RNA INTERACTOME IDENTIFICATION BY NEXT-GENERATION SEQUENCING (RIINGS) – A NOVEL METHOD FOR IDENTIFYING MICRORNA-RNA INTERACTOME IN EXACT CORRESPONDENCE

HUNG WING HONG

Ph.D

The Hong Kong Polytechnic University

2017

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

RNA Interactome Identification by Next-generation Sequencing (RIINGS) – A Novel Method for Identifying MicroRNA-RNA Interactome in Exact Correspondence

HUNG Wing Hong

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

October 2016

Certification of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____

HUNG Wing Hong

Abstract

MicroRNAs are important in epigenetic regulation of gene expression. They are short non-coding RNA molecules, and are able to regulate gene expression at the post-transcriptional level via binding to the target sites of target RNAs. Since target recognition does not require full complementarity between microRNAs and their target sites, a single microRNA can potentially regulate many targets. In addition, the regulation may show conditional dependence as evidenced by the fact that when luciferase reporter assays are used to confirm the inhibition effect of some microRNAs on their targets, the results can be positive or negative in different cell lines. Therefore, microRNA target identification is a very challenging task in microRNA studies.

There was no reliable experimental method to identify microRNA targets in a massive way in the past and hence computational prediction has played an important role in and contributed a lot to this area for a long time. However, the accuracy of computation-based methods is far from desire in identifying the genuine targets. Recently, several high-throughput screening methods have been developed for identifying microRNA targets. These methods can be divided into two major categories. The first category can identify microRNA targets, but cannot reveal the exact correspondence between microRNAs and their targets; methods like **Hi**gh-**T**hroughput **S**equencing of RNAs Isolated by **C**ross**l**inking **I**mmuno**p**recipitation (HITS-CLIP), **P**hoto**a**ctivatable **R**ibonucleoside-enhanced **C**ross**l**inking and **I**mmuno**p**recipitation (PAR-CLIP), and **I**ndividual-nucleotide Resolution **C**ross-**l**inking and **I**mmuno**p**recipitation (iCLIP) fall into this category. The second category can additionally identify the exact correspondence between microRNAs and their targets, and is exemplified by the methods of **C**rosslinking, **L**igation **a**nd **S**equencing of **H**ybrids (CLASH),

Mapping **R**NA **i**nteractome *in viv**o*** (MARIO), and **P**soralen **A**nalysis of **R**NA **I**nteractions and **S**tructures (PARIS). However, these methods are still not capable of detecting all possible interacting pairs because the ligation approaches used in all these methods for generating chimeras are not efficient and only interacting pairs with particular properties can be ligated (i.e. many interacting pairs cannot be ligated) during library preparation. Accordingly, the representation of the library for the entire population of the microRNA-target RNA interacting pairs is poor and may be biased.

With regard to the restrictions and limitations of current experiment-based methods for identifying microRNA target, this study primarily aims at inventing a comprehensive method for identifying miRNA interactome. Based on a different principle, the current method potentially enhances the generation of microRNA-target RNA chimeras from all interacting pairs during library preparation. In order to develop the current method, several critical techniques have been developed and evaluated. First, a novel approach to conjugating the interacting miRNA-target RNA via a stem-loop adaptor has been developed and evaluated in a set of mock experiments. Second, a pair of "repairers" with enhanced ability to repair the mismatches at 5' and 3' ends of miRNA in the miRNA-interacting duplexes has been invented and evaluated in another set of mock experiments. Third, an approach to enhancing the ligation of adaptor to the 3' end of target RNA has been evaluated. This approach not only is suitable for the "Stem-Loop Adaptor (RNA)" ligation in this method, but also can be used in other RNA interactome identification methods, e.g. HITS-CLIP and PAR-CLIP. However, the invention of "Library Enrichment Beads", which are capable of both capturing the interacting microRNA-target RNA duplexes and eluting the captured sequences under an extremely mild and non-denaturing condition, is still in progress. Currently, the capturing ability of the beads has been solved in general whereas the elution part is still under testing. In fact, these beads can also be

applied to other applications that require on-bead manipulations of nucleic acids if they are finally developed. On the other hand, mass spectrometry analysis has revealed that the purity of immunoprecipitated Argonaute complexes is good enough for evaluating the current method directly. Furthermore, melting curve analysis has found that the melting peak of miRNA-target RNA duplexes is about 40℃ and re-annealing of the melted duplexes is probably impossible once they are melted. This evidence supports the feasibility of the current method. The conditions for immunoprecipitation, “Stem-Loop Adaptor (RNA)” ligation, “Re-attaching Adaptor” ligation, RNA fragmentation and UV crosslinking have been well optimized.

This is a method potentially capable of identifying the entire interactome of microRNA in a specific cell type or tissue, and revealing the interactome changes in different physiological states of cells. Furthermore, the method may also be used to identify other protein-mediated nucleic acid interactions. This newly developed method is called **R**NA **I**nteractome **I**dentification by **N**ext **G**eneration **S**equencing (RIINGS).

Acknowledgements

I would like to express my sincere appreciation and gratitude to my Chief Supervisor Professor Shea-ping Yip for his guidance, encouragement and support throughout this research study, and advices on tricky points in some experiments of this study. I also thank Professor Yip and Dr. Cadmon Lim for encouraging me to purse the PhD study.

I sincerely thank Dr. Helen Law, Dr. Daniel Sze and Dr. Tony To for their encouragement and sharing in some technical aspects.

I would like to thank all the people working in the General Office of my department for their administrative supports. Particular thanks go for Mr. David Lau because he has indeed given lots of help in dealing with the orderings of reagents and special chemicals for this study.

I would also like to thank Dr. Christy Yiu, Dr. Edith Cheung, Dr. Tami Zhang, Dr. Isabella Cheuk, Dr. Supin Koh, Mr. Abdul Mannan, Dr. Amiee Chiu, Dr. Yu-lin Kang, Dr. Feng-feng Wang, Dr. Thomas Sin, Dr. Amiee Pei, Ms. Qian-qian Mou, Mr. Angus Yu, Mr. Ka-keung Yip, Ms. Annie Lee, Ms. Bo-bo Sze, Ms, Wan-jung Chang, Ms. Cara Cheung, Ms. Ashley Li, Mr, Kent Wong, Mr. Marco Tam, Ms. Lin-lin Wang, Mr. Rahim Rajwani, Ms. Sharon Ho and Ms. Michelle Ng for their great friendship and supports.

I am very grateful to my family for their support, patience and understanding.

This theis is dedicated to the memory of my father. He has paid lots of efforts and supports in my study and my life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3C	Chromosome conformation capture
4SU	4-thiouridine
A	Adenosine
AP	Antarctic phosphatase
AGO or Ago	Argonaute
AGO1-4	Argonaute 1-4
Ago HITS-CLIP	Argonaute high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation
AMT	4'-aminomethyltrioxsalen
ATP	adenosine triphosphate
B-CLL	B-cell chronic lymphocytic leukemia
B-PLL	B-cell prolymphocytic leukemia
bp	base pair(s)
BSA	Bovine serum albumin
C	Cytosine
CIP	Calf intestinal alkaline phosphatase
CLASH	Crosslinking, ligation, and sequencing of hybrids
CLL	Chronic lymphocytic leukemia
CML	Chronic myelogenous leukemia
Dcr-1	Dicer-1 gene
Dcr-2	Dicer-2 gene
DDX5	DEAD box protein 5
DMSO	Dimethyl sulfoxide

DNMT	DNA methyltransferase
DTT	Dithiothreitol
EBV	Epstein-Barr Virus
EBNA	Epstein-Barr nuclear antigen
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
G	Guanosine
HBSS	Hank's Balanced Salt Solution
HITS-CLIP	High-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation
HSC70	Heat shock cognate protein, 70 kDa
HSP90	Heat hock protein, 90 kDa
I	Inosine
IAA	Iodoacetamide
iCLIP	Individual-nucleotide resolution crosslinking and immunoprecipitation
IDT	Integrated DNA Technologies
IMDM	Iscove's Modified Dulbecco's Medium
IP	Immunoprecipitation
kDa	kilodalton
lncRNA	long non-coding RNA
Loqs	Loquacious
MARIO	**Ma**pping **R**NA **i**nteractome *in viv**o***
MID	Middle (domain)
miRNA	MicroRNA
MS	Mass spectrometry

NEB	New England Biolabs
NGS	Next-generation sequencing
NMPs	Nucleoside monophosphates
NP-40	Nonidet P-40
NR3C1	Nuclear receptor subfamily 3 group C member 1
nt	Nucleotide
PACT	Protein activator of the interferon-induced protein kinase PKR
PAR-CLIP	Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
PARIS	Psoralen analysis of RNA interactions and structures
PAZ	P-element induced wimpy testes/Argonaute/Zwille
PBS	Phosphate-buffered saline
pCp	Cytidine 3’,5’-bis(phosphate)
PCR	Polymerase chain reaction
piRNA	Piwi-interacting RNA
PIWI	P-element induced wimpy testes
pmol	picomole
PNK	Polynucleotide kinase
Pol I	RNA polymerase I
Poll II	RNA polymerase II
Pol III	RNA polymerase III
PVDF	Polyvinylidene difluoride
qPCR	Quantitative (real-time) polymerase chain reaction
R2D2	R2D2 protein or gene
RBPs	RNA-binding proteins

REST	RE1 silencing transcription factor
Rhed	RNA-binding heme domain
RIINGS	RNA Interactome Identification by Next Generation Sequencing
RISC	RNA-induced silencing complex
RLC	RISC-loading complex
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RPMI medium	Roswell Park Memorial Institute medium
rRNA	Ribosomal RNA
RsAgo	*Rhodobacter sphaeroides* Argonaute
R-SBE	RNA-SMAD-binding element
RT	Reverse transcription
SAP	Shrimp alkaline phosphatase
SDN	Small RNA degrading nuclease
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
siRNA	Small (or short) interfering RNA
snoRNA	Small nucleolar RNA
T	Thymine
T4 PNK	T4 Polynucleotide kinase
TGFβ	Transforming growth factor β
TP53	Tumor protein p53
TRBP	HIV transactivation-responsive RNA-binding protein
tRNA	Transfer RNA
TtAgo	*Thermus thermophilus* Argonuate

U	Uracil
Urea-PAGE	Denaturing/urea polyacrylamide gel electrophoresis
UTR	Untranslated region
UV	Ultraviolet (light)
XRN2	Exoribonuclease 2

1. Introduction

1.1 Epigenetics

The prefix "*epi-*" in Greek means above, on, outside of, over; with this meaning, the compound word "epigenetics" means "besides genetics". The term "epigenetics" was first used in early 1940s by Conrad Waddington [1]. He defined this term in a relatively broad sense as mechanisms by which genes guided a genotype into a particular phenotype during development. Notably, DNA was not yet elucidated as the basis of the hereditary material at that time. Since then, many refinements have been made to the term "epigenetics" through our better understanding of many areas related to epigenetics. Riggs et al. have suggested a relatively narrow definition of epigenetics – "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" [2]. In addition, an operational definition was reached in a meeting of chromatin-based epigenetics held by the Banbury Conference Center and Cold Spring Harbor Laboratory in 2008: an epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" [3]. However, these definitions are not comprehensive enough to be accepted by all the people in the community. For instance, the above two definitions exclude the regulation and alteration processes in gene expression that occur in cases without cell division and in non-dividing cells like neurons [4]. Furthermore, the transient change of histone distribution on DNA and changes in chromosome organization during cell cycle must also be excluded from epigenetics according to the above definitions [5-10]. Apparently, not all these exceptional cases show a valid inheritance, but it is not justified to omit these cases from epigenetics when they actually involve the same changes and mechanisms, e.g. DNA methylation.

Indeed, the prefix "*epi-*" in the compound word "epigenetics" has already implied that inheritance, which is essential for genetics, may not be necessary in the combined word.

Regarding the issue that inheritance may not be compulsory for epigenetics, Holliday has defined the epigenetics as (i) "the study of the changes in gene expression that occur in organisms with differentiated cells, and the mitotic inheritance of given patterns of gene expression", and as (ii) "Nuclear inheritance that is not based on differences in DNA sequence" [11]. Likewise, Bird has defined epigenetics as "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" [12]. Indeed, giving a conclusive definition to this subject is not an easy task because our knowledge of biology is now growing rapidly. Therefore, this controversial argument will persist in this field for some time in the future.

Currently, the NIH Roadmap Epigenomics Mapping Consortium refers epigenetics to "both heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable", and adopts this definition in the project for creating an open resource of epigenomic map of selected primary *ex vivo* tissues and stem cells [13]. In my opinion, epigenetics is a subject that studies the effects, regulations and modifications related to genetic materials, i.e. DNA (in most organisms) and RNA (in RNA viruses), influencing the process of life but not the encoding information in the genetic materials. In fact, epigenetics can be regarded as an extension of genetics and is used to study the phenomena around the genes.

1.1.1 *Influencing factors in epigenetics*

If we classify the factors influencing epigenetics by their origin, there are two major kinds of factors affecting how organisms respond at epigenetic level. One of these is environmental factors while the other is internal factors. The latter one can be alterations acquired during the process of life of an organism or inherited from parents.

The first kind of factors is usually demonstrated in organisms in responding to the environmental change or environmental pressure. One of the examples for the environmental factors is childhood abuse that alters the hypothalamic-pituitary-adrenal (HPA) function in stress response via modifying the methylation pattern on the promoter region of a neuro-specific glucocorticoid receptor called nuclear receptor subfamily 3 group C member 1 (*NR3C1*) [14-18]. McGowan and his colleagues compared suicide victims with childhood abuse (here the abuse including severe physical abuse, sexual contact and severe neglect) with either suicide victims without childhood abuse or control cases who died suddenly from unrelated causes [18]. The investigation examined the *NR3C1* gene in hippocampal samples because the expression level of this gene in hippocampus was closely related to HPA responses to stress in rodents [17]. The results revealed that the degree of methylation at *NR3C1* promoter region was elevated only in the group of suicide victim with childhood abuse history. Accordingly, the expression levels of the messenger RNA (mRNA) for the corresponding gene variants were also decreased in the group of suicide victim with childhood abuse history [18]. This example indicates that the environmental pressure (childhood abuse) is a factor that can alter the methylation pattern on DNA (*NR3C1* promoter region) and further affect the phenotype (HPA function in stress response) of an organism. In addition, the hippocampus cells have not changed anything at the DNA sequence level, but modification on DNA alters the expression of a particular gene

and a different phenotype is produced epigenetically in response to the environmental pressure in this example.

The second kind of factors is an internal factor that is displayed in organisms in receiving and inheriting effects, regulations or modifications at a level not related to genetic coding. This factor includes two major sources: alterations acquired during the process of life of an organism or inherited from parents.

Examples for the first kind of internal factors include ageing and tumorigenesis. During the process of life, the levels of methylation on DNA decrease globally by ageing. The related evidence can be traced back to 1973 in a study of albino rats [19]. This investigation revealed that the total amount of 5-methylcytosine in DNA from some of the studied organs, including brain, heart and spleen, decreased gradually with age [19]. This observation has also been recognized in bronchial epithelial cells and leukocytes in humans [20, 21]. However, with evidence from a longitudinal follow-up study of DNA methylation in two cohorts each of more than 100 subjects with blood samples collected at two visits over 10 years apart, global DNA hypomethylation may not be the only truth [22]. On the contrary, global DNA hypomethylation and hypermethylation could occur in different individuals during ageing with nearly the same frequency [22]. Indeed, other than global hypomethylation of genomic DNA during ageing, hypermethylation of specific loci have been reported in various studies (recently reviewed by Zampieri et al. and Jung et al. [23, 24]). Interestingly, some of these local hypermethylation events are detected in the critical regions involved in the pathogenesis of cancer, e.g. the CpG islands in the promoter regions of *RUNX3*, *LOX*, *p16INK4A*, *TIG1*, *RASSF1* and *CHD1* [25, 26]. These genes are known to contribute to tumor suppression and this is perhaps why aged people have higher risks of cancers. In addition to ageing-induced alteration in methylation patterns, cancer can also

be caused by the accumulation of epigenetic changes. These changes include DNA hypomethylation and hypermethylation caused by multiple reasons (e.g. dysfunction of DNA methyltransferases [DNMTs]), histone modifications, deregulation of microRNA (miRNA) expression, chromatin remodeling and deregulation of long non-coding RNA (lncRNA) expression [27, 28].

On the other hand, an example for inherited epigenetic traits that can change phenotypes comes from the work by John Gurdon. As has been reported by him, mature cells can be reprogrammed to become pluripotent again. This finding actually proves the following statement wrong: "a mature or specialized cell cannot return to an immature state". In 1960s, John Gurdon evacuated the nucleus of a fertilized egg cell from a frog and replaced it with a nucleus from an intestinal cell of a tadpole. This modified egg cell subsequently grew into a new frog [29, 30]. This experiment proves that the mature cell still contains the genetic information required for differentiation and the differentiation potential is inherited from the cytoplasm of the fertilized egg, instead of the genetic information from the DNA. Put in his own words, “there may be a memory of a pattern of gene expression characteristic in the enucleated egg” [31]. John Gurdon was awarded in 2012 the Nobel Prize in Physiology and Medicine because of his significant contribution to this field.

In fact, phenotypic changes and differentiation are complex results of epigenetic changes. These epigenetic changes are affected by a combination of influencing factors. For instance, changes in methylation pattern can be caused by ageing, DNMT dysfunction or even environmental pressures. Moreover, the effect of each factor is usually not restricted to one gene or one kind of modification. Therefore, each phenotype should be the result of a crosstalk of multiple influencing factors and understanding of the relationship among these influences is a difficult biological question in current biology.

1.1.2 Epigenetics machinery

As epigenetics concerns a broad range of effects and regulations involved in phenotypic manipulations, the epigenetic machinery must include various mechanisms accordingly. Actually, organisms regulate themselves at epigenetic level via a variety of approaches. Basically, epigenetic regulations are a very complicated pool of chemical reactions within organisms. This chemical reaction pool may involve many clusters of chemical reactions. Unlike the chemical reactions in vitro, these chemical reactions are usually under control and affect each other. For instance, the methylation of DNA can be controlled by several enzymes in the DNMT family. The malfunction of DNMTs would cause the loss of methylation imprinting on DNA [32].

Generally speaking, epigenetic regulations can occur at DNA level via DNA methylation, and chromatin level via histone modification and/or chromatin remodeling. DNA methylation is one of the most well-studied mechanisms in epigenetic machinery. DNA methylation is the addition of a methyl group to the cytosine ring at the 5' carbon. This modification usually occurs at CpG islands located at the promoter regions and telomeric regions [33]. Methylation of the CpG island in the promoter region of a gene usually results in transcriptional repression of that gene via recruiting methyl-CpG binding domain proteins to the methylated site and preventing transcription factors from binding to the promoter region [34, 35]. There are many different types of histone modifications. Common histone modifications include acetylation, methylation, phosphorylation and ubiquitination. Different combinations of modifications in a specific genomic region are capable of repressing or promoting certain nuclear processes, e.g. gene expression and DNA repair [36, 37]. Chromatin remodeling is the topic concerning the conformational change of chromatin. Chromatin can be shaped by ATP-dependent chromatin remodelers [38], transcription

factors [39], and non-coding RNAs [40]. The changes in conformation can also affect nuclear processes and gene expression.

Moreover, epigenetic regulations can also occur at RNA level such as RNA silencing, RNA splicing and RNA methylation. RNA silencing is a group of mechanisms important in gene regulation at post-transcriptional level. These mechanisms control the expression level of genes and in turn affect the phenotype of organisms. A key relevant mechanism is mediated via microRNA (miRNA) [41]. RNA splicing generating different isoforms of mature RNAs from a transcript is essential in RNA biogenesis. The expressed RNA isoforms may have distinct functions or are translated into different protein isoforms with diverse functions; for example, the different isoforms of sex lethal (Sxl) determine the sex of fruit flies [42]. RNA methylation has been identified decades ago. However, our knowledge of RNA methylation is very limited when compared to DNA methylation because we have appropriate methods for investigating DNA methylation very early on. Various types of methylation on RNA bases exist and are important in making RNAs to be functional [43].

