNA: double long hairpin RNA.

Synthetic siRNAs (Figure 1.6 A) have been used to successfully inhibit HIV-1 expression and replication. In a typical efficacy assay in cell culture, RNAi activators are transfected or transduced into cell lines expressing the CD4 receptor and CCR5 or CXCR4 co-receptors, such as HeLa-derived Magi cells (Kimpton and Emerman, 1992) or U87MG cells (Ponten and Macintyre, 1968), before or after infection with HIV. Common laboratory HIV strains that utilise the CCR5 co-receptor are designated as R5-tropic / macrophage-tropic, such as the $HIV_{Ba\text{-}L}$ strain (Gartner et al., 1986); while strains that utilise the CXCR4 co-receptor are X4-tropic / T cell-tropic (Berger et al., 1998), such as the pNL4-3 infectious molecular clone (Adachi et al., 1986). SiRNAs targeting the *tat* and *rev* transcripts successfully inhibited the expression and function of corresponding regulatory proteins and inhibited HIV replication in $CD4^+$ $CCR5^+$ HEK293T cells, T cells and primary lymphocytes in short term cell culture studies (Coburn and Cullen, 2002). SiRNAs against the gag polyprotein (p24 siRNA) reduced p24 expression by at least four-fold, as indicated by p24 ELISA and flow cytometry, with a 10-fold reduction in p24-containing RNAs indicated by northern blot analysis (Novina et al., 2002). Fully complementary siRNAs targeting expression from the *vif*, *nef* and LTR regions of the genome reduced viral production by 30 to 50-fold in Magi cells 24 h after infection (Jacque et al., 2002).

SiRNAs targeting cellular receptors have also been effective. SiRNA-mediated inhibition of the CD4 cellular receptor reduced viral entry and replication by four-fold, as indicated by a viral entry reporter assay (Magi β-gal) and p24 ELISA 48 hours after infection (Novina et al., 2002). However, targeting of the primary CD4 receptor may be limited by its role in normal immune function and the virus itself causes CD4 degradation (Vpu, Nef) during the replicative cycle, suggesting that targeting of CD4 is not ideal. HIV-1 co-receptors are therefore more attractive targets, especially CCR5 because homozygous mutations of this gene are protective (Liu et al., 1996; Samson et al., 1996). SiRNAs against CCR5 and CXCR4 successfully reduced co-receptor expression and HIV entry of R5-tropic and X4-tropic strains, respectively (Martinez et al., 2002b). Other cellular factors have also been targeted, like NF-κB (nuclear factor-κB) which binds to the 5' LTR and activates viral gene expression (Surabhi and Gaynor, 2002). SiRNA-mediated silencing of NF-κB inhibited viral replication for up to 6 days post-infection, but was less effective than a siRNA targeting *tat*.

Synthetic siRNAs mediate potent silencing, but this effect is transient. After the transfection of synthetic siRNAs in infected T cells, levels of the targeted viral mRNA returned to normal by day 4 after transfection and levels of the targeted protein returned to normal by day 9 after transfection (Novina et al., 2002). SiRNAs can therefore be derived from expression cassettes for constitutive gene silencing. Previous strategies to drive the expression of small therapeutic

RNAs from a U6 snRNA or tRNA pol III promoter (Bertrand et al., 1997; Good et al., 1997) were adopted for siRNA expression. Earlier vector-based siRNA expression cassettes used two U6 promoters to drive independent expression of 19 - 21 nt sense and anti-sense strands of the siRNA duplex in tandem (Lee et al., 2002a; Miyagishi and Taira, 2002; Yu et al., 2002). Expressed siRNAs formed staggered RNA duplexes with short 3' overhangs, suitable for RISC incorporation and gene silencing. The U6 termination signal of each strand (4 or 5 thymidine residues) (Bogenhagen and Brown, 1981) generated appropriate 3' uridine overhangs.

Expressed siRNAs were used to effectively silence viral expression. SiRNAs targeting *tat* and *rev* effectively inhibited replication in a 4 day assay (Lee et al., 2002a), while an expressed siRNA targeting *gag* inhibited HIV expression for up to 6 days (Capodici et al., 2002). The use of an expression vector also enables the incorporation of a reporter gene for the selection of transfected cells (Amarziguioui et al., 2005), thus addressing issues of varied transfection efficiency and dose-dependency observed for synthetic siRNAs in different cell lines (Novina et al., 2002). Expression strategies for siRNAs now predominantly rely on hairpin-based formats (section 1.4.3 & 1.4.4), while commercially available synthetic siRNA and miRNA duplexes are widely used in transient gene function studies in cell culture. To improve siRNA function, design criteria were investigated by several groups (Amarzguioui and Prydz, 2004; Hsieh et al., 2004; Reynolds et al., 2004; Ui-Tei et al., 2004) and common properties of effective siRNAs were identified, including a 5' U residue and reduced GC pairing at the 5' terminus. SiRNAs can also be chemically modified to enhance binding affinity, decrease ribonuclease susceptibility and reduce immunogenicity (Behlke, 2008; Burnett and Rossi, 2012) (section 1.4.6).

1.4.3 Short hairpin RNAs (shRNAs)

ShRNAs were initially developed as a facile way to achieve simultaneous, equivalent and stable expression of both strands of a siRNA duplex from the same construct. In a shRNA construct, the two strands (19 - 21 nt) of a siRNA duplex are connected by a short stretch of nucleotides referred to as a spacer or linker. Upon expression, the shRNA transcript folds back on itself to form a hairpin in which the siRNA duplex forms a stem with 5p (5') and 3p (3') arms and the linker forms a terminal loop (Figure 1.6 B). Well characterised U6 or H1 pol III promoters which have a precise starting position for transcription (G) and a precise termination signal (4 or 5 thymidine residues) were adopted for shRNA expression in mammalian cells by several groups (Brummelkamp et al., 2002; Jacque et al., 2002; McManus et al., 2002; Paddison et al., 2002; Paul et al., 2002; Sui et al., 2002; Yu et al., 2002). The pol III termination signal produces a 3' U

overhang beyond the duplex, which is trimmed to between 1 and 4 U residues (Paul et al., 2002). This satisfies the overhang requirement for Dicer processing and protects against further 3' to 5' exonuclease degradation (Good et al., 1997). Like pre-miRNAs, the shRNA terminal loop is cleaved by Dicer (Paddison et al., 2002) to produce a staggered RNA duplex suitable for incorporation into RISC.

Several effective vector expression systems have been described, including pSuper (H1 promoter) (Brummelkamp et al., 2002), pShh-1 (U6 promoter) (Paddison et al., 2002) and pSLOOPIII (H1 promoter) (McManus et al., 2002). A range of terminal loop sizes have also been used successfully, but the use of a 4 nt tetraloop (Paddison et al., 2002; Paul et al., 2002) or 9 nt terminal loop (Brummelkamp et al., 2002; Castanotto et al., 2002) are common. Large terminal loops derived from pre-miRNAs have also been used, including that of pre-miR-26b (Miyagishi et al., 2004) and pre-miR-22 (Gu et al., 2012). The inclusion of three or four G:U wobbles in the passenger strand of a shRNA can improve DNA sequencing without affecting RNAi efficacy (Miyagishi et al., 2004). The development of a facile PCR-based strategy to generate shRNA constructs with U6 promoter enabled widespread use and testing of shRNAs in mammalian cells (Castanotto et al., 2002).

Several effective shRNAs have been developed against HIV. ShRNAs targeting *vif* were expressed using a T7 bacteriophage promoter and suppressed viral expression by 20 to 30-fold in Magi cells 24 h after infection (Jacque et al., 2002). In a large-scale study, shRNAs were designed against 96 different target sequences of the HIV-1 genome selected on the basis of sequence conservation and siRNA design criteria (McIntyre et al., 2009a). Upon expression from an H1 promoter, 65 of these shRNAs were found to be highly effective and reduced viral replication by more than 75 % at 2 days following transfection. However, shRNAs are only effective in long-term viral suppression as part of combinatorial approaches (section 1.4.5) and when stably expressed from viral vectors.

1.4.4 Pri-miRNA mimics

SiRNAs can also be derived as exogenous or synthetic miRNAs from expressed pri-miRNA mimics (Figure 1.6 B). Pri-miRNA mimics are modelled after natural pri-miRNAs and are processed by both Drosha-DGCR8 and Dicer to mediate gene silencing. Pri-miRNA mimics were initially designed by replacing the natural miRNA and corresponding miRNA* of an endogenous pri-miRNA with a 19 bp siRNA duplex (Zeng et al., 2002). A pri-miR-30 mimic

expressed from the cytomegalovirus (CMV) pol II promoter was found to effectively silence the expression of a target gene in mammalian cells (Zeng et al., 2002). In a similar approach, pri-miRNA mimics were created by attaching the loop of pre-miR-26a to one side of a 19 bp siRNA duplex and a ~ 5 bp clamp with an asymmetrical bulge to the other side (McManus et al., 2002). This artificial mimic successfully downregulated gene expression when expressed from an H1 expression cassette (pSLOOPIII), with comparable efficacy to that of a shRNA or siRNA. Both types of pri-miRNA mimic were processed into smaller siRNAs fully complementary to the target mRNA and mediated mRNA degradation. The pri-miR-30 scaffold has since been widely used for pri-miRNA mimic design. In a large-scale study, the pri-miR-30 scaffold was used to generate a library of more than 140 000 mimics (sh-miRs) targeting a large proportion of the human and mouse genomes for gene function studies (Silva et al., 2005).

The function of pri-miRNA mimics was compared with that of shRNAs in several studies, yielding some conflicting results. SiRNAs were about 12 times more abundant when derived from a pri-miRNA mimic than a standard shRNA in a U6 expression format (Silva et al., 2005). A pri-miR-30-based mimic targeting HIV *tat* was also 80 % more effective at inhibiting HIV replication than a corresponding shRNA (Boden et al., 2004). In contrast, shRNAs with a 9 nt loop and 19 bp stem were generally more effective than pri-miR-30-based mimics in target silencing assays (Li et al., 2007). A more definitive study using standardised features for an accurate comparison of shRNAs and pri-miRNA mimics, found that shRNAs mediated more potent gene silencing in short-term assays *in vitro* and *in vivo* (Boudreau et al., 2008). However, this was found to be detrimental in long-term gene silencing applications, where highly expressed shRNAs can cause cellular toxicity (section 1.4.6).

Mimic design requires a careful consideration as structural features can affect mimic function. Pri-miR-30 mimics that closely resembled the original pri-miRNA were more effective than those with significant alterations (Zhou et al., 2005); and altering the position of an asymmetrical bulge in the stem affected mimic function (McManus et al., 2002). The structural and sequence determinants required for efficient processing of natural pri-miRNAs (section 1.3.3 & 1.3.4) also apply to pri-miRNA mimics. Furthermore, the influence of exceptional structural features has not been fully elucidated. Pri-miRNA mimic design has largely relied on several well-characterised pri-miRNAs as scaffolds and the potential of hundreds of other miRNA precursors has not yet been explored in mimic design. This highlights the need for further investigation and the characterisation of novel pri-miRNA mimics with exceptional basal stem lengths was central to the research objectives of this project (section 1.5).

1.4.5 Combinatorial RNAi

RNAi-based strategies have proved effective at inhibiting HIV gene expression and replication in short-term cell culture studies. However, HIV resistance emerges relatively quickly in response to a single therapeutic entity. In a cell line with stable expression of a shRNA targeting *tat*, HIV replication was suppressed by 95 % for 21 days post-infection, but viral escape mutants with a nucleotide substitution in the *tat* target sequence emerged after 25 days and viral suppression was abrogated by day 35 post-infection (Boden et al., 2003). Similarly, HIV replication was suppressed for 14 days post-infection in T cells stably expressing a shRNA against *nef*, but viral escape mutants with a 106 bp deletion in the *nef* gene including the target sequence emerged after 23 days (Das et al., 2004). Multiple other escape mutants with various deletions and substitutions in the *nef* target emerged from day 43 to day 80 post-infection. These findings suggested that, much like conventional anti-retroviral drug therapies, a combination of several different siRNAs is required for effective long-term suppression. These results also suggested that highly conserved regions of the HIV genome should be targeted.

Multi-shRNAs

The simultaneous use of several different shRNAs is a facile option for targeting multiple sequences. In a multi-shRNA approach, three different shRNAs that target highly conserved sequences in the *gag*, *integrase* and *protease* genes were expressed from a single lentiviral vector and generated an additive inhibitory effect with delayed viral resistance (ter Brake et al., 2006). However, viral escape mutants still emerged, suggesting that additional shRNAs are required. An *in silico* model of HIV resistance in response to shRNA-mediated silencing proposed that at least four active shRNAs are required to inhibit viral replication and prevent resistance, provided that the shRNAs are active prior to viral integration (Applegate et al., 2010). Combinations of six or seven highly active shRNAs can theoretically be assembled to include at least 4 shRNAs in each combination that target all known clade B subtypes and up to 87 % of other subtypes (McIntyre et al., 2011). Another possibility to combat viral resistance is to target the wildtype virus and anticipated escape mutants simultaneously. Targeting of the wildtype *int* sequence and two predominant viral escape sequences effectively inhibited the wildtype virus and blocked the emergence of anticipated viral escape mutants (Schopman et al., 2010). However, targeting the wildtype and viral escape variations of the same sequence simultaneously had a negative effect on initial antiviral activity, suggesting that viral targets should be distinct to avoid competition (Nishitsuji et al., 2006).

Long-hairpin RNAs (lhRNAs)

Several siRNAs can also be simultaneously derived from a single lhRNA (Figure 1.6 B) by consecutive Dicer processing, as observed for long dsRNA, but without eliciting an interferon response (section 1.4.6). LhRNAs were initially designed to target contiguous ~ 50 - 300 nt regions of the HIV genome (Konstantinova et al., 2006; Nishitsuji et al., 2006). In comparison to shRNAs, lhRNAs mediated moderate to comparable gene silencing of the same target region, but were more effective at inhibiting viral escape mutants (Konstantinova et al., 2006; Nishitsuji et al., 2006). LhRNAs targeting 50 or 80 bp of the *tat* / *rev* region increased viral suppression from 35 days to 48 days in comparison to a single, potent shRNA (Sano et al., 2008). LhRNAs were also designed to target non-contiguous regions of the HIV genome as part of a combinatorial approach to silence different viral genes (Liu et al., 2007). Single lhRNAs were designed to include two (Liu et al., 2007) or three (Saayman et al., 2008) potent siRNAs in tandem, against *pol* and *nef* genes or *tat*, *rev* and *vif* genes, respectively. These lhRNAs mediated efficient gene silencing and viral suppression comparable to that of the individual shRNAs in short-term assays, while another lhRNA targeting *nef*, *pol* and *rev* / *tat* inhibited viral replication for up to 40 days in an assay with a high viral dose and up to 49 days in an assay with a low viral dose (Liu et al., 2009).

In lhRNAs targeting both contiguous (Barichievy et al., 2007; Sano et al., 2008) and non-contiguous HIV sequences (Liu et al., 2007; Saayman et al., 2008), the processing and silencing efficacy of siRNAs was found to decrease towards the terminal loop. This effect could be diminished by adjusting the distance between individual siRNAs within the lhRNA stem (Saayman et al., 2008) and a distance of 3 nt between 19 nt siRNAs was preferable (Liu et al., 2007). However, this processing effect limited the number of effective siRNAs within a single lhRNA to three (Liu et al., 2009). To overcome this limit, two dual lhRNAs were connected to form a double lhRNA (dhlRNA) capable of expressing four siRNAs (Saayman et al., 2010). Dual lhRNAs were connected by a 2 nt UU bridge, imitating the typical 3' overhang of hairpin RNAs. Again, siRNAs derived from dlhRNAs had comparable function to that of individual shRNAs in short-term assays, but siRNA efficacy was affected by the order and spacing of siRNA sequences (Saayman et al., 2010). LhRNAs are therefore suitable for generating multiple effective siRNAs in a combinatorial strategy against HIV, but efficient conformations must be determined experimentally.

Polycistronic pri-miRNA mimics

Pri-miRNA mimics are particularly well-suited for a combinatorial RNAi approach as multiple pri-miRNA mimics can be co-expressed and processed from a single polycistronic transcript (Figure 1.6 B), as is common for natural pri-miRNAs derived from miRNA clusters (Liang et al., 2007). Polycistronic pri-miRNA constructs can be created by assembling mimics in tandem or by using natural miRNA clusters for mimic design. Successful expression and processing of an assembled polycistronic cluster was demonstrated when pri-miR-30 and a mimic based on pri-miR-30 were expressed from single pol II- and pol III-driven constructs (Zeng and Cullen, 2003). An artificial tri-cistronic mimic targeting *pol* and *vif* transcripts was assembled using two pri-miR-30 mimics and a pri-miR-155 mimic in tandem (Zhang et al., 2012). This construct was processed correctly to generate three miRNAs and inhibited viral replication more effectively than the individual mimics in a three week challenge assay. The pol II-expressed miR-17~92 miRNA cluster was used to design a construct expressing four miRNAs against *gag*, *pol* and *rev* / *tat* and the 5' untranslated leader (ldr) sequence and found to be more effective at inhibiting viral replication than individual mimics (Liu et al., 2008). Similarly, a polycistronic mimic based on the intronic miR-106b cluster and expressing three miRNAs against *tat* and *rev* transcripts was more effective at silencing gene expression than corresponding pol III-expressed shRNAs (Aagaard et al., 2008) and inhibited viral replication for 6 weeks when expressed from a lentiviral vector (Chung et al., 2014). However, the function of individual mimics in these constructs was not initially optimal and was affected by the identity of the mimic (Aagaard et al., 2008), pairing within the stem (Aagaard et al., 2008), the sequence of artificial flanking sequences (Zhang et al., 2012) or the order of individual mimics (Liu et al., 2008; Zhang et al., 2012). This highlights the need for individual anti-HIV mimics with optimal function and more comprehensive rules for pri-miRNA mimic design, both of which are addressed in this project (section 1.5).

1.4.6 Safety concerns of an RNAi modality

Several issues regarding safety and efficacy have been raised in the development of therapeutic RNAi activators. SiRNAs can have various undesirable effects in the mammalian cell, collectively described as off-target effects (OTEs). These effects can be characterised as specific (off-target silencing) or non-specific (interferon response, saturation) responses, but may also occur concurrently. OTEs are described here, including several approaches and modifications that can be used to avoid or reduce the impact of these effects. In particular, pri-miRNA mimics expressed from a lentiviral vector have emerged as a safe approach for RNAi

activation, without eliciting an interferon response or saturation of the endogenous miRNA biogenesis pathway.

Off-target silencing

Specific silencing of a target mRNA by a siRNA or miRNA is largely dictated by perfect complementary pairing with the seed region (section 1.3.9). However, a degree of silencing may also occur for transcripts with partial identity to the intended target, thus creating unintended, sequence-dependent OTEs. A genome wide microarray analysis of siRNA activity in cell culture revealed that potent siRNAs can mediate degradation of off-target transcripts with as few as 11 contiguous pairing interactions and that both strands of a siRNA duplex can affect off-target gene expression (Jackson et al., 2003). The overexpression of a single miRNA can also have dramatic effects on the global gene expression, as illustrated by the overexpression of a miRNA required for cell type maintenance which affected about 100 different messages (Lim et al., 2005). A variety of potential off-targets exist, as G:U wobbles and three to four mismatches are tolerated in guide-target pairing for translational repression and functional target sites are not limited to the 3' UTR (Saxena et al., 2003). However, further analyses revealed that miRNA targets and off-targets are determined in the same manner. Off-target silencing was primarily mediated by perfect matches between a hexamer (2 - 7) or heptamer (2 - 8) seed region and sequences in the 3' UTR of target genes, but not those in the 5' UTR or open reading frame (ORF) (Birmingham et al., 2006; Jackson et al., 2006b; Lin et al., 2005). As for typical miRNAs, not all potential 3' UTR seed-targets were silenced, suggesting that other features can influence the extent of off-targeting (Birmingham et al., 2006; Jackson et al., 2006b). OTEs were also independent of the mode of expression when comparing siRNAs and shRNAs, suggesting that the siRNA sequence was the major determinant of off-target repression (Jackson et al., 2006b).

Off-target silencing is of particular concern in a therapeutic setting, but may be reduced by several siRNA modifications. A balance between effective on-target silencing and reduced off-target silencing must be achieved in any successful modification strategy. Single mismatches in the seed (2 - 8) or supplementary (13 - 16) region abolished off-target silencing with only a reduction in on-target silencing (Jackson et al., 2003). However, a change in the seed region also generated a different subset of off-target transcripts (Jackson et al., 2006b). Single mismatches may therefore be useful to avoid off-target silencing of particular genes, but sequence design is important for a more general improvement. An analysis of all possible 4 096 hexamer seed target sequences revealed a non-uniform distribution in the 3' UTR transcriptome,

indicating that certain siRNA seed sequences have a low frequency of complementary targets and their use will lead to fewer OTEs (Anderson et al., 2008). SiRNAs can also be chemically modified to reduce off-target silencing. A 2'-O-methyl (2'OMe) ribosyl substitution at position 2 of the antisense (guide) siRNA reduced the frequency and magnitude of off-target silencing, without significantly reducing efficacy or creating an additional off-target signature (Jackson et al., 2006a). Other modifications to the backbone of the seed region also reduced off-target silencing, including a locked nucleic acid (LNA) at position 3 and an unlocked nucleic acid (UNA) at position 7 (Bramsen et al., 2010). Target specificity may be further improved by the use of an abasic deoxynucleotide spacer substitution at position 6 of the siRNA (Lee et al., 2015), which makes use of a "pivot pairing rule" for bulged miRNA - mRNA interactions (Chi et al., 2012). Targeting by the passenger strand can also be avoided by segmenting the passenger strand into two pieces and creating a three-stranded, small internally segmented interfering RNA (sisiRNA) (Bramsen et al., 2007). Synthetic guide modifications are limited to transient siRNA strategies and can not be implemented for expressed RNAi activators. Some level of off-target silencing is therefore intrinsic for such an approach, but target prediction (Alexiou et al., 2009) and design algorithms (Anderson et al., 2008) can be used with experimental approaches to identify and avoid potential off-targets with more severe phenotypic effects.

Interferon response

The introduction of dsRNA into a mammalian cell can trigger the innate immune system and a type I interferon (IFN) response (reviewed in (Sadler and Williams, 2008; Stark et al., 1998)). This response functions as a first-line defence mechanism against viral infections, leading to RNA degradation, inhibition of translation and apoptosis. IFN signalling pathways involve numerous components, regulatory mechanisms and positive feedback loops for the amplification of the IFN response. DsRNA, a common intermediate of viral replication, is recognised by the dsRNA-dependent protein kinase R (PKR) (Clemens, 1997), the RIG-I (retinoic acid-inducible gene 1) cytoplasmic receptor (Yoneyama et al., 2004) and the membrane-bound TLR-3 (Toll-like receptor 3) (Alexopoulou et al., 2001). These proteins signal the expression of IFN-α and IFN-β through transcription factors like NF-κB and IRF3 (interferon regulatory factor 3). IFN activates signalling cascades, including the Jak-STAT (Janus kinase - signal transducer and activator of transcription) pathways (Darnell et al., 1994), which signal the upregulation of hundreds interferon-stimulated genes (ISGs) (Schoggins et al., 2011) through transcription factor complexes including ISGF3 (interferon-stimulated gene factor 3) (Fu et al., 1990). ISGs include the oligoadenylate synthetase (*OAS*) locus (Hovnanian et al., 1998), which codes for 2'-5'

oligoadenylate synthetases (OAS1, 2, 3) that function in the 2'-5' A RNA degradation pathway (Player and Torrence, 1998). These enzymes are also activated by direct binding of dsRNA (Hartmann et al., 1998) and synthesise oligoadenylates from ATP, which in turn activate RNAse L to catalyse the general degradation of ssRNA and mRNA. PKR primarily functions in the phosphorylation of EIF2α (eukaryotic translation initiation factor 2 α), resulting in the general inhibition of protein synthesis (Clemens, 1997).

It was initially assumed that small RNAs entering the mammalian miRNA biogenesis pathway were too short to induce an interferon response, as components of the IFN pathway were not typically activated by dsRNA shorter than 30 bp (Manche et al., 1992; Minks et al., 1979). However, a non-specific, dose-dependent IFN response was detected for cells transfected with siRNA duplexes and attributed to activation of the dsRNA-dependent PKR and the Jak-STAT signalling pathway, resulting in a global upregulation of ISGs (Sledz et al., 2003). Similarly, siRNAs and shRNAs were found to induce type I (α, β) IFN, in part, through TLR3 signalling (Kariko et al., 2004). These results were later attributed to the *in vitro* transcription of siRNAs and shRNAs using a T7 bacteriophage RNA polymerase, which adds a 5' triphosphate to transcribed RNAs and thereby stimulates IFN induction (Kim et al., 2004). A subsequent study using only purified, synthetic ~ 21 nt siRNAs failed to detect an IFN response and found that naked siRNAs were well-tolerated in mice (Heidel et al., 2004). However, sequence-dependent immunostimulatory effects were observed for single siRNAs with particular motifs. GU-rich ssRNA motifs are recognised by TLR7 and TLR8 (Diebold et al., 2004; Heil et al., 2004) and thereby induce IFN-α, NF-κB target genes and inflammatory cytokines like TNF-α (tumour necrosis factor α) (Hornung et al., 2005; Judge et al., 2005; Sioud, 2005). The effects of these immunostimulatory RNA (isRNA) motifs were abrogated by avoiding certain motifs (Judge et al., 2005) or the selective incorporation of LNA (Hornung et al., 2005) or 2'OMe (Judge et al., 2006) chemical modifications. The length of a siRNA duplex can also determine IFN induction, as dsRNA greater than 23 bp triggered an IFN response through TLR3 in a cell type-specific manner (Reynolds et al., 2006). Therefore, several factors may contribute to siRNA-mediated IFN induction, including the siRNA sequence, siRNA concentration, cell type and method of siRNA preparation (Heidel et al., 2004); but immunostimulatory effects can be prevented by careful consideration of these factors.

An interferon response was also observed for lentiviral and plasmid-derived shRNAs expressed from a pol III promoter (Bridge et al., 2003), but this was later attributed to deviation from the wildtype promoter sequence and the presence of an AA dinucleotide near the TSS (Pebernard and Iggo, 2004). As for dsRNA, the length of the duplex can also contribute to IFN induction by

shRNAs and lhRNAs. *OAS1* induction was observed for shRNAs with a 21 bp duplex, but not those with a 19 bp duplex when expressed from a lentiviral vector (Fish and Kruithof, 2004). In contrast, shRNAs with a 19 bp stem were found to be more immunostimulatory than those with a 14 bp stem (Pebernard and Iggo, 2004). Amidst conflicting reports on the safe length of shRNAs, a more widely applicable solution was identified to avoid IFN induction. The introduction of G : U wobbles in the sense strand of a hairpin every 4 - 8 bp was found significantly reduce an IFN response for shRNAs and lhRNAs of up to 100 bp in length (Akashi et al., 2005). In addition, these modifications enabled greater stability in *E. coli*-based cloning procedures and easier DNA sequencing. No IFN response was detected for shRNAs or lhRNAs designed with G : U modifications (Liu et al., 2009; Saayman et al., 2010; Sano et al., 2008). Hairpins with G : U modifications were also found to avoid non-specific gene-silencing effects when expressed from a pol III vector, but not when synthesised *in vitro* (Akashi et al., 2005), suggesting an advantage for expressed constructs. Furthermore, an IFN response was avoided for constructs expressed from a lentiviral vector by lowering the vector dose, without a significant reduction in the efficacy of gene silencing (Fish and Kruithof, 2004). Expressed pri-miRNA mimics are therefore a promising option for safe, expressed RNAi constructs, as they can be expressed from an endogenous promoter sequence and contain various mismatches and G : U wobbles, thus preventing the IFN response observed for foreign dsRNA.

Saturation of the endogenous miRNA biogenesis pathway

The effective expression of RNAi constructs is important for successful gene silencing, especially in the treatment of HIV where untargeted viral reservoirs are expected to allow for viral propagation, mutation and resurgence (Leonard and Schaffer, 2005). However, high-level expression of shRNAs can be detrimental in long-term applications and result in cellular toxicity. ShRNAs targeting CCR5 were effectively expressed from the U6 pol III promoter in primary human lymphocytes, but caused cytotoxicity and the gradual decline of transduced cell populations (An et al., 2006). Similarly, the long-term expression of shRNAs from U6 pol III constructs in AAV vectors, led to dose-dependent liver injury and death in adult mice (Grimm et al., 2006). Further investigation suggested that high-level shRNA expression resulted in the saturation of the endogenous miRNA biogenesis pathway and a subsequent downregulation of liver-specific miRNAs, cellular toxicity and liver failure. Synthetic siRNAs and expressed shRNAs were also observed to compete with each other, as well as endogenous miRNAs, for transport and incorporation into RISC (Castanotto et al., 2007). Global changes in gene expression in response to an exogenous siRNA also include gene upregulation, which may indicate dose-

dependent, de-repression of miRNA targets related to saturation of the miRNA pathway (Khan et al., 2009). Exp-5 and the Ago proteins were found to be readily saturated, rate-limiting factors responsible for RNAi efficacy and cellular toxicity (Grimm et al., 2006; Grimm et al., 2010).

Several approaches have been adopted to reduce saturation of the endogenous miRNA pathway and cellular toxicity. Over-expression of Exp-5 and Ago2 increased shRNA efficacy (Diederichs et al., 2008; Yi et al., 2005) and exogenous co-expression of both of these proteins with U6-expressed shRNAs reduced hepatotoxicity (Grimm et al., 2010). However, the use of multiple expression vectors is best avoided in a therapeutic setting and it is easier to adjust the dose or type of RNAi constructs. The use of shorter 19-mer shRNAs and a lower dose of the AAV vector prevented liver failure and death in mice in a 4 month challenge assay against HBV (Grimm et al., 2006). The expression of shRNA constructs from the weaker H1 pol III promoter also enabled the survival of transduced primary human lymphocytes without major cytotoxic effects (An et al., 2006). Similarly, the use of weaker H1 or 7SK pol III promoters for shRNA expression alleviated toxicity *in vivo* and permitted effective target silencing in the mouse liver for more than a year (Grimm et al., 2010). A weaker, inducible pol II promoter can also address issues of cellular toxicity and enable tissue-specific expression of shRNAs (Giering et al., 2008). Most notably, the use of pri-miRNA mimics instead of shRNAs can avoid saturation of the miRNA biogenesis pathway. Pri-miRNA mimics appear to be processed more efficiently in the pathway, thus providing maximal target silencing without an accumulation of unprocessed RNA intermediates (Boudreau et al., 2009; Boudreau et al., 2008) or associated cellular toxicity (McBride et al., 2008).

Overall, pri-miRNA mimics have a number of desirable features in comparison to other RNAi modalities for therapeutic gene-silencing applications. Several effective pri-miRNA mimics can be expressed from a single pol II promoter, which reduces the likelihood of an IFN response, cellular toxicity and saturation of the miRNA biogenesis pathway. The use of an inducible pol II promoter (Lo et al., 2007; Stegmeier et al., 2005; Zeng et al., 2005a; Zhou et al., 2005) can also provide regulated and tissue-specific RNAi activation. A polycistronic pri-miRNA mimic expressed from a single promoter enables multi-targeting, but also avoids recombination and deletion issues associated with repeat sequences in a lentiviral vector expression system (McIntyre et al., 2009b; ter Brake et al., 2008). Pri-miRNA mimics are therefore ideal to mediate effective and safe combinatorial RNAi in long-term gene silencing applications, such as that required for the treatment of HIV-1 infection. The characterisation of novel pri-miRNA mimics is therefore promising and is explored further in this project.

1.5 THESIS OBJECTIVES

Primary microRNA (pri-miRNA) mimics are well suited for mediating RNAi in sustained gene silencing applications and have been shown to avoid side effects associated with more potent RNAi activators. Effective gene silencing by pri-miRNA mimics has also been demonstrated for several diseases and infectious pathogens, including HBV and HIV-1 infection. Despite this success, mimic design has thus far relied on only a handful of natural pri-miRNA scaffolds, including the model pri-miR-30 scaffold. The potential of ~ 1 900 other known human precursors as scaffolds, and in particular, those with exceptional structural features, has not been fully explored in mimic design. Variation in the exact length and conformation of the predicted basal stem is especially common for functional precursors, but may conceivably affect the efficiency of Drosha-DGCR8 processing. The major aim of this project was therefore to explore the use of natural pri-miRNA scaffolds with exceptional predicted stem lengths in mimic design; and from this, to determine whether a particular basal stem length was beneficial to pri-miRNA function or of use in the optimisation of corresponding mimics. This work was intended to contribute to the current understanding of pri-miRNA processing and the development of effective pri-miRNA mimics for future therapeutic applications.

Specific objectives were to:

1) Identify natural pri-miRNAs with exceptional basal stem lengths

2) Examine the processing and target silencing function of selected pri-miRNAs

3) Design corresponding pri-miRNA mimics with exceptional basal stem lengths

4) Examine the processing and target silencing function of pri-miRNA mimics

5) Ascertain whether a particular predicted stem length was beneficial to pri-miRNA or mimic function

6) Propose a more comprehensive set of rules for the design of effective pri-miRNA mimics based on natural scaffolds

CHAPTER TWO

Functional RNA guides derived from naturally occurring pri-miRNAs with exceptional basal stem conformations

2.1 INTRODUCTION

Primary microRNA (pri-miRNA) mimics have been shown to mediate effective gene silencing (Silva et al., 2005; Zeng et al., 2005a; Zeng et al., 2002) and can provide controlled, tissue-specific expression of exogenous miRNAs when expressed from an appropriate polymerase II promoter (Lo et al., 2007; Stegmeier et al., 2005; Zhou et al., 2005). Pri-miRNA mimics are therefore suitable for sustained gene silencing applications, but mimic design has thus far relied on only a handful of precursors (Choi et al., 2015; Chung et al., 2006; Ely et al., 2009; Ely et al., 2008; Huang and Jia, 2013; Liu et al., 2008; Yang et al., 2010b; Zhang et al., 2012). The model miR-30 backbone (Zeng et al., 2005a; Zeng et al., 2002) has been widely used in therapeutic applications against a number of diseases and pathogens, including the hepatitis B virus (HBV) (Ely et al., 2009) and human immunodeficiency virus (HIV) (Boden et al., 2004; Lo et al., 2007; Zhang et al., 2012), and is included in ongoing clinical trials against various cancers (Rao et al., 2010; Senzer et al., 2012). The potential of ~ 1 900 other human precursors (miRBase) (Kozomara and Griffiths-Jones, 2014) and in particular, those with exceptional structural features, has not been fully explored in pri-miRNA mimic design.

Pri-miRNA mimics, like endogenous pri-miRNAs, are processed in the endogenous miRNA biogenesis pathway (section 1. 3) and a precise understanding of processing is crucial in the design of effective mimics. Recognition and processing of pri-miRNAs by the Drosha-DGCR8 complex is modular in nature, where elements of the secondary RNA structure and primary sequence influence processing efficiency to different extents for different pri-miRNAs (Auyeung et al., 2013; Fang and Bartel, 2015). Elements of the secondary RNA structure are crucial for processing by Drosha-DGCR8, while basal, stem and apical sequence motifs can enhance processing (Auyeung et al., 2013; Fang and Bartel, 2015). In productive processing, the Drosha-DGCR8 heterotrimer is orientated correctly through recognition of the basal ssRNA-dsRNA junction and basal UG motif by Drosha; and recognition of the stem, apical ssRNA-dsRNA junction and loop UGU motif by a DGCR8 dimer (Nguyen et al., 2015; Quick-Cleveland et al., 2014). Drosha acts as a molecular “ruler” and cleaves at ~ 11 bp from the ssRNA-dsRNA junction, while DGCR8 enables efficient and accurate cleavage. Non-canonical features within

the stem also contribute to cleavage site selection (Burke et al., 2014; Ma et al., 2013; Starega-Roslan et al., 2011), while structural distortions in proximity to cleavage sites may be required for efficient cleavage (Quarles et al., 2013; Warf et al., 2011). In addition, methylation motifs in the proximity of miRNA genes promote pri-miRNA processing and act, in part, to recruit the HNRNPA2B1 RNA binding protein which interacts with the microprocessor (Alarcon et al., 2015a; Alarcon et al., 2015b). Other accessory proteins also play a role in the processing of specific microRNA precursors (Di Carlo et al., 2013; Pickering et al., 2011; Sakamoto et al., 2009).

The complexity of the current microprocessor model is well suited to explain the processing of a variety of pri-miRNA sequences, provided that a conventional pri-miRNA structure can be adopted. Optimal properties of the secondary RNA structure include a large terminal loop (≥ 10 nt) (Zeng et al., 2005b), moderate basal stem (≥ 8 bp) with an optimal length of ~ 11 bp (Auyeung et al., 2013); and at least 40 nt of flanking sequence, lacking a strong secondary structure, on either side of the pre-miRNA (Chen et al., 2004; Han et al., 2006; Zeng and Cullen, 2005). However, a number of exceptional precursor structures have been predicted for several endogenous miRNA precursors using RNA folding software (Zuker, 2003) and the potential influence of non-canonical structural features has not yet been fully explored. Furthermore, in an artificial expression context, such as that occurring for expressed pri-miRNA mimics, optimal structural features are not necessarily maintained. Variation in the length and conformation of the basal stem is particularly common and could conceivably affect the accuracy and efficiency of cleavage by the Drosha molecular “ruler”. The consequences of basal stem variation have not been fully characterised, especially with regard to the optimal processing of pri-miRNA mimics.

The aims of this chapter were to functionally assess pri-miRNAs with exceptional basal stem conformations and to determine if exceptional features are beneficial or detrimental to pri-miRNA function. Naturally occurring pri-miRNAs with exceptional basal stem conformations were identified and miRNA-mediated silencing and pri-miRNA processing were assessed in cell culture. This work highlighted the potential impact of exceptional stem elements in pri-miRNA function. This chapter also demonstrated the importance of thorough pri-miRNA characterisation prior to mimic design, using both the predicted secondary RNA structure and experimental approaches to identify functional pri-miRNA scaffolds. Several effective pri-miRNAs were identified here and their potential as novel scaffolds in the design of effective pri-miRNA mimics is promising.

2.2 MATERIALS & METHODS

2.2.1Sequence and secondary RNA structure of selected pri-miRNAs

MiRNA precursor stem-loop sequences were accessed from the online miRNA repository miRBase (Release 10.1) (Griffiths-Jones et al., 2008) (http://www.mirbase.org/, accessed February 2008). The Ensembl genome browser (Release 48), (http://www.ensembl.org) (Flicek et al., 2008) was used to extend the genomic co-ordinates by 50 nucleotides (nt) in either direction and incorporate the expected flanking sequence of the full pri-miRNA. Predicted secondary RNA structures were examined using the mfold web server (version 3.5) (Zuker, 2003) with default parameters (37 °C, 5 % sub-optimality). This software utilizes nearest neighbour energy rules to predict secondary RNA structure by free energy minimization according to existing experimental data (Mathews, 1999). Standard extensions of the flanking sequence (5 nt increments) and the vector-derived flanking sequence were used to investigate a range of possible foldings for each pri-miRNA (Appendix A). MiRBase accession codes (Release 21) (Kozomara and Griffiths-Jones, 2014) and updated genomic co-ordinates (Ensembl 2015, Release 81) (Cunningham et al., 2015) are given in Table 2.1.

Table 2.1: Genomic co-ordinates used to extend miRBase precursor sequences.

MiRNA Precursor	MiRBase Accession Number (Release 21)	Genomic co-ordinates (Ensembl 2015, GRCh38, Release 81)
hsa-miR-520b	MI0003155	19: 53 701 227 - 53 701 287 (+)
hsa-miR-198	MI0000240	3: 120 395 668 - 120 395 729 (-)
hsa-miR-147a	MI0000262	9: 120 244 979 - 120 245 050 (-)
hsa-miR-23a	MI0000079	19: 13 836 587 - 13 836 659 (-)
hsa-miR-934	MI0005756	X: 136 550 878 - 136 550 960 (+)
hsa-miR-34a	MI0000268	1: 9 151 668 - 9 151 777 (-)
hsa-miR-513a-2	MI0003192	X: 147 225 826 - 147 225 952 (-)

2.2.2 Generation of pri-miRNA expression plasmids

Design

The extended pri-miRNA precursor sequences were generated through PCR and inserted into the pCI-neo Mammalian Expression Vector (Promega, Madison, WI, USA) to create pri-miRNA expression plasmids. Expression of the pri-miRNA in pCI-neo is driven by the pol II cytomegalovirus (CMV) immediate-early enhancer/promoter, which allows for strong, constitutive expression in a variety of mammalian cell types (Schmidt et al., 1990). Certain pri-miRNA expression vectors were also altered to assess potential promoter-specific effects by replacing the CMV promoter with a simian virus 40 (SV40) enhancer and early promoter region (Benoist and Chambon, 1981), (Fromm and Berg, 1983). The overall cloning process is shown in Figure 2.1.

Construction

Generation of pri-miRNA sequences by PCR

The polymerase chain reaction (PCR) (Mullis et al., 1986) was used to amplify pri-miRNA sequences from genomic DNA or to generate pri-miRNA sequences from synthetic oligonucleotides in a two-step approach, as illustrated in Figures 2.1 A and B, respectively. PCR reactions were performed using a single subunit *Taq (Thermus aquaticus)* DNA polymerase with the following reaction components in a 50 µl volume: 1U *Taq* DNA polymerase, 0.2 mM each dNTP, 1.5 mM $MgCl_2$, 0.4 µM each primer and nuclease-free water (KAPA Taq ReadyMix, KAPA Biosystems Inc, Wilmington, MA, USA). PCR reactions were performed using the Eppendorf Mastercycler Gradient (Eppendorf, Hamburg, Germany) to facilitate multiple annealing temperatures simultaneously. All oligonucleotide primers were designed using the Reverse Complement feature of The Sequence Manipulation Suite (Stothard, 2000) and ordered from Integrated DNA Technologies (IDT, Coralville, IA, USA). Melting temperatures were calculated using OligoAnalyser v3.1 with default parameters and 50 mM Na^+ concentration (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/).

Initially, pri-miRNA sequences were amplified from human genomic DNA using the primers listed in Table 2.2. Genomic DNA was extracted from the Human Embryonic Kidney (HEK) 293-derived 116 cell line, kindly provided by Dr. Abdullah Ely.

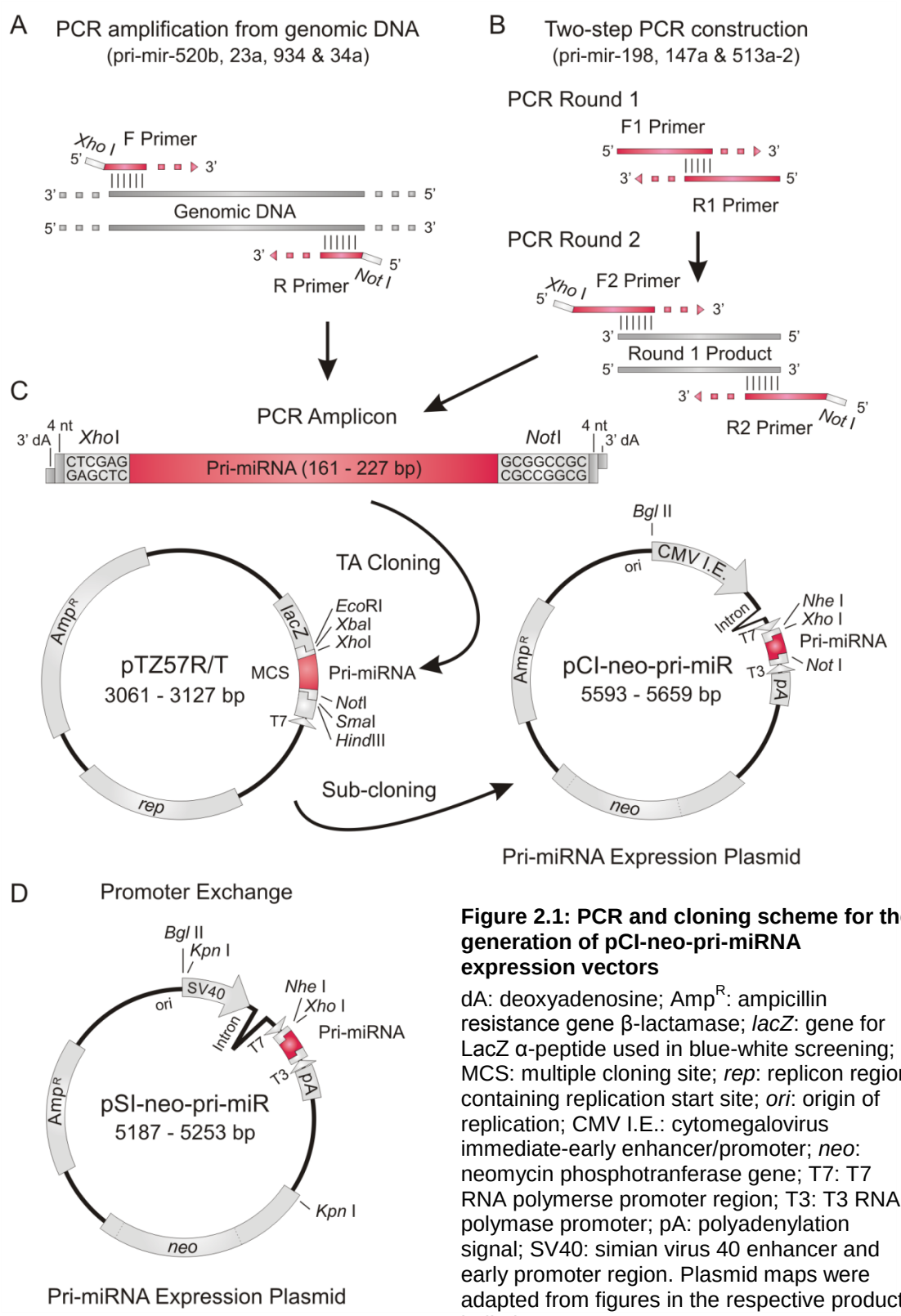


Figure 2.1: PCR and cloning scheme for the generation of pCI-neo-pri-miRNA expression vectors

dA: deoxyadenosine; Amp^R: ampicillin resistance gene β-lactamase; *lacZ*: gene for LacZ α-peptide used in blue-white screening; MCS: multiple cloning site; *rep*: replicon region containing replication start site; *ori*: origin of replication; CMV I.E.: cytomegalovirus immediate-early enhancer/promoter; *neo*: neomycin phosphotranferase gene; T7: T7 RNA polymerse promoter region; T3: T3 RNA polymase promoter; pA: polyadenylation signal; SV40: simian virus 40 enhancer and early promoter region. Plasmid maps were adapted from figures in the respective product bulletins.

Table 2.2: Primers for the amplification of pri-miRNAs from genomic DNA

Primer	Tm (°C)	Sequence (5' to 3' orientation)
Pri-miR-520b F	66.0	*GATC*CTCGAGATTCCAACAAAAAATCCACGGTGCCACAG
Pri-miR-520b R	69.4	*GATC*GCGGCCGCTTATTGCCAAGATCAGCATCCACCTAAACG
Pri-miR-198 F	62.2	*GATC*CTCGAGAGAATGCACTAACCTTTTGATTTGATAAGTTATAAATTCTG
Pri-miR-198 R	69.6	*GATC*GCGGCCGCGGTCATCTTGTTTTTGGCCATCTGACATTTC
Pri-miR-23a F	69.0	*GATC*CTCGAGCAGGTGCCAGCCTCTGGCC
Pri-miR-23a R	75.4	*GATC*GCGGCCGCGCCTCTCTGCCACCCCGTC
Pri-miR-934 F	67.6	*GATC*CTCGAGATCTGGGCCAGCCTTTGATGGTGTG
Pri-miR-934 R	66.1	*GATC*GCGGCCGCTTTTTGAGGCTAAAATGAAACAAATTCTATCTATGAAAACAC
Pri-miR-34a F	68.2	*GATC*CTCGAGGAGAGGCAGGACAGGCCTGTC
Pri-miR-34a R	73.8	*GATC*GCGGCCGCTCCTGGCGTCTCCCACTGGTC
Pri-miR-513a-2 F	64.2	*GATC*CTCGAGACCTATAGGGACTTGAAAAAAAACGGCATGAATAG
Pri-miR-513a-2 R	71.7	*GATC*GCGGCCGCAGGATGCAAAACAGGCATCACCATGCTC

Tm: melting temperature; F: forward primer; R: reverse primer
The initial 4 nucleotide fragment (*GATC*) is shown in italic font.
Flanking *Xho*I (CTCGAG) and *Not*I (GCGGCCGC) restriction sites are underlined.

The PCR programme was as follows: 94 ºC for 5 min; 30 cycles of 94 ºC for 30 s, 62 ºC for 30 s and 72 ºC for 30 s; and a final extension at 72 ºC for 10 min. This approach was successful in the amplification of correct sequences for pri-miR-520b, -23a, -934 and -34a, but sequences obtained for pri-miR-198 and -513a-2 contained several minor variations.

The pri-miR-198 and pri-miR-513a-2 sequences published on miRBase were instead generated by a two-step PCR (Figure 2.1 B) with the extension primers listed in Table 2.3. This approach is an adaptation of the two-step PCR method (Castanotto et al., 2002) used to generate short hairpin RNAs (section 2.2.3). In the first round of PCR, the central portion of each construct was generated using a pair of partially complementary oligonucleotides (F1, R1) which were extended into a complete duplex. The first PCR programme was as follows: 94 ºC for 5 min; 30 cycles of 94 ºC for 10s, 60.5ºC (pri-miR-198), 64.8 °C (pri-miR-147a) or 65.0 ºC (pri-miR-513a-2) for 30 s and 72 ºC for 40s; and final extension at 72 ºC for 7 min. The unique annealing temperatures were estimated at 5 °C below the average melting temperature (Tm) of each primer pair and matched to the closest temperature available on an appropriate gradient setting. This first PCR product was diluted 100 X and 1 µl was used directly as a template in the second round of PCR with another pair of extension primers (F2, R2) to create products with the full pri-miRNA sequence and flanking restriction sites. The second PCR programme was the same as the first, but with unique annealing temperatures as follows: pri-miR-198 (55.5ºC), pri-miR-147a (64.7 °C) and pri-miR-513a-2 (69.7 ºC).

Table 2.3: Extension primers for the generation of pri-miRNAs by a two-step PCR

Primer	Tm (°C)	Sequence (5' to 3' orientation)
Pri-miR-198 F1	67.3	GTTATAAATTCTGTGGTTCTGATCATTGGTCCAGAGGGGAGATAGGTTCCTGTGATTTTTCC
Pri-miR-198 R1	63.2	CTTGTTCTAGTAACAAGATTTCATTTATTCTATAGAGAAGAAGGAAAAATCACAGGAACC
Pri-miR-198 F2	64.2	*GATC*CTCGAGAGAATGCACTAACCTTTTGATTTGATAAGTTATAAATTCTGTGGTTC
Pri-miR-198 R2	69.6	*GATC*GCGGCCGCGGTCATCTTGTTTTTGGCCATCTGACATTTCTTGTTCTAGTAACAAG
Pri-miR-147 F1	69.2	GCAAACACTCCCTGAATCTAAAGACAACATTTCTGCACACACACCAGACTATGGAAGCCA
Pri-miR-147 R1	69.8	GGAAGGAGACCTTCAGAATCTAGCAGAAGCATTTCCACACACTGGCTTCCATAGTCTGGT
Pri-miR-147 F2	68.7	*GATC*CTCGAGACTAGTCCAAATGAAATGTCTTCACTCCTTTCCTTCCATAGAGCAAACACTCCCTGAATC
Pri-miR-147 R2	70.6	GATCGCGGCCGCTAGCAAAATAGAACTCCTCTTCCAGAGGCCATAAATGGAAGGAGACCTTCAGAATC
Pri-miR-147 R2*	70.4	GATCGCGGCCGCAAAATAGAACTCCTCTTCCAGAGGCCATAAATGGAAGGAGACCTTCAGAATC
Pri-miR-513 F1	69.9	CTTGGATGCCACATTCAGCCATTCAGTGTGCAGTGCCTTTCACAGGGAGGTGTCATTTATGTGAACTAAAATATAA
Pri-miR-513 R1	68.3	CAGGGACACCACATATGCAGTGCATGCTGTACATTACCCTTCTCAGAAAGGTGAAATTTATATTTTAGTTCACATAA
Pri-miR-513 F2	69.7	*GATC*CTCGAGACCTATAGGGACTTGAAAAAAAACGGCATGAATAGGGAGCATTTGTTCTTGGATGCCACATTCAG
Pri-miR-513 R2	74.6	*GATC*GCGGCCGCAGGATGCAAAACAGGCATCACCATGCTCGCAGCTACACCCCCATCCTCAGGGACACCACATATG

Tm: melting temperature; F: forward extension primer; R: reverse extension primer.
The added terminal 4 nucleotides (*GATC*) of round two primers are shown in italic font.
Shaded regions of round 1 extension primers are complementary to one another. Shaded regions of round 2 extension primers are complementary to the round 1 PCR product.
Flanking *Xho*I (CTCGAG) and *Not*I (GCGGCCGC) restriction sites of round 2 primers are underlined.
Additional *Spe*I (ACTAGT) and *Nhe*I (GCTAGC) sites in pri-miR-147a round 2 primers are also underlined.
Pri-miR-147 R2* was used in cloning of pSI-neo-pri-miR-147a as it lacks an *NheI* restriction site.

The size and purity of PCR-generated pri-miRNAs was assessed using agarose gel electrophoresis. A 5 µl aliquot of each PCR reaction was resolved on a 1.5 % w/v agarose gel made with 1 X TAE (Tris-Acetate-EDTA) electrophoresis buffer (40 mM Tris-acetate, 2 mM $Na_2EDTA{\cdot}2H_2O$, pH 8.5) (Moore, 2001) and the intercalating agent ethidium bromide (EtBr) (35 µg per 100 ml gel; 10 µg / µl EtBr stock solution). DNA loading dye was added to each sample for tracking purposes (Appendix B.1). A molecular weight marker (O'GeneRuler DNA Ladder Mix, Thermo Fisher Scientific, Waltham, MA, USA) was loaded alongside samples for size estimation. Samples were subjected to electrophoresis for at least 40 min at 80 V (Bio-Rad PowerPac™, Hercules, CA, USA). DNA bands were visualised after exposure to ultraviolet (UV) light, which causes the EtBr stain to fluoresce, using a gel imaging system (G:Box, Syngene, Cambridge, UK) and accompanying software (Gene Snap v. 7.04).

Insertion into a TA cloning vector

PCR-generated pri-miRNAs were initially inserted into the pTZ57R/T TA cloning vector (InsTAclone™ PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA, USA) for convenience (Figure 2.1 C). This vector has single 3' dideoxythymidine (ddT) overhangs at the Multiple Cloning Site (MCS) that allow for direct ligation with PCR products containing complementary 3' deoxyadenosine (dA) overhangs, generated by the terminal transferase activity of *Taq* DNA polymerase (Clark, et. al., 1988). PCR-generated pri-miRNAs were inserted into pTZ57R/T in an overnight ligation reaction at 4ºC with the following components in a 20 µl volume:, 5 U of T4 DNA ligase, 1 X concentration T4 DNA ligase buffer (40 mM Tris-HCl, pH 7.8 at 25°C, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP) (Thermo Fisher Scientific), 1 µl 20 X dilution of second-round PCR reaction, 2 µl (110 ng) pTZ57R/T and nuclease-free water.

A 10 µl aliquot of each ligation reaction was used to transform 100 µl of chemically competent DH5α™ *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA) using heat shock at 42 °C for 90 s (Appendix B.2). The transformation mix was spread on Lysogeny Broth (LB) (Bertani, 1951) agar plates (Appendix B.3) and incubated overnight (~ 16h) at 37 ºC (Labcon incubator, Maraisburg, South Africa) to allow for bacterial replication and colony formation. LB agar plates were substituted with a final concentration of 0.01% w/v ampicillin as a selective agent and IPTG (isopropyl- β -D-thiogalactopyranoside) (40 µl, 100 mM stock) and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (40 µl, 48.9 mM stock) to enable blue-white screening of bacterial colonies. Briefly, the MCS of the pTZ57R/T TA cloning vector is located within the β-galactosidase gene (lacZ) driven by an IPTG-inducible promoter. In the presence of IPTG, lacZ expression will only be disrupted for bacterial colonies harbouring a recombinant plasmid. These bacteria will be unable to produce the β-galactosidase enzyme to metabolise X-Gal and will remain white. Conversely, bacterial colonies harbouring a non-recombinant plasmid will express β-galactosidase and metabolise X-Gal to form an insoluble blue product. Sterile technique was used for all procedures involving bacterial cell culture.

Five white colonies were selected from each agar plate for purification and screening of recombinant pTZ plasmids. Each colony was used to inoculate 5 ml LB broth with a final concentration of 0.01% w/v ampicllin and incubated overnight (~16 h) at 37 °C in a shaking incubator set at 250 rpm (Labcon, Maraisburg, South Africa). A 750 µl aliquot of each bacterial culture was used to make a 1 ml, 25 % v/v glycerol stock and stored at -60 °C. The remaining culture volume was used to purify plasmid DNA (pDNA) by following a preparation protocol for small culture volumes (mini-prep) (Appendix B.4). The pDNA yield was quantified using a

NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), which measures sample absorbance (A) at a wavelength of 260 nm to calculate the amount of DNA present in a sample. pTZ plasmids were initially screened using an economical double restriction enzyme digest, such as *Eco*RI / *Hind*III or *Sma*I / *Xba*I, to confirm the presence of an insert. Restriction enzyme reactions were incubated for 2 -3 h or overnight at 37 °C with the following components in a 10 or 20 µl volume: 5 – 10 U each restriction enzyme (Thermo Fisher Scientific, Waltham, MA, USA), 1X concentration of accompanying buffer (Thermo Fisher Scientific Unique Buffer or Five Buffer System), 1 to 1.5 µg pDNA and sterile water. Restriction digest products were analysed using agarose gel electrophoresis. Recombinant pTZ plasmids were screened in a second restriction enzyme digest, using *Xho*I / *Not*I sites of the PCR-generated pri-miRNA, to assess the size (183 – 249 bp) and integrity of pri-miRNA insert. Suitable recombinant plasmids were again purified from an overnight bacterial culture using a commercially available mini-prep kit (Qiagen, Hilden, Germany) to obtain better quality pDNA for sequencing and further cloning. Recombinant pTZ plasmids with the correct pri-miRNA sequence were identified through fluorescence-based cycle DNA sequencing using the 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) (Inqaba Biotec, Pretoria, South Africa) and primers listed in Table 2.4. The secondary structure of pri-miRNA hairpins commonly disrupted the quality of DNA sequencing results. To obtain reliable sequencing results, a second sequencing reaction with a primer in the reverse orientation was often necessary, while the addition of 5 % v/v dimethylsulfoxide (DMSO) as a denaturant was standard.

Table 2.4: Primers used for DNA sequencing

Plasmid Backbone	Primer	Sequence (5' to 3' orientation)
pTZ57R/T	M13/pUC F (-20)	GTAAAACGACGGCCAGT
	M13/pUC R (-26)	CAGGAAACAGCTATGAC
pCI-neo	T7-5 F	GTACTTAATACGACTCACTATAGG
	T3-3 R	AGCATTAACCCTCACTAAAGGG
psiCHECK2.2	pSiCheck hRluc 3' F	GAGGACGCTCCAGATGAAATG

F: forward primer; R: reverse primer

Insertion into a mammalian expression vector

Pri-miRNA sequences were inserted into the pCI-neo Mammalian Expression Vector (Promega, Madison, WI, USA) to allow for constitutive pri-miRNA expression, driven by the pol II cytomegalovirus (CMV) immediate-early enhancer/promoter (Figure 2.1 C). The correct pri-

miRNA sequences were isolated from corresponding pTZ plasmids through an overnight *Xho*I / *Not*I restriction digest (3 µg pDNA), followed by electrophoresis (1.5 % w/v agarose gel) and extraction with a commercially available kit (Qiagen MinElute Gel Extraction Kit, Hilden, Germany) (Appendix B.5). The pCI-neo Mammalian Expression Vector was prepared in the same manner (0.8 % w/v agarose gel) to produce compatible sticky ends for ligation with the pri-miRNA insert. The mass of pDNA used in restriction digests was calculated according to ligation requirements and assuming 30 % recovery after gel extraction. An overnight ligation reaction was set up at 4ºC with 25 – 30 ng pri-miRNA insert and 215 ng pCI-neo vector to give a molar ratio of 0.36 pmol ends insert : 0.12 pmol ends vector. The amount of insert (pmol) was therefore estimated to be at least 3 times higher than the amount of vector (pmol), which is typically optimal for this type of cloning.

A 10 µl aliquot of each ligation reaction was used to transform chemically competent DH5α™ *E. coli* cells (Invitrogen, Carlsbad, CA, USA). The transformation mix was spread on LB agar plates with 0.01% w/v ampicillin as a selective agent. Three colonies were selected from each plate and used for plasmid purification and screening with an *Xho*I / *Not*I restriction digest as described above. The identity of final pCI-neo-pri-miRNA expression vectors was confirmed after DNA sequencing with primers listed in Table 2.4.

Promoter replacement to create pSI-neo expression vectors

The pCI-neo and pCI-neo-pri-miR-520b, -198, -147a and -34a vectors were altered by replacing the CMV promoter with a simian virus 40 (SV40) enhancer and early promoter region (Benoist and Chambon, 1981), (Fromm and Berg, 1983) to create corresponding pSI-neo vectors (Figure 2.1 D). This was done to assess any potential vector specific effects for non-functional pri-miRNAs, while still allowing for constitutive expression in mammalian cell culture. The CMV promoter was removed by an overnight a *Bgl*II / *Nhe*I restriction digest (3µg pDNA) and the remaining pDNA backbone was isolated through gel electrophoresis (1.2 % w/v agarose gel) and extraction (Qiagen MinElute Gel Extraction Kit, Hilden, Germany). The SV40 promoter fragment was prepared in the same manner from the psiCHECK™-2 Vector (Promega, Madison, WI, USA). An overnight ligation reaction was set up at 4ºC with a molar ratio of 0.36 pmol ends insert : 0.12 pmol ends vector, using 80 ng SV40 insert and 180 ng pCI-neo backbone. Recombinant plasmids were identified following the general transformation and screening protocols described for pCI-neo expression vectors. A *Kpn*I / *Xho*I restriction digest was used for

screening, as a unique *Kpn*I site is located in both the insert and the vector, producing 2 restriction sites in pSI-neo-pri-miRNA expression vectors.

2.2.3 Generation of shRNA expression plasmids

Design

A short hairpin RNA (shRNA) was designed for each of the selected pri-miRNA sequences as a positive control for guide sequence expression and function. The shRNA format used here has been successfully characterised elsewhere (Saayman et al., 2010), and consists of a hairpin with a 5' guide and a 9 nt loop derived from the predicted precursor of hsa-miR-22 (Gu et al., 2012) (Figure 2.2 A, overleaf). The human U6 small nuclear RNA (snRNA) pol III promoter was selected for high level, constitutive expression of shRNAs. The following guide sequences were expected as the major product from each pri-miRNA and used to design corresponding shRNAs: 520b-3p, 198-5p, 147a-3p, 23a-3p, 934-5p, 34-5p and 513-3p.

Construction

Generation of shRNA sequences by a two-step PCR

ShRNAs expression cassettes were constructed using a two-step PCR method (Castanotto et al., 2002) illustrated in Figure 2.2 B with primers listed in Table 2.5. A single forward primer (U6 F) complementary to the 5' end of the human U6 small nuclear RNA (snRNA) pol III promoter (GenBank Accession (Benson et al., 2015) JN255693.1) was used in both rounds of PCR. Additional restriction sites (*Bam*HI / *Xho*I / *Not*I / *Nhe*I) were added to the 5' end of the U6 F primer for screening purposes. In the first PCR step, products containing the U6 promoter and the first part of each short hairpin were generated. The reverse primers (R1) used in this step were complementary to the 3' end of the U6 promoter, but also included the unique 5' and loop sequence of each short hairpin. The pTZ-U6+1 plasmid (10 ng) was used as a template in the first PCR reaction to provide the U6+1 promoter sequence (Bertrand et al., 1997). In the second PCR step, each first round PCR product was extended to include the full sequence of the respective short hairpin. The first round product was diluted 100 X and 1 µl was used directly as a template. The reverse primers (R2) used in the second PCR were complementary to the 3' end of the first PCR product, but also contained the unique 3' sequence of each hairpin to

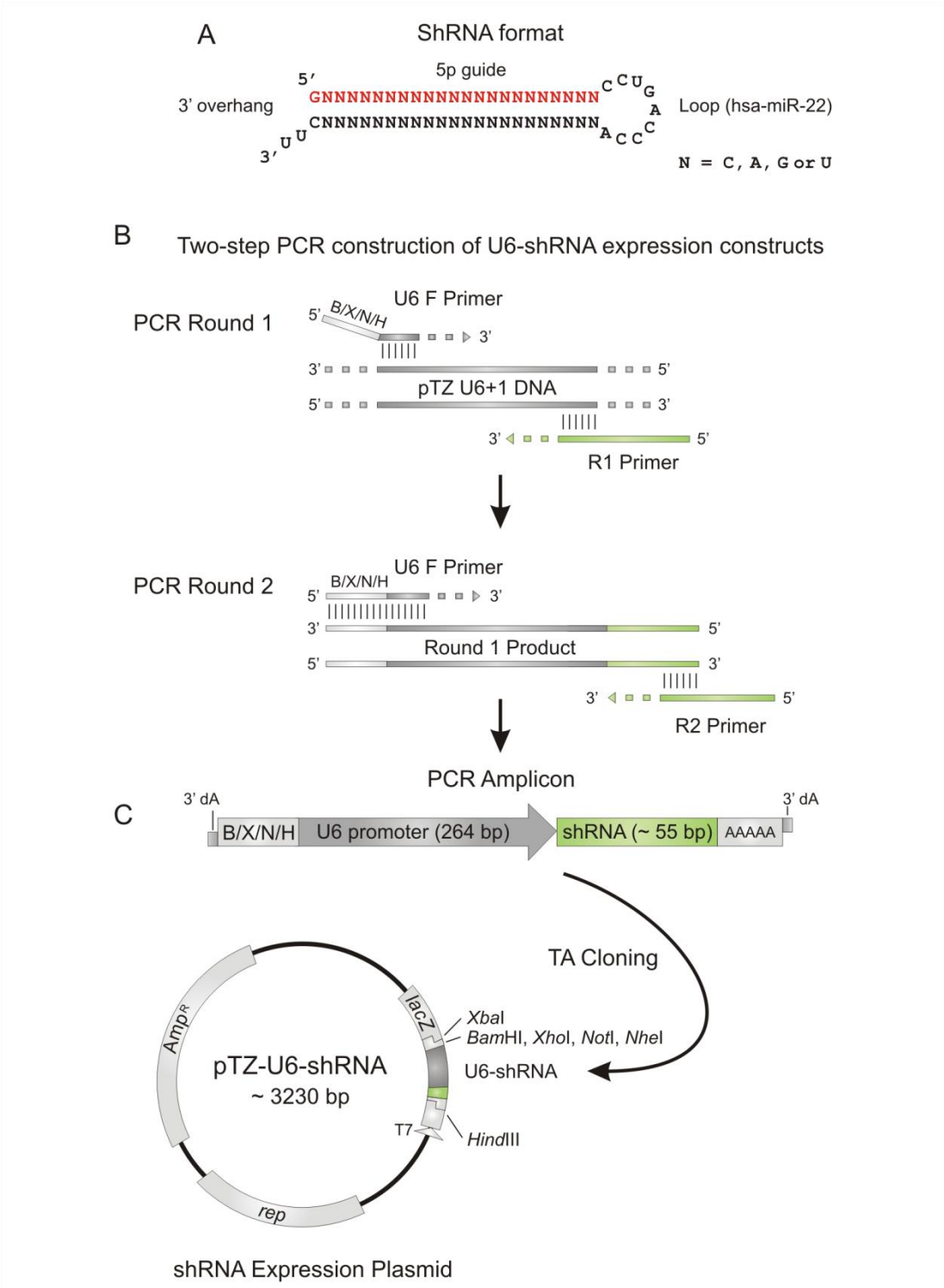


Figure 2.2: PCR and cloning scheme for the generation of shRNA expression vectors

dA: deoxyadenosine; Amp^R: ampicillin resistance gene β-lactamase; *lacZ*: gene for LacZ α-peptide used in blue-white screening; *rep*: replicon region containing replication start site; T7: T7 RNA polymerse promoter region. The plasmid map was adapted from a figure in the product bulletin (Thermo Fisher Scientific).

extend the product. Five deoxyadenosines were included in each R2 primer to add a 3' deoxythymidine pol III transcription termination signal at the end of each short hairpin. The PCR programme for both steps was as follows: 94ºC for 5 min; 30 cycles of 94 ºC for 30s, 55 °C for 30 s and 72 ºC for 40s; and final extension at 72 ºC for 7 min. The size and purity of PCR products was assessed using agarose gel electrophoresis.

Table 2.5: Primers for the generation of U6-driven shRNAs by a two-step PCR

Primer	Tm (°C)	Sequence (5' to 3' orientation)
U6 F B/X/N/H	75.8	GGATCCTCGAGCGGCCGCTAGCAAGGTCGGGCAGGAAGAGGGCC
sh520 R1	72.1	GAGGGTGGGTCAGGCCCTCTAGAAGGAAGCACTTCGGTGTTTCGTCCTTTCCACAA
sh520 R2	66.3	AAAAATAAGTGCTTCCTTTTAGAGGGTGGGTCAGGCCCTC
sh198 R1	72.3	GGTTCTGGGTCAGGGAACCTATCTCCCCTCTGGACCGGTGTTTCGTCCTTTCCACAA
sh198 R2	68.5	AAAAAGGTCCAGAGGGGAGATAGGTTCTGGGTCAGGGAACC
sh147 R1	71.0	TGCTATGGGTCAGGTAGCAGAAGCATTTCCACACACGGTGTTTCGTCCTTTCCACAA
sh147 R2	66.2	AAAAAGTGTGTGGAAATGCTTCTGCTATGGGTCAGGTAGCA
sh23 R1	72.1	TTTCCTGGGTCAGGGGAAATCCCTGGCAATGTGACGGTGTTTCGTCCTTTCCACAA
sh23 R2	66.9	AAAAATTCACATTGCCAGGGATTTCCTGGGTCAGGGGAAA
sh934 R1	72.2	ACTGGTGGGTCAGGCCAGTGTCTCCAGTAGTAGACCGGTGTTTCGTCCTTTCCACAA
sh934 R2	67.4	AAAAATGTCTACTACTGGAGACACTGGTGGGTCAGGCCAGT
sh34 R1	71.7	GTTGTTGGGTCAGGACAACCAGCTAAGACACTGCCCGGTGTTTCGTCCTTTCCACAA
sh34 R2	66.7	AAAAATGGCAGTGTCTTAGCTGGTTGTTGGGTCAGGACAAC
sh513 R1	70.2	GAAGGTGGGTCAGGCCTTCTCAGAAAGGTGAAATTTCGGTGTTTCGTCCTTTCCACA A
sh513 R2	64.3	AAAAATTAATTTCACCTTTCTGAGAAGGTGGGTCAGGCCTTC

Tm: melting temperature; F: forward extension primer; R: reverse extension primer; U6: human U6 snRNA polymerase III promoter;
Shaded region of round 1 primers are complementary to the pTZ U6+1 template. Shaded regions of round 2 extension primers are complementary to the round 1 PCR product.
Flanking *Bam*HI (GGATCC), *Xho*I (CTCGAG), *Not*I (GCGGCCGC) and *Nhe*I (GCTAGC) restriction sites of the U6 F primer are underlined. Sequence corresponding to the 9 nt shRNA loop is also underlined.

Insertion into a TA cloning vector

The U6 expression cassette with the full hairpin sequence was inserted into the pTZ57R/T TA cloning vector (InsTAclone™ PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA, USA) using the same ligation protocol described for pri-miRNA constructs (see 2.2.2). Similarly, each pTZ ligation reaction was used to transform DH5α™ *E. coli* cells and bacterial colonies containing recombinant pTZ were selected by blue-white screening with ampicillin resistance as a selective marker. An *Xba* I / *Nhe*I restriction enzyme digest, followed by gel electrophoresis,

was used to identify plasmids with a U6-shRNA insert of around 345 bp in the same orientation as the *lacZ* gene. An *Xho*I / *Hind*III digest was used to identify U6-shRNA inserts in the opposite orientation to the *lacZ* gene. The U6-shRNA cassette remains functional in both orientations as expression is independent of the pTZ plasmid backbone. Candidate plasmids were again purified from an overnight bacterial culture, inoculated from a glycerol stock, using a commercially available mini-prep kit (Qiagen, Hilden, Germany). Final clones were confirmed after DNA sequencing, often with both primers listed in Table 2.4 and a final concentration of 1 M betaine (Haqqi, 2002). This was necessary as hairpins were designed to be completely complementary, resulting in structures that were often too stable for a continuous sequence output.

2.2.4 Generation of target reporter plasmids

Design

Target reporter plasmids were constructed to assess the silencing function of expressed pri-miRNAs or shRNAs in a dual luciferase reporter assay (see 2.2.5), using the cloning strategy illustrated in Figure 2.3. Target sequences were designed to be complementary to each expected RNA guide and were inserted into the 3' untranslated region (UTR) of the *Renilla reniformis* luciferase gene of the psiCHECK^{TM}-2 Vector (Promega, Madison, WI, USA), which also codes for firely (*Photinus pyralis*) luciferase. The generation of RNA guides from processed pri-miRNAs or shRNAs results in RNAi-mediated degradation of *Renilla* transcripts containing the corresponding target sequence. This leads to decreased expression of *Renilla* luciferase, while the expression of firely luciferase remains unaffected.

Construction

Generation of target duplexes by annealing of complementary oligonucleotides

Complementary target sequences were designed for expected ~ 22 nt RNA guides from both the 5p and 3p arm of each pri-miRNA. Target sequences were extended to include 2 nt on either side of the an anticipated Drosha-DGCR8 and Dicer processing sites and accommodate any minor shifts in guide processing. Additional 5' *Xho*I, *Eco*RV and 3' *Not*I flanking restriction sites were included for ligation and screening purposes. Double-stranded target sequences were assembled by annealing of complementary oligonucleotide pairs (TF, TR) listed in Table 2.6.

Each pair of oligonucleotides was combined with sterile water (10 µM each oligonucleotide, 20 µl volume) and denatured at 80ºC for 5 minutes on an electronic heating block. The heating block was then removed from the heat source and placed at room temperature to allow for gradual cooling and annealing of oligonucleotide pairs. The final reaction was diluted by adding sterile water to a volume of 500 µl.

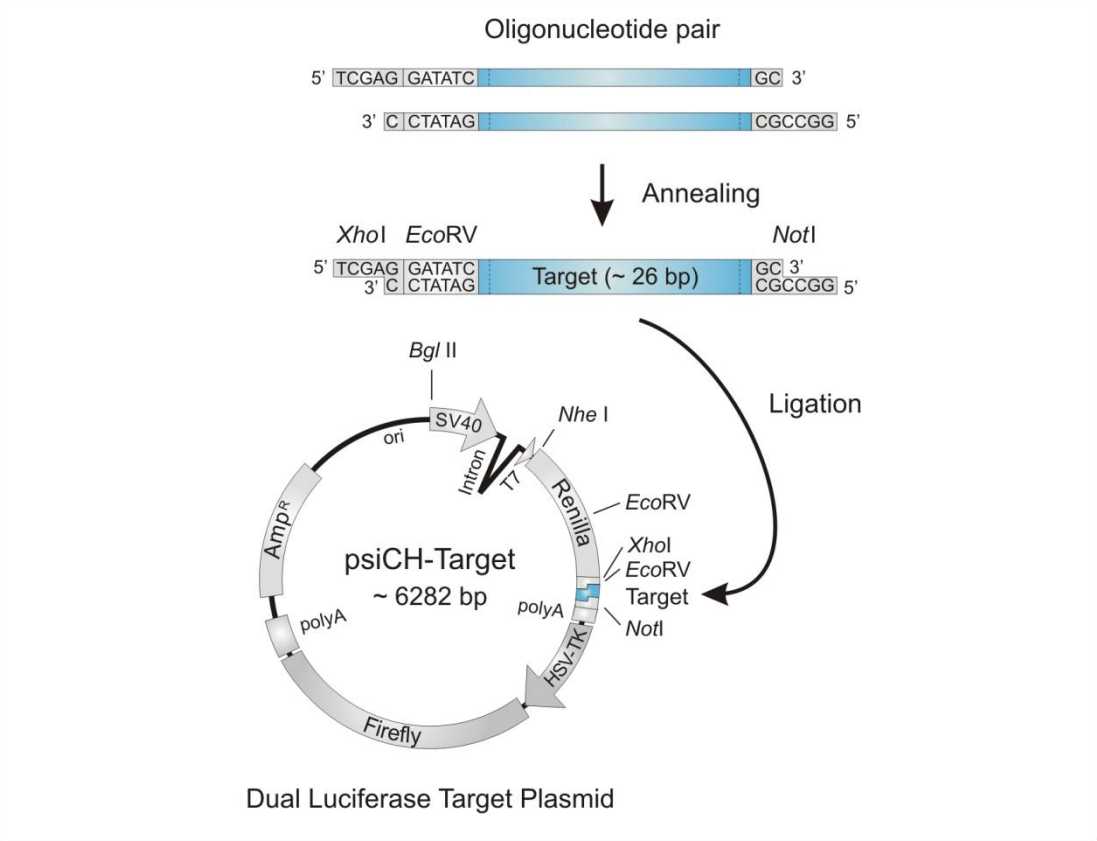


Figure 2.3: Cloning scheme for the generation of dual luciferase target vectors
Amp^R: ampicillin resistance gene β-lactamase; *ori*: origin of replication; SV40: simian virus 40 enhancer and early promoter region; T7: T7 RNA polymerse promoter region; *Renilla*: *Renilla reniformis* luciferase gene; HSV-TK: herpes simplex virus thymidine kinase promoter; Firefly: firefly luciferase; polyA: polyadenylation signal. The plasmid map was adapted from a figure in the product bulletin (Promega).

Table 2.6: Oligonucleotides for generation of a target duplex

Oligonucleotide	Sequence (duplex orientation)
Pri-miR-520b 5p TF Pri-miR-520b 5p TR	5′ TCGAGATATCAGAAAGCGCTTCCCTGTAGAGGGACGC 3′ 3′ CTATAGTCTTTCGCGAAGGGACATCTCCCTGCGCCGG 5′
Pri-miR-520b 3p TF Pri-miR-520b 3pTR	5′ TCGAGATATCAACCCTCTAAAAGGAAGCACTTTCTGC 3′ 3′ CTATAGTTGGGAGATTTTCCTTCGTGAAAGACGCCGG 5′
Pri-miR-198 5p TF Pri-miR-198 5p TR	5′ TCGAGATATCAGGAACCTATCTCCCCTCTGGACCAAGC 3′ 3′ CTATAGTCCTTGGATAGAGGGGAGACCTGGTTCGCCGG 5′
Pri-miR-198 3p TF Pri-miR-198 3p TR	5′ TCGAGATATCTTTATTCTATAGAGAAGAAGGAAAAAGC 3′ 3′ CTATAGAAATAAGATATCTCTTCTTCCTTTTTCGCCGG 5′
Pri-miR-147a 5p TF Pri-miR-147a 5p TR	5′ TCGAGGATATCGTGTGTGTGCAGAAATGTTGTCTTTAGAGC 3′ 3′ CCTATAGCACACACACGTCTTTACAACAGAAATCTCGCCGG 5′
Pri-miR-147a 3p TF Pri-miR-147a 3p TR	5′ TCGAGGATATCAATCTAGCAGAAGCATTTCCACACACTGGC 3′ 3′ CCTATAGTTAGATCGTCTTCGTAAAGGTGTGTGACCGCCGG 5′
Pri-miR-23a 5p TF Pri-miR-23a 5p TR	5′ TCGAGATATCGCAAATCCCATCCCCAGGAACCCCAGGC 3′ 3′ CTATAGCGTTTAGGGTAGGGGTCCTTGGGGTCCGCCGG 5′
Pri-miR-23a 3p TF Pri-miR-23a 3p TR	5′ TCGAGATATCTTGGAAATCCCTGGCAATGTGATTTGGC 3′ 3′ CTATAGAACCTTTAGGGACCGTTACACTAAACCGCCGG 5′
Pri-miR-934 5p TF Pri-miR-934 5p TR	5′ TCGAGATATCTACCAGTGTCTCCAGTAGTAGACAGAGC 3′ 3′ CTATAGATGGTCACAGAGGTCATCATCTGTCTCGCCGG 5′
Pri-miR-934 3p TF Pri-miR-934 3p TR	5′ TCGAGATATCCCCGTCCATTACTGGAGACTCTGGGTGC 3′ 3′ CTATAGGGGCAGGTAATGACCTCTGAGACCCACGCCGG 5′
Pri-miR-34a 5p TF Pri-miR-34a 5p TR	5′ TCGAGATATCCAACAACCAGCTAAGACACTGCCAAAGC 3′ 3′ CTATAGGTTGTTGGTCGATTCTGTGACGGTTTCGCCGG 5′
Pri-miR-34a 3p TF Pri-miR-34a 3p TR	5′ TCGAGATATCCTAGGGCAGTATACTTGCTGATTGCTGC 3′ 3′ CTATAGGATCCCGTCATATGAACGACTAACGACGCCGG 5′
Pri-miR-34a 5p LS TF Pri-miR-34a 5p LS TR	5′ TCGAGGATATCAAGAAACACTCACAGCTGGCCGGTCCGC 3′ 3′ CCTATAGTTCTTTGTGAGTGTCGACCGGCCAGGCGCCGG 5′
Pri-miR-34a 3p LS TF Pri-miR-34a 3p LS TR	5′ TCGAGGATATCCTCTTGGGCCCCACAACGTGCAGCACTTCTAGC 3′ 3′ CCTATAGGAGAACCCGGGGTGTTGCACGTCGTGAAGATCGCCGG 5′
Pri-miR-513a-2 5p TF Pri-miR-513a-2 5p TR	5′ TCGAGATATCCATAAATGACACCTCCCTGTGAAAGGCGC 3′ 3′ CTATAGGTATTTACTGTGGAGGGACACTTTCCGCGCCGG 5′
Pri-miR-513a-2 3p TF Pri-miR-513a-2 3p TR	5′ TCGAGATATCACCCTTCTCAGAAAGGTGAAATTTATAGC 3′ 3′ CTATAGTGGGAAGAGTCTTTCCACTTTAAATATCGCCGG 5′
Pri-miR-513a-2 5p LS TF Pri-miR-513a-2 5p LS TR	5′ TCGAGGATATCGGCACTGCACACTGAATGGCTGAATGTGGCATCCAAGAGC 3′ 3′ CCTATAGCCGTGACGTGTGACTTACCGACTTACACCGTAGGTTCTCGCCGG 5′
Pri-miR-513a-2 3p LS TF Pri-miR-513a-2 3p LS TR	5′ TCGAGGATATCTCAGGGACACCACATATGCAGTGCATGCTGTACATTACCCGC 3′ CCTATAGAGTCCCTGTGGTGTATACGTCACGTACGACATGTAATGGGCGCCGG 3′ 5′

T: Target; F: forward orientation oligonucleotide; R: reverse orientation oligonucleotide
Flanking *Xho*I (CTCGAG), *Eco*RV (GATATC) and *Not*I (GCGGCCGC) restriction sites are underlined.

Insertion into a dual luciferase reporter vector

A modified version of the original psiCHECK^{TM}-2 Vector (Promega, Madison, WI, USA), psiCHECK^{TM}-2.2 (Dr. Marc Weinberg), was used for the construction of target reporter vectors. This vector contains additional *Sac*I, *Spe*I and *Sal*I restriction sites, between *Xho*I and *Not*I of the MCS (Ely et al., 2008), which have no consequence for the overall cloning strategy used here. The psiCHECK^{TM}-2.2 vector was prepared by an overnight *Xho*I / *Not*I restriction digest (3 µg pDNA), followed by electrophoresis (0.8 % w/v agarose gel) and extraction with a commercially available kit (Qiagen MinElute Gel Extraction Kit, Hilden, Germany). The *Xho*I and *Not*I restriction sites are unique to the MCS of the vector and were selected to enable ligation with corresponding sticky ends of a target duplex. An overnight ligation reaction was set up at 4ºC (see 2.2.2) with 200 ng psiCHECK^{TM}-2 backbone (0.1 pmol ends pDNA) and 1 µl annealed target dilution (maximum of 0.8 pmol ends duplex). A 10 µl aliquot of each ligation reaction was used to transform chemically competent DH5α^{TM} *E. coli* cells (Invitrogen, Carlsbad, CA, USA) and ampicillin resistance was used as a selective marker. Plasmids were purified from five bacterial colonies for each target using a protocol for small culture volumes (mini-prep) (Appendix B.4). An *Eco*RV restriction digest was used for screening, where the additional *Eco*RV site of a recombinant plasmid yields a DNA band of ~ 470 bp following a restriction digest and agarose gel electrophoresis. Candidate plasmids were again purified from an overnight bacterial culture, inoculated from a glycerol stock, using a commercially available mini-prep kit (Qiagen, Hilden, Germany). Final clones were confirmed after DNA sequencing with primers listed in Table 2.4.

2.2.5 Assessment of pri-miRNA function in cultured mammalian cells

Mammalian cell culture

Experiments were performed using several common laboratory cell lines suitable for RNAi and gene silencing experiments, namely, HEK293 (human embryonic kidney) (Graham et. al., 1977), Huh-7 (Human hepatoma-7) (Nakabayashi et. al., 1982) and HeLa (Henrietta Lacks, cervical cancer) (Gey et. al., 1952) cell lines. Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Biowhittaker, Lonza, Walkersville, MD, USA), supplemented with 10 % heat-inactivated foetal calf serum (FCS; Delta Bioproducts, Johannesburg, South Africa), penicillin (100 U / ml) and streptomycin (0.1 mg / ml), at 37 °C and 5 % CO_2 in a humidified incubator (Thermo Fisher Scientific, Waltham, MA, USA). Antibiotics were omitted from cell culture medium for all

transfection experiments. Cells were typically sub-cultured every 2 – 3 days at a ratio of 1:4 or 1:5 depending on cell confluency and seeding requirements. The general sub-culture protocol per 25 cm^2 of cultured cells was as follows: removal of existing media; 1 ml rinse with phosphate-buffered saline (PBS); incubation in 1 ml PBS at 37 °C for 3 minutes (HEK293) or a trypsin-EDTA formulation (TrypLE™ Express, Gibco, Thermo Scientific) at 37 °C for 5 minutes (Huh-7 and HeLa) to allow for cell dissociation; re-suspension in 4 – 5 ml DMEM; distribution to new sterile culture flasks; and addition of media to the recommended volume. All cell culture work was performed in a biosafety cabinet disinfected with a 2 % w/v/ Virkon (Du Pont Chemicals, Wilmington, DE, USA) solution and 70 % ethanol. Cell culture stocks were frozen and thawed as described in Appendix B.6.

Transfection of cultured cells

Cells for the dual luciferase assay were seeded at ~ 1.2 x 10^5 cells per well in a 24 well format, 24 h before transfection. Transfections were performed in triplicate with 1 µg total pDNA per well (750 ng pri-miRNA or shRNA expression plasmid; 150 ng target reporter plasmid; 100 ng pCI-eGFP) using Lipofectamine ® 2000 Reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocol (Appendix B.7). pCI-eGFP constitutively expresses enhanced green fluorescent protein (Passman et al., 2000) and was co-transfected as a reporter of transfection efficiency. Culture medium was replaced 24 h post-transfection if necessary, while transfection efficiency was assessed visually at 48 h using fluorescence microscopy. Successfully transfected cells were lysed using 100 µl passive lysis buffer per well (Dual - Luciferase ® Reporter Assay System; Promega, Madison, WI, USA).

Dual luciferase reporter assay

A dual luciferase reporter assay was used to assess the gene silencing function of expressed pri-miRNAs or shRNAs against corresponding target plasmids. Assays were performed using the Dual - Luciferase ® Reporter Assay System (Promega, Madison, WI, USA) and luminescence was measured using a Veritas ™ dual injection Luminometer (Turner Biosystems, Sunnyvale, CA, USA). In this assay, the activity of firefly and *Renilla* luciferase enzymes in bioluminescent reactions was measured sequentially (Figure 2.4). Both enzymes were expressed from the psiCHECK™-2 target vectors, but RNAi-mediated silencing by corresponding pri-miRNA or shRNA guides was only intended to reduce the expression of *Renilla* luciferase (see 2.2.4). In the case of effective RNAi, total activity of *Renilla* luciferase was reduced, while total activity of the firely luciferase was constant. Target silencing mediated by

each pri-miRNA and shRNA was therefore measured as a ratio of *Renilla* luciferase signal to firefly luciferase signal (h*R*luc : hFluc). This dual expression system is ideal to minimize the effects of experimental variation in transfection efficiency, cell number, lysis efficiency and pipetting accuracy.

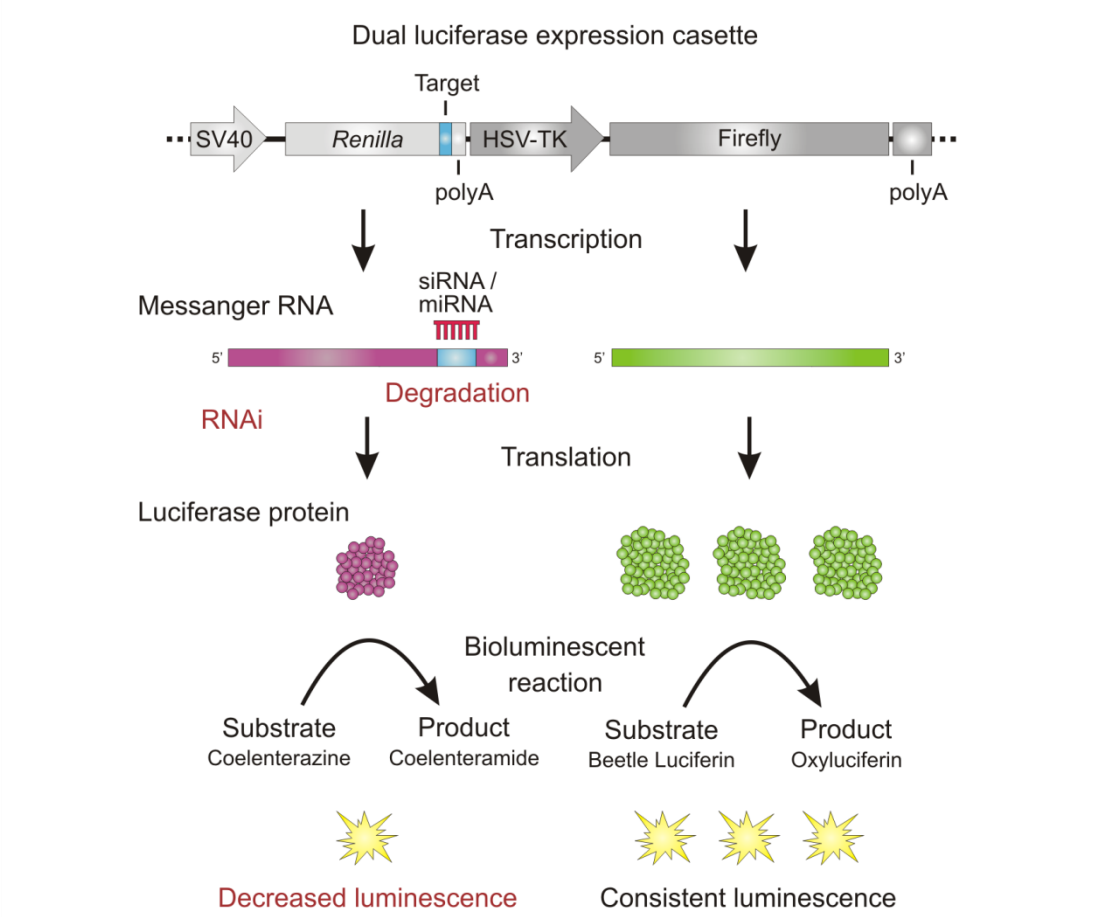


Figure 2.4: Dual luciferase expression system for the detection of specific RNAi activity
The Dual - Luciferase ® Reporter Assay System (Promega, Madison, WI, USA) is designed to detect the effects of RNAi. A shRNA- or pri-miRNA- derived guide complementary to the target sequence in a psiCHECK^{TM}-2 reporter vector, causes RNAi-induced post-transcriptional silencing of the *Renilla reniformis* luciferase gene. This reduces the amount of Renilla luciferase available in subsequent bioluminescent reactions, while the amount of firefly luciferase is unaffected. A measurement of luminescence produced from these reactions indicates the catalytic activity of *Renilla* luciferase relative to firefly luciferase, which in turn reflects the extent of RNAi-mediated silencing of *Renilla* luciferase. SV40: simian virus 40 enhancer and early promoter region; HSV-TK: herpes simplex virus thymidine kinase promoter; Firefly: firefly luciferase; polyA: polyadenylation signal. The schematic was adapted from figures and information in the product bulletins for the plasmid and reporter system (Promega).

The average silencing ratio for each pri-miRNA or shRNA sample triplicate was presented following normalization to the average ratio of a corresponding mock sample triplicate, where the pri-miRNA or shRNA expression vector was replaced by the respective control (empty) expression plasmid (pCI-neo or pTZ57R). A negative control sample was also included for comparison, where the pri-miRNA or shRNA expression vector was replaced by a non-targeting vector, designed to express pri-miR-31-tat (Barichievy, 2009) or shRNA-LTR (Saayman et al., 2010) which were not expected to mediate silencing of experimental targets. In addition, pri-miRNA and shRNA expression vectors were used in a target control assay with an empty reporter plasmid (psiCHECK^{TM}-2.2) to assess for any unanticipated silencing effects.

2.2.6 Assessment of pri-miRNA processing in cultured mammalian cells

Transfection of cultured cells

HEK293 cells for Northern blot analysis were seeded at ~ 1.1 x 10^{6} cells per 6 cm dish 18 - 24 h prior to transfection using Lipofectamine ® 2000 Reagent (Invitrogen, Carlsbad, CA, USA) (Appendix B.7). RNA expression plasmids (9 µg) were co-transfected with pCI-eGFP (1 µg) as a reporter of transfection efficiency. Cell media was replaced 24 h post-transfection if necessary and cells were lysed for RNA extraction 48 h post-transfection using TRI Reagent ® (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's protocol for monolayer cells. RNA-handling techniques were employed to avoid sample degradation (Appendix B.8).

PAGE analysis of total RNA

Northern blot analysis was used to confirm the production of guide sequences from transfected pri-miRNA or shRNA expression constructs in cell culture. Total cellular RNA extracts were re-suspended in 20 µl DEPC-treated water (Appendix B.8) and diluted to give a final concentration of 1 – 1.5 µg / µl as measured by a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The absorbance (A) ratio at wavelengths of 260 nm and 280 nm was used as an indicator of RNA quality (A_{260} / A_{280} ~ 2.0; 1.8 - 2.1 accepted) (Gallagher and Desjardins, 2006). RNA samples (30 µg in 20 - 30 µl) were suspended in an equal volume of 2X RNA loading dye (Appendix B.8) and resolved using denaturing polyacrylamide gel electrophoresis (PAGE). A 15 % w/v polyacrylamide gel was made with the following components: 2 M acrylamide, 0.05 M N,N'-methylenebisacrylamide (acrylamide: bisacrylamide mass ratio of 19:1), 8 M Urea, 1 X concentration of Tris-borate-EDTA (TBE) electrophoresis

buffer (89 mM Tris base, 89 mM boric acid, 20 mM EDTA, pH 8.0) (Moore, 2001) and sterile, DEPC-treated water. Polymerization was initiated by the addition of 250 µl of a 10 % w/v ammonium persulphate (APS) and 25 µl TEMED (N,N,N',N'-tetramethylethylenediamine) to 60 ml of the gel mixture. Gels were cast using a vertical unit (SE400, Hoefer Inc., Holliston, MA, USA; 18 cm x 16 cm glass plates 1.5 mm spacers). Prior to sample loading, gels were subjected to electrophoresis at 150 V (Bio-Rad HC PowerPac^{TM}, Hercules, CA, USA) for 20 - 30 min with 0.5 X concentration of TBE buffer and wells were flushed with a disposable syringe and needle to remove urea and debris. RNA samples were denatured at 80 °C for 5 min, followed by 2 - 5 min on ice before loading with a Hamilton syringe. Decade^{TM} Marker (Ambion, Austin, TX, USA) was also loaded as a molecular weight marker (2 - 5 µl aliquot, 20 µl final volume), following preparation according to the manufacturer's protocol (Appendix B.8). P^{32}-labelled γ-ATP (1 µl; specific activity 3000 Ci/mmol; PerkinElmer, Waltham, MA, USA) was used in the 5' radio-labelling reaction. Gels were typically subjected to electrophoresis for 5 - 6 hours at 100 - 250 V to achieve the desired separation of small RNA products. Gels were stained with 10 µl of an EtBr stock solution (10 µg / µl) in ~ 250 ml electrophoresis buffer if EtBr was absent in the RNA loading dye. RNA samples were visualised using a gel imaging system (G:Box, Syngene, Cambridge, UK). Ribosomal RNA (rRNA) and transfer RNA (tRNA) species, including the cellular 5.8S (152 b) and 5S rRNAs (116 b) were used to assess RNA quality and loading consistency (sizes: Ensembl genome browser, release 78 (Cunningham et al., 2015)). These bands were visible as sharp bands in non-degraded RNA samples (Brown et al., 2004), whereas only a "smear" was visible for degraded RNA samples, which were discarded.

Northern blot transfer of total RNA

RNA was transferred to a Hybond N+ membrane (Amersham, GE Healthcare, Buckinghamshire, UK) using a semi-dry blotter (Sigma-Aldrich, Dorset, UK). Sandwiches were assembled using 3 sheets of Grade 3 MM Whatman® chromatography and blotting paper (0.34 mm) (Whatman®, Maidstone, UK) on either side of the gel and membrane, after a brief soaking in 0.5 X concentration of TBE buffer. An electric current was applied to the blotter for 1 h at 3.3 mA per cm^2 blot or a maximum of 450 mA (Bio-Rad HC PowerPac^{TM}, Hercules, CA, USA). RNA was immobilised on the membrane by UV crosslinking at 254 nm (0.2 J/cm^2) (UVP Crosslinker, Upland, CA, USA) and baking for 1 h at 80 ºC (Hybaid Shake 'n Stack oven, Thermo Fisher Scientific, Waltham, MA, USA).

Detection of RNA guide sequences

RNA blots were hybridized with ~ 22 nt oligonucleotide probes complementary to the anticipated RNA guide sequences (Table 2.7). Locked nucleic acid (LNA™; Exiqon, Woburn, MA, USA) bases were incorporated in certain probes for improved sensitivity and binding affinity (Valoczi et al., 2004). RNA blots were also hybridized with a control probe to detect highly expressed U6 small nuclear RNA (snRNA) as a measure of loading consistency, primarily detecting U6-1 and U6-2 gene transcripts (blastn; http://blast.ncbi.nlm.nih.gov/Blast.cgi, (Johnson et al., 2008)). Hybridisation was performed for ~ 16 h at 42 ºC for standard oligonucleotide probes and at 50 ºC for LNA-modified probes using 10 ml Rapid-hyb buffer (Amersham, GE Healthcare, Buckinghamshire, UK) in a hybridisation bottle (Hybaid, Thermo Fisher Scientific, Waltham, MA, USA) with gradual rotation in a hybridization oven (Hybaid Shake 'n Stack, Thermo Fisher Scientific). Prior to probing, membranes were pre-hybridised at the respective hybridization temperature for 20 minutes.

Table 2.7: Oligonucleotide probes for detection of guides in northern blot analysis.

Probe	RNA Tm (°C)	Sequence (5' to 3' orientation)
U6 snRNA	61.7	TAGTATATGTGCTGCCGAAGCGAGCA
Pri-miR-520b 5p	56.4	AAAGCGCTTCCCTGTAGAGGG
Pri-miR-520b 3p	46.9	CCCTCTAAAAGGAAGCACTTT
Pri-miR-198 5p	62.2	GAACCTATCTCCCCTCTGGACC
Pri-miR-198 3p	35.3	TATTCTATAGAGAAGAAGGAAA
Pri-miR-147a 5p	48.7	GTGTGCAGAAATGTTGTCTTTA
Pri-miR-147a 3p LNA	65.0	T+AGCAGAAGCATTTCCACA+C+AC
Pri-miR-23a 5p	60.1	AAATCCCATCCCCAGGAACCCC
Pri-miR-23a 3p	50.1	GGAAATCCCTGGCAATGTGAT
Pri-miR-934 5p	59.8	CCAGTGTCTCCAGTAGTAGACA
Pri-miR-934 3p	55.5	CGTCCATTACTGGAGACTCTGG
Pri-miR-34a 5p	53.0	ACAACCAGCTAAGACACTGCCA
Pri-miR-34a 3p	53.1	AGGGCAGTATACTTGCTGATTG
Pri-miR-513a-2 5p	53.3	ATGACACCTCCCTGTGAA
Pri-miR-513a-2 3p	44.1	CCTTCTCAGAAAGGTGAAATTTA

+: LNA™ modification of the proceeding base.
RNA Tm: melting temperature for an RNA target, as calculated by the supplier or using OligoAnalyser v3.1.

Probes were 5' radio-labelled in a forward reaction with T4 polynucleotide kinase (PNK) (EC 2.7.1.78) (Maunders, 1993b), (Liebecq, 1997), originally derived from the enterobacteria phage T4, which catalyses the transfer of the γ-phosphate from ATP to the 5'-OH group of single- and

double-stranded DNAs and oligonucleotides. The kinase reaction was incubated at 37 °C for 1h with the following components in a 20 µl volume: 10 U T4 PNK, 2 µl 10 X Buffer A (500mM Tris-HCl, pH7.6 at 25°C, 100mM $MgCl_2$, 50mM DTT, 1mM spermidine) (Thermo Fisher Scientific, Waltham, MA, USA), 2 µl 10 µM probe solution, 1-5 µl P^{32}-labelled γ-ATP (PerkinElmer, Waltham, MA, USA) (Appendix B.9) and nuclease-free water. Probes were purified using a Sephadex ® -G25 spin column to remove unincorporated nucleotides (Appendix B.10). Before addition to the pre-hybridised membrane and buffer, purified probes were denatured at 80 °C for 5 min, followed by 5 min on ice and diluted with 1 ml hybridization buffer to ensure proper mixing.

After hybridization, membranes were washed to reduce background signal with various saline sodium citrate (SSC) and SDS solutions. SSC (20 X stock) comprises 3M sodium chloride and 0.3 M tri-sodium citrate, adjusted to pH 7.0 with 1 M HCl (Moore, 2001). Membranes hybridized with oligonucleotide probes were washed once with 50 ml 5 X SSC, 0.1 % w/v SDS for 20 min at room temperature; followed by 2 washes with 50 ml 1 X SSC, 0.1 % w/v SDS for 15 min at 42 ºC. Membranes hybridised with LNA-modified probes were washed twice with 50 ml 2 X SSC, 0.1 % SDS w/v for 10 min at 50 ºC (Varallyay et al., 2008). After exposure, membranes were stripped in a 30 min wash with 50 ml 1 % w/v/ SDS at 80 °C, before hybridization with a second probe. A maximum of 3 probe hybridisation cycles were permitted per membrane and U6 snRNA was always probed last.

Visualisation and analysis of pri-miRNA guide sequences

Probed membranes were visualised using a phosphor-imager (FLA-7000 Image Reader, Fujifilm, Kanagawa, Japan) and accompanying software. This approach was necessary as guides derived from pol II-driven pri-miRNAs are typically present at a low level and not easily detected using conventional autoradiography with X-ray film. The phosphor-imager technology involves the use of phosphor screens coated with a fine layer of accelerated phosphorescent fluorescent crystals, which can store radiation energy when exposed to a radioisotope signal. The phosphor-imager scans a phosphor screen with an excitation laser (650 nm) (Filter: B390/PMT-1) and fluorescence is emitted in relation to the level of radioisotope exposure and captured as a digital image. The resulting images have a higher sensitivity and resolution than those produced using X-ray film.

Fuji Film Multi Gauge software (v3.1, Science Lab 2005) was used to crop images and to quantify signal intensity of RNA guides. Bands representing RNA guides were identified from a

manually adjusted profile view using the auto-detection feature. This feature enables the detection of weak bands and the inclusion of larger guide signal regions, which may otherwise be partly excluded when using a uniform region of interest (ROI) to identify bands. This is important as both band size and intensity can indicate the amount of RNA guide present. A uniform rectangular ROI was, however, appropriate to quantify the signal intensity of U6 snRNA (loading control) bands which were generally consistent.

Band signal intensity values were expressed as PSL-BG (photo-stimulated luminescence minus background) to correct for background signal intensity, but not for area. Background (BG) signal for RNA guides was determined with an auto-detect function and a spline line setting for a more accurate estimation of the local background signal. BG signal for U6 snRNA bands was estimated from a rectangular ROI adjacent to identified bands. PSL-BG values for pri-miRNA and shRNA bands were then divided by the PSL-BG value of the respective U6 snRNA band, to correct for any minor differences in the amount of total RNA present on the blot. Final signal intensity values for pri-miRNA bands were calculated as a ratio with respect to the corrected PSL-BG value for the shRNA-derived guide on the same blot.

2.2.7 Analysis of pri-miRNA structures *in vitro*

In vitro transcription of pri-miRNAs

To investigate secondary RNA structure and conformation of pri-miRNAs with exceptional structures, pri-miRNAs with short and long stem lengths were transcribed *in vitro* for use in RNase mapping assays. Pri-miRNAs were transcribed from a T7 RNA polymerase promoter located upstream of the pri-miRNA sequence in pCI-neo-pri-miRNA expression vectors (Figure 2.1 C) (Section 2.2.2). pCI-neo-pri-miRNA expression vectors were purified using a commercially available mini-prep kit (Qiagen, Hilden, Germany) and re-suspended in RNAse-free water. Prior to transcription, pCI-neo-pri-miRNA vectors (20 - 30 µg) were linearised in a 3 h - 16 h restriction digest of the single *Not*I site located at the 3' end of the pri-miRNA sequence. A total of 1.2 µg pDNA was digested per 20 µl reaction with 10 U *Not* I. The digest reaction was then incubated with proteinase K (0.05 µg / µl final concentration) (Thermo Fisher Scientific, Waltham, MA, USA) and SDS (0.5 % w/v final concentration) at 50 °C for 45 min to remove potential RNAse contamination from the plasmid preparation and restriction digest steps. Proteinase K was then heat-inactivated by incubation at 65 °C for 20 min. Linearised pDNA was purified through phenol / choloroform / isoamyl alcohol (PCA) extraction, using phenol prepared

with Tris·Cl buffer (pH 8.0), and ethanol precipitation (Appendix B.11). Glycogen (20 µg) (Thermo Fisher Scientific, molecular biology grade), a highly purified polysaccharide derived from oysters, was added as an inert carrier per 120 µl aqueous phase prior to ethanol precipitation. Glycogen is insoluble in ethanol and enhances the visibility and handling of small DNA quantities following precipitation. The purified pDNA was typically re-suspended in a 1.5 µl of 10mM Tris buffer (pH 7.0) per 20 µl restriction digest and the volume was adjusted to give a final DNA concentration of 125 - 165 ng / µl. Total DNA recovery post-extraction was typically 20 - 30 %.

Linearised pri-miRNA expression vectors were used as the DNA template for *in vitro* transcription reactions, with termination by run-off at the cleaved *Not* I site. Pri-miRNA transcripts were expected to contain the full pri-miRNA flanked by an additional 13 bases at the 5' end and about 6 bases at the 3' end. A single transcription reaction (20 µl) was assembled with the following components, in order: DEPC-treated water, 1 X concentration of TranscriptAid reaction buffer, 10 mM of each ribonucleotide triphosphate (ATP, CTP, GTP, UTP), 0.75 - 1.0 µg linearised pDNA template in a volume of 6 µl and 2 µl TranscriptAid Enzyme Mix (T7 RNA polymerase and RiboLock™ RNase inhibitor) (TranscriptAid™ T7 High Yield Transcription Kit; Thermo Fisher Scientific, Waltham, MA, USA). The glycogen concentration in this reaction (~ 4 µg / µl) was below the prescribed maximum of 8 µg / µl and not expected to interfere with transcription. About six transcription reactions were required to ensure a sufficient yield of pri-miRNA transcripts for downstream applications. *In vitro* transcription reactions were incubated at 37 °C for 2.5 h, followed by the addition of 2 U DNase I (Thermo Fisher Scientific) and a further incubation at 37 °C for 15 min to digest the DNA template. RNA-handling techniques were used throughout to avoid sample degradation (Appendix B.8).

To ascertain if full length pri-miRNAs were successfully transcribed, products were analysed using denaturing PAGE, as described in 2.2.6. Polyacrylamide gels (8 % w/v) were cast using a vertical unit (10 cm x 8 cm glass plates; 1.5 mm spacers; SE250, Hoefer Inc, Holliston, MA, USA). A 1 µl aliquot of a transcription reaction was diluted in 20 µl DEPC-treated water and added to an equal volume of 2X RNA loading dye (Thermo Fisher Scientific) prior to loading. RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific) and Low Range ssRNA Ladder (New England Biolabs, Ipswich, MA, USA) were both used as molecular weight markers to provide reference bands of 100, 150, 200 and 300 bases. A 2 µl aliquot of ladder was diluted in 10 µl DEPC-treated water and added to an equal volume of 2X RNA loading dye. Gels were subjected to electrophoresis for 2 - 3 hours at 120 - 140 V to achieve the desired separation of

RNA transcripts. The size, integrity and proportion of full length transcripts in each reaction was assessed visually using a gel imaging system (G:Box, Syngene, Cambridge, UK).

RNA from successful transcription reactions was purified by PCA extraction with acid phenol (pH 4.5) and ethanol precipitation (Appendix B.11). Lithium chloride (2M final concentration) was used instead of sodium acetate to avoid precipitation of DNA contaminants and smaller RNA products of less than about 20 nt. The use of ammonium acetate (0.5 M final concentration) was avoided at this step to prevent inhibition of T4 polynucleotide kinase by ammonium ions in the subsequent radio-labelling reaction. Purified RNA was re-suspended in DEPC-treated water and pooled for each pri-miRNA. The RNA yield and quality was assessed using UV absorbance at 260 nm and 280 nm (NanoDrop 1000 Spectrophotometer) and the final RNA concentration was adjusted to 50 - 550 ng / µl.

Radio-labelling of transcripts

Pri-miRNA transcripts were prepared for 5' radio-labelling by the removal of the 5' phosphate group by alkaline phosphatase (EC 3.1.3.1) (Maunders, 1993a), (Liebecq, 1997). A total of 22 - 32 µg of RNA transcripts were dephosphorylated for each pri-miRNA, thus providing a maximum of 350 - 380 pmol of each full length pri-miRNA. A single dephosphorylation reaction was composed of the following in a 20 µl volume: 7 pmol RNA transcripts (1 - 14 µl), DEPC-treated water, 2 µl 10 X FastAP buffer, 3 U FastAP^{TM} Thermosensitive Alkaline Phosphatase and 20 U RiboLock^{TM} RNase inhibitor (Thermo Fisher Scientific, Waltham, MA, USA). Dephosphorylation reactions were scaled up by 50 - 55 X for each pri-miRNA and incubated at 37 °C for 1 h. Heat inactivation of alkaline phosphatase was omitted to avoid RNA degradation and RNA was purified directly by PCA extraction with acid phenol and ethanol precipitation. RNA-grade glycogen (30 µg) (Thermo Fisher Scientific) was again added as an inert carrier per 120 µl aqueous phase to enhance the visibility and handling of RNA pellets in subsequent reactions. Precipitated RNA from 120 µl aqueous phase was re-suspended in 2.5 µl DEPC-treated water to give a final glycogen concentration of 12 µg / µl. RNA from multiple dephosphorylation reactions was pooled and the yield was assessed by measuring UV absorbance at 260 nm. Final RNA concentrations were adjusted to 0.6 - 20 pmol / µl and total RNA recovery post-extraction was typically 30 - 35 %.

A total of 8.5 - 10 .5 µg of dephosphorylated RNA was radio-labelled for each pri-miRNA, thus providing a maximum of 110 - 135 pmol of each full length pri-miRNA. RNA was 5' radio-labelled using T4 polynucleotide kinase (PNK) in a forward reaction, similar to that used for 5'- radio-

labelling of oligonucleotide probes in northern analysis (2.2.6). T4 PNK can also catalyse the transfer of a γ-phosphate from ATP to the 5'-OH group of single- and double-stranded RNAs. Each labelling reaction was incubated at 37 °C for 1h with the following components in a 20 - 25.5 µl volume: 12 - 20 pmol dephosphorylated RNA in 1 - 20 µl, DEPC-treated water, 2 - 2.5 µl 10 X Buffer A (Thermo Fisher Scientific, Waltham, MA, USA),10 U T4 PNK, 40 U RiboLock™ RNase inhibitor and 1 µl P^{32}-labelled γ-ATP (PerkinElmer, Waltham, MA, USA). A total of 7 - 9 radio-labelling reactions, each with ~ 1.2 µg RNA, was typically required for each pri-miRNA. P^{32}-labelled γ-ATP was used within 4 days of the specific activity calibration date to ensure a strong radioisotope signal in RNase mapping. Labelled transcripts were purified from each reaction by PCA extraction with acid phenol, followed by ethanol precipitation and re-suspension in 15 µl water and 15 µl 2X RNA loading dye (Thermo Fisher Scientific).

PAGE purification of radio-labelled pri-miRNAs

Labelled, full length pri-miRNA transcripts (pri-miR-198 and pri-miR-34a) were isolated and purified from other transcripts by PAGE and gel extraction. Labelled transcription products were resolved on an 8 % w/v polyacrylamide gel, as described in 2.2.6. A 10 µl aliquot of radio-labelled Decade™ Marker was loaded in a 20 µl volume per gel, after preparation according to the manufacturer's protocol. A 5 µl aliquot of radio-labelled RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific) was also loaded in a 20 µl volume per gel, following dephosphorylation and 5' radio-labelling as described in Appendix B.12. Gels were subject to electrophoresis for 20 - 30 min at 100 V before sample loading, followed by 2.5 – 3 hours at 150 V to achieve the desired separation of RNA transcripts. Care was taken to avoid contamination with unincorporated P^{32}-labelled γ-ATP present in the bottom reservoir (anode) of the gel system.

Bands corresponding to full length pri-miRNA transcripts were identified using X-ray film (Fujifilm, Tokyo, Japan). X-ray film was placed on top of a gel for 5 - 10 min, developed (Appendix B.13) and placed back on the gel to mark the location of radio-labelled transcripts. Full length pri-miRNA transcripts were excised using a sterile needle and blade and eluted by a "crush and soak" method (Maxam and Gilbert, 1977). Each gel piece was crushed, re-suspended in 300 µl RNA elution buffer (0.5 M ammonium acetate, 1 mM EDTA, 0.2 % SDS) (Ambion protocol, MEGAshortscript™ kit, AM1354) soaked for 30 min at room temperature. The buffer was then removed and placed on ice and the soaking step was repeated. Pri-miRNA transcripts from a single gel piece were therefore eluted in 600 µl RNA elution buffer and purified by PCA extraction and ethanol precipitation. RNA-grade glycogen (40 µg) was added to the

aqueous phase for each 300 µl volume of extracted elution buffer. Purified transcripts were pooled for each pri-miRNA and re-suspended in a total volume of 28 µl of DEPC-treated water.

RNase mapping of pri-miRNAs

RNase mapping was used to interrogate the folded conformation of pri-miR-198, and pri-miR-34a *in vitro*. Radio-labelled pri-miRNAs were incubated with ribonuclease (RNase) enzymes to identify structurally unconstrained nucleotides. Each pri-miRNA was used in a series of RNase digests with RNase A or RNase T1 under denaturing or non-denaturing conditions and the resulting digest patterns were analysed on a sequencing gel. RNase A (EC 3.1.27.5) (Burrell, 1993), (Liebecq, 1997) specifically cleaves the phosphodiester bond at the 3' end of unpaired uridine (U) and cytidine (C) ribonucleotides, while RNase T1 (EC 3.1.27.3) (Liebecq, 1997) specifically cleaves the phosphodiester bond at the 3' end of unpaired guanosine (G) ribonucleotides. RNase digestion therefore produces a mixture of nucleoside 3'-phosphates and 3'phospho-oligonucleotides ending in a Cp, Up or Gp. The comparison of cleavage points in a non-denaturing reaction with those in a denaturing reaction revealed which nucleotides were structurally unconstrained in the folded pri-miRNA.

RNase A and T1 (Ambion, Austin, TX, USA) digest series were performed with accompanying buffers and according to the manufacturer's protocol with only a few minor modifications. An RNase digest series under denaturing conditions, essentially an RNA sequencing reaction, was carried out as follows. A 4 µl aliquot of radio-labelled pri-miRNA was mixed with 3 µg yeast RNA (3 µg / µl) and 22 µl 1X RNA sequencing buffer (20 mM sodium citrate, pH 5.0, 1 mM EDTA, 7 M urea) in a final volume of 27 µl. This reaction mixture was divided in to three 9 µl aliquots in tubes labelled 1, 2 and 3. All 3 aliquots were heated to 50 °C for 5 min on an electronic heating block to allow for RNA denaturation and then cooled at room temperature for 5 min. RNase A or T1 (1 U) (1 U / µl) was added to tube 2 and mixed by pipette action (10X RNase dilution). A 1 µl volume of the reaction mixture in tube 2 was added to tube 3 mixed by pipette action (100X RNase dilution). No RNase was added to tube 1 and this sample served as a control for undigested RNA. All three tubes were incubated at room temperature for 15 min, followed by the addition of 20 µl inactivation / precipitation buffer (ethanol, isopropanol, guanidine isothiocyanate) and incubation at - 20 °C for 1 - 5 h.

An RNase digest series under non-denaturing conditions was performed with the same reaction scheme, but with the following amendments. The 1X RNA sequencing buffer was replaced with 22 µl 1X RNA structure buffer (100 mM Tris, pH 7.0, 1 M KCl, 100 mM $MgCl_2$). The incubation at

50 °C and subsequent cooling step were omitted. Also, after the addition of the precipitation buffer, the reaction was mixed for 2 - 5 s to ensure RNase inactivation. RNA from both types of RNase digest reactions was collected following centrifugation at 13 200 rpm and 4 °C for 20 -30 min (Eppendorf, 5424R, Hamburg, Germany). RNA pellets were washed with 20 µl 70 % ethanol and air-dried for 2 min, before re-suspension in 7 µl RNA loading dye (Ambion, Austin, TX, USA).

Each pri-miRNA was also used in an alkaline hydrolysis reaction series to produce a hydroxylated RNA ladder with single nucleotide resolution. The 2' hydroxyl (OH) group of the ribose moiety in RNA renders the molecule susceptible to degradation in solutions above pH 8. Increasing pH facilitates equilibrium deprotonation of the 2' OH group, resulting in nucleophilic attack of the phosphate group of the RNA backbone by the remaining 2' oxygen and the subsequent release of individual nucleotides (Donald C. Rio, 2011). A 4 µl aliquot of radio-labelled pri-miRNA was mixed with 3 µg yeast RNA (3 µg / µl) and 10 µl 1X alkaline hydrolysis buffer (50 mM sodium carbonate, pH 9.2) in a final volume of 15 µl. This reaction mixture was divided in to three 5 µl aliquots in tubes labelled 1, 2 and 3. All 3 aliquots were heated to 95 °C. Tube 1 was removed and placed on ice after 2 min, tube 2 after 5 min and tube 3 after 15 min. RNA loading dye (10 µl) was added to each tube before electrophoresis.

RNase-treated pri-miRNAs were resolved on a 5, 8 or 12 % polyacrylamide sequencing gel alongside a hydroxylated ladder of the same transcript and aliquots of radiol-labelled Decade^{TM} Marker (1:2 dilution) and RiboRuler (1:5 dilution of radio-labelled stock, Appendix B.12). Sequencing gels were assembled and run following established protocols (Slatko and Albright, 2001) and recommendations by the apparatus manufacturer (Sequi-Gen GT System, Bio-Rad, Hercules, CA, USA). Denaturing polyacrylamide sequencing gels were prepared as described for PAGE (2.2.6.), but with polymerisation initiated by the addition of 600 µl of a 10 % w/v APS and 60 µl TEMED. A 1X TBE running buffer was used for electrophoresis. Glass plates were generally not siliconised as this was found to be unnecessary, however, they were thoroughly washed with soap paste and wiped with 70 % ethanol prior to pouring the gel. Gel dimensions were approximately 40 cm x 20 cm x 0.4 mm (length x width x thickness). Gels were subject to electrophoresis at 40 - 50 W prior to sample loading to obtain an optimal gel temperature of ~ 50 °C. Urea and debris were flushed from the sample well before insertion of the sharkstooth comb. All RNA samples and ladders were denatured at 85 °C for 5 min prior to loading 3 µl per lane. A gel temperature indicator (Bio-Rad, Hercules, CA, USA) was used to monitor and maintain gel temperature at ~ 50 °C. Gels were subjected to electrophoresis at 40 - 50 W for 2 - 3 h for to

achieve desired resolution. Multiple reaction sets were often staggered by 1 - 2 h on a single gel, increasing the total electrophoresis time.

After electrophoresis, the gel was fixed to remove urea and improve the clarity of the ^{32}P signal. The gel was removed from a single glass plate by floating and soaked in 2 l of fixative solution (5 % v/v methanol, 5% v/v acetic acid) for 10 min in a shallow tray. A larger glass plate (40 cm x 40 cm) (length x width) was placed at the bottom of the tray and used to raise the gel out of the fixative solution. A piece of Grade 3 MM Whatman® paper (Whatman®, Maidstone, UK) (22 cm x 42 cm) was pressed onto the gel and used to peel it off the glass plate, before adding a layer of protective plastic cling wrap and drying at 60 °C for 30 min using a vacuum gel dryer (Hoefer Inc., Holliston, MA, USA). Excess paper was then trimmed and the gel surface was covered with a new layer of plastic cling wrap prior to exposure of a phosphor-imager screen of the same size.

Visualisation and analysis of RNase-generated maps

Signals were detected using a phosphor-imager (FLA-7000 Image Reader, Fujifilm, Kanagawa, Japan) and phosphor-imager screens were exposed to dried sequencing gels for 44 h -72 h. Digest patterns were manually aligned to the pri-miRNA reference sequence and the positions of RNase cleavage sites were identified. This information was used to annotate experimentally unconstrained nucleotides of the existing pri-miRNA structures predicted by mfold (Zuker, 2003). Optimal results were often obtained from hydrolysis reactions with a 5 min incubation at 95 °C, while both 10 X and 100 X RNase digests provided useful results.

2.2.8 Statistical analyses

Microsoft Excel 2007 and GraphPad Prism version 4.03 were used to construct graphs and perform statistical analyses. In the results of dual luciferase assays, standard deviation was used as a measure of sample variation and one-way ANOVA (analysis of variance) was used to compare sample means. A Dunnett's post-test was to compare the means of experimental samples with that of a control (mock) sample. A two-tailed P value was used to assess the probability that the same variation in the means could be expected from random sampling.

2.3 RESULTS

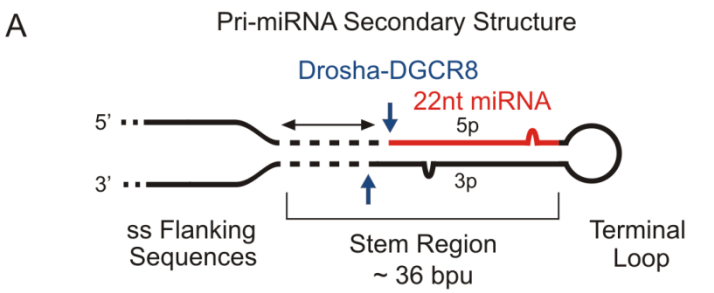
2.3.1 Selection of naturally occurring pri-miRNAs with exceptional basal stems

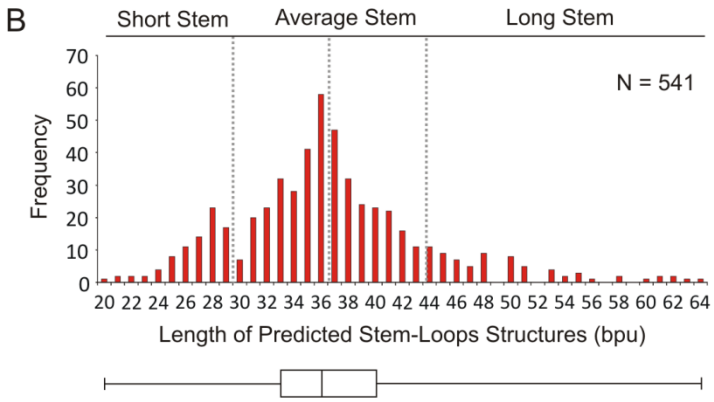
The secondary RNA structure of a typical pri-miRNA (Figure 2.5 A) harbours a stem region of ~ 3 helical turns composed of ~ 36 base pair units (bpu), including both conventional base pairs (~ 33 bp) and internal bulges. To identify pri-miRNAs with potentially exceptional basal stems, a survey of stem length was performed for predicted stem-loop structures on the online miRNA repository miRBase (Release 10.1) (Griffiths-Jones et al., 2008) (http://www.mirbase.org/, accessed February 2008). Stem-loop sequences were assessed for 541 *Homo Sapiens* miRNAs available at the start of the study, which were experimentally identified in humans or related organisms. Predicted stem lengths were counted in base pair units (bpu) to include canonical Watson-Crick base pairs, G - U wobble pairs and both symmetrical and asymmetrical non-terminal mismatches. A basic analysis of the population distribution (Figure 2.5 B) was used to define three stem length categories as short (≤ 29 bpu), average (30 - 43 bpu) and long (≥ 44 bpu). Average length stem-loop structures were defined as those within in one standard deviation of the mean (71 % of entries), while structures more than one standard deviation below or above the mean, were defined as short or long, respectively. Candidate stem-loop structures were selected from each category for the following miRNAs: miR-520b, miR-198, miR-147a (short); miR-23a, miR-934 (average); and miR-34a and miR-513a-2 (long). The criteria for a valid stem-loop structure include free energy of less than - 0.2 kcal / mol / nt and pairing in at least 60 % of the miRNA sequence (Kozomara and Griffiths-Jones, 2014). These criteria were met for the selected stem-loops, with the exception of pairing for miR-198 (59.0 %) (Figure 2.5 C). Stem-loop sequences were extended by 50 nt in either direction to include the anticipated flanking regions of the full pri-miRNA and cloned into a mammalian expression vector (pCI-neo, Promega, Madison, WI, USA) for functional characterisation studies.

2.3.2 Predicted secondary RNA structures of selected pri-miRNAs

An important criterion in the identification miRNAs is the formation of a predicted stem-loop structure in the miRNA-containing sequences. Stem-loop structures available on miRBase were typically generated through computational prediction without further experimental validation and compiled from multiple sources (Griffiths-Jones et al., 2008). As a result, the length of predicted stem-loops on the database is somewhat arbitrary and the stem length of an individual pri-miRNA hairpin may be shorter or longer. Therefore, while the initial survey of stem-loop

structures on miRBase was useful to identify potentially exceptional structures, further validation of the predicted stem length was required.





C

Stem-loop	Stability (kcal / mol / nt)	Pairing (%)
520b	- 0.59	95.0
198	- 0.25	59.0
147a	- 0.34	70.0
23a	- 0.45	85.7
934	- 0.66	90.9
34a	- 0.46	90.9
513a-2	- 0.46	78.3

Figure 2.5: The identification of pri-miRNAs with exceptional stem conformations.

(**A**) The typical secondary RNA structure of a pri-miRNA, including single stranded flanking sequences, a stem region of 36 base pair units (bpu) and a terminal loop. The mature ~ 22nt miRNA (red) is positioned in the upper stem, while the more variable lower stem is indicated (dashed lines). (**B**) The unimodal distribution of 541 stem-loop structures on the miRBase repository (Release 10.1) according to predicted stem length (20 - 64 bpu). Dashed lines indicate the population mean (36.6 bpu) and one standard deviation (7.1 bpu) above and below the mean. Box plot: lower quartile = 33; median = 36; upper quartile = 40 bpu. (**C**) The stability (kcal / mol / nt) and pairing (%) of the mature miRNA sequence for selected stem-loops.

To fully examine the predicted stem length of selected pri-miRNAs, predicted secondary RNA structures were generated using the mfold web server (Zuker, 2003) with standard sequence extensions (15 - 50 nt) flanking the anticipated Drosha-DGCR8 cleavage site (Appendix A). Predicted secondary RNA structures were also generated for RNA sequences derived from the expression vector, to assess folding of the expressed pri-miRNA sequence. Predicted structures were assessed for the formation of a canonical hairpin, defined as a structure with a terminal loop, an upper stem harbouring the miRNA, a basal stem of 8 - 13 bp and at least 4 nt of unstructured residues flanking the expected Drosha-DGCR8 site on one side. While a summed total of 9 - 18 nt flanking residues are optimal for recognition and processing, cleavage can still occur with a minimum of 4 unstructured residues on one side (Auyeung et al., 2013).

Hairpins with canonical basal stem lengths were present in the most stable pri-miRNA structures generated with 15 - 20 nt of flanking sequence, except for pri-miR-198 and pri-miR-147a (Figure 2.6 A). The stem of most selected pri-miRNAs can therefore be considered as typical in structures including the local flanking sequence. Basal stem regions (bp) were counted to include Watson-Crick pairs and G - U wobble pairs, while the potential contribution of single symmetrical mismatches was also considered (Table 2.8). The basal stem of pri-miR-520b, for example, only contains 7 bp, but two non-Watson-Crick pairs may also stack within the stem and contribute to its length (~ 9 bp). The basal stem of pri-miR-34a can be counted as 9 bp to include only single G - U wobble pairs, but an internal bulge formed by two consecutive G - U pairs may also contribute to its length (~ 11 bp). Predicted structures of pri-miR-198 and pri-miR-147a can only be considered as canonical if triple internal bulges, like those found in the upper stem, are tolerated to give basal stems of ~ 10 bp.

In contrast, canonical hairpins were not necessarily present in structures generated with the extended flanking sequence (30 - 50 nt) or vector-derived sequence (Figure 2.6 B) (Appendix A). Only a truncated hairpin was generated for pri-miR-520b, while an extended basal stem of ~ 18 - 22 bp was common for pri-miR-34a, in agreement with the initial “short” and “long” stem-loop classifications. Similarly, the short basal stem of pri-miR-147a was followed by a larger loop that could not be considered as part of a typical stem, with or without tolerance of triple internal bulges. The basal stem of pri-miR-198 was also truncated by an alternative structural element, although the hairpin itself was commonly disrupted. The expected stem-loop structure of pri-miR-513a-2 with extended pairing interactions was observed, but the basal stem was more easily defined as conventional (~ 10 bp) or truncated. In fact, canonical hairpin structures were only well-maintained in the expressed pri-miRNA sequence for pri-miR-23a and pri-miR-934.

Predicted interactions with the extended flanking sequence therefore either supported or disrupted the formation of a conventional basal stem and canonical hairpin.

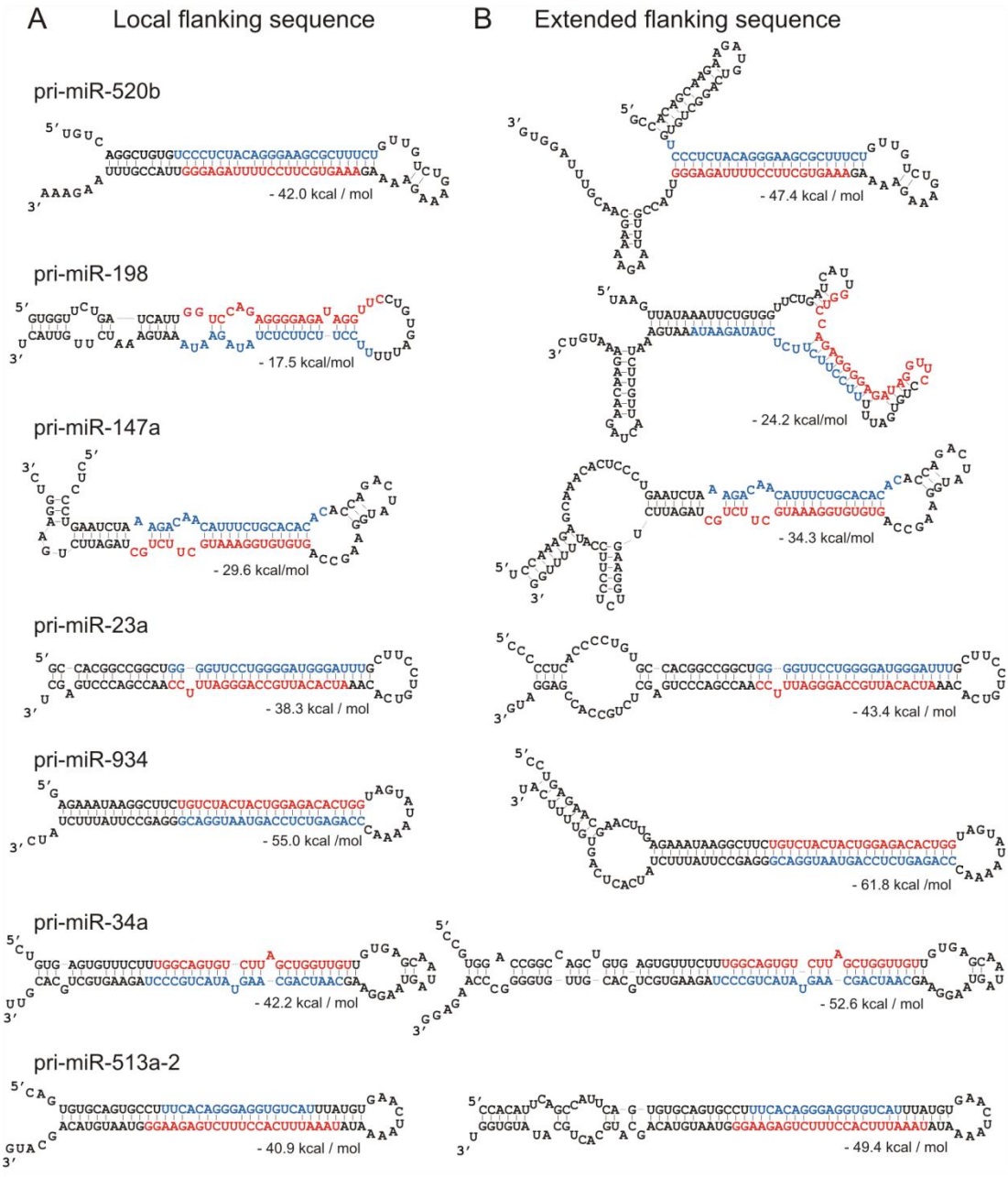


Figure 2.6: Predicted secondary RNA structures of pri-miRNAs with exceptional basal stem conformations.

Predicted pri-miRNA structures generated with both (**A**) local (15 nt) and (**B**) extended (30 nt) sequence flanking the anticipated Drosha-DGCR8 cleavage site. The major guide sequence (red), minor guide sequence (blue) and Gibbs free energy (dG) of each structure are indicated.

Table 2.8: The basal stem length (BSL) and free energy (dG) of predicted secondary RNA structures of pri-miRNAs.

Input Sequence	Pri-miR-520b		Pri-miR-198		Pri-miR-147a		Pri-miR-23a		Pri-miR-934		Pri-miR-34a		Pri-miR-513a-2	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
+ 15 nt	9	- 42.0•	5	-17.5•	7	-29.6•	13	- 38.3•	12	-55.0•	11	-42.2•	10	-40.9•
+ 20 nt	9 4 9	- 42.0• - 40.0 - 39.7	5 NH	-20.3• -17.1	7 7	-29.6• -28.7	13	- 38.3•	12	-55.0•	15•	-45.2•	10 7 10	-40.9• -40.5 -39.9
+ 50 nt	0 0 0 0	- 59.9• - 56.0 - 53.9 - 52.8	NH	-28.5•	7 7	-48.2• -47.0	13 13 13 5 0 13 5	- 66.5• - 62.9 - 61.2 -61.7 - 61.5 - 63.8 -63.5	12	-68.4•	15 22 18 15 15 19	-63.8• -59.9 -59.6 -60.5 -58.6 -56.4	13 10 10	-65.9• -60.3 -59.0
Full pri-miRNA	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	- 215.4 -221.0• - 212.4 - 216.0 - 212.5 - 204.3 - 204.4 - 212.9 - 217.1 - 211.6 - 206.6 - 211.2 - 195.5 - 200.7 - 199.4 - 199.9 - 205.8 - 206.6 - 203.4 - 200.2 - 208.1	NH NH NH 3 3 NH 3 NH 3 NH NH NH NH NH 3 3 3 3 3 NH 3 NH	-188.9 -185.1 -187.4 -200.2• -196.7 -198.6 -192.5 -193.9 -183.5 -174.2 -172.9 -188.2 -168.8 -185.0 -186.8 -195.3 -186.5 -189.3 -185.1 -188.9 -193.1 -179.1 -187.0 -184.4	7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 NH	-224.7• -215.2 -211.5 -210.4 -213.4 -212.3 -205.4 -204.2 -213.8 -216.3 -220.3 -208.7 -214.7 -199.8 -206.0 -200.6 -199.3 -209.6 -200.4 -206.3	8 8 13 13 13 8 0 13 0 13 5 13 13 0 8 8 0 8 13 NH 13 0 NH 8 8 NH NH 13	- 241.0• - 237.2 - 238.4 - 236.7 - 237.4 - 229.9 - 239.7 -231.0 - 233.2 - 225.0 -235.1 - 228.4 - 233.5 - 235.1 - 226.3 - 225.0 - 234.9 - 220.9 - 231.8 -234.3 - 235.7 - 230.7 -228.6 - 224.2 -220.4 -235.6 -224.6 -222.7	12 12 12 12 12 12 12 12 12 12 11 11 12 12 12 12 11 NH	-240.8 -237.0 -237.1 -229.7 -241.2• -239.4 -237.2 -243.3 -222.5 -224.8 -232.2 -234.0 -227.8 -236.3 -237.4 -234.6 -227.5 -214.9	11 11 11 18 11 11 11 11 11 11 11 11 18 11 4 22 NH 22 NH NH	-245.8• -240.5 -245.1 -240.4 -232.4 -234.8 -240.9 -245.6 -233.4 -242.4 -238.6 -241.8 -236.7 -238.3 -232.2 -244.7 -232.2 -236.9 -239.5 -229.1	0 0 0 0 10 0 10 0 10 0 0 0 13 0	-241.3 -237.5 -248.1• -246.3 -238.9 -230.2 -245.2 -227.7 -236.7 -239.4 -235.7 -225.3 -242.2 -233.0

The input sequence used for mfold structure prediction included a specific length of flanking sequence on either side of the anticipated Drosha-DGCR8 cleavage site (e.g. + 15 nt) or the full pri-miRNA sequence of the expected transcript. The BSL was defined as the duplex region between the anticipated Drosha-DGCR8 cleavage site and the nearest disruptive element (a fork or any bulge of 3 nt or more). The contribution of G-U wobble pairs and non-canonical pairs was also included where appropriate. A BSL entry with a solid underline indicates that the secondary RNA structure adopted a characteristic hairpin form, including a conventional basal stem (8 - 13 bp) flanked by at least 4 nt of consecutive ssRNA on one side. A BSL entry with a dashed underline indicates that the length of the flanking sequences was insufficient to enable a characteristic hairpin form. The most stable predicted secondary RNA structure in each data set is indicated (•). NH: No Hairpin.

2.3.3 Functional characterisation of pri-miRNAs with different basal stem conformations

Predicted secondary RNA structures are not necessarily a perfect indication of dynamic RNA conformations in the cell. However, structural prediction results indicated that the predicted basal stem lengths were not necessarily optimal at ~ 11 bp and ranged from 5 - 15 bp in predicted structures with 15 or 20 nt of flanking sequence. Canonical hairpins were also more successfully maintained for certain pri-miRNAs in the expressed pri-miRNA sequence, while non-canonical, or even non-hairpin, structures of similar stability were predicted for other pri-miRNAs. Variation in the length and conformation of the basal stem could conceivably affect the accuracy and efficiency of processing by the Drosha molecular "ruler". This could result in reduced miRNA-mediated silencing by certain pri-miRNAs, making them undesirable candidates in the design of effective pri-miRNA mimics. The potential effects of basal stem variation were examined through the functional characterisation of the selected pri-miRNAs.

Target-specific silencing was investigated in cell culture through dual luciferase reporter assays and indicated by a reduction in the ratio of *Renilla* luciferase : Firefly luciferase activity (h*R*luc : hFluc). A short hairpin RNA (shRNA) expressing the major guide sequence was included for each pri-miRNA as a positive control of guide function. A non-targeting pri-miRNA (pri-non) and non-targeting shRNA (sh-non) were also included as negative controls. Effective target silencing (63 - 91 %) by the major miRNAs was observed for all pri-miRNAs, except pri-miR-198 and pri-miR-147a where no target silencing was detected (Figure 2.7 A). The most effective target silencing was observed for pri-miR-934, while silencing of ~ 80 % or more was observed for pri-miR-520b, pri-miR-34a and pri-miR-513a-2. Effective target silencing was only observed for miR-34a-3p (68 %) and miR-513a-2-5p (92 %) minor miRNAs* derived from the complementary arm of the pri-miRNA (Figure 2.7 B). Reporter gene silencing was specific to individual guide-target interactions, as no silencing was observed in the absence of a target sequence (Figure 2.7 C). As the function of certain pri-miRNAs can be regulated by specific accessory factors, silencing assays were repeated in a second cell line (Huh-7). Similar results were obtained with effective target silencing (59 - 94 %) for major miRNAs, except those derived from pri-miR-198 and pri-miR-147a, and minor miRNAs* miR-34a-3p (73 %) and miR-513a-2-5p (92 %) (Figure 2.8 A, B).

Stability and pairing of the pri-miR-198 and pri-miR147a hairpins were lower than that of other functional hairpins, which had stability of less than - 0.44 kcal / mol / nt and > 75 % of pairing in the mature sequence (Figure 2.5 C). However, the lack of function for pri-miR-198 and pri-miR-147a was unexpected as the full pri-miR-198 sequence was successfully transcribed *in vitro*

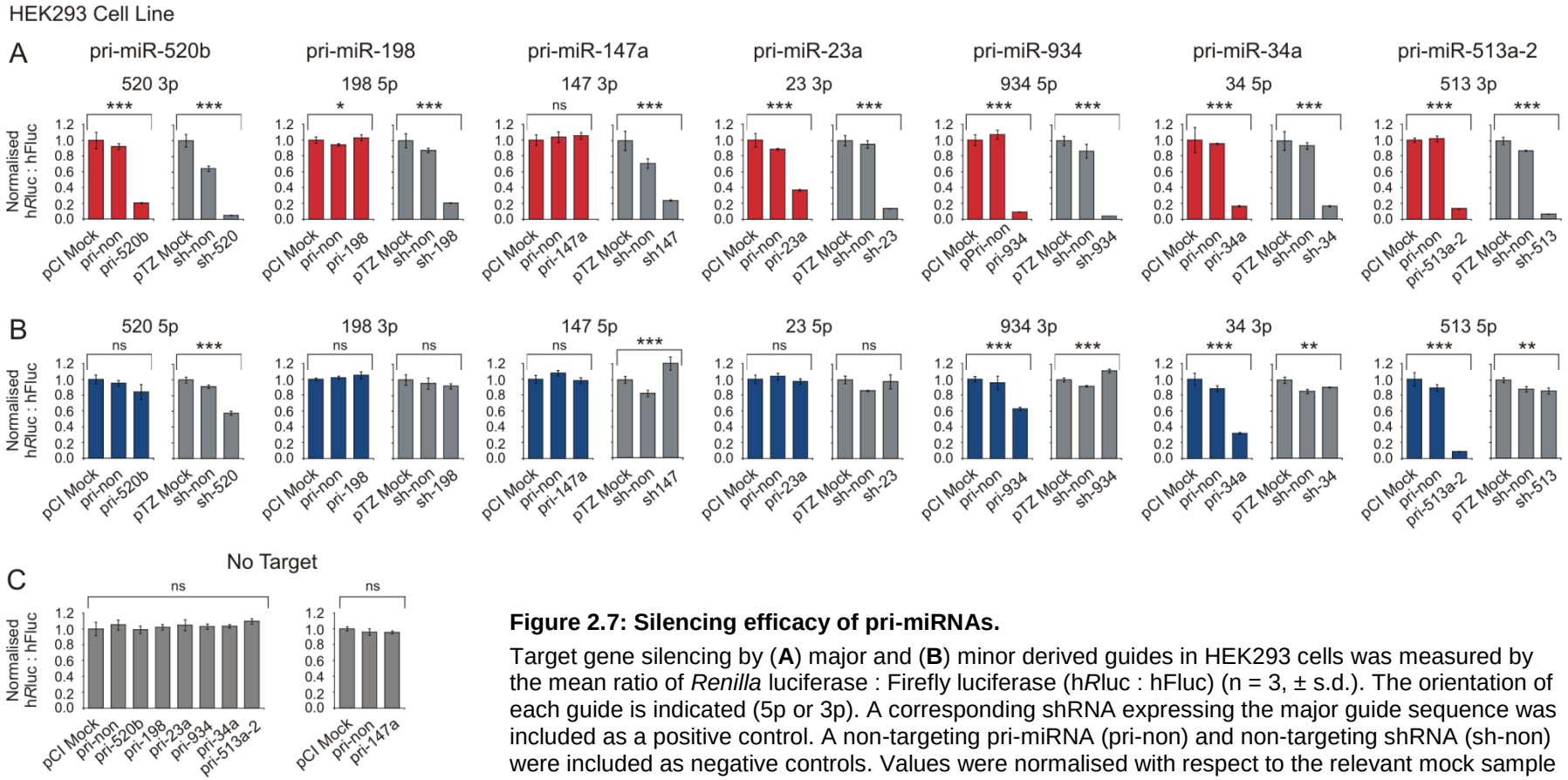


Figure 2.7: Silencing efficacy of pri-miRNAs.

Target gene silencing by (**A**) major and (**B**) minor derived guides in HEK293 cells was measured by the mean ratio of *Renilla* luciferase : Firefly luciferase (h*R*luc : hFluc) (n = 3, ± s.d.). The orientation of each guide is indicated (5p or 3p). A corresponding shRNA expressing the major guide sequence was included as a positive control. A non-targeting pri-miRNA (pri-non) and non-targeting shRNA (sh-non) were included as negative controls. Values were normalised with respect to the relevant mock sample (pCI or pTZ Mock). (**C**) Dual luciferase assays were repeated with a control luciferase vector lacking the relevant target sequence (empty psiCHECK^{TM}-2). Variation of sample means was significant (one-way ANOVA, ***P < 0.001, **P < 0.01, *P < 0.05) or not significant (ns).

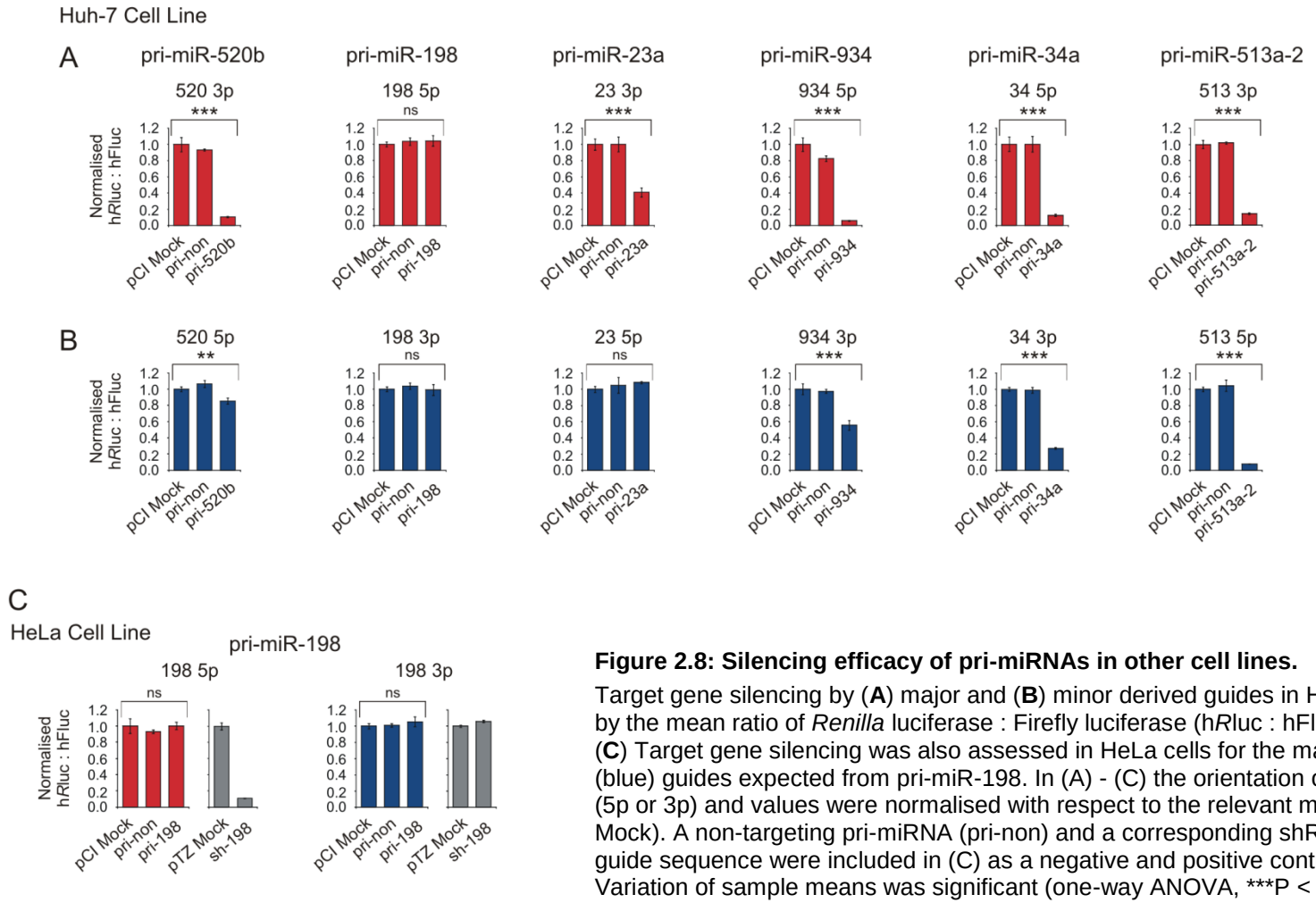


Figure 2.8: Silencing efficacy of pri-miRNAs in other cell lines.

Target gene silencing by (**A**) major and (**B**) minor derived guides in Huh-7 cells was measured by the mean ratio of *Renilla* luciferase : Firefly luciferase (h*R*luc : hFluc) (n = 3, ± s.d.). (**C**) Target gene silencing was also assessed in HeLa cells for the major (red) and minor (blue) guides expected from pri-miR-198. In (A) - (C) the orientation of each guide is indicated (5p or 3p) and values were normalised with respect to the relevant mock sample (pCI or pTZ Mock). A non-targeting pri-miRNA (pri-non) and a corresponding shRNA expressing the major guide sequence were included in (C) as a negative and positive control, respectively. Variation of sample means was significant (one-way ANOVA, ***P < 0.001, **P < 0.01) or not significant (ns).

from the T7 promoter (see 2.3.6) ; the miR-198 stem-loop was processed *in vitro* using cytoplasmic HEK293T cell extracts (Sundaram et al., 2013); and miR-198 and miR-147a guide sequences were functional when derived from Dicer-dependent shRNAs or introduced as synthetic miRNAs: miR-198 (Sundaram et al., 2013; Sung and Rice, 2009; Tan et al., 2011; Ye et al., 2013), miR-147a (Bertero et al., 2012; Lee et al., 2014; Uhlmann et al., 2012). A KH-type splicing regulatory protein (KSRP) has recently been shown to enhance processing of miR-198 from the endogenous transcript by binding to a terminal GUG loop motif (Sundaram et al., 2013). As KSRP-mediated processing was observed in both HEK293-derived and HeLa cell lines (Sundaram et al., 2013; Trabucchi et al., 2009), the function of pri-miR-198 was also assessed in the HeLa cell line, but target silencing was still not observed (Figure 2.8 C).

To investigate potential inhibitory effects of the expression cassette for pri-miR-198 and pri-miR-147a, alternative polymerase II expression constructs were generated for both functional and non-functional pri-miRNA sequences (Figure 2.9 A). Similar results were obtained for both major and minor miRNAs without any improvement in target silencing by pri-miR-198 or pri-miR-147a (Figure 2.9 B, C). Target silencing by effective miRNAs derived from pri-miR-520b and pri-miR-34a was generally reduced by ~ 20 - 30 %, presumably as a consequence of slightly reduced expression from the SV40 promoter, while silencing by positive control shRNAs was consistent (Figure 2.9 D).

The long-stem conformation of pri-miR-34a could, in principle, enable alternative processing sites. Drosha-DGCR8 cleavage at ~ 11 bp from the basal ssRNA - dsRNA junction of this conformation could result in an alternative miRNA species (Figure 2.10 A). Alternatively, consecutive Dicer processing could produce multiple ~ 22 nt miRNAs from one arm of the pri-miRNA, as has been observed for a murine precursor (mmu-miR-3102) (Chiang et al., 2010) and long hairpin RNAs (lhRNAs) (Liu et al., 2007; Saayman et al., 2008; Sano et al., 2008) (Figure 2.10 B). The potential generation of alternative, functional miRNAs from the lower stem (LS) of pri-miR-34a was assessed through dual luciferase assays, but no silencing of corresponding target sequences was observed (Figure 2.10 C).

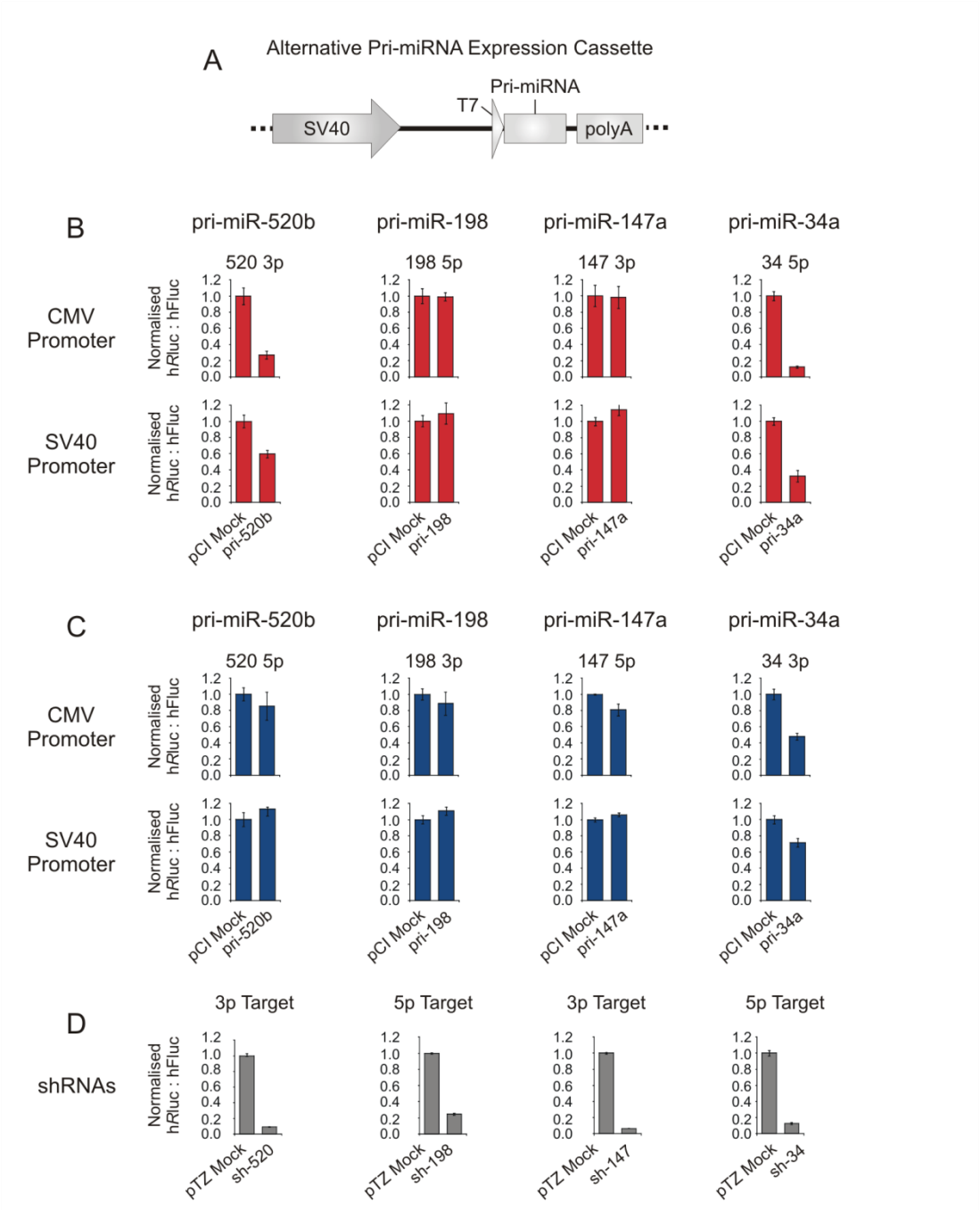


Figure 2.9: Silencing efficacy of pri-miRNAs generated from an SV40 expression cassette.
(**A**) An alternative pol II expression cassette was generated with an SV40 (simian virus 40) enhancer and early promoter region. Target gene silencing by (**B**) major and (**C**) minor guides derived from a CMV or SV40 pol II expression cassette was assessed in HEK293 cells. Silencing was measured by the mean ratio of *Renilla* luciferase : Firefly luciferase (h*R*luc : hFluc) (n = 3, ± s.d.). (**D**) ShRNAs expressing each major guide sequence were included as positive controls. All values were normalised with respect to the relevant mock sample (pCI or pTZ Mock).