Furthermore, epigenetic regulations can occur at protein level, e.g. protein ubiquitination, protein phosphorylation, protein glycosylation, protein lipidation, prionic transmission, etc. [44]. In a broad sense, the epigenetic regulations also include signal transmission by some small molecules. For example, the hormone estradiol is an inducer for the expression of vitellogenin by inducing a specific demethylation at the estradiol response element of DNA [45].

In brief, epigenetic regulations at the above-mentioned four different levels always cooperate and cross-talk with each other and result in the regulation of gene expression, biomolecule activation and inactivation, modifying biomolecules and other small molecules,

transportation of biomolecules and other small molecules, signal transmission, etc. Therefore, epigenetics is fundamentally a subject studying the regulation of chemical reactions in life.

1.1.3 Study of epigenetics

Indeed, epigenetic regulations are complex chemical reactions from the perspective of molecular biology. With the intention of investigating the regulation of epigenetics, we have to examine all the interactions between molecules in cells, i.e. the interactions between reactants and catalysts. However, this work is extremely difficult because the reactants usually interact in a many-to-many manner and change under different dynamic conditions, e.g. protein-protein interaction, protein-nucleic acid interaction, chromatin interaction and gene expression regulation by miRNAs. That is why we need to spend a very long time to understand a single biochemical pathway, e.g. the Krebs cycle. Unfortunately, epigenetic regulation is not just a single pathway, but rather a complex array of chemical reactions involving many pathways and reactants. This is the reason why the study of epigenetics is difficult and usually stays at the level of types of regulations with lots of uncertainties about the details for a particular type of regulation. For instance, we know miRNAs are able to regulate the expression of genes, but we do not know the *whole* picture of which particular miRNA interacts with which target(s) and which particular target is regulated by which miRNA(s).

1.2 Gene – the basic element of life

Genes are the fundamental units of heredity in any organism. The term “gene” was first introduced by Johannsen in 1909. However, the foundation of the mechanism of inheritance in modern genetics was introduced by Mendel in the nineteenth century [46]. One of the most important and well-studied functions of genes is encoding proteins. In this case, the sequence of nucleotides determines the sequence of a peptide product or the primary structure of a protein because protein synthesis is directed by the genomic information. In eukaryotes, the process of gene expression starts with transcription of the template DNA into RNA in the nucleus and then proceeds to translation of RNA into peptide or protein in the cytoplasm. Other than expressing proteins, genes also encode non-coding RNAs (RNAs are not translated to produce proteins) such as ribosomal RNA (rRNA), transfer RNA (tRNA), lncRNA and miRNA. The final products in these situations are RNAs only.

Genes exist in every organism and are the elementary units of life. In addition, genes collaborate with each other in every living organism and construct an elegant collaborating network for making life different from death. The expression of each particular gene is under a stringent control. This control can be a single regulation or a cascade of regulations. These regulations can act at transcriptional level, translational level or both levels. They can be achieved by various means, e.g. DNA methylation and demethylation, and miRNAs. In summary, the controls are crucial to the collaborating network of genes. With the genes and the controlling system, life can have specialties. This machinery drives all the cellular processes from life-maintaining metabolisms in cells to the construction of the whole body of multicellular organisms, e.g. cell differentiation and morphogenesis in embryos. Different types of cells possess different profiles of expressed genes from the same

genome. Therefore, the existence of genes and the regulation of gene expression are of supreme importance to all life forms and generate life-specific processes.

1.2.1 Expression of genes

In all prokaryotes, eukaryotes and archaea, gene expression refers to the process of producing proteins or RNAs from the information encoded in the DNA of that gene. Therefore, the genes are either protein-coding genes or non-protein-coding genes (also called non-coding genes). Protein-coding genes require two individual steps to express the encoded proteins. These steps are transcription and translation. Alternatively, non-coding genes only involve the transcription step to achieve the expression of non-coding RNA [47, 48].

In eukaryotes, the process of transcription takes place in the nucleus and on the DNA. In transcription, only one strand of the double-stranded DNA is used as the template to synthesize the RNA transcript. RNA polymerase I (Pol I), RNA Polymerase II (Pol II) and RNA polymerase III (Pol III) are the three DNA-dependent RNA Polymerases responsible for converting the DNA messages into different types of RNAs in eukaryotes [49]. Additionally, all these polymerases are primer-independent polymerase and synthesize RNA complementary to the antisense DNA template in a 5'-to-3' direction [49]. Pol I is responsible for synthesizing the precursors of most rRNAs except 5S rRNA [50]. Pol II is responsible for synthesizing the precursors of messenger RNAs (mRNAs), small nuclear RNAs and microRNAs [51, 52]. Pol III is responsible for synthesizing the precursors of 5S rRNA, tRNA and some small RNAs [53, 54].

1.2.1.1 Transcription

The transcription process can be subdivided into three major steps: initiation, elongation and termination [49].

First, the initiation of transcription requires the formation of a transcription complex at the promoter region. RNA polymerase first binds to the promoter of the gene to be transcribed. Then, the polymerase needs to recruit other accessory proteins to form the transcription complex known as “closed complex” because the DNA in these complexes is in duplex form at this stage. The closed complex is usually located at ten to several tens of bases upstream from the transcription start site and some special sequence elements exist in the promoter regions very often for recognition by the transcription complex. For example, the TATA box is usually located at position -25 to -30 (i.e. 25-30 bases upstream from the transcription start site), and is a common elementary sequence for Pol II involved transcription. Then, the RNA polymerase unwinds the DNA, usually less than 20 base pairs (bp), under the assistance of one or more transcription factors to form a so-called “open complex”. After the open complex is formed, ribonucleotides (ATP, UTP, CTP and GTP) are sequentially bound to the template DNA and condense together in a 5’-to-3’ direction to initiate the synthesis of complementary RNA from the transcription start site (position +1). This initial process is a primer-independent process [49].

Second, the newly synthesized RNA chain is elongated by the same RNA polymerase through the addition of ribonucleotides to the 3’ end of the growing RNA chain. During the elongation, the RNA chain grows in the 5’-to-3’ direction while the RNA polymerase reads and moves along the antisense DNA template in a 3’-to-5’ direction. Additionally, the double-stranded DNA is locally unwound in agreement with the movement of the RNA polymerase to expose the template strand for transcription. The synthesized RNA is

released from the template DNA during the elongation step and the DNA duplex is reformed again upon departure of the RNA polymerase [49].

Third, the transcription in prokaryotes terminates at some special DNA sequences known as terminator. These DNA sequences usually contain self-complementary components so that the components are able to form hairpin or stem-loop structure. These secondary structures are very important to the transcription halt and the dissociation of RNA polymerase. Furthermore, some terminations are accomplished by the terminator sequences alone while other terminations require Rho factor as a co-factor to terminate the transcription in prokaryotes [49]. However, the mechanism of transcription termination in eukaryotes is less studied than that in prokaryotes [49, 55].

1.2.1.2 Translation

Translation is the process of synthesizing a polypeptide or the primary structure of a protein by using the mature mRNA as the guiding template. Like transcription, the translation process can also be divided into three major steps: initiation, elongation and termination.

First, the initiation of translation is the assembly of a ribosome, which includes a 40S subunit and a 60S subunit, on an mRNA in eukaryotes. In prokaryotes, the two subunits of the ribosome are 30S subunit and 50S subunit [49]. The assembly of a ribosome is different between prokaryotes and eukaryotes. The overall process involves 3 initiation factors in prokaryotes, but at least 12 factors in eukaryotes. However, the major route of the two processes is still common in general. The assembly of a ribosome on an mRNA begins with the formation of small subunit-mRNA complex (called pre-initiation complex). The complex includes a ribosome small subunit (30S subunit in prokaryotes and 40S subunit in

eukaryotes), an initiator tRNA and a mature mRNA. The pre-initiation complex then moves to the start codon of the mRNA with the assistance of the initiator tRNA. The start codon is AUG. After the small subunit correctly locates itself at the translation start site, the large subunit (50S subunit in prokaryotes and 60S subunit in eukaryotes) binds to the pre-initiation complex to complete the assembly of a ribosome [49].

Second, the elongation of the peptide sequence immediately follows the assembly of ribosome. Amino acid delivery, formation of peptide bond and movement of the ribosome along the mRNA (so-called translocation) are sequentially repeated during the elongation process. Each ribosome has two critical sites, the P-site and A-site, for mounting the two aminoacyl-tRNAs inside the ribosome. At the beginning, the P-site of the ribosome has been already occupied by the initiator tRNA and the A-site is free. The initiator tRNA recognizes the start codon and is charged with methionine at its 3' end. The growing of the peptide requires a second aminoacyl-tRNA, which recognizes the codon in the A-site and moves into the A-site. Then, the two adjacent amino acids on the two tRNAs are joined together via the formation of a peptide bond and the peptide grows in a direction from N terminal to C terminal. Meanwhile, the aminoacyl-tRNA in the P-site discharges the amino acid it owns. Subsequently, EF-G-GTP complex (in prokaryotes) or eEF2-GTP complex (in eukaryotes) binds to the ribosome. This motion causes the ejection of the discharged tRNA from the P-site and the translocation of the peptidyl-tRNA from the A-site to the P-site by moving one mRNA codon in a 5'-to-3' direction relative to the ribosome. This peptide construction cycle is repeated until a stop codon is encountered in the A-site of the ribosome. Like the initiation step, elongation requires different sets of factors to assist in the process in prokaryotes and eukaryotes [49].

Third, termination is carried out by the binding of a GTP-dependent release factor in eukaryotes. This factor is called eukaryotic release factor (eRF), which can recognize all the three stop codons. In contrast, prokaryotes require three release factors to recognize the three different stop codons and release the completed peptide chain. Release factor 1 (RF1) recognizes the stop codons UAG and UAA while release factor 2 (RF2) recognizes the stop codons UGA and UAA. Both the RF1 and RF2 need the assistance of release factor 3 (RF3) when functioning. In both prokaryotes and eukaryotes, the release factor(s) lead the C terminal of the polypeptide to react with a water molecule, instead of the usual aminoacyl-tRNA, and finally release the synthesized peptide from the ribosome [49].

1.2.2 Modifications of genes

The expressed products, including proteins and RNAs, are the fundamental, functional units in all lives. Before they function, modification may be required to reshape their conformation or molecular architecture to finalize their precise function. On the one hand, change in conformation only requires altering the conformation of the expressed products and does not need to change the chemical formula and structural formula of the products. An example for this case is prion. Prions, also known as prion proteins, possess isoforms with infectious property. The prion forms of proteins can promote the conversion of the properly folded normal proteins into the misfolded prion protein forms and further expand in a chain-reaction manner [56, 57]. Indeed, prion proteins are the guided proteins to make more normal proteins to re-fold into prion protein forms and thus change the function of the protein by changing the conformation. Additionally, the proper folding of proteins is usually driven by molecular chaperones (a class of proteins helping the folding of proteins) and begins co-translationally [58]. Without the assistance of molecular chaperones, nascent proteins probably cannot fold into their correct conformation. On the other hand,

change in molecular architecture requires altering the chemical structure of the expressed products, including change in the primary sequence of the products, attachment of some chemical functional group(s) to the products, removal of some group(s) from the products, and change of some bonds of the products intra- or inter-molecularly without altering the chemical formula. Several important machineries are described below.

RNA splicing is a mechanism that edits the sequence of RNA transcripts, and exists in all the kingdoms or even in all the domains of lives. This process usually occurs in nascent mRNA precursors in eukaryotes, but in non-coding RNAs (e.g. rRNA and tRNA) in prokaryotes, [49, 59, 60]. After RNAs are transcribed from DNA, some of the internal regions of the pre-mature RNAs are to be removed. These internal regions are called introns. After the removal of introns, the adjacent portions of the cleaved RNA are joined together to generate the mature RNA. These retained portions are called exons [49]. This splicing process can occur simultaneously during the process of transcription [61, 62]. The process of RNA splicing can be achieved in multiple ways: spliceosome-mediated splicing [63], self-splicing [64], and ribozyme-catalyzed splicing [59, 65]. For most eukaryotic mRNAs, the region(s) to be spliced together on each particular mRNA can vary. As a result, more than one spliced form of mature mRNAs can be generated from the same transcript of mRNA precursor and these mature mRNA variants can be used to produce different isoforms of proteins accordingly. This phenomenon is called alternative splicing [42].

RNA modifications can affect the stability, localization, structure and functions of both coding and non-coding RNA transcripts. To date, more than a hundred kinds of RNA modifications have been identified in distinct species of RNA transcripts, e.g. tRNA, mRNA, rRNA, small nucleolar RNA (snoRNA), small nuclear RNA, miRNA [66, 67]. However, the exact functions of most of these modifications are not clear. Among the modifications, N^7-

methylguanosine is one of the most understood modifications. It is a necessary modification at the 5' end of mRNAs and plays an important role in mRNA stability [68], exportation [69] and RNA splicing [70]. Other examples of modified bases include pseudouridine, N^6-methyladenosine and 5-methylcytosine. These three modifications can alter the meaning of the three stop codons to amino-acid-coding codons on mRNAs [71], and contribute to mRNA splicing [72], stability [73], translation efficiency [74], pri-microRNA processing [75], etc. They can increase the force of base-stacking via altering the hydrophobic property and alter the folding of RNA [76-78].

Protein modifications include various types of modifications to the polypeptide chains of proteins. These modifications are usually enzymatic covalent modifications that are made during peptide synthesis or after protein synthesis. Protein modifications are essential in the maturation and the functions of proteins. By changing the phosphorylation status of a protein, the function of this protein can be activated or deactivated. Phosphorylation is the most frequent modification (Table 1.2.1), as is evident from the results obtained from the Proteome-Wide Post-Translational Modification Statistics Curator, which automatically counts the number of each reported post-translational modification in the Swiss-Prot Knowledgebase by a computation-based method [79]. Many proteins also undergo glycosylation during their maturation. This is a process of attaching a carbohydrate molecule to a targeted site of a protein. A single protein can have several glycosylation modifications. Glycans, which can be used to describe the attached carbohydrate moieties, influence the folding and stability of a protein; for example, the ABO blood group antigens are based on carbohydrates and play a crucial role in immunity [80]. Moreover, glycosylation is also used by viruses to evade the immune recognition by the host, e.g. human immunodeficiency virus [81]. Table 1.2.1 is a summary statistics of the frequencies of experimentally identified protein modifications. Data were retrieved from Proteome-

Wide Post-Translational Modification Statistics Curator on 7 March 2017 via http://selene.princeton.edu/PTMCuration [79].

Table 1.2.1 Frequencies of experimentally identified protein modifications. (Khoury GA, Baliban RC, Floudas CA: Proteome-wide post-translational modification statistics: frequency analysis and curation of the Swiss-Prot database. Scientific Reports 2011, 1:90. http://selene.princeton.edu/PTMCuration, retrieved on 7 March 2017)

Modification	Frequency
Phosphorylation	58383
Acetylation	6751
N-linked glycosylation	5526
Amidation	2844
Hydroxylation	1619
Methylation	1523
O-linked glycosylation	1133
Ubiquitylation	878
Pyrrolidone Carboxylic Acid	826
Sulfation	504
Gamma-Carboxyglutamic Acid	450
Sumoylation	413
Palmitoylation	305
Myristoylation	178
ADP-ribosylation	152
C-linked glycosylation	147
Farnesylation	81
Nitration	65
S-nitrosylation	62
Geranyl-geranylation	56
Citrullination	55
Formylation	55
Deamidation	53
S-diacylglycerol cysteine	37
GPI anchoring	34
Bromination	33
FAD	19
Others	7208
Total Processed	89390

1.2.3 Regulation of gene expression

Regulation of gene expression is the term describing the control of increase or decrease of the production of gene products, including proteins and RNAs. The control can be accomplished by a wide range of regulatory mechanisms. Generally speaking, regulation is carried out at transcriptional and post-transcriptional levels. It also includes the process of degradation of mRNAs and gene products.

First, the expression of genes begins with transcription. At the transcriptional level, the expression of genes can be regulated by modifying the DNA or chromatin, such as DNA methylation, histone modification and chromatin structure. These machineries belong to epigenetic regulation and have been mentioned in Section 1.1.2. Moreover, the regulation can also be achieved via the binding of some factors to the DNA to be transcribed. These factors are known as activators and repressors. Activators usually enhance the interaction between the transcription initiation site on DNA and the enzyme RNA polymerase, and promote the expression of the corresponding RNA product. Activators do this via binding to specific DNA sites called enhancers [82]. On the other hand, repressors prevent the interaction between the transcription initiation site (on DNA) and RNA polymerase, and inhibit the expression of the corresponding RNA product. Repressors do this via binding to specific DNA sites called silencers [82]. Besides, a group of transcriptional regulatory elements and DNA loci also serves to regulate or alter the expression of genes at transcriptional level. Indeed, regulation of gene expression at the transcriptional level covers a broad spectrum of control in which different elements on DNA, their binding factors and co-factors work together and contribute their impact on the promoter of a particular gene. Ultimately, the final expression level of this particular gene is determined.

Second, the transcribed RNA products are subject to a second level of regulation after transcription – the so-called post-transcriptional regulation of gene expression. MicroRNAs play a key role at this level of regulation. They are also the transcriptional products from DNA and usually regulate the stability of their targets or silence them. They function through a protein-mediated direct binding onto their targets. Afterwards, different sets of assisting factors are recruited for initializing the degradation or silence of their target [41, 83].

In gene expression studies, the regulatory machinery at the transcriptional level has been identified early on. However, this machinery is not enough to explain all phenomena in gene expression. For example, we sometimes find that the expression level of the mRNA of a gene is not changed although the protein level of that particular gene is down-regulated. The regulation of expression of these genes seems to violate our common knowledge of gene expression because the expression levels of an mRNA and its corresponding protein do not correlate. Furthermore, the response in the expression of a particular gene is expected to be slow if the regulatory machinery only acts at the transcriptional level. It is because the cell must first express the repressor/activator (proteins) for inhibiting/promoting this particular gene. Then, the expressed repressor/activator can bind to the silencer/enhancer (elementary sequence on DNA) to inhibit/promote the production of mRNA of that particular gene. Therefore, this is not an effective way of providing cells a fast response. Considering all these implications, it should not be too difficult for us to believe that there should be additional machinery for regulating the expression of gene at another level. However, this had been a difficult task until the two key discoveries of miRNA in early 1990s [84] and in 2000 [85]. These two key findings have motivated us to study the post-transcriptional regulation of gene expression — the miRNA machinery. To

date, we still do not fully understand the miRNA machinery in the regulation of gene expression such as target identification of miRNAs.

Third, degradation is also important in regulating gene expression. Both the RNA products and protein products of genes are bound to be degraded eventually. In general, many of the mechanisms involving protein factors are responsible for degrading coding RNAs or non-coding RNAs or both. Some examples of mechanisms functioning to degrade both coding and non-coding RNAs in eukaryotes include deadenylation, uridylation, decapping, XRN1-dependent 5'-to-3' decay, DIS3L2-dependent 3'-to-5' decay, AU-rich element decay, nonsense-mediated decay, and small RNA-induced decay [86-88]. Indeed, the degradation processes of some kinds of RNAs are known, but others are still unclear or not known. On the other hand, proteolysis (or protein degradation) is critical to protein turnover. Eukaryotic cells achieve proteolysis in two major ways: ubiquitin-dependent protein degradation and proteolysis in lysosome. Being a selective process, ubiquitin-dependent pathway starts with marking targeted proteins by covalently labeling them with ubiquitin. The ubiquitinated proteins are then transported to proteasomes for degradation by their hydrolytic activity and ubiquitin is released for reuse [89]. Proteolysis in lysosome is a non-selective process, but can become selective in some kinds of autophagy processes such as chaperone-mediated autophagy [90]. The degradation takes place in lysosomes, which are membrane-bound vesicles with various hydrolytic enzymes inside [91]. In fact, both RNA degradation processes and protein degradation processes play an essential role in maintaining the dynamic balance between production and decay of gene products, which provides an equilibrated environment for all living cells.

All in all, physiological changes requires regulation of gene expression, such as differentiation, cell division, apoptosis, cell signaling, and immune response and even

responses to the attack of diseases as in tumorigenesis and viral infections. Therefore, the regulation of gene expression is crucial to all kinds of lives.

1.3 A brief introduction of microRNAs

Generally speaking, miRNAs are small non-coding RNAs expressed in a broad range of eukaryotic cells from animals to plants and even in virus, and are 21 to 25 nucleotides (nt) long [92, 93]. They mainly serve to repress target mRNAs from being translated into proteins although miRNAs have also recently been shown to stimulate mRNA expression into protein via binding to 3' end AU-rich region of mRNA [94]. In animal cells, miRNA achieves the inhibition by binding to the 3' untranslated region (UTR) of target mRNAs via partial complementarity, and subsequent silencing of the protein translation of mRNAs or degradation of mRNAs. The copy number of mRNAs is not changed in the former situation, but is reduced in the latter situation. Both of these situations are the *direct* effects of miRNAs. In fact, a particular miRNA may interact with a single target RNA or many other target RNAs (transcribed from different genes) via partial complementarity. On the other hand, a particular gene can be regulated by many different miRNAs too. Since the pairing between miRNAs and target RNAs requires only partial complementarity, we cannot reliably identify their interactions through sequence alignment. Therefore, the target identification of miRNA interactome is difficult. Here, a term "identifying miRNA-target RNA interactome in exact correspondence" has been used to describe the exact mapping of this matching network in this thesis.

Furthermore, miRNAs can also affect secondary targets through the primary targets because genes regulate and interact with each other. A change of expression of one gene

will always trigger some other responses. These responses can be the change of expression of other gene(s), activation or modification of gene(s), change of something at DNA or RNA level, etc. All of these effects can be defined as *indirect* effects. In addition, there are an unpredictable large number of targets for the indirect effects. This means that a single miRNA can regulate the expression of many genes at a post-transcriptional level including primary and secondary targets. Thus, it has an important role in regulating gene expression and is involved in many cellular processes, including embryonic development, cell differentiation and tumorigenesis.

1.3.1 Unmasking the cryptic microRNAs and the establishment of miRBase

The first microRNA (miRNA) was discovered in the early 1990s in *Caenorhabditis elegans* by Lee et. al. and named *lin-4* [84]. Interestingly, the coding region of *lin-4* does not encode any protein, but is transcribed to give two short RNAs that are complementary to the 3' UTR of *lin-14* mRNA and inhibit the expression of lin-14 protein at mRNA level [84]. This was the first discovery of post-transcriptional regulation of gene expression. Coincidentally, Wightman et. al. also identified the same phenomenon at nearly the same time [95]. Because *lin-4* was conserved within the nematode species only, the lin-4 post-transcriptional regulatory phenomenon was thought to be a distinctive regulatory mechanism restricted to nematodes only in the early years. The second miRNA, *let-7*, was found in 2000 [85]. Unlike *lin-4*, *let-7* is widely conserved in different species including *Homo sapiens* [96]. Moreover, another hint also suggested that the miRNA machinery was not an oddity: the maturation of *lin-4* and *let-7* were found to be related to Dicer, an enzyme phylogenetically conserved in different species [97, 98]. In fact, these findings provided a new insight into miRNA as a widespread phenomenon in biology, rather than

just a specific trait of nematodes. Since then, scientists spent another decade to unmask other cryptic miRNAs and understand their functions, biogenesis and mechanisms of action.

The early years of miRNA study witnessed a lot of difficulty because miRNA was a kind of RNAs very new to biological research. One very distinctive property of miRNAs is their short length – just 21 to 25 nucleotides long [92, 93]. This specific feature made many conventional methods and tools commonly used in RNA research inapplicable to miRNA study. For instance, the ordinary column-based RNA extraction method, which is often used in total RNA extraction, is not suitable for the extraction of miRNAs because the silica membrane used to capture the ordinary RNA (long RNA) is unable to capture small-sized RNAs [99]. Moreover, we cannot use the conventional method of reverse transcription (RT) coupled with polymerase chain reaction (PCR) (or RT-PCR) to amplify such short RNA sequence either because the length of the primers are nearly equal to the length of miRNAs. In fact, the above two experimental techniques are the indispensable, fundamental procedures in nucleic acid research. Since total RNA extracted by column-based methods do not contain miRNAs and investigation of target nucleic acid sequences are very difficult without amplification, the progress of miRNA research was very restricted in the early years. This is the reason why our knowledge of miRNAs was very limited for a long period, which is a good example of "technologies limit our knowledge".

In the early years of studying RNA interference, a research group led by Tuschl T developed a cloning-based method to explore the mechanism of RNA interference induced by small interfering RNA (siRNA) [100]. In brief, short RNAs were prepared from long double-stranded RNA treated with *Drosophila* embryo lysate, and then purified. Two adaptors were ligated to the 3’ and the 5’ ends of the purified product by T4 RNA ligase. RT-PCR was used to amplify these target sequences. Subsequently, concatemerization, cloning and

sequencing were used to identify the sequence of these short RNAs (Figure 1.3.1) [100]. In addition to the actual sequence of the siRNAs, the study also identified a large number of endogenous *Drosophila* short RNAs (~21 nt). Initially believed to be retrotransposons [100], these 21-nt RNAs were in fact 16 novel miRNAs [101]. Moreover, these researchers also identified 21 more novel human miRNAs in HeLa cells with the same technique [101]. In the meantime, 52 and 15 novel miRNAs were also identified in *Caenorhabditis elegans* by two other research teams led by Bartel and Ambros, respectively, using the same strategy [102, 103].

However, one significant drawback of this cloning method is that it cannot identify miRNAs on a massive scale. Moreover, this method is also labor- and time-intensive. Thus, the work of identifying the cryptic miRNAs in early years was particularly difficult. Furthermore, there is another important problem always being neglected in this cloning strategy: the short RNAs are easily self-ligated during ligation process [100, 102]. As a result, a lot of circularized miRNAs form and hence cannot be cloned and sequenced. With this limitation, only a small proportion of highly expressed miRNAs can still be detected, but others are not. To overcome this obstacle, Tuschl T applied a dephosphorylation step catalyzed by phosphatase to the purified ~21-nt RNA sample before ligating adaptors [100]. This step removes the 5' phosphate group from the miRNAs and prevents their self-ligation. However, this step also removes 3' phosphate group from those partially-degraded RNAs and makes these unwanted RNAs to be the substrates for T4 RNA ligase. On the other hand, Bartel used a more favored approach to solving this problem. This approach relies on the ligation of a pre-adenylated adaptor to the 3' end of miRNAs in an ATP-independent manner [102]. As such, the ligation does not require ATP as the activator in the T4 RNA ligase-mediated ligation. The ATP-free reaction prevents the miRNAs from self-ligating during the ligation step. This approach essentially prevents both self-ligation of miRNAs and

ligation of adaptors to partially degraded RNAs. Both strategies are illustrated in Figure 1.3.1.

Since then, these two different strategies of investigation have attracted many groups of researchers for identifying novel miRNAs. Some try to make improvements on the experimental method and even couple it with next-generation sequencing (the latest sequencing technology that can sequence millions of sequences in parallel) for enriching the content of miRNAs in the prepared library and increasing the throughput of library sequencing [104-107]. Others try to explore the novel miRNAs in different kinds of samples: different species of organisms, different organs and cell types, normal and malignant cells and tissues from various kinds of cancers, tissues of different differentiation states during development and responses during infections [108-113], and even viruses [114, 115]. Figure 1.3.2 shows the growth in the total number of miRNAs from multiple species from 2002 (version 1 of the database miRBase) to 2014 (version 21).

As miRNAs are confirmed to be a new class of functional non-coding RNAs and new identities of this class of RNAs have been discovered rapidly, a miRNA database was developed in late 2003 with the miRNA registry coming from Rfam (a database of non-coding RNAs released in early 2003) [116]. The database developed by Griffiths-Jones was a separate database collecting miRNAs only [117], was finally named miRBase [118]. The development of miRBase has two major objectives. The first aim is to provide a service of gene name assignment for novel miRNAs before they are published. This can prevent the overlap of gene names. The second aim is to provide comprehensive and searchable resources of published miRNAs to the public via the internet [117-121]. Thus, we can now search and download information of miRNAs collected in the miRBase, such as the name and the sequence of the miRNAs, their precursors (including pre-microRNA and pri-

microRNA), annotations, and deep sequencing data. The latest version of the miRBase (release 21.0; June 2014) contains 28645 miRNA entries. These entries include the miRNAs from plants to animals and viruses including organisms from chromalveolata and mycetozoa. The miRBase database is available at http://microrna.sanger.ac.uk/ [117-121].

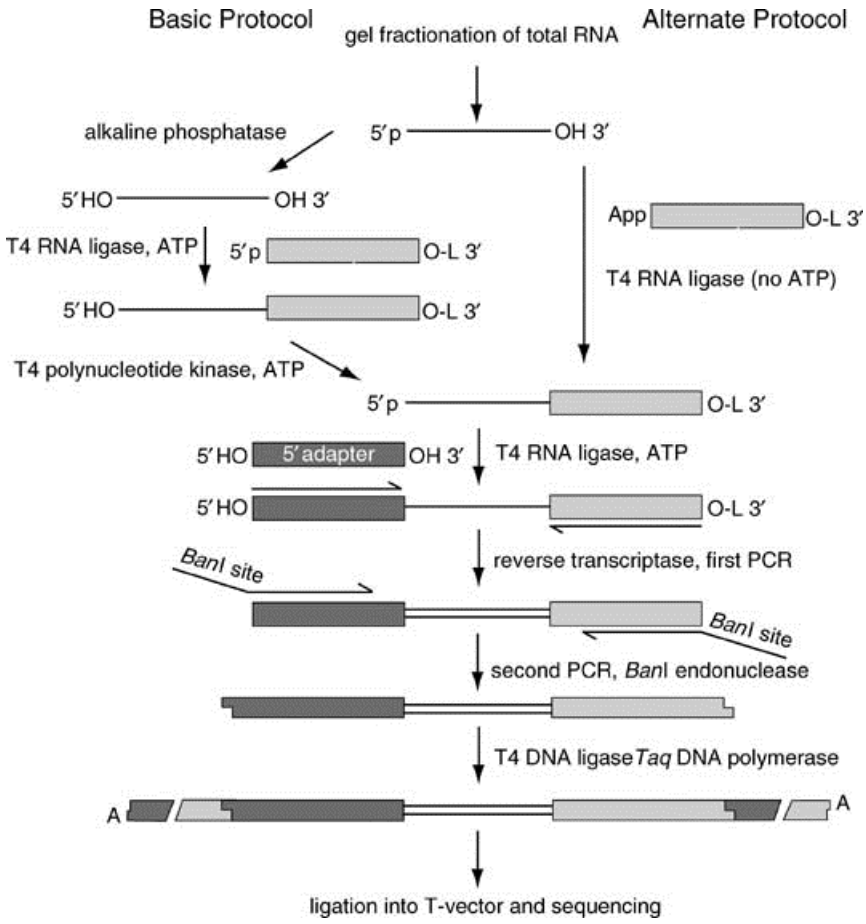


Figure 1.3.1 The two approaches to miRNA identification. The basic protocol was developed by Tuschl T in 2001 [100] while the alternate protocol was developed by Bartel DP in 2001 [102]. **(Adopted from Current Protocols in Molecular Biology. 2005; Unit 26.4)**

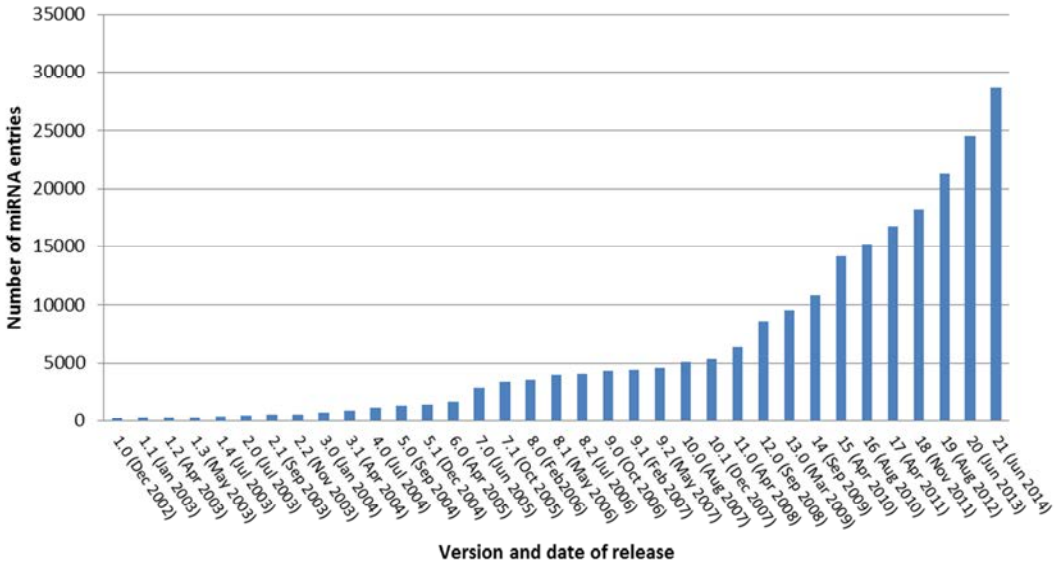


Figure 1.3.2 The total number of miRNAs from multiple species from 2002 to 2014. Data are obtained from miRBase (http://microrna.sanger.ac.uk/) and retrieved on 19 March 2017.

1.3.2 MicroRNA biogenesis, function and degradation in cells

The primary precursors of miRNAs, known as pri-microRNA or pri-miRNA, are the miRNA precursors that are each approximately one thousand nucleotides long, and sometimes contain several clustered miRNAs [122]. In the canonical pathway, pri-microRNAs are transcribed by RNA polymerase II and III [51, 54, 123] from either introns of protein-coding genes and non-coding genes (co-transcribed with the host-gene) or independent genes [124]. Following transcription, pri-microRNAs are cut by the enzyme Drosha, a member of the RNase III superfamily consisting of double-stranded RNA-specific endoribonucleases, in the nucleus to become relatively short miRNA precursors (~70 nt) [125]. These shortened miRNA precursors with stem-loop secondary structure are known as pre-microRNA. These cleaved products are then transported to cytoplasm by exportin-5 [126, 127]. The final step of miRNA maturation is achieved through a second cleavage: the stem-loop pre-microRNAs are cut into small duplexed RNAs of 21-25 bp by Dicer (also a member of the RNase III family) [97, 98]. Each of the Dicer-cleaved products is eventually loaded onto an Argonaute protein to form the core of the effector complex called RNA-induced silencing complex (RISC) [128, 129]. One of the strands of each Dicer-cleaved product is the mature miRNA and is known as the guide strand while the other strand is called the passenger strand. The passenger strands are removed from the duplexes by either Argonaute cleavage or an ambiguous unwinding mechanism before the miRNAs become functional [130, 131]. The mature miRNAs have 5' phosphate end and 3' hydroxyl end [100] (see also Figure 1.3.1). These two ends are inserted into two distinct binding pockets of the Argonaute protein [132, 133]. The mature miRNA located inside the Argonaute protein functions as the guide molecule in targeting the target RNAs. With the assistance of Argonaute protein and other factors of the RISC, translation repression or degradation of the target mRNAs is effected.

1.3.2.1 Transcription of miRNA genes

MicroRNAs are the regulator of many genes including protein-coding genes and non-coding genes [134]. The expression of miRNAs is crucial for modulating numerous cellular processes and even cell differentiation and pathological responses. More than 60% of protein-coding genes are predicted to be regulated by miRNAs targeting the binding sites at the 3' UTR of mammalian mRNAs [135]. Thus, the regulation of miRNA expression is substantially critical. The expression of miRNAs is tissue-specific and developmental-stage-specific [136-138]. These two specific features in the expression of miRNAs may explain why they are able to regulate so many cellular processes in organisms at particular site and time. In an analysis of about 500 mammalian miRNA entities by Saini et al. in 2007, around half of the miRNAs are transcribed from intergenic regions (DNA loci between genes) while the rest are often located at the intronic region of protein-coding genes [139] and occasionally at the intronic region of non-coding genes [140]. In addition, several miRNAs are sometimes grouped in a cluster and transcribed as a single polycistronic transcript [51].

The majority of miRNAs are transcribed by RNA polymerase II (Pol II) [51, 123]. According to the evidences from a correlation study between neighboring miRNAs and host genes examined by microarray, intronic miRNAs are co-expressed (and usually co-transcribed) together with their host gene [141]. With exception of intronic miRNAs, the primary transcripts of other Pol II-transcribed miRNAs are also capped with N^7-methylguanosine at the 5' end and polyadenylated at the 3' end [51, 123]. Therefore, both intronic miRNAs and intergenic miRNAs share the same terminal structures (5' N^7-methylguanosine cap and 3' poly-A tail) in their initial transcripts. Not surprisingly, the promoter regions of intergenic miRNAs are found upstream of the coding regions of the primary transcripts of these intergenic miRNAs [139]. In an extensive analysis of multiple data sets, Saini et al.

confidently derived the general structure of the primary transcripts of mammalian intergenic miRNAs [139]. Their prediction analysis was similar to that applied to coding gene prediction [142]. They analyzed six important features of genes for neighboring regions of intergenic miRNAs, including transcription start sites, transcription factor binding sites, CpG islands, poly-A signals, data of gene identification signature paired-end ditags, and data of expressed sequence tags. Some intergenic miRNAs were found to have multiple predicted transcription start sites while many intergenic miRNAs had their transcription start sites within 2 kb upstream of the pre-miRNAs. Additionally, CpG islands were commonly identified in the surrounding region of transcription start sites. Moreover, the distribution of poly-A signals reached the peak at 2 kb downstream of the pre-miRNAs. Accordingly, they suggested that a canonical primary transcript of human intergenic miRNA had a span of 3-4 kb in length [139]. Figure 1.3.3 illustrates the predicted structure of a typical intergenic miRNA gene.

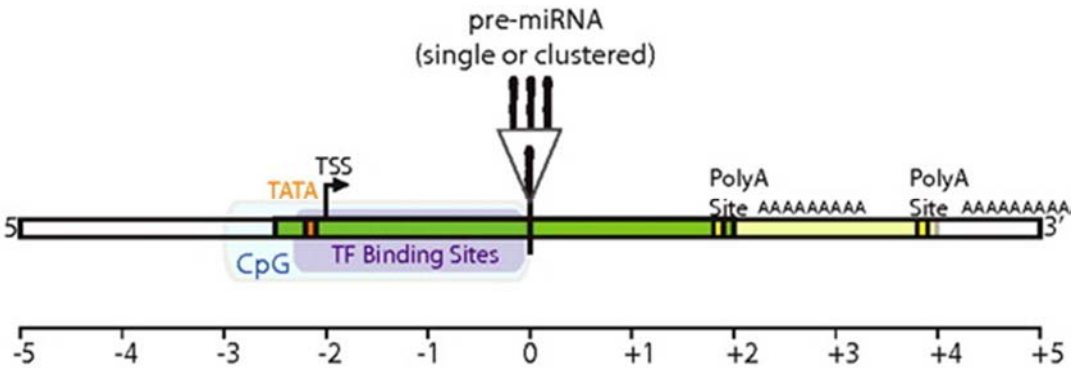


Figure 1.3.3 The predicted structure of a typical miRNA gene as suggested by Saini et al [139]. The pre-miRNA coding region is located at position 0. Transcription start site is about 2 kb upstream of the pre-miRNA coding region while the poly-A site is about 2 kb downstream of the pre-miRNA coding region. The complete primary transcript of a miRNA or miRNA cluster spans about 3 to 4 kb. CpG, CpG island; TATA, TATA box; TF Binding Sites, transcription factor binding sites; TSS, transcription start site. **(Adopted from Proceedings of the National Academy of Sciences 2007, 104(45):17719-17724)**

The majority of miRNA genes structurally contain the elements and features for Pol II transcription. In other words, intronic miRNAs are transcribed from protein-coding genes and non-coding genes whereas intergenic miRNAs are transcribed from its own mRNA gene-like promoters. It is therefore convincing to believe that the expression of miRNAs is subject to the canonical regulatory machinery of protein-coding genes, including epigenetic and genetic regulation.