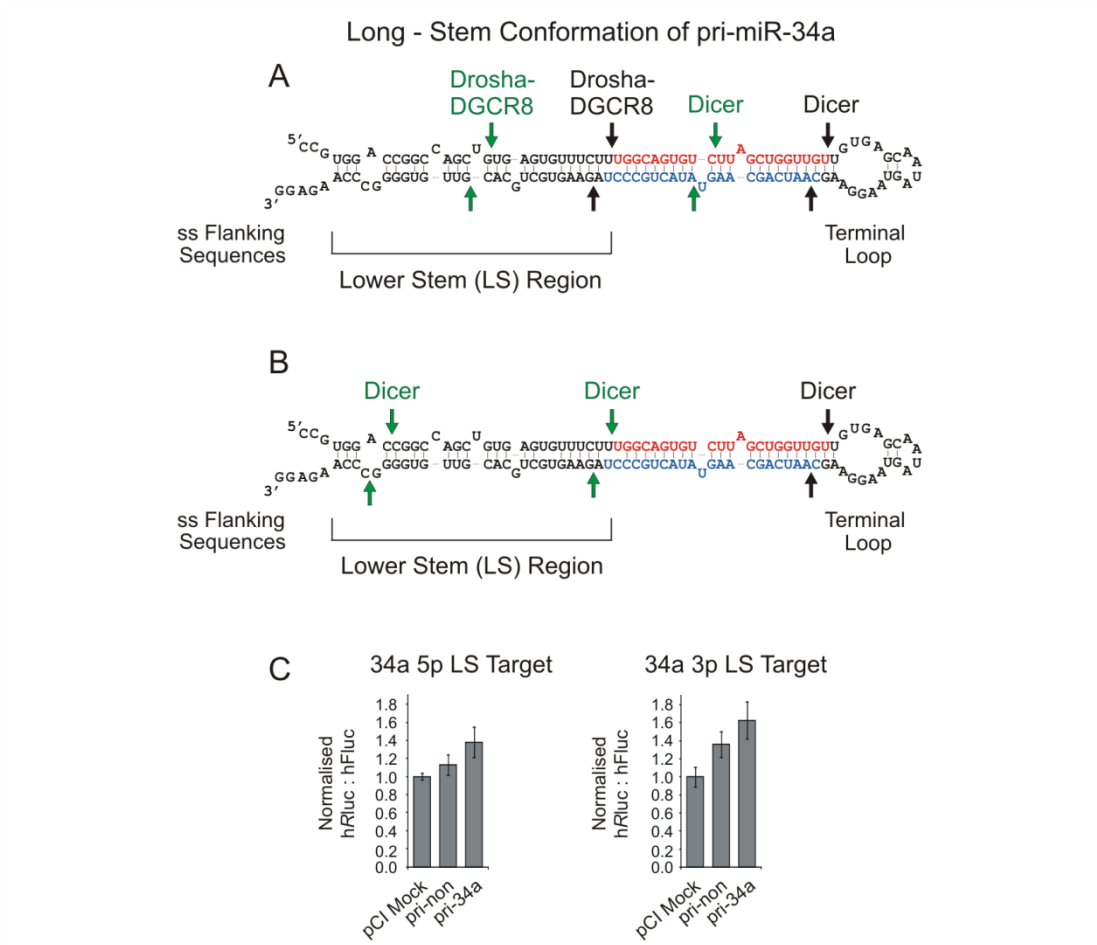


Figure 2.10: Potential alternative processing of a pri-miRNA with a longer stem.

The longer stem of pri-miR-34a could enable (**A**) a shift in the Drosha-DGCR8 processing site or (**B**) consecutive Dicer processing as observed for long hairpin RNAs (lhRNAs). (**C**) Silencing of lower stem (LS) target sequences was assessed in HEK293 cells and measured by the mean ratio of *Renilla* luciferase : Firefly luciferase (h*R*luc : hFluc) (n = 3, ± s.d.). Values were normalised with respect to the mock sample (pCI Mock) and a non-targeting pri-miRNA (pri-non) was included as a negative control.

2.3.4 Processing of miRNA guides from pri-miRNAs with different basal stem conformations.

Results from silencing assays suggested that pri-miRNA function was not restricted by minor variation in the length of the basal stem or additional non-canonical conformations of the expressed pri-miRNA sequence, provided that a canonical hairpin was predicted for the local pri-miRNA sequence. To investigate pri-miRNA processing, northern blot analysis was performed on total RNA extracted from HEK293 cells transfected with the mock, non-targeting pri-miRNA, pri-miRNA or shRNA expression vectors. Derived RNA sequences were detected through hybridisation with complementary oligonucleotide probes. Mature ~ 22 nt miRNA guides were detected for each pri-miRNA that mediated effective target silencing (Figure 2.11). The relative signal intensity of pri-miRNA-derived guides differed, but the use of different probes meant that these values were not a comparable indicator of the magnitude of target silencing. The relative detection ratio of miR-23a-3p (3.6), for example, was higher than that of miR-513-3p (0.9), even though miR-23a-3p mediated less effective target silencing. Complementary strands of the miRNA duplex were also detected for each functional pri-miRNA (Figure 2.11 B). However, only miR-34a-3p and miR-513a-2-5p appeared to function as effective miRNAs* in target silencing assays, implying that these minor miRNAs can also be effectively loaded into RISC. Pre-miRNAs (~ 65 nt) were detected for effective pri-miRNAs, thus supporting both Drosha-DGCR8 and Dicer cleavage at the anticipated processing sites.No pre-miRNA or miRNA products were detected for pri-miR-198 or pri-miR-147a, suggesting that these sequences were not processed in the miRNA biogenesis pathway. RNA bands detected in all four samples of a blot were considered as non-specific endogenous products, while probed U6 small nuclear RNAs (snRNAs) of ~ 107 nt indicated loading consistency.

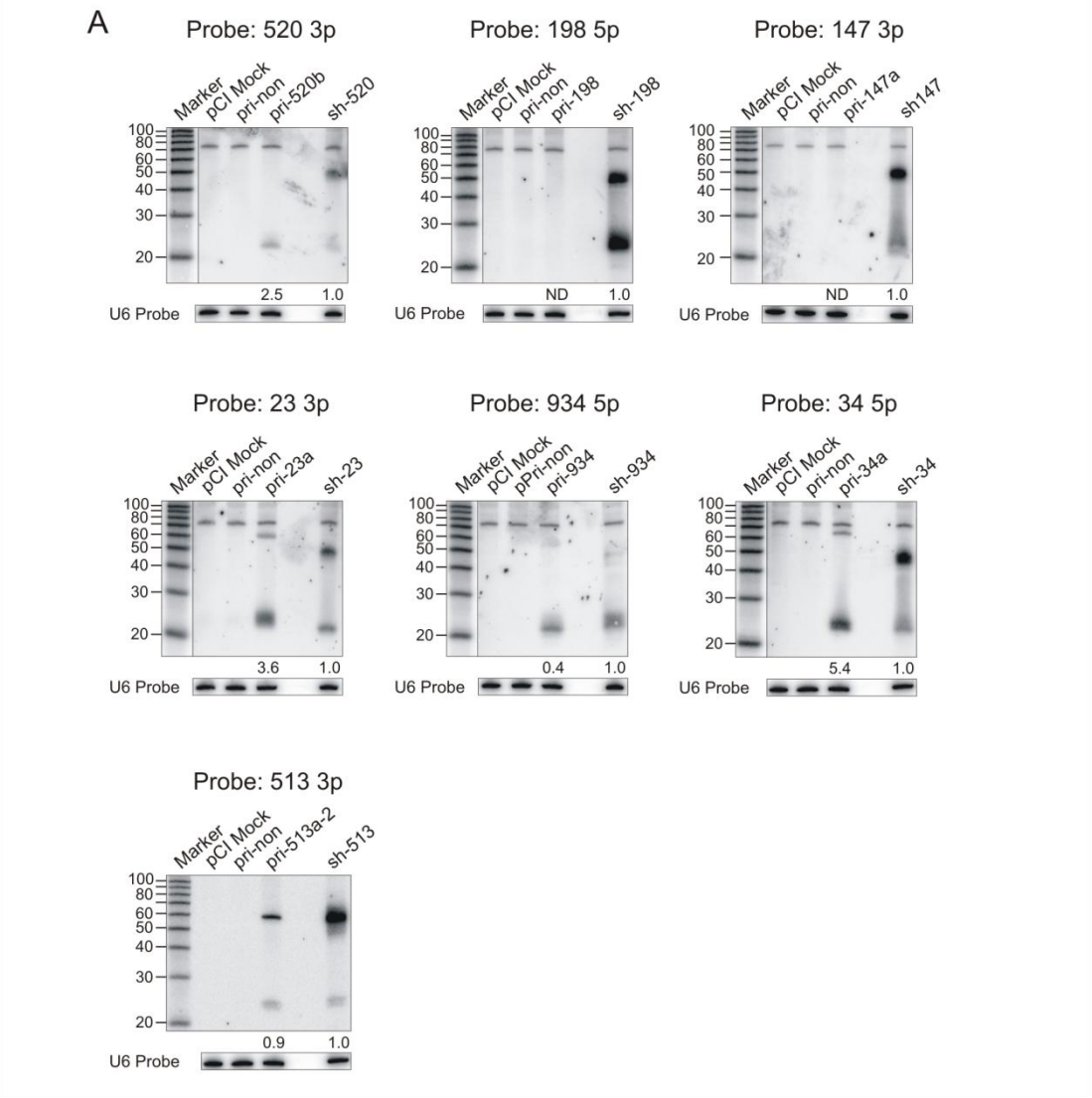
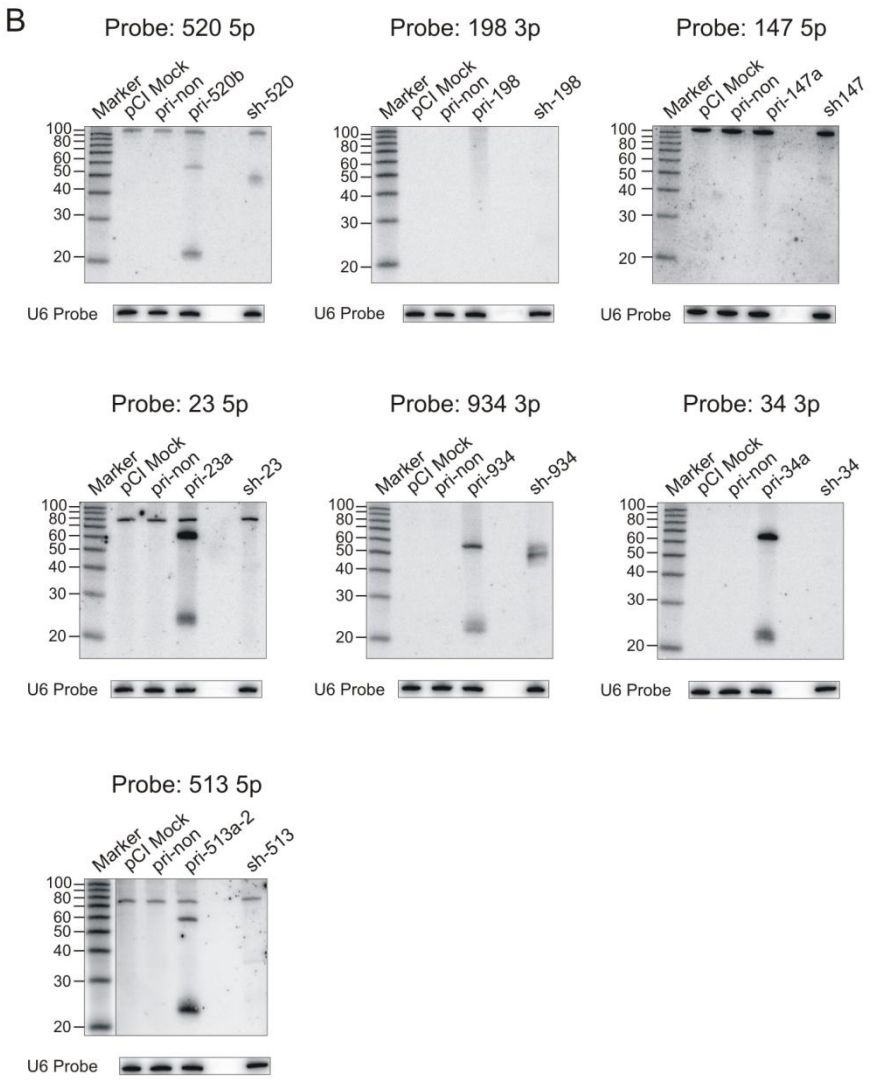


Figure 2.11: Processing of major and minor miRNA guides.

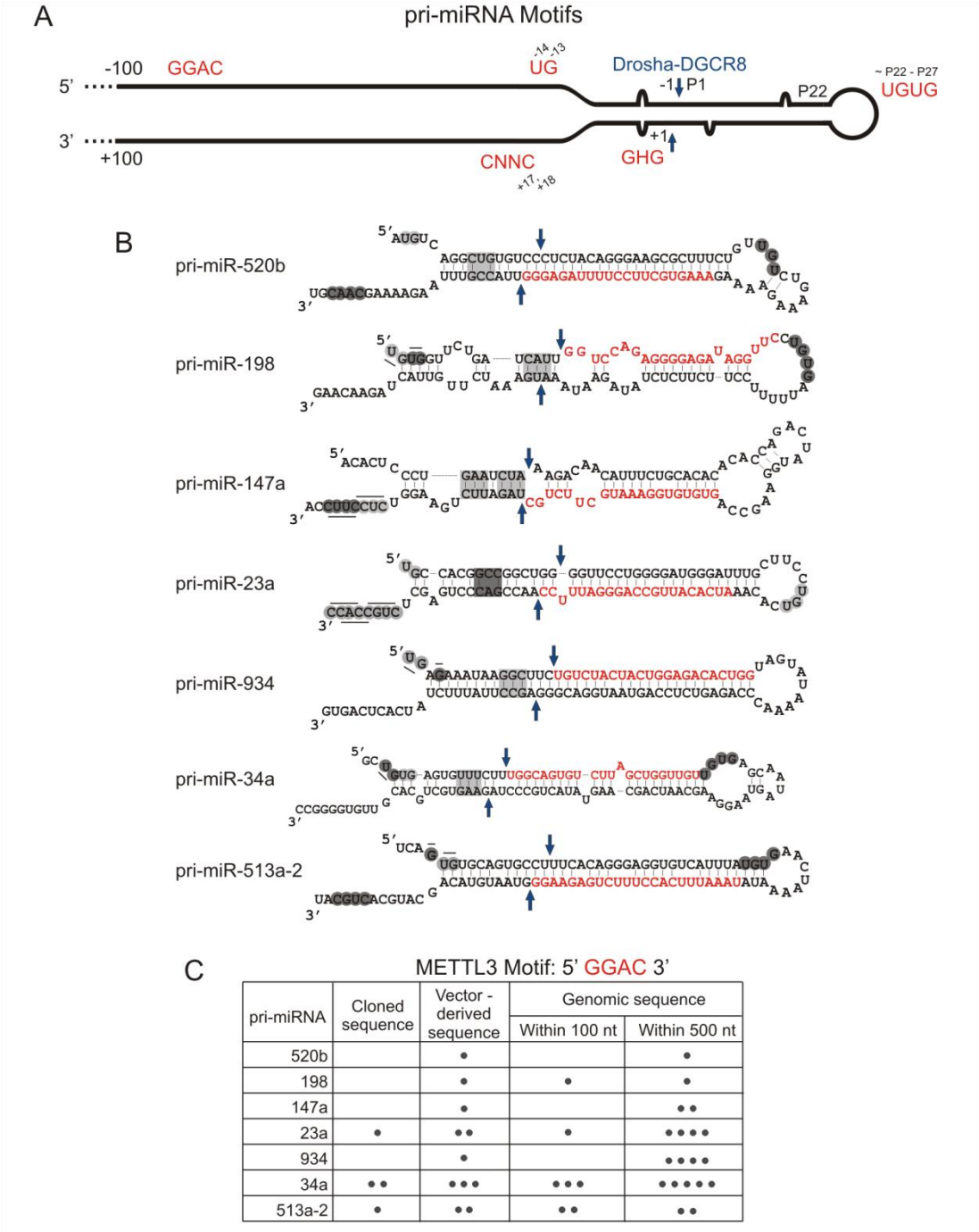
(**A**) Major and (**B**) minor guide sequences were detected through northern blot analyses, following transfection of HEK293 cells with the corresponding pri-miRNA or shRNA (positive control) expression vectors. A mock expression vector (pCI mock) and non-specific pri-miRNA expression vector (pri-non) were included as negative controls. Signal intensities of the major guides were quantified relative to U6 signal intensity and are shown as a ratio of pri-miRNA-derived guide to shRNA-derived guide. A similar quantification was not given for the minor miRNAs as shRNAs were not designed to generate these guides. ND: not-detected.


B
Probe: 520 5p
Marker
pCI Mock
pri-non
pri-520b
sh-520
100
80
60
50
40
30
20
U6 Probe
Probe: 198 3p
Marker
pCI Mock
pri-non
pri-198
sh-198
U6 Probe
Probe: 147 5p
Marker
pCI Mock
pri-non
pri-147a
sh147
U6 Probe
Probe: 23 5p
Marker
pCI Mock
pri-non
pri-23a
sh-23
U6 Probe
Probe: 934 3p
Marker
pCI Mock
pri-non
pri-934
sh-934
U6 Probe
Probe: 34 3p
Marker
pCI Mock
pri-non
pri-34a
sh-34
U6 Probe
Probe: 513 5p
Marker
pCI Mock
pri-non
pri-513a-2
sh-513
U6 Probe

2.3.5 Potential pri-miRNA sequence motifs

Silencing and processing activity was only observed for expressed pri-miRNAs that had a conventional basal stem in predicted structures of the local sequence. However, non-structural determinants may also contribute to the observed pri-miRNA function and recently described pri-miRNA sequence motifs have been shown to rescue Drosha-DGCR8 processing of pri-miRNAs with structural defects or non-optimal conformations (Auyeung et al., 2013; Fang and Bartel, 2015). Four sequence motifs with a direct effect on Drosha-DGCR8 processing have been described (section 1.3.3), namely, the basal 5'-$U^{-14}G^{-13}$-3' motif in the 5p arm preceding (-) the Drosha-DGCR8 cleavage site; the 5'-UGUG-3' terminal loop motif at ~ 24 nt after the Drosha-DGCR8 cleavage site; the mismatched 3'-GHG-5' duplex motif in the 3p arm of the basal stem at positions 7 - 9 from the basal junction; and the basal 5'-CNNC-3' motif in the 3p arm starting at 17 or 18 nt after (+) the Drosha-DGCR8 cleavage site (Auyeung et al., 2013; Fang and Bartel, 2015) (Figure 2.12 A). These motifs all appear to enhance Drosha-DGCR8 processing and are important for orientation and cleavage by Drosha-DGCR8. Specifically, Drosha recognises the 5p UG basal motif and DGCR8 recognises the apical UGU loop motif (Nguyen et al., 2015). In addition, methyltransferase-like 3 (METTL3) recognition motifs occurring within 100 bp of the adjacent flanking sequence also enable the selective methylation of pri-miRNAs by creating m^6A (N^6-methyladenosine) marks that enhance recognition and processing by DGCR8 (Alarcon et al., 2015b).

Pri-miRNA sequences were analysed to determine if the observed pri-miRNA function could be explained by the presence or absence of sequence motifs. Each pri-miRNA was found to have a subset of potential sequence motifs, as is typical for natural pri-miRNAs (Fang and Bartel, 2015), with between two and four potential basal, stem or loop motifs (Figures 2.12 B). Motifs located at precise positions (dark grey) were expected to contribute to processing efficiency, while those located within a few nucleotides of the expected position (light grey) may also play a role, but the tolerance and contribution of imprecisely positioned motifs has not been fully explored at the level of individual pri-miRNAs. The number or type of potential motifs did not necessarily dictate the level of pri-miRNA function. Pri-miR-520b, for example, potentially has all four motifs and mediated target silencing of 79 %, while pri-miR-934 has only a potential basal motif (UG) and a weak GHG motif with a 5p-G / 3p-C base pair at position 7, but mediated target silencing of 91 %. Potential motifs were also identified for the non-functional pri-miR-198 and pri-miR-147a, but the results of characterisation studies suggested that these motifs were not sufficient to enable processing. However, in both cases, bulges in the secondary RNA structure made it difficult to identify stem motifs at the expected position from the basal junction. The lack of well positioned



pri-miRNA	Cloned sequence	Vector - derived sequence	Genomic sequence	
			Within 100 nt	Within 500 nt
520b		•		•
198		•	•	•
147a		•		••
23a	•	••	•	••••
934		•		••••
34a	••	•••	•••	•••••
513a-2	•	••	••	••

Figure 2.12: Potential primary sequence motifs.

(**A**) A typical pri-miRNA showing the sequence and location of previously described primary sequence motifs in relation to the Drosha-DGCR8 site (arrows). Motifs can occur in flanking regions before (-) or after (+) the Drosha-DGCR8 cleavage site, within the basal stem or in the terminal loop of the pre-miRNA (P). Methyltransferase-like 3 (METTL3) recognition motifs may also occur at non-specific locations in the pri-miRNA flanking sequence. (**B**) Potential motifs in the basal flanks, basal stem or loop at precise (dark grey) or slightly shifted (light grey) positions.

Potential mismatched GHG motifs in the 3p basal stem that start with a 5p-C or U/G-3p pair are expected to have a greater effect on processing than those that start with a 5p-G/C-3p. (**C**) Potential METTL3 recognition motifs occurring in the cloned or vector-derived RNA sequence or within 100 or 500 nt on either side of the expected pre-miRNA in the genomic context (Ensembl Release 81).

stem motifs could contribute to the observed lack of function for pri-miR-198 and pri-miR-147a, but the absence of a single motif is not expected to abrogate function as ~ 21 % of human pri-miRNAs lack primary sequence motifs altogether (Auyeung et al., 2013). In the case of pri-miR-520b, the shifted position of a potential GHG motif could also suggest that a more conventional basal stem length is plausible. The motif is positioned at ~ 2 - 4 nt from the Drosha-DGCR8 cleavage site as expected for a conventional hairpin, but the distance between the motif and the basal junction should be ~ 7 bp instead of 4 bp, suggesting that a transient, unpredicted interaction of stem residues may extend pairing of the basal stem. Alternatively, a more likely explanation is that the presence of several motifs could compensate for a sub-optimal basal stem and permit pri-miR-520b processing, as has been described for other pri-miRNAs (Fang and Bartel, 2015). Overall, these results are in agreement with the current model of pri-miRNA processing, where primary sequence motifs were found to contribute to pri-miRNA function in a modular manner specific to individual pri-miRNAs (Auyeung et al., 2013).

Pri-miRNA sequences were also analysed for the presence of the 5'-GGAC-3' METTL3 motif (Alarcon et al., 2015b) in the context of both the expression vector and genomic sequence. The motif status within 100 nt of the pre-miRNA in the genomic context was maintained in the cloned sequence for four pri-miRNAs, excluding pri-miR-198, pri-miR-34a and pri-miR-513a-2 where a single motif was omitted. This was unlikely to have a major impact on the function of pri-miR-34a or pri-miR-513a-2, which still retained one or two original motifs and were processed to give effective target silencing (> 80 %). The absence of a METTL3 motif within a certain distance could contribute to the observed lack of function for pri-miR-198. An additional motif was included at ~ 120 nt from the pre-miRNA in the 3p flank of the vector-derived RNA sequence, but this differed from the genomic context where a single motif was located at ~ 80 nt from the pre-miRNA in the 5p sequence. However, a requirement for such precise positioning of a METTL3 motif has not been described for other miRNAs and the absence of m^6A marks is not expected to completely abolish pri-miRNA processing.

2.3.6 RNase mapping of functional and non-functional pri-miRNA structures *in vitro*

Predicted secondary RNA structures do not necessarily indicate the exact conformation of certain residues and interactions within the terminal loop and flanking sequences can differ from experimentally determined structures (Krol et al., 2004). However, the percentage of correctly predicted base pairs is typically high (88.4 %) (Krol et al., 2004), thus supporting the predicted conformation of paired stem regions. To further investigate secondary RNA structure experimentally, RNase mapping assays were performed for a functional (pri-miR-34a) and non-functional (pri-miR-198) pri-miRNA. Pri-miRNA sequences were transcribed *in vitro* from the T7 RNA polymerase promoter and partial RNAse digests were used to interrogate the secondary RNA structure under denaturing and non-denaturing conditions. Digest patterns were analysed on an RNA sequencing gel to identify unconstrained C, U (RNase A) and G (RNase T1) residues, which were mapped on to the predicted secondary RNA structure. *In vitro* transcribed pri-miRNA sequences were shorter than those expected from the pol II expression vector and generated fewer predicted secondary RNA structures for the full sequence (Table 2.9) (Appendix A). However, these structures contained commonly predicted hairpin conformations and all predicted structures generated with 15 - 50 nt of flanking sequence were identical.

Table 2.9: The basal stem length (BSL) and free energy (dG) of predicted secondary RNA structures of expected transcripts.

Input Sequence	Pri-miR-198 (CMV)		Pri-miR-198 (T7)		Pri-miR-34a (CMV)		Pri-miR-34a (T7)	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
Full pri-miRNA	NH	-188.9	3	-40.7•	11	-245.8•	11	-84.4
	NH	-185.1	NH	-37.3	11	-240.5	22	-83.4
	NH	-187.4			11	-245.1	22	-83.0
	3	-200.2•			18	-240.3	22	-84.0
	3	-196.7			11	-232.4	11	-79.1
	NH	-198.6			11	-234.8	15	-87.2•
	3	-192.5			11	-240.9	19	-80.5
	NH	-193.9			11	-245.6		
	3	-183.5			11	-233.4		
	NH	-174.2			11	-242.4		
	NH	-172.9			11	-238.6		
	NH	-188.2			11	-241.8		
	NH	-168.8			18	-236.7		
	NH	-185.0			11	-238.3		
	3	-186.8			4	-232.2		
	3	-195.3			22	-244.7		
	3	-186.5			NH	-232.2		
	3	-189.3			22	-236.9		
	3	-185.1			NH	-239.5		
	NH	-188.9			NH	-229.1		
	3	-193.1						
	NH	-179.1						
	3	-187.0						
	NH	-184.4						

The input sequence used for mfold structure prediction was the full pri-miRNA sequence of the expected transcript derived from a CMV or T7 expression cassette. The table format follows that of Table 2.8. The BSL was defined as the duplex region between the anticipated Drosha-DGCR8 cleavage site and the nearest disruptive element (a fork or any bulge of 3 nt or more). The contribution of G-U wobble pairs and non-canonical pairs was also included where appropriate. A BSL entry with a solid underline indicates that the secondary RNA structure adopted a characteristic hairpin form, including a conventional basal stem (8 - 13 bp) flanked by at least 4 nt of consecutive ssRNA on one side. The most stable predicted secondary RNA structure in each data set is indicated (•). NH: No Hairpin.

Structurally unconstrained residues were identified in the flanking sequences, large internal bulges and terminal loop of pri-miR-198 as expected (Figure 2.13). However, flexibility was also detected in regions of both the upper and lower stem predicted to contain paired residues. Slightly unconstrained residues were detected in paired regions corresponding to the miRNA guide, including U^{58}, C^{59}-G^{103} and the paired region between G^{64} and G^{73}. The latter is not supported in any predicted hairpin or non-hairpin conformation. Although flexibility at the 5' end of the miRNA (G^{56}) is in agreement with previously described pri-miRNA features (Han et al., 2006; Khvorova et al., 2003). More unconstrained residues were detected in the basal stem, suggesting that pairing in this region might not be supported *in vitro*. Flexibility at U^{51} and C^{52}-G^{111} does not support the formation of a 3 bp basal stem; additional flexibility at G^{49} does not support a 5 bp basal stem; and additional flexibility of both canonical and G - U wobble pairs between G^{41} and G^{44} does not support a 10 bp basal stem. In general, the pri-miR-198 structure appears to be more flexible than expected *in vitro* and this might contribute to the observed lack of function in the artificial expression context.

Structurally unconstrained residues were also identified in the flanking sequences and terminal loop of pri-miR34a, but flexibility within the stem was largely limited to asymmetrical bulges (U^{83}; G^{124}, U^{125}) and the end of the basal stem (Figure 2.14). Although cleavage at C^{75} indicated that this residue might also be accessible *in vitro*. In contrast to pri-miR-198, mapping results supported the formation of a canonical hairpin with a basal stem of at least 6 bp, while pairing of G^{59}-C^{168} and G^{61} supports the formation of a 9 bp basal stem. Flexibility between C^{54} and U^{58}, as well as G^{149} and G^{152}, does not support the formation of a 13 bp basal stem, although this region may be a flexible part of a longer 17 bp stem conformation, supported by pairing between C^{49} and G^{53}. Further extension of the basal stem to 20 bp seems unlikely, as residues U^{45} - G^{47} were unconstrained. In general, the predicted structures of pri-miR34a appeared to be a fair reflection of the hairpin *in vitro*, thus supporting a processive conformation and the observed function of pri-miR-34a.

(Figure 2.13 overleaf)

Figure 2.13: RNase structure map of pri-miR-198.

(**A**) RNA sequencing gels showing partial RNAse digest patterns of *in vitro*-transcribed, 5'-radiolabeled pri-miR-198. OH^-: alkaline hydrolysis ladder; A: RNase A digest; T1: RNase T1 digest; (N) non-denaturing conditions. (**B**) Corresponding structure maps of pri-miR-198. Black arrows indicate the 3' cleavage sites of Y (C or U) or G residues in single stranded regions that are recognised by RNase A or T1, respectively. Green arrows indicate the beginning and end of the mapped region.

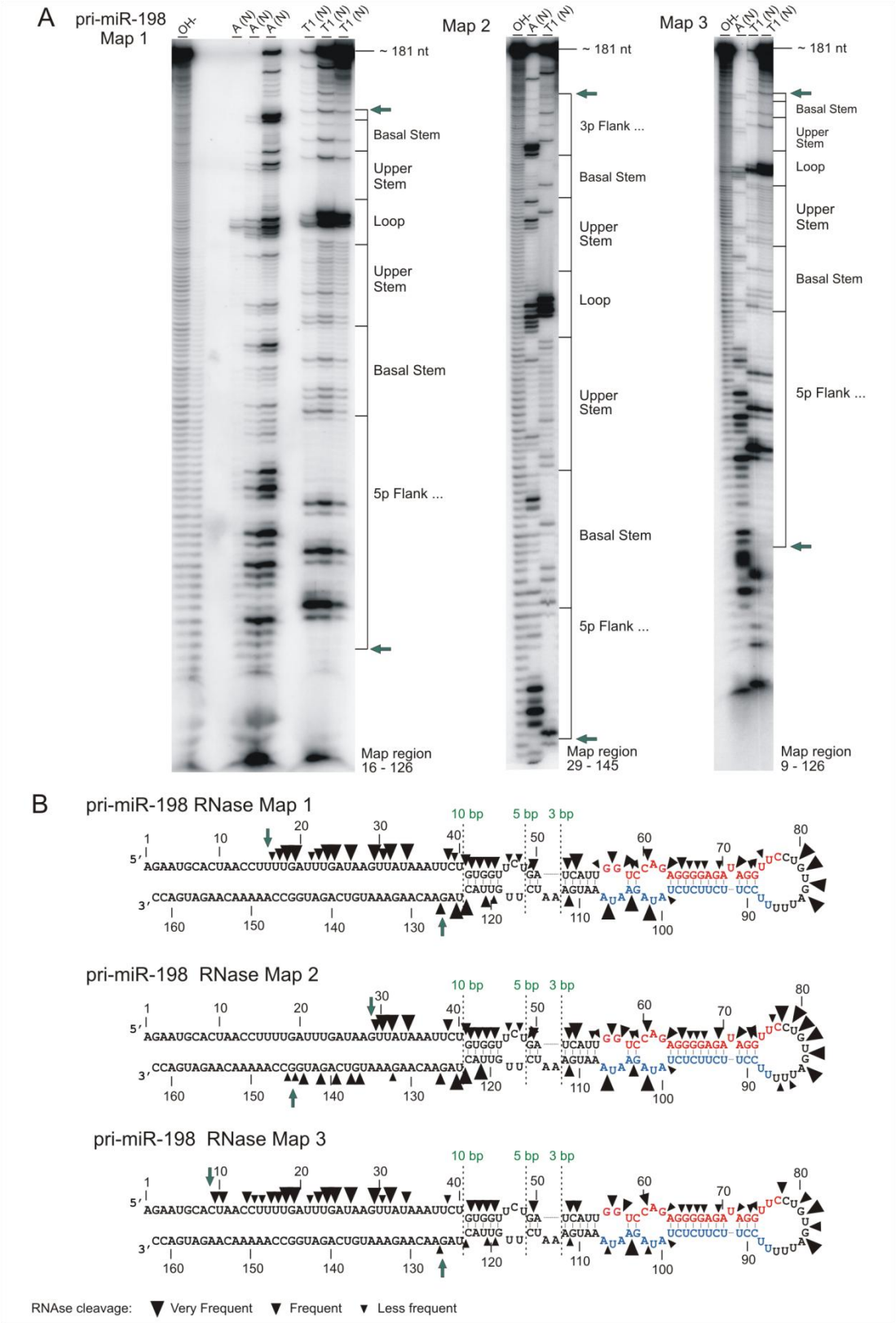


Figure 2.13: RNase structure map of pri-miR-198.

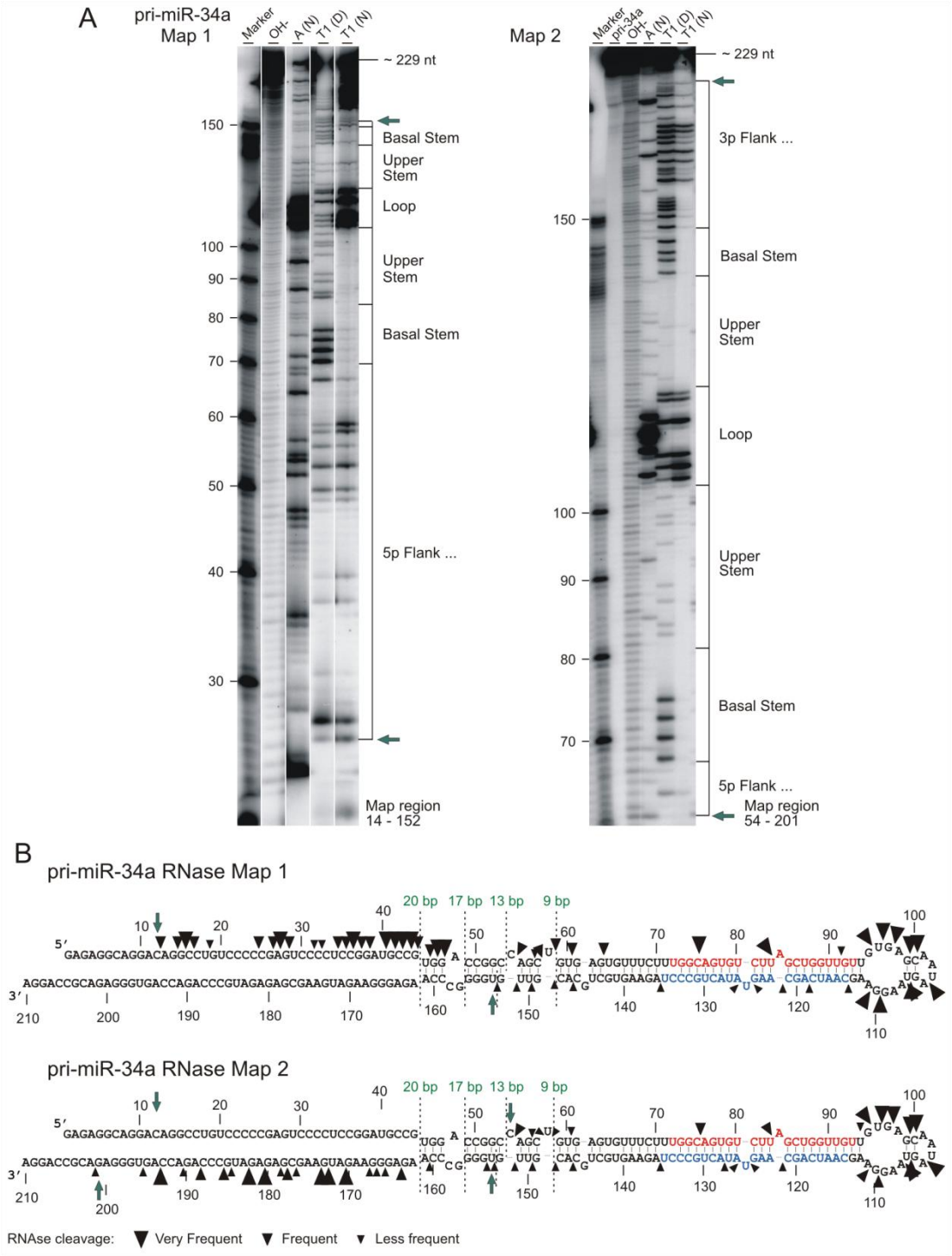


Figure 2.14: RNase structure map of pri-miR-34a.

(**A**) RNA sequencing gels showing partial RNAse digest patterns of *in vitro*-transcribed, 5'-radiolabeled pri-miR-34a. OH^-: alkaline hydrolysis ladder; A: RNase A digest; T1: RNase T1 digest; (N) non-denaturing conditions; (D) denaturing conditions. (**B**) Corresponding structure maps of pri-miR-34a. Black arrows: 3' cleavage sites, green arrows: mapped region.

2.4 DISCUSSION

The processing and silencing efficacy of naturally occurring pri-miRNAs with various basal stem conformations was examined. Variation in the length of a canonical basal stem, predicted to occur in secondary RNA structure of the local pri-miRNA sequence, did not appear to influence general pri-miRNA function. Functional miRNAs were effectively derived from pri-miRNAs with basal stems of 9 -13 bp, implying that minor variations in the basal stem length can be tolerated by the Drosha molecular "ruler". However, no function was observed for pri-miRNAs where the basal stem was non-canonical and contained a short duplex and large, internal bulges. This suggested that exceptional features of the basal stem might be detrimental to Drosha processing. Variation in the length and conformation of the basal stem was also expected to occur in predicted structures of the expressed pri-miRNA sequences generated in the artificial expression context. However, disruptive interactions predicted to occur between the basal stem and extended flanking sequences had no appreciable effect on general pri-miRNA function. Comparable function (≥ 80 %) was observed for pri-miRNAs where the canonical hairpin was preserved in most (pri-miR-23a, 934), a few (pri-miR-34a, 513a-2) or none (pri-miR-520b) of the predicted structures including the extended or expressed sequence. A subset of potential pri-miRNA sequence motifs was observed for most functional pri-miRNAs, as is typical for natural pri-miRNAs, and effective function was not dictated by the presence of a particular motif. Together these results suggested that canonical folding of the local pri-miRNA sequence was more indicative of the processive hairpin conformation and was sufficient to permit conventional Drosha-DGCR8 processing, while various sequence motif combinations were expected to enhance processing of individual pri-miRNAs in a modular manner.

The comparable and effective processing of pri-miRNAs with a range of basal stem lengths of 9 - 13 bp was initially surprising as the optimal basal stem length has previously been described as ~ 11 bp (Auyeung et al., 2013) and very short (3 bp) or long (18 bp) basal stem lengths are deleterious to pri-miRNA function (Zeng et al., 2005b). However, a narrow preference for a total stem length of 35 ±1 bp has recently been identified in a subsequent study by the Bartel group (Fang and Bartel, 2015). Pri-miRNAs, including pri-miR-23a, 934, 34a and 513a-2, may therefore have a sub-optimal basal stem length, but an optimal total stem length of 35 or 36 bp which permits effective Drosha-DGCR8 processing and can explain the pri-miRNA function observed here. This finding may also explain the lack of function observed for pri-miR-198 and pri-miR-147a, which both have non-optimal basal (5 or 7 bp) and total (25 and 26 bp) stem lengths. Pri-miR-520b, however, also has a sub-optimal basal stem length of ~ 9 bp and a sub-optimal total stem length of ~ 32 bp, but functions effectively. A shorter stem length of ~ 30 bp

was found to severely reduce productive Drosha-DGCR8 cleavage and the generation of mature miRNAs and this effect was not remedied by the presence of all four pri-miRNA sequence motifs (Fang and Bartel, 2015). The effective function and processing of pri-miR-520b therefore remains surprising, but also suggests that processing of a pri-miRNA with a total stem length of ~ 32 bp may be enabled in the presence of all four pri-miRNA sequence motifs. The potential GHG stem motif of pri-miR-520b is not positioned correctly with respect to the basal junction as the result of a shorter basal stem, but is at an expected distance of 2- 4 nt from the Drosha-DGCR8 cleavage site, suggesting that this measurement may be more appropriate to describe the position of the GHG motif for pri-miRNAs with exceptional basal stems. The processing of pri-miRNAs with various basal stem conformations can therefore presumably be explained by an optimal total stem length or the compensatory role of sequence motifs.

Pri-miRNAs with various non-canonical basal stem conformations might also be expected to show abortive or non-productive processing. The flanking segments of pri-miR-30a, like those of pri-miR-34a, were predicted to form additional duplex pairing beyond the conventional basal stem (~ 11 bp). The absence of a clear basal junction was shown to support unproductive processing of pri-miR-30a (Nguyen et al., 2015), where Drosha cleavage occurred at ~ 11 bp from the terminal loop, thus cleaving the miRNA. Similar processing might also occur for pri-miR-34a, pri-miR-513a-2 and pri-miR-520b where the basal junction was distorted in certain predicted structures of the expressed pri-miRNA sequence. The sub-optimal total stem length of pri-miR-520b is also expected to promote non-productive processing, as has been observed for a pri-miRNA with a shorter stem (Fang and Bartel, 2015). However, functional miRNAs were effectively derived from productive processing of these pri-miRNAs, as seen in silencing assay and northern blot results, so any non-productive processing is likely to be minor. It would be interesting to investigate the possibility of unproductive processing further, but productive processing was prominent in pri-miRNA characterisation and supported the subsequent generation of mimics.

Functional miRNAs were not processed from the pri-miR-198 and pri-miR-147a sequences, which may be explained by non-optimal basal and total stem lengths. However, this result was still surprising as experimental evidence supports the role of Drosha-DGCR8 and Dicer in miR-198 processing. SiRNA-mediated knockdown of Drosha causes a simultaneous decrease in the abundance of miR-198 and an increase the abundance of the mRNA (FSTL1 transcript) from which this exonic miRNA is derived in a “see-saw” manner; while *in vitro* processing of the miR-198 stem-loop using a HEK293T cytoplasmic extract supports Dicer function (Sundaram et al., 2013). In addition, miR-198 and miR-147a guides appear to enable RNAi-mediated silencing

when derived from Dicer-dependent shRNAs (sh-198, sh-147) or synthetic miRNA duplexes (Bertero et al., 2012; Lee et al., 2014; Sundaram et al., 2013; Sung and Rice, 2009; Tan et al., 2011; Uhlmann et al., 2012; Ye et al., 2013). MiR-198 has also been found to be part of a subset of miRNAs where binding of a KH-type splicing regulatory protein (KSRP) to a G-containing terminal loop motif facilitates processing by interacting with both Drosha and Dicer complexes (Sundaram et al., 2013; Trabucchi et al., 2009). KSRP-mediated processing has not been described for miR-147a, but the G doublet or GAAG motif in the terminal loop of pri-miR-147a could potentially allow for KSRP-binding (Trabucchi et al., 2009). KSRP-mediated processing has been demonstrated in HEK293T and HeLa cell lines (Trabucchi et al., 2009) and it was therefore expected that pri-miR-198 processing was supported in cell culture assays; but KSRP-mediated processing may also be unexpectedly downregulated by miR-181a in response to signalling by transforming growth factor-β1 (TGF-β1) (Sundaram et al., 2013).

In addition, other disruptive elements of the secondary RNA structure may also contribute to the lack of function of pri-miR-198 and pri-miR-147a. Large internal bulges have been identified as as detrimental to conventional processing (Fang and Bartel, 2015; Lee et al., 2003b; Zeng and Cullen, 2003). In a recent study, large asymmetrical (2-0) and symmetrical (3-3) bulges within the stem were described as structural defects and found to dramatically reduce Drosha-DGCR8 processing and abolish the generation of mature miRNA guides (Fang and Bartel, 2015). The presence of all four basal, stem and loop motifs could rescue Drosha-DGCR8 processing for both types of defects, but did not rescue miRNA generation for a pri-miRNA with a large 3-3 bulge adjacent to the Dicer cleavage site, implying that such bulges can inhibit Drosha and / or Dicer RNAse III processing. The presence of several similar internal 3 nt bulges, along with a non-optimal stem length, is therefore expected to reduce or even abolish conventional processing of guides from pri-miR-198 and pri-miR-147a, in spite of potential basal, stem and loop motifs. The mechanism of processing for pri-miRNAs with such disruptive internal bulges is currently unknown, but seems likely to involve additional processing, signalling or regulatory factors. Further to this, miR-198 and 147a may even be part of a functional miRNA subset and reliant on common signalling pathways or additional processing factors, as both miRNAs play a role in inhibiting cell migration. MiR-198 regulates several proteins involved in cell migration, including serine proteases responsible for degradation of the extracellular matrix and a protein required for actin polymerisation (Sundaram et al., 2013); and miR-147a regulates genes involved in cell proliferation and migration, including cyclin proteins and cyclin-dependent kinases responsible for cell cycle progression (Bertero et al., 2012; Lee et al., 2014; Uhlmann et al., 2012). Alternatively, non-conventional cleavage mechanisms may be involved, particularly for pri-miR-147a where experimental evidence to support conventional processing is scant.

Drosha-independent, splicing-dependent processing has been described for intronic mammalian miRNAs (mirtrons), such as miR-877 and miR-1226 (Berezikov et al., 2007; Schamberger et al., 2012; Sibley et al., 2012b); while Drosha-dependent, splicing-independent processing has been shown for mirtron-like miRNAs, such as miR-1225 and miR-1228 (Havens et al., 2012). Dicer-independent processing has also been described for miR-451, which relies on the catalytic activity of Ago2 for cleavage (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010a). Another alternative explanation for the lack of miR-198 function here, may be that the miR-198 hairpin functions primarily as a mRNA destabilising element with only a minimal role as a miRNA precursor (Ha and Kim, 2014). It would be interesting to further investigate the various possibilities behind the lack of pri-miR-198 and pri-miR-147a function, but regardless of the explanation, these sequences were unattractive for use as scaffolds in subsequent pri-miRNA mimic design.

In conclusion, pri-miRNAs with basal stems in the range of 9 - 13 bp were processed effectively to mediate gene silencing, even though a basal stem length of ~ 11 bp has previously been described as optimal for processing of pri-miR-30a, 125a, 16-1 and 223 (Auyeung et al., 2013). The Drosha molecular “ruler” therefore appears to tolerate minor variations in the basal stem length, while Drosha-DGCR8 processing of pri-miRNAs with sub-optimal basal stem lengths was presumably enabled by an optimal total stem length of 35 ±1 bp or the presence of all four pri-miRNA sequence motifs in the instance of a slightly shorter stem (~ 32 bp) (Fang and Bartel, 2015). The lack of function for expressed pri-miR-198 and pri-miR-147a sequences was likely a result of non-optimal stem lengths and large internal bulges, both of which have recently been described as detrimental to conventional Drosha-DGCR8 processing (Fang and Bartel, 2015). However, as evidence supports the function of both miR-198 and miR-147a, the lack of processing of the respective pri-miRNAs here may also be the absence of additional processing or regulatory signals and factors in the artificial expression context. Canonical folding of the local pri-miRNA sequence with 15 - 20 nt of flanking sequence was found to be a reliable indicator of functional pri-miRNAs and should be a priority in the identification of functional scaffolds. This study has contributed to our current understanding of the effects of non-canonical structural features in the processing of expressed pri-miRNAs. Furthermore, these results also suggest that pri-miRNAs with various basal stem conformations and motif combinations can be used for the generation of exogenous RNA guides. This concept was investigated further in Chapter 3, where characterised pri-miRNAs were used as novel scaffolds in the design of corresponding pri-miRNA mimics.

CHAPTER THREE

Therapeutic RNA guides derived from pri-miRNA mimics with exceptional basal stem conformations

3.1 INTRODUCTION

Expressed primary microRNA (pri-miRNA) mimics have been shown to mediate effective gene silencing and are an important component of therapeutic RNAi constructs (Chung et al., 2014; Chung et al., 2012). Pri-miRNA mimics are well suited for therapeutic RNAi applications and have several advantages over conventional short hairpin RNAs (shRNAs). Mimics based on the pri-miR-30a scaffold were found to mediate maximal target silencing without the side effects associated with more potent shRNAs in an equivalent expression format, including disruption of miRNA biogenesis *in vitro* and cellular toxicity *in vivo* (Boudreau et al., 2009). Exogenous RNA guides are typically expressed at a lower level when derived from pri-miRNA mimics, but are processed more efficiently in the miRNA biogenesis pathway with two cleavage events, thus avoiding the accumulation of unprocessed precursors (Boudreau et al., 2009). The saturation of pathway components is an important consideration *in vivo*, where competition with highly expressed shRNAs can lead to the downregulation of endogenous miRNAs, resulting in organ failure and death (Grimm et al., 2006). In addition, while highly expressed therapeutic RNA activators are more effective in short-term assays, they can be disadvantageous in long-term studies of stably transduced cells as a result of cellular toxicity (Chung et al., 2014).

Pri-miRNA mimics are also well suited for combinatorial RNAi approaches, as several mimics can be generated from a single polycistronic miRNA cluster to silence multiple therapeutic targets. A multi-targeting approach is required in the treatment of a rapidly evolving pathogen like HIV, where at least four viral sequences should be targeted simultaneously to successfully inhibit viral replication and prevent viral escape (Applegate et al., 2010; Leonard and Schaffer, 2005); and near-complete coverage of all HIV-1 strains can be achieved by targeting seven conserved viral sequences (McIntyre et al., 2011). Several polycistronic constructs based on naturally-occurring units, like the miR-17~92 (Liu et al., 2008) and miR-106b clusters (Aagaard et al., 2008; Chung et al., 2014), or artificial polycistronic units (Zhang et al., 2012), have been used to successfully inhibit viral replication and suppress viral escape. The use of a single inducible pol II promoter allows for controlled, tissue-specific expression of exogenous RNA guides (Lo et al., 2007; Stegmeier et al., 2005; Zhou et al., 2005) and, in the context of a

clinically relevant lentiviral vector delivery system, can also provide equimolar mimic expression without the re-arrangement and deletion issues observed for repeated expression cassettes (ter Brake et al., 2008).

Polycistronic pri-miRNA mimics have shown much promise in combinatorial RNAi against HIV (Chung et al., 2014), but the optimisation of individual mimics within a polycistronic construct is often required. Ideally, each individual mimic should function effectively, but this is not necessarily the case. In the full miR-17~92 construct (Liu et al., 2008), not all mimics were functional, while processing of mimics from the initial miR-106b cluster was not equivalent (Aagaard et al., 2008) and the efficacy of artificially-linked mimics was affected by the order of mimics and the identity of the flanking sequences (Zhang et al., 2012). These issues have been addressed by omitting or rearranging ineffective mimics (Liu et al., 2008; Zhang et al., 2012) or by improving the function of individual mimics through the inclusion of original guide region mismatches (Aagaard et al., 2008) or flanking sequences (Zhang et al., 2012). However, the generation of improved polycistronic constructs against HIV would be advanced by an expanded repertoire of individual mimics with highly effective function and more comprehensive rules for the design of individual mimics. The use of different, but effective pri-miRNA sequences is also required in the context of a lentiviral vector delivery system to avoid homologous recombination of repeat sequences (ter Brake et al., 2008). Mimic design has typically relied on a few, well characterised, conventional precursors like that of miR-30, but the potential of hundreds of other pri-miRNAs has not been fully explored. In particular, the potential of pri-miRNAs with exceptional or non-canonical features has not yet been described in mimic design.

The aims of this chapter were to design and characterise a panel of anti-HIV pri-miRNA mimics based on effective and naturally-occurring pri-miRNAs with various basal stem conformations that were characterised in Chapter 2. The RNAi-mediated silencing and processing of exogenous RNA guides was assessed in cell culture. Common properties of the exogenous RNA guide and the pri-miRNA guide region were identified for effective and less effective pri-miRNA mimics, highlighting the importance of the guide region in agreement with previous studies. This work enabled a more comprehensive, updated set of rules to be proposed for the design of effective pri-miRNA mimics based on natural scaffolds. In addition, several novel pri-miRNA scaffolds and effective pri-miRNA mimics were characterised here and are suitable for use in therapeutic constructs and future gene silencing applications.

3.2 MATERIALS & METHODS

3.2.1 Generation of anti-HIV pri-miRNA mimic expression plasmids

Design

Pri-miRNA mimics were created by replacing the guide sequence of each original pri-miRNA with an anti-HIV guide sequence. Four highly effective anti-HIV guide sequences designed to target expression from respective regions of the HIV-1 genome were selected from previous studies: anti-*tat* / *rev* (Lee et al., 2002a), anti-*int* (Nishitsuji et al., 2006), anti-*gag* (McIntyre et al., 2009a) and anti-LTR (McIntyre et al., 2009a) (Table 3.1). The predicted secondary structure of mimics was kept as close as possible to the original pri-miRNA, in both form and overall stability, using the mfold web server (version 3.5) (Zuker, 2003) with default parameters (37 °C, 5 % sub-optimality). The complementary arm of the stem region was altered to maintain canonical base pairs, G:U wobble pairs and both symmetrical and asymmetrical mismatches, while no alterations were made to the loop or flanking sequences. To examine and compare the range of possible foldings for each pri-miRNA mimic, predicted structures were again generated with standard extensions (15 - 50 nt) of the sequence flanking the anticipated Drosha-DGCR8 processing site (Appendix A).

Table 3.1: Anti-HIV guide sequences

Anti-HIV guide	Sequence (5' to 3' orientation)
anti-*tat* / *rev* (Lee et al., 2002a)	CUCUUCGUCGCUGUCUCCGC(UU)
anti-*int* (Saayman et al., 2010)	ACUAGCCAUUGCUCUCCGGC(UU)
anti-*gag* (McIntyre et al., 2009a) (gag 532-21)	UAAAUCUUGUGGGGUGGCUCC(UU)
anti-LTR (McIntyre et al., 2009a) (LTR 510-21)	UAUUGAGGCUUAAGCAGUGGG(UU)

Construction

Pri-miRNA mimic sequences and expression vectors were generated using the same materials and methods as described for the generation of endogenous pri-miRNAs (See 2.2.2). Briefly, a two-step PCR approach (Castanotto et al., 2002) with extension primers was used to generate all mimics sequences. Extension primers used in the first round of PCR (F1, R1) (Table 3.2) were unique to each mimic, while common extension primers were used in the second round of PCR (F2, R2) (Table 2.2). After initial TA cloning into the pTZ57R/T vector (InsTAclone™ PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA, USA), pri-miRNA mimic sequences were verified by DNA sequencing with the primers listed in Table 2.3 and inserted into the

Table 3.2: Extension primers for the generation of pri-miRNA mimics by a two-step PCR

Primer	Tm (°C)	Sequence (5' to 3' orientation)
Pri-miR-520-tat F1	72.4	CAAGAAGATGTCAGGCTGTGTCGTGGAGCCGGCGGCGGAGAGCTGTTGTCTGAAAGAAAAG
Pri-miR-520-tat R1	69.1	GTTGCTTTTCTTAAACGGTAACGCGGAGACAGCGACGAAGAGCTTTTCTTTCAGACAACAG
Pri-miR-198-tat F1	67.9	GTTATAAATTCTGTGGTTCTGATCATTCTCTTCGTCGCTGTCTCCGCTCCTGTGATTTTTCG
Pri-miR-198-tat R1	66.8	CTTGTTCTAGTAACAAGATTTCATTTTACTGTTTCACCGTCACCGAAAAATCACAGGAGCG
Pri-miR-23-tat F1	78.6	CCTCACCCCTGTGCCACGGCCGGCTGGGGGGGCAGCAGCGGTGGGTTGCTTCCTGTCACAAC
Pri-miR-23-tat R1	74.2	CGGTGGCAGAGCTCAGGGTCGGTTGGCGGAGACAGCGACGAAGAGTTGTGACAGGAAGCAAC
Pri-miR-934-tat F1	69.8	CTGAGAACGAACTTGAGAAATAAGGCTTCCTCTTCGTCGCTGTCTCCGCGGTAGTATAAAACCCG
Pri-miR-934-tat R1	69.7	GAAAACACTGAGTGATAGAAATAAGGCTCCCCCTCCTTCGCTGTCTCTGCGGGTTTTATACTACCG
Pri-miR-34-tat F1	72.6	CCGTGGACCGGCCAGCTGTGAGTGTTTCTTCTCTTCGTCGCTGTCTCCGCGTTGTGAGCAATAGTAAGGAAG
Pri-miR-34-tat R1	73.5	TCCCTCTTGGGCCCCACAACGTGCAGCACTTCTAATCTTCGCCAGCTTCTCCGCGCTTCCTTACTATTGCTCAC
Pri-miR-513-tat F1	70.3	CTTGGATGCCACATTCAGCCATTCAGTGTGCAGTGCCTTGCAGAGGTAGTGAATAGGAGTGTGAACTAAAATACTC
Pri-miR-513-tat R1	71.0	CAGGGACACCACATATGCAGTGCATGCTGTACATTACCCTGCGGAGACAGCGACGAAGAGTATTTTAGTTCACACTC
Pri-miR-520-int F1	67.3	CAAGAAGATGTCAGGCTGTGTAGCCGGATAGTAATGGTTAGTCTGTTGTCTGAAAGAAAAG
Pri-miR-520-int R1	68.1	GTTGCTTTTCTTAAACGGTAAAGCCGGAGAGCAATGGCTAGTCTTTTCTTTCAGACAACAG
Pri-miR-198-int F1	67.4	GTTATAAATTCTGTGGTTCTGATCATTACTAGCCATTGCTCTCCGGCTCCTGTGATTTTACC
Pri-miR-198-int R1	65.2	CTTGTTCTAGTAACAAGATTTCATTTCTTATTTATCGCTCTTCGGTAAAATCACAGGAGCC
Pri-miR-23-int F1	76.5	CCTCACCCCTGTGCCACGGCCGGCTGGTGGGGAGCACTGGTGGGTTTGCTTCCTGTCACAAAC
Pri-miR-23-int R1	73.8	CGGTGGCAGAGCTCAGGGTCGGTTGGCCGGAGAGCAATGGCTAGTTTGTGACAGGAAGCAAAC
Pri-miR-934-int F1	68.5	CTGAGAACGAACTTGAGAAATAAGGCTTCACTAGCCATTGCTCTCCGGCTGTAGTATAAAACCAG
Pri-miR-934-int R1	67.1	GAAAACACTGAGTGATAGAAATAAGGCTCCACTAACTATTGCTCTCCCGCTGGTTTTATACTACAG
Pri-miR-34-int F1	72.0	CCGTGGACCGGCCAGCTGTGAGTGTTTCTTACTAGCCATTGCTCTCCGGCTTTGTGAGCAATAGTAAGGAAG
Pri-miR-34-int R1	72.8	TCCCTCTTGGGCCCCACAACGTGCAGCACTTCTAGCTAGCCATATGCCTCCAGCTCTTCCTTACTATTGCTCAC
Pri-miR-513-int F1	70.6	CTTGGATGCCACATTCAGCCATTCAGTGTGCAGTGCCTTGCGGGAAGGCAATAACTAGTTGTGAACTAAAATAAC
Pri-miR-513-int R1	70.8	CAGGGACACCACATATGCAGTGCATGCTGTACATTACCCTGCCGGAGAGCAATGGCTAGTTATTTTAGTTCACAAC
Pri-miR-520-gag F1	68.7	CAAGAAGATGTCAGGCTGTGTGGAGCCAACTCACAAGGTTTACTGTTGTCTGAAAGAAAA
Pri-miR-520-gag R1	67.7	GTTGCTTTTCTTAAACGGTAAGGAGCCACCCCACAAGATTTACTTTTCTTTCAGACAACA
Pri-miR-147-gag F1	71.2	GCAAACACTCCCTGAATCAGCGCCCAACCCACAGAATTTAACACCAGACTATGGAAGCCA
Pri-miR-147-gag R1	70.4	GGAAGGAGACCTTCAGAATCAGGAGCCACCCCACAAGATTTATGGCTTCCATAGTCTGGT

Pri-miR-23-gag F1	75.3	CCTCACCCCTGTGCCACGGCCGGCAGGGTCGCTCCAGGAGAGTTATTGCTTCCTGTCAC
Pri-miR-23-gag R1	72.0	CGGTGGCAGAGCTCAGGGTCGGTAGGAGCCACCCCACAAGATTTATTGTGACAGGAAGCAATAAC
Pri-miR-934-gag F1	68.5	CTGAGAACGAACTTGAGAAATAAGGCTTCTAAATCTTGTGGGGTGGCTCCGTAGTATAAAACCGG
Pri-miR-934-gag R1	69.6	GAAAACACTGAGTGATAGAAATAAGGCTCCCAAACCGTGTGGGGTGGTTCCGGTTTTATACTACGG
Pri-miR-34-gag F1	71.8	CCGTGGACCGGCCAGCTGTGAGTGTTTCTTTAAATCTTGTGGGGTGGCTCCTTGTGAGCAATAGTAAGGAAG
Pri-miR-34-gag R1	73.3	TCCCTCTTGGGCCCCACAACGTGCAGCACTTCTAGAAATCTCGATGGTGTGGCCCCCTTCCTTACTATTGCTCAC
Pri-miR-513-gag F1	69.5	CTTGGATGCCACATTCAGCCATTCAGTGTGCAGTGCCGGGACCCATACCACATTATTTATGTGAACTAAAATATAAATC
Pri-miR-513-gag R1	69.9	CAGGGACACCACATATGCAGTGCATGCTGTACATTACCAGGAGCCACCCCACAAGATTTATATTTTAGTTCAC
Pri-miR-520-LTR F1	67.8	CAAGAAGATGTCAGGCTGTGTCCCACTGATTGAGCCTTAATACTGTTGTCTGAAAGAAAA
Pri-miR-520-LTR R1	67.1	GTTGCTTTTCTTAAACGGTAACCCACTGCTTAAGCCTCAATACTTTTCTTTCAGACAACA
Pri-miR-147-LTR F1	68.2	GCAAACACTCCCTGAATCACAACTAGATTAAGCTCCAATAACACCAGACTATGGAAGCCA
Pri-miR-147-LTR R1	69.8	GGAAGGAGACCTTCAGAATCACCCACTGCTTAAGCCTCAATATGGCTTCCATAGTCTGGT
Pri-miR-23-LTR F1	72.5	CCTCACCCCTGTGCCACGGCCGGCACCTATGCTTAACCCTCCATATTGCTTCCTGTCACAATATTG
Pri-miR-23-LTR R1	72.1	CGGTGGCAGAGCTCAGGGTCGGTACCCACTGCTTAAGCCTCAATATTGTGACAGGAAGC
Pri-miR-934-LTR F1	67.0	CTGAGAACGAACTTGAGAAATAAGGCTTCTATTGAGGCTTAAGCAGTGTGGTAGTATAAAACCC
Pri-miR-934-LTR R1	66.9	GAAAACACTGAGTGATAGAAATAAGGCTCCCATTAATGCTTAAGCAGAGTGGGTTTTATACTACCAC
Pri-miR-34-LTR F1	71.3	CCGTGGACCGGCCAGCTGTGAGTGTTTCTTTATTGAGGCTTAAGCAGTGGGTTGTGAGCAATAGTAAGGAAG
Pri-miR-34-LTR R1	71.4	TCCCTCTTGGGCCCCACAACGTGCAGCACTTCTAGATTGAGACATTATGCAGTAGGCTTCCTTACTATTGCTCAC
Pri-miR-513-LTR F1	70.8	CTTGGATGCCACATTCAGCCATTCAGTGTGCAGTGCCGCCCTCTGTCTAAGCGCCAATATGTGAACTAAAATATATTG
Pri-miR-513-LTR R1	69.4	CAGGGACACCACATATGCAGTGCATGCTGTACATTACCACCCACTGCTTAAGCCTCAATATATTTTAGTTCAC
Pri-miR-520 F2	69.6	*GATC*CTCGAGACTAGTATTCCAACAAAAAATCCACGGTGCCACAGCAAGAAGATGTCAGGCTG
Pri-miR-520-R2	70.3	*GATC*GCGGCCGCTAGCTTATTGCCAAGATCAGCATCCACCTAAACGTTGCTTTTCTTAAACG
Pri-miR-23-F2	77.9	*GATC*CTCGAGACTAGTCAGGTGCCAGCCTCTGGCCCCGCCCGGTGCCCCCCTCACCCCTGTGCCAC
Pri-miR-23-R2	78.5	*GATC*GCGGCCGCTAGCGCCTCTCTGCCACCCCGTCCCCGGGCAGCATCCTCGGTGGCAGAGCTCAGG
Pri-miR-934-F2	70.1	*GATC*CTCGAGACTAGTATCTGGGCCAGCCTTTGATGGTGTGTGTCTGTATCCTGAGAACGAACTTGAG
Pri-miR-934-R2	67.7	*GATC*GCGGCCGCTAGCTTTTTGAGGCTAAAATGAAACAAATTCTATCTATGAAAACACTGAGTGATAG
Pri-miR-34-F2	76.2	*GATC*CTCGAGACTAGTGAGAGGCAGGACAGGCCTGTCCCCCGAGTCCCCTCCGGATGCCGTGGACCGGCCAGCTG
Pri-miR-34-R2	75.8	*GATC*GCGGCCGCTAGCTCCTGGCGTCTCCCACTGGTCTGGGCATCTCTCGCTTCATCTTCCCTCTTGGGCCCCAC

F: forward extension primer; R: reverse extension primer. The terminal 4 nucelotide fragment (*GATC*) of round two primers is shown in italic font.
Flanking *Xho*I (CTCGAG), *Spe*I (ACTAGT), *Nhe*I (GCTAGC) and *Not*I (GCGGCCGC) restriction sites of round 2 primers are underlined.
Tm: melting temperature; calculated using OligoAnalyser v3.1 with default parameters.

pCI-neo Mammalian Expression Vector (Promega, Madison, WI, USA) for constitutive expression from the pol II cytomegalovirus (CMV) immediate-early enhancer/promoter.

3.2.2 Short hairpin RNA expression plasmids

ShRNAs expressing each anti-viral guide sequence were used as positive controls for guide expression and function (Figure 3.1). These shRNAs were previously constructed using a two-step PCR method (as in 2.2.3), inserted into the pTZ vector and characterized by other members of the lab: sh-tat (Saayman et al., 2008), sh-int (Saayman et al., 2010), sh-gag-1, -2 (Nhlabatsi, 2014) & sh-LTR-1, -2 (Nhlabatsi, 2014). ShRNA sequences were verified through fluorescence-based cycle DNA sequencing (Inqaba Biotec, Pretoria, South Africa) with the primers listed in Table 2-3.

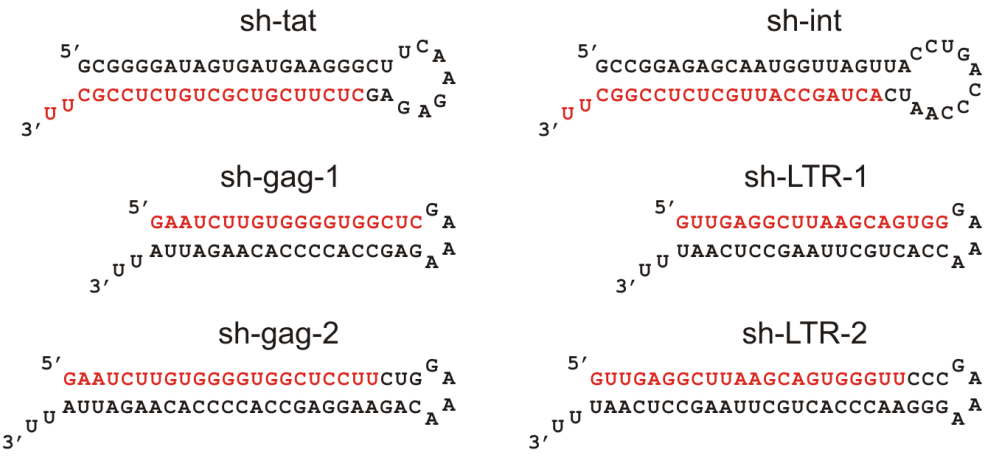


Figure 3.1: ShRNAs used as positive controls for exogenous RNA guides.
Previously constructed shRNAs designed to express functional RNA guides targeting the *tat*, *int*, *gag* and LTR regions of the HIV genome. The anticipated guide sequences are shown in red.

All six hairpins were suitable for use as positive controls for guide production and function, but there were small differences in the shRNA format. Firstly, an RNA pol III promoter was used for high level, constitutive expression of each hairpin, but sh-tat and sh-int were expressed from the human U6 small nuclear RNA (snRNA) promoter (as in 2.2.3), while sh-gag and sh-LTR hairpins were expressed from the human H1 RNA promoter (Brummelkamp et al., 2002). Secondly, different loop sequences were used: the effective 9 nt loop of the pSuper vector (Brummelkamp

et al., 2002) was used for sh-tat; a loop based on the miR-22 precursor was used for sh-int (as in 2.2.3); and a tetraloop (Vlassov et al., 2007) was used for sh-gag and sh-LTR hairpins. Thirdly, guide sequences were placed in the 3p arm of the hairpin for sh-tat and sh-int, while the guide was in the 5p arm for sh-gag and sh-LTR hairpins to provide greater 5' sequence consistency following Dicer processing (Gu et al., 2012). Lastly, sh-gag-1 and sh-LTR-1 were designed with 19 bp stem regions for Dicer-independent processing (Dallas et al., 2012); while sh-gag-2 (25 bp) and sh-LTR-2 (25 bp), as well as sh-tat (22 bp) and sh-int (22 bp), were designed for Dicer-dependent processing. Sh-gag-1 and sh-LTR-1 mediated more effective target silencing and were used as positive controls in dual luciferase reporter assays, while sh-gag-2 and sh-LTR-2 generated ~ 22 nt guide RNAs and were suitable as positive controls in northern blot analysis. Overall, the differences in format, while notable, were not of great concern here in the use of these shRNAs as positive control constructs.

3.2.3 Target reporter plasmids

Reporter plasmids containing viral target sequences (Table 3.3) were constructed previously by other members of the lab using the psiCHECK^{TM}-2 Vector (Promega, Madison, WI, USA): psiCHECK-*tat* (Saayman et al., 2008), psiCHECK-*int* (Saayman et al., 2010), psiCHECK-*gag* (Nhlabatsi, 2014) and psiCHECK-LTR (Nhlabatsi, 2014). The *gag* and LTR genomic target sequences used by Mcintyre *et. al*. (McIntyre et al., 2009a) were extended by N. Nhlabatsi to accommodate silencing by sh-gag-2 and sh-LTR-2 and the *int* genomic target sequence (Saayman et al., 2010) was modified to contain hairpin-specific residues. The identity of target sequences was confirmed by sequencing with primers listed in Table 2.3 and alignment with the HIV genome from the pNL4-3 infectious molecular clone (Adachi et al., 1986) (Genbank Accession (Benson et al., 2015) AF324493.2).

Table 3.3: HIV target sequences

HIV-1 target	Sequence (5' to 3' orientation)
tat (Lee et al., 2002a)	GCGGAGACAGCGACGAAGAGC
int (Saayman et al., 2010)	AGCCGGAGAGCAATGGCTAGT
gag (Nhlabatsi, 2014)	CAGAAGGAGCCACCCCACAAGATTA
LTR (Nhlabatsi, 2014)	GGGAACCCACTGCTTAAGCCTCAAT

Extended or hairpin-specific nucleotides of the target sequence are underlined.

3.2.4 Assessment of pri-miRNA mimic function in cultured mammalian cells

The function of pri-miRNA mimics in cell culture was assessed using the Dual - Luciferase ® Reporter Assay System (Promega, Madison, WI, USA) as described in section 2.2.5.

3.2.5 Assessment of pri-miRNA mimic processing in cultured mammalian cells

Northern blot analysis

The generation of guide sequences was assessed through northern blot analysis, following the protocol described in section 2.2.6. Locked nucleic acid (LNA™; Exiqon, Woburn, MA, USA) bases were incorporated in oligonucleotide probes (Table 3.4) for improved detection of tat, int, gag and LTR guide sequences.

Table 3.4: LNA™-modified probes for detection of guides in northern blot analysis.

Probe	RNA Tm (°C)	Sequence (5' to 3' orientation)
tat	78	GC+GGA+GACA+GCGACGAAG
int	80	CC+GGA+GAGCAATG+GCTAG
gag (Nhlabatsi, 2014)	80	GAG+CCAC+CCCAC+AAGATTC
LTR (Nhlabatsi, 2014)	81	C+C+A+CTGCTTAAGCCTC+A+AC

+ indicates LNA™ modification of the proceeding base (Exiqon, Woburn, MA, USA).
RNA Tm: melting temperature of the LNA probe with an RNA target, as indicated by the supplier.

Deep sequencing

The generation of guide sequences from pri-miR-34a and pri-miR-513a-2 and respective anti-*int* pri-miRNA mimics was examined more closely through deep sequencing of small RNAs. HEK293 cells were cultured as described in section 2.2.5 and seeded in a 6 well format (~ 4.8 x 10^5 cells / well) 24 h prior to co-transfection with 5 µg total pDNA (4.5 µg pri- / shRNA expression plasmid; 0.5 µg pCI-eGFP). Cell culture medium was replaced 24 h post-transfection and total RNA was extracted 48 h post- transfection using TRI Reagent ® (Sigma-Aldrich, St. Louis, MO, USA).

RNA samples were sequenced using the Illumina Genome Analyzer II platform. The 3'-adapter of raw reads were trimmed and reads were aligned to the full pri-miRNA reference sequence using Novoalign (http://www.novocraft.com). Aligned reads were converted to .bam format and indexed using Samtools (Li et al., 2009) and visualized using Integrative Genomics Viewer (IGV, ver. 2.3) (MacRae et al., 2008). Further quantification analysis was performed by Dr. Jonathan Hart of the Scripps Research Institute using miRDeep2 (Friedlander et al., 2012). Identifier sequences of 17 nt, starting with the third nucleotide from the 5' end of each expected 22 nt miRNA sequence, were used to identify RNA species aligning to the core of each miRNA. The expected miRNA sequences were adjusted in relation to the Drosha-DGCR8 cleavage site of the major characterized miRNA species. Reads were ranked according to frequency and quantified in Reads Per Million (RPM) for a single experiment.

3.2.6 Analysis of thermodynamic properties of predicted pri-miRNA mimic structures

Thermodynamic properties were examined over the first 20 nt of the guide sequence for pri-miRNA and pri-miRNA mimic subsets. This region was commonly replaced in mimic design and corresponds to the core of the processed miRNA duplex. All free energy values (dG, kcal/mol) were generated by the mfold web server (version 3.5) (Zuker, 2003) with default parameters at 37 ºC. Total stability of the guide region was calculated by summation of the free energy values for individual structural elements. Average thermodynamic stability profiles were constructed in a similar manner to previous studies (Han et al., 2006; Khvorova et al., 2003; Krol et al., 2004). Three approaches were used to estimate the stability of positions within internal loops, which can not be assigned a precise free energy value using the nearest neighbour method for structural prediction. The free energy of internal loops was treated by (1) assigning an average loop value to all positions within the loop; (2) assigning the loop value only to the closing pair and omitting internal loop positions from calculations (Han et al., 2006); or (3) assigning the average value of a 3 nt window to the central position of that window, where the loop is considered as a single position (Krol et al., 2004). The difference in free energy between position 1 and 20 (ddG (1-20)) was used to assess the bias in guide region stability.

3.2.7 Analysis of thermodynamic properties of predicted miRNA duplexes

Conceptual dicing (Schwarz et al., 2003) was used to assess the relative terminal stability of anticipated miRNA duplexes. In an approach described by Krol *et. al.* (Krol et al., 2004), the stability of each terminus was assessed in the context of a predicted hairpin structure generated by the mfold web server (version 3.5) (Zuker, 2003) with default parameters. To generate a hairpin sequence for a particular terminus, the opposite terminus of the anticipated miRNA duplex was modified by removing the 2 nt 3' overhang and attaching a standard tetraloop (5' GAAA 3') (Vlassov et al., 2007). Terminal mismatches were assigned the value of the external loop feature.

3.2.8 Statistical analyses

Microsoft Excel 2007 and GraphPad Prism version 4.03 were used to construct graphs and perform statistical analyses. In the results of dual luciferase assays, standard deviation was used as a measure of sample variation and one-way ANOVA (analysis of variance) was used to compare sample means. A Dunnett's post-test was to compare the means of experimental samples with that of a control (mock) sample. A two-tailed P value was used to assess the probability that the same variation in the means could be expected from random sampling. Correlation in several figures was assessed using the Spearman's (nonparametric) rank coefficient (r_s) as x and y values were assumed to be independent and no assumption was made about the distribution of values. A value of zero indicates no correlation, while a value of + 1.0 or - 1.0 indicates a perfect positive or negative correlation, respectively. A two-tailed P value was used to assess the probability that the same correlation could be expected from random sampling.

3.3. RESULTS

3.3.1 Design of pri-miRNA mimics

Functional pri-miRNAs with various basal stem conformations and motif combinations (Chapter 2) were used in the design of corresponding pri-miRNA mimics. The endogenous miRNA guide sequences were replaced with one of four previously characterized, highly effective therapeutic siRNA guides (Figure 3.2 A). The anti-*tat* / *rev* (Lee et al., 2002a), anti-*int* ((Nishitsuji et al., 2006), anti-*gag* (McIntyre et al., 2009a) and anti-LTR (McIntyre et al., 2009a) guide sequences were designed to target expression from corresponding regions of the HIV-1 genome which code for the trans-activator and regulatory proteins (Tat and Rev), integrase (Int), gag polyprotein (Gag) and long terminal repeats (LTR); and were selected for their ability to effectively inhibit HIV-1 replication in short-term cell culture studies when expressed from a pol-III-driven shRNA. The original miRNA sequence was replaced with 20 - 22nt of a therapeutic guide, depending on the length of the original miRNA. The predicted secondary structure of mimics was kept as close as possible to that of the original pri-miRNA as in previous approaches (Ely et al., 2008; Liu et al., 2008), because structural features like internal bulges may be crucial to maintain productive processing of the intended guide (Aagaard et al., 2008; Burke et al., 2014; Han et al., 2006; Ma et al., 2013) (Figure 3.2 B). Alterations were made to the complementary arm of the stem region to maintain canonical base pairs, G:U wobble pairs and both symmetrical and asymmetrical mismatches, while no alterations were made to the loop or flanking sequences. To assess potential changes in the pri-miRNA conformation, predicted secondary RNA structures were again generated with the local and extended flanking sequence (Appendix A). Canonical hairpins were maintained in predicted structures generated with the local flanking sequence (Figure 3.3), although a single, alternative non-canonical form was also generated for pri-miR-520-tat (Table 3.5). The proportion of canonical foldings in structures generated with the extended flanking sequence was similar to that of the original pri-miRNA and only varied slightly for most mimics (Figure 3.3).

(Figure overleaf)

Figure 3.2 Design and predicted secondary RNA structures of pri-miRNA mimics.
(**A**) Pri-miRNA mimics were designed by replacing the original guide of a pri-miRNA with one of four artificial RNA guides (red) targeting the *tat*, *int*, *gag* or LTR regions of the HIV genome. (**B**) Predicted secondary RNA structures generated for anti-*tat*, anti-*int*, anti-*gag* and anti-LTR pri-miRNA mimics with 15 nt of local sequence flanking the anticipated Drosha-DGCR8 cleavage site. The guide sequence (red) and Gibbs free energy (dG) of each structure are indicated.

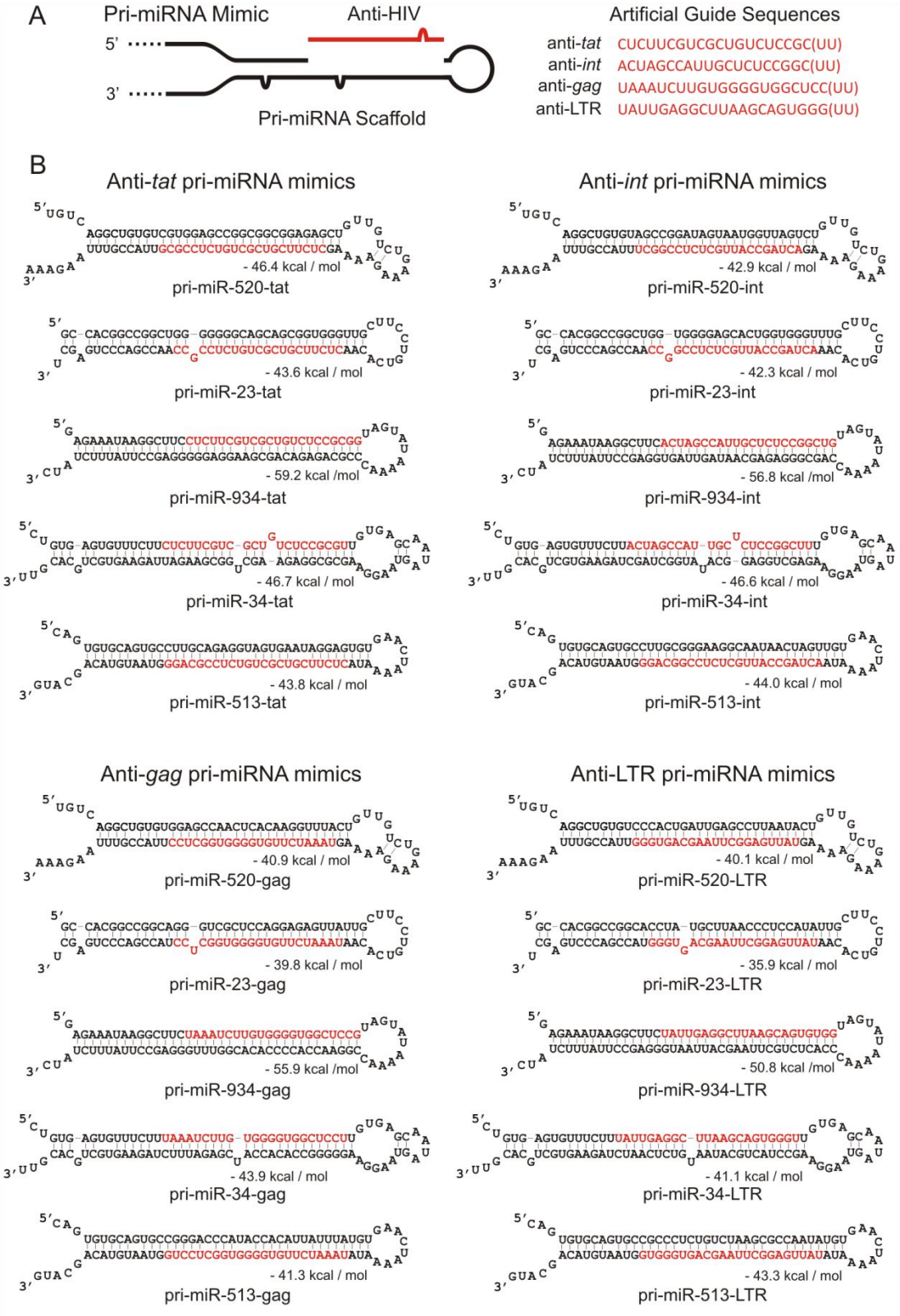


Figure 3.2: Design and predicted secondary RNA structures of pri-miRNA mimics.

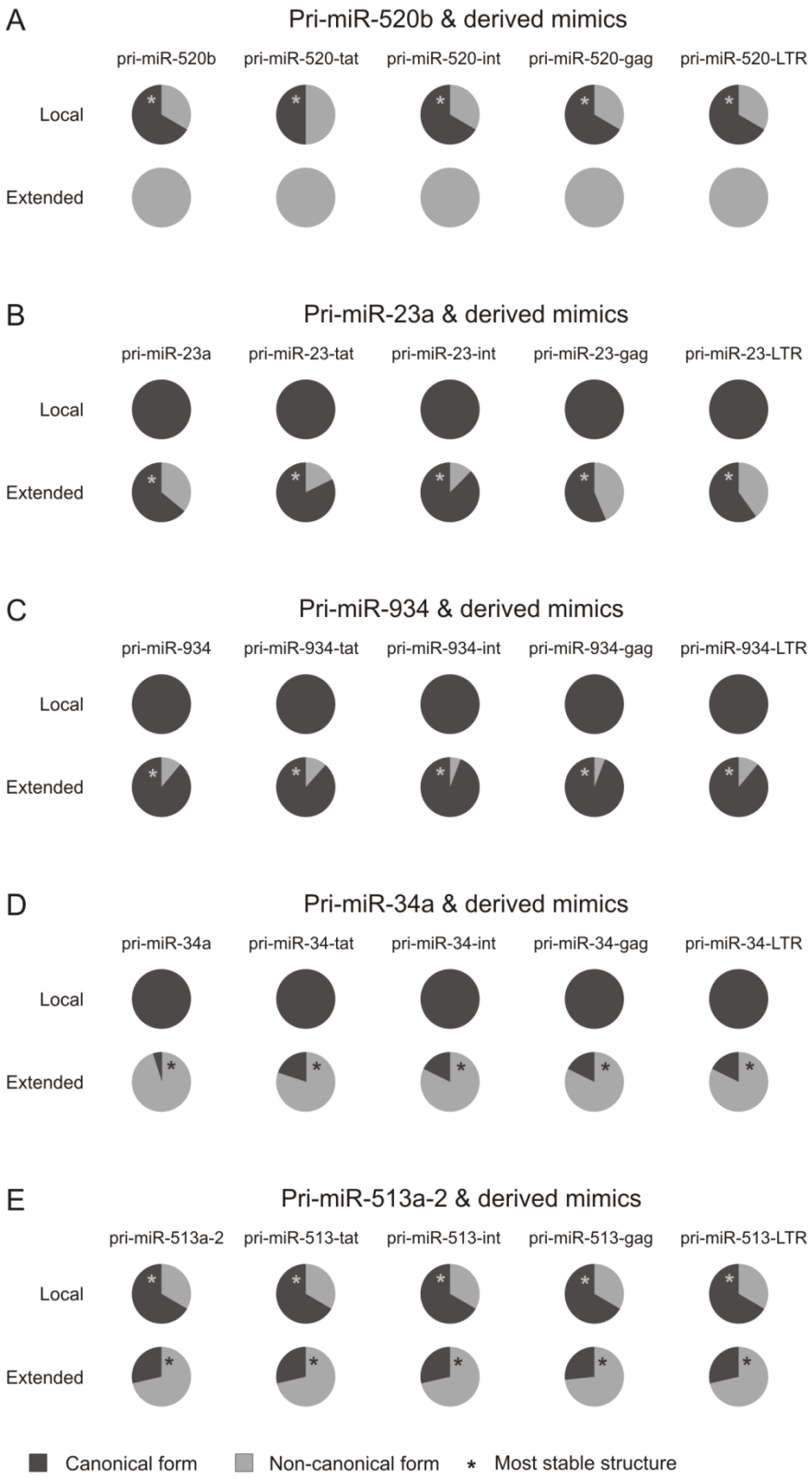


Figure 3.3: The proportion of canonical and non-canonical predicted structures.

(Figure 3.3 on previous page)

Figure 3.3 The proportion of canonical and non-canonical predicted structures. The proportion of canonical and non-canonical hairpin structures for (**A**) pri-miR-520b, (**B**) pri-miR-23a, (**C**) pri-miR-934, (**D**) pri-miR-34a and (**E**) pri-miR-513a-2 and derived mimics generated with local (20 nt) flanking sequence or the extended flanking sequence of the expressed pri-miRNA (Table 3.5 values).

Table 3.5: The basal stem length (BSL) and free energy (dG) of predicted secondary RNA structures of pri-miRNA mimics.

A

Input Sequence	Pri-miR-520b		Pri-miR-520-tat		Pri-miR-520-int		Pri-miR-520-gag		Pri-miR-520-LTR	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
+ 15 nt	9	- 42.0•	9	- 46.4•	9	-42.9•	9	-40.9•	9	-40.1•
+ 20 nt	9	- 42.0•	9	-46.4•	9	-42.9•	9	-40.9•	9	-40.1•
	4	- 40.0	9	-44.1	4	-40.9	4	-38.9	4	-38.1
	9	- 39.7	4	-44.4	9	-40.6	9	-38.6	9	-37.8
			4	-44.5						
+ 50 nt	0	- 59.9•	0	-64.3•	0	-61.0•	0	-58.8•	0	-58.0•
	0	- 56.0	0	-60.4	0	-57.1	0	-54.9	0	-54.1
	0	- 53.9	0	-58.3	0	-55.0	0	-52.8	0	-52.0
	0	- 52.8	0	-57.2	0	-53.9	0	-51.7	0	-50.9
Full pri-miRNA	0	- 215.4	0	-224.0	0	-220.7	0	-218.5	0	-217.7
	0	-221.0•	0	-220.2	0	-216.9	0	-214.7	0	-213.9
	0	- 212.4	0	-225.3	0	-222.0	0	-219.8	0	-219.0
	0	- 216.0	0	-228.6•	0	-225.3	0	-223.1•	0	-222.3•
	0	- 212.5	0	-225.3	0	-222.0	0	-219.8	0	-219.0
	0	- 204.3	0	-226.1	0	-222.8•	0	-220.6	0	-219.8
	0	- 204.4	0	-212.9	0	-209.6	0	-207.4	0	-206.6
	0	- 212.9	0	-222.5	0	-219.2	0	-217.0	0	-216.2
	0	- 217.1	0	-219.4	0	-216.1	0	-213.9	0	-213.1
	0	- 211.6	0	-217.5	0	-214.2	0	-212.0	0	-211.2
	0	- 206.6	0	-220.0	0	-216.7	0	-214.5	0	-213.7
	0	- 211.2	0	-204.5	0	-201.2	0	-199.0	0	-198.2
	0	- 195.5	0	-209.4	0	-206.1	0	-203.9	0	-203.1
	0	- 200.7	0	-208.0	0	-204.7	0	-202.5	0	-201.7
	0	- 199.4	0	-220.4	0	-217.1	0	-214.9	0	-214.1
	0	- 199.9	0	-215.5	0	-211.9	0	-210.5	0	-209.2
	0	- 205.8	0	-204.2	0	-200.9	0	-198.7	0	-197.9
	0	- 206.6	0	-207.6	NH	-203.7	0	-202.1	0	-201.3
	0	- 203.4	NH	-208.2	0	-204.3	NH	-200.0	NH	-206.4
	0	- 200.2			NH	-203.8	NH	-192.3	NH	-201.2
	0	- 208.1			NH	-197.5	NH	-194.7	NH	-201.5
									NH	-191.1
									NH	-182.0

A summary of the BSL and free energy of secondary RNA structures predicted for (**A**) pri-miR-520b, (**B**) pri-miR-23a, (**C**) pri-miR-934, (**D**) pri-miR-34a and (**E**) pri-miR-513a-2 and derived mimics. The table format follows that of Table 2.8. The input sequence used for mfold structure prediction included a specific length of flanking sequence on either side of the anticipated Drosha-DGCR8 cleavage site (e.g. + 15 nt) or the full pri-miRNA sequence of the expected transcript. The BSL was defined as the duplex region between the anticipated Drosha-DGCR8 cleavage site and the nearest disruptive element (a fork or any bulge of 3 nt or more). The contribution of G-U wobble pairs and non-canonical pairs was also included where appropriate. A BSL entry with a solid underline indicates that the secondary RNA structure adopted a characteristic hairpin form, including a conventional basal stem (8 - 13 bp) flanked by at least 4 nt of consecutive ssRNA on one side. A BSL entry with a dashed underline indicates that the length of the flanking sequences was insufficient to enable a characteristic hairpin form. The most stable predicted secondary RNA structure in each data set is indicated (•). NH: No Hairpin.

B

Input Sequence	Pri-miR-23a		Pri-miR-23-tat		Pri-miR-23-int		Pri-miR-23-gag		Pri-miR-23-LTR	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
+ 15 nt	13	- 38.3•	13	-43.6•	13	-42.3•	13	-39.8•	13	-35.9•
+ 20 nt	13	- 38.3•	13	-43.6•	13	-42.3•	13	-39.8•	13	-35.9•
+ 50 nt	13	- 66.5•	13	-71.8•	13	-70.5•	13	-68.0•	0	-58.3
	13	- 62.9	13	-68.2	13	-66.9	13	-64.4	13	-64.1•
	13	- 61.2	13	-66.5	13	-65.2	13	-62.3	0	-61.6
	5	-61.7	5	-67.0	5	-65.7	5	-63.2	13	-60.5
	0	- 61.5	13	-69.1	13	-67.8	13	-65.3	0	-59.2
	13	- 63.8	5	-65.1	5	-67.5	5	-61.3	13	-58.8
	5	-63.5					0	-61.8		
Full pri-miRNA	8	- 241.0•	8	-241.6	8	-240.3	NH	-243.9	8	-233.9
	8	- 237.2	8	-237.8	8	-236.5	8	-237.8	8	-230.1
	13	- 238.4	8	-240.6	8	-239.3	8	-234.0	NH	-236.2
	13	- 236.7	8	-238.9	8	-237.6	8	-236.8	8	-232.9
	13	- 237.4	8	-247.5	8	-246.3	8	-243.8	8	-231.2
	8	- 229.9	8	-230.5	8	-229.2	8	-226.7	8	-239.9•
	0	- 239.7	8	-243.9	13	-250.5•	NH	-242.6	0	-235.5
	13	-231.0	13	-251.8•	0	-241.4	NH	-230.4	8	-222.8
	0	- 233.2	0	-242.7	NH	-244.1	13	-248.0•	0	-234.6
	13	- 225.0	NH	-249.0	8	-226.0	NH	-230.5	0	-234.5
	5	-235.1	8	-227.3	8	-234.5	NH	-239.0	8	-219.6
	13	- 228.4	8	-235.8	8	-220.2	8	-217.7	8	-228.1
	13	- 233.5	8	-221.5	8	-232.7	8	-224.5	8	-213.8
	0	- 235.1	8	-234.0	8	-227.0	8	-221.0	8	-226.3
	8	- 226.3	NH	-245.3	8	-223.5	NH	-229.4	NH	-238.6
	8	- 225.0	8	-228.38	13	-227.3	NH	-230.4	NH	-226.2
	0	- 234.9	8	-224.8					8	-220.6
	8	- 220.9							8	-217.1
	13	- 231.8							NH	-226.0
	NH	-234.3							NH	-221.3
	13	- 235.7								
	0	- 230.7								
	NH	-228.6								
	8	- 224.2								
	8	-220.4								
	NH	-235.6								
	NH	-224.6								
	13	-222.7								

C

Input Sequence	Pri-miR-934		Pri-miR-934-tat		Pri-miR-934-int		Pri-miR-934-gag		Pri-miR-934-LTR	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
+ 15 nt	12	-55.0•	12	-59.2•	12	-56.8•	12	-55.9•	12	-50.8•
+ 20 nt	12	-55.0•	12	-59.2•	12	-56.8•	12	-55.9•	12	-50.8•
+ 50 nt	12	-68.4•	12	-72.6•	12	-70.2•	12	-69.3•	12	-64.2•
Full pri-miRNA	12	-240.8	12	-251.5•	12	-249.1•	12•	-248.2•	12	-243.1•
	12	-237.0	12	-247.7	12	-245.3	12	-244.4	12	-239.3
	12	-237.1	12	-250.3	12	-247.9	12	-247.0	12	-241.9
	12	-229.7	12	-248.6	12	-246.2	12	-245.3	12	-240.2
	12	-241.2•	12	-240.4	12	-238.0	12	-237.1	12	-232.0
	12	-239.4	12	-244.7	12	-246.8	12	-241.4	12	-240.8
	12	-237.2	12	-244.3	12	-241.9	12	-241.0	12	-235.9
	12	-243.3	12	-243.0	12	-240.6	12	-239.7	12	-234.6
	12	-222.5	12	-227.0	12	-224.6	12	-223.7	12	-218.6
	12	-224.8	11	-241.3	11	-238.9	11	-238.0	11	-232.9
	11	-232.2	12	-235.5	12	-233.1	12	-232.2	12	-227.1
	11	-234.0	12	-236.9	12	-234.5	12	-233.6	12	-228.5
	12	-227.8	12	-245.4	12	-243.0	12	-242.1	12	-237.0
	12	-236.3	11	-240.2	11	-237.8	11	-236.9	11	-231.8
	12	-237.4	12	-239.1	12	-236.7	12	-235.8	12	-230.7
	12	-234.6	12	-243.7	12	-241.3	12	-240.4	12	-235.3
	11	-227.5	12	-235.5	12	-233.1	12	-232.2	12	-227.1
	NH	-214.9							NH	-221.1

D

Input Sequence	Pri-miR-34a		Pri-miR-34-tat		Pri-miR-34-int		Pri-miR-34-gag		Pri-miR-34-LTR	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
+ 15 nt	11	-42.2•	11	-46.7•	11	-46.4•	11	-43.9•	11	-41.1•
+ 20 nt	15	-45.2•	15	-49.7•	15	-49.6•	15	-46.9•	15	-44.1•
+ 50 nt	15	-63.8•	15	-68.3•	15	-68.2•	15	-65.5•	15	-62.7•
	22	-59.9	22	-64.4	22	-64.3	22	-61.6	22	-58.8
	18	-59.6	18	-64.1	18	-64.0	18	-61.3	18	-58.5
	15	-60.5	15	-65.0	15	-64.9	15	-62.2	15	-59.4
	15	-58.6	15	-63.1	15	-63.0	15	-60.3	15	-57.5
	19	-56.4	19	-60.9	19	-60.9	19	-58.1	19	-55.3
Full pri-miRNA	11	-245.8•	11	-252.4•	11	-252.3•	11	-249.6•	11	-246.8•
	11	-240.5	11	-246.0	11	-245.9	11	-243.2	11	-240.4
	11	-245.1	11	-250.0	11	-249.9	11	-247.2	11	-244.4
	18	-240.4	11	-245.7	11	-245.6	11	-242.9	11	-240.1
	11	-232.4	11	-243.7	11	-243.6	11	-240.9	11	-238.1
	11	-234.8	11	-240.0	11	-239.9	11	-237.2	11	-234.4
	11	-240.9	4	-246.8	4	-246.7	4	-244.0	4	-241.2
	11	-245.6	11	-239.1	11	-239.0	11	-236.3	11	-233.5
	11	-233.4	11	-247.6	11	-247.5	11	-244.8	11	-242.0
	11	-242.4	11	-238.7	11	-238.6	11	-235.9	11	-233.1
	11	-238.6	18	-242.3	18	-242.2	18	-239.5	18	-236.7
	11	-241.8	22	-249.7	22	-249.6	22	-246.9	22	-244.1
	18	-236.7	18	-228.9	0	-248.7	18	-226.1	18	-223.3
	11	-238.3	11	-240.7	18	-228.8	11	-237.9	11	-235.1
	4	-232.2	18	-241.3	11	-240.6	18	-238.5	NH	-238.9
	22	-244.7			18	-241.2	NH	-242.2	18	-235.7
	NH	-232.2			NH	-236.3	NH	-230.8	NH	-230.5
	22	-236.9								
	NH	-239.5								
	NH	-229.1								

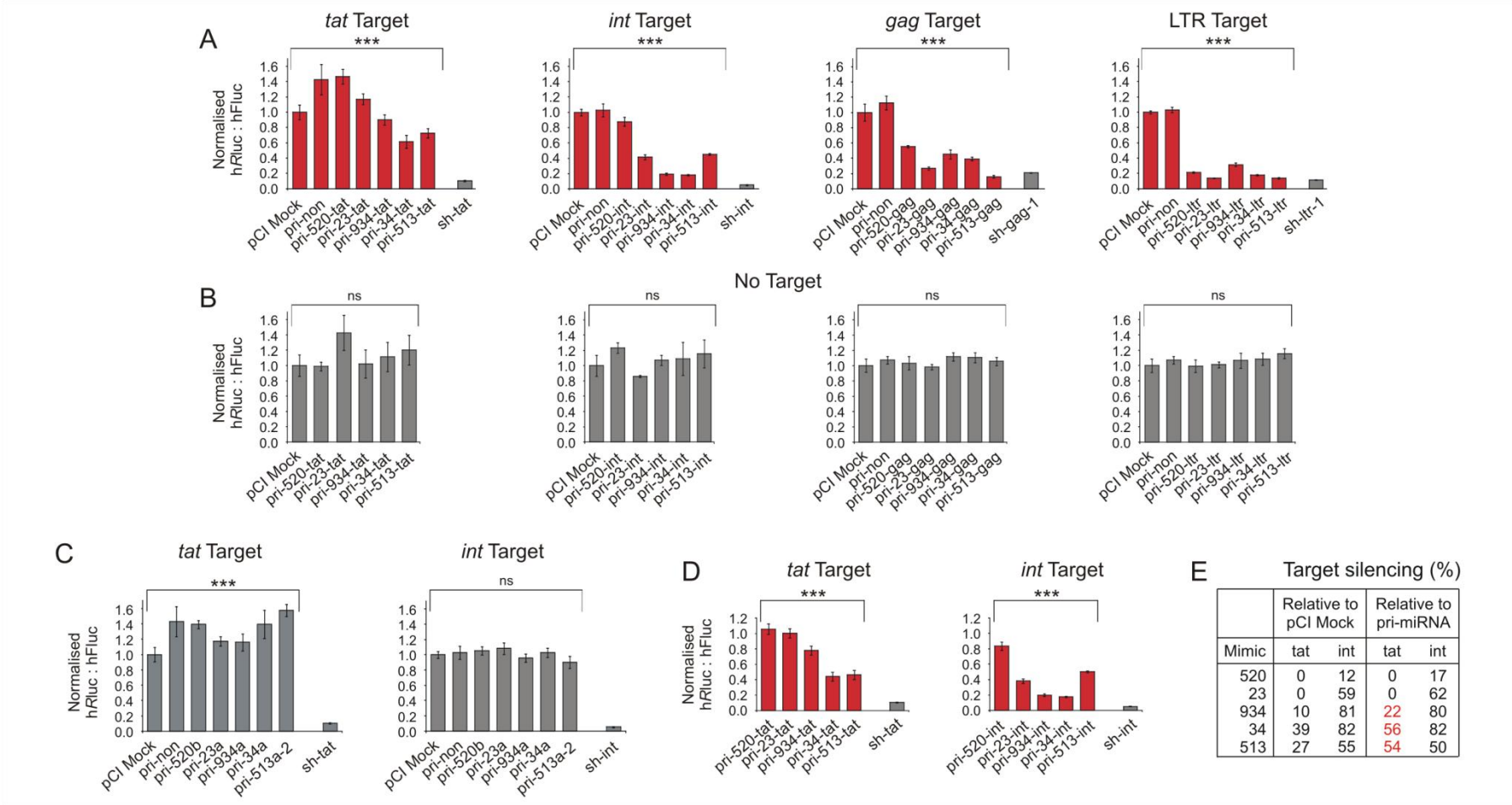
E

Input Sequence	Pri-miR-513a-2		Pri-miR-513-tat		Pri-miR-513-int		Pri-miR-513-gag		Pri-miR-513-LTR	
	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)	BSL (bp)	dG (kcal/mol)
+ 15 nt	10	-40.9•	10	-43.8•	10	-44.0•	10	-41.3•	10	-43.3•
+ 20 nt	10	-40.9•	10	-43.8•	10	-44.0•	10	-41.3•	10	-43.3•
	7	-40.5	7	-43.4	7	-43.6	7	-40.9	7	-42.9
	10	-39.9	10	-42.8	10	-43.0	10	-40.3	10	-42.3
+ 50 nt	13	-65.9•	13	-68.8•	13	-69.0•	13	-66.3•	13	-68.3•
	10	-60.3	10	-63.2	10	-63.4	10	-60.7	10	-62.7
	10	-59.0	10	-61.9	10	-62.1	10	-59.4	10	-61.4
Full pri-miRNA	0	-241.3	0	-244.2	0	-244.4	0	-241.7	0	-243.7
	0	-237.5	0	-240.4	0	-240.6	0	-237.9	0	-239.9
	0	-248.1•	0	-251.0•	0	-251.2•	0	-248.5•	0	-250.5•
	0	-246.3	0	-249.2	0	-249.4	0	-246.7	0	-248.7
	10	-238.9	10	-241.8	10	-242.0	10	-239.3	10	-241.3
	0	-230.2	0	-233.1	0	-233.3	0	-230.6	0	-232.6
	10	-245.2	10	-248.1	10	-248.3	10	-245.6	10	-247.6
	0	-227.7	0	-230.6	0	-230.8	0	-228.1	0	-230.1
	10	-236.7	10	-239.6	10	-239.8	10	-237.1	10	-239.1
	0	-239.4	0	-242.3	0	-242.5	0	-239.8	0	-241.8
	0	-235.7	0	-238.6	0	-238.8	0	-236.1	0	-238.1
	0	-225.3	0	-228.2	0	-228.4	0	-225.7	0	-227.7
	13	-242.2	13	-245.1	13	-245.3	13	-242.6	13	-244.6
	0	-233.0	0	-235.9	0	-236.1	0	-233.4	0	-235.4
							NH	-227.3		

3.3.2 Functional characterisation of pri-miRNA mimics

The silencing efficacy of guides derived from pri-miRNA mimics was assessed by dual luciferase reporter assays. ShRNAs expressing each guide sequence were included as positive controls: sh-tat (Saayman et al., 2008), sh-int (Saayman et al., 2010), sh-gag-1 (Nhlabatsi, 2014) and sh-LTR-1 (Nhlabatsi, 2014). In contrast to the original pri-miRNAs, effective target silencing was not observed for all pri-miRNA mimics (Figure 3.4 A). Effective target silencing was observed more frequently for mimics based on a particular pri-miRNA scaffold. Target silencing of ~ 80 % or more was observed for two mimics based on both the pri-miR-34a and 513a-2 scaffolds, but for only one mimic based on each of the pri-miR-520b, 23a and 934 scaffolds. As a group, mimics based on the pri-mR-34a scaffold had the highest average target silencing (66 %), followed closely by mimics based on pri-miR-513a-2 (63 %), while mimics based on the pri-miR-520b scaffold had the lowest average target silencing (34 %). Effective target silencing was also observed more frequently for mimics within a particular guide sequence set. Target silencing of ~ 80 % or more was observed for four of the anti-LTR mimics, two of both the anti-*gag* and anti-*int* mimics, but none of the anti-*tat* mimics. In fact, three of the anti-*tat* mimics (pri-miR-520-tat, 23-tat, 934-tat) were essentially non-functional with target silencing of 10 % or less, while target silencing by pri-miR-34-tat and pri-miR-513-tat was poor (< 40 %). As a group, the anti-LTR mimic set was generally the most successful with the highest average target silencing (80 %), followed by the anti-*gag* (64 %), anti-*int* (58 %) and anti-*tat* (15 %) mimic sets, respectively. This ranking was reflected in target silencing by the set of pri-miR-520b mimics (pri-miR-520-tat: 0 %, int: 12 %, gag: 45 %, ltr: 79 %).

Reporter gene silencing was specific to individual guide-target interactions and no significant silencing was observed in the absence of a target sequence (Figure 3.4 B). However, slightly elevated h*R*luc : hFluc values for the anti-*tat* mimic set (Figure 3.4 A, B) were re-examined. Raw values of the average silencing ratio for control samples of the anti-*tat* and anti-LTR mimic sets were almost identical (tat mock: 0.67, LTR mock: 0.73) (sh-tat: 0.07, sh-ltr-1: 0.08), suggesting that values obtained for the anti-*tat* mimic set were valid. To identify potential non-specific background effects, assays for *tat* and *int* targets were repeated using the original pri-miRNAs (Figure 3.4 C). No significant effects were observed between the original pri-miRNAs and the *int* target, but a positive effect was observed for the *tat* target, suggesting that silencing by anti-*tat* mimics in Figure 3.4 A may be undervalued. To show guide-specific silencing, h*R*luc : hFluc values for each anti-*tat* mimic were presented relative to the respective original pri-miRNA (Figure 3.4 D). In contrast to the anti-*int* mimic set, target silencing values improved by 12 - 27 %



	Relative to pCI Mock		Relative to pri-miRNA	
Mimic	tat	int	tat	int
520	0	12	0	17
23	0	59	0	62
934	10	81	22	80
34	39	82	56	82
513	27	55	54	50

Figure 3.4: Silencing efficacy of pri-miRNA mimics.

(Figure 3.4 on previous page)

Figure 3.4: Silencing efficacy of pri-miRNA mimics.
(**A**) Target gene silencing by guides derived from pri-miRNA mimics in HEK293 cells was measured by the mean ratio of *Renilla* luciferase : Firefly luciferase (h*R*luc : hFluc) (n = 3, ± s.d.). A corresponding shRNA (positive control) and a non-targeting pri-miRNA (pri-non) (negative control) were included. Values were normalised with respect to the pCI Mock sample. Sample means varied significantly (one-way ANOVA, ***P < 0.0001) (**B**) Assays were repeated with a control luciferase vector lacking the relevant target sequence (psiCHECK^{TM}-2) (n = 3, ± s.d.). Variation of sample means was not significant (ns) (one-way ANOVA, P > 0.05). (**C**) Assays with the *tat* and *int* targets were also repeated with the original pri-miRNAs (n = 3, ± s.d.). Sample means varied significantly for the *tat* target set (one-way ANOVA, *** P < 0.001, Dunnett's post-test), but not for the *int* target set (P > 0.05). (**D**) Target silencing values for anti-*tat* and anti-*int* mimics were normalised with respect to the original pri-miRNAs instead of the mock sample. Means varied significantly (one-way ANOVA, ***P < 0.0001) (**E**) A comparison of target silencing values normalised to either pCI Mock or the original pri-miRNA. Revised values that differed by more than 5 % are indicated (red).

for pri-miR-934-tat, 34-tat and 513-tat, but average silencing by the anti-*tat* mimic set remained poor (26 %) (Figure 3.4 E).

In the design of pri-miRNA mimics, the original miRNA sequence was replaced with 20 - 22nt of a therapeutic guide depending on miRNA length. Resulting minor variations in the last 1 - 3 nt of exogenous guides were not expected to drastically influence mimic function within each mimic set, where there was a minimum of 20 nt of common 5' guide sequence, including the miRNA seed region (2 - 8 nt) important for target recognition (Brennecke et al., 2005; Lewis et al., 2003); and a minimum of 19 complementary guide - target interactions (Figure 3.5 A). Non-complementary interactions were also expected between the 5' 1 - 2 nt of anti-*gag* and anti-LTR guides and the respective target sequence as a consequence of target design. These non-complementary interactions may be expected to slightly reduce target silencing, but occurred uniformly within each mimic set. In addition, anti-*gag* and anti-LTR mimic sets with a minimum of 19 complementary guide-target interactions were generally more effective than anti-*tat* and anti-*int* mimic sets with a minimum of 20 complementary guide-target interactions. Effective target silencing was also observed for sh-gag and sh-LTR positive controls, where only 18 complementary guide-target interactions were expected. Overall, the exact number of guide-target interactions was expected to vary, but no correlation was observed between the number of total (19 - 22 bp) or consecutive (18 - 22 bp) guide-target interactions and target silencing (Figure 3.5 B).

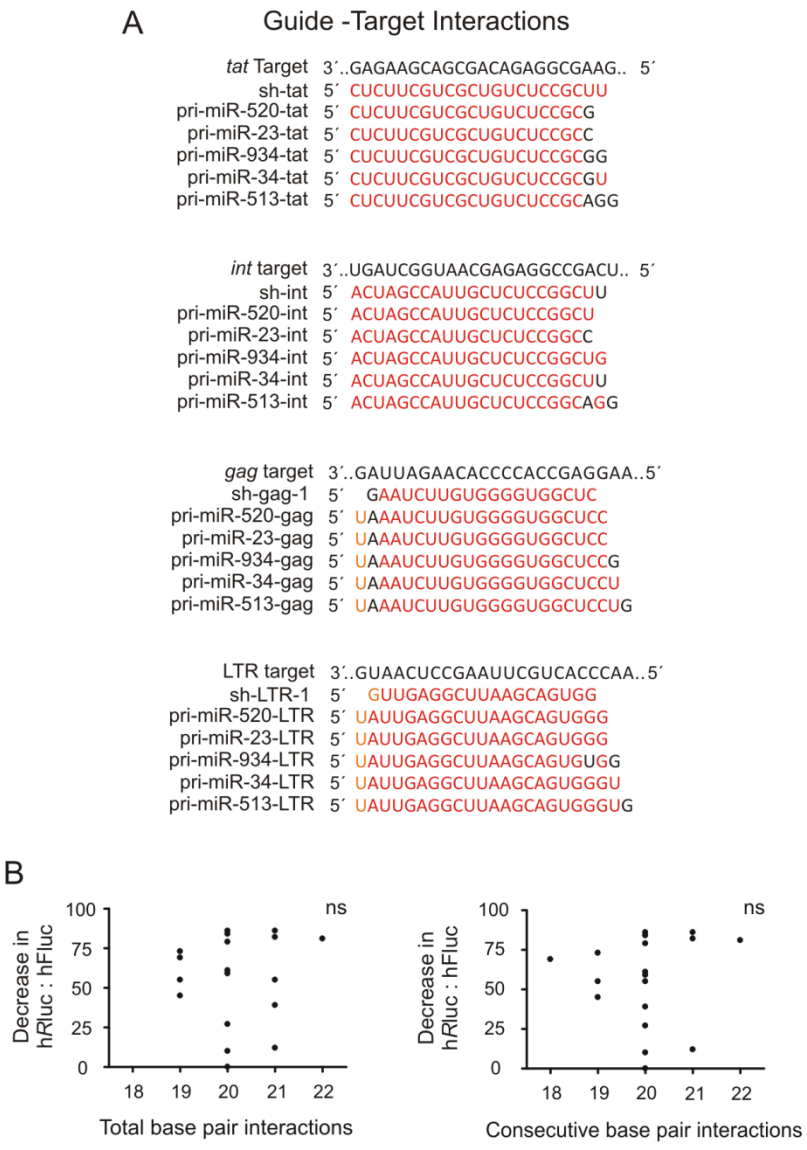


Figure 3.5: Expected guide-target interactions.

(**A**) The guide-target interactions expected to occur in target gene silencing assays for anti-*tat*, anti-*int*, anti-*gag* and anti-LTR mimic sets. Complementary base pairing (red) and G:U wobble pairs (orange) are indicated. (**B**) The total or consecutive number of expected base pair interactions was plotted against the corresponding decrease in h*R*luc : hFluc, (n = 20). No significant (ns) correlation was observed. Spearman (ranked) correlation co-efficient (rs^{TOTAL} = 0.19), (rs^{CONS}. = 0.27) (two-tailed P value > 0.25).

3.3.3 Processing of exogenous RNA guides from pri-miRNA mimics

Silencing assay results suggested that functional pri-miRNA mimics, like those of the anti-LTR set, could be generated using natural pri-miRNA scaffolds with a range of basal stem lengths (9 - 13 bp) and motif combinations, where the canonical hairpin was present in almost all, most or none of the predicted structures including the extended flanking sequence. To investigate the processing of pri-miRNA mimics, northern blot analysis was performed on total RNA extracted from HEK293 cells transfected with corresponding expression vectors and guides were detected by hybridisation with LNA^{TM}-modified oligonucleotide probes. In agreement with silencing assay results, processed ~ 22 nt guides were detected for functional pri-miRNA mimics (Figure 3.6). Guides were detected for pri-miR-934-tat, pri-miR-34-tat and pri-miR-513-tat, which supported revised target silencing values. MiRNAs of ~ 21 - 24 nt were also detected for several mimics, suggesting that the observed target silencing was mediated by several miRNA species. This also implied that exact processing sites might vary, although the detection of several ~ 65 nt pre-miRNAs supported cleavage at or near to the anticipated Drosha-DGCR8 sites. The signal intensity ratio was typically higher for the most effective mimics within each set, including guides derived from pri-miR-34-tat, pri-miR-513-tat, pri-miR-34-int, pri-miR-934-int, pri-miR-23-gag, pri-miR-23-LTR and pri-miR-513-LTR. However, signal intensity values did not necessarily reflect the order or magnitude of target silencing. For example, the signal intensity ratio calculated for pri-miR-513-gag was lower than that of pri-miR-934-gag, but target silencing mediated by these mimics was 84 % and 55 % respectively. The use of different probes also meant that signal intensity values could not be compared between different mimic sets, while relative values were higher for guides derived from anti-*gag* and anti-LTR mimics as a result of weaker sh-gag-2 and sh-LTR-2 controls. RNA bands detected across all samples of a blot were considered as non-specific endogenous products. Small nuclear U6 RNAs of ~ 107 nt were probed as a measure of loading consistency.

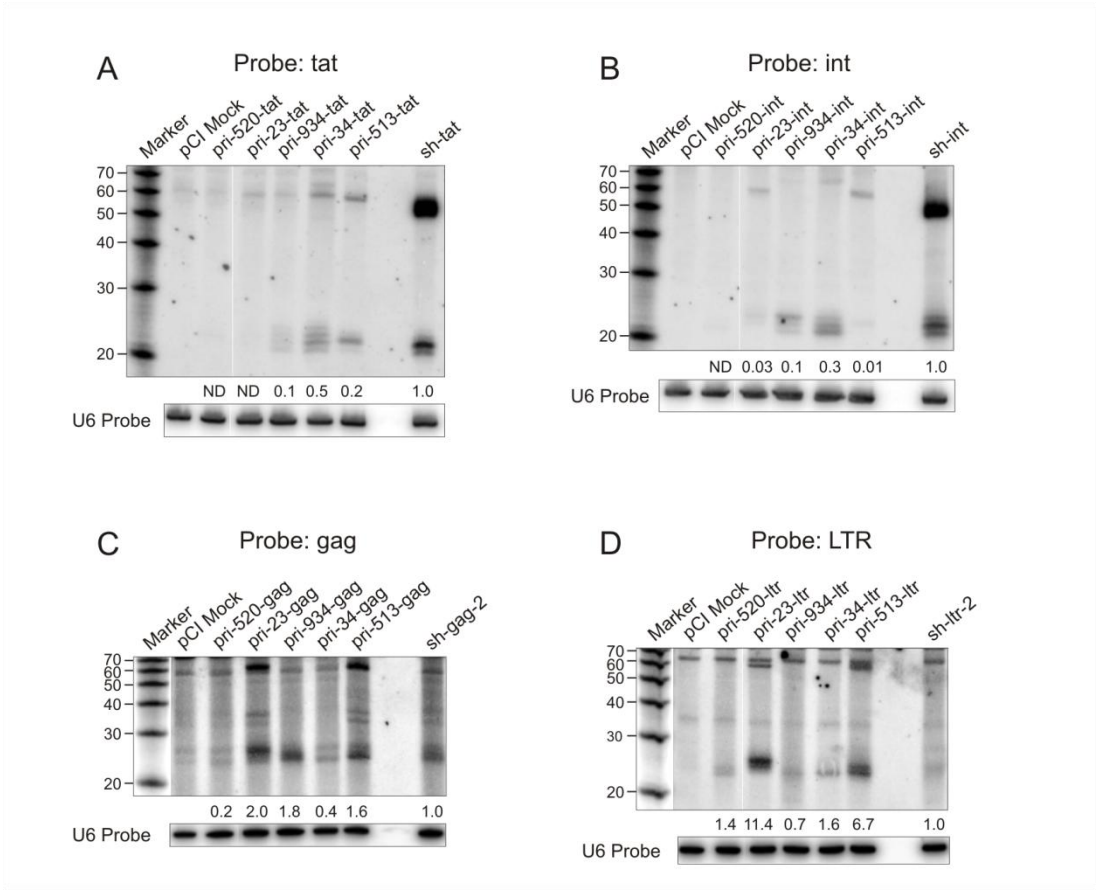


Figure 3.6: Processing of exogenous RNA guides from pri-miRNA mimics.

Exogenous guide sequences were detected through northern blot analyses for (**A**) anti-*tat*, (**B**) anti-*int*, (**C**) anti-*gag* and (**D**) anti-LTR mimics. RNA guides were detected following the transfection of HEK293 cells with corresponding pri-miRNA mimic or shRNA (positive control) expression vectors. A mock expression vector (pCI mock) was also included (negative control). Signal intensities of the RNA guides were quantified relative to U6 signal intensity (loading control) and are shown as a ratio of mimic-derived guide to shRNA-derived guide. ND: not-detected.

3.3.4 Deep sequencing of processed RNA guides

The results of characterisation studies indicated that the anticipated, functional RNA guides were successfully derived from pri-miRNA mimics, but a certain degree of guide heterogeneity exists. Deep sequencing was therefore used to investigate miRNA processing in more detail. Total RNA was extracted from HEK293 cells transfected with the relevant expression construct and filtered small RNA reads were aligned to a 17 nt identifier of each miRNA core sequence. The top ten reads aligned to pri-miR-34a, pri-miR-513a-2 and their respective anti-*int* mimics are shown in Figure 3.7 A - D. Overall, Drosha-DGCR8 and Dicer processing sites were well maintained between expressed pri-miRNAs and anti-*int* mimics, as indicated by the 5' ends of derived miRNAs. Variation of 1 - 2 nt was observed for both the Drosha and Dicer cleavage sites as cleavage can be imprecise (Starega-Roslan et al., 2011). However, variation was more frequent at the Dicer cleavage sites, which occur at ~ 22 bp from the end of the pre-miRNA and might be shifted as a result of 3' editing (Newman et al., 2011). This is in contrast to the Drosha-DGCR8 site, which occurs at ~ 11 bp from the ssRNA-dsRNA junction of the secondary RNA structure and is therefore not affected by minor 3' editing. This suggested that RNA guides with more consistent 5' seed regions (2 - 8 nt) (Lewis et al., 2003) were generated from the 5p arm. Most sequence heterogeneity was limited to the 3' ends of miRNAs and was indicative of the 3' trimming and editing (uridylation and adenylation) commonly observed in small RNA sequencing data (Landgraf et al., 2007) and in the generation of isomiRs (Neilsen et al., 2012). The anticipated 22 nt miRNA was observed in the top three reads for pri-miR-34a (93.5 RPM) and pri-miR-34-int (78.6 RPM) with similar abundance, but the anticipated ~ 23 nt miRNA was only observed for pri-miR-513a-2 and not pri-miR-513-int. However, the top reads for pri-miR-513-int-5p and pri-miR-513-int-3p were conventional 22 nt miRNAs derived from the same processing sites, suggesting that the Dicer processing site is shifted by 1 nt away from the terminal loop for pri-miR-513-int. MiRNAs derived from the 5p arm were also more abundant for pri-miR-513a-2 and pri-miR-513-int, suggesting that these are in fact the major miRNA species. This is in agreement with functional characterisation results for pri-miR-513a-2 (Figures 2.7 & 2.11), suggesting that the miRBase stem-loop should be re-annotated and that a more effective mimic could be designed by replacing the 5p guide of pri-miR-513a-2 scaffold. It should be noted that reads aligned to pri-miR-34a and pri-miR-513a-2 were expected to originate from both endogenous miRNA genes and the expressed pri-miRNA sequences; however, as the expressed pri-miRNAs were derived from the genomic sequence, processing is expected to be consistent. Fewer reads were expected to originate from endogenous miRNA genes, as relevant species were not detected in northern analysis (pCI Mock sample) (Figure 2.11).

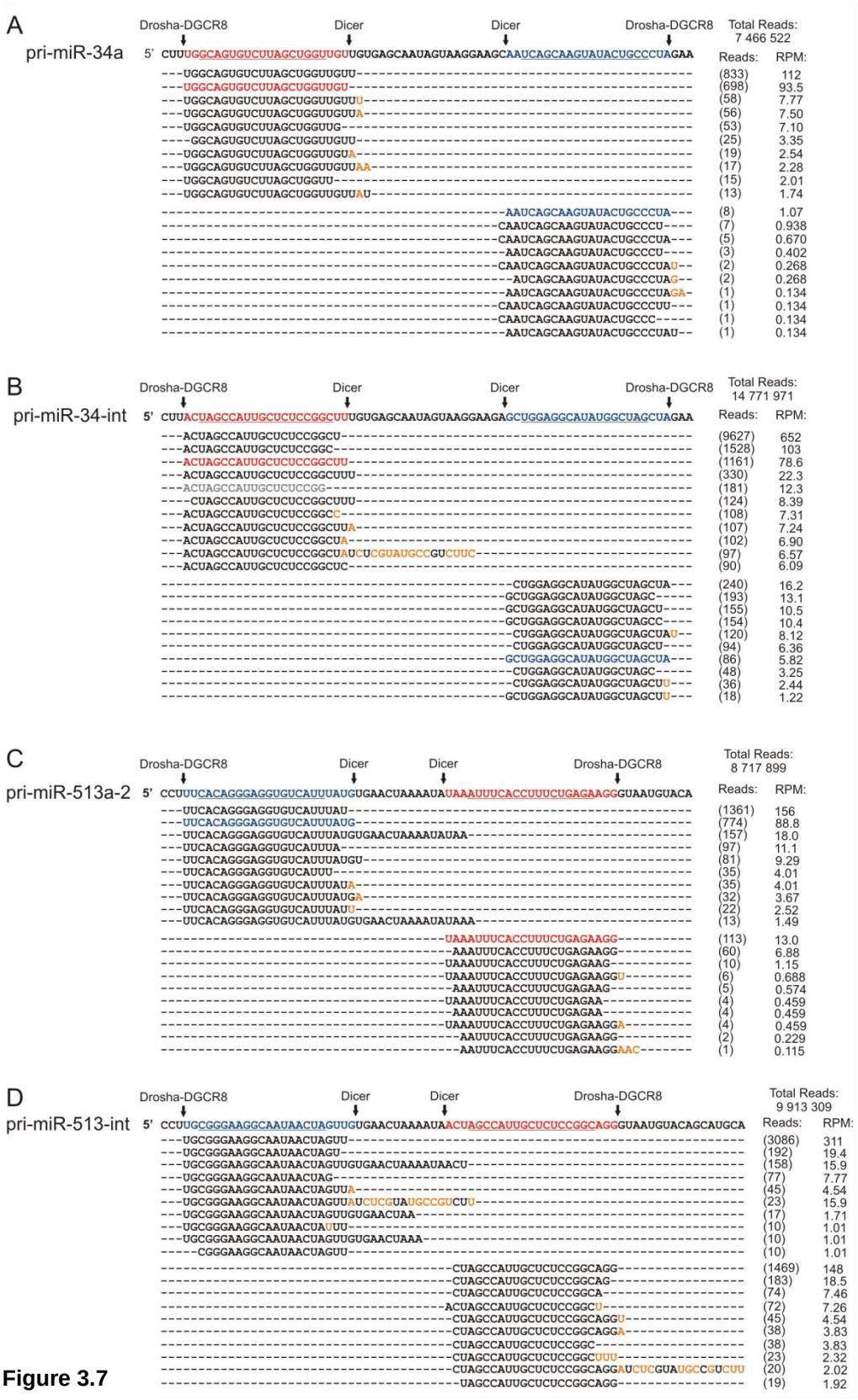

A
pri-miR-34a
Drosha-DGCR8
Dicer
Dicer
Drosha-DGCR8
Total Reads: 7 466 522
Reads:
RPM:
B
pri-miR-34-int
Total Reads: 14 771 971
C
pri-miR-513a-2
Total Reads: 8 717 899
D
pri-miR-513-int
Total Reads: 9 913 309

Figure 3.7

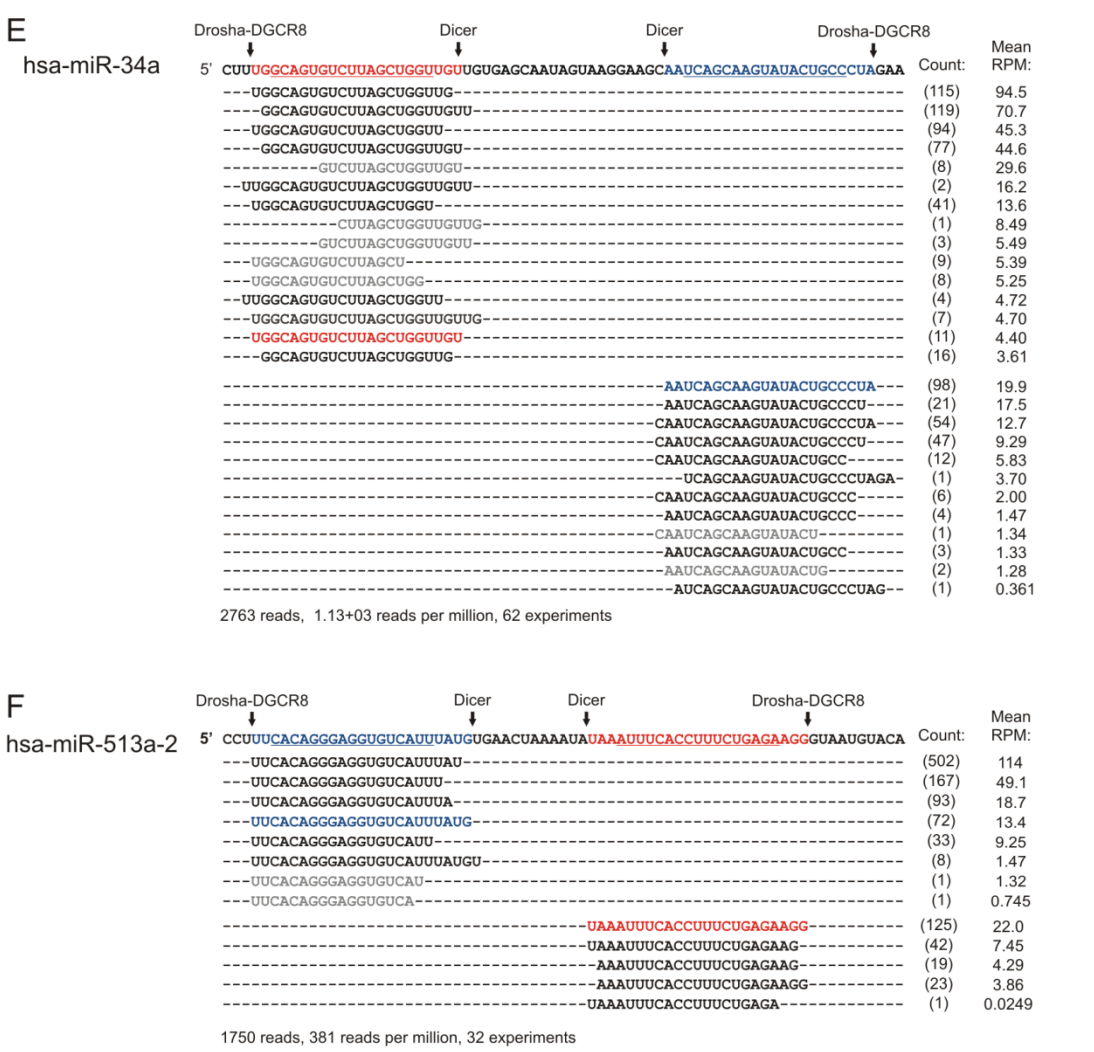


Figure 3.7: Precise sequences of processed miRNAs and exogenous RNA guides.

(**A** - **D**) Deep sequencing was used to assess guide production from expressed pri-miR-34a, pri-miR-513a-2 and corresponding anti-*int* pri-miRNA mimics. The top ten reads aligned to each 17 nt identifier (underlined) for major (red) and minor miRNA (blue) sequences are shown and were quantified in reads per million (RPM, 3 significant figures). Non-templated nucleotides (orange) and sequences that do not contain the full identifier (grey) are indicated. The anticipated processing sites corresponding to the major miRNA (red) species are marked. (**E**, **F**) Deep sequencing data compiled from multiple experiments was available on miRBase for hsa-miR-34a and hsa-miR-513a-2 stem-loops. The top ten reads aligned to the identifier sequences are shown if available.

Deep sequencing data now available on miRBase (Kozomara and Griffiths-Jones, 2014) was also examined to identify the top reads aligned to miR-34a and miR-513a-2 stem loops (Figure 3.7 E, F). The top reads were similar to those generated here and the anticipated miRNA species were identified, again supporting the ~ 22 nt pri-miR-513a-2-5p species as a major miRNA. Differences in the order of miRNA abundance were also observed, but miRBase data is compiled from multiple experiments in different cell lines and may be less reliable.

3.3.5 Guide region stability of pri-miRNA mimics

Together, these results indicated that effective miRNA guides can be derived and consistently processed from expressed pri-miRNA mimics with a range of basal stem lengths (9 - 13 bp) and predicted conformations. However, differences in mimic functionality could not be attributed to changes in the overall form of the predicted secondary RNA structure or primary sequence motifs, which were unaltered in mimic design. To identify a possible cause for the varied function of mimics, guide region stability was examined across the first 20 nt of the guide commonly replaced in mimic design. Three approaches based on previously described methods were used to estimate the stability of positions within internal loops, which can not be assigned a precise free energy value using the nearest neighbour method for structural prediction (mfold) (Zuker, 2003). The free energy of internal loops was treated by (1) assigning an average loop value to all positions within the loop; (2) assigning the loop value only to the closing pair and omitting internal loop positions from calculations (Han et al., 2006); or (3) assigning the average value of a 3 nt window to the central position of that window, where the loop is considered as a single position (Krol et al., 2004). A weak, but statistically significant, positive correlation (r_s: 0.46 - 0.59) was observed between the free energy of the guide region and target silencing using all three approaches (Figure 3.8 A). Free energy of the guide region was expected to be largely influenced by a common guide sequence for each mimic set and the change in free energy incurred in mimic design was therefore examined. A weak positive correlation between the change in free energy and target silencing was again observed using all three approaches (Figure 3.8 B). The free energy of the guide region for the more effective anti-LTR mimics was similar or slightly positive compared to the original guide region (0 to 2.5 kcal / mol; graph 1, 2), while the free energy of the guide region decreased for less effective anti-*tat* mimics (-1.7 to - 6.2 kcal / mol). This suggested that the identity and subsequent stability of the guide region might be a contributing factor to the general function of pri-miRNA mimic sets.

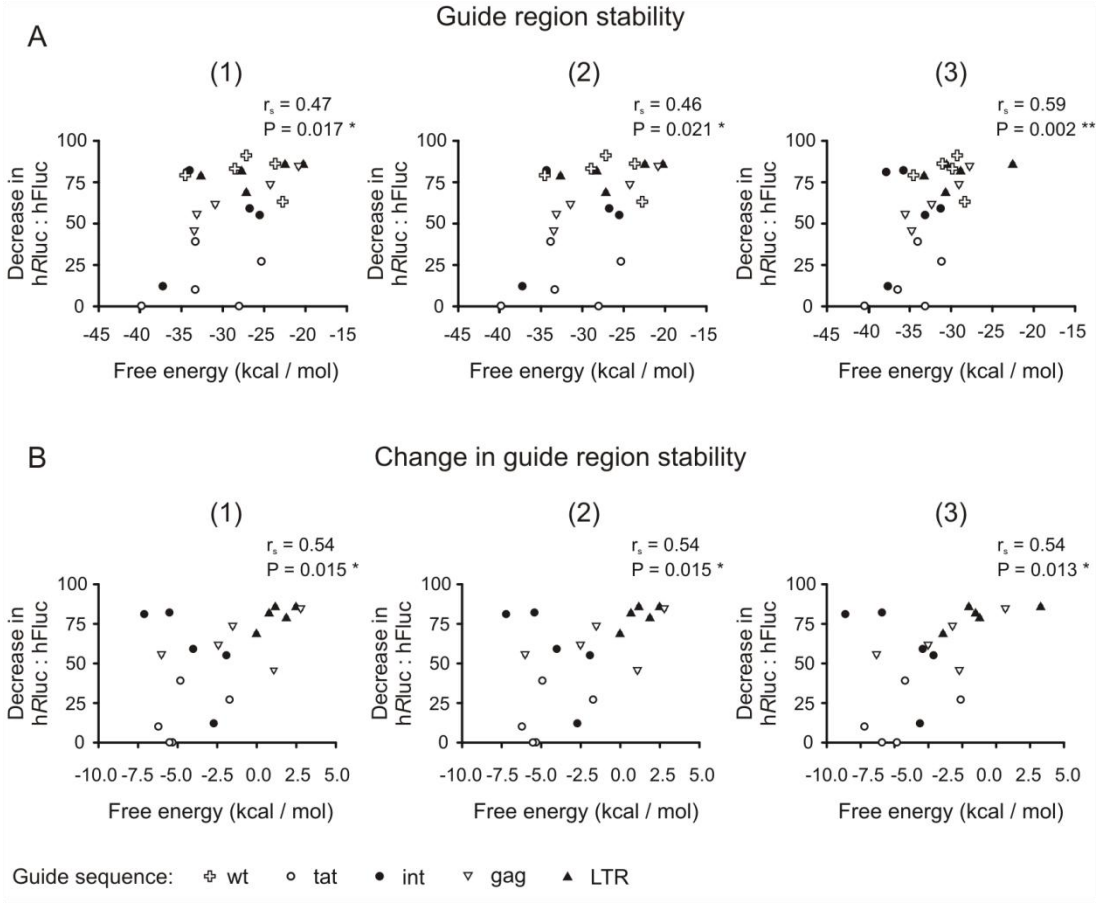


Figure 3.8: Predicted stability of the guide region.

(**A**) Stability of the first 20 nt of the guide region for pri-miRNAs and corresponding mimics (n = 25). (**B**) The change in stability of the guide region for each mimic in (A) compared to the original pri-miRNA (n = 20). In both panels, values were calculated using three different approaches to estimate the free energy of internal loop positions, by (1) assigning an average value for the loop; (2) omitting values at internal loop positions (Han, *et. al.*, 2006); (3) or assigning the average value of a 3 nt window (Krol, *et. al*., 2004). The Spearman (ranked) correlation co-efficient (r_s) and P value are indicated (* $P < 0.05$, ** $P < 0.01$).

Thermodynamic profiles were constructed to examine stability of the guide region in more detail. This approach has been used previously to describe a characteristic bias in stability and common thermodynamic features of functional, natural pri-miRNAs (Han et al., 2006; Khvorova et al., 2003; Krol et al., 2004; Warf et al., 2011). The region corresponding to the 5' end of the mature miRNA is expected to be less stable than the region corresponding to its 3' end (Krol et al., 2004) and the position corresponding to the first nucleotide of the guide is expected to be particularly unstable (Han et al., 2006; Khvorova et al., 2003). Profiles were constructed from the 5' end of each guide, as stability features vary with respect to the 5p or 3p orientation of the miRNA (Han et al., 2006). Average thermodynamic profiles were generated for pri-miRNA and

mimic subsets grouped according to target silencing function, a common exogenous guide or a common pri-miRNA scaffold, using approach (1) (Figure 3.9). Common features were identified for pri-miRNA subsets with effective (>80 %), functional (> 50 %) or ineffective (< 50 %) target silencing (Figure 3.9 A), including a decrease in relative stability at positions 4 and 9, as well as an increase in the relative stability at positions 11, 16 and 20. However, the guide region of more effective mimics was generally less stable than that of ineffective mimics. In particular, the free energy at position 1 of the stem, which corresponds to the 5' end of the miRNA, increased with the average target silencing of each subset. In turn, the expected bias in guide region stability, as indicated by the difference in free energy between position 1 and 20 (ddG (1-20)), was highest for the group of effective pri-miRNAs (1.7 kcal / mol) and lowest for the group of ineffective pri-miRNAs (0.3 kcal / mol). A similar profile for the original pri-miRNAs had a ddG (1-20) value slightly lower than expected (0.7 kcal / mol) because of a less favourable value for pri-miR-23a, which demonstrated more moderate function (3p miRNA: 63 %; 5p miRNA: 3 %).

Average thermodynamic profiles were also generated for mimic subsets with a common exogenous guide sequence (Figure 3.9 B). Again, the ddG (1 -20) signature increased with the average target silencing of each mimic subset: anti-*tat* (0.2 kcal / mol), anti-*int* (0.9 kcal / mol), anti-*ga*g (1.3 kcal / mol) and anti-LTR (1.8 kcal / mol). This suggested that inherent properties of the exogenous guide were key in determining a favourable ddG (1-20) signature and subsequently influenced general mimic function. Profiles were also constructed for mimic subsets with the same pri-miRNA scaffold (Figure 3.9 C). A trend between guide region stability and the average target silencing was not observed for each group, but the ddG (1-20) signature was still highest for the most effective scaffold subset (pri-miR34a: 3.3 kcal / mol) and low for the least effective scaffold subset (pri-miR-520b: 0.4 kcal / mol). This suggested that unique properties of a particular scaffold, like the mismatch at position 1 of pri-miR34a, may be more supportive of a bias in guide region stability. Pri-miR-23a is unusual in that the ddG (1-20) value is negative. This is a consequence of the bulge next to position 20 which decreases local stability and implies that the minor 5p miRNA may be favoured in downstream strand selection. However, while miR-23-5p was detected in northern blot analysis, no target silencing was observed for this guide (Figures 2.7, 2.8 & 2.11). This suggested that stable pairing at position 21 of the guide region may compensate for decreased stability at position 20 to facilitate a bias in guide region stability and enable selection of the major, functional 3p miRNA for pri-miR-23a and derived mimics.

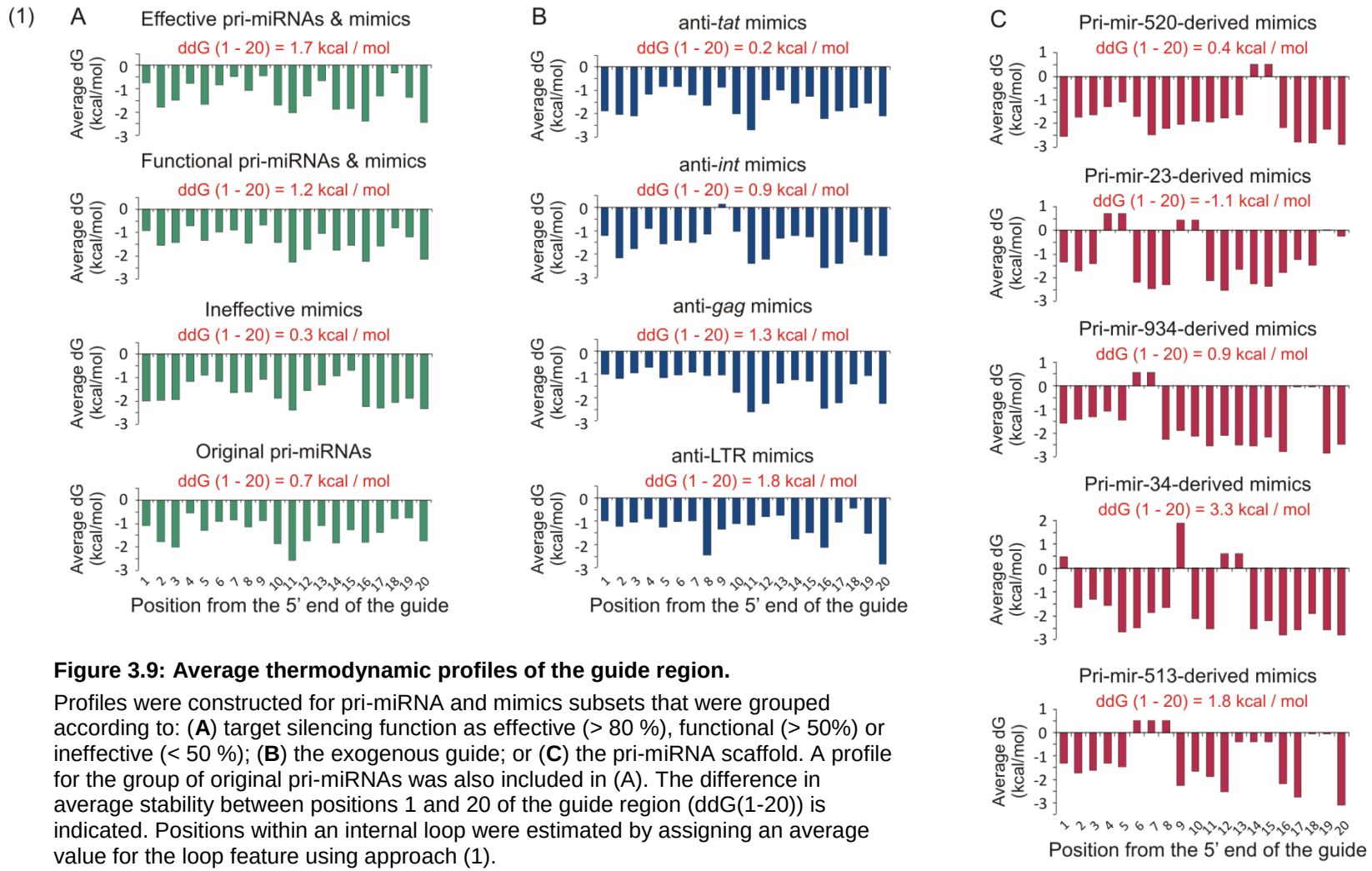


Figure 3.9: Average thermodynamic profiles of the guide region.

Profiles were constructed for pri-miRNA and mimics subsets that were grouped according to: (**A**) target silencing function as effective (> 80 %), functional (> 50%) or ineffective (< 50 %); (**B**) the exogenous guide; or (**C**) the pri-miRNA scaffold. A profile for the group of original pri-miRNAs was also included in (A). The difference in average stability between positions 1 and 20 of the guide region (ddG(1-20)) is indicated. Positions within an internal loop were estimated by assigning an average value for the loop feature using approach (1).

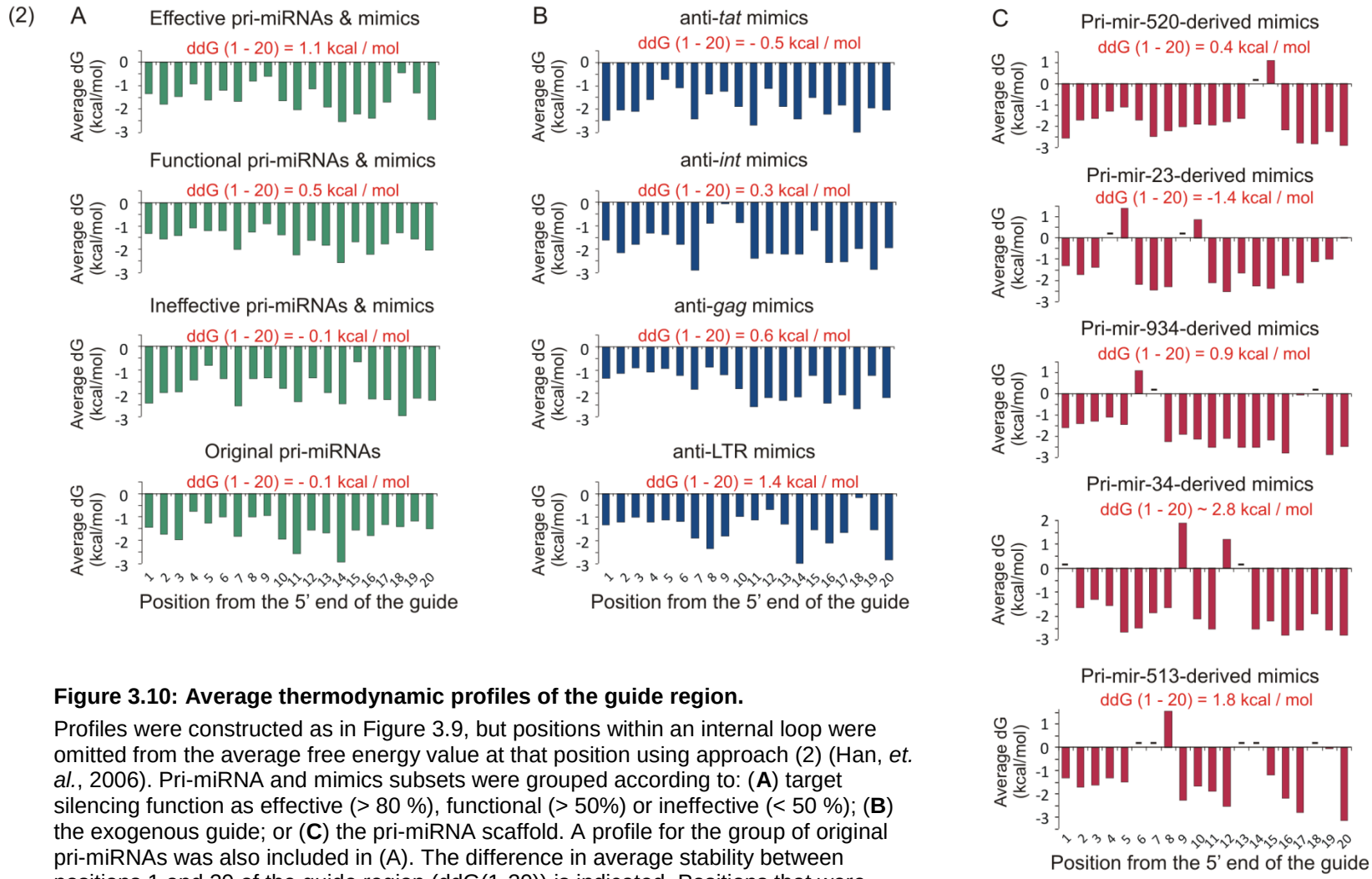


Figure 3.10: Average thermodynamic profiles of the guide region.

Profiles were constructed as in Figure 3.9, but positions within an internal loop were omitted from the average free energy value at that position using approach (2) (Han, *et. al.*, 2006). Pri-miRNA and mimics subsets were grouped according to: (**A**) target silencing function as effective (> 80 %), functional (> 50%) or ineffective (< 50 %); (**B**) the exogenous guide; or (**C**) the pri-miRNA scaffold. A profile for the group of original pri-miRNAs was also included in (A). The difference in average stability between positions 1 and 20 of the guide region (ddG(1-20)) is indicated. Positions that were omitted from average calculations (**-**) are indicated in (C).

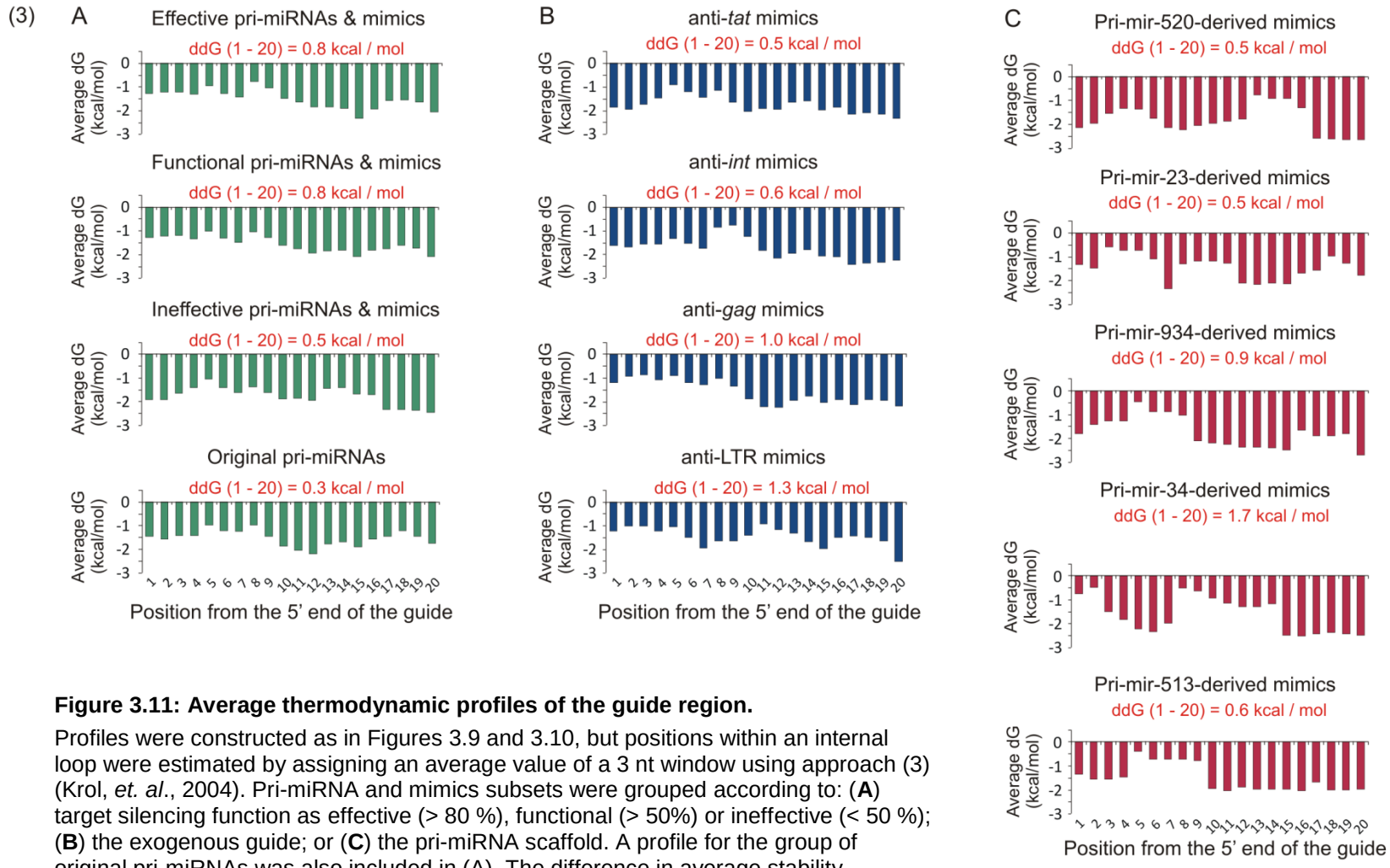


Figure 3.11: Average thermodynamic profiles of the guide region.

Profiles were constructed as in Figures 3.9 and 3.10, but positions within an internal loop were estimated by assigning an average value of a 3 nt window using approach (3) (Krol, *et. al*., 2004). Pri-miRNA and mimics subsets were grouped according to: (**A**) target silencing function as effective (> 80 %), functional (> 50%) or ineffective (< 50 %); (**B**) the exogenous guide; or (**C**) the pri-miRNA scaffold. A profile for the group of original pri-miRNAs was also included in (A). The difference in average stability between positions 1 and 20 of the guide region (ddG(1-20)) is indicated.

Together, these results indicated that mimics with more effective target silencing had a more pronounced bias in guide region stability. This bias appears to be largely defined by the identity and inherent properties of the exogenous guide, but also affected by the unique properties of individual scaffolds. Similar findings were also observed for average thermodynamic profiles constructed using approach (1) and (2) to estimate internal loop stability (Figures 3.10 & 3.11). Additional points of increased stability were observed at positions 7 and 14 using the approach by Han *et. al*. (Figure 3.10 A), while stability profiles were more gradual using the approach by Krol. *et. al.* (Figure 3.11). In both approaches, the ddG(1 -20) signature again increased with average target silencing for pri-miRNA and mimic subsets grouped according to (A) target silencing function or (B) a common exogenous guide; and was highest for the mimics based on the most effective pri-miR-34a scaffold and low for mimics based on the least effective pri-miR-520b scaffold (C). In Figure 3.11 A, the ddG(1-20) value of effective and functional groups was the same (0.8 kcal / mol), but still higher than that of the ineffective group (0.5 kcal / mol). Again, an unusual, negative ddG(1-20) value was observed for the pri-miR23a scaffold subset (Figure 3.10), but was compensated for by stable pairing at position 21, as indicated by a positive ddG(1-20) value when using a 3 nt assessment window (Figure 3.11, 0.5 kcal / mol). Similarly, ddG(1-20) values of the original pri-miRNA and anti-*tat* mimic subsets negatively affected by the pri-miR23a or pri-miR-23-derived mimics in Figure 3.10, were improved in Figure 3.11.