A large group of transcription factors that originally regulates the transcription of protein-coding genes is able to regulate the expression of miRNAs epigentically. For instance, the regulator gene *MYC* proto-oncogene (*c-Myc*) encodes a transcription factor involved in the control of cell cycle, cell transformation and apoptosis via regulating the expression of other genes [143]. In mammals, c-Myc interacts with 10–15% of genomic loci. The genes located at these loci are predominately involved in biosynthesis, signal transduction, metabolism and cell cycle [143]. Dysregulation of c-Myc can cause cancer. This protein binds to enhancer box sequences (E-boxes) and activates the expression of the miR-17-92 cluster [144]. Elevated expression of miR-17-92 induced by c-Myc can be detected in cancer cells [144]. Moreover, genomic amplification of 13q31-32, which includes the miR-17-92 cluster, is frequently identified in c-Myc-rearranged lymphomas, suggesting that c-Myc and miR-19-92 synergistically induce cell transformation during cancer development [145]. Cancers under this synergistic induction are usually more aggressive than those induced by either factor alone [145]. On the other hand, c-Myc also inhibits the expression of several miRNA genes that serve as tumor suppressors, such as miR-15a, miR-29 and let-7 [146]. The results from chromatin immunoprecipitation assay have shown that the inhibition is mediated by the direct binding of c-Myc to the promoter regions of these miRNAs [146]. We use tumor protein p53 (TP53) as a second example. TP53 activates the expression of miR-34 family, the family of miRNAs participating in cell cycle arrest. The down-regulation

of TP53 reduces the expression of miR-34 and causes the loss of cell cycle arrest [147]. We use RE1 silencing transcription factor (REST) as a third example. REST regulates a brain-specific miRNA – miR-124a. The introduction of miR-124a into HeLa cells stimulates a neuron-like gene expression profile through down-regulating a large number of non-neuronal transcripts [148]. REST inhibits the expression of miR-124a. When the neuron progenitors transform to mature neurons, REST leaves the promoter region of the *mir-124a* locus and promotes the inhibition of non-neuronal transcripts [149].

Moreover, the transcription of miRNAs can also be regulated by methylation in CpG islands and histone modifications [150, 151]. These are the epigenetic mechanisms involved in regulatory mRNA genes. The CpG islands of *mir-148a*, *mir-34b/c* and *mir-9* genes have been found hypermethylated in cancer cells when compared with normal tissues. The expression of these miRNAs can be restored by treating with a DNA demethylating agent (5-aza-2'-deoxycytidine) [152]. In addition, treatment of a breast cancer cell line (SKBr3) with a histone deacetylase inhibitor (LAQ824) alters the expression of 27 out of 67 tested miRNAs, suggesting that histone acetylation may control the expression of many miRNA genes [153].

On the other hand, the expression of miRNAs can also be genetically regulated. For example, the miRNA cluster miR-15a/16-1 produces two miRNA precursors: *hsa-mir-15a* and *hsa-mir-16-1*. The mature miRNAs are *hsa-miR-15a-5p* and *hsa-miR-16-5p* respectively. The genomic region of this miRNA cluster is deleted in about 50% of patients (both Chinese and Caucasian) with chronic lymphocytic leukemia (CLL). This deletion often exists as a unique aberration in leukemic B-cells detectable by fluorescence in situ hybridization [154-158]. Not surprisingly, these two miRNAs are found to be absent or down-regulated in CLL patients with this deletion [159-162]. This could be due to mono-allelic or bi-allelic loss. In addition, down-regulation of these two miRNAs in CLL patients was also linked to a

germline point mutation at the 3' flanking region of the *hsa-mir-16-1* gene-coding region [163]. In brief, no matter the expression of *hsa-miR-15a-5p* and *hsa-miR-16-5p* is down-regulated by deletion of the coding region or is inhibited by a mutation, it seems that the down-regulation of these two miRNAs is found in about half of CLL patients. Similarly, a point mutation at the 3' end region of the *mmu-miR-16-1* gene-coding region also caused down-regulation of *mmu-mir-16-5p* in New Zealand Black mice strain that developed B-cell lymphoproliferative disease naturally [164].

All in all, the above evidences dictate that the transcription of miRNAs is regulated by both epigenetic and genetic mechanisms. Transcription factors participate in the regulation of mRNA transcription via direct binding to particular elements on DNA. The binding of these factors can result in promoting or inhibiting the expression of the corresponding mRNA. These controls can be extended to miRNA expression. It may be that miRNA genes structurally contain the same elements and features for Pol II transcription machinery, which is the same machinery for mRNA transcription. Moreover, the expression of miRNAs can also be regulated by epigenetic mechanisms such as DNA methylation and histone modifications. Furthermore, genetical changes, such as deletion and point-mutation, also regulate the expression of miRNAs. These observations suggest that the regulation of both mRNA and miRNA transcriptions is under a very sophisticated and precise control through the shared transcription machinery in cells. Dysregulation of this machinery causes aberrant expression profiles of mRNAs and miRNAs, and can ultimately result in disease development.

1.3.2.2 MicroRNA processing

Drosha and Dicer are the two central proteins in the process of miRNA maturation. Both of the enzymes are the members of RNase III superfamily [165, 166]. The RNase III superfamily is a class of ribonucleases that recognizes and cleaves double-stranded RNA endogenously, and leaves 5' phosphate and 3' hydroxyl termini at the newly generated ends. Thus, its members are also known as endoribonucleases [165]. Drosha and Dicer participate in two individual cleavage steps of miRNA precursor processing in the nucleus and cytoplasm, respectively. These two endoribonucleases also require accessory proteins to perform their function and generally operate in complex forms. In addition, the maturation process of miRNAs can be regarded as an additional layer of regulating miRNA function. A halt of the maturation process of a particular miRNA can result in loss of function of that miRNA. Indeed, a group of accessory proteins regulates the activities of Drosha and Dicer in both positive and negative ways. Some of these proteins only regulate the maturation of specific miRNA families while others are able to regulate a relative large number of miRNA precursors. These regulations are discussed in the following sections.

Drosha and Drosha processing

The first maturation step after the pri-miRNA transcription is a cleavage process by Drosha in nucleus [125]. Drosha is a protein of about 160 kilodalton (kDa) and belongs to the RNase III superfamily [166]. This protein contains two RNase III catalytic domains, one double-stranded RNA-binding domain and an arginine-/serine-rich domain [166, 167]. By generating mutants of human Drosha, Han et al. have dissected the action mechanism of this enzyme [168]. The two RNase III catalytic domains, domain A and domain B, form an intra-molecular dimer and cleave the 5' strand and 3' strand of the stem of *pri*-miRNA, respectively [168]. This

processing step generates a product RNA of around 70 nt, known as *pre*-miRNA, with a stem-loop structure [168]. In addition, since the positions of the two RNase III catalytic domains are not on the same plane, a 2-nt 3' overhang is usually found on the cleavage product [169, 170]. In reality, Drosha has to form a heterotrimeric complex of about 650 kDa together with two protein molecules of DiGeorge syndrome critical region gene 8 (DGCR8) for functioning (Figure 1.3.4) [169, 170]. The complex is called Microprocessor [168, 171-173]. The two DGCR8-binding sites of Drosha are located at each of the two RNase III catalytic domains. The C-terminal helix of each DGCR8 binds to one of these binding sites and forms the final Microprocessor heterotrimeric complex. In addition, the RNA-binding heme domain (Rhed) on each DGCR8 also anchors with each other at an additional point [169, 170]. These three anchoring points stabilize the structure of Microprocessor. The stem-loop RNA is loaded into the double-stranded RNA-binding pocket formed by two Rhed subunits from the two DGCR8 molecules and one double-stranded RNA-binding domain from Drosha during the cleavage process (Figure 1.3.4) [169, 170]. Furthermore, Arg 914 in the Bump helix located at the central domain of Drosha protrudes towards the basal junction of the stem-loop RNA, which results in fixing the stem-loop RNA in a precise position. Near the Arg 914, a conserved surface with positive charge and other related structures further strengthen the interaction between the basal segments (single-stranded RNA) and Drosha. Remarkably, the distance between the RNase III catalytic domain and the Bump is about 28 Å, which is equivalent to ~11 bp or one helix turn of double-stranded RNA (Figure 1.3.4) [169, 170]. This explains why Microprocessor makes cleavage at ~11 bp away from the basal junction on pri-miRNAs [169, 170]. Figure 1.3.4 shows the structure of Drosha.

Interestingly, Drosha and DGCR8 regulate each other in a negative feedback loop for maintaining the level of Microprocessor in a homeostatic point. In particular, DGCR8 is able to stabilize Drosha through a direct protein-protein interaction even though the mechanism of stabilizing Drosha by DGCR8 and the degradation process of Drosha are not yet known [174]. Conversely, the Drosha/DGCR8 heterotrimeric complex recognizes and cleaves two hairpins on DGCR8 mRNA. The two hairpins form around the 5' UTR (hairpin A) and the coding region close to start codon (hairpin B) of the DGCR8 mRNA. The cleavage, similar to the cleavage on typical pri-miRNAs, generates two hairpin products of ~60 nt and ~76 nt. Most importantly, the cleavages induce the degradation of DGCR8 mRNA in the end [174, 175].

Notably, regulation of Microprocessor's processing activity is governed by a large number of factors including kinases, acetyltransferases and other accessory proteins. The regulation can be either positive or negative. Additionally, some of these regulatory factors can affect the maturation of individual specific miRNAs, and thus these factors do not form a universal pathway to govern the biogenesis of all miRNAs [176].

LIN-28 is one of the most studied proteins and specifically inhibits the expression of let-7 family at both pri-miRNA level and pre-miRNA level [177, 178]. Binding of LIN-28 to the apical loop of *pri*-let-7 can result in blocking the cleavage process by Drosha while binding of LIN-28 to the apical loop of *pre*-let-7 can result in blocking the cleavage process by Dicer and induces uridylation on pre-let-7 [177, 178]. The inhibition of let-7 miRNAs is critical for blocking differentiation and maintains the pluripotency of embryonic stem cells. During cell differentiation, mature let-7 miRNAs accumulate owing to the decreased level of LIN-28 [179]. Interestingly,

LIN-28 itself is regulated by let-7 miRNAs [180]. Thus, LIN-28 and let-7 regulate the levels of each other in an inverse manner.

The p68, also known as DEAD box protein 5 (DDX5), and p72 helicases contribute to the maturation of a subset of miRNAs via interacting with Drosha [172]. The inhibition of miRNA production may be achieved by impairing the pri-miRNA processing. SMAD proteins are a group of signal transducer proteins that transduce extracellular signal from transforming growth factor β (TGFβ) or other similar ligands to nucleus and stimulate gene expression [181]. Intriguingly, SMADs also directly bind to a consensus sequence, RNA-SMAD-binding element (R-SBE), found at the stem region of a group of miRNA primary transcripts induced by signals from TGFβ and bone morphogenetic protein. Microprocessor with p68 is motivated by the direct binding of SMADs to R-SBE, and accelerates the maturation of this subset of miRNAs [182, 183].

General regulatory factors also exist to modulate the expression of a broader range of miRNAs. Histone deacetylase I deacetylates DGCR8 to increase its binding affinity for pri-miRNAs [184]. Glycogen synthase kinase 3β contributes to the nuclear localization of Drosha by phosphorylating Drosha protein at serine 300 and 302 [185]. ERK/MAPKs can stabilize DGCR8 by phosphorylating DGCR8 and increase the total miRNA expression [186].

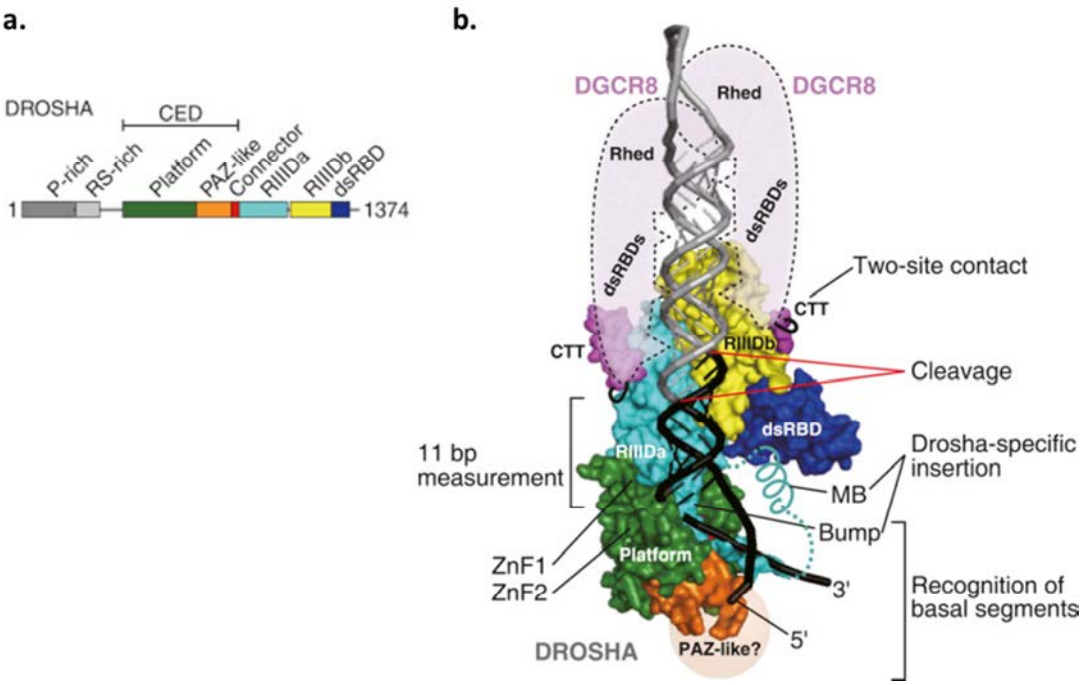


Figure 1.3.4 The structure of human Drosha. (a) Domain organization of the human Drosha protein. The color boxes indicate the position of each functional domain on the primary structure of Drosha. (b) The figure shows the three-dimensional structure of Drosha and the interactions between Drosha, two molecules of DGCR8 and pre-miRNA. CED, Central domain; CTT, C-terminal tail; dsRBD, Double-stranded RNA-binding domain; Rhed, RNA-binding heme domain; RIIID, RNase III catalytic domain. **(Adopted from Cell 2016, 164(1):81-90)**

Dicer and Dicer processing

The second cleavage step in miRNA maturation is carried out by Dicer in cytoplasm [97, 98]. Being also a member of the RNase III endoribonuclease family, Dicer is homologous to the structure of Drosha, but with distinctive amino acid sequence [170]. The molecular weight of human Dicer is about 200 kDa. Similar to Drosha, the functional core of human Dicer contains at its C-terminal two RNase III catalytic domains, which form an intra-molecular dimer (Figure 1.3.5a and b) [187]. Moreover, a P-element induced wimpy testes/Argonaute/Zwille (PAZ) domain is also found in Dicer and functions as the anchoring site for the 2-nt 3' overhang (generated by Drosha cleavage) on pre-miRNAs [188]. The PAZ domain has two separate binding pockets: the phosphate pocket binds the 5' phosphate end of pre-miRNAs, and the 3' pocket binds the 2-nt 3' overhang of pre-miRNAs [188]. The RNase III catalytic domains and the PAZ domain are located at two distinct regions in Dicer. These two regions are separated by a positively charged connecting helix (Figure 1.3.5b). The connecting helix generates a gap of about 65 Å between the two regions. This distance is equivalent to about 25 bp in length. Thus, the connecting helix can also be regarded as a "molecular ruler" for measuring the correct cleavage site on pre-miRNAs [189, 190]. Furthermore, the N-terminal of the Dicer consists of 3 helicase domains, which interact with the apical loop of pre-miRNAs and are thus important for accurate cleavage of pre-miRNAs by Dicer (Figure 1.3.5a and b) [191, 192].

In a way similar to Drosha, Dicer also requires accessory proteins for proper function. Human Dicer interacts with HIV transactivation-responsive RNA-binding protein (TRBP), a double-stranded RNA binding protein [193-195]. This protein has

three double-stranded RNA-binding domains (Figure 1.3.5b) [196, 197]. In *Drosophila*, the protein Loquacious (Loqs) interacts with *Drosophila* Dicer-1. Two variants of Loqs are responsible for interacting with *Drosophila* Dicer-1, the Loqs-PA and Loqs-PB. The latter one is the homolog of human TRBP [198, 199]. Dicer-1 direct binds to either one of the two variants when facilitates the cleavage process [198, 199]. The absence of Loqs results in loss of function in Dicer-1 and accumulation of pre-miRNAs in *Drosophila* S2 cells [198]. Unlike the Loqs in *Drosophila*, the human TRBP is not essential in Dicer cleavage process. The absence of TRBP does not affect the cleavage activity of Dicer in pre-miRNA processing [195]. However, the TRBP protein and another partner protein called protein activator of the interferon-induced protein kinase (PACT) assist in the assembly of RISC [193, 195, 200]. Crystal structure analysis has revealed that the binding interface of TRBP and PACT are highly conserved with each other, suggesting that the two partner proteins bind to Dicer in a similar manner [197]. Moreover, the fidelity of miRNA maturation is disrupted as a result of the aberrant Dicer cleavage of pre-miRNAs when the TRBP is absent during the cleavage process. This eventually alters the sequence of the mature miRNAs (so-called isomiR), and influences the selection of guide strand and passenger strand by altering thermodynamic stability of the two ends of miRNA duplexes (see below for strand selection in thermodynamic stability rule) (Figure 1.3.5c) [197, 201]. On the other hand, many studies have investigated the contribution of TRBP and PACT to Argonuate loading. The results indicate that both TRBP and PACT are dispensable for miRNA loading onto Argonaute protein from Dicer [197, 201, 202]. In RNA-crowded environment, however, TRBP can act as a gatekeeper for preventing Dicer from engaging pre-miRNA-like RNAs, e.g. tRNAs. It is also possible that TRBP

acquires the selectivity for genuine miRNA precursors by using the PAZ domain of Dicer while Dicer reduces the RNA-binding affinity of TRBP for fast turnover when pre-miRNA-like RNA is encountered. These two proteins cooperate with each other to create an efficient machinery of discarding pre-miRNA-like RNAs [203].

Strand selection is a persistent question in miRNA machinery. After the cleavage by Dicer, the double-stranded product is loaded onto Argonaute protein for the subsequent assembly of RISC [130, 131]. In the loading, the 5' end and 3' end of one strand of a cleaved product are loaded into two individual binding pockets at the middle (MID) domain and PAZ domain of Argonaute protein respectively [133]. This strand is the guide strand and is used for guiding the RISC to home in on correct target sites. The other strand is the passenger strand and would be removed from the duplexes either by Argonaute cleavage or an ambiguous unwinding mechanism before the miRNAs starts functioning [130, 131]. The strand selection is not a random process, but its mechanism is still unclear. According to a study conducted by Wilson et al., 13% of miRNA duplexes have different strand selection *in vivo* in the absence of TRBP and PACT [197]. This phenomenon is most likely due to the alteration on the cleavage site of the Dicer cleavage that has generated reversed relative thermodynamic stability on the cleaved product [197, 201]. Indeed, each double-stranded RNA product from Dicer has two 5' recessive ends on the tails. Previous studies has found in *Drosophila* that Dicer-2 tends to choose the strand whose 5' end is thermodynamically less stable [204, 205]. Further investigation has found that this rule is consistent with miRNAs in animals [206]. This is the origin of the thermodynamic stability rule for the strand selection in Argonaute. However, we still do not fully understand the genuine mechanism of

strand selection in the miRNA machinery as the altered cleavage and thermodynamic stability rule only partly explain miRNA strand selection.