The identity of the exogenous guide appeared to be a major contributing factor in the function of pri-miRNA mimics. However, pairing of the common guide sequence was not necessarily consistent between mimics, as small differences in base pairing were used to maintain the internal mismatches of individual scaffolds. Base pairing profiles were therefore generated to examine base pairing variation in the same pri-miRNA and mimic subsets. The frequency of GC base pairing was lower at the 5’ side of the guide region for effective and functional pri-miRNA and mimic groups, while no GC base pairing was observed in position 1 for these subsets or for the original, functional pri-miRNAs (Figure 3.12 A). In contrast, GC pairing was observed in the first position for the ineffective subset and more frequently at positions 3 and 6. This bias in GC pairing was primarily defined by the identity of the exogenous guide and common base pairing was typically observed for between 3 and 5 mimics within each mimic subset (Figure 3.12 B). The common 5’ to 3’ bias in stable GC pairing therefore supported the observed bias in guide region stability. As expected, this bias was more pronounced for the anti-*gag* and anti-LTR mimic groups and no GC base pairs were observed in the first four positions as a consequence of siRNA design (McIntyre et al., 2009a; Reynolds et al., 2004). The base pairing profiles for mimic subsets with the same pri-miRNA scaffold would be identical, if not for unique features of

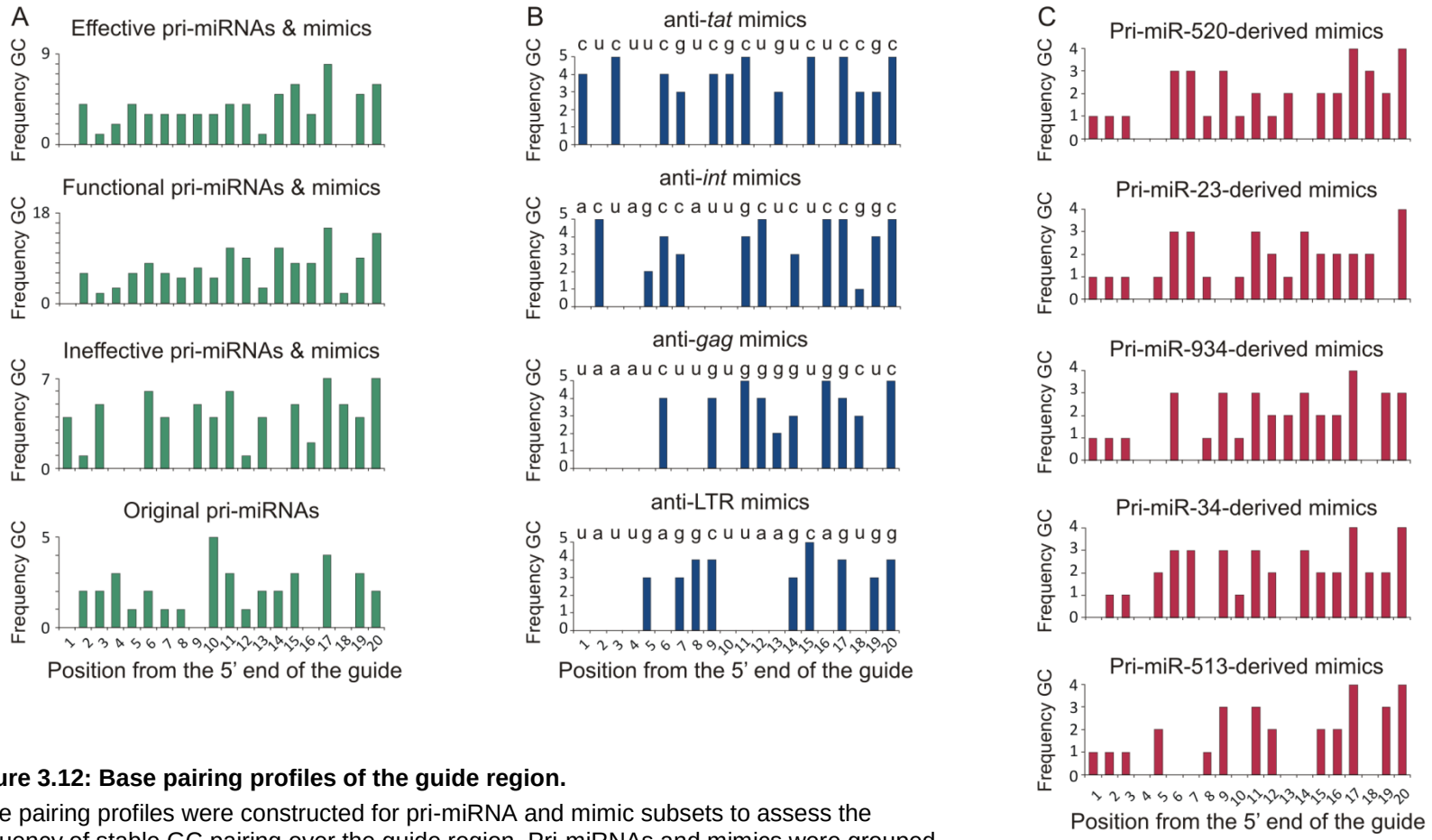


Figure 3.12: Base pairing profiles of the guide region.

Base pairing profiles were constructed for pri-miRNA and mimic subsets to assess the frequency of stable GC pairing over the guide region. Pri-miRNAs and mimics were grouped according to were grouped according to: (**A**) target silencing function as effective (> 80 %), functional (> 50%) or ineffective (< 50 %); (**B**) the exogenous guide; or (**C**) the pri-miRNA scaffold. A profile for the group of original pri-miRNAs was also included in (A).

the guide region present in different scaffolds. Resulting variations in base pairing across the guide region were minor, but the 5' mismatch in the pri-miR-34a scaffold resulted in no GC base pairs at position 1 in derived mimics. This supported the more pronounced stability bias of pri-miR-34-tat compared to other anti-*tat* mimics.

These results suggested that the average bias in guide region stability, indicated by ddG (1-20), was important and may explain the general function of pri-miRNA and mimic subsets. To investigate whether this finding could be applied to individual mimics, the relationship between ddG (1-20) and target silencing was examined for all pri-miRNAs and mimics (Figure 3.13 A). However, no correlation was observed between this measure of the bias in stability and target silencing (r_s: 0.28). This might be expected as a result of unique scaffold features (e.g. pri-miR-23a) and guide-scaffold combinations. In fact, average thermodynamic profiling is typically adopted to identify common features from a population of highly variable pri-miRNA structures. Nonetheless, the ddG(1-20) was positive for 8 out of 9 effective pri-miRNAs and mimics with target silencing of > 80 %; and for 14 out of 18 functional pri-miRNAs and mimics with target silencing of > 50 %. Functional mimics with a negative ddG(1-20) value included pri-miR-23-int and pri-miR-23-gag, which had a similar value to that of the original pri-miRNA (- 2.1 kcal / mol). In contrast, pri-miR-23-LTR had a positive ddG (1-20) value and improved target silencing (86 %) compared to the original pri-miRNA (63 %), which was facilitated by moving the bulge at position 19 to position 17 with a subsequent increase in the stability of position 20. This suggested that an improvement of the guide region bias, even through the use of non-natural features, may be beneficial in mimic design. Unlike most effective pri-miRNAs and mimics, the ddG(1-20) value of pri-miR-934-int was low (-0.1 kcal / mol), as a result of omitting a G-U wobble pair at position 1 or 2 and including a less stable base pair at position 21. The effective function of pri-miR-934-int, however, suggested that these modifications were not deleterious and a positive ddG (1-20) value obtained using a 3 nt window of assessment implied that the stability of neighbouring residues may be compensatory for this mimic. An increase in ddG(1-20) appeared particularly beneficial for anti-*tat* mimics. The relationship between the change in ddG (1-20) and the change in target silencing was also examined for individual mimics (Figure 3.13 B). A weak positive correlation (r_s = 0.48) was observed, suggesting that the change in the bias of the guide region incurred in mimic design might also contribute to the observed change in mimic function. A zero or positive ddG(1-20) value was observed for 7 out of 9 mimics with a functional decrease of up to 10 % compared to the original pri-miRNA; and for 12 out of 15 mimics with a functional decrease of up to 50 %. In contrast, most of the mimics with a functional decrease of more than 50 % had a negative ddG(1-20) value associated with the tat guide sequence.

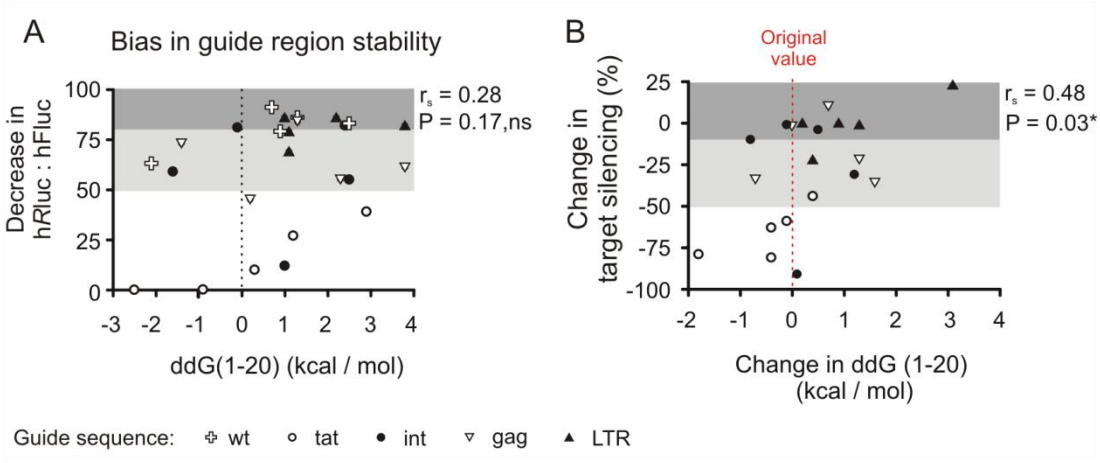


Figure 3.13: Guide region stability of individual pri-miRNAs and mimics.

(**A**) The bias in guide region stability (ddG(1-20)) was plotted against target silencing for pri-miRNAs and mimics. Shaded regions indicate pri-miRNAs and mimics with effective (> 80 %, dark grey) or basic (>50 %, grey) target silencing function. (**B**) The change in the bias of guide region stability incurred in mimic design was plotted against the corresponding change in target silencing function. Shaded regions indicate mimics with a decrease in function of no more than 10 % (dark grey) or 50 % (grey) when compared to the original pri-miRNA. Values in (A) and (B) were calculated by assigning an average loop value to internal loop positions using approach (1). The Spearman (ranked) correlation co-efficient (r_s) and P value are given (* $P < 0.05$); ns: not significant.

Overall, these results indicated that a pronounced, positive bias in guide region stability appears to be favourable for mimic function. While a particular ddG (1-20) value can not be used to predict or guarantee the precise function of a particular mimic in the context of various pri-miRNA scaffolds, a positive bias in guide region stability was observed for most functional and effective mimics. The bias in guide region stability appears to be largely defined by the sequence of the exogenous guide, but is also affected by the unique properties of individual scaffolds and neighbouring positions within the stem. The contribution of both the exogenous guide sequence and pri-miRNA scaffold to guide region stability should therefore be considered in mimic design.

3.3.6 Stability of expected miRNA duplexes

Thermodynamic profiling suggested that a bias in stability of the pri-miRNA guide region may explain the general function of pri-miRNA mimics. To determine whether a similar bias in stability could be discerned directly for processed miRNA duplexes, conceptual dicing (Schwarz et al., 2003) was applied to the expected duplexes using the approach by Krol *et. al.* (Krol et al., 2004). Again, in most cases, the 5' end of the miRNA is expected to be less stable, affecting strand selection and RISC loading (Gregory et al., 2005; Khvorova et al., 2003; Schwarz et al., 2003). The average terminus stability was examined for pri-miRNA and mimic subsets with effective (>80 %), functional (> 50 %) or ineffective (< 50 %) target silencing using a 1 nt, 2 nt or 3 nt window (Figure 3.14 A). As expected, the average stability was lower for the terminus corresponding to the 5' end of the major miRNA (red) than that of the minor miRNA (blue). However, in contrast to the results of thermodynamic profiling, the difference in the average terminus stability (ddG(termini)) did not increase with the average target silencing of each group. The ddG(termini) values were similar when using a 1 nt window, but actually higher for the group of ineffective pri-miRNAs and mimics when using a 2 nt or 3 nt window. This does not support the optimal selection of the intended miRNA in effective or functional groups, but the increased stability of both termini for the ineffective group of precursors may also be unfavourable. The average terminus stability was also examined for mimic subsets with the same exogenous guide (Figure 3.14 B). Again, the ddG(termini) values did not necessarily increase with average target silencing, but the stability of the miRNA 5' terminus was higher for the less effective mimic subset. The ddG(termini) did increase with average target silencing for the anti-*tat*, anti-*int* and anti-*gag* mimic groups using a 2 nt or 3 nt window, but ddG(termini) of the anti-LTR mimic subset was expected to be higher than that of the anti-*gag* mimic subset.

Differences in the order of magnitude of the stability bias were therefore evident in the results of thermodynamic profiling and conceptual dicing. These differences in the conceptual dicing results may have originated from the inclusion of extra residues in the assessment of terminus stability using 2 nt and 3 nt windows. For example, the ddG (termini) values were higher than expected for duplexes derived from pri-miR-520-gag and pri-miR-34-gag when compared to the respective anti-LTR mimics. The inclusion of additional residues indicated more stable pairing near the 3' terminus (~position 20) for these anti-*gag* mimics, resulting in a larger bias in duplex stability for the anti-*gag* subset. The bias in terminus stability was also lower than expected for the group of original pri-miRNAs using a 2 nt and 3 nt window, as stable base pairs near the 5' terminus of the miR-34a duplex and a mismatch near the 3' terminus of the miR-934 duplex

affected terminus stability values. This was surprising as the 2 nt window was previously described as the most reliable for miRNA discrimination for other pri-miRNAs (Krol et al., 2004).

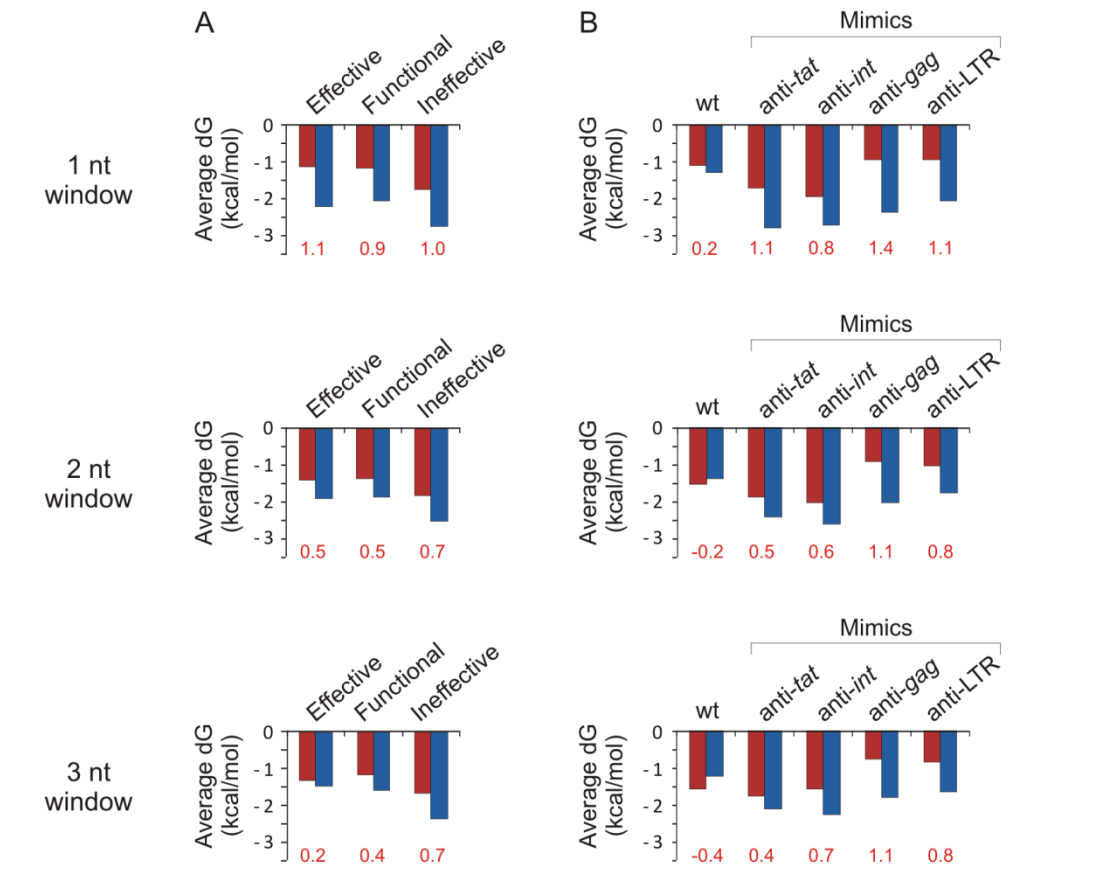


Figure 3.14: Terminus stability of expected miRNA duplexes.

Average terminus stability values were calculated for duplexes derived from conceptual dicing of pri-miRNA and mimic subsets using a 1, 2 or 3 nt window. (**A**) The average terminus stability of duplexes from pri-miRNAs and mimics with effective (> 80 %), functional (> 50%) or ineffective (< 50 %) target silencing. Average stability for the 5' terminus of the major miRNA (red) and minor miRNA (blue) are shown. The change in average terminus stability (ddG(termini)) is indicated in red. (**B**) The average terminus stability of duplexes from different mimic subsets and the group of original, wildtype (wt) pri-miRNAs.

Differences in the conceptual dicing results may also have originated from the nature of terminus dynamics compared to inner stem dynamics. For example, the 5’ residue of miR-34a-5p has a lower free energy value as a terminal mismatch of the processed RNA duplex (- 0.8 kcal / mol), compared to its value within the internal loop of the pri-miRNA stem (0.4 kcal / mol, approach 1). The positioning of an internal bulge can also affect inner stem stability and terminus stability differently, as seen in the bias in stability of the miR-23-gag duplex which was larger than expected compared to the respective LTR duplex. The unmatched residue at position 19 of pri-miR-23-gag decreased 3’ stability of the pri-miRNA guide region (0.25 kcal / mol), but not the processed duplex (- 2.5 kcal / mol); resulting in a larger bias for the processed duplex (ddG(termini), 1 nt window: 1.2 kcal / mol) than for the pri-miRNA guide region (ddG(1-20): - 1.4 kcal / mol). In contrast, the unmatched residue at position 17 of pri-miR-23-LTR did not affect 3’ stability of the pri-miRNA guide region (- 2.1 kcal / mol), but disrupted 3’ pairing and stability of the processed duplex (- 0.6 kcal / mol); resulting in a smaller bias for the processed duplex (ddG(termini), 1 nt window: - 0.7 kcal / mol) than for the pri-miRNA guide region (ddG(1-20): 1.0 kcal / mol). Overall, the relationship between the magnitude of thermodynamic asymmetry and average target silencing was not as clear when examining the terminus stability of the anticipated miRNA duplexes in conceptual dicing using mfold. As 5’ pairing of the major and minor miRNAs can be more difficult to assess in this approach, thermodynamic profiling may provide a more reliable indication of relative pairing stability of terminal residues. In both methods, the stability of the 5’ terminus was generally lower for pri-miRNA and mimic subsets with more effective target silencing, again suggesting the importance of this feature for miRNA function.

3.4 DISCUSSION

Several effective pri-miRNA scaffolds with varied basal stem lengths (9 -13 bp) were used in the design of corresponding pri-miRNA mimics. In contrast to the original pri-miRNAs, RNA guides were not effectively derived from all pri-miRNA mimics in functional characterisation studies. Effective target silencing and processing of RNA guides was observed more frequently for mimics with particular scaffolds or exogenous guide sequences. Mimics based on the pri-miR-34a scaffold were generally more effective with average target silencing of 70 %, while those based on the pri-miR-520b scaffold were generally less effective with average target silencing of 34 %. Mimics containing the anti-LTR guide sequence were also generally more effective (70 - 86 %), while those containing the anti-*tat* guide sequence functioned poorly or not at all (0 - 56 %). These differences in mimic function could not be explained by pri-miRNA sequence motifs, which were present in the original, effective pri-miRNAs and were unaltered in mimic design. Similarly, the form of the pri-miRNA hairpin was not appreciably altered in the predicted secondary RNA structures of mimics. However, more subtle features of the guide region may explain the general function of pri-miRNA mimics.

The results of thermodynamic profiling showed that the bias in guide region stability, indicated by the ddG (1-20) signature, increased with the average target silencing of pri-miRNA and mimic subsets. This trend could be attributed to the free energy of the first position of the guide region, which also increased with the average target silencing of pri-miRNA and mimic subsets. Stability of the guide region appeared to be largely defined by the sequence and inherent properties of the exogenous guide, but was also affected by the unique features of individual scaffolds. A positive bias in guide region stability was observed for most functional and effective mimics, but a strong correlation was not observed between the ddG (1-20) value and target silencing of individual mimics. This suggested that while a bias in guide region stability appeared to be an important property for mimic function, it could not be used to predict precise mimic function across different pri-miRNA scaffolds. The relationship between the magnitude of thermodynamic asymmetry and average target silencing was not as clearly reflected in the terminus stability of corresponding miRNA duplexes assessed by conceptual dicing. This suggested that thermodynamic profiling may be a more reliable approach to assess the relative 5' stability of the expected major and minor miRNAs. However, in results generated by both approaches, the average free energy of the 5' terminus was typically higher for pri-miRNA and mimic subsets with more effective target silencing. Small RNA sequencing indicated that processing of selected pri-miRNAs and their respective anti-*int* mimics was consistent with the generation of 22 nt therapeutic guides, while minor sequence heterogeneity was commonly the result of typical 3'

RNA modifications. Overall, these results demonstrated that exogenous, therapeutic guides could be successfully and consistently derived from mimics with a range of basal stem lengths and motif combinations, but a pronounced thermodynamic asymmetry was commonly observed for more effective mimics and appears to be a priority in mimic design.

Potential consequences of thermodynamic asymmetry

A pronounced, positive bias in guide region stability may be advantageous to mimic function by affecting several downstream processing events in the RNAi pathway. Properties of the pri-miRNA also determine properties of the derived pre-miRNA following Drosha-DGCR8 processing and the derived miRNA duplex after subsequent Dicer processing. In the thermodynamic profiling approach used here, a comparison of stability at position 1 and 20 of the guide region corresponds to a comparison of 5' stability for standard 22 nt major and minor miRNAs. The strand of the miRNA duplex that has a less stable 5' region is expected to function as the mature miRNA in the miRNA-induced silencing complex (miRISC) complex, while the passenger strand is degraded (Khvorova et al., 2003; Schwarz et al., 2003). The major functional significance of a positive bias in guide region stability in these results was therefore expected to be the favoured selection and function of the intended miRNA. This can explain why properties of the exogenous guide sequence and the guide region of the pri-miRNA scaffold that supported decreased stability at position 1 and a more positive ddG (1-20) value were observed more frequently for effective mimics. Thermodynamic stability at the 5' ends of the miRNA duplex is also affected by pairing over the first five bases at each terminus (Khvorova et al., 2003; Schwarz et al., 2003). Selection of the intended guide was therefore supported by the absence of GC pairing in the first four or five positions of anti-*gag* and anti-LTR miRNAs. These guides were originally designed with respect for siRNA design principles (McIntyre et al., 2009a; Reynolds et al., 2004) and were generally more effective in the context of a pri-miRNA scaffold than those selected solely on the basis of target site accessibility (anti-*tat*) (Lee et al., 2002a) or conservation (anti-*int*) (Nishitsuji et al., 2006). Reliable strand selection is particularly important in a therapeutic setting to avoid off target effects of the passenger strand (Aagaard et al., 2008; Khvorova et al., 2003; Schwarz et al., 2003).

Structural features of the guide region

In addition to enhancing thermodynamic asymmetry of the processed miRNA duplex, flexibility at position 1 of the guide region may also indicate enhanced flexibility at Drosha-DGCR8 or Dicer cleavage sites. Structural distortions are over-represented at positions adjacent to Drosha and Dicer cleavage sites in *C. elegans* (Warf et al., 2011) and deformable regions adjacent to the Drosha cleavage site are required for the efficient processing of human pri-miR-16-1, pri-miR-30a and pri-miR-107 (Quarles et al., 2013). This may explain why a scaffold like pri-miR-34a, with a double mismatch at the Drosha cleavage site was generally more effective than scaffolds with more moderate features near this site, like single symmetrical mismatches (pri-miR-520b & 23a) or G-U wobble pairs (pri-miR-934 & 513a-2), particularly when the 5' sequence and stability of an exogenous guide were sub-optimal. This suggests a thermodynamic requirement for Drosha-DGCR8 in addition to basic structural elements and recognition motifs. DGCR8 can not sense deformable regions adjacent to the cleavage site (Quarles et al., 2015), suggesting that any thermodynamic requirement of the microprocessor would involve Drosha. These results are also in agreement with previous proposals for siRNA and mimic design that suggested a mismatch should be included at the first position of a 5p guide (Han et al., 2006; Schwarz et al., 2003).

Other structural features may also contribute to the efficiency of processing and strand selection. Common structural preferences of the miRNA duplex have been identified for all four human Argonaute proteins (Ago 1 - 4) (Yoda et al., 2010). In addition to a 5' mismatch at position 1 to promote thermodynamic asymmetry, single mismatches in the central region (position 8 - 11) of the miRNA duplex promote RISC loading to form pre-RISC, while single mismatches in the seed region (position 2 - 7) and to a lesser degree, the 3' mid-region (position 12 - 15), promote unwinding of the duplex to form mature RISC. Single mismatches in both the seed and central regions therefore promote efficient RISC function and may contribute to the observed function of pri-miR-23a and derived mimics (mismatches at 4 and 9), especially as thermodynamic asymmetry between positions 1 and 20 was unfavourable. A mismatch at position 4 has also been shown to enhance strand selection and target cleavage by an Ago2 complex (Noland and Doudna, 2013). Asymmetrical mismatches in the central and 3' mid-regions of pri-miR-34a and derived mimics may also facilitate RISC function. However, as these mismatches are found in both the original pri-miRNAs and derived mimics which have varied efficacy, other properties like thermodynamic asymmetry appear to be of greater consequence in mimic function here. In addition to mismatches, decreased stability at position 12 of the guide region was commonly observed for natural pri-miRNAs and may suppress abortive processing where Drosha-DGCR8

cleaves the stem at ~ 11 bp from the terminal loop and fails to generate the intended miRNA (Han et al., 2006). This idea may be supported by the absence of GC pairing at position 12 for anti-LTR mimics and structural distortions adjacent to this position in the pri-miR-34a scaffold.

Sequence of the RNA guide

A U in position 1 of the guide is common for natural miRNA sequences (Krol et al., 2004; Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001; Warf et al., 2011; Zhang et al., 2009) and is particularly useful in mimic design to enable weaker 5' U-A or 5' U-G base pairs for thermodynamic asymmetry. However, there are other functional implications for this feature. Drosha and Dicer may have a preference for a U in the first position of the miRNA (Starega-Roslan et al., 2015b), while Dicer may specifically select cleavage positions that generate a 5' U and not a 5' G for a 3p miRNA (Starega-Roslan et al., 2015a). The human Argonaute 2 (Ago2) protein has a preference for miRNAs with a 5' U or A which make specific contacts with the middle (MID) domain (Frank et al., 2010). The binding affinity for a 5' UMP was also found to be around 2-fold higher than that of a 5' AMP and around 30-fold higher than that of a 5' CMP or 5' GMP. As Ago2 is a core component of miRISC (Gregory et al., 2005), this implies that miRNAs with a 5' U or A are loaded more efficiently into the complex. The identity of the 5′-terminal nucleotide of the exogenous guide was therefore expected to influence the efficiency of target silencing and was particularly favourable for anti-*gag* and anti-LTR mimics. The mechanism of target silencing in dual luciferase assays was not investigated further, but cleavage of the target mRNA by an Ago2-containing complex (Meister et al., 2004) was expected to be the major mechanism here as a consequence of near-perfect and central complementarity between exogenous miRNAs and target sequences (Hutvagner and Zamore, 2002; Zeng et al., 2003). The observed target silencing may also involve a minor translational repression effect mediated by miRISC containing Ago2 or other members of the human Argonaute family (Ago1, 3 or 4), as all four proteins are expressed in HEK293 cells (Meister et al., 2004); miRNAs have been shown to associate with all four Argonaute proteins to mediate translational repression (Meister et al., 2004; Pillai et al., 2004); and general sorting of miRNAs between different Ago proteins according to their sequence has not been demonstrated in humans (Burroughs et al., 2011; Dueck et al., 2012). Although not yet described, a similar 5' nucleotide preference may also exist for other members of the conserved Argonaute family (Carmell et al., 2002) and could contribute to the preferential function of certain miRNA sequences in a minor translational repression effect here.

Poor performance of anti-tat mimics

The general function of pri-miRNA mimics can be explained with respect to the sequence and stability of the exogenous guide, as well as properties of the pri-miRNA scaffold, in agreement with a number of previous publications. The poor performance of the anti-*tat* mimic set might have therefore been expected, but was initially surprising as an effective anti-*tat* mimic was previously generated by another group to give target silencing of ~ 75 % in the same dual luciferase assay system (Aagaard et al., 2008). In addition, this previously described anti-*tat* mimic was included in several effective combinatorial RNAi constructs that inhibited HIV replication in long-term challenge assays (Aagaard et al., 2008; Chung et al., 2014; Chung et al., 2012). However, in contrast to anti-*tat* mimics designed here, Aagaard *et. al*. positioned the anti-*tat* / *rev* guide sequence (Lee et al., 2002a) within the 3p arm of the miR-106b scaffold so that the processed 21 nt guide started with a 5' U and ended with an additional 3' GC. These differences are expected to enhance Ago2 binding affinity and thermodynamic asymmetry, thus enhancing the efficiency of RISC loading, strand selection and target silencing. As the tat guide sequence used here (5' C) mediated potent target silencing when derived from the pol III shRNA positive control (sh-tat) (Saayman et al., 2008), which was also expected to direct target cleavage through Ago2 (Wang et al., 2009a; Yoda et al., 2010), it seems that sub-optimal properties of a processed RNA duplex may be more evident in a less robust pol II expression system.

Desirable features for effective pri-miRNA mimics

Together, these results highlight the importance of a careful approach to mimic design and suggest that, while individual features may fine-tune mimic function, thermodynamic asymmetry of the guide region is important. In addition to basic elements of the pri-miRNA secondary structure and processing motifs, features that promote guide region asymmetry should be considered in the selection of an exogenous guide and pri-miRNA scaffold in order to design effective pri-miRNA mimics (Figure 3.15 A).

The following is an updated list of desirable features for individual pri-miRNA mimics, including previously proposed features and design principles: (**1**) a 22 nt exogenous guide that begins with a 5' U, lacks G or C bases in the first four positions and has a pronounced bias in GC pairing over the first 20 nt, in agreement with siRNA design criteria (Amarzguioui and Prydz, 2004; Hsieh et al., 2004; Reynolds et al., 2004; Ui-Tei et al., 2004); (**2**) flexible pairing at the Drosha-

DGCR8 cleavage site, which may include the first position of the guide (P1) (Han et al., 2006; Schwarz et al., 2003) and the preceding position (-1), as observed for pri-miR-34a; (**3**) a 5p exogenous guide where the 5' end is defined by Drosha processing and is less susceptible to variation in the Dicer cleavage site (Zhou et al., 2012) with subsequent changes in the identity of the 5' nucleotide and the "seed" sequence; (**4**) a pri-miRNA scaffold predicted to have an optimal total stem length of 35 ± 1 bp (Fang and Bartel, 2015) which may include a basal stem length of 9 - 13 bp, counted to include G-U wobble pairs and single symmetrical mismatches; (**5**) a terminal loop predicted to include 8 - 11 unpaired nucleotides, similar to a previous suggestion of ≥ 10 nt (Zeng et al., 2005b); (**6**) 40 - 50 nt of original flanking sequences on either side of the pre-miRNA, in agreement with a previously described requirement for at least 40 nt of flanking sequences (Chen et al., 2004); (**7**) no large bulges in the pri-miRNA stem (pri-miR-198 and 147a), which have recently been described as structural defects that negatively affect processing (Fang and Bartel, 2015); (**8**) all original sequence motifs of an effective scaffold, including the basal UG, basal CNNC and terminal UGU loop motifs (Auyeung et al., 2013) and a mismatched GHG stem motif (Fang and Bartel, 2015); (**9**) METTL3 methylation motifs (Alarcon et al., 2015b) within 25 - 100 nt of the pre-miRNA hairpin (pri-miR-34a); and (**10**) expression from an inducible pol II promoter (Zeng et al., 2005a). Pri-miRNA mimics designed in accordance with these features are expected to have enhanced processing fidelity, strand selection and target specificity and efficient Drosha-DGCR8 processing and Ago2 binding; whilst avoiding issues of pathway saturation, cellular toxicity and off-target effects. In accordance with these features, a pri-miRNA scaffold like pri- miR-34a was particularly well suited for use in mimic design. These features should also be included for the individual pri-miRNA mimics of a polycistronic therapeutic construct (Figure 3.15 B).

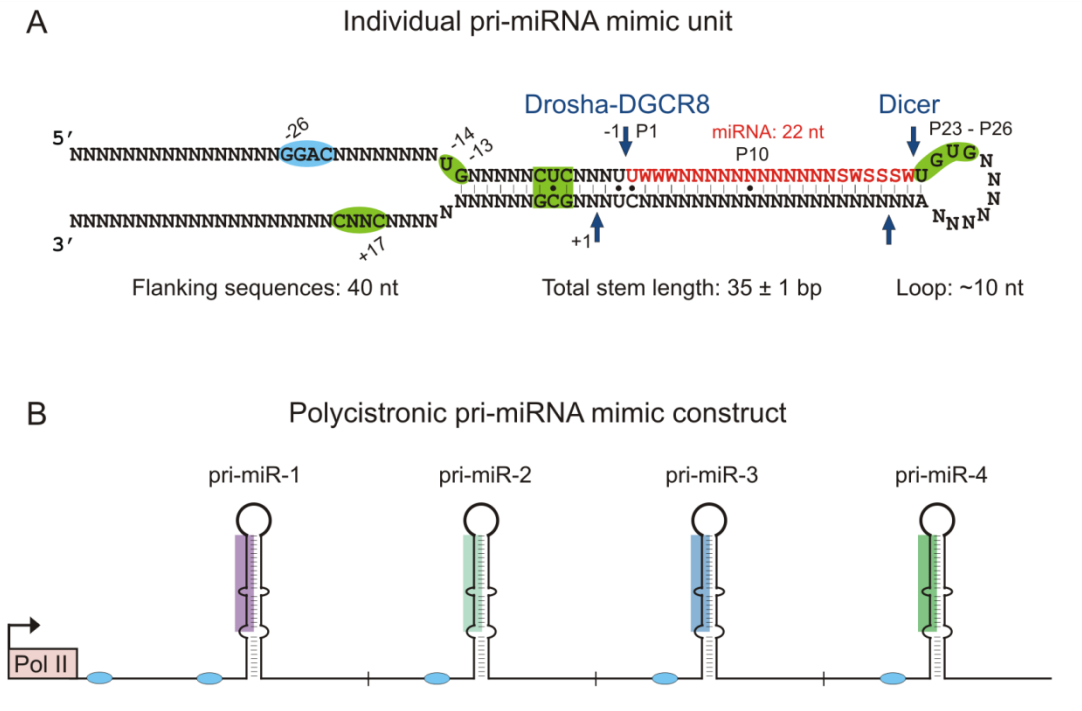


Figure 3.15: Desirable features for pri-miRNA mimics and therapeutic constructs.

(**A**) An ideal pri-miRNA mimic with optimal sequence and structure features. Ideal features of the predicted secondary RNA structure include 40 nt of sequence flanking either side of the pre-miRNA, a total stem length of 35 ± 1 bp with a basal stem of ~ 11 bp and a large loop with ~ 10 unstructured nucleotides. Sequence features include basal, stem and loop motifs (green) and methylation motifs in the vicinity of the pre-miRNA (blue). Ideal sequence and pairing properties of the 22 nt 5p miRNA (red) are indicated, including a 5' U, 5' and 3' pairing preferences (S:G or C; W: A or U) and mismatches (•) at positions -1, 1 and 10. The expected Drosha-DGCR8 and Dicer cleavage sites are indicated (blue arrows). This figure was adapted from (Fang et al., 2015) to include a methylation motif from (Alarcon et al., 2015b). (**B**) A schematic illustrating an ideal polycistronic pri-miRNA construct with at least four different pri-miRNA scaffolds (1 - 4), each with a unique exogenous guide sequence (highlighted). Ideal properties of individual pri-miRNA units are maintained and expression is driven by an inducible pol II promoter. Additional METTL3 methylation motifs (blue) can be added to the expressed sequence.

In conclusion, natural pri-miRNA scaffolds with various basal stem conformations and motif combinations were successfully used to generate functional pri-miRNA mimics. However, in contrast to the original pri-miRNAs, not all derived mimics mediated effective target silencing. The general function of pri-miRNA mimics appeared to be affected by features of both the exogenous guide sequence and pri-miRNA scaffold, in agreement with several previous and concurrent findings. Features that contributed to a decreased 5' stability of the intended miRNA appeared to be particularly beneficial. In combination with the 5' nucleotide preference of Ago2, this suggested that the major factor underlying mimic function here was the favoured selection and function of the intended therapeutic guide. However, several stability features of the pri-miRNA may also contribute to the efficiency of Drosha-DGCR8 processing. Processing of RNA guides from pri-miRNAs therefore appears to be affected by several sequence and stability features. An updated set of desirable features and rules was proposed for the design of pri-miRNA mimics with effective processing at several steps in the mammalian RNAi pathway. This study has therefore provided insight for the design of effective pri-miRNA mimics and has identified several novel scaffolds suitable for use in therapeutic constructs. Several promising pri-miRNA mimics were also generated here and are ideal for future use and testing in gene silencing applications against HIV.

CHAPTER FOUR

DISCUSSION

4.1 Summary of results

The major aim of this project was to explore the use of natural pri-miRNA scaffolds with exceptional basal stem lengths in mimic design; and from this, to determine whether a particular basal stem length was beneficial to mimic function or of use in the optimisation of therapeutic mimics. Natural pri-miRNAs with varied basal stem lengths were therefore used as scaffolds in the design of corresponding pri-miRNA mimics and both sets of constructs were characterised in functional studies. Functional miRNAs were effectively derived from natural pri-miRNAs predicted to form canonical hairpins with minor variations in the basal stem length (9 - 13 bp). However, pri-miRNAs predicted to form non-canonical basal stems containing a short duplex region (5 - 7 bp) and large internal bulges were not functional. In the artificial expression context, disruptive interactions predicted to occur between the basal stem and flanks of the expressed sequence had no appreciable effect on pri-miRNA function. A subset of potential processing sequence motifs was identified for each pri-miRNA, but the presence of a particular motif did not dictate function. The characterisation of natural pri-miRNAs suggested that canonical hairpins covering a range of sub-optimal basal stem lengths and a variety of sequence motifs could be used as scaffolds in the design of effective pri-miRNA mimics. However, in contrast to the original pri-miRNAs, not all corresponding pri-miRNA mimics were effective or even functional. Effective target silencing and processing of RNA guides was observed more frequently for mimics with particular scaffolds or exogenous guide sequences. Differences in mimic function could not be attributed to changes in the overall form of the predicted secondary RNA structure or sequence motifs which were unaltered in mimic design; but may be explained by the sequence and stability of the guide region which can affect both pri-miRNA processing and guide strand selection. In particular, a 5' U and increased free energy at the first position of the miRNA were commonly observed for more effective mimics, along with a corresponding increase in thermodynamic asymmetry of the guide region. Together, these results were used to propose a more comprehensive, updated set of rules for the design of effective pri-miRNA mimics based on natural scaffolds. Predicted stem length is an important consideration in mimic design, but a basal stem of 11 bp is not an absolute requirement and scaffolds with a small range of similar basal stem lengths (9 - 13 bp) were used successfully. However, a recent report has suggested that the total stem length may be a more important consideration. Other features

of the stem region also appear to contribute to mimic function, including the sequence and stability over the guide region and internal bulges. This work has contributed to the current understanding of pri-miRNA processing and expanded the current repertoire of natural pri-miRNA scaffolds suitable for use in future therapeutic applications.

4.2 The current model of pri-miRNA processing by Drosha-DGCR8

A comprehensive model of pri-miRNA processing is now apparent from several recent biochemical, sequencing and structural studies (Figure 4.1). Biochemical analysis of the microprocessor revealed that it is a ~ 364 kDa heterotrimeric complex composed of a Drosha monomer and a DGCR8 dimer (Nguyen et al., 2015). The complex is likely to have an elongated shape and span the full length of the pri-miRNA (Nguyen et al., 2015). Preferred sequence motifs were identified at the apical and basal junctions of the pri-miRNA using a high-throughput sequencing strategy (Auyeung et al., 2013), suggesting a "unifying model" in which sequence motifs further distinguish the apical and basal junctions to enable the correct orientation of the microprocessor (Fang and Bartel, 2015). Characterisation of the microprocessor revealed that it can indeed recognise sequence motifs to facilitate processing (Nguyen et al., 2015). In productive processing, Drosha specifically binds to the ssRNA-dsRNA basal junction and interacts with the basal UG motif; and the DGCR8 dimer specifically binds to the apical stem and junction and recognises the apical UGU motif. The current model suggests that the microprocessor functions as "molecular callipers" to measure the full length of the pri-miRNA stem and has a narrow preference for a total stem length of 35 ± 1 bp (Fang and Bartel, 2015), including a basal stem length of ~ 11bp (Auyeung et al., 2013). The stem length preference includes wobble pairs and mismatches, but not bulges. Drosha functions as a "molecular ruler" to cleave the pri-miRNA stem at ~ 11 bp from the basal junction (Nguyen et al., 2015). This role was supported by the crystal structure of Drosha which indicated that the bump helix, MB helix, platform and PAZ-like domains of Drosha are positioned to allow for the 11 bp measurement and work together to recognise the basal junction and ssRNA flanking sequences (Kwon et al., 2015). In particular, the distance between the bump helix and the RIIIDa catalytic site is ~ 28 Å which corresponds to ~ 11 bp.

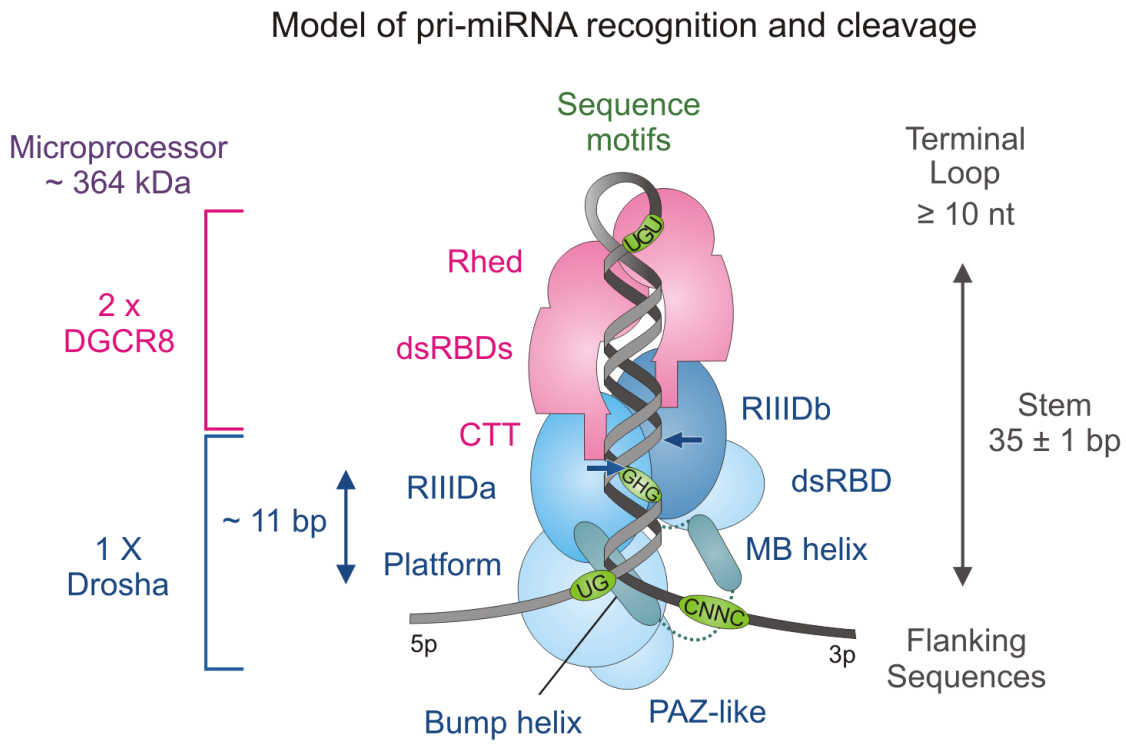


Figure 4.1: A comprehensive model of pri-miRNA recognition and cleavage.

This current model was compiled from several recent reports and features a heterotrimeric microprocessor and optimal pri-miRNA substrate. Drosha functions as a “molecular ruler” to cleave the pri-miRNA at ~ 11 bp from the basal junction. Basal and apical pri-miRNA sequence motifs enhance productive processing. A large terminal loop and total stem length of 35 ± 1 bp are preferred pri-miRNA features. Each DGCR8 monomer consists of an RNA-binding haem domain (Rhed), double-stranded RNA-binding domain (dsRBD) and a C terminal tail (CTT). Drosha is composed of two RNAse III domains (RIIIDa, RIIIDb), a dsRBD and a central domain (CED) composed of a “connector helix”, PAZ-like sub-domain and N terminal region known as the “platform”. Each RIIID has a DGCR8 binding site and interacts with the CTT of a DGCR8 monomer, likely staggering the DGCR8 monomers by a half-helical turn on the upper pri-miRNA stem. The RIIIDs form a single catalytic core and each domain cleaves one strand of the pri-miRNA (RIIIDa: 3p; RIIIDb: 5p). RIIIDa contains a unique “bump helix” and “mobile-basic” (MB) helix. The microprocessor composition was taken from models by (Nguyen et al., 2015) and (Kwon et al, 2015), while motifs and the optimal pri-miRNA stem length were taken from models by (Auyeung et al., 2013) and (Fang and Bartel, 2015).

In light of the specific stem length requirements of the microprocessor, the comparable and effective function of natural pri-miRNAs with a range of basal stem lengths (9 - 13 bp) was initially surprising. This raised the question of how these various pri-miRNA substrates are accommodated by the microprocessor. Most of these pri-miRNAs (pri-miR-23a, -934, -34a and -513a-2) had an optimal total stem length of 35 ± 1 bp, suggesting that they could be accommodated by the microprocessor to enable effective processing. Pri-miR-520b, on the other hand, had a non-optimal basal stem length (9 bp) and a non-optimal total stem length (32 bp), but still functioned effectively. This implied that the Drosha molecular ruler is able to tolerate basal stem lengths of up to 11 ± 2 bp. However, this may not be necessary as transient pairing or unpairing may occur upon binding to satisfy the narrow stem length preference (Fang and Bartel, 2015). An additional 2 bp may occur transiently upon binding for pri-miR-520b, increasing the basal stem length to 11 bp and the total stem length to 34 bp. The stooped conformation of Drosha and the position of the bump helix suggest that the dsRNA basal stem can only be accommodated up to ~ 11 bp, after which the stem should bifurcate into ssRNA (Kwon et al., 2015). This can explain why pri-miRNAs predicted to have longer basal stem lengths may still be processed effectively. For example, a bulge at the end of the 13 bp basal stem of pri-miR-23a may facilitate bifurcation at this point. This would result in a basal stem of exactly 11 bp, while the total stem length remains optimal at 34 bp. Similarly, disruption of the final base pair in the 12 bp basal stem of pri-miR-934 would create a basal stem of exactly 11 bp and a total stem length of 35 bp. Bifurcation of the stem could also explain why longer stem conformations predicted to occur in the expressed sequence do not prevent processing. For example, when additional flanking sequence was considered for pri-miR-34a, basal stem lengths of up to 22 bp were possible, but this pri-miRNA functioned effectively. However, exceptional interactions of the extended sequence are also likely to be more transient.

All four sequence motifs are known to enhance processing and interact with the microprocessor (Nguyen et al., 2015), but may also rescue processing of pri-miRNAs with sub-optimal stem lengths (Fang and Bartel, 2015). This effect is limited, however, as the presence of all four motifs did not effectively restore processing of a pri-miRNA with a stem length of 30 bp (Fang and Bartel, 2015). In the context of the current processing model, it seems likely that the rescue or compensatory effect is mediated by motif recognition and correct positioning of the pri-miRNA within the complex. This could allow for appropriate transient pairing or bifurcation of the basal stem to satisfy the stem length requirement. A pri-miRNA with a short stem, like pri-miR-520b, may therefore be more reliant on the presence of sequence motifs for processing; and, in particular, the apical UGU motif for correct positioning within the complex. This is in agreement with the description of pri-miRNA processing as a modular phenomenon, in which different pri-

miRNAs are reliant on different structural elements and motifs to different extents (Auyeung et al., 2013). Here, the processing of pri-miRNAs with less consistent basal junctions may be more reliant on the recognition of apical elements and motifs by the DGCR8 dimer. In line with this suggestion, functional pri-miRNAs with various non-canonical basal stem conformations (pri-miR-520b, 34a, 513a-2) had potential loop motifs in the precise position; while functional pri-miRNAs with more consistent basal stem conformations (pri-miR-23a, 934) lacked typical loop motifs. Motifs have also been shown to rescue processing of pri-miRNAs with structural defects in the stem, such as large internal bulges (Fang and Bartel, 2015). However, a compensatory effect was not observed here for pri-miR-198 and pri-miR-147a which contain large internal bulges and were not functional.

The modular nature of pri-miRNA processing and the current model of Drosha-DGCR8 processing can also explain the discrepancies in earlier reports. Strategic mutations of the pri-miR-30a structure led investigators to conclude that the Drosha-DGCR8 cleavage site is determined by a distance of ~ 22 bp from the terminal loop (Zeng et al., 2005b). However, similar experiments on pri-miR-16-1 led to the conclusion that the cleavage site is instead primarily determined by a distance of ~ 11 bp from the basal junction (Han et al., 2006). These seemingly contradictory results can now be explained. Pri-miR-30a has a terminal loop motif, but lacks a distinct ssRNA-dsRNA basal junction or basal motif, which may explain why it is dependent on DGCR8 for productive processing (Nguyen et al., 2015). In contrast, pri-miR-16-1 lacks a loop motif, but has a distinct ssRNA-dsRNA basal junction and basal motif and is less reliant on DGCR8 for productive processing (Nguyen et al., 2015). This can explain why the cleavage site of pri-miR-30a was not greatly affected by mutations to the basal stem (Zeng et al., 2005b); and why the cleavage site of pri-miR-16-1 was not greatly affected by mutations of the upper stem and terminal loop (Han et al., 2006). This supports the modular nature of pri-miRNA processing and implies that recognition by DGCR8 or Drosha may be more important for precise processing of certain pri-miRNAs. This also implies that recognition by either component of the microprocessor can result in processing, thus supporting a pre-formed heterotrimeric complex (Herbert et al., 2015) or perhaps a mutual recruitment model, as opposed to a specific, sequential assembly (Han et al., 2006).

Recent studies have therefore provided much insight into the recognition and cleavage of pri-miRNAs by the Drosha-DGCR8 microprocessor and resolved several long-standing issues. However, a number of minor issues must still be addressed. Firstly, it is still unclear how certain motifs are recognised by the microprocessor. The apical UGU motif is recognised by the DGCR8 Rhed, but the Drosha domain involved in the recognition of the basal UG motif is currently

unknown. The mismatched GHG stem motif also enhances pri-miRNA processing (Fang and Bartel, 2015), but it is not yet known whether this motif can interact directly with the microprocessor. The location of the GHG motif suggests a possible interaction with Drosha, but the domains involved in the recognition of this motif have not yet been identified (Kwon et al., 2015). It has been suggested that flexible regions such as the PAZ-like domain or MB helix may be able recognise the UG and GHG motifs; or that the Drosha dsRBD may re-position itself upon binding of the pri-miRNA to interact with these motifs (Kwon et al., 2015). Secondly, it is not clear how the microprocessor interacts with other proteins to enhance recognition and processing efficiency. The basal CNNC motif enhances pri-miRNA processing by recruiting the SRp20 (Serine / Arginine-rich protein) (Auyeung et al., 2013), but it is not clear if SRp20 interacts directly with the microprocessor. Other accessory proteins, like KSRP (Trabucchi et al., 2009), hnRNPA1 (Guil and Caceres, 2007) and nucleolin (Pickering et al., 2011) are known to bind to the pri-miRNA terminal loop or Drosha-DGCR8, but details of these interactions are currently unknown. Thirdly, ~ 21 % of human pri-miRNAs lack known primary sequence motifs (Auyeung et al., 2013), which suggests that additional processing determinants may exist. However, conventional processing can still occur in the absence of motifs. Current evidence indicates that Drosha-DGCR8 processing in the absence of these motifs is not necessarily abolished, but may be less efficient and more susceptible to the negative effects of a sub-optimal stem length or internal bulges (Fang and Bartel, 2015). As most natural pri-miRNAs lack a full set of motifs, it has been suggested that decreased processing efficiency may in fact be desirable to enable other modes of post-transcriptional regulation (Fang and Bartel, 2015). The basic requisites for pri-miRNA processing by the Drosha-DGCR8 complex are now known, but it will be interesting to elucidate these remaining details.

4.3 Modular processing of pri-miRNAs

Overall, the processing of both pri-miRNAs and derived mimics appears to be affected by several structure, sequence and stability features that may promote efficiency at different processing events. Firstly, elements of the secondary RNA structure are known to improve the efficiency of Drosha-DGCR8 processing, including a large terminal loop (≥ 10 nt) (Zeng et al., 2005b), a total stem length of 35 ±1 bp (Fang and Bartel, 2015) and at least 40 nt of unstructured flanking sequences in an artificial expression context (Chen et al., 2004). Secondly, the 5’ nucleotide of the miRNA guide sequence is expected to affect the efficiency of guide strand loading into RISC, as the Ago2 core component has a higher binding affinity for a 5’ U or A (Frank et al., 2010). Thirdly, decreased stability at the first position of the guide region is

primarily expected to enhance thermodynamic asymmetry of the miRNA duplex and strand selection by RISC (Khvorova et al., 2003; Schwarz et al., 2003). This feature may also contribute to a deformable region adjacent to the Drosha cleavage site that enhances processing (Quarles et al., 2013), but flexibility at Drosha and Dicer cleavage sites (Warf et al., 2011) may also be an artefact of decreased 5' stability of the guide. In general, it is therefore difficult to predict the function of a pri-miRNA or mimic on the basis on a single feature or parameter, such as the ddG (1-20) signature used here to indicate the relative thermodynamic asymmetry of the guide region. A specific loop size, stem length, motif, thermodynamic asymmetry or 5' sequence alone can not guarantee efficient pri-miRNA or mimic function.

The influence of several factors suggests that processing in general may be considered as a modular phenomenon. For example, the pri-miR-520 scaffold has a sub-optimal basal (9 bp) and total stem length (32 bp) and derived mimics appeared particularly susceptible to changes in the guide region. Derived mimics had reduced target silencing when properties of the guide region were sub-optimal (pri-miR-520-tat: 0 %; pri-miR-520-int: 17 %), but effective target silencing when properties of the guide region were more optimal (pri-miR-520-gag: 45 %; pri-miR-520-LTR: 79 %). This suggested that an optimal guide region may compensate for sub-optimal stem length. Conversely, the pri-miR-34a scaffold has both an optimal basal (11 bp) and total stem length (35 bp) and derived mimics were less susceptible to sub-optimal properties of the guide region (pri-miR-34-tat: 56 %; pri-miR-34-int: 82 %). This suggested that an optimal stem length may also help to compensate for a sub-optimal guide region. Properties of the guide region are primarily expected to affect strand selection of the miRNA duplex, so a modular perspective (Auyeung et al., 2013) could therefore apply to pri-miRNA processing as a whole; and the sequence and subsequent stability of the guide region may be considered as additional modules.

4.4 Alternative approaches to pri-miRNA mimic design

Early mimics were designed by replacing the guide region with a near-perfect duplex (Silva et al., 2005), but the original mismatches and bulges were later found to be important for maintaining strand selection (Aagaard et al., 2008). In retrospect, these original features may have improved mimic function by satisfying certain criteria, for example, a 5p guide with a 5' mismatch for a miR-93-based mimic (Aagaard et al., 2008). In the cautionary approach used here and in other studies (Ely et al., 2008; Liu et al., 2008), original features of the pri-miRNA structure and guide region were preserved without discrimination in an attempt to maintain the

original function. However, specific features and parameters appear to promote efficient pri-miRNA function. A list of optimal features was compiled here in accordance with this study and several others. Optimal features include a total stem length of 35 ± 1 bp (Fang and Bartel, 2015), a terminal loop of 8 - 11 nt (≥ 10 nt) (Zeng et al., 2005b); 40 -50 nt of flanking sequences (Chen et al., 2004); original basal, apical and stem sequence motifs (Auyeung et al., 2013; Fang and Bartel, 2015); METTL3 methylation motifs (Alarcon et al., 2015b) in the proximity of the pre-miRNA; a 5p guide with a 5' U (Amarzguioui and Prydz, 2004; Hsieh et al., 2004; Reynolds et al., 2004; Ui-Tei et al., 2004); and flexible pairing at the first position of the guide to promote decreased stability.

Together, these features indicate that a scaffold like pri-miR-34a was particularly well suited for use in mimic design. However, most natural scaffolds are unlikely to satisfy all of these criteria. This suggests a move towards the rational design of mimics where natural scaffolds can be modified to include optimal features. In a recent study (Choi et al., 2015), effective pri-miRNA mimics based on the precursors for miR-30a, -21, -20a, -16-1, -122, -185 and -150 were designed by replacing the original sequence, bulges and mismatches of the guide region with a 22 nt exogenous 5p guide with artificial mismatches only at positions 1 and 10 (Wu et al., 2011b). A mismatch at position 1 is primarily expected to enhance strand selection, while a mismatch at position 10 is expected to promote RISC loading to form pre-RISC (Yoda et al., 2010). In a second recent study, completely artificial pri-miRNA mimics were created in a de novo approach to mimic design (Fang and Bartel, 2015) proposed over a decade earlier (Zeng et al., 2005b). Arbitrary sequences were used except for optimal features, which included a 5p guide with a 5' U; all four apical, basal and stem sequence motifs, a total stem length of 35 bp, a loop of 12 nt and a mismatch or G - U wobble pair at position 1 of the guide (P1) or the preceding position (-1), as observed for pri-miR-34a. A total of 2 - 4 single G - U wobble pairs and 2 - 3 single mismatches were also included in the stem. This reflects a natural stem configuration, which has on average 3.7 wobble pairs and 2.7 single mismatches and, in comparison to a perfect duplex, facilitates cloning without a decrease in processing (Fang and Bartel, 2015).

Constructs expressing several pri-miRNA mimics can also be modified for optimal function. Anti-HIV mimics based on natural polycistronic clusters, such as the miR-106b and miR-17~92 clusters, were initially used for the expression of three or four exogenous guides (Aagaard et al., 2008; Liu et al., 2008). An artificial polycistronic construct was also generated by using tandem repeats of the pri-miR-30a and pri-miR-155 scaffolds (Zhang et al., 2012). However, the function of individual mimics within these clusters was not necessarily equivalent (Aagaard et al., 2008;

Liu et al., 2008) and could be affected by the order of individual mimics and the identity of the flanking sequences (Zhang et al., 2012). In the context of a lentiviral vector delivery system, multiple copies of the same scaffold should also be avoided to prevent homologous recombination of repeat sequences (ter Brake et al., 2008). This effect was recently demonstrated in a multiplex strategy where up to eighteen tandem mimics, all based on the pri-miR-155 scaffold, were assembled in a single construct using golden-gate cloning; but only constructs expressing up to four mimics were successfully integrated into the genome using a lentiviral vector (Wang et al., 2015). A more effective artificial multiplex construct was recently described, where seven optimised mimics based on different scaffolds were co-expressed from a single construct (Choi et al., 2015). The intended guide sequence of each mimic was accurately expressed without a decrease in guide expression with distance from the promoter; and only 30 nt of flanking sequence was required for each pri-miRNA (Choi et al., 2015). This approach is promising, as the use of tandem optimised mimics can avoid variations in the level of miRNA processing inherent for the different scaffolds of a natural polycistron. Targeting seven conserved viral sequences can also theoretically provide near-complete coverage of all HIV-1 strains (McIntyre et al., 2011), while targeting the host CCR5 gene and six viral genes successfully inhibited viral replication in immunodeficient mice for 6 weeks following transplantation with HIV-infected donor PBMCs (peripheral blood mononuclear cells) expressing the polycistronic construct (Choi et al., 2015).

The current, more comprehensive understanding of pri-miRNA processing therefore signals a shift in mimic design from the reliance on natural scaffolds and polycistronic clusters, to the generation of artificially optimised and de novo structures co-expressed from high-order multiplex constructs. Mimics designed in accordance with recent strategies (Choi et al., 2015; Fang and Bartel, 2015) and the criteria compiled here are expected to have enhanced processing, strand selection and target specificity. The potential redundancy of optimal features may be particularly useful if a sub-optimal feature can not be avoided. For example, if it is necessary to target a particular gene sequence, the corresponding miRNA guide may have unfavourable sequence and stability features, but these can be compensated for by including other optimal features. In addition, the presence of several optimal features may enable sufficient mimic processing in the context of endogenous regulatory networks. However, while it is now possible to optimise pri-miRNA mimics for efficient expression, this might not always be preferable. Decreased processing efficiency may be desirable to enable other modes of post-transcriptional regulation (Fang and Bartel, 2015). Mimics based on natural pri-miRNA scaffolds have also been used successfully in effective gene silencing strategies (Chung et al., 2014; Chung et al., 2012) and clinical trials (Senzer et al., 2012). The design of pri-miRNA mimics

based on natural scaffolds therefore remains a facile and adaptable approach and effective mimics described here (target silencing ≥ 80 %) could be used as alternative units in a multiplex strategy.

4.5 Alternative RNA and combinatorial constructs

In addition to RNAi-based modalities, other RNA-based modalities may also be used to inhibit gene expression and protein function. Prior to the advent of RNAi and microRNAs, ribozymes and RNA aptamers were investigated in several therapeutic approaches against HIV and derived from similar expression constructs (Good et al., 1997). Ribozymes (ribonucleic acid enzymes) are catalytic RNAs that typically mediate cleavage or ligation reactions and can been altered to specifically cleave HIV sequences (Sarver et al., 1990). Effective ribozymes that inhibit HIV-1 replication include hammerhead and hairpin ribozymes targeting the 5' LTR (Yamada et al., 1994; Yu et al., 1993), *tat, rev* (Zhou et al., 1994) and *gag* regions (Unwalla et al., 2008) of the HIV genome. Ribozymes have also been designed to target cellular co-factors like CCR5 to provide resistance against viral infection (Cagnon and Rossi, 2000). RNA aptamers are single-stranded RNAs that form stable three dimensional structures and bind specifically to protein targets to modulate protein activity (Bouchard et al., 2010). The Systematic Evolution of Ligands by EXponential enrichment (SELEX) (Tuerk and Gold, 1990) method was used to identify high affinity therapeutic aptamers, such as TAR and RBE (RRE) RNA decoys that bind and sequester HIV Tat and Rev proteins, respectively (Tuerk and MacDougal-Waugh, 1993). Aptamers have also been designed against other HIV-1 proteins, including the reverse transcriptase (Joshi and Prasad, 2002), gp120 (Khati et al., 2003) and viral protease (Duclair et al., 2015).

RNAi and RNA constructs may be particularly effective when used together in a combinatorial approach. Combinatorial constructs with two (Li, 2003), three (Li et al., 2005) or four (Aagaard et al., 2008) RNA components were developed by the Rossi research group and found to be more effective at inhibiting HIV replication than constructs expressing the individual components. For example, the tricistronic miR-106b cluster was used to design a combinatorial construct (Aagaard et al., 2008) expressing three pri-miRNA mimics against *tat* and *rev* and a U6-expressed chimeric U16 snoRNA-TAR decoy (U16TAR) which localises to the nucleolus (Michienzi et al., 2002). The combinatorial construct containing all four activators was the most effective at inhibiting HIV-1 replication in challenge assays (Aagaard et al., 2008), in agreement with a previously described synergistic effect (Li et al., 2005). The intronic miR-106b cluster was also used to create various triple constructs with co-expression from a single pol II promoter, by

replacing individual pri-miRNAs with a chimeric U16 anti-U5 ribozyme (U16U5RZ) and / or U16TAR or U16RBE decoys (Chung et al., 2012). An additional pol III-expressed U16TAR decoy and anti-CCR5 tRNA-shRNA were also added to the miR-106b platform to create vectors with up to five RNA components (Chung et al., 2014).

In spite of using various RNA components with different processing mechanisms, pri-miRNA mimics were found to be important for therapeutic efficacy. A tricistronic construct expressing only the pri-miRNA mimics was among the most effective triple constructs and mediated potent viral inhibition in a one month challenge assay (Chung et al., 2012).This triple pri-miRNA mimic construct was also more effective than combinatorial constructs with four or five components at inhibiting viral replication in a 42 day challenge assay, with only limited viral break-through at the final time-point (Chung et al., 2014). The less effective function of other combinatorial constructs seems to be a consequence of cellular toxicity or pathway saturation resulting from the use of pol III-expressed decoys and shRNAs (Chung et al., 2014).Therefore, pol II-expressed polycistronic pri-miRNA mimics appear to be an integral component of safe and effective long-term therapeutic constructs targeting HIV-1. Pri-miRNA mimic efficacy was also demonstrated by Choi *et. al.* in their six week study targeting seven anti-HIV sequences with an artificial polycistronic construct (Choi et al., 2015). Nonetheless, targeting of multiple components and stages of the HIV-1 replication cycle is likely to be a necessary approach to prevent viral escape and further optimisation of combinatorial constructs to safely express other components is imminent.

Several other types of components may also be suitable for use in future combinatorial approaches. RNAi components with alternative processing mechanisms may have unique advantages over conventional constructs. ShRNAs with short stems of 17 -19 bp resemble the Dicer-independent pre-miR-451 and are processed by Ago2 (Liu et al., 2013). These Ago-shRNAs are a promising alternative as the sense strand is cleaved by Ago2 and degraded, thus preventing its contribution to off-target silencing (Liu et al., 2013; Yang et al., 2012). Drosha-independent tRNA-shRNA chimeras may avoid saturation of the endogenous miRNA biogenesis pathway as expression from pol III tRNA promoters is typically lower than that from U6 or H1 pol III promoters (Scherer et al., 2007b). Mimics based on endogenous mirtrons, like miR-877, miR-1226 and miR-1224, may also avoid saturation issues as they appear to be partially dependent on Exp-5 (Seow et al., 2012) and several mirtrons can be derived from a single pol II-expressed therapeutic transgene (Sibley et al., 2012a). Another benefit of the mirtron format is expected to be consistency of the pre-miRNA 5' processing site, which is defined by the splicing mechanism instead of Drosha RNAse III cleavage (Kock et al., 2015; Sibley et al., 2012a). However, greater

5’ and 3’ end heterogeneity have also been observed for endogenous mirtrons as a result of trimming or non-templated additions (Wen et al., 2015), suggesting that this approach may be limited by an increased risk of off-target silencing.

Additional components can also be included to improve an RNAi approach. Off-target silencing by shRNAs or mimics may be prevented by a co-expressed tough decoy (TuD) RNA, which binds and sequesters the sense strand to improve specificity (Mockenhaupt et al., 2015). Cell-specific delivery and internalisation of siRNAs can be achieved through the use of aptamer-siRNA chimeras, like the gp120 aptamer - *tat / rev* siRNA chimera (Zhou et al., 2008) or the CCR5 aptamer - transportin 3 siRNA chimera (Zhou et al., 2015), which block viral entry and inhibit viral replication. Artificial decoy targets or “sponges” may be included to reduce the undesired activity of certain miRNAs by competing with endogenous targets (Ebert et al., 2007). Non-RNA moieties can also be added to combinatorial expression constructs, such as transdominant negative mutants of HIV-1 proteins, like Tat (Pearson et al., 1990), Rev (Malim et al., 1989), and Gag (Trono et al., 1989), which inhibit function of the wildtype protein. It would be interesting to investigate the possibility of an anti-HIV construct containing a transdominant HIV-1 or cellular co-factor protein with intronic RNAi components, like several mirtrons or an intronic pri-miRNA mimic cluster. In an alternative approach to conventional RNAi, introns of a single protein coding gene can be modified to express snoRNA modulators of gene expression (snoMEN) for gene silencing (Ono et al., 2010). This approach may prove useful for therapeutic protein replacement strategies (Ono et al., 2010). Overall, there are several different options for novel combinatorial approaches against HIV, as well as other disease conditions. It will be interesting to see which combinations of components are most effective and whether additional components can improve the function of polycistronic pri-miRNA mimics in long-term applications.

4.6 Delivery of expressed RNA constructs

The efficient delivery of therapeutic constructs is important for any successful gene therapy and is of particular concern in the treatment of HIV-1 to prevent untargeted viral reservoirs (Leonard and Schaffer, 2005). Therapeutic RNAi and RNA constructs can be stably expressed in human cells through the use of viral vectors. Vectors based on adenovirus (Ad), adeno-associated virus (AAV), retrovirus or lentivirus genomes are commonly used, each with particular advantages and disadvantages (Liu and Berkhout, 2011). Ad vectors are preferred for applications requiring short-term hepatic delivery, such as RNAi activators against the Hepatitis B virus (HBV) (Mowa

et al., 2010), but their use in HIV applications is limited by immunostimulatory effects and the inability to integrate into the host genome. Other viral vectors have therefore been used in HIV-1 applications. AAV vectors were used to express a highly effective shRNA and pri-miR-30-based mimic against *tat* (Boden et al., 2003; Boden et al., 2004) and retroviral (RV) vectors were used to express an anti-*nef* shRNA (Das et al., 2004), as well as various ribozymes and aptamers (Bai et al., 2002). Lentiviral (LV) vectors are particularly well-suited for HIV-1 applications, as they can transduce non-dividing cells and integrate into the genome, preferentially within introns (Ustek et al., 2012), for persistent expression (Naldini et al., 1996). LV vectors have been used in a number of studies to express anti-HIV shRNAs (Lee et al., 2003a), multi-shRNAs (ter Brake et al., 2006), lhRNAs (Liu et al., 2009; Nishitsuji et al., 2006) and polycistronic pri-miRNA mimics (Liu et al., 2008; Zhang et al., 2012). Repeat sequences can lead to the recombination and deletion of individual cassettes in a lentiviral vector expression system, but this can be remedied by using several different promoters in a combinatorial construct (ter Brake et al., 2008). An easier option is to use a combinatorial construct with a single promoter sequence, highlighting another advantage of a pol II-expressed polycistronic pri-miRNA mimic.

In a clinical setting, the *ex vivo* transduction of a patient's own T lymphocytes or haematopoietic stem cells (HSCs) followed by an autologous cell transfer is a promising approach for HIV-1 gene therapy (Rossi et al., 2007). T cells transduced with a therapeutic vector can provide immediate resistance against HIV infection; while modified HSCs are capable of self-renewal and can differentiate into other blood cell types, including T lymphocytes, macrophages and dendritic cells. A cell-based approach could provide a lifelong source of HIV-resistant cells and act as a functional cure for HIV infection. Advances in vector design have therefore been made to improve transduction efficiency and clinical safety towards these therapeutic goals. An HIV-based LV vector optimised for transduction of HSCs and primary T lymphocytes (pHIV7) (Yam et al., 2002) has been used in several recent studies to express combinatorial constructs (Chung et al., 2014; Chung et al., 2012; Li et al., 2005). This is a self-inactivating (SIN) retroviral vector with deletions in the enhancer and promoter sequences that prevent active transcription of the provirus following integration, thus improving vector safety by minimising the risk of vector mobilisation upon infection (Bukovsky et al., 1999) and insertional activation of proto-oncogenes (Miyoshi et al., 1998; Zufferey et al., 1998). LV vectors are commonly pseudotyped with VSV-G (vesicular stomatitis virus envelope G glycoprotein) to expand cell tropism and enable the transduction of CD34+ HSCs (Akkina et al., 1996; Naldini et al., 1996). Alternatively, a lentiviral vector pseudotyped with HIV-1 envelope protein can limit transduction to HIV-1 susceptible cells and enable transduction of both resting and activated T cells (Agosto et al., 2009; Choi et al., 2015). A pol II-expressed polycistronic mimic was recently identified as an ideal RNAi activator

in the context of a lentiviral vector expression system and did not induce cellular toxicity or limit the differentiation potential of HSCs (Chung et al., 2014). The individual pri-miRNA mimic units characterised here are therefore suitable for testing in a polycistronic format in a lentiviral vector and it would be interesting to see if the effective target silencing observed in cell culture assays would translate in a pre-clinical expression system.

4.7 Pre-clinical and clinical studies of expressed RNA constructs

Anti-HIV RNA and RNAi constructs are commonly tested in *in vivo* pre-clinical studies using humanised mouse models, which may contain human genes, cells or tissue. In the study of HIV-1 infection and treatment, humanised mouse models that can be engrafted with HSCs are particularly useful (Berges and Rowan, 2011; Denton and Garcia, 2011). Following sub-lethal irradiation, transplanted HSCs can reconstitute a functional immune system with multi-lineage human haematopoietic cells, thus creating a model system for HIV-1 infection in human cells. There are four basic mouse models that have been engrafted with a functional human immune system and used in HIV studies, each with specific advantages and limitations (Shultz et al., 2012). These are (1) Hu-PBL-SCID, created by injection of human peripheral blood lymphocytes (PBLs) into SCID (severe combined immunodeficiency) mice (Mosier et al., 1988); (2) Hu-SRC-SCID, created by injection of human HSCs as SCID-repopulating cells (SRCs) (Lapidot et al., 1992); (3) SCID-hu, created by implantation of human foetal liver and thymus under the renal capsule (McCune et al., 1988); and (4) BLT, created by implantation of human foetal liver and thymus under the renal capsule of NOD/SCID (non-obese diabetic) mice, followed by injection of human HSCs from the same foetal liver (Melkus et al., 2006). The anti-HIV efficacy of lentiviral vectors expressing RNA and RNAi constructs has been assessed in several models. For example, a lentiviral vector expressing a polycistronic pri-miRNA mimic was tested in Hu-PBL mice (Choi et al., 2015); and lentiviral vectors expressing shRNAs have been tested in BLT mice (Burke et al., 2015; Ringpis et al., 2012). Combinatorial RNAi constructs expressed from a lentiviral vector have also been tested in other variants of humanised mice which permit HSC engraftment and T cell development. For example, the NRG mouse model (Brehm et al., 2010) was used to test a construct expressing a shRNA and TAR decoy (Walker et al., 2012); and the NSG mouse model (Shultz et al., 2005) was used to test combinations of pri-miRNA mimics, ribozymes and decoys (Chung et al., 2014). However, BLT mice have been emphasised as an infection model that more closely resembles that in humans, as T cell development occurs in human thymus tissue, mucosal transmission of HIV-1 is readily supported and CD4+ T cells are rapidly depleted upon infection (Denton and Garcia, 2011; Shultz et al., 2012).

Several RNA constructs targeting HIV have also progressed to clinical studies using *ex vivo* cell transduction and transfer methods. In general, these studies have served as “proof of concept” for a cell-based gene therapy targeting HIV infection. Treatments have proven to be well-tolerated in subjects with long-term persistence, but significant efficacy has been difficult to demonstrate with a low percentage of modified cells. In a well characterised example, a ribozyme targeting the overlapping *tat* and *vpr* reading frames was shown to be effective at inhibiting HIV replication *in vitro* (Sun et al., 1995; Wang et al., 1998) and was expressed from a Moloney murine leukaemia virus (MLV) vector (OZ1) for assessment in phase I and phase II clinical trials. In a phase I trial using *ex vivo* transduction of CD4+ T lymphocytes, the ribozyme construct was detected for up to 44 months (Macpherson et al., 2005); while in a similar phase I trial using CD34+ HPCs (haematopoietic progenitor cells), the construct was detected in multiple cell lineages with de novo development of modified T lymphocytes (Amado et al., 2004). In a subsequent placebo-controlled phase II trial using CD34+ HPCs, CD4+ lymphocyte counts were higher for the OZ1 treatment group over 100 weeks, but there was no statistically significant difference in viral loads, presumably as a consequence of low engraftment efficiency in the absence of strong selective pressure (Mitsuyasu et al., 2009). A long-term, follow-up study to assess the safety of this treatment is ongoing (NCT01177059) (https://clinicaltrials.gov, accessed 19 April 2016).

A similar outcome was obtained in clinical trials for a triple combinatorial RNA construct. A lentiviral vector expressing an anti-*tat* / *rev* shRNA, a TAR decoy and an anti-CCR5 ribozyme (pHIV7-shI-TAR-CCR5RZ) was found to mediate long-term inhibition of HIV-1 replication in primary haematopoietic cells (Li et al., 2005). This construct was subsequently assessed in a clinical trial of 4 patients with AIDS-related non-Hodgkins lymphoma (NHL), where ablation of the bone marrow prior to HPC transplantation was ethically acceptable (DiGiusto et al., 2010). The treatment was well-tolerated and siRNA and ribozyme expression was detected at low - levels for up to 24 months in multiple cell lineages, demonstrating long-term expression of therapeutic RNA constructs. However, patients were transplanted with both modified and unmodified HPCs for ethical reasons, resulting in a low percentage of modified cells and difficulty in assessing anti-HIV efficacy. Subsequent trials involving the same triple RNA construct (NCT02337985; NCT01961063; NCT00569985) are ongoing.

Completed clinical trials of anti-HIV RNA constructs have thus far shown much promise and demonstrated the safety and feasibility of a cell-based gene approach for gene therapy. However, a true indication of clinical efficacy will only be possible in future studies using a higher

dose of transduced cells and selective markers to improve the percentage of modified cells *in vivo.* A drug resistance marker was included in the effective combinatorial pri-miRNA mimic, lentiviral expression vector developed by the Rossi research group (Chung et al., 2014), making this vector a particularly promising candidate for future gene therapy. The marker, a mutant form of the human methylguanine methyltransferase gene ($MGMT^{P140K}$), allows for the positive selection of cells that survive treatment with O^6-BG / BCNU (O^6-benzylguanine / bis-chloroethylnitrosourea) (Davis et al., 1999). The use of pri-miRNA mimics in future gene therapy applications is also supported by superior *in vivo* efficacy in animal models (Choi et al., 2015; Chung et al., 2014), as well as established safety and efficacy in phase I trials against various cancers (Rao et al., 2010; Senzer et al., 2012). In addition, significant progress has been made in the application of viral vectors. Since 2003, several therapies using Ad and AAV vectors have received market approval in China and the EU (Wirth et al., 2013), while several therapies using LV vectors are currently in Phase I and II clinical trials (Booth et al., 2016). Most notably, in April 2016, a gene therapy (Strimvelis) using *ex vivo* transduction of CD34+ HSCs with a retrovirus expressing wildtype adenosine deaminase (ADA) was recommended for approval in the EU for use in children with ADA deficiency where no matched stem cell donor is available (EMA, 2016). The progress of RNA-based gene therapies (Burnett and Rossi, 2012) and HSC *ex vivo* gene therapies (Booth et al., 2016) lends promise to similar strategies for the treatment of HIV-1.

4.8 Genome-editing: a new approach to gene therapy

Another promising avenue in anti-HIV gene therapy is that of genome editing. Several tools for gene editing have been developed in recent years and can be used to permanently disrupt or alter human genes. The three main tools are zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the CRISPR-Cas system (clustered regularly interspaced short palindromic repeats - CRISPR associated nuclease), reviewed in (Urnov et al., 2010), (Joung and Sander, 2013) and (Sander and Joung, 2014), respectively. All three nuclease approaches can be used to introduce a double strand break (DSB) at specific DNA sequences, which is then repaired by non-homologous end-joining (NHEJ), resulting in targeted mutagenesis; or homology directed repair (HDR) in the presence of a donor DNA template, resulting in targeted gene correction or replacement. Targeting of a gene using the CRISPR-Cas method only requires the generation of a complementary single guide RNA (sgRNA), making this approach easier than those using ZFNs or TALENs which require the generation of custom proteins. The CRISPR-Cas system can also be modified to regulate gene expression through CRISPR interference (CRISPRi) or CRISPR activation (CRISPRa), in which deactivated

versions of the Cas9 nuclease (dCas9) are fused to repressor or activator domains, respectively, and bind to target DNA to regulate transcription (Dominguez et al., 2016).

Gene-editing tools can be used to permanently disrupt or replace cellular co-factors or the reverse-transcribed proviral DNA before or after integration (Hu et al., 2014). A number of *ex vivo* gene editing strategies have been developed for HIV-1 infection and are currently in phase I and phase II clinical trials (Maeder and Gersbach, 2016). In particular, the success of the "Berlin patient" has made the CCR5 gene an attractive target (Allers and Schneider, 2015). In this case, an HIV positive patient with acute myeloid leukaemia received a stem cell transplant from a donor homozygous for the CCR5Δ32 deletion and no viral rebound was detected following the transplant and discontinuation of HAART (Hutter et al., 2009). However, this result has not been repeated in other patients and the introduction of cells with the CCR5Δ32 allele can result in a shift in viral tropism and viral rebound (Kordelas et al., 2014). Therefore, like conventional drug treatments and proposed RNA approaches, multiple genes need to be targeted consistently and simultaneously to prevent viral escape (Wang et al., 2016). However, the continued expression of nucleases in the cell is undesirable (Swamy et al., 2016). Gene-editing methods also share similar concerns to RNAi-based approaches, such as those concerning off-target effects, safety and delivery. It has therefore been argued that an RNAi is the only practical approach to suppress the function of several host and viral genes in both infected and uninfected cells, which is essential for an effective prevention and treatment strategy for HIV-1 infection (Swamy et al., 2016). Nonetheless, a gene-editing approach may prove to be a useful adjunct to existing drug regimens or an RNA-based gene therapy and, crucially, may be used to target latent viral genomes and extinguish viral reservoirs (Saayman et al., 2016).

4.9 Concluding remarks

RNAi has come a long way in nearly two decades since its first description. This field has evolved from the discovery of small RNAs in numerous organisms and processes, to the description of intricate gene regulation pathways, to the development of RNAi activators for experimental analyses and gene therapy applications in humans. We now have a more complete knowledge of the various aspects of mammalian microRNA biogenesis, including a comprehensive model of pri-miRNA processing by the Drosha-DGCR8 complex. Recent structural and large-scale sequencing studies have provided much insight into pri-miRNA processing, thus providing a likely explanation for the function of pri-miRNAs with various basal stem lengths. This knowledge also enables a shift towards rational and de novo design of pri-

miRNA mimics with optimal processing and function. The updated set of rules outlined here may prove useful in this endeavour. Polycistronic pri-miRNA mimics now also appear to be ideal for the safe, effective and long -term silencing of HIV-1 replication when expressed from a lentiviral vector. The anti-HIV pri-miRNA mimics characterised here are suitable for incorporation into such an expression vector and may be useful RNAi components in a future therapeutic strategy. Advances in the design of RNA constructs, lentiviral vectors, humanised mouse models, *ex vivo* transduction strategies and gene-editing tools have lent much promise to the development of an anti-HIV gene therapy. A number of RNA, RNAi and gene-editing constructs that target HIV are currently in preclinical and clinical studies towards this goal. In light of this progress and the success of gene therapies for primary immunodeficiency diseases, a successful *ex vivo* gene therapy for HIV-1 infection appears imminent. This is an exciting era in the fields of RNAi and gene therapy, with a variety of tools now available to regulate gene function. As such, we are closer than ever before to achieving a more general, functional cure for HIV-1 infection, as well as effective treatments for a number of other diseases.

REFERENCES

(HHS), U.S.Department of Health and Human Services (2016). Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. In AIDSinfo.

Aagaard, L.A., Zhang, J., von Eije, K.J., Li, H., Sætrom, P., Amarzguioui, M., and Rossi, J.J. (2008). Engineering and optimization of the miR-106b cluster for ectopic expression of multiplexed anti-HIV RNAs. Gene Ther, 1 - 14.

Adachi, A., Gendelman, H.E., Koenig, S., Folks, T., Willey, R., Rabson, A., and Martin, M.A. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol *59*, 284-291.

Agostini, M., Tucci, P., Steinert, J.R., Shalom-Feuerstein, R., Rouleau, M., Aberdam, D., Forsythe, I.D., Young, K.W., Ventura, A., Concepcion, C.P.*, et al.* (2011). microRNA-34a regulates neurite outgrowth, spinal morphology, and function. Proc Natl Acad Sci U S A *108*, 21099-21104.

Agosto, L.M., Yu, J.J., Liszewski, M.K., Baytop, C., Korokhov, N., Humeau, L.M., and O'Doherty, U. (2009). The CXCR4-tropic human immunodeficiency virus envelope promotes more-efficient gene delivery to resting CD4+ T cells than the vesicular stomatitis virus glycoprotein G envelope. J Virol *83*, 8153-8162.

Agranat-Tamir, L., Shomron, N., Sperling, J., and Sperling, R. (2014). Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. Nucleic Acids Res *42*, 4640-4651.

Akashi, H., Miyagishi, M., Yokota, T., Watanabe, T., Hino, T., Nishina, K., Kohara, M., and Taira, K. (2005). Escape from the interferon response associated with RNA interference using vectors that encode long modified hairpin-RNA. Mol Biosyst *1*, 382-390.

Akkina, R.K., Walton, R.M., Chen, M.L., Li, Q.X., Planelles, V., and Chen, I.S. (1996). High-efficiency gene transfer into CD34+ cells with a human immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular stomatitis virus envelope glycoprotein G. J Virol *70*, 2581-2585.

Alarcon, C.R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., and Tavazoie, S.F. (2015a). HNRNPA2B1 Is a Mediator of mA-Dependent Nuclear RNA Processing Events. Cell *162*, 1-10.

Alarcon, C.R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S.F. (2015b). N6-methyladenosine marks primary microRNAs for processing. Nature *519*, 482-485.

Alexiou, P., Maragkakis, M., Papadopoulos, G.L., Reczko, M., and Hatzigeorgiou, A.G. (2009). Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics *25*, 3049-3055.

Alexopoulou, L., Holt, A.C., Medzhitov, R., and Flavell, R.A. (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. Nature *413*, 732-738.

Alizon, M., Sonigo, P., Barre-Sinoussi, F., Chermann, J.C., Tiollais, P., Montagnier, L., and Wain-Hobson, S. (1984). Molecular cloning of lymphadenopathy-associated virus. Nature *312*, 757-760.

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., and Berger, E.A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science *272*, 1955-1958.

Allers, K., and Schneider, T. (2015). CCR5Delta32 mutation and HIV infection: basis for curative HIV therapy. Curr Opin Virol *14*, 24-29.

Allo, M., Agirre, E., Bessonov, S., Bertucci, P., Gomez Acuna, L., Buggiano, V., Bellora, N., Singh, B., Petrillo, E., Blaustein, M.*, et al.* (2014). Argonaute-1 binds transcriptional enhancers and controls constitutive and alternative splicing in human cells. Proc Natl Acad Sci U S A *111*, 15622-15629.

Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M.J., Tuschl, T., and Margalit, H. (2005). Clustering and conservation patterns of human microRNAs. Nucleic Acids Res *33*, 2697-2706.

Amado, R.G., Mitsuyasu, R.T., Rosenblatt, J.D., Ngok, F.K., Bakker, A., Cole, S., Chorn, N., Lin, L.S., Bristol, G., Boyd, M.P.*, et al.* (2004). Anti-human immunodeficiency virus hematopoietic progenitor cell-delivered ribozyme in a phase I study: myeloid and lymphoid reconstitution in human immunodeficiency virus type-1-infected patients. Hum Gene Ther *15*, 251-262.

Amarzguioui, M., and Prydz, H. (2004). An algorithm for selection of functional siRNA sequences. Biochem Biophys Res Commun *316*, 1050-1058.

Amarziguioui, M., Rossi, J.J., and Kim, D. (2005). Approaches for chemically synthesised siRNA and vector-mediated RNAi. FEBS Letters *579*, 5974-5981.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

An, D.S., Qin, F.X., Auyeung, V.C., Mao, S.H., Kung, S.K., Baltimore, D., and Chen, I.S. (2006). Optimization and functional effects of stable short hairpin RNA expression in primary human lymphocytes via lentiviral vectors. Mol Ther *14*, 494-504.

Anderson, E.M., Birmingham, A., Baskerville, S., Reynolds, A., Maksimova, E., Leake, D., Fedorov, Y., Karpilow, J., and Khvorova, A. (2008). Experimental validation of the importance of seed complement frequency to siRNA specificity. RNA *14*, 853-861.

Applegate, T.L., Birkett, D.J., McIntyre, G.J., Jaramillo, A.B., Symonds, G., and Murray, J.M. (2010). In silico modeling indicates the development of HIV-1 resistance to multiple shRNA gene therapy differs to standard antiretroviral therapy. Retrovirology *7*, 83.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Current biology : CB *11*, 1017-1027.

Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell *152*, 844-858.

Azuma-Mukai, A., Oguri, H., Mituyama, T., Qian, Z.R., Asai, K., Siomi, H., and Siomi, M.C. (2008). Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci U S A *105*, 7964-7969.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev *22*, 2773-2785.

Baccarini, A., Chauhan, H., Gardner, T.J., Jayaprakash, A.D., Sachidanandam, R., and Brown, B.D. (2011). Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. Current biology : CB *21*, 369-376.

Bae, Y., Yang, T., Zeng, H.C., Campeau, P.M., Chen, Y., Bertin, T., Dawson, B.C., Munivez, E., Tao, J., and Lee, B.H. (2012). miRNA-34c regulates Notch signaling during bone development. Hum Mol Genet *21*, 2991-3000.

Bai, J., Banda, N., Lee, N.S., Rossi, J., and Akkina, R. (2002). RNA-based anti-HIV-1 gene therapeutic constructs in SCID-hu mouse model. Mol Ther *6*, 770-782.

Ballarino, M., Pagano, F., Girardi, E., Morlando, M., Cacchiarelli, D., Marchioni, M., Proudfoot, N.J., and Bozzoni, I. (2009). Coupled RNA processing and transcription of intergenic primary microRNAs. Molecular and cellular biology *29*, 5632-5638.

Barichievy, S. (2009). Post-transcriptional inhibition of human immunodeficiency virus type 1 (HIV-1) using combinatorial RNA interference (RNAi) expression vectors. In Department of Molecular Medicine and Haematology (Johannesburg: University of the Witwatersrand), pp. 215.

Barichievy, S., Saayman, S., von Eije, K.J., Morris, K.V., Arbuthnot, P., and Weinberg, M.S. (2007). The inhibitory efficacy of RNA POL III-expressed long hairpin RNAs targeted to untranslated regions of the HIV-1 5' long terminal repeat. Oligonucleotides *17*, 419-431.

Barr, I., Smith, A.T., Chen, Y., Senturia, R., Burstyn, J.N., and Guo, F. (2012). Ferric, not ferrous, heme activates RNA-binding protein DGCR8 for primary microRNA processing. Proc Natl Acad Sci U S A *109*, 1919-1924.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C.*, et al.* (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science *220*, 868-871.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Baskerville, S., and Bartel, D.P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA *11*, 241-247.

Behlke, M.A. (2008). Chemical modification of siRNAs for in vivo use. Oligonucleotides *18*, 305-319.

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

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promotor region. Nature *290*, 304-310.

Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2015). GenBank. Nucleic Acids Res *43*, D30-35.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol Cell *28*, 328-336.

Berger, E.A., Doms, R.W., Fenyo, E.M., Korber, B.T., Littman, D.R., Moore, J.P., Sattentau, Q.J., Schuitemaker, H., Sodroski, J., and Weiss, R.A. (1998). A new classification for HIV-1. Nature *391*, 240.

Berges, B.K., and Rowan, M.R. (2011). The utility of the new generation of humanized mice to study HIV-1 infection: transmission, prevention, pathogenesis, and treatment. Retrovirology *8*, 65.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. Journal of bacteriology *62*, 293-300.

Bertero, T., Grosso, S., Robbe-Sermesant, K., Lebrigand, K., Henaoui, I.S., Puissegur, M.P., Fourre, S., Zaragosi, L.E., Mazure, N.M., Ponzio, G.*, et al.* (2012). "Seed-Milarity" confers to hsa-miR-210 and hsa-miR-147b similar functional activity. PLoS ONE *7*, e44919.

Bertrand, E., Castanotto, D., Zhou, C., Carbonnelle, C., Lee, N.S., Good, P., Chatterjee, S., Grange, T., Pictet, R., Kohn, D.*, et al.* (1997). The expression cassette determines the functional activity of ribozymes in mammalian cells by controlling their intracellular localization. RNA *3*, 75-88.

Betancur, J.G., and Tomari, Y. (2012). Dicer is dispensable for asymmetric RISC loading in mammals. RNA *18*, 24-30.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J.*, et al.* (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nature methods *3*, 199-204.

Boden, D., Pusch, O., Lee, F., Tucker, L., and Ramratnam, B. (2003). Human immunodeficiency virus type 1 escape from RNA interference. J Virol *77*, 11531-11535.

Boden, D., Pusch, O., Silbermann, R., Lee, F., Tucker, L., and Ramratnam, B. (2004). Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. Nucleic Acids Res *32*, 1154-1158.

Boele, J., Persson, H., Shin, J.W., Ishizu, Y., Newie, I.S., Sokilde, R., Hawkins, S.M., Coarfa, C., Ikeda, K., Takayama, K.*, et al.* (2014). PAPD5-mediated 3' adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. Proc Natl Acad Sci U S A *111*, 11467-11472.

Bogenhagen, D.F., and Brown, D.D. (1981). Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell *24*, 261-270.

Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlers, M., and Cullen, B.R. (2010). A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. Mol Cell *37*, 135-142.

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

Boon, R.A., Iekushi, K., Lechner, S., Seeger, T., Fischer, A., Heydt, S., Kaluza, D., Treguer, K., Carmona, G., Bonauer, A.*, et al.* (2013). MicroRNA-34a regulates cardiac ageing and function. Nature *495*, 107-110.

Booth, C., Gaspar, H.B., and Thrasher, A.J. (2016). Treating Immunodeficiency through HSC Gene Therapy. Trends Mol Med *22*, 317-327.

Borchert, G.M., Lanier, W., and Davidson, B.L. (2006). RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol *13*, 1097-1101.

Bortolamiol-Becet, D., Hu, F., Jee, D., Wen, J., Okamura, K., Lin, C.J., Ameres, S.L., and Lai, E.C. (2015). Selective Suppression of the Splicing-Mediated MicroRNA Pathway by the Terminal Uridyltransferase Tailor. Mol Cell *59*, 217-228.

Bortolin-Cavaille, M.L., Dance, M., Weber, M., and Cavaille, J. (2009). C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. Nucleic Acids Res *37*, 3464-3473.

Bouchard, P.R., Hutabarat, R.M., and Thompson, K.M. (2010). Discovery and development of therapeutic aptamers. Annu Rev Pharmacol Toxicol *50*, 237-257.

Boudreau, R.L., Martins, I., and Davidson, B.L. (2009). Artificial MicroRNAs as siRNA Shuttles: Improved Safety as Compared to shRNAs In vitro and In vivo. Mol Ther *17*, 169-175.

Boudreau, R.L., Monteys, A.M., and Davidson, B.L. (2008). Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. RNA *14*, 1834-1844.

Bouhallier, F., Allioli, N., Lavial, F., Chalmel, F., Perrard, M.H., Durand, P., Samarut, J., Pain, B., and Rouault, J.P. (2010). Role of miR-34c microRNA in the late steps of spermatogenesis. RNA *16*, 720-731.

Bour, S., Geleziunas, R., and Wainberg, M.A. (1995). The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. Microbiol Rev *59*, 63-93.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Res *39*, 675-686.

Bramsen, J.B., Laursen, M.B., Damgaard, C.K., Lena, S.W., Babu, B.R., Wengel, J., and Kjems, J. (2007). Improved silencing properties using small internally segmented interfering RNAs. Nucleic Acids Res *35*, 5886-5897.

Bramsen, J.B., Pakula, M.M., Hansen, T.B., Bus, C., Langkjaer, N., Odadzic, D., Smicius, R., Wengel, S.L., Chattopadhyaya, J., Engels, J.W.*, et al.* (2010). A screen of chemical modifications identifies position-specific modification by UNA to most potently reduce siRNA off-target effects. Nucleic Acids Res *38*, 5761-5773.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science *319*, 921-926.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell *44*, 120-133.

Brehm, M.A., Cuthbert, A., Yang, C., Miller, D.M., DiIorio, P., Laning, J., Burzenski, L., Gott, B., Foreman, O., Kavirayani, A.*, et al.* (2010). Parameters for establishing humanized mouse models to study human immunity: analysis of human hematopoietic stem cell engraftment in three immunodeficient strains of mice bearing the IL2rgamma(null) mutation. Clin Immunol *135*, 84-98.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol *3*, e85.

Bridge, A.J., Pebernard, S., Ducraux, A., Nicoulaz, A.L., and Iggo, R. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. Nat Genet *34*, 263-264.

Brown, T., Mackey, K., and Du, T. (2004). Analysis of RNA by northern and slot blot hybridization. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 4*, Unit 4 9.

Brownawell, A.M., and Macara, I.G. (2002). Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. J Cell Biol *156*, 53-64.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science *296*, 550-553.

Bukovsky, A.A., Song, J.P., and Naldini, L. (1999). Interaction of human immunodeficiency virus-derived vectors with wild-type virus in transduced cells. J Virol *73*, 7087-7092.

Burke, B.P., Levin, B.R., Zhang, J., Sahakyan, A., Boyer, J., Carroll, M.V., Colon, J.C., Keech, N., Rezek, V., Bristol, G.*, et al.* (2015). Engineering Cellular Resistance to HIV-1 Infection In Vivo Using a Dual Therapeutic Lentiviral Vector. Molecular therapy Nucleic acids *4*, e236.

Burke, J.M., Kelenis, D.P., Kincaid, R.P., and Sullivan, C.S. (2014). A central role for the primary microRNA stem in guiding the position and efficiency of Drosha processing of a viral pri-miRNA. RNA *20*, 1068-1077.

Burnett, J.C., and Rossi, J.J. (2012). RNA-based therapeutics: current progress and future prospects. Chem Biol *19*, 60-71.

Burrell, M.M. (1993). RNase A (EC 3.1.27.5). Methods in molecular biology *16*, 263-270.

Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2011). Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol *8*, 158-177.

Cagnon, L., and Rossi, J.J. (2000). Downregulation of the CCR5 beta-chemokine receptor and inhibition of HIV-1 infection by stable VA1-ribozyme chimeric transcripts. Antisense Nucleic Acid Drug Dev *10*, 251-261.

Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA *10*, 1957-1966.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proc Natl Acad Sci U S A *98*, 9742-9747.

Capodici, J., Kariko, K., and Weissman, D. (2002). Inhibition of HIV-1 infection by small interfering RNA-mediated RNA interference. J Immunol *169*, 5196-5201.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev *16*, 2733-2742.

Castanotto, D., Li, H., and Rossi, J.J. (2002). Functional siRNA expression from transfected PCR products. RNA *8*, 1454-1460.

Castanotto, D., Sakurai, K., Lingeman, R., Li, H., Shively, L., Aagaard, L., Soifer, H., Gatignol, A., Riggs, A., and Rossi, J.J. (2007). Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. Nucleic Acids Res *35*, 5154-5164.

Castellano, L., and Stebbing, J. (2013). Deep sequencing of small RNAs identifies canonical and non-canonical miRNA and endogenous siRNAs in mammalian somatic tissues. Nucleic Acids Res *41*, 3339-3351.

Cerutti, L., Mian, N., and Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends Biochem Sci *25*, 481-482.

Chang, T.C., Wentzel, E.A., Kent, O.A., Ramachandran, K., Mullendore, M., Lee, K.H., Feldmann, G., Yamakuchi, M., Ferlito, M., Lowenstein, C.J.*, et al.* (2007). Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. Mol Cell *26*, 745-752.

Chang, T.C., Yu, D., Lee, Y.S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A., and Mendell, J.T. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. Nat Genet *40*, 43-50.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T.*, et al.* (2011). Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. RNA Biol *8*, 1105-1114.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584-589.

Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83-86.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Chi, S.W., Hannon, G.J., and Darnell, R.B. (2012). An alternative mode of microRNA target recognition. Nat Struct Mol Biol *19*, 321-327.

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

Choi, J.G., Bharaj, P., Abraham, S., Ma, H., Yi, G., Ye, C., Dang, Y., Manjunath, N., Wu, H., and Shankar, P. (2015). Multiplexing seven miRNA-Based shRNAs to suppress HIV replication. Mol Ther *23*, 310-320.

Chomczynski, P. (1992). Solubilization in formamide protects RNA from degradation. Nucleic Acids Res *20*, 3791-3792.

Chung, J., Scherer, L.J., Gu, A., Gardner, A.M., Torres-Coronado, M., Epps, E.W., Digiusto, D.L., and Rossi, J.J. (2014). Optimized lentiviral vectors for HIV gene therapy: multiplexed expression of small RNAs and inclusion of MGMT(P140K) drug resistance gene. Mol Ther *22*, 952-963.

Chung, J., Zhang, J., Li, H., Ouellet, D.L., DiGiusto, D.L., and Rossi, J.J. (2012). Endogenous MCM7 microRNA cluster as a novel platform to multiplex small interfering and nucleolar RNAs for combinational HIV-1 gene therapy. Hum Gene Ther *23*, 1200-1208.

Chung, K.H., Hart, C.C., Al-Bassam, S., Avery, A., Taylor, J., Patel, P.D., Vojtek, A.B., and Turner, D.L. (2006). Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. Nucleic Acids Res *34*, e53.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D.*, et al.* (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694-1698.

Clemens, M.J. (1997). PKR--a protein kinase regulated by double-stranded RNA. Int J Biochem Cell Biol *29*, 945-949.

Coburn, G.A., and Cullen, B.R. (2002). Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference. J Virol *76*, 9225-9231.

Corbo, C., Orru, S., and Salvatore, F. (2013). SRp20: an overview of its role in human diseases. Biochem Biophys Res Commun *436*, 1-5.

Corcoran, D.L., Pandit, K.V., Gordon, B., Bhattacharjee, A., Kaminski, N., and Benos, P.V. (2009). Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. PLoS One *4*, e5279.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S.*, et al.* (2015). Ensembl 2015. Nucleic Acids Res *43*, D662-669.

Curtis, H.J., Sibley, C.R., and Wood, M.J. (2012). Mirtrons, an emerging class of atypical miRNA. Wiley interdisciplinary reviews RNA *3*, 617-632.

Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R.*, et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

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

Dallas, A., Ilves, H., Ge, Q., Kumar, P., Shorenstein, J., Kazakov, S.A., Cuellar, T.L., McManus, M.T., Behlke, M.A., and Johnston, B.H. (2012). Right- and left-loop short shRNAs have distinct and unusual mechanisms of gene silencing. Nucleic Acids Res *40*, 9255-9271.

Darnell, J.E., Jr., Kerr, I.M., and Stark, G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. Science *264*, 1415-1421.

Das, A.T., Brummelkamp, T.R., Westerhout, E.M., Vink, M., Madiredjo, M., Bernards, R., and Berkhout, B. (2004). Human immunodeficiency virus type 1 escapes from RNA interference-mediated inhibition. J Virol *78*, 2601-2605.

Das, S.K., Sokhi, U.K., Bhutia, S.K., Azab, B., Su, Z.Z., Sarkar, D., and Fisher, P.B. (2010). Human polynucleotide phosphorylase selectively and preferentially degrades microRNA-221 in human melanoma cells. Proc Natl Acad Sci U S A *107*, 11948-11953.

Davis, B.M., Roth, J.C., Liu, L., Xu-Welliver, M., Pegg, A.E., and Gerson, S.L. (1999). Characterization of the P140K, PVP(138-140)MLK, and G156A O6-methylguanine-DNA methyltransferase mutants: implications for drug resistance gene therapy. Hum Gene Ther *10*, 2769-2778.

De, N., Young, L., Lau, P.W., Meisner, N.C., Morrissey, D.V., and MacRae, I.J. (2013). Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. Mol Cell *50*, 344-355.

de la Mata, M., Gaidatzis, D., Vitanescu, M., Stadler, M.B., Wentzel, C., Scheiffele, P., Filipowicz, W., and Grosshans, H. (2015). Potent degradation of neuronal miRNAs induced by highly complementary targets. EMBO reports *16*, 500-511.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M.*, et al.* (1996). Identification of a major co-receptor for primary isolates of HIV-1. Nature *381*, 661-666.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Denton, P.W., and Garcia, J.V. (2011). Humanized mouse models of HIV infection. AIDS Rev *13*, 135-148.

Di Carlo, V., Grossi, E., Laneve, P., Morlando, M., Dini Modigliani, S., Ballarino, M., Bozzoni, I., and Caffarelli, E. (2013). TDP-43 regulates the microprocessor complex activity during in vitro neuronal differentiation. Molecular neurobiology *48*, 952-963.

Dickson, J.R., Kruse, C., Montagna, D.R., Finsen, B., and Wolfe, M.S. (2013). Alternative polyadenylation and miR-34 family members regulate tau expression. J Neurochem *127*, 739-749.

Diebold, S.S., Kaisho, T., Hemmi, H., Akira, S., and Reis e Sousa, C. (2004). Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. Science *303*, 1529-1531.

Diederichs, S., and Haber, D.A. (2007). Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell *131*, 1097-1108.

Diederichs, S., Jung, S., Rothenberg, S.M., Smolen, G.A., Mlody, B.G., and Haber, D.A. (2008). Coexpression of Argonaute-2 enhances RNA interference toward perfect match binding sites. Proc Natl Acad Sci U S A *105*, 9284-9289.

DiGiusto, D.L., Krishnan, A., Li, L., Li, H., Li, S., Rao, A., Mi, S., Yam, P., Stinson, S., Kalos, M.*, et al.* (2010). RNA-based gene therapy for HIV with lentiviral vector-modified CD34(+) cells in patients undergoing transplantation for AIDS-related lymphoma. Sci Transl Med *2*, 36ra43.

Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev *17*, 438-442.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev *18*, 504-511.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nature reviews Molecular cell biology *17*, 5-15.

Donald C. Rio, M.A., Gregory Hannon, Timothy Nilsen (2011). RNA: A Laboratory Manual (Cold Spring Harbour, New York, USA: Cold Spring Harbour Laboratory Press).

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P.*, et al.* (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature *381*, 667-673.

Duclair, S., Gautam, A., Ellington, A., and Prasad, V.R. (2015). High-affinity RNA Aptamers Against the HIV-1 Protease Inhibit Both In Vitro Protease Activity and Late Events of Viral Replication. Molecular therapy Nucleic acids *4*, e228.

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

Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nature methods *4*, 721-726.

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

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

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001c). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

Elbing, K., and Brent, R. (2002). Media preparation and bacteriological tools. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 1*, Unit 1 1.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100-110.

Ely, A., Naidoo, T., and Arbuthnot, P. (2009). Efficient silencing of gene expression with modular trimeric Pol II expression cassettes comprising microRNA shuttles. Nucleic Acids Res *37*, e91.

Ely, A., Naidoo, T., Mufamadi, S., Crowther, C., and Arbuthnot, P. (2008). Expressed anti-HBV primary microRNA shuttles inhibit viral replication efficiently in vitro and in vivo. Mol Ther *16*, 1105-1112.

EMA, E.M.A. (2016). Committee for Medicinal Products for Human Use (CHMP): Summary of opinion for Strimvelis. In EMA/CHMP/160482/2016.

Ender, C., Krek, A., Friedlander, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. Mol Cell *32*, 519-528.

Engelman, A., and Cherepanov, P. (2012). The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol *10*, 279-290.

Eulalio, A., Behm-Ansmant, I., and Izaurralde, E. (2007a). P bodies: at the crossroads of post-transcriptional pathways. Nature reviews Molecular cell biology *8*, 9-22.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007b). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes Dev *21*, 2558-2570.

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

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A.*, et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

Faller, M., Matsunaga, M., Yin, S., Loo, J.A., and Guo, F. (2007). Heme is involved in microRNA processing. Nat Struct Mol Biol *14*, 23-29.

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

Fang, W., and Bartel, D.P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. Mol Cell *60*, 131-145.

Filippov, V., Solovyev, V., Filippova, M., and Gill, S.S. (2000). A novel type of RNase III family proteins in eukaryotes. Gene *245*, 213-221.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503-514.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Fish, R.J., and Kruithof, E.K. (2004). Short-term cytotoxic effects and long-term instability of RNAi delivered using lentiviral vectors. BMC Mol Biol *5*, 9.

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell *155*, 807-816.

Flicek, P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T.*, et al.* (2008). Ensembl 2008. Nucleic Acids Res *36*, D707-714.

Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D., and Lai, E.C. (2010). MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol Cell *38*, 900-907.

Forman, J.J., Legesse-Miller, A., and Coller, H.A. (2008). A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proc Natl Acad Sci U S A *105*, 14879-14884.

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

Freed, E.O. (2001). HIV-1 replication. Somat Cell Mol Genet *26*, 13-33.

Freed, E.O. (2015). HIV-1 assembly, release and maturation. Nat Rev Microbiol *13*, 484-496.

Friedlander, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res *40*, 37-52.

Fromm, M., and Berg, P. (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. Molecular and cellular biology *3*, 991-999.

Fu, X.Y., Kessler, D.S., Veals, S.A., Levy, D.E., and Darnell, J.E., Jr. (1990). ISGF3, the transcriptional activator induced by interferon alpha, consists of multiple interacting polypeptide chains. Proc Natl Acad Sci U S A *87*, 8555-8559.

Gallagher, S.R., and Desjardins, P.R. (2006). Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Appendix 3*, Appendix 3D.

Gartner, S., Markovits, P., Markovitz, D.M., Kaplan, M.H., Gallo, R.C., and Popovic, M. (1986). The role of mononuclear phagocytes in HTLV-III/LAV infection. Science *233*, 215-219.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z.*, et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Giering, J.C., Grimm, D., Storm, T.A., and Kay, M.A. (2008). Expression of shRNA from a tissue-specific pol II promoter is an effective and safe RNAi therapeutic. Mol Ther *16*, 1630-1636.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Good, P.D., Krikos, A.J., Li, S.X., Bertrand, E., Lee, N.S., Giver, L., Ellington, A., Zaia, J.A., Rossi, J.J., and Engelke, D.R. (1997). Expression of small, therapeutic RNAs in human cell nuclei. Gene Ther *4*, 45-54.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. Cell *123*, 631-640.

Gregory, R.I., Yan, K.-p., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res *32*, D109-111.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res *36*, D154-D158.

Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature *441*, 537-541.

Grimm, D., Wang, L., Lee, J.S., Schurmann, N., Gu, S., Borner, K., Storm, T.A., and Kay, M.A. (2010). Argonaute proteins are key determinants of RNAi efficacy, toxicity, and persistence in the adult mouse liver. J Clin Invest *120*, 3106-3119.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.

Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. Genes Dev *20*, 1709-1714.

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

Guil, S., and Caceres, J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat Struct Mol Biol *14*, 591-596.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

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

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

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Han, B.W., Hung, J.H., Weng, Z., Zamore, P.D., and Ameres, S.L. (2011). The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. Current biology : CB *21*, 1878-1887.

Han, B.W., Wang, W., Li, C., Weng, Z., and Zamore, P.D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science *348*, 817-821.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary miRNA processing. Genes Dev *18*, 3016-3027.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., and Kim, V.N. (2006). Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. Cell *125*, 887-901.

Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.K., Yeom, K.H., Yang, W.Y., Haussler, D., Blelloch, R., and Kim, V.N. (2009). Posttranscriptional crossregulation between Drosha and DGCR8. Cell *136*, 75-84.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384-388.

Hansen, T.B., Veno, M.T., Jensen, T.I., Schaefer, A., Damgaard, C.K., and Kjems, J. (2016). Argonaute-associated short introns are a novel class of gene regulators. Nat Commun *7*, 11538.

Haqqi, T., Zhao, X., Panciu, A., Yadav, S. P. (2002). Sequencing in the presence of betaine: improvement in sequencing of the localised repeat sequence regions. Journal of Biomolecular Techniques *13*, 265-271.

Hartmann, R., Norby, P.L., Martensen, P.M., Jorgensen, P., James, M.C., Jacobsen, C., Moestrup, S.K., Clemens, M.J., and Justesen, J. (1998). Activation of 2'-5' oligoadenylate synthetase by single-stranded and double-stranded RNA aptamers. J Biol Chem *273*, 3236-3246.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., and Kay, M.A. (2010). Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA *16*, 673-695.

Havens, M.A., Reich, A.A., Duelli, D.M., and Hastings, M.L. (2012). Biogenesis of mammalian microRNAs by a non-canonical processing pathway. Nucleic Acids Res *40*, 4626-4640.

Hawkins, P.G., Santoso, S., Adams, C., Anest, V., and Morris, K.V. (2009). Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. Nucleic Acids Res *37*, 2984-2995.

He, L., He, X., Lim, L.P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D*., et al.* (2007). A microRNA component of the p53 tumour suppressor network. Nature *447*, 1130-1134.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J*., et al.* (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828-833.

He, Y., and Smith, R. (2009). Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. Cellular and molecular life sciences : CMLS *66*, 1239-1256.

Heidel, J.D., Hu, S., Liu, X.F., Triche, T.J., and Davis, M.E. (2004). Lack of interferon response in animals to naked siRNAs. Nat Biotechnol *22*, 1579-1582.

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S. (2004). Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. Science *303*, 1526-1529.

Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H., and Kim, V.N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell *151*, 521-532.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell *32*, 276-284.

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

Herbert, K.M., Pimienta, G., DeGregorio, S.J., Alexandrov, A., and Steitz, J.A. (2013). Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile. Cell reports *5*, 1070-1081.

Herbert, K.M., Sarkar, S.K., Mills, M., Delgado De la Herran, H.C., Neuman, K.C., and Steitz, J.A. (2015). A heterotrimer model of the complete Microprocessor complex revealed by single-molecule subunit counting. RNA.

Hermeking, H. (2010). The miR-34 family in cancer and apoptosis. Cell Death Differ *17*, 193-199.

Hong, E.J., Villen, J., Gerace, E.L., Gygi, S.P., and Moazed, D. (2005). A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. RNA Biol *2*, 106-111.

Horman, S.R., Janas, M.M., Litterst, C., Wang, B., MacRae, I.J., Sever, M.J., Morrissey, D.V., Graves, P., Luo, B., Umesalma, S.*, et al.* (2013). Akt-mediated phosphorylation of argonaute 2 downregulates cleavage and upregulates translational repression of MicroRNA targets. Mol Cell *50*, 356-367.

Hornung, V., Guenthner-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., Noronha, A., Manoharan, M., Akira, S., de Fougerolles, A.*, et al.* (2005). Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. Nat Med *11*, 263-270.

Horvitz, H.R., Brenner, S., Hodgkin, J., and Herman, R.K. (1979). A uniform genetic nomenclature for the nematode Caenorhabditis elegans. Mol Gen Genet *175*, 129-133.

Hovnanian, A., Rebouillat, D., Mattei, M.G., Levy, E.R., Marie, I., Monaco, A.P., and Hovanessian, A.G. (1998). The human 2',5'-oligoadenylate synthetase locus is composed of three distinct genes clustered on chromosome 12q24.2 encoding the 100-, 69-, and 40-kDa forms. Genomics *52*, 267-277.

Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., and Sellers, W.R. (2004). A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucleic Acids Res *32*, 893-901.

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., Luo, B., Alvarez-Carbonell, D., Garcia-Mesa, Y., Karn, J.*, et al.* (2014). RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. Proc Natl Acad Sci U S A *111*, 11461-11466.

Huang, H., Gao, Q., Peng, X., Choi, S.Y., Sarma, K., Ren, H., Morris, A.J., and Frohman, M.A. (2011). piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. Dev Cell *20*, 376-387.

Huang, V., Zheng, J., Qi, Z., Wang, J., Place, R.F., Yu, J., Li, H., and Li, L.C. (2013). Ago1 Interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cancer cells. PLoS genetics *9*, e1003821.

Huang, X., and Jia, Z. (2013). Construction of HCC-targeting artificial miRNAs using natural miRNA precursors. Experimental and therapeutic medicine *6*, 209-215.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nature reviews Genetics *12*, 99-110.

Huntzinger, E., Kuzuoglu-Ozturk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2013). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res *41*, 978-994.

Hutter, G., Nowak, D., Mossner, M., Ganepola, S., Mussig, A., Allers, K., Schneider, T., Hofmann, J., Kucherer, C., Blau, O.*, et al.* (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med *360*, 692-698.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

Hwang, H.W., Wentzel, E.A., and Mendell, J.T. (2007). A hexanucleotide element directs microRNA nuclear import. Science *315*, 97-100.

Ipsaro, J.J., Haase, A.D., Knott, S.R., Joshua-Tor, L., and Hannon, G.J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature *491*, 279-283.

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

Iwasaki, Y.W., Kiga, K., Kayo, H., Fukuda-Yuzawa, Y., Weise, J., Inada, T., Tomita, M., Ishihama, Y., and Fukao, T. (2013). Global microRNA elevation by inducible Exportin 5 regulates cell cycle entry. RNA *19*, 490-497.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol *21*, 635-637.

Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K.*, et al.* (2006a). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. RNA *12*, 1197-1205.

Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. (2006b). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA *12*, 1179-1187.

Jacque, J.M., Triques, K., and Stevenson, M. (2002). Modulation of HIV-1 replication by RNA interference. Nature *418*, 435-438.

Janas, M.M., Khaled, M., Schubert, S., Bernstein, J.G., Golan, D., Veguilla, R.A., Fisher, D.E., Shomron, N., Levy, C., and Novina, C.D. (2011). Feed-forward microprocessing and splicing activities at a microRNA-containing intron. PLoS genetics *7*, e1002330.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res *36*, W5-9.

Jones, D., Li, Y., He, Y., Xu, Z., Chen, H., and Min, W. (2012). Mirtron microRNA-1236 inhibits VEGFR-3 signaling during inflammatory lymphangiogenesis. Arterioscler Thromb Vasc Biol *32*, 633-642.

Joshi, P., and Prasad, V.R. (2002). Potent inhibition of human immunodeficiency virus type 1 replication by template analog reverse transcriptase inhibitors derived by SELEX (systematic evolution of ligands by exponential enrichment). J Virol *76*, 6545-6557.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nature reviews Molecular cell biology *14*, 49-55.

Judge, A.D., Bola, G., Lee, A.C., and MacLachlan, I. (2006). Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. Mol Ther *13*, 494-505.

Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K., and MacLachlan, I. (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. Nat Biotechnol *23*, 457-462.

Juvvuna, P.K., Khandelia, P., Lee, L.M., and Makeyev, E.V. (2012). Argonaute identity defines the length of mature mammalian microRNAs. Nucleic Acids Res *40*, 6808-6820.

Kalantari, R., Chiang, C.M., and Corey, D.R. (2016). Regulation of mammalian transcription and splicing by Nuclear RNAi. Nucleic Acids Res *44*, 524-537.

Kariko, K., Bhuyan, P., Capodici, J., and Weissman, D. (2004). Small interfering RNAs mediate sequence-independent gene suppression and induce immune activation by signaling through toll-like receptor 3. J Immunol *172*, 6545-6549.

Kataoka, N., Fujita, M., and Ohno, M. (2009). Functional association of the Microprocessor complex with the spliceosome. Molecular and cellular biology *29*, 3243-3254.

Katoh, T., Hojo, H., and Suzuki, T. (2015). Destabilization of microRNAs in human cells by 3' deadenylation mediated by PARN and CUGBP1. Nucleic Acids Res *43*, 7521-7534.

Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R., and Nishikura, K. (2007a). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO reports *8*, 763-769.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007b). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137-1140.

Kawamata, T., and Tomari, Y. (2010). Making RISC. Trends Biochem Sci *35*, 368-376.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793-797.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature *431*, 211-217.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-1026.

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

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

Khan, A.A., Betel, D., Miller, M.L., Sander, C., Leslie, C.S., and Marks, D.S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. Nat Biotechnol *27*, 549-555.

Khati, M., Schuman, M., Ibrahim, J., Sattentau, Q., Gordon, S., and James, W. (2003). Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2'F-RNA aptamers. J Virol *77*, 12692-12698.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kim, D.H., Longo, M., Han, Y., Lundberg, P., Cantin, E., and Rossi, J.J. (2004). Interferon induction by siRNAs and ssRNAs synthesized by phage polymerase. Nat Biotechnol *22*, 321-325.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nature reviews Molecular cell biology *10*, 126-139.

Kim, Y.K., and Kim, V.N. (2007). Processing of intronic microRNAs. EMBO J *26*, 775-783.

Kimpton, J., and Emerman, M. (1992). Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene. J Virol *66*, 2232-2239.

Kloosterman, W.P., Wienholds, E., Ketting, R.F., and Plasterk, R.H. (2004). Substrate requirements for let-7 function in the developing zebrafish embryo. Nucleic Acids Res *32*, 6284-6291.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Kock, K.H., Kong, K.W., Hoon, S., and Seow, Y. (2015). Functional VEGFA knockdown with artificial 3'-tailed mirtrons defined by 5' splice site and branch point. Nucleic Acids Res *43*, 6568-6578.

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

Konstantinova, P., de Vries, W., Haasnoot, J., ter Brake, O., de Haan, P., and Berkhout, B. (2006). Inhibition of human immunodeficiency virus type 1 by RNA interference using long-hairpin RNA. Gene Ther *13*, 1403-1413.

Kordelas, L., Verheyen, J., Beelen, D.W., Horn, P.A., Heinold, A., Kaiser, R., Trenschel, R., Schadendorf, D., Dittmer, U., Esser, S.*, et al.* (2014). Shift of HIV tropism in stem-cell transplantation with CCR5 Delta32 mutation. N Engl J Med *371*, 880-882.

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

Krell, J., Stebbing, J., Carissimi, C., Dabrowska, A., de Giorgio, A., Frampton, A.E., Harding, V., Fulci, V., Macino, G., Colombo, T.*, et al.* (2015). TP53 regulates miRNA association with AGO2 to remodel the miRNA-mRNA interaction network. Genome research.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nature reviews Genetics *11*, 597-610.

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

Kwon, S.C., Nguyen, T.A., Choi, Y.G., Jo, M.H., Hohng, S., Kim, V.N., and Woo, J.S. (2015). Structure of Human DROSHA. Cell.

Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O., and Lai, E.C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. Genome research *22*, 1634-1645.

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

Lai, E.C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet *30*, 363-364.

Lai, E.C., Tam, B., and Rubin, G.M. (2005). Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. Genes Dev *19*, 1067-1080.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M.*, et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401-1414.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis. Current biology : CB *14*, 2162-2167.

Lapidot, T., Pflumio, F., Doedens, M., Murdoch, B., Williams, D.E., and Dick, J.E. (1992). Cytokine stimulation of multilineage hematopoiesis from immature human cells engrafted in SCID mice. Science *255*, 1137-1141.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Lau, P.W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B., and MacRae, I.J. (2012). The molecular architecture of human Dicer. Nat Struct Mol Biol *19*, 436-440.

Lee, C.G., McCarthy, S., Gruidl, M., Timme, C., and Yeatman, T.J. (2014). MicroRNA-147 induces a mesenchymal-to-epithelial transition (MET) and reverses EGFR inhibitor resistance. PLoS One *9*, e84597.

Lee, H.S., Seok, H., Lee, D.H., Ham, J., Lee, W., Youm, E.M., Yoo, J.S., Lee, Y.S., Jang, E.S., and Chi, S.W. (2015). Abasic pivot substitution harnesses target specificity of RNA interference. Nat Commun *6*, 10154.

Lee, H.Y., and Doudna, J.A. (2012). TRBP alters human precursor microRNA processing in vitro. RNA *18*, 2012-2019.

Lee, H.Y., Zhou, K., Smith, A.M., Noland, C.L., and Doudna, J.A. (2013). Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. Nucleic Acids Res *41*, 6568-6576.

Lee, M.T., Coburn, G.A., McClure, M.O., and Cullen, B.R. (2003a). Inhibition of human immunodeficiency virus type 1 replication in primary macrophages by using Tat- or CCR5-specific small interfering RNAs expressed from a lentivirus vector. J Virol *77*, 11964-11972.

Lee, N.S., Dohjima, T., Bauer, G., Li, H., Li, M.J., Ehsani, A., Salvaterra, P., and Rossi, J. (2002a). Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. Nat Biotechnol *20*, 500-505.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.

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

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

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. EMBO *25*, 522-532.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002b). MicroRNA maturation: stepwise processing and subcellular localization. EMBO *21*, 4663-4670.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO *23*, 4051-4060.

Leonard, J.N., and Schaffer, D.V. (2005). Computational design of antiviral RNA interference strategies that resist human immunodeficiency virus escape. J Virol *79*, 1645-1654.

Leuschner, P.J., Ameres, S.L., Kueng, S., and Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO reports *7*, 314-320.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, J., Kim, T., Nutiu, R., Ray, D., Hughes, T.R., and Zhang, Z. (2014a). Identifying mRNA sequence elements for target recognition by human Argonaute proteins. Genome research *24*, 775-785.

Li, L., Lin, X., Khvorova, A., Fesik, S.W., and Shen, Y. (2007). Defining the optimal parameters for hairpin-based knockdown constructs. RNA *13*, 1765-1774.

Li, M. (2003). Inhibition of HIV-1 infection by lentiviral vectors expressing pol III-promoted anti-HIV RNAs. Mol Ther *8*, 196-206.

Li, M.J., Kim, J., Li, S., Zaia, J., Yee, J.K., Anderson, J., Akkina, R., and Rossi, J.J. (2005). Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy. Mol Ther *12*, 900-909.

Li, T., Wu, M., Zhu, Y.Y., Chen, J., and Chen, L. (2014b). Development of RNA Interference-Based Therapeutics and Application of Multi-Target Small Interfering RNAs. Nucleic acid therapeutics *24*, 302-312.

Liang, Y., Ridzon, D., Wong, L., and Chen, C. (2007). Characterization of microRNA expression profiles in normal human tissues. BMC Genomics *8*, 166.

Liebecq, C. (1997). IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB). Biochemistry and molecular biology international *43*, 1151-1156.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of Caenorhabditis elegans. Genes Dev *17*, 991-1008.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development *124*, 2463-2476.

Lin, X., Ruan, X., Anderson, M.G., McDowell, J.A., Kroeger, P.E., Fesik, S.W., and Shen, Y. (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. Nucleic Acids Res *33*, 4527-4535.

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

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

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, N., Abe, M., Sabin, L.R., Hendriks, G.J., Naqvi, A.S., Yu, Z., Cherry, S., and Bonini, N.M. (2011). The exoribonuclease Nibbler controls 3' end processing of microRNAs in Drosophila. Current biology : CB *21*, 1888-1893.

Liu, N., Landreh, M., Cao, K., Abe, M., Hendriks, G.J., Kennerdell, J.R., Zhu, Y., Wang, L.S., and Bonini, N.M. (2012a). The microRNA miR-34 modulates ageing and neurodegeneration in Drosophila. Nature *482*, 519-523.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., and Landau, N.R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell *86*, 367-377.

Liu, W.M., Pang, R.T., Chiu, P.C., Wong, B.P., Lao, K., Lee, K.F., and Yeung, W.S. (2012b). Sperm-borne microRNA-34c is required for the first cleavage division in mouse. Proc Natl Acad Sci U S A *109*, 490-494.

Liu, X., Jin, D.Y., McManus, M.T., and Mourelatos, Z. (2012c). Precursor microRNA-programmed silencing complex assembly pathways in mammals. Mol Cell *46*, 507-517.

Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V., and Liu, Q. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science *325*, 750-753.

Liu, Y.P., and Berkhout, B. (2011). miRNA cassettes in viral vectors: problems and solutions. Biochimica et biophysica acta *1809*, 732-745.

Liu, Y.P., Haasnoot, J., and Berkhout, B. (2007). Design of extended short hairpin RNAs for HIV-1 inhibition. Nucleic Acids Res *35*, 5683-5693.

Liu, Y.P., Haasnoot, J., ter Brake, O., Berkhout, B., and Konstantinova, P. (2008). Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. Nulceic Acids Res *36*, 2811-2824.

Liu, Y.P., Schopman, N.C., and Berkhout, B. (2013). Dicer-independent processing of short hairpin RNAs. Nucleic Acids Res *41*, 3723-3733.

Liu, Y.P., von Eije, K.J., Schopman, N.C., Westerink, J.T., ter Brake, O., Haasnoot, J., and Berkhout, B. (2009). Combinatorial RNAi against HIV-1 using extended short hairpin RNAs. Mol Ther *17*, 1712-1723.

Liu, Z., Wang, J., Li, G., and Wang, H.W. (2015). Structure of precursor microRNA's terminal loop regulates human Dicer's dicing activity by switching DExH/D domain. Protein Cell *6*, 185-193.

Lo, H.L., Chang, T., Yam, P., Marcovecchio, P.M., Li, S., Zaia, J.A., and Yee, J.K. (2007). Inhibition of HIV-1 replication with designed miRNAs expressed from RNA polymerase II promoters. Gene Ther *14*, 1503-1512.

Lucas, K., and Raikhel, A.S. (2013). Insect microRNAs: biogenesis, expression profiling and biological functions. Insect Biochem Mol Biol *43*, 24-38.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.

Ma, H., Wu, Y., Choi, J.G., and Wu, H. (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. Proc Natl Acad Sci U S A *110*, 20687-20692.

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

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

Macpherson, J.L., Boyd, M.P., Arndt, A.J., Todd, A.V., Fanning, G.C., Ely, J.A., Elliott, F., Knop, A., Raponi, M., Murray, J.*, et al.* (2005). Long-term survival and concomitant gene expression of ribozyme-transduced CD4+ T-lymphocytes in HIV-infected patients. The journal of gene medicine *7*, 552-564.

MacRae, I.J., Ma, E., Zhou, M., Robinson, C.V., and Doudna, J.A. (2008). In vitro reconstitution of the human RISC-loading complex. Proc Natl Acad Sci U S A *105*, 512-517.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Maeder, M.L., and Gersbach, C.A. (2016). Genome-editing Technologies for Gene and Cell Therapy. Mol Ther *24*, 430-446.

Malim, M.H., Bohnlein, S., Hauber, J., and Cullen, B.R. (1989). Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function. Cell *58*, 205-214.

Manche, L., Green, S.R., Schmedt, C., and Mathews, M.B. (1992). Interactions between double-stranded RNA regulators and the protein kinase DAI. Molecular and cellular biology *12*, 5238-5248.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes Dev *19*, 2979-2990.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002a). Sinlge-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

Martinez, M.A., Gutierrez, A., Armand-Ugon, M., Blanco, J., Parera, M., Gomez, J., Clotet, B., and Este, J.A. (2002b). Suppression of chemokine receptor expression by RNA interference allows for inhibition of HIV-1 replication. AIDS *16*, 2385-2390.

Martinez, N.J., and Gregory, R.I. (2013). Argonaute2 expression is post-transcriptionally coupled to microRNA abundance. RNA *19*, 605-612.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nature reviews Molecular cell biology *8*, 209-220.

Mathews, D.H., Sabina, J., Zuker, M & Turner, D. H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Imporves Prediction of RNA Secondary Structure. Journal of molecular biology *288*, 911-940.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Maunders, M.J. (1993a). Alkaline phosphatase (EC 3.1.3.1). Methods in molecular biology *16*, 331-341.

Maunders, M.J. (1993b). Polynucleotide kinase (EC 2.7.1.78). Methods in molecular biology *16*, 343-356.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc Natl Acad Sci U S A *74*, 560-564.

McBride, J.L., Boudreau, R.L., Harper, S.Q., Staber, P.D., Monteys, A.M., Martins, I., Gilmore, B.L., Burstein, H., Peluso, R.W., Polisky, B.*, et al.* (2008). Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. Proc Natl Acad Sci U S A *105*, 5868-5873.

McCune, J.M., Namikawa, R., Kaneshima, H., Shultz, L.D., Lieberman, M., and Weissman, I.L. (1988). The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function. Science *241*, 1632-1639.

McIntyre, G.J., Groneman, J.L., Yu, Y.H., Jaramillo, A., Shen, S., and Applegate, T.L. (2009a). 96 shRNAs designed for maximal coverage of HIV-1 variants. Retrovirology *6*, 55.

McIntyre, G.J., Groneman, J.L., Yu, Y.H., Tran, A., and Applegate, T.L. (2011). Multiple shRNA combinations for near-complete coverage of all HIV-1 strains. AIDS Res Ther *8*, 1.

McIntyre, G.J., Yu, Y.H., Tran, A., Jaramillo, A.B., Arndt, A.J., Millington, M.L., Boyd, M.P., Elliott, F.A., Shen, S.W., Murray, J.M.*, et al.* (2009b). Cassette deletion in multiple shRNA lentiviral vectors for HIV-1 and its impact on treatment success. Virology journal *6*, 184.

McManus, M.T., Petersen, C.P., Haines, B.B., Chen, J., and Sharp, P.A. (2002). Gene silencing using micro-RNA designed hairpins. RNA *8*, 842-850.

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

Melamed, Z., Levy, A., Ashwal-Fluss, R., Lev-Maor, G., Mekahel, K., Atias, N., Gilad, S., Sharan, R., Levy, C., Kadener, S.*, et al.* (2013). Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. Mol Cell *50*, 869-881.

Melkus, M.W., Estes, J.D., Padgett-Thomas, A., Gatlin, J., Denton, P.W., Othieno, F.A., Wege, A.K., Haase, A.T., and Garcia, J.V. (2006). Humanized mice mount specific adaptive and innate immune responses to EBV and TSST-1. Nat Med *12*, 1316-1322.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M.*, et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333-338.

Michienzi, A., Li, S., Zaia, J.A., and Rossi, J.J. (2002). A nucleolar TAR decoy inhibitor of HIV-1 replication. Proc Natl Acad Sci U S A *99*, 14047-14052.

Minks, M.A., West, D.K., Benvin, S., and Baglioni, C. (1979). Structural requirements of double-stranded RNA for the activation of 2',5'-oligo(A) polymerase and protein kinase of interferon-treated HeLa cells. J Biol Chem *254*, 10180-10183.

Mitsuyasu, R.T., Merigan, T.C., Carr, A., Zack, J.A., Winters, M.A., Workman, C., Bloch, M., Lalezari, J., Becker, S., Thornton, L.*, et al.* (2009). Phase 2 gene therapy trial of an anti-HIV ribozyme in autologous CD34+ cells. Nat Med *15*, 285-292.

Mitxelena, J., Apraiz, A., Vallejo-Rodriguez, J., Malumbres, M., and Zubiaga, A.M. (2016). E2F7 regulates transcription and maturation of multiple microRNAs to restrain cell proliferation. Nucleic Acids Res.

Miyagishi, M., Sumimoto, H., Miyoshi, H., Kawakami, Y., and Taira, K. (2004). Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. The journal of gene medicine *6*, 715-723.

Miyagishi, M., and Taira, K. (2002). U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. Nat Biotechnol *20*, 497-500.

Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H., and Verma, I.M. (1998). Development of a self-inactivating lentivirus vector. J Virol *72*, 8150-8157.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

Mockenhaupt, S., Grosse, S., Rupp, D., Bartenschlager, R., and Grimm, D. (2015). Alleviation of off-target effects from vector-encoded shRNAs via codelivered RNA decoys. Proc Natl Acad Sci U S A *112*, E4007-4016.

Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science *348*, 812-817.

Monteys, A.M., Spengler, R.M., Wan, J., Tecedor, L., Lennox, K.A., Xing, Y., and Davidson, B.L. (2010). Structure and activity of putative intronic miRNA promoters. RNA *16*, 495-505.

Monteys, A.M., Wilson, M.J., Boudreau, R.L., Spengler, R.M., and Davidson, B.L. (2015). Artificial miRNAs Targeting Mutant Huntingtin Show Preferential Silencing In Vitro and In Vivo. Molecular therapy Nucleic acids *4*, e234.

Moore, D.D. (2001). Commonly used reagents and equipment. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Appendix 2*, Appendix 2.

Moore, D.D., and Dowhan, D. (2002). Preparation and analysis of DNA. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 2*, 2.1.1 -2.1.10.

Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. Nat Struct Mol Biol *15*, 902-909.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science *305*, 1289-1292.

Mosier, D.E., Gulizia, R.J., Baird, S.M., and Wilson, D.B. (1988). Transfer of a functional human immune system to mice with severe combined immunodeficiency. Nature *335*, 256-259.

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

Mowa, M.B., Crowther, C., and Arbuthnot, P. (2010). Therapeutic potential of adenoviral vectors for delivery of expressed RNAi activators. Expert Opin Drug Deliv *7*, 1373-1385.

Mueller, G.A., Miller, M.T., Derose, E.F., Ghosh, M., London, R.E., and Hall, T.M. (2010). Solution structure of the Drosha double-stranded RNA-binding domain. Silence *1*, 2.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology *51 Pt 1*, 263-273.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263-267.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. The Plant cell *2*, 279-289.

Neilsen, C.T., Goodall, G.J., and Bracken, C.P. (2012). IsomiRs--the overlooked repertoire in the dynamic microRNAome. Trends in genetics : TIG *28*, 544-549.

Nepal, C., Coolen, M., Hadzhiev, Y., Cussigh, D., Mydel, P., Steen, V.M., Carninci, P., Andersen, J.B., Bally-Cuif, L., Muller, F.*, et al.* (2015). Transcriptional, post-transcriptional and chromatin-associated regulation of pri-miRNAs, pre-miRNAs and moRNAs. Nucleic Acids Res.

Newman, M.A., Mani, V., and Hammond, S.M. (2011). Deep sequencing of microRNA precursors reveals extensive 3' end modification. RNA *17*, 1795-1803.

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

Nhlabatsi, N. (2014). Comparing the silencing efficacy of Dicer-independent and dependent shRNAs. In Department of Molecular Medicine and Haematology (Johannesburg: University of the Witwatersrand).

Nielsen, C.B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., and Burge, C.B. (2007). Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA *13*, 1894-1910.

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M.K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J.*, et al.* (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature *491*, 284-287.

Nishitsuji, H., Kohara, M., Kannagi, M., and Masuda, T. (2006). Effective Suppression of Human Immunodeficiency Virus Type 1 through a Combination of Short- or Long-Hairpin RNAs Targeting Essential Sequences for Retroviral Integration. J Virol *80*, 7658-7666.

Noguer-Dance, M., Abu-Amero, S., Al-Khtib, M., Lefevre, A., Coullin, P., Moore, G.E., and Cavaille, J. (2010). The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. Hum Mol Genet *19*, 3566-3582.

Noland, C.L., and Doudna, J.A. (2013). Multiple sensors ensure guide strand selection in human RNAi pathways. RNA *19*, 639-648.

Novina, C.D., Murray, M.F., Dykxhoorn, D.M., Beresford, P.J., Riess, J., Lee, S.K., Collman, R.G., Lieberman, J., Shankar, P., and Sharp, P.A. (2002). siRNA-directed inhibition of HIV-1 infection. Nat Med *8*, 681-686.

O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V., and Mendell, J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. Nature *435*, 839-843.

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

Okamura, K. (2013). Argonaute reformatting. Mol Cell *50*, 305-306.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89-100.

Okamura, K., Liu, N., and Lai, E.C. (2009). Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. Mol Cell *36*, 431-444.

Ono, M., Yamada, K., Avolio, F., Scott, M.S., van Koningsbruggen, S., Barton, G.J., and Lamond, A.I. (2010). Analysis of human small nucleolar RNAs (snoRNA) and the development of snoRNA modulator of gene expression vectors. Mol Biol Cell *21*, 1569-1584.

Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S., and Fisher, D.E. (2008). Chromatin structure analyses identify miRNA promoters. Genes Dev *22*, 3172-3183.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev *16*, 948-958.

Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature *475*, 201-205.

Park, J.H., and Shin, C. (2015). Slicer-independent mechanism drives small-RNA strand separation during human RISC assembly. Nucleic Acids Res *43*, 9418-9433.

Parker, J.S., Roe, S.M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J *23*, 4727-4737.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

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

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P.*, et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86-89.

Passman, M., Weinberg, M., Kew, M., and Arbuthnot, P. (2000). In situ demonstration of inhibitory effects of hammerhead ribozymes that are targeted to the hepatitis Bx sequence in cultured cells. Biochem Biophys Res Commun *268*, 728-733.

Paul, C.P., Good, P.D., Winer, I., and Engelke, D.R. (2002). Effective expression of small interfering RNA in human cells. Nat Biotechnol *20*, 505-508.

Pawlicki, J.M., and Steitz, J.A. (2008). Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. J Cell Biol *182*, 61-76.

Pearson, L., Garcia, J., Wu, F., Modesti, N., Nelson, J., and Gaynor, R. (1990). A transdominant tat mutant that inhibits tat-induced gene expression from the human immunodeficiency virus long terminal repeat. Proc Natl Acad Sci U S A *87*, 5079-5083.

Pebernard, S., and Iggo, R.D. (2004). Determinants of interferon-stimulated gene induction by RNAi vectors. Differentiation *72*, 103-111.

Peng, Z., Cheng, Y., Tan, B.C., Kang, L., Tian, Z., Zhu, Y., Zhang, W., Liang, Y., Hu, X., Tan, X.*, et al.* (2012). Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol *30*, 253-260.

Pickering, B.F., Yu, D., and Van Dyke, M.W. (2011). Nucleolin protein interacts with microprocessor complex to affect biogenesis of microRNAs 15a and 16. J Biol Chem *286*, 44095-44103.

Pillai, R.S., Artus, C.G., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. RNA *10*, 1518-1525.

Player, M.R., and Torrence, P.F. (1998). The 2-5A system: modulation of viral and cellular processes through acceleration of RNA degradation. Pharmacol Ther *78*, 55-113.

Ponten, J., and Macintyre, E.H. (1968). Long term culture of normal and neoplastic human glia. Acta Pathol Microbiol Scand *74*, 465-486.

Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science *224*, 497-500.

Qi, H., Watanabe, T., Ku, H.Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. J Biol Chem *286*, 3789-3797.

Qi, H.H., Ongusaha, P.P., Myllyharju, J., Cheng, D., Pakkanen, O., Shi, Y., Lee, S.W., Peng, J., and Shi, Y. (2008). Prolyl 4-hydroxylation regulates Argonaute 2 stability. Nature *455*, 421-424.

Quarles, K.A., Chadalavada, D., and Showalter, S.A. (2015). Deformability in the cleavage site of primary microRNA is not sensed by the double-stranded RNA binding domains in the microprocessor component DGCR8. Proteins *83*, 1165-1179.

Quarles, K.A., Sahu, D., Havens, M.A., Forsyth, E.R., Wostenberg, C., Hastings, M.L., and Showalter, S.A. (2013). Ensemble analysis of primary microRNA structure reveals an extensive capacity to deform near the Drosha cleavage site. Biochemistry *52*, 795-807.

Quick-Cleveland, J., Jacob, J.P., Weitz, S.H., Shoffner, G., Senturia, R., and Guo, F. (2014). The DGCR8 RNA-binding heme domain recognizes primary microRNAs by clamping the hairpin. Cell reports *7*, 1994-2005.

Ramachandran, P.S., Bhattarai, S., Singh, P., Boudreau, R.L., Thompson, S., Laspada, A.R., Drack, A.V., and Davidson, B.L. (2014). RNA interference-based therapy for spinocerebellar ataxia type 7 retinal degeneration. PLoS One *9*, e95362.

Ramalingam, P., Palanichamy, J.K., Singh, A., Das, P., Bhagat, M., Kassab, M.A., Sinha, S., and Chattopadhyay, P. (2014). Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. RNA *20*, 76-87.

Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621-629.

Rao, D.D., Maples, P.B., Senzer, N., Kumar, P., Wang, Z., Pappen, B.O., Yu, Y., Haddock, C., Jay, C., Phadke, A.P*., et al.* (2010). Enhanced target gene knockdown by a bifunctional shRNA: a novel approach of RNA interference. Cancer gene therapy *17*, 780-791.

Reimao-Pinto, M.M., Ignatova, V., Burkard, T.R., Hung, J.H., Manzenreither, R.A., Sowemimo, I., Herzog, V.A., Reichholf, B., Farina-Lopez, S., and Ameres, S.L. (2015). Uridylation of RNA Hairpins by Tailor Confines the Emergence of MicroRNAs in Drosophila. Mol Cell *59*, 203-216.

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

Reynolds, A., Anderson, E.M., Vermeulen, A., Fedorov, Y., Robinson, K., Leake, D., Karpilow, J., Marshall, W.S., and Khvorova, A. (2006). Induction of the interferon response by siRNA is cell type- and duplex length-dependent. RNA *12*, 988-993.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., and Khvorova, A. (2004). Rational siRNA design for RNA interference. Nat Biotechnol *22*, 326-330.

Ringpis, G.E., Shimizu, S., Arokium, H., Camba-Colon, J., Carroll, M.V., Cortado, R., Xie, Y., Kim, P.Y., Sahakyan, A., Lowe, E.L*., et al.* (2012). Engineering HIV-1-resistant T-cells from short-hairpin RNA-expressing hematopoietic stem/progenitor cells in humanized BLT mice. PLoS One *7*, e53492.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol *12*, 340-349.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. Genome research *14*, 1902-1910.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Rossi, J.J., June, C.H., and Khon, D.B. (2007). Genetic therapies against HIV. Nat Biotechnol *25*, 1444-1453.

Roth, B.M., Ishimaru, D., and Hennig, M. (2013). The core microprocessor component DiGeorge syndrome critical region 8 (DGCR8) is a nonspecific RNA-binding protein. J Biol Chem *288*, 26785-26799.

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

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

Rudel, S., Wang, Y., Lenobel, R., Korner, R., Hsiao, H.H., Urlaub, H., Patel, D., and Meister, G. (2011). Phosphorylation of human Argonaute proteins affects small RNA binding. Nucleic Acids Res *39*, 2330-2343.

Ruegger, S., and Grosshans, H. (2012). MicroRNA turnover: when, how, and why. Trends Biochem Sci *37*, 436-446.

Rybak, A., Fuchs, H., Hadian, K., Smirnova, L., Wulczyn, E.A., Michel, G., Nitsch, R., Krappmann, D., and Wulczyn, F.G. (2009). The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. Nat Cell Biol *11*, 1411-1420.

Saayman, S., Arbuthnot, P., and Weinberg, M.S. (2010). Deriving four functional anti-HIV siRNAs from a single Pol III-generated transcript comprising two adjacent long hairpin RNA precursors. Nucleic Acids Res *38*, 6652-6663.

Saayman, S., Barichievy, S., Capovilla, A., Morris, K.V., Arbuthnot, P., and Weinberg, M.S. (2008). The Efficacy of Generating Three Independent Anti-HIV-1 siRNAs from a Single U6 RNA Pol III-Expressed Long Hairpin RNA. PLoS ONE *3*, e2602.

Saayman, S.M., Lazar, D.C., Scott, T.A., Hart, J.R., Takahashi, M., Burnett, J.C., Planelles, V., Morris, K.V., and Weinberg, M.S. (2016). Potent and Targeted Activation of Latent HIV-1 Using the CRISPR/dCas9 Activator Complex. Mol Ther *24*, 488-498.

Sadler, A.J., and Williams, B.R. (2008). Interferon-inducible antiviral effectors. Nature reviews Immunology *8*, 559-568.

Saini, H.K., Enright, A.J., and Griffiths-Jones, S. (2008). Annotation of mammalian primary microRNAs. BMC Genomics *9*, 564.

Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T., and Agata, Y. (2009). The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. Molecular and cellular biology *29*, 3754-3769.

Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C.*, et al.* (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature *382*, 722-725.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

Sano, M., Li, H., Nakanishi, M., and Rossi, J.J. (2008). Expression of long anti-HIV-1 hairpin RNAs for the generation of multiple siRNAs: advantages and limitations. Mol Ther *16*, 170-177.

Saraiya, A.A., and Wang, C.C. (2008). snoRNA, a novel precursor of microRNA in Giardia lamblia. PLoS Pathog *4*, e1000224.

Sarver, N., Cantin, E.M., Chang, P.S., Zaia, J.A., Ladne, P.A., Stephens, D.A., and Rossi, J.J. (1990). Ribozymes as potential anti-HIV-1 therapeutic agents. Science *247*, 1222-1225.

Sasaki, T., Shiohama, A., Minoshima, S., and Shimizu, N. (2003). Identification of eight members of the Argonaute family in the human genome. Genomics *82*, 323-330.

Saxena, S., Jonsson, Z.O., and Dutta, A. (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. J Biol Chem *278*, 44312-44319.

Schamberger, A., Sarkadi, B., and Orban, T.I. (2012). Human mirtrons can express functional microRNAs simultaneously from both arms in a flanking exon-independent manner. RNA Biol *9*, 1177-1185.

Scherer, L., Rossi, J.J., and Weinberg, M.S. (2007a). Progress and prospects: RNA-based therapies for treatment of HIV infection. Gene Ther *14*, 1057-1064.

Scherer, L.J., Frank, R., and Rossi, J.J. (2007b). Optimization and characterization of tRNA-shRNA expression constructs. Nucleic Acids Res *35*, 2620-2628.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

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

Schmidt, E.V., Christoph, G., Zeller, R., and Leder, P. (1990). The cytomegalovirus enhancer: a pan-active control element in transgenic mice. Molecular and cellular biology *10*, 4406-4411.

Schoggins, J.W., Wilson, S.J., Panis, M., Murphy, M.Y., Jones, C.T., Bieniasz, P., and Rice, C.M. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. Nature *472*, 481-485.

Schopman, N.C., ter Brake, O., and Berkhout, B. (2010). Anticipating and blocking HIV-1 escape by second generation antiviral shRNAs. Retrovirology *7*, 52.

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

Schwarz, D.S., Hutvagner, G., Haley, B., and Zamore, P.D. (2002). Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol Cell *10*, 537-548.

Schwarz, D.S., Tomari, Y., and Zamore, P.D. (2004). The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Current biology : CB *14*, 787-791.

Senturia, R., Faller, M., Yin, S., Loo, J.A., Cascio, D., Sawaya, M.R., Hwang, D., Clubb, R.T., and Guo, F. (2010). Structure of the dimerization domain of DiGeorge critical region 8. Protein science : a publication of the Protein Society *19*, 1354-1365.

Senzer, N., Barve, M., Kuhn, J., Melnyk, A., Beitsch, P., Lazar, M., Lifshitz, S., Magee, M., Oh, J., Mill, S.W.*, et al.* (2012). Phase I trial of "bi-shRNAi(furin)/GMCSF DNA/autologous tumor cell" vaccine (FANG) in advanced cancer. Mol Ther *20*, 679-686.

Seow, Y., Sibley, C.R., and Wood, M.J. (2012). Artificial mirtron-mediated gene knockdown: functional DMPK silencing in mammalian cells. RNA *18*, 1328-1337.

Shen, J., Xia, W., Khotskaya, Y.B., Huo, L., Nakanishi, K., Lim, S.O., Du, Y., Wang, Y., Chang, W.C., Chen, C.H.*, et al.* (2013). EGFR modulates microRNA maturation in response to hypoxia through phosphorylation of AGO2. Nature *497*, 383-387.

Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2003). Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. Biochem Biophys Res Commun *304*, 184-190.

Shultz, L.D., Brehm, M.A., Garcia-Martinez, J.V., and Greiner, D.L. (2012). Humanized mice for immune system investigation: progress, promise and challenges. Nature reviews Immunology *12*, 786-798.

Shultz, L.D., Lyons, B.L., Burzenski, L.M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S.D., King, M., Mangada, J.*, et al.* (2005). Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. J Immunol *174*, 6477-6489.

Sibley, C.R., Seow, Y., Curtis, H., Weinberg, M.S., and Wood, M.J. (2012a). Silencing of Parkinson's disease-associated genes with artificial mirtron mimics of miR-1224. Nucleic Acids Res *40*, 9863-9875.

Sibley, C.R., Seow, Y., Saayman, S., Dijkstra, K.K., El Andaloussi, S., Weinberg, M.S., and Wood, M.J. (2012b). The biogenesis and characterization of mammalian microRNAs of mirtron origin. Nucleic Acids Res *40*, 438-448.

Silva, J.M., Li, M.Z., Chang, K., Ge, W., C., G.M., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R.*, et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nat Genet *37*, 1281-1288.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology *12*, 246-258.

Sioud, M. (2005). Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization. Journal of molecular biology *348*, 1079-1090.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell *5*, 659-669.

Slatko, B.E., and Albright, L.M. (2001). Denaturing gel electrophoresis for sequencing. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 7*, Unit7 6.

Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., and Williams, B.R. (2003). Activation of the interferon system by short-interfering RNAs. Nat Cell Biol *5*, 834-839.

Slezak-Prochazka, I., Kluiver, J., de Jong, D., Kortman, G., Halsema, N., Poppema, S., Kroesen, B.J., and van den Berg, A. (2013). Cellular localization and processing of primary transcripts of exonic microRNAs. PLoS ONE *8*, e76647.

Smibert, P., Yang, J.S., Azzam, G., Liu, J.L., and Lai, E.C. (2013). Homeostatic control of Argonaute stability by microRNA availability. Nat Struct Mol Biol *20*, 789-795.

Sohn, S.Y., Bae, W.J., Kim, J.J., Yeom, K.-H., Kim, V.N., and Cho, Y. (2007). Crystal structure of human DGCR8 core. Nat Struct Mol Biol *14*, 847-853.

Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Starega-Roslan, J., Galka-Marciniak, P., and Krzyzosiak, W.J. (2015a). Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer. Nucleic Acids Res.

Starega-Roslan, J., Krol, J., Koscianska, E., Kozlowski, P., Szlachcic, W.J., Sobczak, K., and Krzyzosiak, W.J. (2011). Structural basis of microRNA length variety. Nucleic Acids Res *39*, 257-268.

Starega-Roslan, J., Witkos, T.M., Galka-Marciniak, P., and Krzyzosiak, W.J. (2015b). Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. Int J Mol Sci *16*, 8110-8127.

Stark, G.R., Kerr, I.M., Williams, B.R., Silverman, R.H., and Schreiber, R.D. (1998). How cells respond to interferons. Annu Rev Biochem *67*, 227-264.

Stegmeier, F., Hu, G., Rickles, R.J., Hannon, G.J., and Elledge, S.J. (2005). A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. Proc Natl Acad Sci U S A *102*, 13212-13217.

Stein, P., Rozhkov, N.V., Li, F., Cardenas, F.L., Davydenko, O., Vandivier, L.E., Gregory, B.D., Hannon, G.J., and Schultz, R.M. (2015). Essential Role for endogenous siRNAs during meiosis in mouse oocytes. PLoS genetics *11*, e1005013.

Steiner, F.A., Hoogstrate, S.W., Okihara, K.L., Thijssen, K.L., Ketting, R.F., Plasterk, R.H., and Sijen, T. (2007). Structural features of small RNA precursors determine Argonaute loading in Caenorhabditis elegans. Nat Struct Mol Biol *14*, 927-933.

Stothard, P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques *28*, 1102-1104.

Sui, G., Soohoo, C., Affar el, B., Gay, F., Shi, Y., Forrester, W.C., and Shi, Y. (2002). A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. Proc Natl Acad Sci U S A *99*, 5515-5520.

Sun, L.Q., Wang, L., Gerlach, W.L., and Symonds, G. (1995). Target sequence-specific inhibition of HIV-1 replication by ribozymes directed to tat RNA. Nucleic Acids Res *23*, 2909-2913.

Sundaram, G.M., Common, J.E., Gopal, F.E., Srikanta, S., Lakshman, K., Lunny, D.P., Lim, T.C., Tanavde, V., Lane, E.B., and Sampath, P. (2013). 'See-saw' expression of microRNA-198 and FSTL1 from a single transcript in wound healing. Nature *495*, 103-106.

Sung, T.-L., and Rice, A.P. (2009). miR-198 Inhibits HIV-1 Gene Expression and Replication in Monocytes and Its Mechanism of Action Appears To Involve Repression of Cyclin T1. PLoS Pathog *5*, e1000263.

Surabhi, R.M., and Gaynor, R.B. (2002). RNA interference directed against viral and cellular targets inhibits human immunodeficiency Virus Type 1 replication. J Virol *76*, 12963-12973.

Suzuki, H.I., Yamagata, K., Sugimoto, K., Iwamoto, T., Kato, S., and Miyazono, K. (2009). Modulation of microRNA processing by p53. Nature *460*, 529-533.

Svoboda, P., Stein, P., Hayashi, H., and Schultz, R.M. (2000). Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. Development *127*, 4147-4156.

Swamy, M.N., Wu, H., and Shankar, P. (2016). Recent advances in RNAi-based strategies for Therapy and Prevention of HIV-1/AIDS. Adv Drug Deliv Rev.

Sylvestre, Y., De Guire, V., Querido, E., Mukhopadhyay, U.K., Bourdeau, V., Major, F., Ferbeyre, G., and Chartrand, P. (2007). An E2F/miR-20a autoregulatory feedback loop. J Biol Chem *282*, 2135-2143.

Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. RNA *15*, 1233-1240.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M.*, et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

Tan, S., Li, R., Ding, K., Lobie, P.E., and Zhu, T. (2011). miR-198 inhibits migration and invasion of hepatocellular carcinoma cells by targeting the HGF/c-MET pathway. FEBS Letters *585*, 2229-2234.

Tang, X., Li, M., Tucker, L., and Ramratnam, B. (2011). Glycogen synthase kinase 3 beta (GSK3beta) phosphorylates the RNAase III enzyme Drosha at S300 and S302. PLoS One *6*, e20391.

Tang, X., Wen, S., Zheng, D., Tucker, L., Cao, L., Pantazatos, D., Moss, S.F., and Ramratnam, B. (2013). Acetylation of drosha on the N-terminus inhibits its degradation by ubiquitination. PLoS ONE *8*, e72503.

Taylor, D.W., Ma, E., Shigematsu, H., Cianfrocco, M.A., Noland, C.L., Nagayama, K., Nogales, E., Doudna, J.A., and Wang, H.W. (2013). Substrate-specific structural rearrangements of human Dicer. Nat Struct Mol Biol *20*, 662-670.

TenoRes Study, G. (2016). Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. Lancet Infect Dis.

ter Brake, O., Konstantinova, P., Ceylan, M., and Berkhout, B. (2006). Silencing of HIV-1 with RNA Interference: A Multiple shRNA Approach. Mol Ther *14*, 883-892.

ter Brake, O., t Hooft, K., Liu, Y.P., Centlivre, M., von Eije, K.J., and Berkhout, B. (2008). Lentiviral vector design for multiple shRNA expression and durable HIV-1 inhibition. Mol Ther *16*, 557-564.

Thomson, D.W., and Dinger, M.E. (2016). Endogenous microRNA sponges: evidence and controversy. Nature reviews Genetics *17*, 272-283.

Tian, Y., Simanshu, D.K., Ma, J.B., Park, J.E., Heo, I., Kim, V.N., and Patel, D.J. (2014). A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. Mol Cell *53*, 606-616.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M.G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature *459*, 1010-1014.

Trono, D., Feinberg, M.B., and Baltimore, D. (1989). HIV-1 Gag mutants can dominantly interfere with the replication of the wild-type virus. Cell *59*, 113-120.

Truscott, M., Islam, A.B., and Frolov, M.V. (2016). Novel regulation and functional interaction of polycistronic miRNAs. RNA *22*, 129-138.

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

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Tuerk, C., and MacDougal-Waugh, S. (1993). *In vitro* evolution of functional nucleic acids: high-affinity RNA ligands of HIV-1 proteins. Gene *137*, 33-39.

Uhlmann, S., Mannsperger, H., Zhang, J.D., Horvat, E.A., Schmidt, C., Kublbeck, M., Henjes, F., Ward, A., Tschulena, U., Zweig, K.*, et al.* (2012). Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. Mol Syst Biol *8*, 570.

Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R., and Saigo, K. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. Nucleic Acids Res *32*, 936-948.

Unwalla, H.J., Li, H., Li, S.Y., Abad, D., and Rossi, J.J. (2008). Use of a U16 snoRNA-containing ribozyme library to identify ribozyme targets in HIV-1. Mol Ther *16*, 1113-1119.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature reviews Genetics *11*, 636-646.

Ustek, D., Sirma, S., Gumus, E., Arikan, M., Cakiris, A., Abaci, N., Mathew, J., Emrence, Z., Azakli, H., Cosan, F.*, et al.* (2012). A genome-wide analysis of lentivector integration sites using targeted sequence capture and next generation sequencing technology. Infect Genet Evol *12*, 1349-1354.