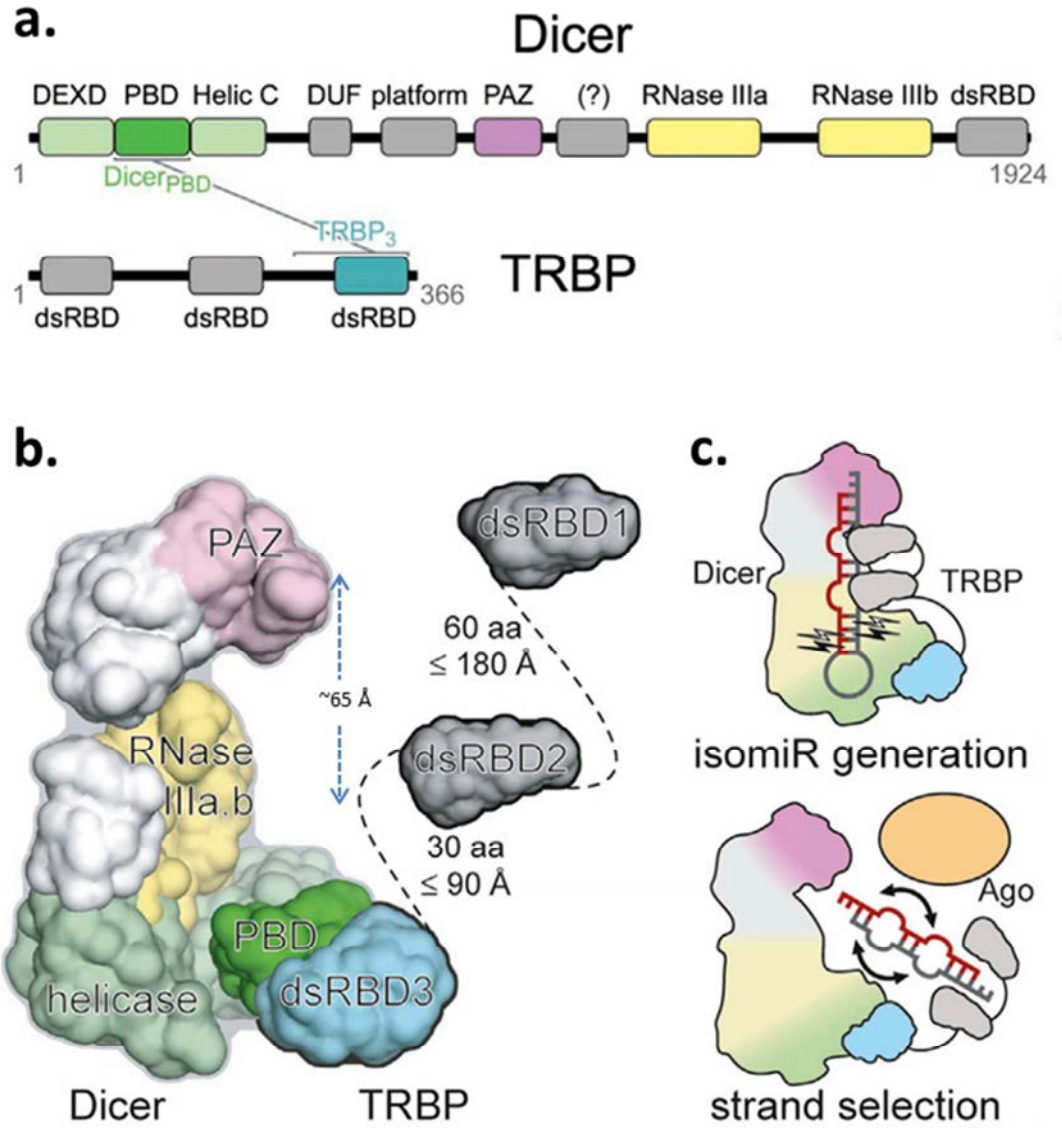


Figure 1.3.5 Cooperation between human Dicer and TRBP in miRNA maturation. (a) Domain organization of human Dicer protein and TRBP. The color boxes have indicated the position of each functional domain on the primary structure. (b) The figure shows the three-dimensional structure of Dicer and the three dsRBDs of TRBP. The dsRBD3 of TRBP interacts with Dicer at PBD. The distance between the binding pocket for the pre-miRNA 2-nt 3' overhang in the PAZ domain and RNase IIIa/RNase IIIb is ~65 Å, which is equivalent to ~25 bp. (c) Models of Dicer partner proteins contributing to the formation of isomiR (top) and the fidelity in strand selection during transfer of a product duplex from Dicer to Ago (bottom). The coloring in (a), (b) and (c) are consistent. Ago, Argonaute; dsRBD, Double-stranded RNA-binding domain; PAZ, P-element induced wimpy testes/Argonaute/Zwille domain; PBD, Protein-binding domain; RNase III, RNase III catalytic domain. **(Adopted from Molecular Cell 2015, 57(3):397-407)**

1.3.2.3 Argonaute proteins and formation of RNA-induced silencing complex for miRNA functioning

Argonaute family proteins are widespread in all domains of life. In eukaryotes, the Argonaute (AGO) proteins are the core proteins of RISCs and mediate several small RNA-dependent gene silencing mechanisms; these small RNAs include miRNA, siRNA and piwi-interacting RNA (piRNA). While more attention is paid to eukaryotes for better understanding of miRNA- or siRNA-mediated gene silencing, little is known about the mechanisms in prokaryotes.

Interestingly, Ago proteins are also present in many bacteria and archaea even though they do not have the kinds of small RNAs found in eukaryotes [207]. Two important studies have revealed that Ago proteins in bacteria may function through other mechanisms. The Ago protein in *Rhodobacter sphaeroides*, a kind of gram-negative bacterium, most likely protects the bacteria against foreign genomic elements, e.g. plasmid DNA. The *Rhodobacter sphaeroides* Argonaute (RsAgo) binds target DNA under the guidance of small RNAs (15-19 nt), which are potentially derived from long single-stranded RNA transcripts, and induce DNA cleavage at the both ends three nt outside the region targeted by the small RNA. As the RsAgo does not contain the residues essential for Ago endonuclease activity, the cleavage may be mediated by a partner endonuclease [208]. Meanwhile, the Ago protein in *Thermus thermophilus* also functions as a protector against foreign genomic elements. However, the *Thermus thermophilus* Argonaute (TtAgo) uses a different machinery to achieve the protection. TtAgo integrates with *small DNAs* as the guide strand to mediate the cleavage of target DNA by its own endonuclease activity [209]. Although the Ago proteins in bacteria use very different mechanisms to achieve the gene silencing, the structures of Argonaute proteins in eukaryotes and prokaryotes are comparable with each

other, as has been investigated in studies of their crystal structures. In fact, the production of pure Argonaute protein from mammalian cells has been a difficult task in the past. Therefore, the earliest investigations of Argonuate protein structures usually used their bacterial and archaeal counterparts for characterizing the structural mechanism of Ago in its process [210-213]. Subsequent studies show that the structure of human AGO2 is nearly the same as that of the prokaryotic Ago although the two analogous proteins still have some differences; for example, the bacterial Ago proteins are able to bind DNA molecules [133, 214, 215].

In humans, the Argonaute protein family consists of eight different Argonaute proteins [216]. These eight members are under 2 subfamilies, the PIWI subfamily and eIF2C/AGO subfamily [216]. Previous studies have found that all the members in the eIF2C/AGO subfamily, i.e. AGO1 to AGO4, are able to incorporate miRNAs in an indiscriminate manner and AGO2 is the only member responsible for inducing target RNA cleavage [217, 218]. Crystal structure investigations have revealed that AGO2 protein has four major domains. From N-terminal to C-terminal of the AGO2 primary structure, the four domains are N-terminal (N) domain, P-element induced wimpy testes/Argonaute/Zwille (PAZ) domain, Middle (MID) domain and P-element induced wimpy testes (PIWI) domain (Figure 1.3.6a and b), in which the N domain and the PAZ domain are connected by a linker (L1) while the PAZ domain and the MID domain are connected by another linker (L2) (Figure 1.3.6a and b) [133, 214].

When a miRNA is loaded onto a human AGO protein, the process involves two sequential steps: (1) the loading of the double-stranded RNA product from Dicer, and (2) the removal of the passenger strand from AGO protein.

During the loading, the 5’ phosphate end of the guide strand is bound into the 5’ phosphate binding pocket located in between the MID domain and PIWI domain (Figure 1.3.6b) [133, 214, 215]. In the meantime, the 3’ hydroxyl end of the guide strand is anchored into the PAZ domain (Figure 1.3.6b) [133, 214]. Multiple residues in the MID domain and the PIWI domain form hydrogen bonds with the 5’ phosphate of the guide strand, suggesting that the 5’ phosphate group of the guide strand is tightly anchored to the 5’ phosphate binding pocket (Figure 1.3.6c and d) [133, 214]. On the other hand, the observation from Schirle et al. has indicated that the 3’ binding pocket only contains some weak electron density [214]. This observation agrees with the detailed interactions described by Elkayam et al., in which only a few hydrogen- bond and stacking interactions are involved in the 3’ end anchoring (Figure 1.3.6d) [133]. In addition, AGO2 anchors the seed region of the guide strand (positions ~2-8) through hydrogen bonds (main) and salt linkages (supplementary) to the phosphate groups on the backbone of the guide strand and Van der Waals interactions with the ribose on the backbone (Figure 1.3.6d) [133, 214]. Obviously, the binding force between AGO2 and guide strand gradually decreases from the 5’ end to the 3’ end of the guide strand as indicated by the findings (Figure 1.3.6d). Furthermore, according to a study of the cleavage mechanism of TtAgo, the 3’ end of the guide strand can most likely be released from the binding pocket of the PAZ domain, and the base-pairing at guide strand positions 17–19 is blocked by the N-terminal domain owing to a conformational change during the cleavage process induced by the guide sequence (Figure 1.3.7) [219]. These evidences indicate that the binding on the 3’ end is weak (relative to the 5’ end). However, the 3’ end of the miRNA is still substantially anchored into the binding pocket in the PAZ domain during the translational repression processes [219, 220]. On the other hand, nucleotide positions 2 to 6 of the guide sequence (usually the seed region of miRNAs) are pre-arranged in an A-form conformation by the

support from the anchoring residues [214]. As the natural conformation of RNA duplex is A-form duplex, this pre-arrangement reduces the entropic cost in forming double-stranded seed region with complementary target RNA during target binding. As a result, this conformation in the seed region of the loaded miRNA largely facilitates the scanning of target RNAs for complementary regions, which supports the "seed match model" of miRNA targeting [221, 222].

In the process of loading miRNA onto AGO, the RISC-loading complexes (RLCs) control the loading of miRNA from Dicer to AGO proteins. Each RISC-loading complex (RLC) consists of an AGO protein, a double-stranded small RNA, a double-stranded RNA-binding protein, a Dicer and other potential proteins assisting in the loading of small RNA onto AGO. In *Drosophila*, each organism has two Dicer genes (*Dcr-1* and *Dcr-2*), and the Dcr-2 protein is required to cooperate with R2D2 for loading of an siRNA onto *Drosophila* **Ago-2** protein [205, 223]. R2D2 is a double-stranded RNA-binding protein and is so called because of its two double-stranded RNA-binding domains (R2) and its association with Dcr-2 (D2) [224]. R2D2 contributes to orienting the Dcr-2/R2D2 heterodimer on the siRNA duplex by binding to the more stable end of the duplex. This action further determines which one of the strands loads onto the *Drosophila* Ago-2 [205, 223]. On the other hand, the components of RLC responsible for loading miRNAs onto *Drosophila* **Ago-1** and human **AGO1 to 4** are not fully understood and are still controversial. TRBP, PACT and Loqs are three different double-stranded RNA-binding proteins. The first two proteins interact with human AGOs while the last protein interacts with *Drosophila* Ago-1. They may resemble the function of R2D2 in *Drosophila* siRNA loading [193, 198, 200, 225]. However, both TRBP and PACT are dispensable for miRNA loading onto Argonaute protein from Dicer [197, 201, 202]. Surprisingly, Dicers are not required for loading small RNAs onto Drosophila Ago-1 [226] and mammalian Argonaute proteins (including human AGOs) [202, 227] *in vitro* and *in vivo*.

Remarkably, the loading of small RNA onto Ago protein for forming pre-RISC (the Ago protein loaded with a small RNA duplex) is an ATP-consuming process [226, 228, 229]. The hydrolysis of ATP is aided by heat shock 70 kDa protein 8 (HSC70) and heat shock protein 90 (HSP90), both of which are chaperon proteins assisting nascent polypeptides to fold into correct architectures through active processes (consuming energy via ATP hydrolysis) [228, 230, 231]. This observation is consistent from fruit fly to humans [228, 231] and even in plant (involving HSP90 only) [230]. Kawamata et al. suggest that the consumption of ATP through HSC70 and HSP90 produces energy and triggers a conformational opening of the double-stranded RNA-binding pocket in AGO proteins, which is like a stretched rubber band. In other words, structural tension is introduced to the opened binding pocket of AGO proteins. Their hypothesis elucidates how a stiff double-stranded RNA can be loaded into the rigid binding pocket in AGO protein [229]. This hypothesis is called the 'rubber band' model (Figure 1.3.8c). Kawamata et al. have also suggested that the release of the above structural tension may be the driven force for the ATP-independent unwinding of the duplex. Thus, a passenger strand is released from AGO [229].

Duplex unwinding and disposal of the passenger strand for degradation are passive processes and do not require energy from ATP hydrolysis [131, 226, 229]. The 2 Ago proteins (Ago1 and Ago2) in *Drosophila* possess cleavage activity and are able to cleave target sequences accordingly when the central sequence of the guide strand is perfectly matched to the target [232]. However, among the four human AGO proteins (AGO1-4), only AGO2 possesses this cleavage activity [217, 218]. Therefore, depending on which Ago protein (with or without cleavage activity) that a nascent miRNA or siRNA (both in duplex form) is incorporated inside and the central pairing status of the small RNA duplex itself, the passenger strand may or may not experience the cleavage process. In other words, passenger strand cleavage occurs when a small RNA duplex contains central

complementary and is incorporated into an Ago protein with cleavage activity; however, the passenger strand is not cleaved when either one of the criteria is not fulfilled. Therefore, the passenger strands of nascent miRNAs and siRNAs are probably subject to two different processes for their removal: with [233] or without [131, 234] cleavage.

Cleavage-dependent ejection of passenger strand

When a small RNA duplex is incorporated into an Ago protein, only the guide strand is securely anchored to the binding pocket (also called binding channel) inside the Ago protein through multiple interactions between the guide strand and amino acid residues of the protein (see above in this section; Figure 1.3.6c and d). Indeed, the binding of the duplex pushes the N domain and L1/L2 linkers (Figure 1.3.6a and b) outward because the duplex is too bulky for the binding channel. As a result, the expanded binding channel produces an inward pressure owing to the increased enthalpy of the Ago protein. During the cleavage-dependent ejection of passenger strand, the functionally wedge-like N domain and α7 (the seventh α-helix from the N-terminal) in L2 push and destabilize two different regions of the duplex respectively. The N domain destabilizes the base pairings between the 5' end of the passenger strand and the 3' end of the guide strand whereas the α7 destabilizes the duplex at positions 2-7 of the guide strand through pushing on the minor groove of this site [235, 236]. Additionally, when Ago protein (with cleavage activity) cleaves the passenger strand at its central region, the nicked passenger strand becomes more unstable and no longer withstands the inward pressure from N domain and α7. Eventually, the two half-passenger strands are squeezed out from the Ago and only the guide strand is retained inside the Ago [235, 236]. The Ago protein loaded with a single strand small RNA is the mature RISC.

Cleavage-independent ejection of passenger strand

Ago proteins without cleavage activity (human AGO1, AGO3 and AGO4) are not able to create nick on the passenger strand and are only able to eject the passenger strand of fully complementary RNA duplexes at 37℃, but not at 25℃ [237]. Similarly, a human AGO2 mutant D597A, which is defective in its cleavage activity, is consistent with the results from human AGO1, AGO3 and AGO4 [237]. These evidences imply that the ejection of passenger strand under the cleavage-independent model most likely utilizes the same mechanism as the cleavage-dependent ejection except the passenger cleavage. However, the passenger strand is not unstable enough to escape from the Ago protein when the strand is intact. For compensation, an elevated temperature (37℃) is required for the ejection. Thus, the cleavage-independent ejection heavily depends on the thermostability of the loaded duplex. Under this circumstance, mismatches in the seed region (positions 2-8 of the guide strand) and/or the 3' middle region (positions 12-16 of the guide strand) largely facilitate the cleavage-independent ejection and these two regions make direct contact with the two wedges (N domain and α7) [131, 234]. This explains why miRNA duplexes are able to form RISCs at both 25℃ and 37℃ in vitro while siRNA duplexes are able to form RISCs at 37℃ only [237]. Moreover, small RNA duplexes containing wobble base pairs (G:U pairs) are also essential for efficient cleavage-independent unwinding [131, 226]. On the other hand, truncated human AGO2 protein that lacks the PAZ domain is still able to load miRNA duplexes, but demonstrates a severe defect in duplex unwinding. This result indicates that PAZ domain is fully necessary for cleavage-independent unwinding [237, 238].

In humans, after a nascent siRNA duplex or miRNA duplex is loaded onto an AGO protein and the passenger strand is ejected from the pre-RISC through either cleavage-dependent or cleavage-independent mechanism, a functional RISC is formed. A functional RISC is an AGO protein loaded with a single-stranded guide RNA in which the AGO protein can be anyone of AGO1-4. Like the removal of the passenger strand, the incorporation of RISC to the target site is also an ATP-independent process. Therefore, Kawamata et al. conceive that the cleavage-independent ejection of the passenger strand is simply the "mirror-image" process of target incorporation (Figure 1.3.8b) [226]. They also suggest that the guide strand inside the binding channel of Ago protein has already pre-organized the channel for target strand incorporation (lowering the activation energy for opening the channel). Therefore, the process of target recognition does not consume ATP to open the binding channel. In contrast, the incorporation of a small RNA duplex into the binding channel is an ATP-dependent process because the double-stranded RNA is too bulky for loading [229].

In the RISC-induced gene silencing processes, including target RNA degradation and translation inhibition, different AGO proteins may be involved in different silencing processes. For example, the *Drosophila* Ago1 is primarily coupled with miRNA pathway while the *Drosophila* Ago2 is participated in siRNA pathway [198, 239, 240]. In contrast, the four human AGO proteins do not show any preferences for siRNAs or miRNAs. However, AGO2 is the only one with cleavage activity among AGO1-4 and is the only one AGO capable of inducing cleavage at the target site [217, 218].

On the other hand, RISCs usually work together with a large number of proteins. These proteins are associated with positive or negative control of miRNA effects. GW182 proteins cooperate with different AGO proteins very often and are essential for RISC-induced gene silencing [241]. This is a family of proteins that contain numerous glycine/tryptophan (GW)

repeats at their N-terminal, a central glutamine-rich region, an RNA recognition motif at their C-terminal and multiple phosphorylation sites. Additionally, their calculated molecular weight is about 182 kDa [242, 243]. GW182 family consists of three paralogues in mammals, i.e. trinucleotide repeat-containing proteins 6A to 6C (TNRC6A, TNRC6B and TNRC6C) [244]. By co-immunoprecipitation targeting a series of truncated *Drosophila* Ago1 and hemagglutinin-tagged GW182 proteins, a study has identified that the PIWI domain of Ago1 interacts with the N-terminal GW repeats of GW182 [245]. Furthermore, gene expression analysis by microarray has discovered that depletion of GW182 or Ago1 in *Drosophila* S2 cells produces similar differential mRNA expression profiles [245]. These findings suggest that both Ago protein and GW182 are linked to the RISC-induced gene silencing process. The process can be impaired when either one of them is missing, which indeed indicates their strong relationship. Moreover, phosphorylation and ubiquitination of GW182 can affect the activity of this protein and further positively or negatively affect the process of RISC-induced gene silencing [246, 247].

In addition to the GW182 family, many other accessory proteins also promote or inhibit RISC-induced gene silencing process. However, little is known about their functions and modes of action. Below are some examples. Accessory proteins of the DExD/H box RNA helicase family are ATP- or NTP-dependent RNA helicases involved in unwinding RNA duplex or remodeling RNA structures, and examples include DExH-box helicase 9 (DHX9), Mov10 RISC complex RNA helicase (MOV10) and DEAD-box helicase 6 (DDX6). Such helicases play an important role in miRNA-RISC assembly, target RNA recognition and dissociation, and even RISC disassembly and turnover [248-251]. RNA-binding proteins (RBPs) act as modulators in enhancing and diminishing the effect of miRNAs via binding to the same RNA molecule targeted by RISC, and include HuR, DND1, fragile X mental retardation protein and RBM4. Ubiquitin E3 ligases of the TRIM-NHL family contribute to

the *activation* of RISC-induced gene silencing process; for instance, TRIM32 and NHL-2 promote miRNA activity in mammals and *Caenorhabditis elegans* respectively. On the contrary, this family of proteins also takes part in *inhibition* of RISC-induced gene silencing process; for example, Mei-P26 is able to decrease the expression level of miRNAs in *Drosophila* [252-254]. However, the full mechanism is still not clear. Nuclear import receptor importin 8 (IMP8) is a miRNA-RISC interacting protein that interacts with all four human AGO proteins and is mandatory for efficient target recognition of a distinct set of mRNAs by human AGO2 [255]. In summary, RISC accessory proteins include a large number of proteins involved in modulating RISC-induced gene silencing. The interplay between miRNA RISCs and their accessory proteins can either counteract or facilitate the effects of miRNAs.

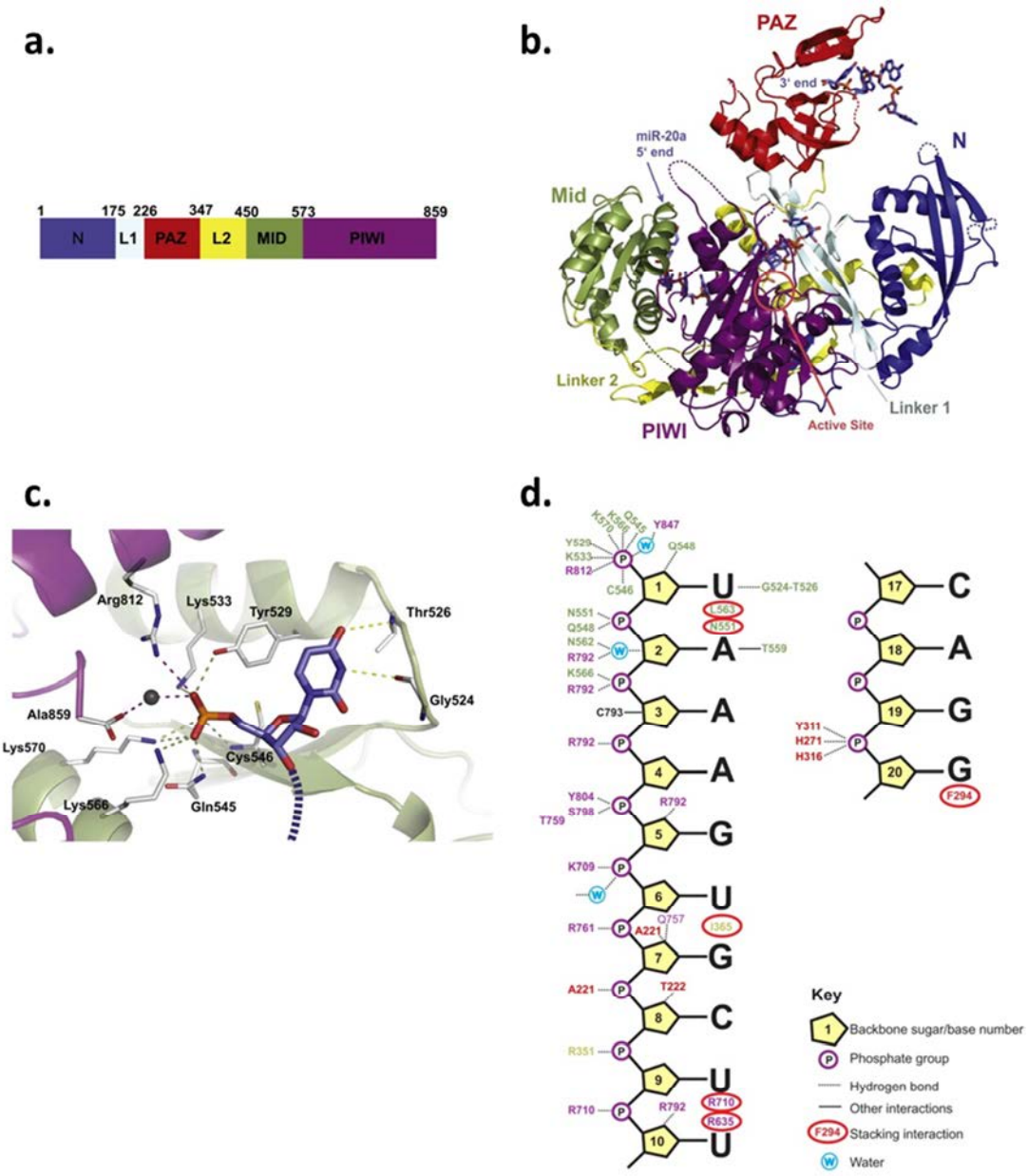


Figure 1.3.6 Structure of human AGO2. (a) Domain organization of human AGO2 protein. The color boxes indicate the position of each functional domain on the primary structure of AGO2. (b) The figure shows the three-dimensional structure of AGO2 and the interaction between AGO2 (shown in ribbon style) and hsa-miR-20a-5p (shown in stick style). (c) The figure shows a three-dimensional close-up capture of the interactions between the 5' end of hsa-miR-20a-5p (the 5' phosphate group and the first nucleotide shown here) and the amino acid residues from the 5' binding pocket of AGO2. Interacting protein side chains and hsa-miR-20a-5p are shown in stick style. (d) A simplified diagram shows the interactions between hsa-miR-20a-5p (positions 11-16 are hidden here for simplicity) and the amino acid residues of the AGO2 protein. All the interactions are indicated according to the key. The same consistent coloring scheme is used throughout. **(Adopted from Cell 2012, 150(1):100-110)**

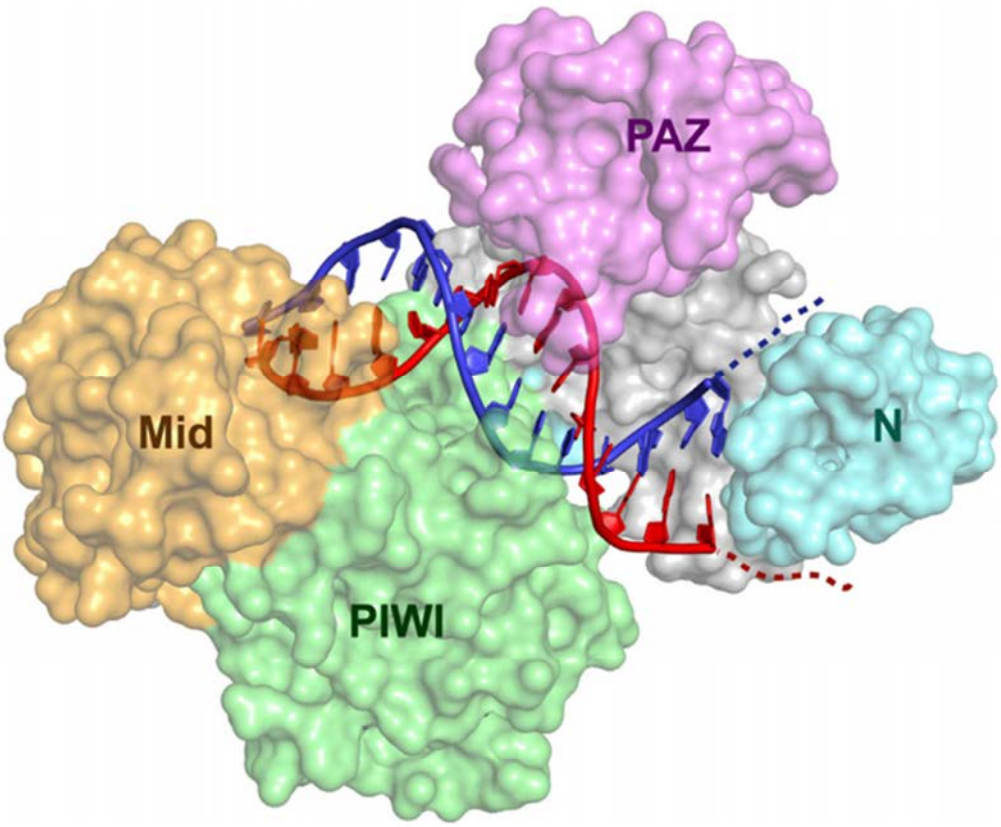


Figure 1.3.7 An anticipated trajectory of guide (DNA) and target (RNA) strands beyond position 16 of guide strand in an Ago protein mutant D478N (lack of cleavage activity). The figure illustrates the spatial structure of the duplexed guide strand (21-nt DNA in red color) and target strand (19-nt RNA in blue color) in a *Thermus thermophilus* Ago mutant. Base-pairing of the guide strand at positions 17–19 (shown in dash line) is blocked by the N-terminal domain of Ago. Thus, a fork forms at this position. **(Adopted from Nature 2009, 461(7265):754-761)**

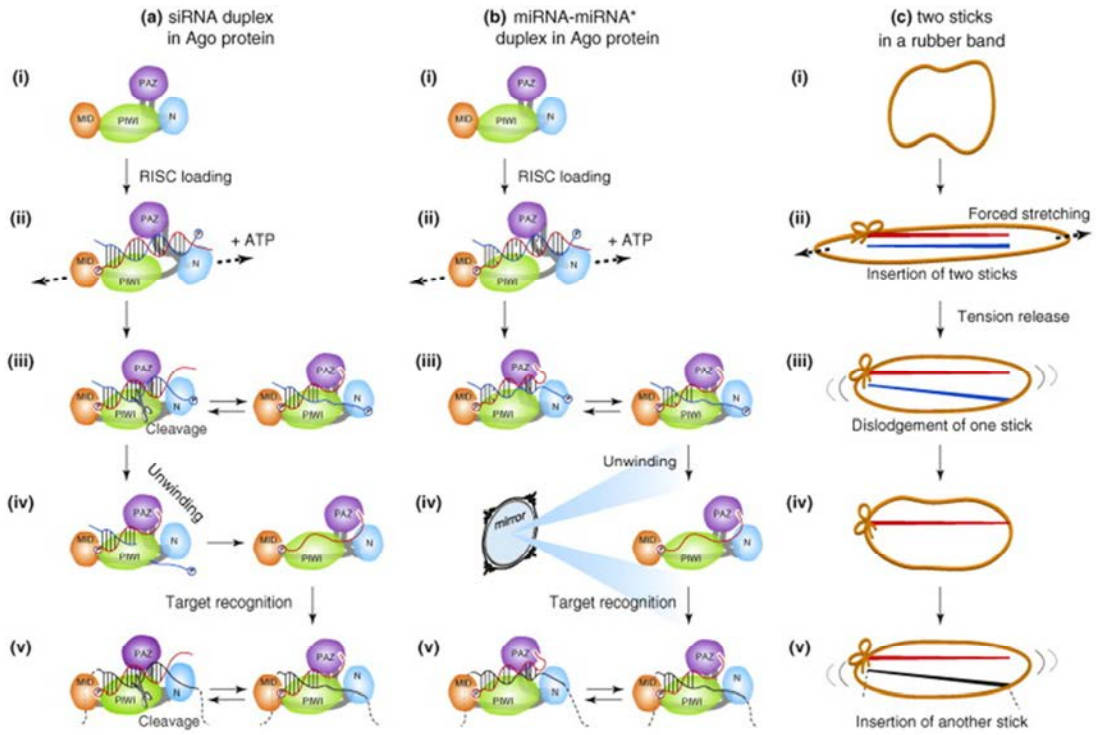


Figure 1.3.8 The "rubber band model" for small RNA duplex loading onto Ago protein. The three columns illustrate the process of loading, unwinding and target binding for different small RNA duplexes: (a) siRNA duplex with perfect complementarity, and (b) miRNA-miRNA* duplex with partial complementarity. Column c shows the rubber band model; note that stretching of the tied rubber band requires energy. **(Adopted from Trends in Biochemical Sciences 2010, 35(7):368-376)**

1.3.2.4 MicroRNA decay

Compared with studies of miRNA biogenesis and functions, studies of miRNA turnover and decay are relatively fewer. As a result, only an incomplete picture of miRNA turnover and degradation is now available. Previous studies have indicated that the stability of miRNA molecules is relatively high. After the depletion of miRNA processing enzymes or the inhibition of RNA polymerase II, the measured half-lives of mature miRNAs in tissues or cell lines can be many hours or up to many days, whereas the half-lives of miRNA precursors is few minutes only [256-258].

Some specific miRNAs have been found to decay at an elevated rate in particular circumstances. For example, miRNAs in retinal neurons degraded much faster than those in other non-neuronal cells in a study of miR-183/96/182 cluster, miR-204 and miR-211 in mouse retina in response to different light levels [258]. In fact, the rapid turnover of miRNAs in neurons was a common property of these cells and was linked to neuronal activity [258]. Another miRNA, miR-29b, was also found to decay rapidly in cycling mammalian cells, but not in mitosis-phase arrested cells [259]. This phenomenon may depend on a sequence motif at the 3' end of miR-29b, which primarily directs miRNAs to nucleus. Interestingly, the nuclear localized miR-29b also showed a lower stability. In contrast, a paralogue miR-29a, which was co-transcribed with miR-29b and did not have such motif, did not undergo rapid degradation [259].

Furthermore, viral infection can also affect the stability of miRNAs. The level of miR-27a, but not two other co-expressed miRNAs, was dramatically reduced upon the infection of mouse cells by murine cytomegalovirus [260]. This suggests that the removal of miR-27a occurred after miRNA maturation. This miRNA exhibits an antiviral activity against murine

cytomegalovirus and its rapid degradation is most likely due to the induction of either virus-encoded or host-encoded factors [260].

Deep sequencing analysis of miRNA profiles in various cells has revealed that many of miRNAs have an untemplated addition of uracil or adenosine at their 3’ end post-transcriptionally. MUT68 is a nucleotidyltransferase responsible for uridylating 3’ ends of small RNAs without requiring a template [261]. The enzyme exhibits *in vivo* uridylation activity in *Chlamydomonas reinhardtii* (an alga). The uridylated miRNAs can be degraded by RRP6 exosome *in vitro*. Impaired expression of either *MUT68* or *RRP6* also causes accumulation of miRNAs in *Chlamydomonas reinhardtii*. These results suggest that the two enzymes cooperate in promoting the decay of mature miRNAs [261]. Likewise, uridylation of miR-26a at its 3’ end by Zcchc11 nucleotidyltransferase results in functional repression of this miRNA by promoting degradation of this miRNA [262].

The 3’ end adenylation is another common modification of miRNA molecules. Addition of a single adenosine at the 3’ end of miR-122 by GLD-2 prevents the miRNA from degradation in human and mouse liver cells [263]. Similarly, ptc-MIR397 and ptc-MIR1447 families of *Populus trichocarpa* (a plant) miRNAs are also protected from exonucleolytic degradation by addition of adenosines at their 3’ end [264]. Moreover, the 3’ end of miRNAs with 2’-O-methylation is also able to prevent miRNA degradation and uridylation in plants [265].

In reality, miRNAs can be degraded from the 5’ ends by 5’-to-3’ exonucleases or from the 3’ ends by 3’-to-5’ exonucleases. The enzyme 5'-3' exoribonuclease 2 (XRN2) possesses 5’-to-3’ exonuclease activity, and has been found to catalyze the decay of mature miRNAs in *Caenorhabditis elegans* [266]. Likewise, small RNA degrading nuclease (SDN) family possesses 3’-to-5’ exonuclease activity, and its members SDN1, SDN2 and SDN3 have been

found to play a role in mature miRNA degradation in *Arabidopsis thaliana* [267]. In fact, the degradation of miRNAs requires the release of miRNAs from the RISCs. Chatterjee et al. have revealed in *Caenorhabditis elegans* that XRN2 not only promotes the degradation of miRNAs, but also supports the release of the mature miRNAs from RISCs when miRNA target abundance is low [266]. These properties of XRN2 facilitate the turnover of Ago protein in RISC-induced gene silencing process and prevent the accumulation of miRNAs in cells [266].

1.3.3 Nomenclature of microRNAs and microRNA precursors

The miRNA nomenclature scheme is based on suggestions given by Griffiths-Jones et al. in two publications [118, 119]. The primary rules of the nomenclature scheme are summarized as follows. (i) All the names of miRNAs start with a three- or four-letter prefix to indicate the species (e.g. hsa-miR-101-3p). (ii) This is followed by "miR" for mature sequences or "mir" for precursor stem-loops. (iii) Next, numerical identifiers are assigned to miRNAs sequentially. (iv) Mature miRNAs with only one or two positional differences are paralogous sequences. A lettered suffix immediately following the numerical identifier distinguishes these sequences (e.g. hsa-miR-15a-5p and hsa-miR-15b-5p). (v) A numbered suffix is given to miRNA precursors to differentiate stem-loop loci that encode the same mature miRNA (e.g. hsa-mir-16-1 and hsa-mir-16-2). (vi) The two different mature miRNAs originating from opposite arms of the same stem-loop precursor carry a suffix to indicate their relationship with the precursor: "5p" for the miRNA close to 5' end of the precursor, and "3p" for the miRNA close to 3' end of the precursor (e.g. hsa-miR-15a-5p and hsa-miR-15a-3p).

1.3.4 MicroRNA related diseases

The first miRNA-related disease, chronic lymphocytic leukemia (CLL), was discovered more than a decade ago by Calin's group [159]. CLL is a hematological malignancy characterized by accumulation of phenotypically mature, but functionally impaired B-lymphocytes. These cells are activated B-lymphocytes that are defective in apoptosis and have low proliferation rate [268].

1.3.4.1 The discovery of two cancer associated miRNAs

In early studies of CLL, genomic aberration is one of the major concerns. In practice, assessing the genomic aberration status is useful for disease classification. The genomic aberration del13q14 is the most frequent aberration (~50%) among all the aberrations identified in CLL. Moreover, the frequency of this deletion is similar in both Caucasian and Chinese CLL patients [154-158]. Furthermore, this deletion can also be found in many other types of cancers including solid cancers (e.g. prostate cancer, and head and neck squamous cell carcinoma) and other hematological malignancies (e.g. myeloma and lymphoma) [269]. These suggest that del13q14 is critical in cancer development and the genes located within this critical region may have an important function in tumor suppression. Several groups attempted to determine the minimal deleted region and characterize the genes located in this region [270-275]. One of the suspected components located within this region is a miRNA cluster. It encodes two miRNAs: *hsa-miR-15a-5p* and *hsa-miR-16-5p* [159]. Subsequent studies have confirmed the association of these two miRNAs with cancer development.

1.3.4.2 Despairing in miRNA target identification

BCL2 is an anti-apoptotic protein, which is over-expressed in B-lymphocytes from CLL patients. This gene is one of the suspected targets for *hsa-miR-15a-5p* and *hsa-miR-16-5p*. First described by Cimmino et al. in 2005 [160], *BCL2* is negatively regulated by *hsa-miR-15a-5p* and *hsa-miR-16-5p* in CLL patients' B-lymphocytes. Thereafter, this finding has also been confirmed by other groups [162, 276]. However, some other studies suggested that *BCL2* was not the target of *hsa-miR-15a-5p* and *hsa-miR-16-5p*, or was only partially suppressed by these two microRNAs [161, 277, 278]. Several other genes regulated by these miRNAs in cell models or patients were also demonstrated by other groups although it was certain whether the inhibition was direct or indirect effect [162, 278, 279]. In other words, an ultimate answer for the direct targets of *hsa-miR-15a-5p* and *hsa-miR-16-5p* and the related mechanisms in leukemogenesis is still unavailable.