Valoczi, A., Hornyik, C., Varga, N., Burgyan, J., Kauppinen, S., and Havelda, Z. (2004). Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res *32*, e175.

Varallyay, E., Burgyan, J., and Havelda, Z. (2008). MicroRNA detection by northern blotting using locked nucleic acid probes. Nat Protoc *3*, 190-196.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W.S., Karpilow, J., and Khvorova, A. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. Rna *11*, 674-682.

Vlassov, A.V., Korba, B., Farrar, K., Mukerjee, S., Seyhan, A.A., Ilves, H., Kaspar, R.L., Leake, D., Kazakov, S.A., and Johnston, B.H. (2007). shRNAs targeting hepatitis C: effects of sequence and structural features, and comparision with siRNA. Oligonucleotides *17*, 223-236.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669-687.

Voinnet, O., and Baulcombe, D.C. (1997). Systemic signalling in gene silencing. Nature *389*, 553.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. Cell *40*, 9-17.

Walker, J.E., Chen, R.X., McGee, J., Nacey, C., Pollard, R.B., Abedi, M., Bauer, G., Nolta, J.A., and Anderson, J.S. (2012). Generation of an HIV-1-resistant immune system with CD34(+) hematopoietic stem cells transduced with a triple-combination anti-HIV lentiviral vector. J Virol *86*, 5719-5729.

Wang, B., Li, S., Qi, H.H., Chowdhury, D., Shi, Y., and Novina, C.D. (2009a). Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. Nat Struct Mol Biol *16*, 1259-1266.

Wang, G., Zhao, N., Berkhout, B., and Das, A.T. (2016). CRISPR-Cas9 Can Inhibit HIV-1 Replication but NHEJ Repair Facilitates Virus Escape. Mol Ther *24*, 522-526.

Wang, H.-W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., and Nogales, E. (2009b). Structural insights into RNA processing by the human RISC-loading complex. Nat Struct Mol Biol *16*, 1148-1153.

Wang, L., Fu, C., Fan, H., Du, T., Dong, M., Chen, Y., Jin, Y., Zhou, Y., Deng, M., Gu, A.*, et al.* (2013). miR-34b regulates multiciliogenesis during organ formation in zebrafish. Development *140*, 2755-2764.

Wang, L., Witherington, C., King, A., Gerlach, W.L., Carr, A., Penny, R., Cooper, D., Symonds, G., and Sun, L.Q. (1998). Preclinical characterization of an anti-tat ribozyme for therapeutic application. Hum Gene Ther *9*, 1283-1291.

Wang, T., Xie, Y., Tan, A., Li, S., and Xie, Z. (2015). Construction and Characterization of a Synthetic MicroRNA Cluster for Multiplex RNA Interference in Mammalian Cells. ACS Synth Biol.

Wang, X., Xu, X., Ma, Z., Huo, Y., Xiao, Z., Li, Y., and Wang, Y. (2011). Dynamic mechanisms for pre-miRNA binding and export by Exportin-5. RNA *17*, 1511-1528.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Warf, M.B., Johnson, W.E., and Bass, B.L. (2011). Improved annotation of C. elegans microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. RNA *17*, 563-577.

Watanabe, T., and Lin, H. (2014). Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs. Mol Cell *56*, 18-27.

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev *20*, 1732-1743.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T*., et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Weber, B., Stresemann, C., Brueckner, B., and Lyko, F. (2007). Methylation of human microRNA genes in normal and neoplastic cells. Cell Cycle *6*, 1001-1005.

Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. RNA *12*, 256-262.

Wen, J., Ladewig, E., Shenker, S., Mohammed, J., and Lai, E.C. (2015). Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates. PLoS Comput Biol *11*, e1004441.

Westholm, J.O., Ladewig, E., Okamura, K., Robine, N., and Lai, E.C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. RNA *18*, 177-192.

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

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

Wirth, T., Parker, N., and Yla-Herttuala, S. (2013). History of gene therapy. Gene *525*, 162-169.

Wu, C., So, J., Davis-Dusenbery, B.N., Qi, H.H., Bloch, D.B., Shi, Y., Lagna, G., and Hata, A. (2011a). Hypoxia potentiates microRNA-mediated gene silencing through posttranslational modification of Argonaute2. Molecular and cellular biology *31*, 4760-4774.

Wu, H., Ma, H., Ye, C., Ramirez, D., Chen, S., Montoya, J., Shankar, P., Wang, X.A., and Manjunath, N. (2011b). Improved siRNA/shRNA functionality by mismatched duplex. PLoS One *6*, e28580.

Wu, H., Ye, C., Ramirez, D., and Manjunath, N. (2009). Alternative Processing of Primary microRNA Transcripts by Drosha Generates 5' End Variation of Mature microRNA. PLoS ONE *4*, e7566.

Wu, L., Fan, J., and Belasco, J.G. (2008). Importance of translation and nonnucleolytic ago proteins for on-target RNA interference. Current biology : CB *18*, 1327-1332.

Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J., and Tewari, M. (2011). Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome research *21*, 1450-1461.

Yam, P.Y., Li, S., Wu, J., Hu, J., Zaia, J.A., and Yee, J.K. (2002). Design of HIV vectors for efficient gene delivery into human hematopoietic cells. Mol Ther *5*, 479-484.

Yamada, O., Yu, M., Yee, J.K., Kraus, G., Looney, D., and Wong-Staal, F. (1994). Intracellular immunization of human T cells with a hairpin ribozyme against human immunodeficiency virus type 1. Gene Ther *1*, 38-45.

Yamakuchi, M., Ferlito, M., and Lowenstein, C.J. (2008). miR-34a repression of SIRT1 regulates apoptosis. Proc Natl Acad Sci U S A *105*, 13421-13426.

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

Yang, J.S., Maurin, T., and Lai, E.C. (2012). Functional parameters of Dicer-independent microRNA biogenesis. RNA *18*, 945-957.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010a). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A *107*, 15163-15168.

Yang, X., Haurigot, V., Zhou, S., Luo, G., and Couto, L.B. (2010b). Inhibition of hepatitis C virus replication using adeno-associated virus vector delivery of an exogenous anti-hepatitis C virus microRNA cluster. Hepatology *52*, 1877-1887.

Ye, L., Li, S., Ye, D., Yang, D., Yue, F., Guo, Y., Chen, X., Chen, F., Zhang, J., and Song, X. (2013). Livin expression may be regulated by miR-198 in human prostate cancer cell lines. European journal of cancer *49*, 734-740.

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

Yeom, K.H., Lee, Y., Han, J., Suh, M.R., and Kim, V.N. (2006). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res *34*, 4622-4629.

Yi, R., Doehle, B.P., Qin, Y., Macara, I.G., and Cullen, B.R. (2005). Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. RNA *11*, 220-226.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.

Yoda, M., Cifuentes, D., Izumi, N., Sakaguchi, Y., Suzuki, T., Giraldez, A.J., and Tomari, Y. (2013). Poly(A)-specific ribonuclease mediates 3'-end trimming of Argonaute2-cleaved precursor microRNAs. Cell reports *5*, 715-726.

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

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nature immunology *5*, 730-737.

Yu, J.Y., DeRuiter, S.L., and Turner, D.L. (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. Proc Natl Acad Sci U S A *99*, 6047-6052.

Yu, M., Ojwang, J., Yamada, O., Hampel, A., Rapapport, J., Looney, D., and Wong-Staal, F. (1993). A hairpin ribozyme inhibits expression of diverse strains of human immunodeficiency virus type 1. Proc Natl Acad Sci U S A *90*, 6340-6344.

Zamore, P.D. (2001). Thirty-three years later, a glimpse at the ribonuclease III active site. Mol Cell *8*, 1158-1160.

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

Zeng, Y., Cai, X., and Cullen, B.R. (2005a). Use of RNA polymerase II to transcribe artificial microRNAs. Methods Enzymol *392*, 371-380.

Zeng, Y., and Cullen, B.R. (2003). Sequence requirements for micro RNA processing and function in human cells. RNA *9*, 112-123.

Zeng, Y., and Cullen, B.R. (2004). Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res *32*, 4776-4785.

Zeng, Y., and Cullen, B.R. (2005). Efficient Processing of Primary microRNA Hairpins by Drosha Requires Flanking Nonstructured RNA Sequences. J Biol Chem *280*, 27595-27603.

Zeng, Y., Sankala, H., Zhang, X., and Graves, P.R. (2008). Phosphorylation of Argonaute 2 at serine-387 facilitates its localization to processing bodies. Biochem J *413*, 429-436.

Zeng, Y., Wagner, E.J., and Cullen, B.R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell *9*, 1327-1333.

Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A *100*, 9779-9784.

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

Zhang, B., Stellwag, E.J., and Pan, X. (2009). Large-scale genome analysis reveals unique features of microRNAs. Gene *443*, 100-109.

Zhang, H., Artiles, K.L., and Fire, A.Z. (2015). Functional relevance of "seed" and "non-seed" sequences in microRNA-mediated promotion of C. elegans developmental progression. RNA *21*, 1980-1992.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J *21*, 5875-5885.

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

Zhang, T., Cheng, T., Wei, L., Cai, Y., Yeo, A.E., Han, J., Yuan, Y.A., Zhang, J., and Xia, N. (2012). Efficient inhibition of HIV-1 replication by an artificial polycistronic miRNA construct. Virology journal *9*, 118.

Zhang, X., and Zeng, Y. (2010). The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res *38*, 7689-7697.

Zhou, C., Bahner, I.C., Larson, G.P., Zaia, J.A., Rossi, J.J., and Kohn, E.B. (1994). Inhibition of HIV-1 in human T-lymphocytes by retrovirally transduced anti-tat and rev hammerhead ribozymes. Gene *149*, 33-39.

Zhou, H., Arcila, M.L., Li, Z., Lee, E.J., Henzler, C., Liu, J., Rana, T.M., and Kosik, K.S. (2012). Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res *40*, 5864-5875.

Zhou, H., Xia, X.G., and Xu, Z. (2005). An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi. Nucleic Acids Res *33*, e62.

Zhou, J., Li, H., Li, S., Zaia, J., and Rossi, J.J. (2008). Novel dual inhibitory function aptamer-siRNA delivery system for HIV-1 therapy. Mol Ther *16*, 1481-1489.

Zhou, J., Satheesan, S., Li, H., Weinberg, M.S., Morris, K.V., Burnett, J.C., and Rossi, J.J. (2015). Cell-specific RNA aptamer against human CCR5 specifically targets HIV-1 susceptible cells and inhibits HIV-1 infectivity. Chem Biol *22*, 379-390.

Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L., and Trono, D. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J Virol *72*, 9873-9880.

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

APPENDIX A

PREDICTED SECONDARY RNA STRUCTURES GENERATED BY MFOLD

This appendix contains predicted secondary RNA structures generated by the mfold web server (version 3.5) (Zuker *et. al.*, 2003) for pri-miRNAs and derived mimics. Predicted structures were generated with 15 to 50 nt of flanking sequence (5 nt increments) and for the vector-derived RNA sequence. Only the most stable structures with 20 nt, 50 nt or vector-derived flanking sequences are shown here for each pri-miRNA or mimic (thumbnails) (**A1 - 7**). The predicted structures of the T7-expressed pri-miR-198 and pri-miR-34a sequences used in RNase structure mapping are also given (**A8 & 9**).The pre-miRNA in each structure is highlighted in green.

A-1 **Pri-miR-520b & derived anti-*tat*, -*int*, -*gag* & -LTR mimics**

A-2 **Pri-miR-198**

A-3 **Pri-miR-147a**

A-4 **Pri-miR-23a & derived anti-*tat*, -*int*, -*gag* & -LTR mimics cs**

A-5 **Pri-miR-934 & derived anti-*tat*, -*int*, -*gag* & -LTR mimics**

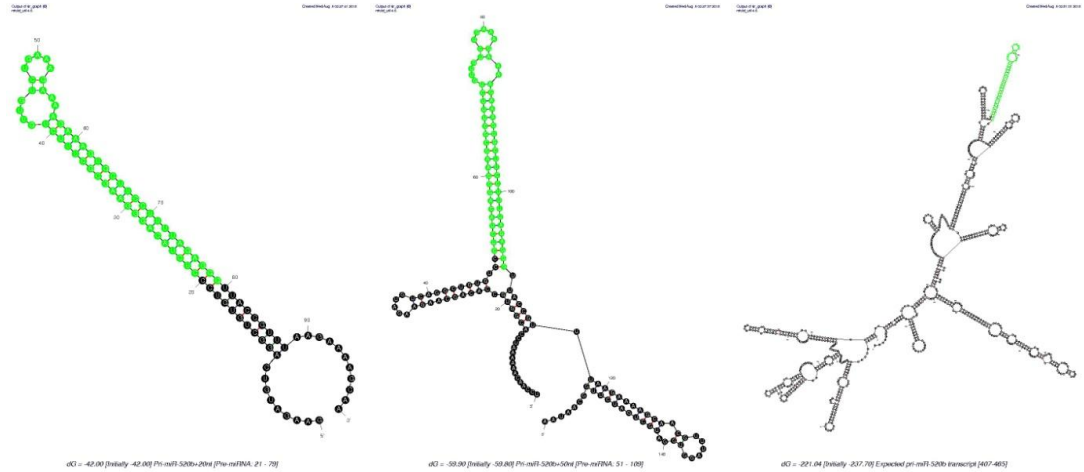
A-6 **Pri-miR-34a & derived anti-*tat*, -*int*, -*gag* & -LTR mimics**

A-7 **Pri-miR-513a-2 & derived anti-*tat*, -*int*, -*gag* & -LTR mimics**

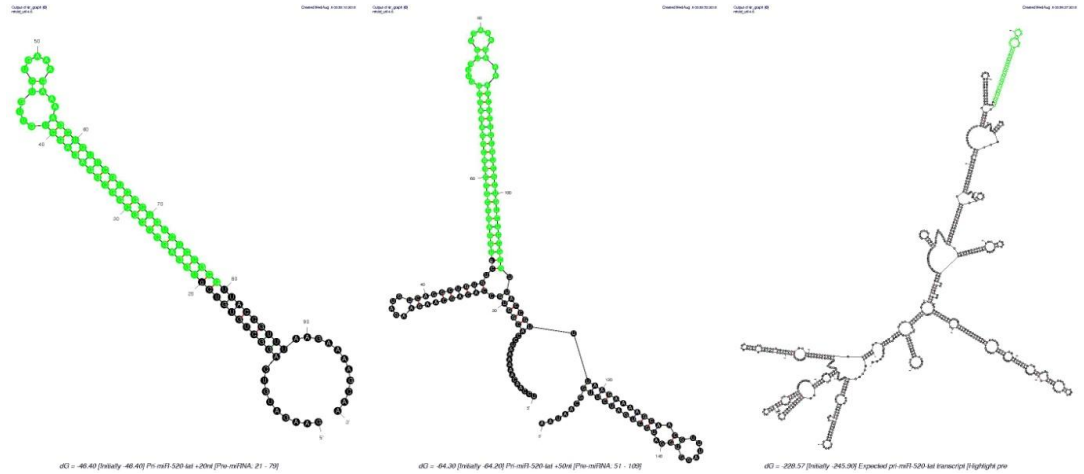
A-8 **T7-expressed pri-miR198**

A-9 **T7-expressed pri-miR-34a**

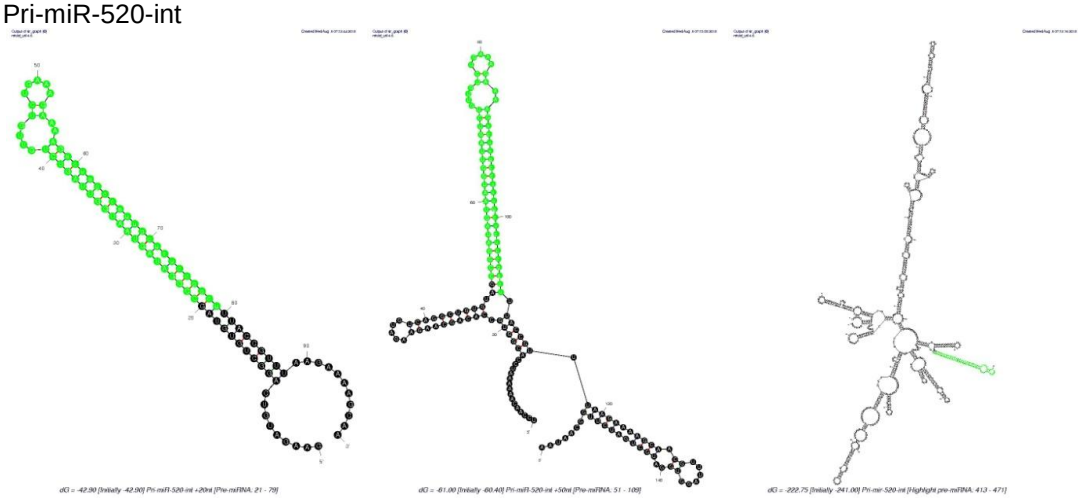
Pri-miR-520b



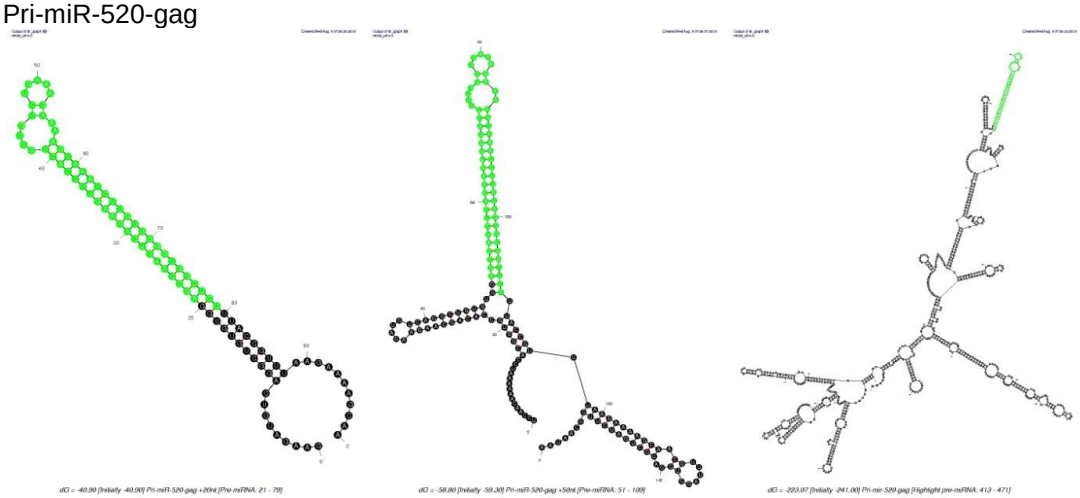
Pri-miR-520-tat



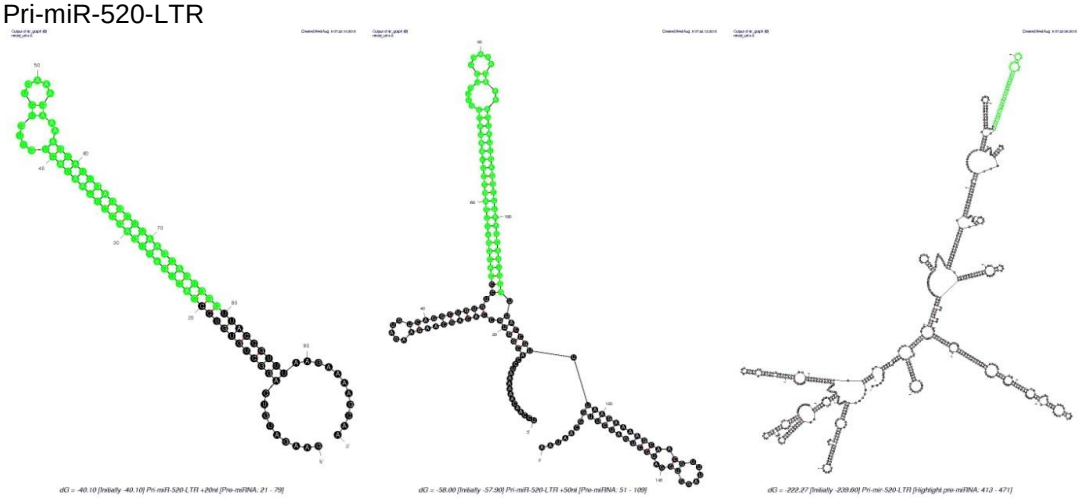
Pri-miR-520-int



Pri-miR-520-gag

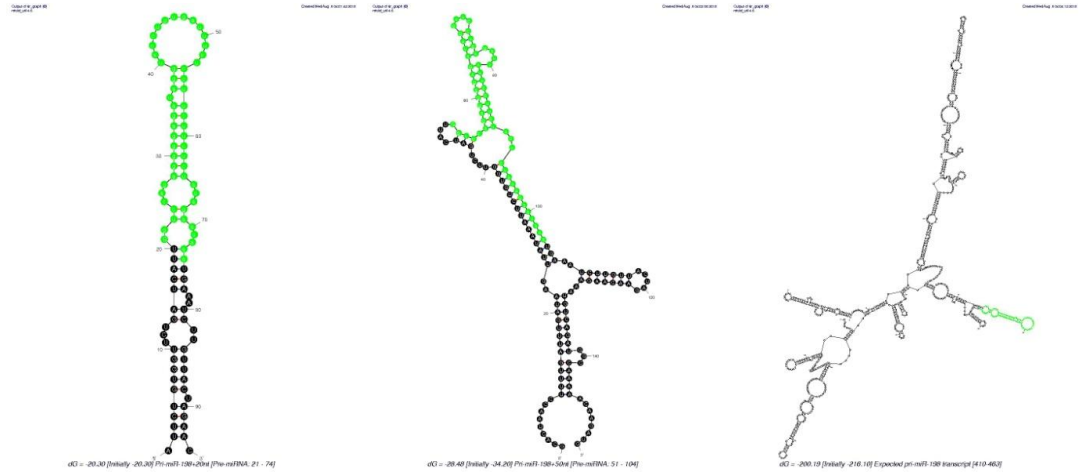


Pri-miR-520-LTR



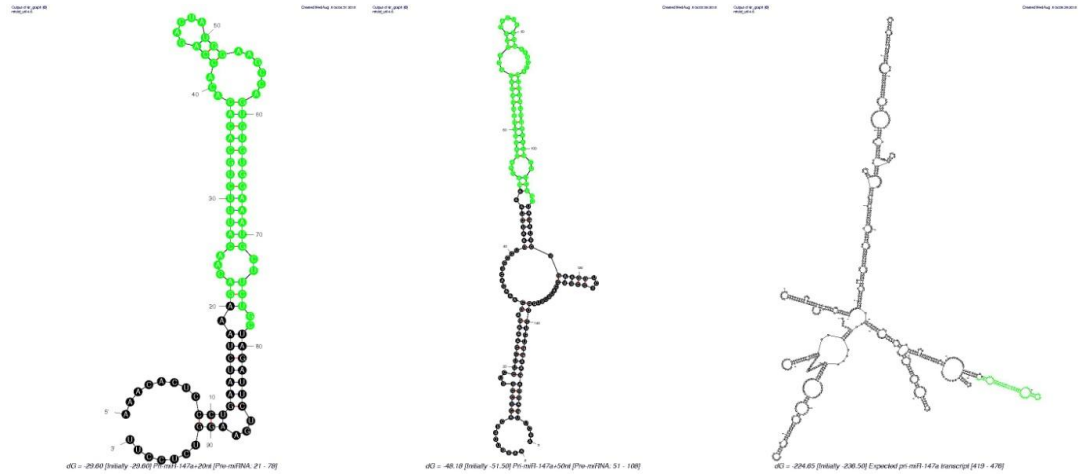
APPENDIX A-2

Pri-miR-198

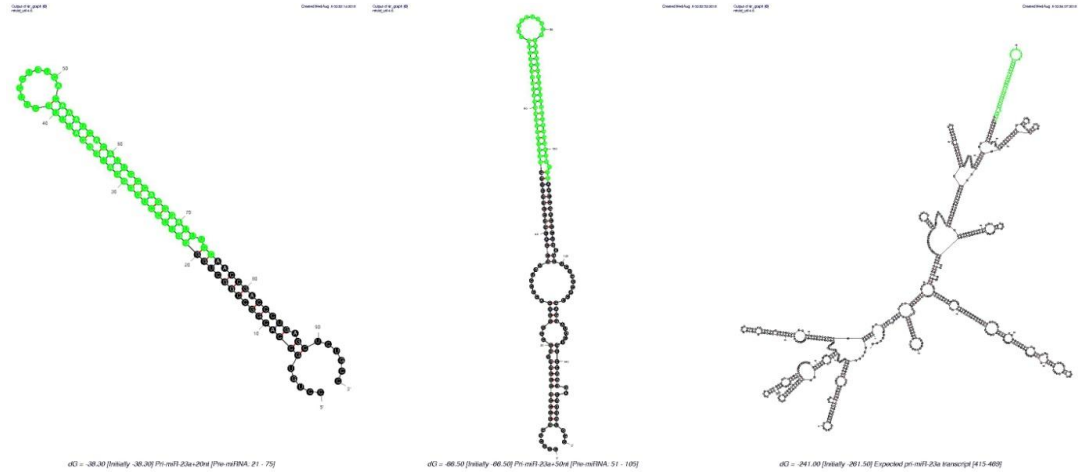


APPENDIX A-3

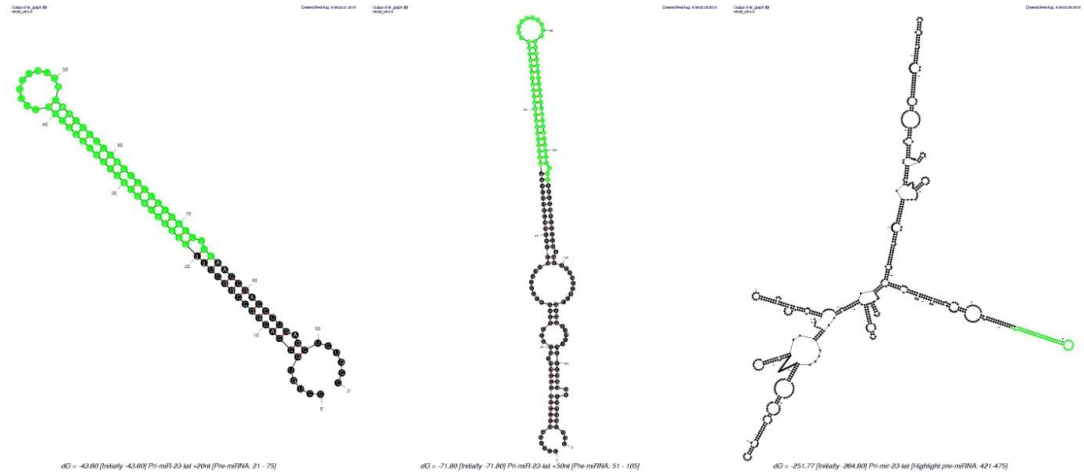
Pri-miR-147a



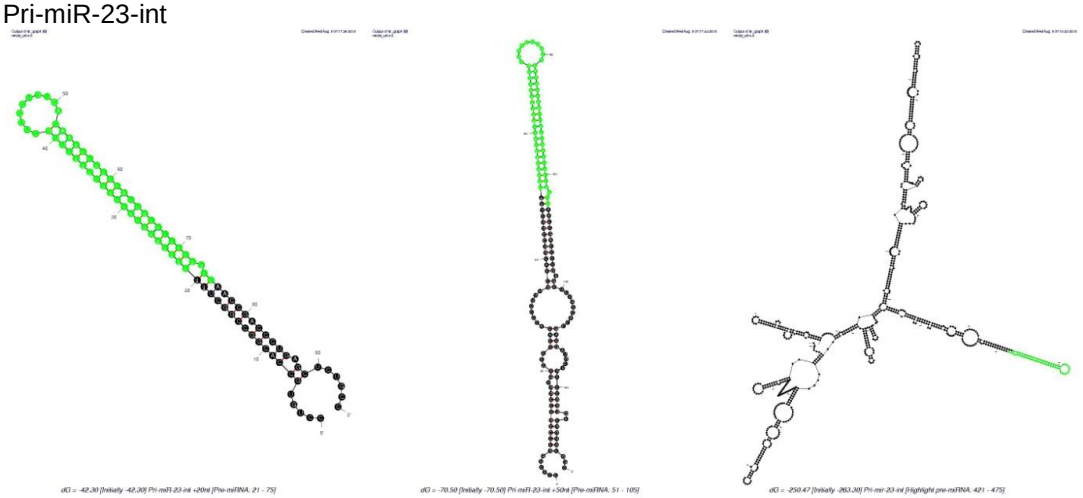
Pri-miR-23a



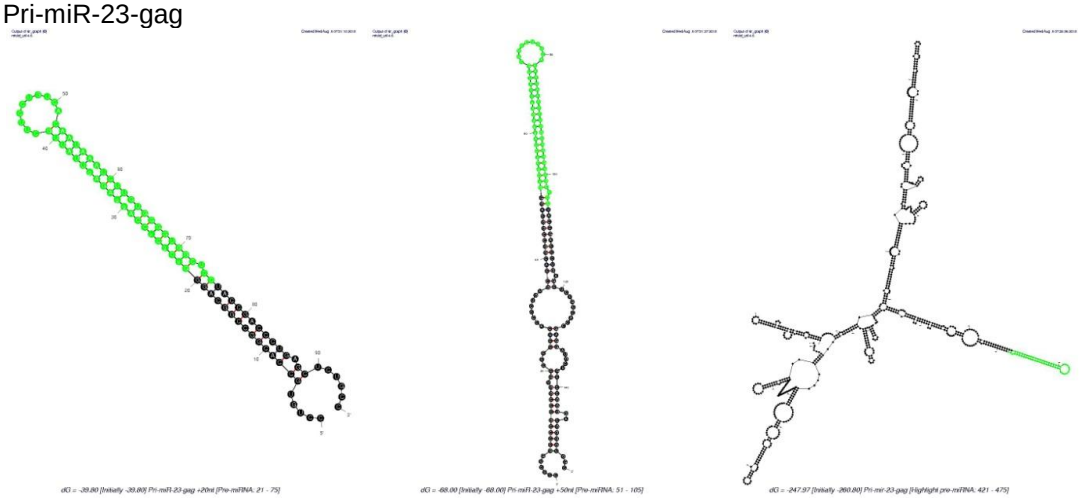
Pri-miR-23-tat



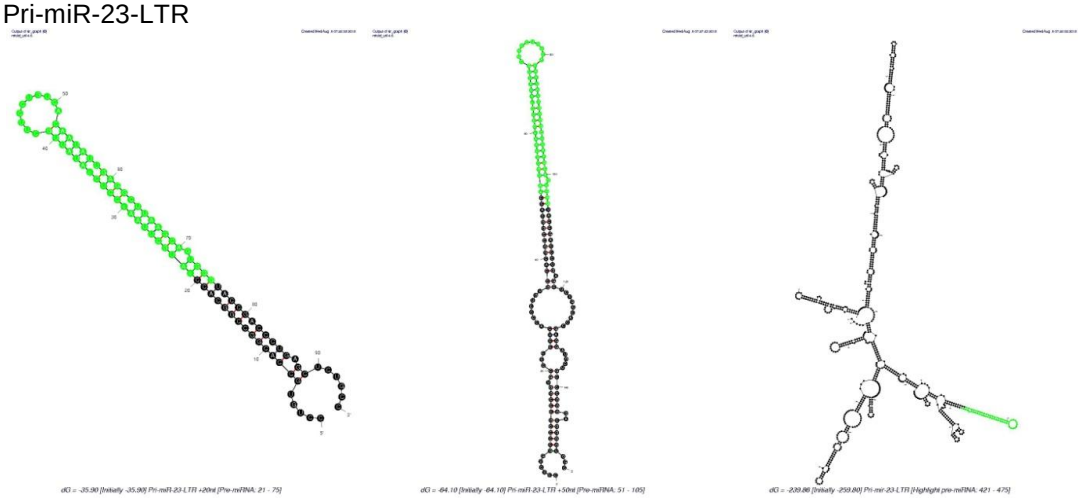
Pri-miR-23-int



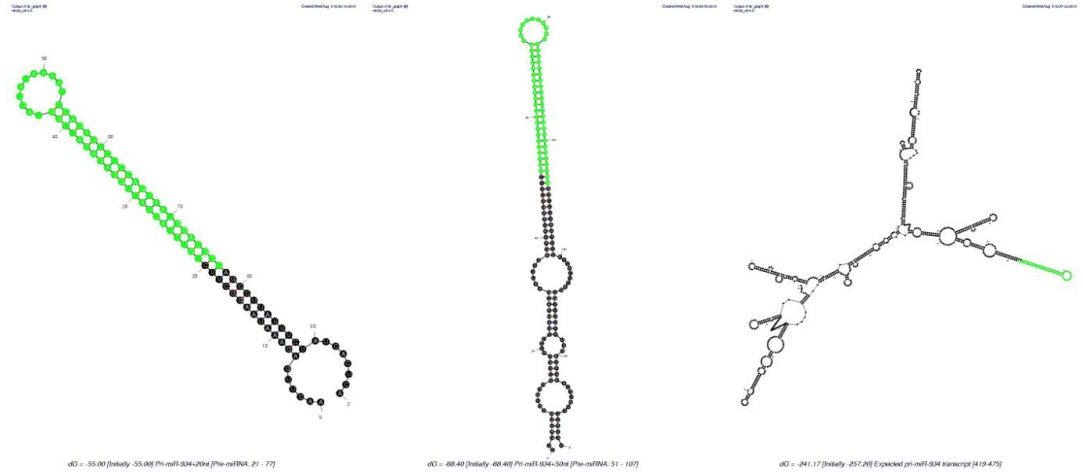
Pri-miR-23-gag



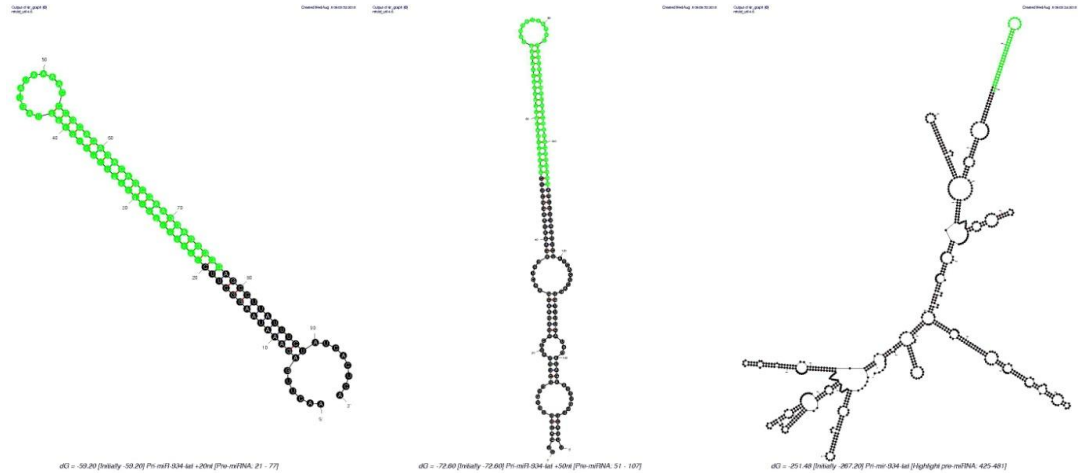
Pri-miR-23-LTR



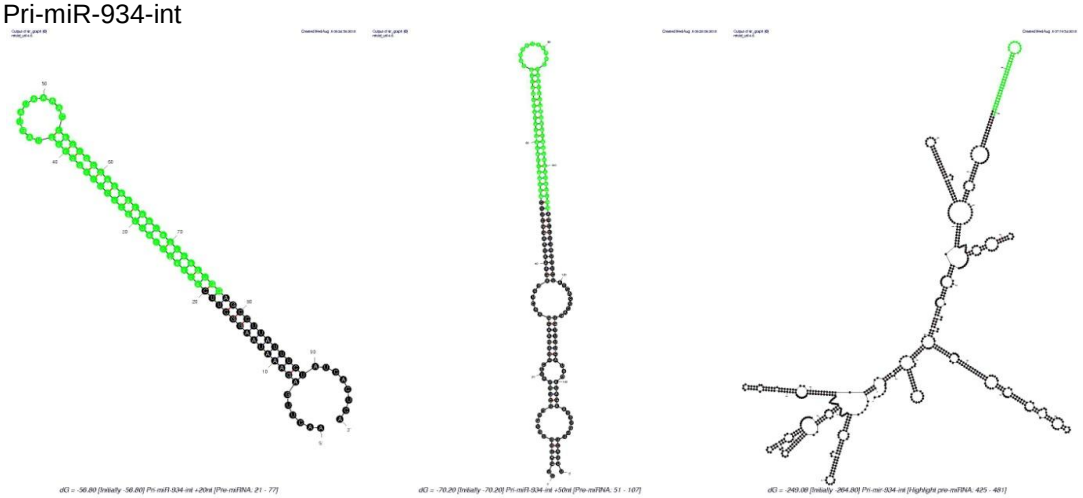
Pri-miR-934



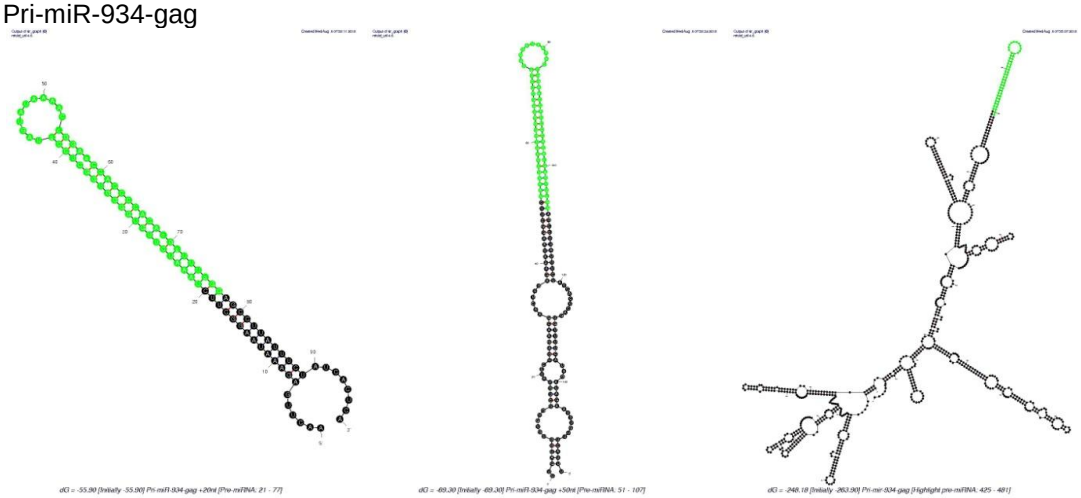
Pri-miR-934-tat



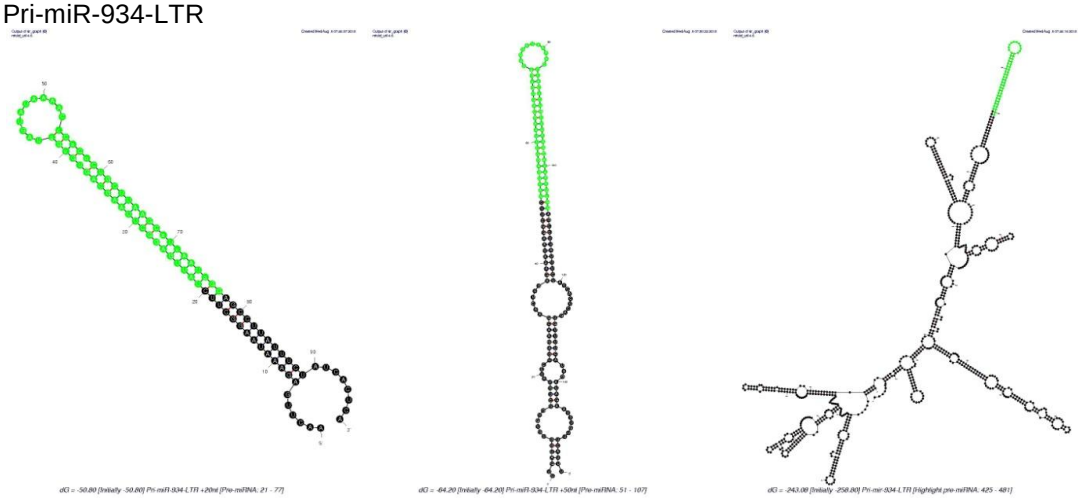
Pri-miR-934-int



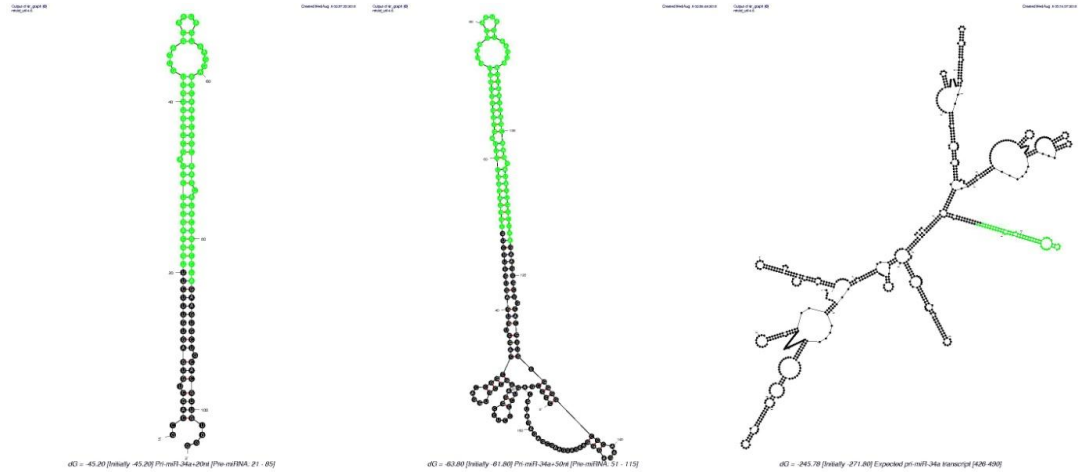
Pri-miR-934-gag



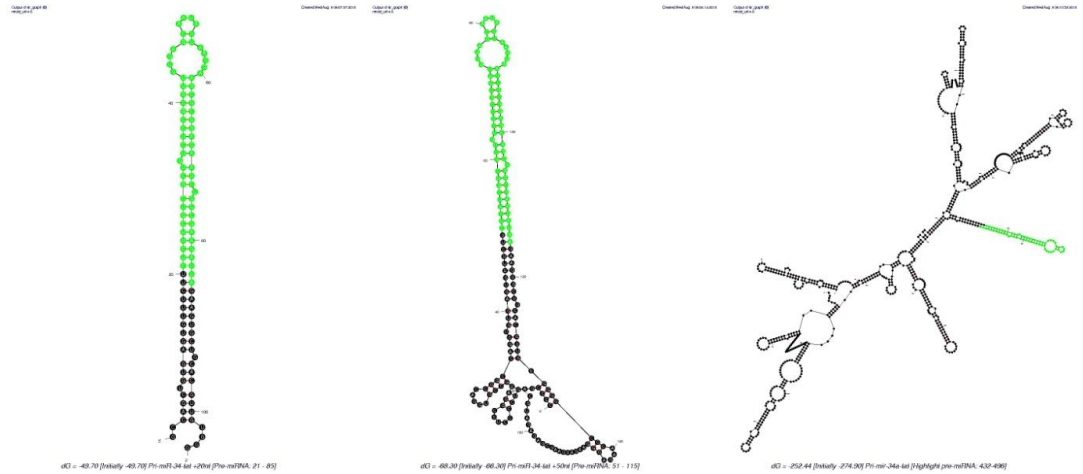
Pri-miR-934-LTR



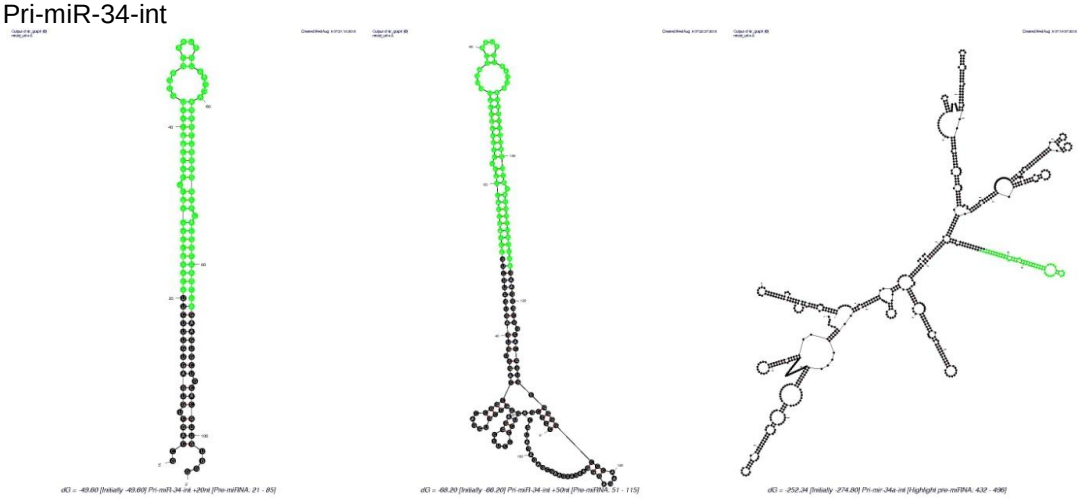
Pri-miR-34a



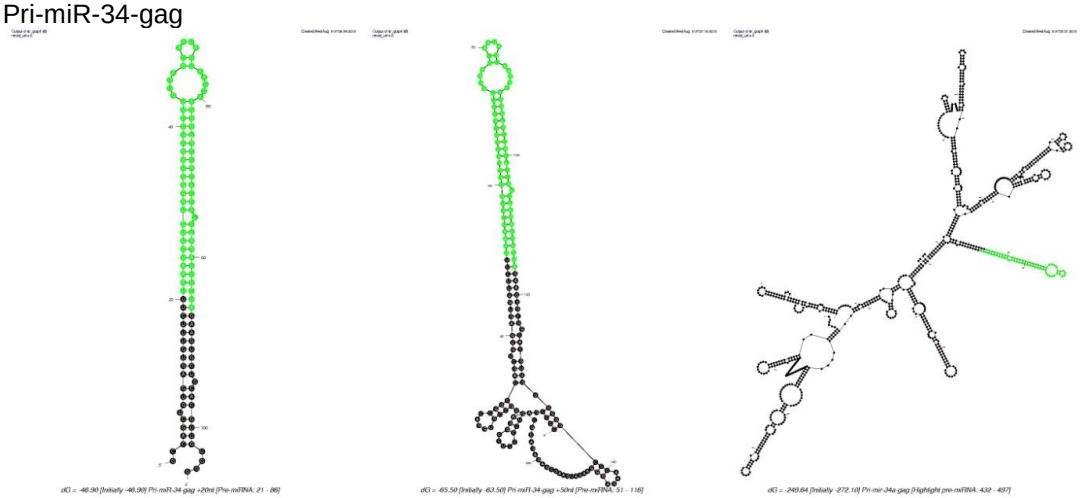
Pri-miR-34-tat



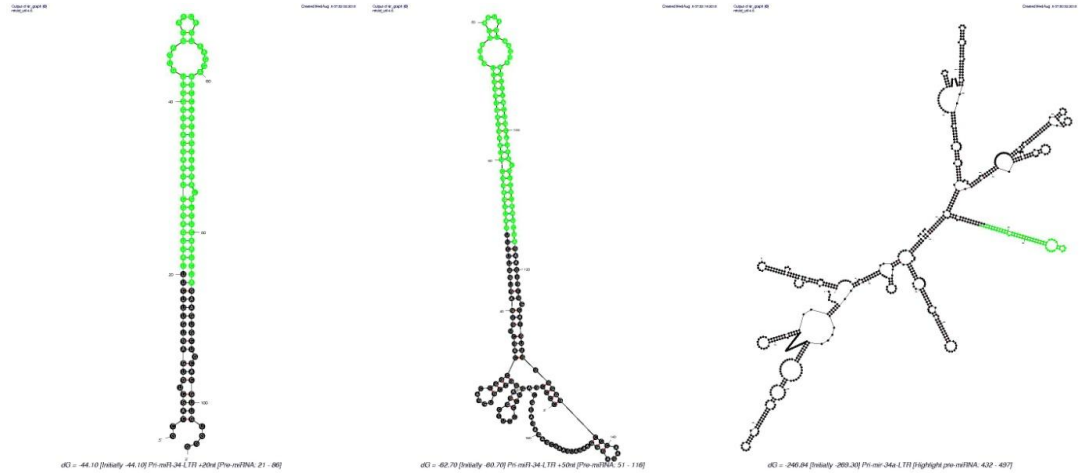
Pri-miR-34-int



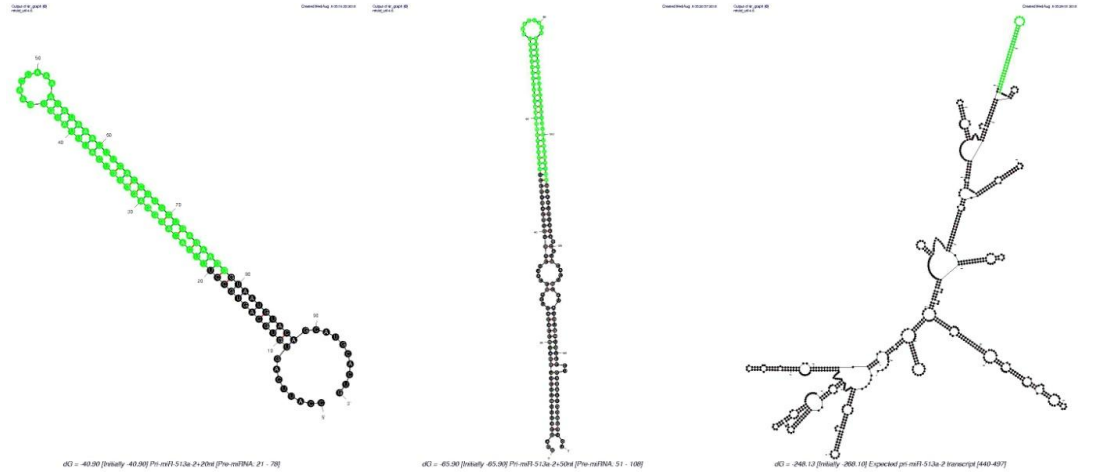
Pri-miR-34-gag



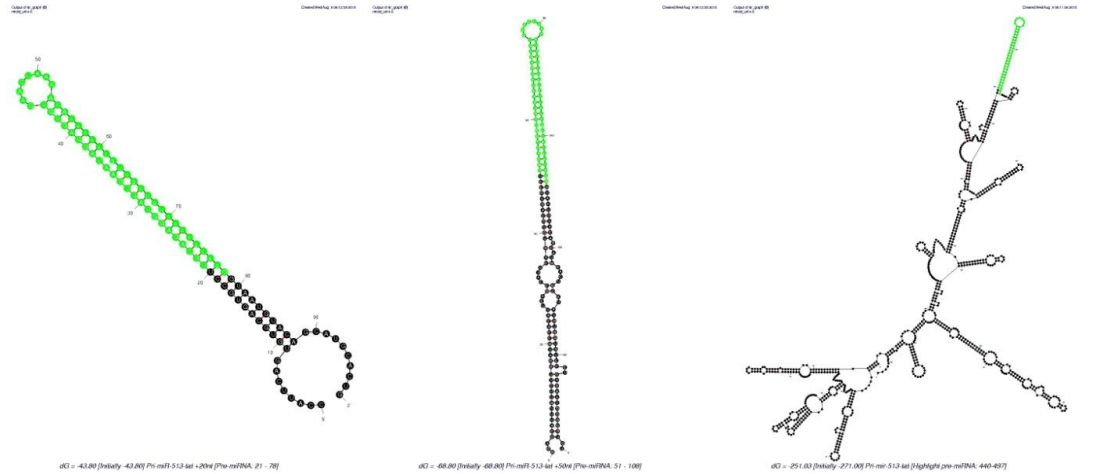
Pri-miR-34-LTR



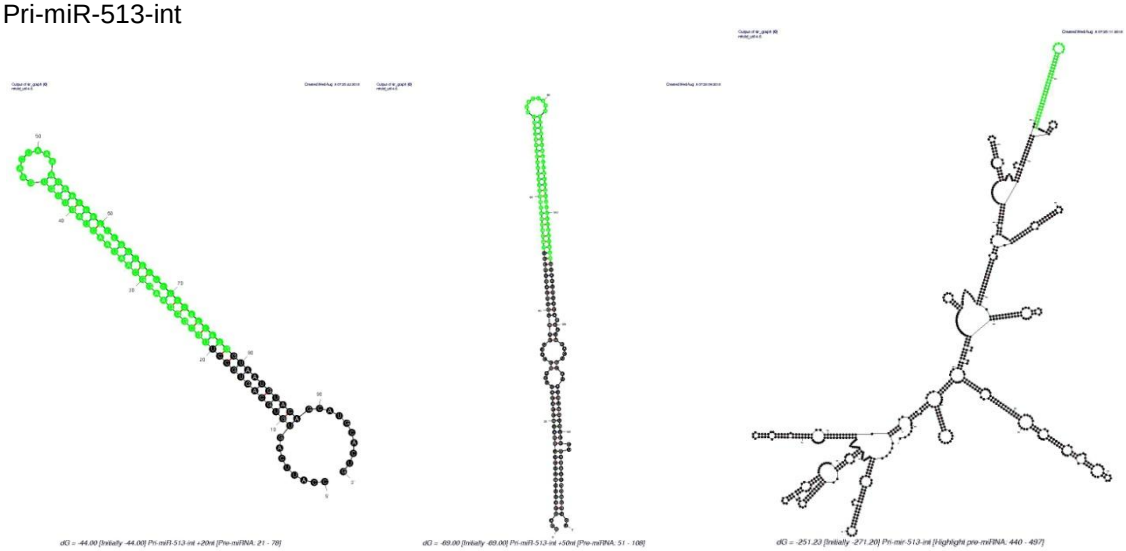
Pri-miR-513a-2



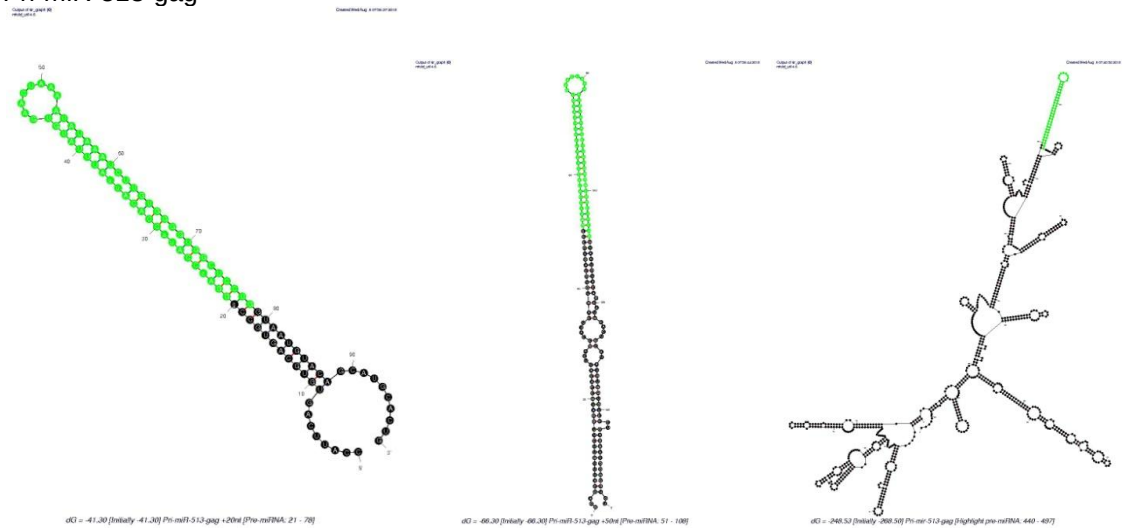
Pri-miR-513-tat



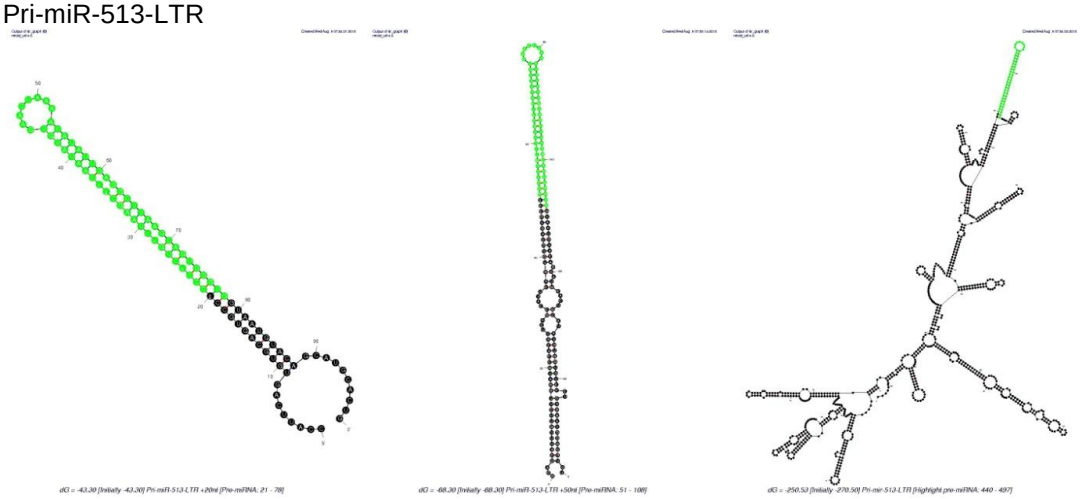
Pri-miR-513-int



Pri-miR-513-gag

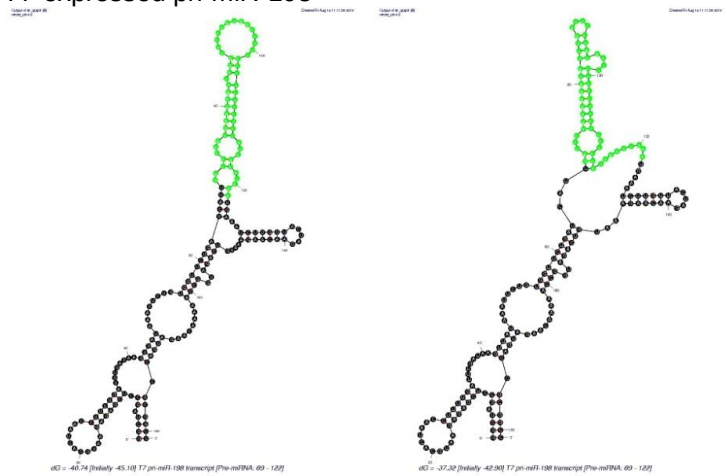


Pri-miR-513-LTR

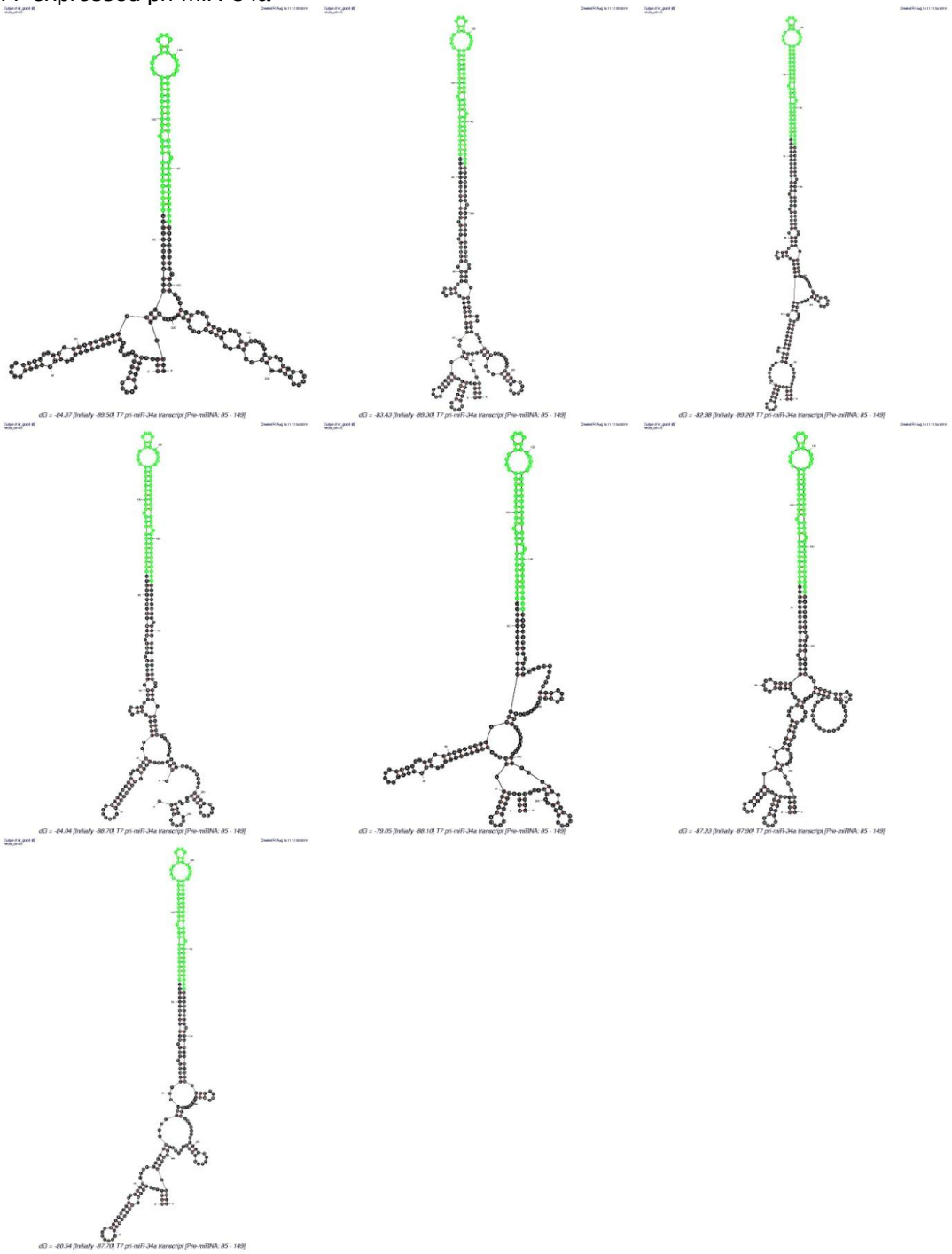


APPENDIX A-8

T7-expressed pri-miR-198



T7-expressed pri-miR-34a



APPENDIX B

ADDITIONAL MATERIALS & METHODS

All basic methods are from standard AGTRU laboratory protocols or derived from manufacturer's protocols or those described in Current Protocols in Molecular Biology (Edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, Kevin Struhl; 2003 edition; John Wiley & Sons; ISBN: 047150338X).

B.1. 6X DNA loading dye

DNA loading dye was purchased (Thermo Fisher Scientific, Waltham, MA, USA) or prepared in the laboratory (10 mM Tris-HCl pH 7.6, 0.03 % w/v bromophenol blue, 0.03 % w/v xylene cyanol FF, 60 % glycerol, 60 mM EDTA). Xylene cyanol FF migrates at approximately 4 160 bp and bromophenol blue migrates at approximately 370 bp in a 1 % agarose gel and 1 X TAE buffer.

B.2. Preparation and transformation of chemically competent cells

To prepare chemically competent cells, a 150 µl aliquot of a DH5α™ *E. coli* (Invitrogen, Carlsbad, CA, USA) glycerol stock was used to inoculate150 ml antibiotic-free LB and the culture was incubated overnight at 37 °C, 250 rpm in a shaking incubator (Labcon, Maraisburg, South Africa). The following day, 2 ml of the overnight culture was used to inoculate each of four 100 ml LB media (inoculation ratio of 1:50). The growth of these bacterial cultures was assessed every 30 min - 1 h by measuring the optical density (O.D) at 600 nm using a spectrophotometer. At an $O.D._{600}$ of 0.5 - 0.6, which approximates the log phase of growth, cells were harvested by centrifugation (Eppendorf, 5810 R, Hamburg, Germany) in eight non-sterile 50 ml Nunc™ tubes (Thermo Fisher Scientific) at 2 500 rpm, 4 °C for 15 min. The supernatant was discarded and gently replaced with 5 ml, pre-chilled PIPES buffer (15 % v/v glycerol, 0.1 M $CaCl_2 \cdot H_2O$, 10 mM PIPES (piperazine-N,N'-bis(2-ethanesulfonic acid), pH 7.0). Samples were pooled into 2 equivalent volumes and incubated on ice for 30 min. This was followed by centrifugation at 1 500 rpm, 4 °C for 10 min. Supernatant was carefully aspirated and discarded, followed by gentle re-suspension in 1.5 ml chilled PIPES buffer per pellet. Competent cells were distributed into 110 µl aliquots and stored at - 70 °C.

Prior to transformation, an aliquot of competent cells was thawed on ice. To transform competent cells, the cells were gently added to a 10 µl aliquot of a ligation reaction or purified plasmid DNA (10 - 20 ng pDNA). The mixture was incubated on ice for 30 min, followed by heat shock at precisely 42 °C for 90 s using an electronic heating block. Cells were immediately returned to ice for 5 min, followed by spreading on an appropriate LB agar plate for overnight incubation at 37 °C.

B.3. Bacterial culture media

All bacterial culture was prepared using sterile technique and methods described by Elbing and Brent (Elbing and Brent, 2002). All solutions and glassware were sterilised in an autoclave at 121 °C for 20 min, while thermolabile solutions were filter-sterilised (syringe filter, 0.2 µm pore size, Nalgene^{TM}, Thermo Fisher Scientific). Lysogeny broth (LB) (Bertani, 1951) (1 % w/v tryptone, 0.5 % w/v yeast extract, 0.5 % w/v NaCl) was used for bacterial growth media. Agar (1.5 % w/v) was added to create solid growth media for agar plates (25 ml media / plate). Ampicillin (1 000 X stock: 100 mg / ml, 50 % v/v ethanol) was added to all media in plasmid preparation for resistance selection at a final concentration of 100 µg / ml Ampicillin.

B.4. Purification of plasmid DNA from small culture volumes (pDNA mini-prep)

In a single mini-prep, 5 ml of LB medium (5 ml, 0.01 % w/v Ampicillin) was dispensed into a 10 ml culture tube and inoculated with a single bacterial colony from an LB agar plate or glycerol stock (5 µl). This culture was left to incubate overnight at 37 °C, 250 rpm in a shaking incubator. The cells were harvested the next day by centrifugation at 4 000 rpm, 4 °C for 10 min (Eppendorf, 5810 R). The supernatant was discarded and the cell pellet was re-suspended in 250 µl chilled RNAse buffer P1 (50 mM Tris·Cl, pH 8.0, 10 mM EDTA, 100 µg / ml RNase A) and transferred to a 1.5 ml microcentrifuge tube. Lysis buffer P2 (250 µl; 0.2 M NaOH, 1 % SDS) was added to the tube, followed by inversion 4 - 6 times to mix and incubation at room temperature for 4 min (maximum 5 min). Chilled neutralisation buffer P3 (350 µl; 3 M potassium acetate, pH 5.5) was added to the tube, followed by inversion 4 - 6 times to mix and incubation for 5 min on ice. A white precipitate started to form and the sample was centrifuged at 13 200 rpm, 4 °C for 10 min (Eppendorf, 5424R) to form a compact white pellet. The supernatant (~ 700 - 800 µl) was removed and transferred to a clean 1.5 ml microcentrifuge tube. An equal volume of isopropanol

was added and the tube was inverted to mix. To precipitate the pDNA, the tube was incubated at - 20 °C for 1 - 3 h or overnight. The sample was then centrifuged at 13 200 rpm, 4° C for 10 min to form a DNA pellet. The isopropanol was carefully discarded. The pellet was rinsed with 100 µl of 70 % ethanol, followed by centrifugation at 13 200 rpm, 4° C for 2 - 5 min. The ethanol was discarded by aspiration and the pellet was air-dried at room temperature for 5 - 10 min. The pDNA pellet was then re-suspended in 100 µl of sterile water and incubated at room temperature for 5 min - 1h. The pDNA yield (typically ~ 100 ng / µl) was assessed using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific) to measure absorbance (A) at 260 nm. Plasmids for DNA sequencing or *in vitro* transcription reactions were purified from bacterial culture using a commercially available mini-prep kit with spin columns (QiaPrep Spin; Qiagen, Hilden, Germany). This protocol was scaled up for larger culture volumes (100 - 150 ml) to gain higher plasmid yields, but commercially available spin columns (Qiagen-tip 100,midi-prep; Qiagen-tip 500,maxi-prep) and accompanying equilibration (QBT: 750 mM NaCl, 50 mM MOPS (3-(*N*-morpholino) propanesulfonic acid), pH 7.0, 15 % v/v isopropanol, 0.15 % v/v Triton X-100), wash (QC: 1 M NaCl, 50 mM MOPS, pH 7.0, 15 % v/v isopropanol) and elution (QF: 1.25 M NaCl, 50 mM Tris-Cl, pH 8.5, 15 % v/v isopropanol) buffers were used after the centrifugation step following buffer P3 incubation.

B.5. Agarose gel extraction

A commercially available kit was used for DNA extraction from agarose gels (Qiagen MinElute Gel Extraction Kit). Briefly, the desired DNA bands were visualised on a UV transilluminator while wearing a protective face shield. Bands were excised using a scalpel blade, transferred to a 1.5 ml microcentrifuge tube and weighed to get an approximate mass. Gel pieces were melted in 3 volumes (300 µl per 100 mg gel) of a guanidine thiocyanate buffer (QG) at 50 °C for 10 min with vortexing every 2 - 3 min. DNA was then purified using a spin column with accompanying wash (PE) and elution (EB) buffers. The kit protocol was followed, but additional or optional steps were taken to increase the DNA yield and purity. The selected spin columns were incubated at room temperature for 1 h before use. After the addition of the PE wash buffer, an optional incubation at room temperature for 5 min was included to remove salts. The required volume of the final elution buffer (TE) or water was heated to 50 °C for 1 min prior to use and incubated on the silica membrane of the spin column for 2 min before the final spin.

B.6 Freezing and thawing of mammalian cell culture stocks

In preparation for creating a cell culture stock, cells were seeded in a 25 cm^2 tissue culture flask 24 - 48 h prior to freezing and grown to ~ 80 % confluency. Cells were dissociated from the flask surface using standard sub-culture protocol (PBS rinse, trypsin-EDTA / PBS incubation), gently re-suspended in 5 ml standard cell culture medium (DMEM, 10 % FCS, 100 U / ml penicillin, 0.1 mg / ml streptomycin) and transferred to a to a 15 ml Nunc™ tube (Thermo Fisher Scientific). Cells were collected by centrifugation at 800 rpm for 3 min (Eppendorf, 5810 R). The supernatant was carefully removed and the cells were gently re-suspended in 2 -3 ml (roughly 1 ml per million cells) of freezing solution (DMEM, 40 % FCS, 10 % DMSO (dimethyl sulfoxide), 100 U / ml penicillin, 0.1 mg / ml streptomycin) and 1 ml was transferred to a 1.5 ml cryogenic tube (Nalgene™, Thermo Fisher Scientific). Tubes were then placed in a Mr. Frosty™ freezing container (Nalgene™) which contains isopropanol to slow the rate of freezing to about -1 °C / min and placed at - 70 °C overnight. Samples were transferred to a designated storage box in a liquid nitrogen tank for long-term storage, while wearing personal protective equipment (face shield, lab coat, closed shoes and protective gloves).

To start a new cell culture, a sample was removed from the liquid nitrogen storage tank. The cap of the tube was immediately loosened to prevent damage from rapid expansion. The sample was thawed for a few minutes at 37 °C in a water bath. Cells were gently transferred to a 15 ml Nunc™ tube containing 10 ml of standard DMEM culture medium and collected by centrifugation at 800 rpm for 3 min (Eppendorf, 5810 R). The supernatant was carefully removed and cells were re-suspended in 1 ml standard medium and transferred to a 25 cm^2 tissue culture flask for further culture.

B.7. Lipofectamine transfection protocol

Lipofectamine ® 2000 Reagent (Invitrogen, Carlsbad, CA, USA), a cationic lipid reagent, was used for transfections according to manufacturer's protocol. A ratio of 1 µl Lipofectamine ®: 1 µg pDNA : 100 µl Opti-MEM ® I Reduced Serum Medium (Gibco™, Thermo Fisher Scientific) was used for all transfections. Single transfections for a dual luciferase assay (1 µg pDNA per well, 24 well format) were performed in triplicate using a 3.4 X mastermix (Lipofectamine ®,pDNA, Opti-MEM ®); while individual mixes were used for single transfections for RNA extraction (10 µg pDNA, 6 cm dish). The total volume of Lipofectamine ® for a transfection series was mixed with half the total volume of Opti-MEM ® in a Nunc™ tube (L + O). The other half of the total

volume of Opti-MEM ® was distributed to pDNA of each individual mix or mastermix (D + O). Both (L + O) and (D + O) solutions were incubated for 10 min (range: 5 - 25 min) at room temperature and covered to reduce exposure to light. The (L + O) solution was then distributed to each (D + O) sample, covered and incubated for 20 min at room temperature to allow DNA-lipid complexes to form. Complexes are stable for 6 h at room temperature. The (D + O + L) solution was then distributed drop-wise to cells in wells or dishes for transfection.

B.8. RNA protocols

Procedures to prevent RNA degradation

RNA is more susceptible to degradation in the presence of RNAse enzymes and hydrolysis in the presence of alkali pH (pH > 8) or divalent metal ions. Various precautions were therefore taken to avoid these sources of degradation. To avoid degradation by ubiquitous RNAse enzymes, all equipment and surfaces were cleaned thoroughly with concentrated detergent and 70 % ethanol before use. All glassware and thermostable solutions were autoclaved, while thermolabile solutions were filter-sterilised (0.2 µm filter). Glassware, stirrer bars, spatulas and gel equipment were soaked in a 0.5 % w/v SDS (sodium dodecyl sulfate) detergent solution and rinsed with filtered Millipore water before use. RNA work was performed at a dedicated RNA bench with dedicated pipettes and filter tips. Nitrile gloves were worn for all procedures to prevent RNAse contamination. RiboLock^{TM} RNase inhibitor (ThermoFisher Scientific), a non-competitive 1 : 1 inhibitor of RNAses, was added to certain reactions as described in the main text. To avoid hydrolysis, the pH of buffers was monitored carefully and a metal chelator (EDTA; ethylenediaminetetraacetic acid) was also present in the electrophoresis buffer (TBE buffer). When possible, RNA work was carried out on ice and RNA samples were suspended in RNA loading dye and stored at - 60 to -80 °C.

DEPC-treated water

Diethylpyrocarbonate (DEPC) was used to treat water for use in RNA buffers. DEPC inactivates RNases by covalent modification of amino acid residues. DEPC was added to water (0.2 % v/v or 2 ml / l) in the fume hood as DEPC is a suspected carcinogen. The bottle was then shaken vigorously and left to incubate in the shaking incubator at room temperature for at least 2h or

overnight. The DEPC-treated water was autoclaved for 30 - 40 min to inactivate the DEPC, producing carbon dioxide and ethanol byproducts.

RNA loading dye (2X)

Deionised formamide is a crucial component of RNA loading dye as it deionises and stabilises the RNA for storage (Chomczynski, 1992). The composition of RNA loading dye differed slightly according to the manufacturer. RNA loading dye (2 X) (Ambion, Austin, TX, USA): 95% w/v formamide, 18 mM EDTA, 0.025% w/v each of SDS, bromophenol blue and xylene cyanol. RNA loading dye (2 X) (Thermo Fisher Scientific): the same formula but with 0.5 mM EDTA and an additional 0.025% w/v ethidium bromide. Ethidium bromide staining of an RNA gel was therefore not required for samples suspended in the latter dye.

Decade Marker

Decade^{TM} Marker RNA ladder (Ambion) was prepared according to the manufacturer's protocol with provided reagents. The ladder was radio-labelled in a reaction with T4 polynucleotide kinase (PNK) with components in the following order: 1 µl Decade^{TM} Marker RNA, 6 µl nuclease-free water, 1 µl 10 X Kinase reaction buffer, 1 µl T4 PNK and 1 µl P^{32}-labelled γ-ATP. The reaction was incubated at 37 °C for 1 h using a heating block. A 10 X cleavage reagent (2 µ) and 8 µl nuclease-free water were added to the reaction, followed by incubation at room temperature for just under 5 min (~ 4 min 40 - 50 s). The reaction was then placed on ice and re-suspended in 20 µl 2 X RNA loading dye. The Decade^{TM} marker ladder was distributed into four 10 µl aliquots and stored at - 60 to - 80 °C. Prior to use, an aliquot was thawed on ice and 4 µl was removed and added to 10 µl 1 X RNA loading dye (5 µl 2X loading dye, 5 µl nuclease-free water) for easier loading. The ladder sample was heated to 80 - 95 °C and returned to ice for 2 min before loading onto the gel. The unused ladder aliquot was returned to the freezer for future use. If necessary, a smaller volume of radio-labelled ladder was used to reduce the strength of the radio-isotope signal and prevent signal "bleeding" in the final gel or blot image.

B. 9. Note on work with the P^{32} radioactive isotope

P^{32}-labelled γ-ATP (3000 Ci / mmol; PerkinElmer, Waltham, MA, USA) was used for 5' radio-labelling reactions. All experiments were performed in a designated "hot lab" in accordance with procedural protocol of the Department of Molecular Medicine and Haematology. Protective plastic shields (acrylic / Perspex, 1 cm thick) were used for protection and a personal film badge dosimeter was worn and monitored by the Radiation and Health Physics Unit. Personal protective equipment also included a lab coat, safety glasses and double nitrile gloves for gel procedures involving potential contact. A two-piece Geiger-Müller counter was used to monitor for potential contamination. Waste was disposed of in storage drums for liquids and solids, the latter of which had acrylic shielding. Surfaces were cleaned with Ekon D or Radiacwash (BioDex, Shirley, NY, USA) detergents.

B.10. Sephadex® G-25 spin column

A Sephadex® G-25 spin column was constructed to remove unincorporated nucleotides and dye from labelling reactions by gel filtration (Brown et al., 2004). Sephadex® G-25 consists of cross-linked dextran gel beads which exclude RNA of more than ~ 12 nt. To allow the beads to swell, 5 g of Sephadex® G-25 was suspended in 50 ml TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 7.5) (Moore, 2001) and left to rotate overnight at room temperature. The buffer was then removed and replaced twice, with a 1 - 2 h rotation in between. This solution was ready for use, while the final solution was again left to rotate overnight for re-suspension in future use. To construct the spin column, a small piece of compact filter fibre was inserted into a 1 ml syringe (Neoject, Neomedic, Rickmansworth, UK) with the plunger removed. A 1 ml volume of Sephadex® G-25 solution was added to the syringe and the column was subject to centrifugation at 4 000 rpm for 2 min at 4 °C (Eppendorf, 5810 R) in a 15 ml Nunc™ tube. The eluate was discarded. Nuclease-free water (30 µl) was added to the reaction, followed by loading of the reaction onto the column. The centrifugation step was repeated with a sterile tube to collect the radio-labelled sample.

B.11. Phenol / chloroform / isoamyl alcohol (PCA) extraction of nucleic acids

A phenol / chloroform / isoamyl alcohol (PCA; 25:24:1 volume ratio) was used to extract RNA or DNA from enzyme reactions and remove protein contaminants. Phenol was prepared with a Tris·Cl buffer (pH 8.0) for DNA extractions and sodium acetate buffer (pH 4.5) for RNA

extractions (Moore and Dowhan, 2002). Prior to DNA or RNA extraction from a single 20 µl reaction in a 1.5 ml microcentrifuge tube, 115 µl of nuclease-free water and 15 µl 3M sodium acetate pH 5.2 were added to the reaction (final concentration 0.3 M). An equal volume (150 µl) of PCA was added and the solution was mixed thoroughly for 5 s using a vortex mixer. The solution was allowed to stand at room temperature for 1 min and then centrifuged for 2 min, 4 °C, 13 200 rpm (Eppendorf, 5424 R) to allow for phase separation. The top, clear aqueous phase (~ 140 µl) was removed and transferred to a new microcentrifuge tube, while the organic phase was discarded with phenol chemical waste. An equal volume of chloroform was added to the aqueous phase, followed by vortex mixing for 5 s and incubation at room temperature for 1 min. The solution was again centrifuged for 2 min, 4 °C, 13 200 rpm (Eppendorf, 5424 R) to allow for phase separation. The top aqueous phase (~ 120 µl) was removed to a clean microcentrifuge tube and the remaining organic phase was discarded with the chloroform chemical waste. If required, glycogen (Thermo Fisher Scientific) was added to the aqueous phase at this stage as an inert carrier for pellet visibility. Ethanol (2.5 volumes, ~ 300 µl) was added to precipitate the DNA or RNA and the solution was mixed by inverting 4 - 6 times. Samples were incubated at - 20 °C for at least 2h, but preferably overnight. The samples were centrifuged for 30 min, 4 °C, 13 200 rpm to collect the DNA or RNA, visible as a small white or clear pellet. The 100 % ethanol was discarded and the pellet was rinsed with 20 µl of 70 % ethanol. The 70 % ethanol rinse was discarded and the pellet was air-dried for 5 min at room temperature. An appropriate volume of nuclease-free water or buffer was added and the sample was incubated at room temperature for 1 - 2 min before gently re-suspending the pellet. This protocol was scaled up to extract DNA or RNA from 40 µl reactions in a 1.5 ml microcentrifuge tube, while larger reactions volumes require larger tubes. The amount of ethanol added can also be reduced to 2 volumes if required.

B.12. Preparation of RiboRuler Low Range RNA Ladder

Dephosphorylation

The RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific) was dephosphorylated in a four-fold reaction incubated at 37 °C for 1h, with the following components per single 20 µl reaction: 1 µl RiboRuler (~ 490 ng, 5.1 pmol 5' ends), 10.5 µl DEPC-treated water, 2 µl 10 X FastAP buffer, 6 U FastAP™ Thermosensitive Alkaline Phosphatase and 20 U RiboLock™ RNase inhibitor (Thermo Fisher Scientific). Dephosphorylated RiboRuler was purified by PCA extraction with acid phenol and ethanol precipitation. RNA-grade glycogen (80 µg) was added

per 120 µl aqueous phase prior to ethanol precipitation and re-suspension in 12 µl DEPC-treated water. The RNA yield was assessed by measuring UV absorbance at 260 nm and was typically ~ 20 ng / µl in a 48 µl volume (~ 50% recovery).

Radio-labelling

Dephosphorylated RiboRuler was 5'radio-labelled in a forward kinase reaction at 37 °C for 1h in a 25.5 µl volume with the following components: 20 µl dephosphorylated RiboRuler (~ 400 ng, ~ 4.1 pmol 5' ends), 2.5 µl 10 X Buffer A (Thermo Fisher Scientific), 10 U T4 PNK, 40 U RiboLock^{TM} RNase inhibitor and 1 µl P^{32}-labelled γ-ATP (PerkinElmer). Radio-labelled RiboRuler was again purified by PCA and ethanol extraction and re-suspended in 30 µl of DEPC-treated water with 1 X concentration of RNA loading dye. RiboRuler aliquots were denatured at 80 °C for 5 min, followed by 2 - 5 min on ice before loading.

B.13. X-ray film development

A blue-sensitive, universal X -ray film was used for radioisotope exposure (Agfa, CP-Bu, Medical X-ray film, 18 x 24 cm, 100 NIF) (Agfa, Mortsel, Belgium). X-ray films were developed in a dark room to prevent light exposure. Prior to development, the top left corner of the film was cut while still in the development cassette to mark film orientation in relation to the sample. X-ray film was washed for 2 min each in baths containing 1 l of (1) developer solution (2) fixer solution and (3) tap water; rinsed off with running tap water and hung up to dry overnight. Developer and fixer chemicals were supplied as 5 l concentrates (Axim, Midrand, South Africa) or powder (Kodak, Rochester, New York, USA).

APPENDIX C

PUBLICATION DERIVED FROM THIS WORK

Citation: Molecular Therapy—Nucleic Acids (2016) 5, e278; doi:10.1038/mtna.2015.53
Official journal of the American Society of Gene & Cell Therapy All rights reserved 2162-2531/16
www.nature.com/mtna

Design of Effective Primary MicroRNA Mimics With Different Basal Stem Conformations

Fiona T van den Berg[1], John J Rossi[2], Patrick Arbuthnot[1] and Marc S Weinberg[1,3,4]

Primary microRNA (pri-miRNA) mimics are important mediators of effective gene silencing and are well suited for sustained therapeutic applications. Pri-miRNA mimics are processed in the endogenous miRNA biogenesis pathway, where elements of the secondary RNA structure are crucial for efficient miRNA production. Cleavage of the pri-miRNA to a precursor miRNA (pre-miRNA) by Drosha-DGCR8 typically occurs adjacent to a basal stem of ~11 bp. However, a number of pri-miRNA structures are expected to contain slightly shorter or longer basal stems, which may be further disrupted in predicted folding of the expressed pri-miRNA sequence. We investigated the function and processing of natural and exogenous RNA guides from pri-miRNAs with various basal stems (9–13 bp), where a canonical hairpin was predicted to be well or poorly maintained in predicted structures of the expressed sequence. We have shown that RNA guides can be effectively derived from pri-miRNAs with various basal stem conformations, while predicted guide region stability can explain the function of pri-miRNA mimics, in agreement with previously proposed design principles. This study provides insight for the design of effective mimics based on naturally occurring pri-miRNAs and has identified several novel scaffolds suitable for use in gene silencing applications.

Molecular Therapy—Nucleic Acids (2016) **5**, e278; doi:10.1038/mtna.2015.53; published online 12 January 2016
Subject Category: siRNAs, miRNAs and shRNAs

Introduction

Primary microRNA (pri-miRNA) mimics have been shown to mediate effective gene silencing,[1–3] while providing several advantages over conventional short hairpin RNAs (shRNAs) and avoiding pathway saturation[4] and cellular toxicity.[5] Pri-miRNA mimics derived from endogenous miRNA precursors and expressed from a polymerase II promoter[6–8] can provide controlled, tissue-specific expression of exogenous guide RNAs, suitable for sustained posttranscriptional gene silencing. About 4,500 human miRNA precursors are available on the online miRNA repository,[9,10] but mimic design has thus far relied on only several precursors.[11–18] In particular, the model miR-30 backbone[1,2] has been widely used in therapeutic applications against a number of diseases and pathogens, including the hepatitis B virus[14] and human immunodeficiency virus,[8,16,19] and is included in ongoing clinical trials against various cancers.[20,21] The potential of thousands of other pri-miRNAs as scaffolds in mimic design has not yet been explored.

Pri-miRNA mimics, like endogenous pri-miRNAs, are processed in the miRNA biogenesis pathway.[22] Pri-miRNAs fold into a characteristic hairpin structure, with a terminal loop, imperfect duplex stem region of ~33 bp and unstructured flanking sequences.[23] Pri-miRNA processing is typically modular in nature, where cleavage by the Drosha-DGCR8 microprocessor[24–26] is reliant on both elements of the secondary RNA structure and pri-miRNA sequence, but to different extents for different pri-miRNAs.[27–29] In productive processing, Drosha recognizes the basal ssRNA-dsRNA junction and basal UG motif, cleaving as a molecular "ruler" at ~11 bp from the basal junction; and a DGCR8 dimer recognizes the stem, apical ssRNA-dsRNA junction, and loop UGU motif to facilitate processing fidelity.[28,30] Noncanonical features within the stem also contribute to cleavage site selection,[31,32] while structural distortions in proximity to cleavage sites may facilitate efficient cleavage.[33,34] In addition, methylation motifs in the proximity of miRNA genes promote processing.[35] The precursor miRNA (pre-miRNA) produced by Drosha-DGCR8 cleavage is subsequently cleaved by the RNase III enzyme Dicer,[36–39] to produce a staggered 5p/3p miRNA duplex with 3′ dinucleotide overhangs. One strand of the duplex is incorporated into the RNA-induced silencing complex[40] where it acts as a mature ~22 nt miRNA and directs posttranscriptional gene silencing.

Optimal properties of the pri-miRNA secondary RNA structure have been described that enable efficient recognition and processing. These include a large terminal loop (≥10 nt),[41] moderate basal stem (≥8 bp) with an optimal length of ~11 bp,[27] and at least 40 nt of flanking sequence on either side of the pre-miRNA, that lack a strong secondary structure.[23,42,43] These structural properties should be implemented in the design of effective pri-miRNA mimics. However, a number of pri-miRNA structures are expected to contain slightly shorter or longer basal stems, which are not necessarily maintained when mimics are generated from an expression construct. In particular, interactions between the hairpin and incorporated flanking sequences may potentially

[1]Wits/SAMRC Antiviral Gene Therapy Research Unit, Department of Molecular Medicine and Haematology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa; [2]Department of Molecular and Cellular Biology, Beckman Research Institute of City of Hope, Duarte, California, USA; [3]HIV Pathogenesis Research Unit, Department of Molecular Medicine and Haematology, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa; [4]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California, USA. Correspondence: Marc S Weinberg, Department of Molecular Medicine and Haematology, School of Pathology, The University of the Witwatersrand Medical School, 7 York Road, Parktown, Johannesburg 2193, South Africa. E-mail: marcow@scripps.edu or marc.weinberg@wits.ac.za
Keywords: artificial microRNAs; primary microRNA mimics; RNAi therapeutics
Received 17 October 2015; accepted 30 November 2015; advance online publication 12 January 2016. doi:10.1038/mtna.2015.53

support or disrupt the formation of a conventional basal stem and hairpin.

Here, we investigated the function and processing of miRNAs from expressed pri-miRNAs with various basal stem conformations (9–13 bp), where the canonical hairpin was either well or poorly maintained in the predicted secondary RNA structure of the expressed pri-miRNA sequence. The suitability of such pri-miRNAs as scaffolds was assessed through the design and characterization of corresponding pri-miRNA mimics. In their design, we have shown that minor variations in the basal stem length of pri-miRNA mimics were tolerated, while established stability features of the guide region were important and should be considered in mimic design. Several novel pri-miRNA scaffolds were characterized and are ideal for applications requiring exogenous guide sequences.

Results

Predicted secondary RNA structure of pri-miRNAs with various basal stem conformations

MicroRNA precursors with various basal stem conformations were selected from the online miRNA repository,[9,10] which lists experimentally identified miRNAs and associated stem-loop structures, and five example precursors for human miRNAs were investigated further: miR-520b, miR-23a, miR-934, miR-34a, and miR-513a-2. To assess pri-miRNA folding conformations, predicted secondary RNA structures were generated for each naturally occurring pri-miRNA with 15–50 nt of sequence flanking the anticipated Drosha-DGCR8 cleavage site, using the mfold web server[44] with default parameters (**Supplementary Figures S1** and **S2**). Predicted secondary RNA structures were also generated for RNA sequences derived from the expression vector, to assess folding of the expressed pri-miRNA. Canonical hairpins were present in the most stable pri-miRNA structures generated with 15–20 nt of flanking sequence (**Figure 1a**). These hairpins were defined as structures with a terminal loop, an upper stem harboring the miRNA, a basal stem (8–13 bp), and at least 4 nt of unstructured residues flanking the expected Drosha-DGCR8 site on one side.[27] Stem regions were counted to include Watson-Crick pairs and G–U wobble pairs, while the potential contribution of single symmetrical mismatches was also considered. The basal stem region of pri-miR-520b, for example, was slightly shorter (7 bp), but two non-Watson-Crick pairs can also stack within the stem and contribute to its length (~9 bp). Interestingly, canonical hairpin structures were not necessarily maintained in predicted pri-miRNA conformations generated with 30–50 nt of sequence flanking the cleavage site or with full flanks of the expressed pri-miRNA (**Figure 1a**,**b**). Canonical hairpins were commonly generated for pri-miR-23a and pri-miR-934 structures with extended flanking sequences but were not generated at all for corresponding pri-miR-520b structures, where predicted interactions with the flanking sequences produced a truncated hairpin. Predicted structures for pri-miR-34a typically had an extended ~16–19 bp basal stem or a canonical ~9 bp basal stem lacking adjacent unstructured residues. In addition, predicted structures encompassing a canonical hairpin were not always the most stable (*), as noticed for pri-miR-34a and pri-miR-513a-2.

Functionality of endogenous pri-miRNAs with various basal stem conformations

Predicted secondary RNA structures are not a perfect indication of dynamic RNA conformations in the cell. However, in predicted structures of the expressed pri-miRNAs that included the extended flanking sequences, the ratio of canonical to noncanonical structures with similar levels of stability was lower for specific pri-miRNAs (**Figure 1b**). Therefore, in addition to variation of the basal stem length, the basal stem may be further disrupted in the expressed pri-miRNA sequence. Expressed pri-miRNAs with additional, noncanonical hairpin conformations could conceivably show reduced miRNA-mediated silencing, making these structures less suitable as mimic scaffolds. We therefore proceeded to functionally characterize selected pri-miRNAs.

The processing and silencing capabilities of the selected pri-miRNAs were investigated in cell culture through northern blot analyses and luciferase reporter assays. A shRNA designed to express each major guide sequence was included for each corresponding pri-miRNA as a positive control for guide production and function. Mature ~22 nt guides were detected for exogenously expressed pri-miRNAs with a range of basal stem lengths (9–13 bp) and efficient silencing (63–91 %) of corresponding targets was observed (**Figure 2a**). Pre-miRNAs (55–65 nt) were also detected, thus supporting cleavage at the anticipated processing sites. The most effective target silencing was observed for pri-miR-934, while target silencing of over 80% was observed for both pri-miR-34a and pri-miR-513a-2. Complementary strands of the miRNA duplex were detected for each pri-miRNA, but effective target silencing was only observed for miR-34a-3p (68%) and miR-513a-2-5p (92%) (**Figure 2b**), implying that these minor miRNAs can also be effectively loaded into RNA-induced silencing complex. Larger (75–120 nt) RNA products were also detected across all samples in several blots and were presumed to be endogenous background RNAs that bind nonspecifically to the probe. Reporter gene suppression was specific to individual guide–target interactions, as no silencing was observed in the absence of a target sequence (**Supplementary Figure S3**). Similar assay results were also obtained in the Huh-7 cell line (**Supplementary Figure S4**). These data suggest that minor variation in the basal stem length and additional, noncanonical folding of the basal stem were not restrictive in conventional pri-miRNA processing. Functional miRNAs were effectively derived from pri-miRNAs where the canonical hairpin was present in almost all (pri-miR-934), most (pri-miR-23a), or none (pri-miR-520b) of the predicted structures including the extended flanking sequence of the expressed pri-miRNA. However, as canonical hairpins were present in predicted structures generated with 15–20 nt of flanking sequence for all pri-miRNAs, these data also suggest that structures including the local flanking sequence were more useful in assessing the processive form.

Nonstructural determinants are also expected to contribute to the observed pri-miRNA function. Pri-miRNA sequences were analyzed for the presence of recently described sequence motifs, including the basal UG and CNNC motifs, the loop GUG motif, the mismatched GHG stem motif, and

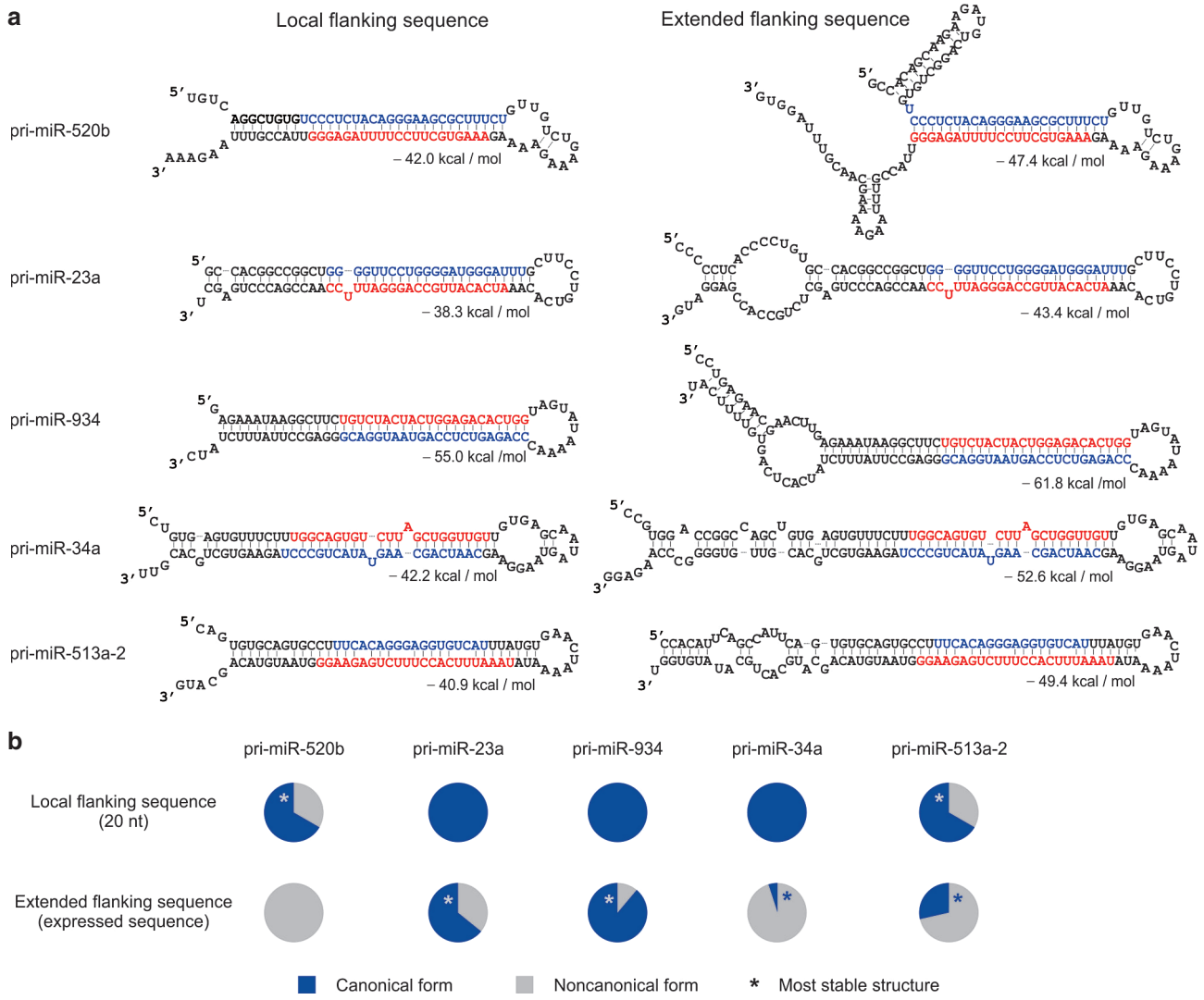


Figure 1 Predicted secondary RNA structures of selected pri-miRNAs. (**a**) (Left) Structures were generated with the local flanking sequence (15 nt) and are representative of the conventional hairpins predicted for each pri-miRNA. (Right) Structures were also generated with extended regions of the flanking sequence (30 nt) and show the additional alternative structures predicted for some pri-miRNAs. The anticipated (miRBase) major (red) or minor (blue) 5p or 3p miRNA guide sequences, and total free energy of the predicted structure (mfold ver. 3.5) are indicated. (**b**) The proportion of predicted secondary RNA structures with a canonical pri-miRNA hairpin generated using local (20 nt) or extended flanking sequences of the expressed pri-miRNA.

nearby METTL3 motifs[27,29,35] (**Figure 3a**). Several potential motifs were identified for each pri-miRNA sequence and are expected to contribute to the observed function in a modular manner specific to each pri-miRNA. The contribution of individual pri-miRNA sequence motifs was not assessed, as these motifs were initially well characterized.[27,29] However, it is interesting to note that pri-miR-520b potentially has all four pri-miRNA sequence motifs (**Figure 3b**) which are expected to compensate for the suboptimal basal stem length and permit conventional Drosha-DGCR8 processing.[29] Similarly, pri-miR-23a has all four potential pri-miRNA sequence motifs, which may permit processing in the context of various basal stem conformations (**Supplementary Figure S1**); while pri-miR-934 only has two potential pri-miRNA sequence motifs, but a well-maintained conventional basal stem. Pri-miRNA sequences were also analyzed for the presence of the 5′-GGAC-3′ METTL3 motif[35] in the context of both the expressed and genomic sequences (**Figure 3c**). Motifs occurring within 100 nt of the pre-miRNA in the genomic context were typically preserved in the cloned pri-miRNA sequence, with the exception of pri-miR-34a and pri-miR-513a-2 where a single motif was omitted. However, this was unlikely to have a major impact on function, as one or two original motifs were still retained in pri-miR-34a and pri-miR-513a-2 sequences, which were processed to give effective target silencing (>80%). In addition, an extra motif was present in the expressed pri-miRNA sequence of all five pri-miRNAs. In principle, these data suggest that effective miRNAs can be embedded and processed from within exogenously expressed pri-miRNAs with various basal stem lengths and

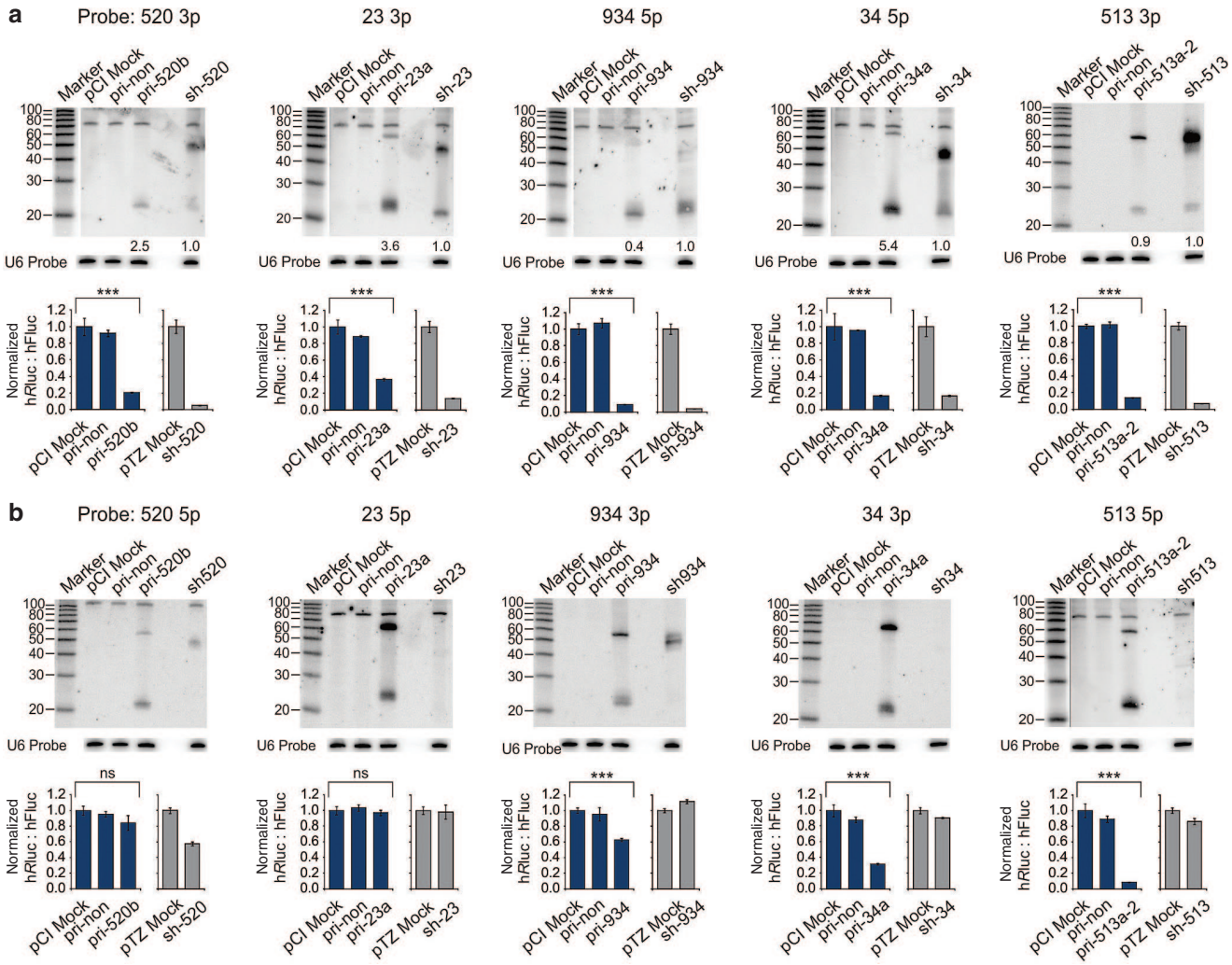


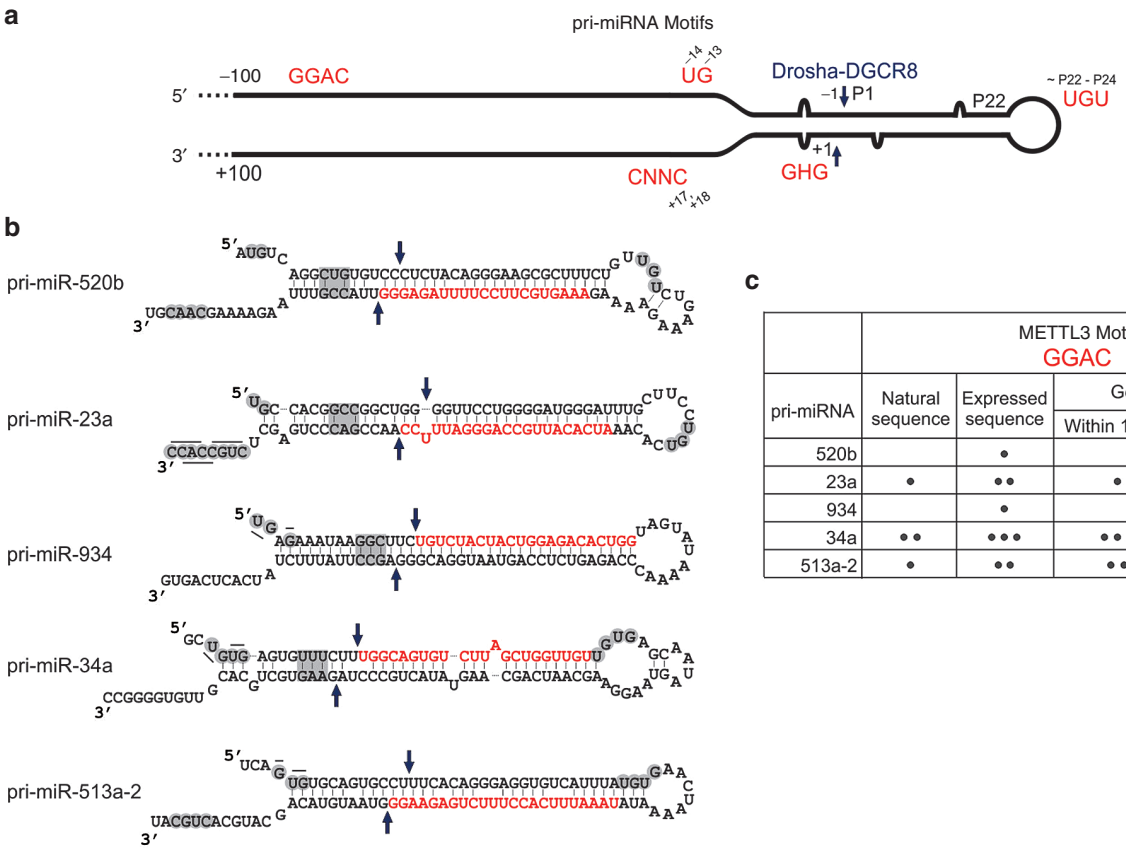
Figure 2 Processing and silencing efficacy of expressed pri-miRNAs. (**a**) (Upper panel) Major guide sequences were detected through northern blot analyses, following transfection of HEK293 cells with the corresponding pri-miRNA or shRNA (positive control) expression vectors. A mock expression vector (pCI/pTZ Mock) and nonspecific pri-miRNA expression vector (pri-non) were included as negative controls. Guide signal intensities were quantified relative to U6 signal intensity and are shown as a ratio of pri-miRNA-derived guide to shRNA-derived guide. (Lower panel) Target gene silencing by derived guides was measured by the mean ratio of *Renilla* luciferase: Firefly luciferase (h*R*luc: hFluc) ($n = 3$, ± SD). Values were normalized with respect to the relevant mock sample. Sample means differed significantly (one-way analysis of variance (ANOVA), ***$P < 0.0001$). (**b**) The processing and function of minor miRNA sequences was similarly assessed through northern blot analyses (upper panel) and target gene silencing (lower panel). Included shRNAs were only designed to express the major miRNA guide. ns: no significant difference between sample means (one-way ANOVA, $P > 0.05$); 934-3p: $P = 0.0003$.

motif combinations, while potentially disruptive interactions predicted to occur between the basal stem and flanking sequences appear to be of little consequence.

Design and functionality of pri-miRNA mimics

Since each characterized natural pri-miRNA appeared suitable for use as a scaffold into which functional exogenous miRNA/siRNA sequences can be embedded, we proceeded to determine the efficacy of corresponding pri-miRNA mimics. Pri-miRNA mimics were designed by replacing the endogenous miRNA guide sequences with one of four previously characterized therapeutic siRNA guides[45–47] targeting the expression of *tat*, *int*, *gag*, or long terminal repeat (LTR) regions of the human immunodeficiency virus genome (**Figure 4a**). The original miRNA sequence was replaced with 20–22 nt of a therapeutic guide, depending on the length of the original miRNA. Resulting minor variations in the 3′-terminal 1–3 nt of the miRNA were not expected to drastically influence the function of mimics within each guide sequence set, where there was a minimum of 20 nt of common 5′ guide sequence, including the seed region,[48] and a minimum of 19 complementary guide–target interactions.

The predicted secondary structure (mfold)[44] of pri-miRNA mimics was kept as close as possible to that of the endogenous pri-miRNAs, as structural features, like internal bulges, may be crucial to maintain productive processing[23,31,32,49] (**Supplementary Figure S5**). Alterations to the complementary strand of the guide region were used to maintain these structural features. To assess potential changes in the pri-miRNA conformation, the predicted secondary RNA structure



pri-miRNA	METTL3 Motif GGAC			
	Natural sequence	Expressed sequence	Genomic sequence	
			Within 100 nt	Within 500 nt
520b		•		•
23a	•	••	•	••••
934		•		••••
34a	••	•••	•••	•••••
513a-2	•	••	••	••

Figure 3 Potential pri-miRNA sequence motifs. (**a**) Schematic illustration of a typical pri-miRNA showing the location of potential sequence motifs. Primary sequence motifs[27] may occur in the 5p arm preceding (–) the Drosha-DGCR8 cleavage site (arrows), the 3p arm after (+) the Drosha-DGCR8 cleavage site or the terminal loop within the pre-miRNA (P) sequence. A mismatched GHG-type motif[29] may also be present at position 7–9 of the 3p basal stem. Methyltransferase-like 3 (METTL3) recognition motifs[35] may occur at nonspecific locations of the pri-miRNA sequence outside of the pre-miRNA. (**b**) Potential basal, stem, or loop motifs. (**c**) Potential METTL3 recognition motifs in the natural or expressed pri-miRNA sequence or within 100 or 500 nt of sequence flanking either side of the expected pre-miRNA in the genomic context.

of each mimic was again examined with both the local and extended sequence flanking the anticipated Drosha-DGCR8 site (**Supplementary Figure S6**). Although structural prediction results differed slightly for some pri-miRNA mimics, such as pri-miR-520-tat (local sequence) and pri-miR-23, -934 and 34a derivatives (extended sequence) (**Supplementary Figure S7**), the ratio of canonical to noncanonical forms was typically similar to that of the original pri-miRNA.

The processing and silencing efficacy of guides derived from pri-miRNA mimics was again assessed by northern blot analyses and luciferase reporter assays (**Figure 4b–e**). ShRNAs expressing each guide sequence, sh-tat,[50] sh-int,[51] sh-gag,[52] and sh-LTR,[52] were included as positive controls. In contrast to the original pri-miRNAs, functional miRNAs were not effectively derived from all pri-miRNA mimics. Exogenous ~22 nt guides with effective target silencing were detected more frequently for certain pri-miRNA scaffolds. Target silencing of ~80% or more was observed for two mimics based on both the pri-miR-34a and 513a-2 scaffolds, but for only one mimic based on each of the pri-miR-520b, 23a, and 934 scaffolds. As a group, mimics based on the pri-mR-34a scaffold had the highest average target silencing, while mimics based on the pri-miR-520b scaffold had the lowest average target silencing. Effective therapeutic guides were also detected more frequently for certain mimic sets. Target silencing of ~80% or more was observed for four of the LTR mimics, but none of the tat mimics. As a group, the LTR mimic set was generally the most successful with the greatest average decrease in target silencing, followed by the gag, int, and tat mimic sets, respectively. This ranking was reflected in target silencing by the set of pri-miR-520b mimics (pri-miR-520-tat: 0%, int: 12%, gag: 45%, ltr: 79%). Luciferase assay results were specific to individual guide–target interactions and no significant difference was observed between sample means (**Supplementary Figure S8a**). Guide sequences were detected for pri-mir-34-tat and pri-mir-513-tat despite poor target silencing. However, recalculating target silencing values relative to each respective endogenous pri-miRNA only produced slightly improved values for the tat mimic set (**Supplementary Figure 8b,c**). Minor variations in mimic design did not significantly affect target silencing results. While the exact number of complementary guide–target interactions was expected to vary, no correlation was observed between the total (19–22 bp) or consecutive (18–22 bp) number of guide–target interactions and target silencing (**Figure 4f**). Noncomplementary 5′ guide–target interactions of gag and LTR mimics (1–2 nt) may be expected to slightly reduce target silencing but occurred uniformly within each set

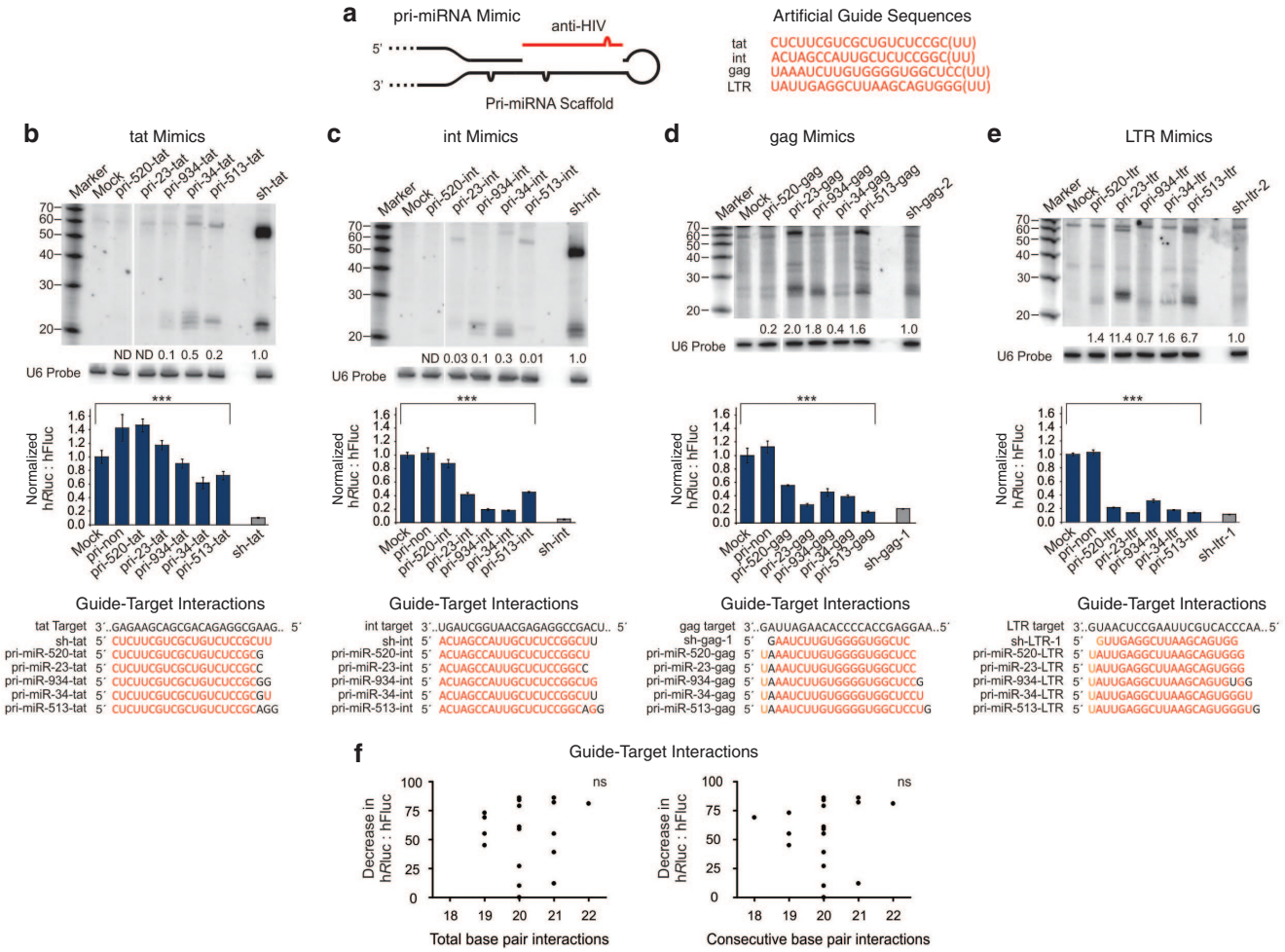


Figure 4 The design and functional characterization of expressed pri-miRNA mimics. (**a**) Each original guide sequence was replaced with a therapeutic guide against tat, int, gag, or LTR (long terminal repeat) regions of the HIV-1 (human immunodeficiency virus-1) genome. (**b–e**) The functional characterization of pri-miRNA mimic sets. (Upper panel) Processed antiviral guide sequences were detected in northern blot analyses, following transfection with the corresponding pri-miRNA mimic, shRNA (positive control) or mock (negative control) expression vectors. Guide signal intensities were again quantified relative to U6 signal intensity and are shown as a ratio of pri-miRNA-derived guide to shRNA-derived guide. ND: not detected. (Middle panel) Dual luciferase reporter assays show silencing of respective target sequences in cell culture, measured by mean h*R*luc: hFluc ($n = 3$, ± SD). A nonspecific pri-miRNA (pri-non) expression vector was included as a negative control. Sample means differed significantly for each set of mimics (one-way analysis of variance, $^{***}P < 0.0001$). (Lower panel) Expected guide–target interactions showing complementary base pairing (red). (**f**) The total or consecutive number of expected base pair interactions was plotted against the corresponding decrease in h*R*luc: hFluc ($n = 20$). No significant (ns) correlation was observed. Spearman (ranked) correlation coefficient ($rs^{TOTAL} = 0.19$), ($rs^{CONS.} = 0.27$) (two-tailed P value > 0.2).

of mimics and were also expected for the respective shRNA controls. Pre-miRNAs (55–65 nt) were also detected for several mimics, again supporting cleavage at the anticipated Drosha-DGCR8 sites. Together, these results indicated that exogenous miRNAs can indeed be effectively derived from expressed pri-miRNAs with a variety of basal stem lengths and motifs, where the canonical hairpin was present in almost all, most, or none of the predicted structures including the extended flanking sequence, but the level of target silencing differed for mimics with the same scaffold or core guide sequence.

Stability of the pri-miRNA guide region

Differences in pri-miRNA mimic functionality could not be attributed to changes in the overall form of predicted secondary RNA structure or known primary sequence motifs, which were unaltered in mimic design. However, we wondered whether these differences could be explained by the more subtle thermodynamic changes of the pri-miRNA stem incurred in guide replacement. We therefore investigated the stability of predicted pri-miRNAs over the guide region. A characteristic bias in pri-miRNA stability and common thermodynamic features of functional pri-miRNAs have been described in previous studies.[23,33,49,53] The region corresponding to the 5′ end of the mature miRNA is less stable than the region corresponding to its 3′ end,[49] as revealed by thermodynamic profiling of pri-miRNAs using mfold values. In particular, the most unstable position of the stem corresponds to the first position of the expected miRNA.[23,53] This approach relies on the predicted secondary RNA structures of pri-miRNAs,

which can differ from experimentally determined structures, especially at positions within the terminal loop and flanking sequences.[49] However, the percentage of correctly predicted base pairs tends to be high (88.4%),[49] supporting the predicted conformation of the upper pri-miRNA stem. We applied thermodynamic profiling to investigate stability of the pri-miRNA stem along the guide region for various pri-miRNA and mimic groups. In this approach, precise values cannot be assigned to positions within internal loops and we therefore estimated values at these positions using an average loop value, but similar results were also obtained using previously described methods[23,49] (**Supplementary Figure S9**). As profiles differ for pri-miRNAs with 5p and 3p miRNAs in accordance with the miRNA orientation, we assessed each guide region from the 5′ end of the expected miRNA. The guide region was also limited to the portion of the pri-miRNA stem corresponding to the first 20 nt of the exogenous guide, as this region was replaced for all mimics, with the exception of pri-mir-934-LTR where 19 nt of the first 20 nt were replaced. The difference in free energy between the position 1 and 20 (ddG (1–20)) was used to assess the bias in stability expected over the guide region.

Average thermodynamic profiles were constructed for functional groups of pri-miRNAs with effective (>80%), functional (>50%), or ineffective (<50%) target silencing (**Figure 5a**). Common features were identified, including a decrease in the relative stability at positions 4 and 9, as well as an increase in the relative stability at positions 11, 16, and 20. However, the stability at position 1 of the stem, which corresponds to the 5′ end of the miRNA, decreased with increasing target silencing of each pri-miRNA group. In turn, the expected bias in pri-miRNA stability, as indicated by the ddG (1–20) value, was highest for the group of effective pri-miRNAs (1.7 kcal/mol) and lowest for the group of ineffective pri-miRNAs (0.3 kcal/mol). A similar profile for the group of original pri-miRNAs had a ddG (1–20) value slightly lower than expected (0.7 kcal/mol, lower panel), as position 20 of the pri-miR-23a guide region was relatively unstable; but stable pairing at position 21 may facilitate a bias in guide region stability.

Average thermodynamic profiles were also constructed for groups of pri-miRNAs with the same exogenous guide (**Figure 5b**). Again, the ddG (1–20) signature increased with the average target silencing of each group and was pronounced for the LTR mimic set (1.8 kcal/mol). The stability of the guide region for each group was primarily defined by a common guide sequence, but small differences in base pairing were used to maintain the internal mismatches of individual scaffolds. To examine the extent of base pairing variation within each mimic set, we constructed base pairing profiles (**Figure 5c**). Common base pairing was observed with minor variation, thus supporting a common 5′ to 3′ bias in stable GC pairing for each mimic set, in agreement with the thermodynamic analysis. This bias was more pronounced for the gag and LTR mimic sets, where no GC base pairs were observed in the first four positions as a consequence of siRNA design principles used by McIntyre *et al.*[47,54] Similarly, no GC base pairs were expected in the first position for the original, functional pri-miRNAs, as is common for natural miRNAs and supported by an Ago-2 preference for uracil in the first position of a mature miRNA.[33,55]

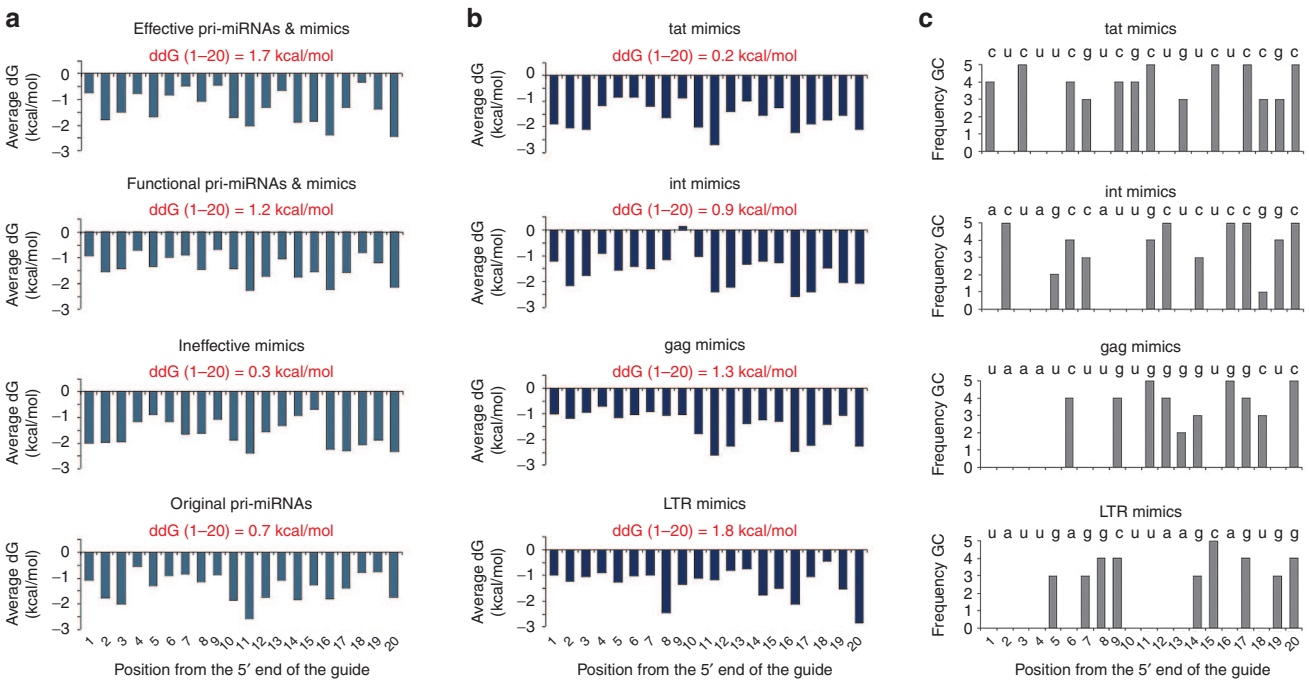


Figure 5 Predicted stability of the pri-miRNA or mimic guide region. (**a**) Average thermodynamic profiles of the guide region were constructed for pri-miRNA/mimic subsets with effective (>80%), functional (>50%), or ineffective (<50%) target silencing and for the group of original pri-miRNAs. The difference in average stability between positions 1 and 20 of the guide region (ddG (1–20)) is indicated (red). (**b**) Average thermodynamic profiles were also constructed for different pri-miRNA mimic subsets. (**c**) Base pairing profiles for the guide region of different pri-miRNA mimic subsets, indicating the frequency of GC base pairing.

Subtle thermodynamic changes of the pri-miRNA stem incurred in guide replacement may therefore explain the overall function of mimic sets. We could detect no difference in the form of the predicted secondary RNA structure, including flanking sequences, or in various motifs of pri-miRNA mimics; but the average target silencing of each mimic set increased with the magnitude of the bias in stability, as indicated by ddG (1–20). This confirms the importance of siRNA design and previously described thermodynamic features of the guide region in the generation of effective mimics. In fact, pri-miR-23-gag and pri-miR-23-LTR mimics had an increase in both ddG (1–20) and target silencing compared to pri-miR-23a. These results are in agreement with design principles suggested by Han *et al.*,[23] stating that the first position of an artificial guide should be unstable, while position ~20 should be relatively stable.

Duplex stability

To determine whether the expected miRNA duplexes also displayed thermodynamic asymmetry, we used conceptual dicing to assess the relative terminus stability.[49,56] In most cases, the 5′ end of the miRNA is expected to be less stable, affecting strand selection and RNA-induced silencing complex loading.[53,56,57] The average terminus stability for pri-miRNA groups with effective (>80%), functional (>50 %), or ineffective (<50 %) target silencing was calculated using a 2 nt window, which has previously been described as the most informative window for miRNA identification[49] (**Figure 6a**). As expected, the average free energy was higher for the terminus corresponding to the 5′ end of the major miRNA (red) than that of the minor miRNA (blue). The average free energy of both termini was lower for the group of ineffective pri-miRNAs; but, in contrast to the results of thermodynamic profiling, a similar difference in terminus stability was observed for all functional groups (~0.5 kcal/mol). Similarly, the difference in average terminus stability for duplexes derived from each mimic set did not increase with the average target silencing (**Figure 6b**). The average free energy of the miRNA 5′ terminus was lower than expected for the group of original (wt) pri-miRNAs using a 2 nt window, as a result of a mismatched residue near the 5′ terminus of the minor miRNA for pri-miR-934 and a noninternal mismatch at the 5′ terminus of the major miRNA for pri-miR-34a. Similar results were obtained using a 1 nt or 3 nt window (**Supplementary Figure S10**).

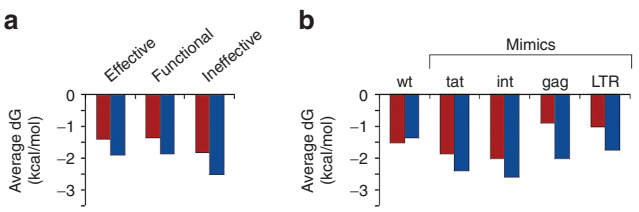


Figure 6 Terminus stability of expected miRNA duplexes. (**a**) The average terminus stability of duplexes derived from conceptual dicing of pri-miRNA/mimic subsets with effective (>80 %), functional (>50%), or ineffective (<50%) target silencing. Average free energy of the major (red) and minor (blue) miRNA termini are shown. (**b**) The average terminus stability of duplexes derived from conceptual dicing of different pri-miRNA mimic subsets and the group of original, wild-type (wt) pri-miRNAs. LTR, long terminal repeat.

In contrast to thermodynamic profiling, the results of conceptual dicing did not clearly identify a link between the magnitude of thermodynamic asymmetry and the overall function of mimic sets. This may be because only 20 nt of the guide region were commonly replaced. However, stability of the both termini generally increased with a decrease in the average target silencing of different mimic sets. This suggests that while duplex asymmetry was observed for less effective pri-miRNA groups, terminus stability was generally unfavorable. This again supports the importance of relative instability at the 5′ of the miRNA in mimic design.

Precise miRNA sequences

The results of characterization studies and guide region analyses rely on the successful derivation of the expected guide sequences and consistent processing of derived mimics. The expected miRNAs were detected in northern analysis, but we also sought to examine the derived miRNA species in more detail through deep sequencing. The top three reads aligning to pri-miR-34a, pri-miR-513a-2, and their respective int mimics are shown in **Figure 7** (top 10: **Supplementary Figure S11a**). We expect that reads aligning to pri-mir-34a and pri-mir-513a-2 originate from both endogenous miRNA genes and the respective expression vectors. However, as the expressed pri-miRNAs were derived from genomic sequence, we expect the miRNA species to be processed in the same manner. Endogenous miRNA genes were expected to generate a smaller fraction of miRNAs, as relevant species were not detected in northern analysis (pCI Mock sample). Overall, the miRNA processing sites were well maintained between expressed pri-miRNAs and int mimics. In particular, Drosha-DGCR8 processing sites were more precisely maintained, as indicated by the consistent 5′ ends of the 5p guides. A 1 nt 5′ variation was observed at Dicer cleavage sites, although Dicer processing of a 22 nt miR-513-int-3p was more consistent. Evidence of 3′ trimming and editing (uridylation and adenylation) of miRNAs was observed as is common for small RNA sequencing data.[58] The annotated miRNA guides of the original pri-miRNAs (**Figure 1**) were present in the top three reads, with the exception of miR-513a-2-5p, where several ~22 nt miRNAs were identified. This finding is in agreement with the collective deep sequencing data now available on miRBase (**Supplementary Figure S11b,c**). These data also suggest that more conventional miR-513a-2-5p species are in fact the major, functional miRNA products of pri-miR-513a-2, consistent with the anticipated processing sites of the 3p miRNA and characterization studies (**Figure 2**). An effective mimic may therefore also be designed by replacement of the 5p miRNA in the pri-miR-513a-2 scaffold.

Discussion

In this study, we have characterized several novel and effective naturally occurring pri-miRNA scaffolds, ideal for use in applications requiring exogenous guide sequences for posttranscriptional gene silencing. Our results indicated that suboptimal basal stem lengths and potentially disruptive interactions of the basal stem region in the expressed pri-miRNA did not appear to restrict pri-miRNA function or

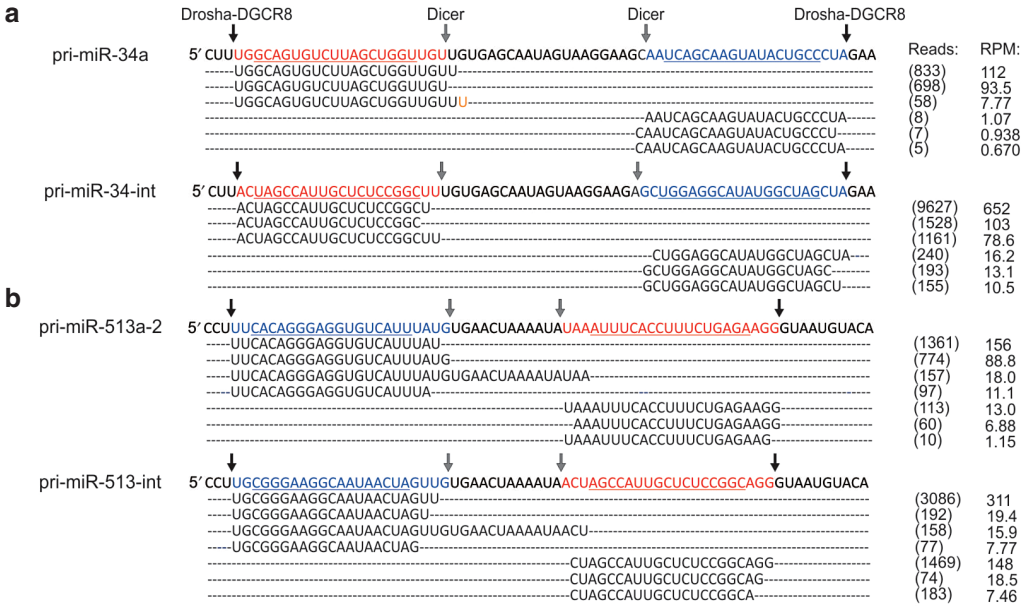


Figure 7 Exact miRNA species and implied processing sites of pri-miRNA mimics. Deep sequencing was used to assess guide production from expressed pri-miR-34a, pri-miR-513a-2 and corresponding int mimics. The top three reads aligned to each 17 nt identifier (underlined) for major miRNA (red) and minor miRNA (blue) sequences are shown and were quantified in reads per million (RPM, three significant figures). Nontemplated nucleotides are indicated (orange).

conventional processing. Therefore, these factors do not appear of great concern or consequence in the design of effective pri-miRNA mimics. This is presumably because a suboptimal basal stem length may be compensated for by an optimal total stem length. The total stem length of pri-miR-23a, -934, -34a, and -513a-2 is 35 ± 1 bp which has recently been described by Fang and Bartel[29] as optimal for pri-miRNA processing by Drosha-DGCR8. Sequence motifs may also compensate for suboptimal or varied basal stem conformations. In the case of a pri-miR-520b with a suboptimal basal (9 bp) and total (32 bp) predicted stem length, the presence of all four potential pri-miRNA sequence motifs is expected to facilitate conventional Drosha-DGCR8 recognition and processing.[27,29] Our data also suggested that hairpin structures including the local flanking sequence were more indicative of the conventional, processive form. Naturally occurring pri-miRNAs with a range of basal stem lengths and motif combinations were therefore suitable for mimic design, provided that the original sequence and secondary RNA structural features supported pri-miRNA processing, in agreement with the current modular understanding of pri-miRNA processing.[27,29]

Our results also indicated that while naturally occurring pri-miRNAs were effective at silencing relevant targets, derived mimics were not necessarily as effective, or even functional. Thermodynamic profiling showed that the most effective mimic groups were those where the guide sequence permitted a decrease in relative stability at the first position of the miRNA sequence and an increased bias in pri-miRNA stability over the first 20 nt of the guide region. A zero or positive change in ddG (1–20) was also observed for 11 out of 13 individual functional mimics when compared to the original pri-miRNA. These results are in agreement with a number of previous studies describing both common features of functional pri-miRNAs,[23,49,53] including a U at the first position of a mature miRNA,[55] and principles of effective siRNA design.[54,59,60] Guide sequences selected with respect for siRNA design principles (gag, LTR) were generally more effective in the context of a pri-miRNA scaffold than those selected solely on the basis of target accessibility and conservation (tat, int), despite the fact that all four guides were effective when derived from a pol III shRNA format. Suboptimal properties of a processed RNA duplex may also be more evident in a less robust pol II expression system. In a previous study by Aagaard *et al.*,[61] an effective tat mimic based on the miR-106b precursor was created, where the core guide sequence was positioned within the scaffold to generate a ~21 nt guide with a U at position 1 and an additional 3′ GC. The function of this mimic contrasts with the generally poor performance of the tat mimics created here, again highlighting the importance of these sequence features in mimic design.

This work provides a practical example of effective mimic design using naturally occurring pri-miRNA scaffolds with a range of predicted basal stem conformations and sequence motif combinations. Effective pri-miRNA mimics satisfied basic requirements of the secondary RNA structure in folding of the local sequence, had features commonly observed for functional pri-miRNAs, and guide sequences selected with regard for established siRNA design principles. In addition, ddG (1–20) proved to be a useful parameter in pri-miRNA mimic design. Several highly effective pri-miRNA mimics have recently been generated by utilizing artificial guide region features within a natural scaffold[18] and a *de novo* design approach to optimize mimic function,[29] but our work illustrates that the use of naturally occurring scaffolds with original guide region features remains a facile and effective approach for the design of pri-miRNA mimics.

Materials and methods

pri-miRNA and mimic expression plasmids. Stem-loop precursors available on the miRBase repository[9,10] were inspected manually and five were selected as examples of precursors with exceptional basal stems. The Ensembl genome browser[62] was used to extend the miRBase genomic coordinates by ± 50 nt to incorporate the expected pri-miRNA flanking regions. Sequences for pri-miR- 520b, -23a, -934, and -34a were amplified by PCR from genomic DNA (HEK116 extract), using primers listed in **Supplementary Table S1**. Sequences for pri-miR-513a-2, and all pri-miRNA mimics were constructed from a two-step PCR with oligonucleotides (**Supplementary Table S2**). After initial TA cloning (InsTAclone PCR Cloning Kit, Thermo Fisher Scientific, Waltham, MA), sequences were inserted into the pCI-neo Mammalian Expression Vector (Promega, Madison, WI) and expressed from the pol II cytomegalovirus immediate-early enhancer/promoter. All oligonucleotides and primers were from Integrated DNA Technologies (IDT, Coralville, IA). The sequences of all RNA expression and target constructs were verified by standard dideoxy chain termination DNA sequencing (Inqaba Biotec, Pretoria, South Africa).

Short hairpin RNA expression plasmids. shRNAs were constructed for each pri-miRNA as a positive control of guide function, using a two-step PCR method[63] with primers listed in **Supplementary Table S3**. ShRNA expression cassettes included a U6 pol III promoter sequence and a transcription termination signal of five deoxythymidines. These expression cassettes were inserted into a TA cloning vector (InsTAclone PCR Cloning Kit, Thermo Fisher Scientific). Antiviral shRNAs were previously constructed and characterized by our lab: sh-tat,[50] sh-int,[51] sh-gag, and sh-LTR.[52]

Target reporter plasmids. Target sequences were fully complementary to the anticipated ~22 nt miRNA, as well as 2 nt before and after the anticipated guide to accommodate minor processing variation, and cloned into the 3′ UTR of the *Renilla* luciferase gene of the psiCHECK-2 Vector (Promega). Cleavage or translational suppression mediated by an effective miRNA results in decreased *Renilla* luciferase expression, while expression of Firefly luciferase is unaffected. Target duplexes were constructed by annealing complementary oligonucleotides (**Supplementary Table S4**). Antiviral reporter plasmids were constructed by others: psiCHECK-*tat*,[50] psiCHECK-*int*,[51] psiCHECK-*gag*,[52] and psiCHECK-LTR.[52]

Cell culture and transfections. HEK293, Huh-7, and HeLa cells were cultured in Dulbecco's modified Eagle medium (Biowhittaker, Lonza, Walkersville, MD), supplemented with 10% heat-inactivated fetal calf serum, penicillin (100 U/ml), and streptomycin (0.1 mg/ml), at 37 °C, 5% CO_2. Antibiotic-free media was used for transfections. Cells for northern blot analysis were seeded in a 6-cm dish (~1.1×10^6 cells) 24 hours prior to cotransfection with 10 µg total pDNA (9 µg pri-/shRNA expression plasmid; 1 µg pCI-eGFP) using Lipofectamine 2000 Reagent (Invitrogen, Carlsbad, CA) according to manufacturer's protocol. Cells for the dual luciferase assay were seeded in a 24-well format (~1.2×10^5 cells/well) 24 hours prior to triplicate cotransfections with 1 µg total pDNA (750 ng pri-/shRNA expression plasmid; 150 ng target reporter plasmid; 100 ng pCI-eGFP). Cells for deep sequencing were seeded in a 6-well format (~4.8×10^5 cells/well) 24 hours prior to cotransfection with 5 µg total pDNA (4.5 µg pri-/shRNA expression plasmid; 0.5 µg pCI-eGFP). pCI-eGFP was constructed previously by our lab[64] to expresses an enhanced green fluorescent protein and was included as a reporter of transfection efficiency visualized by fluorescence microscopy. Cell culture medium was replaced 24 hours after transfection. At 48 hours after transfection, cells were lysed for the dual luciferase assay or total RNA was extracted for northern blot or deep sequencing samples by phase separation using TRI Reagent (Sigma-Aldrich, St. Louis, MO) according to the manufacturers' protocol for monolayer cells.

Northern blot analysis. Total RNA (30 µg) was resolved using 15% denaturing PAGE with Decade Marker (Ambion, Austin, TX). RNA was transferred to a Hybond N+ membrane (Amersham, GE Healthcare, Buckinghamshire, UK) using a semi-dry blotter (Sigma-Aldrich, Dorset, UK). Blots were UV cross-linked at 254 nm (0.2 J/cm^2) (UVP Crosslinker, Upland, CA) and baked for 1 hour at 80 °C. Probes (**Supplementary Table S5**) were 5′ radiolabeled by T4 polynucleotide kinase with 1 µl γ-32P ATP (PerkinElmer, Waltham, MA) and purified using a Sephadex-G25 spin column. Hybridization was performed for 16 hours at 42 °C for standard probes and at 50 °C for locked nucleic acid-modified probes (Exiqon, Woburn, MA) using Rapid-hyb buffer (Amersham, GE Healthcare). Membranes were washed as described in the Rapid-hyb protocol or elsewhere.[65] Signals were quantified using a phosphor-imager (FLA-7000 Image Reader, Fuji Film, Kanagawa, Japan) and accompanying software (Fuji Film Multi Gauge, ver. 3.0, Science Lab 2005).

Dual luciferase reporter assay. Assays were performed in triplicate using the Dual-Luciferase Reporter Assay System (Promega) and a Veritas dual injection Luminometer (Turner Biosystems, Sunnyvale, CA). Silencing was measured as a ratio of target-specific *Renilla* luciferase signal to background Firefly luciferase signal (h*R*luc: hFluc). The average ratio of each sample was normalized with respect to the average ratio of a control sample (empty expression plasmid).

Thermodynamic profiling and conceptual dicing. Average thermodynamic stability profiles were constructed in a similar manner to previous studies[23,33,49,53] using free energy values (kcal/mol) from the mfold web server (version 3.5)[44] with default parameters at 37 °C. In this approach, precise values cannot be assigned to positions within internal loops as the values for loop features are assigned to the preceding closing pair. We therefore estimated values at these positions based on previously described methods. Positions within internal loops were assigned an average value for the loop feature; assigned the average value of a 3 nt window, including one position on either side, where all positions within a loop are assigned a single value[49]; or designated as blank positions and omitted from the average free energy value at that position.[23] Conceptual dicing was performed as previously described,[49] where the stability of each duplex terminus was estimated from the predicted mfold structure of a corresponding hairpin. Terminal mismatches were assigned the value of the external loop feature.

Deep sequencing. Samples were sequenced using the Illumina Genome Analyzer II platform. The 3′-adapter of raw reads were trimmed and reads were aligned to the full pri-miRNA reference sequence using Novoalign (http://www.novocraft.com). Aligned reads were converted to .bam format and indexed using SAMtools[66] and visualized using Integrative Genomics Viewer (IGV, ver. 2.3).[67] Further quantification analysis was performed using miRDeep2.[68] Identifier sequences of 17 nt, starting with the third nucleotide from the 5′ end of each expected 22 nt miRNA sequence, were used to identify RNA species aligning to the core of each miRNA. The expected miRNA sequences were adjusted in relation to the Drosha-DGCR8 cleavage site of the major characterized miRNA species. Reads were ranked according to frequency and quantified in reads per million (RPM) for a single experiment.

Supplementary material

Figure S1. Summary of folding results for pri-miRNAs.
Figure S2. Predicted secondary RNA structures of selected pri-miRNAs.
Figure S3. Target gene silencing of pri-miRNAs: control assays.
Figure S4. Target gene silencing in a second cell line.
Figure S5. Predicted secondary RNA structures of pri-miRNA mimics.
Figure S6. Summary of folding results for pri-miRNA mimics.
Figure S7. The proportion of canonical hairpins predicted for each set of pri-miRNA mimics.
Figure S8. Target gene silencing of mimics: control assays.
Figure S9. Predicted stability of the pri-miRNA or mimic guide region.
Figure S10. Terminus stability of expected miRNA duplexes.
Figure S11. Deep sequencing results of pri-miRNAs and mimics.
Table S1. Primers for PCR amplification of pri-miRNA sequences from genomic DNA.
Table S2. Oligonucleotides for two-step PCR construction of pri-miRNAs and pri-miRNA mimics.
Table S3. Primers for PCR construction of shRNAs.
Table S4. Oligonucleotides for construction of target duplexes.
Table S5. Probes for detection of guide RNAs in northern analyses.

Acknowledgments We gratefully acknowledge grant support to M.S.W. by the South African National Research Foundation (NRF) and the South African Medical Research Council (MRC). F.T.v.d.B. acknowledges scholarship support from the NRF, Andrew W Mellon Foundation, the South African Poliomyelitis Research Foundation, the Freda Lawenski Foundation, and the Professional Provident Society of South Africa Chairman's Bursary. We thank Sheena Saayman and Neli Nhlabatsi for antiviral shRNA constructs used as positive controls and corresponding target reporter plasmids; Samantha Barichievy for the nonspecific pri-miRNA construct used as a negative control; and Jiehua Zhou, Xiwei Wu, and Jonathan Hart for assistance with the generation and analysis of deep sequencing data.

1. Zeng, Y, Wagner, EJ and Cullen, BR (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9**: 1327–1333.
2. Zeng, Y, Cai, X and Cullen, BR (2005). Use of RNA polymerase II to transcribe artificial microRNAs. *Methods Enzymol* **392**: 371–380.
3. Silva, JM, Li, MZ, Chang, K, Ge, W, Golding, MC, Rickles, RJ *et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet* **37**: 1281–1288.
4. Grimm, D, Streetz, KL, Jopling, CL, Storm, TA, Pandey, K, Davis, CR *et al.* (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**: 537–541.
5. Boudreau, RL, Martins, I and Davidson, BL (2009). Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs *in vitro* and *in vivo*. *Mol Ther* **17**: 169–175.
6. Zhou, H, Xia, XG and Xu, Z (2005). An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi. *Nucleic Acids Res* **33**: e62.
7. Stegmeier, F, Hu, G, Rickles, RJ, Hannon, GJ and Elledge, SJ (2005). A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc Natl Acad Sci USA* **102**: 13212–13217.
8. Lo, HL, Chang, T, Yam, P, Marcovecchio, PM, Li, S, Zaia, JA *et al.* (2007). Inhibition of HIV-1 replication with designed miRNAs expressed from RNA polymerase II promoters. *Gene Ther* **14**: 1503–1512.
9. Griffiths-Jones, S (2004). The microRNA Registry. *Nucleic Acids Res* **32**(Database Issue): D109–D111.
10. Kozomara, A and Griffiths-Jones, S (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**(Database Issue): D68–D73.
11. Chung, KH, Hart, CC, Al-Bassam, S, Avery, A, Taylor, J, Patel, PD *et al.* (2006). Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. *Nucleic Acids Res* **34**: e53.
12. Liu, YP, Haasnoot, J, ter Brake, O, Berkhout, B and Konstantinova, P (2008). Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. *Nucleic Acids Res* **36**: 2811–2824.
13. Ely, A, Naidoo, T, Mufamadi, S, Crowther, C and Arbuthnot, P (2008). Expressed anti-HBV primary microRNA shuttles inhibit viral replication efficiently *in vitro* and *in vivo*. *Mol Ther* **16**: 1105–1112.
14. Ely, A, Naidoo, T and Arbuthnot, P (2009). Efficient silencing of gene expression with modular trimeric Pol II expression cassettes comprising microRNA shuttles. *Nucleic Acids Res* **37**: e91.
15. Yang, X, Haurigot, V, Zhou, S, Luo, G and Couto, LB (2010). Inhibition of hepatitis C virus replication using adeno-associated virus vector delivery of an exogenous anti-hepatitis C virus microRNA cluster. *Hepatology* **52**: 1877–1887.
16. Zhang, T, Cheng, T, Wei, L, Cai, Y, Yeo, AE, Han, J *et al.* (2012). Efficient inhibition of HIV-1 replication by an artificial polycistronic miRNA construct. *Virol J* **9**: 118.
17. Huang, X and Jia, Z (2013). Construction of HCC-targeting artificial miRNAs using natural miRNA precursors. *Exp Ther Med* **6**: 209–215.
18. Choi, JG, Bharaj, P, Abraham, S, Ma, H, Yi, G, Ye, C *et al.* (2015). Multiplexing seven miRNA-based shRNAs to suppress HIV replication. *Mol Ther* **23**: 310–320.
19. Boden, D, Pusch, O, Silbermann, R, Lee, F, Tucker, L and Ramratnam, B (2004). Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. *Nucleic Acids Res* **32**: 1154–1158.
20. Rao, DD, Maples, PB, Senzer, N, Kumar, P, Wang, Z, Pappen, BO *et al.* (2010). Enhanced target gene knockdown by a bifunctional shRNA: a novel approach of RNA interference. *Cancer Gene Ther* **17**: 780–791.
21. Senzer, N, Barve, M, Kuhn, J, Melnyk, A, Beitsch, P, Lazar, M *et al.* (2012). Phase I trial of "bi-shRNAi(furin)/GMCSF DNA/autologous tumor cell" vaccine (FANG) in advanced cancer. *Mol Ther* **20**: 679–686.
22. Bartel, DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
23. Han, J, Lee, Y, Yeom, KH, Nam, JW, Heo, I, Rhee, JK *et al.* (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**: 887–901.
24. Denli, AM, Tops, BB, Plasterk, RH, Ketting, RF and Hannon, GJ (2004). Processing of primary microRNAs by the microprocessor complex. *Nature* **432**: 231–235.
25. Gregory, RI, Yan, KP, Amuthan, G, Chendrimada, T, Doratotaj, B, Cooch, N *et al.* (2004). The microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235–240.
26. Han, J, Lee, Y, Yeom, KH, Kim, YK, Jin, H and Kim, VN (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**: 3016–3027.
27. Auyeung, VC, Ulitsky, I, McGeary, SE and Bartel, DP (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**: 844–858.
28. Nguyen, TA, Jo, MH, Choi, YG, Park, J, Kwon, SC, Hohng, S *et al.* (2015). Functional anatomy of the human microprocessor. *Cell* **161**: 1374–1387.
29. Fang, W and Bartel, DP (2015). The menu of features that define primary microRNAs and enable *de novo* design of microRNA genes. *Mol Cell* **60**: 131–145.
30. Quick-Cleveland, J, Jacob, JP, Weitz, SH, Shoffner, G, Senturia, R and Guo, F (2014). The DGCR8 RNA-binding heme domain recognizes primary microRNAs by clamping the hairpin. *Cell Rep* **7**: 1994–2005.

31. Ma, H, Wu, Y, Choi, JG and Wu, H (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc Natl Acad Sci USA* **110**: 20687–20692.
32. Burke, JM, Kelenis, DP, Kincaid, RP and Sullivan, CS (2014). A central role for the primary microRNA stem in guiding the position and efficiency of Drosha processing of a viral pri-miRNA. *RNA* **20**: 1068–1077.
33. Warf, MB, Johnson, WE and Bass, BL (2011). Improved annotation of *C. elegans* microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. *RNA* **17**: 563–577.
34. Quarles, KA, Sahu, D, Havens, MA, Forsyth, ER, Wostenberg, C, Hastings, ML *et al.* (2013). Ensemble analysis of primary microRNA structure reveals an extensive capacity to deform near the Drosha cleavage site. *Biochemistry* **52**: 795–807.
35. Alarcón, CR, Lee, H, Goodarzi, H, Halberg, N and Tavazoie, SF (2015). N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**: 482–485.
36. Bernstein, E, Caudy, AA, Hammond, SM and Hannon, GJ (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366.
37. Grishok, A, Pasquinelli, AE, Conte, D, Li, N, Parrish, S, Ha, I *et al.* (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**: 23–34.
38. Hutvágner, G, McLachlan, J, Pasquinelli, AE, Bálint, E, Tuschl, T and Zamore, PD (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.
39. Ketting, RF, Fischer, SE, Bernstein, E, Sijen, T, Hannon, GJ and Plasterk, RH (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev* **15**: 2654–2659.
40. Hammond, SM, Bernstein, E, Beach, D and Hannon, GJ (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**: 293–296.
41. Zeng, Y, Yi, R and Cullen, BR (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **24**: 138–148.
42. Chen, CZ, Li, L, Lodish, HF and Bartel, DP (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**: 83–86.
43. Zeng, Y and Cullen, BR (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**: 27595–27603.
44. Zuker, M (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.
45. Lee, NS, Dohjima, T, Bauer, G, Li, H, Li, MJ, Ehsani, A *et al.* (2002). Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol* **20**: 500–505.
46. Nishitsuji, H, Kohara, M, Kannagi, M and Masuda, T (2006). Effective suppression of human immunodeficiency virus type 1 through a combination of short- or long-hairpin RNAs targeting essential sequences for retroviral integration. *J Virol* **80**: 7658–7666.
47. McIntyre, GJ, Groneman, JL, Yu, YH, Jaramillo, A, Shen, S and Applegate, TL (2009). 96 shRNAs designed for maximal coverage of HIV-1 variants. *Retrovirology* **6**: 55.
48. Lewis, BP, Shih, IH, Jones-Rhoades, MW, Bartel, DP and Burge, CB (2003). Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.
49. Krol, J, Sobczak, K, Wilczynska, U, Drath, M, Jasinska, A, Kaczynska, D *et al.* (2004). Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J Biol Chem* **279**: 42230–42239.
50. Saayman, S, Barichievy, S, Capovilla, A, Morris, KV, Arbuthnot, P and Weinberg, MS (2008). The efficacy of generating three independent anti-HIV-1 siRNAs from a single U6 RNA Pol III-expressed long hairpin RNA. *PLoS One* **3**: e2602.
51. Saayman, S, Arbuthnot, P and Weinberg, MS (2010). Deriving four functional anti-HIV siRNAs from a single Pol III-generated transcript comprising two adjacent long hairpin RNA precursors. *Nucleic Acids Res* **38**: 6652–6663.
52. Nhlabatsi, N (2014). *Comparing the Silencing Efficacy of Dicer-Independent and Dependent shRNAs*. Dissertation, Department of Molecular Medicine and Haematology, University of the Witwatersrand: Johannesburg, South Africa.
53. Khvorova, A, Reynolds, A and Jayasena, SD (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209–216.
54. Reynolds, A, Leake, D, Boese, Q, Scaringe, S, Marshall, WS and Khvorova, A (2004). Rational siRNA design for RNA interference. *Nat Biotechnol* **22**: 326–330.
55. Frank, F, Sonenberg, N and Nagar, B (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**: 818–822.
56. Schwarz, DS, Hutvágner, G, Du, T, Xu, Z, Aronin, N and Zamore, PD (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199–208.
57. Gregory, RI, Chendrimada, TP, Cooch, N and Shiekhattar, R (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**: 631–640.
58. Landgraf, P, Rusu, M, Sheridan, R, Sewer, A, Iovino, N, Aravin, A *et al.* (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**: 1401–1414.
59. Ui-Tei, K, Naito, Y, Takahashi, F, Haraguchi, T, Ohki-Hamazaki, H, Juni, A *et al.* (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* **32**: 936–948.
60. Amarzguioui, M and Prydz, H (2004). An algorithm for selection of functional siRNA sequences. *Biochem Biophys Res Commun* **316**: 1050–1058.
61. Aagaard, LA, Zhang, J, von Eije, KJ, Li, H, Saetrom, P, Amarzguioui, M *et al.* (2008). Engineering and optimization of the miR-106b cluster for ectopic expression of multiplexed anti-HIV RNAs. *Gene Ther* **15**: 1536–1549.
62. Cunningham, F, Amode, MR, Barrell, D, Beal, K, Billis, K, Brent, S *et al.* (2015). Ensembl 2015. *Nucleic Acids Res* **43**(Database Issue): D662–D669.
63. Castanotto, D, Li, H and Rossi, JJ (2002). Functional siRNA expression from transfected PCR products. *RNA* **8**: 1454–1460.
64. Passman, M, Weinberg, M, Kew, M and Arbuthnot, P (2000). *In situ* demonstration of inhibitory effects of hammerhead ribozymes that are targeted to the hepatitis Bx sequence in cultured cells. *Biochem Biophys Res Commun* **268**: 728–733.
65. Várallyay, E, Burgyán, J and Havelda, Z (2008). MicroRNA detection by northern blotting using locked nucleic acid probes. *Nat Protoc* **3**: 190–196.
66. Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N *et al.*; 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
67. Robinson, JT, Thorvaldsdóttir, H, Winckler, W, Guttman, M, Lander, ES, Getz, G *et al.* (2011). Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
68. Friedländer, MR, Mackowiak, SD, Li, N, Chen, W and Rajewsky, N (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**: 37–52.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/

Supplementary Information accompanies this paper on the Molecular Therapy–Nucleic Acids website (http://www.nature.com/mtna)