There are several possible explanations for these discrepancies. First, differences between individuals may cause failure to replicate the findings in different cohorts of patients. Biological variation together with small sample size can be one of the reasons for the contradictory results. Another notable possibility is the presence of some unknown subtypes of patients within the group of patients with these two miRNAs. Second, the accuracy and coverage of prediction methods vary. Seed region perfect match, conservation between species, free energy of the miRNA-mRNA duplex, etc., are the frequent criteria considered in various miRNA target prediction algorithms. Each of these algorithms predicts several hundreds of targets for each miRNA. Because different algorithms have different combinations of criteria with different weights for each criterion, each prediction algorithm has their own set of predicted targets for each miRNA. However, many of these predicted targets are not the *bona fide* targets because these criteria cannot

precisely reflect the real situation in biology. For example, seed-region pairing is the criterion commonly considered in prediction algorithms. However, it was found that only 55-75% of miRNA targets are seed-matched targets in recent high-throughput analyses of miRNA-mRNA duplexes associated with RISC [134, 280]. Third, prediction algorithms always pay a lot of attention to the sequence of miRNAs and its corresponding target sites; however, issues related to genes expressed specifically only in certain tissues and the sophisticated regulation of miRNA machinery are neglected. In addition, the regulatory mechanisms may vary with the conditions under study. For example, lin-4 miRNA only inhibits the expression of lin-14 during the early stages of larval development in *Caenorhabditis elegans* [281]. These may be the reasons for the difficulties encountered during the discovery of the miRNA targets.

Indeed, it is difficult and challenging to identify and define the genes and the pathways that are regulated by miRNAs. However, this is the key to revealing the genuine function of miRNAs. Many different strategies have been proposed for miRNA target prediction, such as TargetScan [282] and PicTar [283]. These prediction methods are entirely based on computational algorithms only, and not supported by experimental data. Thus, if only computational algorithms are used to predict the target genes of *hsa-miR-15a-5p* and *hsa-miR-16-5p*, the genuine biological processes involved may not be revealed.

1.4 Computational prediction of microRNA targets

Identification of miRNA targets in animal cells is one of the biggest challenges in miRNA research because miRNAs do *not* require perfect match to their targets to carry out their functions. Therefore, target identification is an extremely difficult task because we cannot

identify the target sites by simply matching the miRNA sequences to the transcriptome or genome of the corresponding species. Accordingly, target identification has usually been achieved by different prediction algorithms until quite recently. Then, the predicted targets are examined by *in vivo* experiments individually in the model(s) of interest. However, the work for miRNA target identification is still very difficult because the conventional *experimental* strategy (e.g. luciferase reporter assay) for demonstrating the exact binding relationship between miRNAs and target RNAs employs low-throughput and labor-intensive experimental methods, and each miRNA may have several hundreds of predicted targets.

Target prediction is only based on several criteria at all times. These criteria are usually determined by observations or common knowledge. This prediction strategy has limited accuracy and sensitivity, i.e. some of the predicted targets are not *bona fide* targets and some genuine targets are missed. For example, seed-region pairing is the criterion commonly considered in prediction algorithms. However, recent high-throughput analyses of RISC-associated miRNA-mRNA duplexes show that only 55-75% of miRNA targets are seed-matched targets [134, 280]. In addition, expression of miRNA and target RNAs is usually tissue-specific. This phenomenon implies that the targets confirmed in one tissue type may not be re-confirmed in another tissue type. Therefore, a correct tissue type is necessary when confirmation is attempted.

Indeed, it is difficult and challenging to predict the targets of miRNAs. However, this is one of the most important strategies for revealing the genuine function of miRNAs. Many different prediction strategies have been proposed for target prediction for miRNAs, such as TargetScan and PicTar. Notably, these prediction methods are based only on computational algorithms and not supported by experimental data. If only computational

algorithms are used to predict the target genes of miRNAs, the genuine biological processes involved are most likely not revealed. Moreover, all computational prediction methods are not capable of identifying any synergetic effects or additive effects of miRNAs. Several important prediction algorithms are discussed in the following sections.

1.4.1 TargetScan

TargetScan is the first developed prediction algorithm for predicting mammalian miRNA targets. It was developed by Lewis et al in 2003 [282]. Its early release adopted two basic criteria. First, the algorithm seeks for seed matches at 3' UTR of mRNA of the query organism such as humans. The definition of seed match is the full complementarity of miRNA positions 2 to 8 (the 7-mer seed region of miRNA) to anywhere in the 3' UTR of candidate target mRNAs. The base-pairing does not consider wobble base-pairing. Second, the algorithm only accepts those predicted target sites that have orthologs in reference organisms (humans, mouse, rat and *Fugu rubripes* in its early release) [282]. Afterwards, TargetScan undertook a series of refinements for adapting the increased number of new miRNAs from different species and within species. These refinements increase the power of TargetScan in target identification, minimize false discovery, and provide a more user-friendly interface (TargetScanS, a simplified version of TargetScan since release 3.0) [135, 222, 284-287] The major changes are listed below. (i) Change the criteria of "seed match" from 7-mer only to accept 8-mer seed match and 6-mer match plus an adenosine at target site opposite to miRNA position 1 as well [222]. (ii) Add a context score to the prediction as a reference for reliability. This score considers 14 features (in release 7.1) in the surrounding region of candidate target sites, e.g. supplementary base-pairing, local AU content, and the length of 3' UTR [284, 287]. (iii) Overhaul the searching preference for conservation of sequence motifs and add the probability of preferentially conserved

targeting (P_{CT}) score to each candidate target site. This score reflects the estimated Bayesian probability that a target site is conserved across species due to selective maintenance of miRNA targeting [135]. With all these changes, the program is maintained as one of the most useful algorithm for target prediction to date.

1.4.2 PicTar

Probabilistic identification of combinations of Target sites (PicTar) is the second launched algorithm for predicting miRNA targets. This algorithm is built on hidden Markov model with the approach of maximum likelihood [283]. Like TargetScan, PicTar also considers the seed matches and target site conservation across different species in its algorithm. However, PicTar allows imperfect matching at the seed region of the miRNA and at the same time imposes a stringent free energy cutoff for imperfect seed pairings. In addition, PicTar has the added advantage that its algorithm takes into account the coordinate actions of multiple miRNAs: synergistic effect of miRNAs is assumed if predicted target sites are optimally spaced with each other, but antagonistic effect of miRNAs is assumed if predicted target sites overlap. PicTar's scoring reflects these considerations, and is defined as likelihood ratio of observed/background [283]. The last update of PicTar is in March 2007.

1.4.3 miRanda

The miRanda algorithm is also commonly used for predicting target sites of miRNAs. This algorithm also uses criteria similar to those of TargetScan, and was initially built for predicting miRNA target sites in fly and humans [283, 288]. It first aligns miRNAs (the whole sequence of each miRNA) to the 3' UTR of target mRNAs to recognize highly complementary sequences. Seed match is not compulsory in the alignment, but seed match has greater weight than pairing in other parts of each miRNA. Moreover, wobble

base pairing (GU base pairing) is also allowed. Then, the extracted candidates of target sites are filtered by heteroduplex free energy (evaluating the stability of the duplex). Lastly, conservation between species is also considered. All predicted targets are filtered based on evolutionary distance [283, 288]. This is the first algorithm considering the wobble base pairing in the prediction [283, 288].

1.4.4 PITA

Although the stability between the miRNA and suspected target site is worth taking into consideration in the target site prediction, the accessibility of each suspected target is also important for prediction algorithm to consider. The accessibility of suspected target is reflected from the energy required for opening or unfolding the surrounding secondary structure of the target site. Lower energy input makes it easier to access the target site and vice versa. On top of the miRNA-target site stability, the Probability of Interaction by Target Accessibility (PITA) prediction also considers the unfolding free energy cost in its algorithm. Thus, the PITA algorithm takes into account two different stages of energy requirement. Similar to TargetScan, PicTar and miRanda, the PITA algorithm also considers the seed match. However, it allows users to specify the number of mismatches in the seed region. This additional feature enables flexibility in seed pairing and thus makes the algorithm potentially suitable for exploring new types of miRNA targets. The algorithm disregards cross-species conservation. In addition, PITA prediction is not restricted to 3' UTR of the target mRNAs. It predicts target sites outside the 3' UTR as well [289].

1.4.5 Rna22

Similar to PITA, rna22 also does not consider cross-species conservation and allows users to specify the seed match criteria. Therefore, the algorithm enables the identification of non-

conserved miRNA binding sites and allows the discovery of binding sites other than the type with seed match only [290]. Likewise, the rna22 algorithm permits the searching of miRNA target sites outside the 3' UTR of genes [290]. Nevertheless, it has a fundamental difference in its target discovery procedure. It is a pattern-based algorithm for miRNA target identification. The concept is based on the use of salient sequence features, which is identified via comparing all known miRNAs, to identify the miRNA targets in the transcriptome [290]. Briefly, the algorithm first obtains the sequence information from Rfam database (an RNA database that collects non-coding RNAs including miRNAs). All the identical and nearly identical entries are first removed before the preparation of salient sequence features. Then, patterns of salient sequence features are recognized via intra- and inter-species comparisons. Only statistically significant sequence features are chosen for the next step in target identification. Subsequently, the selected patterns of sequences are converted to reverse complementary sequences and used to identify "target islands" from transcriptome via sequence alignment. Finally, the "target islands" are paired up with miRNAs. Stability of the resulted miRNA-target site heteroduplexes are considered in the algorithm as well [290].

1.5 Experimental identification of microRNA targets

The effect of miRNAs is only demonstrated when they are used as the guided strands in RISC for guiding the complex to bind to the specified target sites. This binding may then recruit some other accessory factors, and elicit other subsequent changes and responses. Therefore, correct identification of genuine miRNA targets is critically important for studying the functions and effects of miRNAs. For miRNA target identification,

computational algorithms have contributed a lot because they are able to massively screen and predict miRNA target sites from the whole transcriptome. However, accuracy and sensitivity are the major drawbacks of computational prediction [134, 280] because these *in silico* prediction algorithms are based on criteria derived from previous observations or common knowledge only. A predicted miRNA target may not have any experimental support. Therefore, computational prediction algorithms can only be used as the tools for shortlisting the candidates of miRNA target for subsequent confirmation and characterization by appropriate experimental methods in a relevant model.

Remarkably, different algorithms have different combinations of criteria with different weights for each criterion. As a result, each prediction algorithm produces its own set of predicted targets for each miRNA. This causes a huge problem in selecting targets of interesting miRNA(s) for validation. If we consider all the predictions from different algorithms, the number of the predicted targets will be very large. In contrast, if only the common targets predicted by all prediction algorithms are considered, some targets will be missed. Accordingly, sometimes we also combine gene ontology analysis in target selection. Gene ontology analysis helps to categorize the gene functions of the predicted targets and also provides hints in selection of candidate targets. However, gene ontology is not a perfect analysis because many genes are not fully studied. In addition, prediction algorithms always focus their attention to the sequence of miRNAs and its corresponding target sites while neglecting the real situation in biology. Thus, a predicted target confirmed in a particular model may not be confirmed in another model because the action of miRNAs is a complicated process and is regulated in many steps by many regulatory machineries (see Section 1.3.2). Therefore, validation of individual miRNA target by conventional experimental methods (e.g. luciferase reporter assay) is difficult because

there are so many uncertainties over inaccurate predictions. That is why our understanding of the effects of miRNAs is extremely limited.

1.5.1 The beginning of experiment-based identification of microRNA targets

Experimental approaches (e.g. microarray analysis and qPCR) that only demonstrate the *association* between miRNAs and mRNAs can be used to discover the relationships between miRNAs and the target genes. The principle of these approaches is based on two assumptions. First, the expression profile of cells in a steady state is quite stable and can be used as a reference. Second, miRNAs usually function as negative regulators for inhibiting the expression of target genes. With these assumptions, some research groups have applied microarray analysis to screen for miRNA targets on a massive scale. They first over-express (knock-in) or inhibit (knock-down) the expression of a miRNA of interest in a selected model. They then compare the expression profiles between the miRNA-manipulated cells and the reference (cells not experienced any manipulations) by microarray analysis to pinpoint the target genes of that miRNA [148, 291-293]. Other than the microarray experiment, 2-D gel electrophoresis has also been used for identifying miRNA targets in a similar fashion [291]. However, these experiments cannot discriminate direct and indirect effects of miRNAs (see Section 1.3 for miRNA direct and indirect effects).

In contrast, experimental approaches that demonstrate the *binding* between miRNAs and the 3'UTR of mRNAs can be used to confirm whether it is a direct effect or not, and are exemplified by luciferase reporter assay and Argonaute high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (Ago HITS-CLIP) (see below). The luciferase reporter assay is a conventional approach for validating miRNA targets. Researchers usually use luciferase reporter assay to confirm whether their finding is a

direct or an indirect effect. The basic principle of this method involves the transfection of a plasmid into a selected model; the plasmid carries a luciferase gene as the reporter gene and the sequence of the putative miRNA target site at the 3' UTR of the luciferase gene. Under the additional transfection of corresponding mature miRNA (duplex form) or other relevant construct (e.g. pre-miRNA or plasmid encoding mature miRNA) into the above transfected model, the expression of luciferase will be reduced when the target site of interest is truly for that particular miRNA. The expression of luciferase can be quantified by measuring the bioluminescence (yellow-green light) produced from its substrate (luciferin). Unfortunately, a miRNA may have several hundreds or even thousands of predicted direct targets [93, 294]. Therefore, validation of miRNA targets by luciferase reporter assay is a time-consuming and labor-intensive approach.

In the second approach, several groups attempt to invent some new methods to identify miRNA targetome experimentally. The suffix "-ome" refers to a concept of totality in biology. The word "miRNA targetome" describes all the targets of miRNAs in a specific cell type, tissue or even organism. Different cell types or a specific cell type under different conditions may have different miRNA targetomes. Therefore, studying the targetome is much more advanced than studying miRNA targets one by one. Moreover, large-scale studies are the trend in future investigations, just like studying genome, transcriptome, etc. Several important methods have been developed for studying miRNA targetome or RNA-RNA interactome. These methods are summarized in

Table 1.5.1 and described in the following sections.

Table 1.5.1 A summary of methods for identifying genome-wide miRNA targets.

Method [a]	Transcriptome-wide detection of miRNA targets	Detecting exact correspondence between miRNAs & targets	UV Wavelength for crosslinking	Special chemical used (Chemical name)
HITS-CLIP	Yes	No	254nm	No
PAR-CLIP	Yes	No	365nm	Yes (4-thiouridine)
iCLIP	Yes	No	254nm	No
CLASH	Yes	Yes [b]	254nm	No
MARIO	No [c]	Yes [d]	254nm	No
PARIS	Uncertain	Uncertain [e]	365nm [f]	Yes (4'-aminomethyltrioxsalen)

[a] HITS-CLIP High-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation
PAR-CLIP Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
iCLIP Individual-nucleotide resolution crosslinking and immunoprecipitation
CLASH Crosslinking, ligation, and sequencing of hybrids
MARIO Mapping RNA interactome in vivo
PARIS Psoralen analysis of RNA interactions and structures

[b] For CLASH, the representativeness of exact correspondence between miRNAs and their target is doubtful.

[c] For MARIO, the method discards all the non-chimeric sequences when the library is prepared.

[d] For MARIO, the representativeness of exact correspondence between miRNAs and their target is doubtful.

[e] For PARIS, the method intends to discover the RNA structurome (or structome). Theoretically, it can be used to detect most of the RNA-RNA interactions.

[f] For PARIS, light of 254 nm is used to reverse the crosslinking.

1.5.2 HITS-CLIP

Next-generation sequencing (NGS) is a sequencing method able to sequence an extremely large number of pooled sequences in parallel. It has significantly improved the current sequencing method and enabled sequencing on a genomic scale in a much shorter time and at a much lower cost. By taking advantage of NGS, Darnell's group successfully developed the method **Argonaute Hi**gh-**T**hroughput **S**equencing of RNAs Isolated by **C**ross**l**inking **I**mmuno**p**recipitation (Ago HITS-CLIP) in 2009 – the first method of identifying miRNA targetome experimentally [134].

This method combines ultraviolet-crosslinking techniques, protein immunoprecipitation, and NGS [134]. Briefly, the method covalently crosslinks *in vivo* interacting RNA and RNA-binding protein (close enough for crosslinking) by 254 nm ultraviolet (UV) light stimulation. This crosslinking step also crosslinks the effecting Argonaute complex (the interacting miRNA-target RNA-Argonaute protein) covalently. The experiment starts with cells or tissues that have been disrupted into small pieces. After the UV light crosslinking step, the cells or tissues are lysed in an appropriate buffer that is compatible with immunoprecipitation in the next step (Figure 1.5.1). Immunoprecipitation is carried out by adding an appropriate amount of magnetic beads coated with anti-Ago antibody. These magnetic beads serve to capture the Argonaute protein from the protein lysate. Therefore, the Argonaute protein that is crosslinked with miRNA and target RNA is also captured by the beads. Subsequently, RNase A is used to trim away the target RNA segments protruding from the Argonaute protein complexes (on-bead process). Ligation of radiolabeled 3' adaptor to the 3' end of the target RNAs is performed after the 3' end of the target RNAs have been repaired (dephosphorylated) by calf intestinal alkaline phosphatase (on-bead process). The 5' end of the target RNAs is then phosphorylated by T4 polynucleotide kinase

(on-bead process). Next, the Argonaute protein complexes are eluted from the beads and the eluate is then separated by protein gel electrophoresis. The gel band (a smear) with the expected size and containing the Argonaute protein complexes is cut out for recovering the targeted Argonaute protein complexes. Then, the Argonaute protein in the complex is digested by proteinase K. Next, the 5’ adaptor is ligated to the 5’ end of the target RNA. Finally, RT and PCR amplification are performed to finish the targetome library preparation. The library is ultimately sequenced by NGS [134]. Figure 1.5.1 shows the procedure of HITS-CLIP.

In fact, this technique has provided a great opportunity to discover miRNA target sites directly, which circumvents failure caused by predictions. However, this method is unable to detect the exact correspondence between miRNAs and their targets.

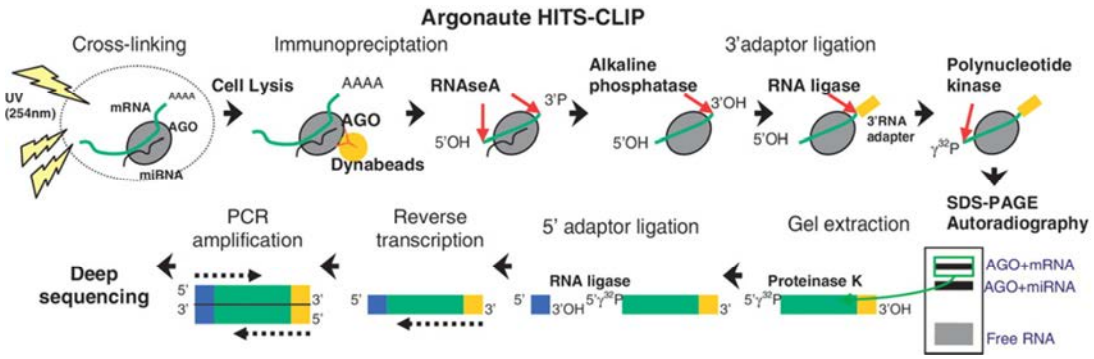


Figure 1.5.1 The experimental scheme of Ago HITS-CLIP. See Section 1.5.2 for details. **(Adopted from Nucleic Acids Res 2011; 39 (16): 6845-6853)**

1.5.3 PAR-CLIP

Photo**a**ctivatable **R**ibonucleoside-enhanced **C**ross**l**inking and **I**mmuno**p**recipitation (PAR-CLIP) is a modified version of HITS-CLIP. The general procedure of PAR-CLIP is very similar to that of HITS-CLIP, though with some minor revisions. The method also includes UV-crosslinking, protein immunoprecipitation, and NGS [295, 296]. The major difference between PAR-CLIP and HITS-CLIP is that the former approach uses 4-thiouridine (4SU) (Figure 1.5.2a), a photoactivatable ribonucleoside analogous to uridine, to enhance the crosslinking efficiency in covalent crosslinks between the interacting RNA and RNA-binding protein under the induction of 365 nm UV light [295, 296]. As the result, the yield of RNA-protein crosslinked complexes is much higher in PAR-CLIP than in HITS-CLIP and the prepared library is much more representative in PAR-CLIP than in HITS-CLIP accordingly [295]. This uridine analog can be taken up by cells and further utilized by RNA polymerases to produce RNAs. Therefore, the uridine residues in the produced RNAs can be replaced by 4SU incidentally.

In this method, an appropriate concentration of 4SU is added to the cells during culturing. After the addition of 4SU, the cells are allowed to grow in culture for enough time so that they can uptake and utilize the 4SU for RNA synthesis (Figure 1.5.2b). The final concentration of 4SU in the cultured cells and the incubation time have to be optimized for each particular cell type. Normally, the optimal 4SU concentration and incubation time are 100 μM and 14 hours respectively [295, 296]. Figure 1.5.2b shows the complete scheme of PAR-CLIP experimental protocol.

Other than the advantage of enhanced crosslinking efficiency, the method also reveals the exact binding site of the RNA-binding proteins. Through data analysis, the team inventing PAR-CLIP has discovered that only the crosslinking sites in isolated RNAs are apparently

converted from T to C during cDNA synthesis with the isolated RNA templates. Therefore, PAR-CLIP can be used to locate the crosslinking sites precisely via mutational analysis. The crosslinking sites actually represent the binding sites of RNA-binding proteins [295, 296].

Like HITS-CLIP, this technique can be used to discover miRNA target sites directly in a much precise manner while preventing failure caused by computational predictions. However, this method is not capable of detecting the exact correspondence between miRNAs and their targets either.

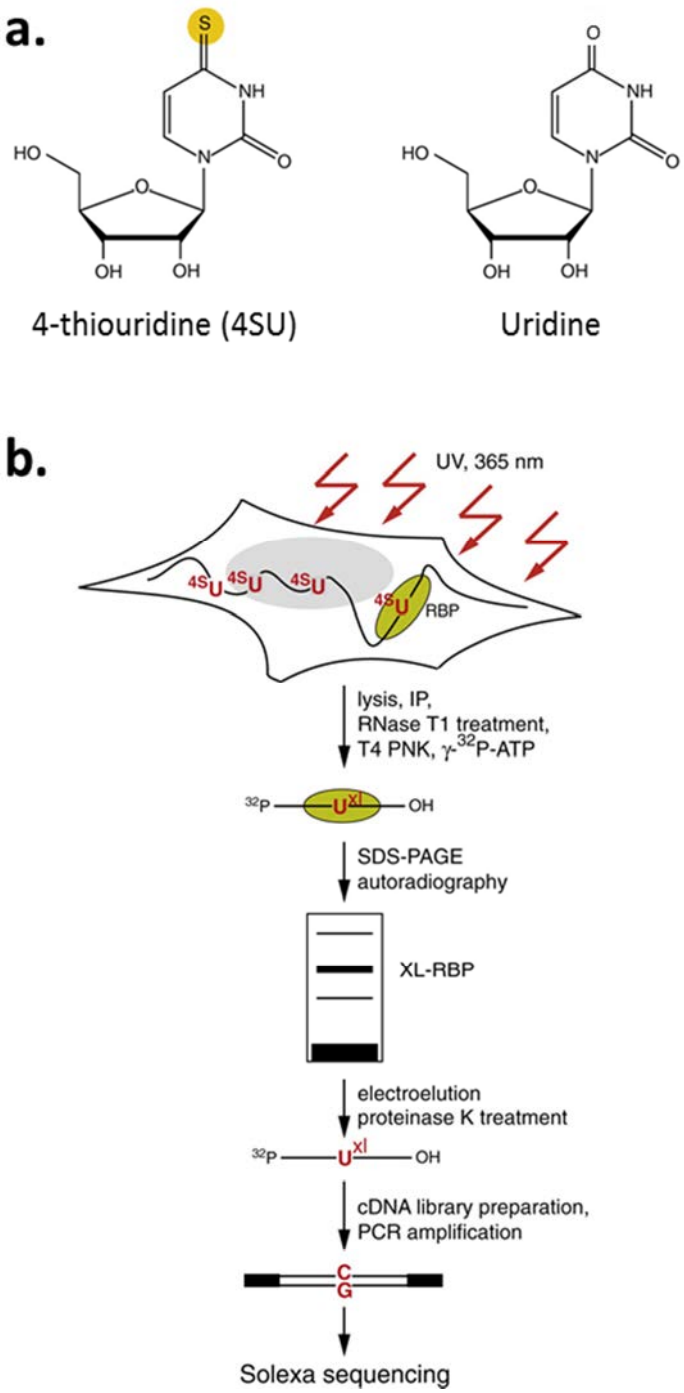


Figure 1.5.2 The experimental scheme of PAR-CLIP. (a) The chemical structure of 4-thiouridine and the ordinary uridine. (b) The exact experimental procedure of PAR-CLIP. **(Adopted from Cell 2010, 141(1):129-141)**

1.5.4 iCLIP

Individual-nucleotide Resolution **C**ross-**l**inking and **I**mmuno**p**recipitation (iCLIP) is a method intended for identifying exact binding sites of RNA-binding proteins [297, 298]. In fact, this primary purpose happens to be one of the advantages of PAR-CLIP. Similar to the above two methods, the procedure of iCLIP also includes UV crosslinking and immunoprecipitation in the library preparation [297, 298]. However, there are some differences in iCLIP. After the effecting Argonaute protein complexes have been captured by anti-Argonaute antibody-coated magnetic beads and the flanking RNAs trimmed away by RNase I, a 3' adaptor is ligated to the 3' end of miRNA targets (Figure 1.5.3). This adaptor is complementary to the primer for reverse transcription (RT) and hence enables the following RT step to occur. Subsequently, proteinase K is used to remove the Argonaute protein from the complexes. In fact, the removal of the Argonaute protein by proteinase K is an incomplete process. A small residue is usually retained at each crosslinking site on the isolated RNAs. These small residues are big hindrances to any reverse transcriptase used in cDNA synthesis and thus terminate RT immediately in front of the crosslinking sites. The cDNAs are then circularized by self-ligation. The circularized cDNAs are cleaved by a restriction enzyme at an internal site of the RT primer. Finally, the linearized cDNAs are amplified by PCR and the amplified products (the library) sequenced by NGS; note that the library represents the isolated target RNAs [297, 298]. Figure 1.5.3 shows the scheme of iCLIP experimental protocol.

An advantage of iCLIP is the simplified ligation steps. It replaces the 5' adaptor ligation step with the self-ligation step. In this ligation step, no matter the reverse transcribed cDNAs are self-ligated or ligated to other cDNA molecules, the ligated products can still be amplified by PCR and sequenced by NGS. Therefore, the ligation efficiency is much better than 5'

adaptor ligation because only inter-molecular ligation is valid in 5’ adaptor ligation. Theoretically, the number of representative molecules can be improved under this setting.

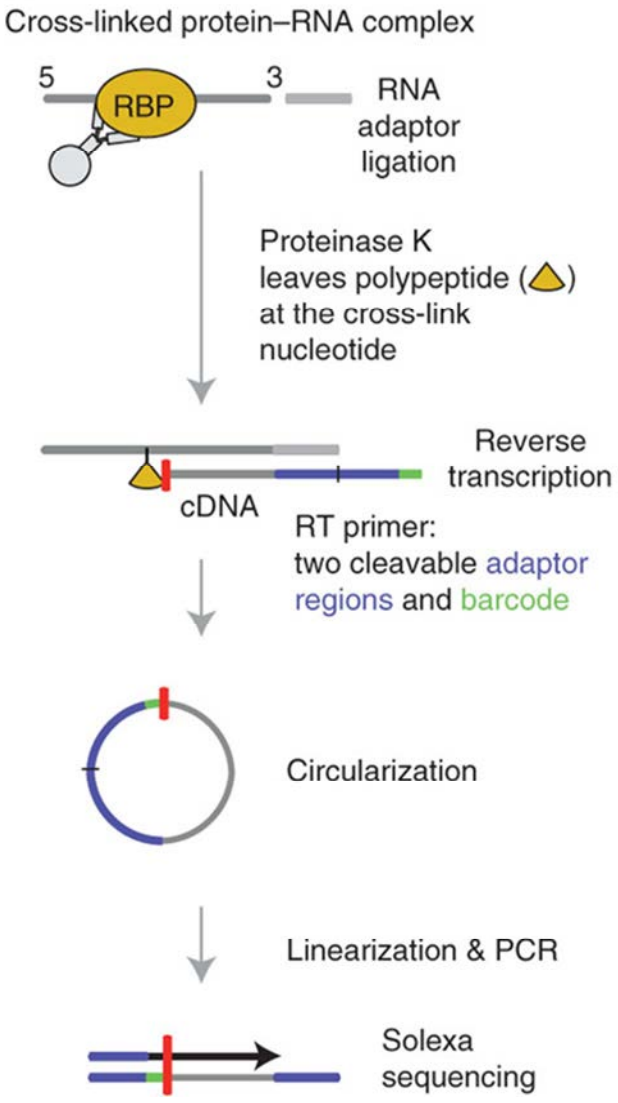


Figure 1.5.3 The experimental scheme of iCLIP. See Section 1.5.4 for details. **(Adopted from Nat Struct Mol Biol 2010, 17(7):909-915)**

1.5.5 CLASH

Unlike the three methods described above, the method **C**rosslinking, **L**igation, **a**nd **S**equencing of **H**ybrids (CLASH) is the first method that not only discovers the target sites of effecting Argonaute complexes, but also detects the exact correspondence between miRNAs and their target sites [299]. This method was first developed for studying protein-mediated RNA-RNA interactions by Tollervey's group in 2011 [300]. With some minor modifications in 2013, the method was adapted for the investigation of interactions between miRNAs and their target sites [299]. The concept of CLASH is highly related to the chromosome conformation capture (3C) method (a chromosome conformation mapping technique) [301]. The CLASH method can be regarded as an RNA version of 3C for identifying the protein-mediated interactions between RNA segments instead of DNA segments.

UV crosslinking is also required in CLASH (Figure 1.5.4). After the crosslinking step, the effecting Argonaute complexes are isolated by anti-Argonaute antibody-coated magnetic beads. The target RNA segments protruding from the Argonaute protein are then removed by RNase digestion. Next, the 5' hydroxyl end of the target RNAs is repaired (phosphorylated) by T4 polynucleotide kinase. Within each individual complex, the interacting RNA-RNA *inter-molecular* ligation is subsequently accomplished by the use of T4 RNA Ligase 1 (key step). To prevent inter-complex ligation, the ligation must be performed under an appropriately dilute condition in which the distance between complexes is large enough to prevent such unwanted ligation. Then, ligation of 3' adaptor and 5' adaptor is sequentially carried out. Separate treatment with shrimp alkaline phosphatase and T4 polynucleotide kinase is required to convert the 3' and the 5' ends of target RNAs to the suitable forms for ligation respectively and individually before each ligation. The Argonaute

protein is removed by proteinase K. Reverse transcription and library amplification by PCR are performed to complete the library preparation. Finally, the prepared library is sequenced by NGS (Figure 1.5.4). There are three kinds of reads in the sequencing results: (i) the chimeric reads in which each of them represents one pair of interacting miRNA and target RNA-binding site; (ii) miRNA reads, which represent miRNAs; and (iii) target-site reads, which represent the target sites of miRNAs [299].

Obviously, CLASH has the advantage that it is able to identify the exact correspondence between miRNAs and miRNA target sites through analyzing the sequence information from chimeric reads. However, only less than 2% of reads are chimeric reads and more than 10% of these chimeric reads are fake chimeras as reported by the research team [299]. In fact, the researchers have conducted a method evaluation to study the reliability of the chimeric reads because inter-complex ligation could still occur even in a very dilute condition. To investigate this uncertainty, they mixed two isolated Argonaute complexes from two distinct species and then proceeded with the CLASH protocol. The results showed that more than 10 % of chimeric reads were cross-species chimeric reads, indicating that more than 10% of the miRNA targets identified by CLASH are not genuine [299].

In summary, the CLASH method is the first method developed for identifying miRNA target sites with the advantage of showing the exact correspondence between miRNAs and their targets. However, the ability of CLASH to achieve its goal is quite limited because only an extremely small fraction of interacting miRNA-target site duplexes can be ligated by CLASH. The minute fraction of chimeric reads and the presence of fake chimeric reads dictate that the miRNA target sites identified by CLASH are not good enough for representing the whole profile of the exact interacting Argonaute-miRNA-target site complexes. This is most probably caused by the inefficient ligation step (the interacting RNA-RNA inter-molecular

ligation). In fact, unlike the setting in studying chromosome conformation, studying the effecting RISC complex is a very different situation. Two major differences are also the problems underlying CLASH. First, both the 5’ and the 3’ ends of a miRNA are anchored inside two individual binding pockets in a RISC (see Section 1.3.2 and Figure 1.3.6). Even though the 3’ end of the miRNA may be transiently released from the 3’ binding pocket during RISC-mediated cleavage of target RNA, the short length of released segment and the huge spatial hindrance will most likely prevent the 3’ end of the miRNA from undergoing ligation. In addition, most of the miRNA-target RNA duplexes in animals contain central mismatches, which prevent RISC-induced target cleavage. Thus, the 3’ end of the miRNAs in the effecting RISC complexes is under conditions unfavorable for ligation. Second, unlike the 3C method for studying the protein-mediated DNA-DNA interactions, CLASH focuses on protein-mediated RNA-RNA interactions. In fact, the flanking DNA segments are in duplex form while the protruding RNA segments are single-stranded. Therefore, the protruding RNA segments are potentially able to form some unexpected intra- or inter-molecular secondary structures. These secondary structures may pose physical hindrance to the ends for ligation. The most apparent secondary structure that prevents the ligation between the 3’ end of the miRNA and the 5’ end of the target site is the miRNA 3’ end base-pairing, i.e. the 3’ end of the miRNA is complementary to its target. Indeed, these two major problems limit the ability of CLASH to reveal the miRNA-target site exact correspondences. Likewise, these are the critical difficulties pending for solutions in CLASH.

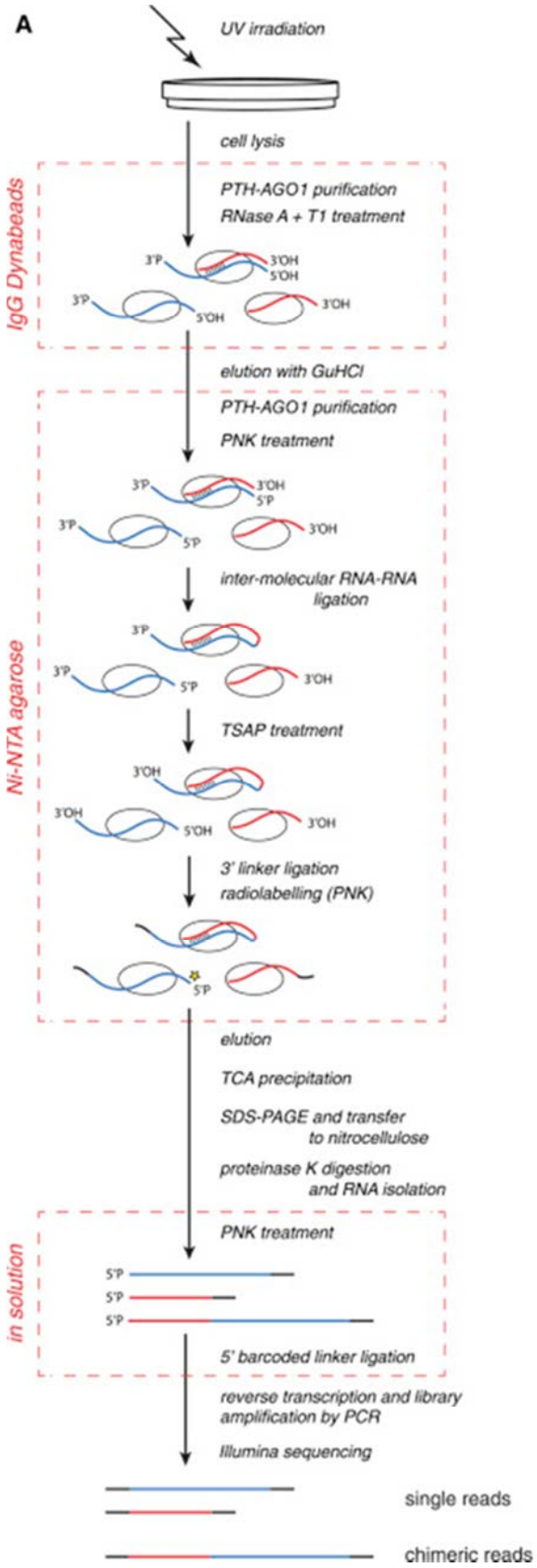


Figure 1.5.4 The experimental scheme of CLASH. See Section 1.5.5 for details. **(Adopted from Cell 2013, 153(3):654-665)**

1.5.6 MARIO

Mapping **R**NA **i**nteractome *in viv**o*** (MARIO) is a method proposed to study protein-mediated RNA-RNA interactions. This method was first reported in 2016 [302], five years after CLASH. Like CLASH, MARIO can also potentially be used to identify miRNA-target site interactions in exact correspondence. However, unlike CLASH, MARIO requires the insertion of a biotinylated short linker between the two interacting RNAs in RNA chimera construction. The chimeras can thus be enriched from the sample by this biotinylated short linker (Figure 1.5.5) [302]. The chimera enrichment is a very distinctive advantage of MARIO.

First of all, the experiment starts with crosslinking the interacting RNA-protein complexes by UV light *in vivo*. Then, the cells are lysed in a lysis buffer. RNA digestion, protein denaturation and protein biotinylation are performed successively (Figure 1.5.5). In order to prevent the RNA-protein complexes of interest from coming too close with each other on the capturing beads, immunoprecipitation is carried out by adding excess amount of streptavidin beads and under a low protein concentration condition. Subsequently, a biotinylated RNA linker is ligated to the 5' end of the co-purified RNAs. Next comes the proximity ligation of the linker-ligated RNA to the 3' end of the interacting partner. Proteinase is used to remove the proteins from the protein-RNA complexes and then all the remaining RNAs are purified by conventional phenol/chloroform/isoamyl alcohol method. With the intention of chimera enrichment, a DNA oligo complementary to the biotinylated RNA linker is added to the purified RNAs. This DNA oligo would form duplex with the RNA linker. T7 exonuclease (an *exo*nuclease with RNase H activity) is added to the mixture to remove all the end-located biotinylated RNA linkers, but not the linkers located internally. Then, all the chimeras can be easily enriched by streptavidin beads. Finally, the enriched

chimeras are processed for adaptor ligation, reverse transcription, PCR and sequencing (Figure 1.5.5) [302]. The cross-species analysis conducted by the original group has found that about 2.5% to 6.8% of chimeras are fakes [302].

Compared with CLASH, MARIO has the advantage that it is specifically design to enrich the chimeras. This feature dramatically reduces the consumption of sequencing capacity for the non-chimeras, which is the major RNA population in the reaction mixture. Moreover, MARIO treats the cell lysate under a harsh condition before the ligation step of generating chimeras. Under such a condition, the proteins of the protein-RNA complexes may be denatured. Therefore, this treatment may release the RNAs from the complexes and facilitate the subsequent ligation. However, this strategy demands that both of the two interacting RNAs are crosslinked to the protein. Regrettably, current crosslinking methods are not efficient. Therefore, this approach fails to capture a large portion of the complexes in which only one of the RNA strands is crosslinked. Indeed, the structure of these complexes may still be stable enough for analysis because the non-crosslinked strand is stabilized by base-pairing with the crosslinked strand.

On the other hand, MARIO used by the original group has discovered only two hundred miRNA-target RNA interactions in the test cell model (including 100 miRNA-mRNA interactions and 100 miRNA-snoRNA interactions). This study actually investigated the complete interactome of all protein-mediated RNA-RNA interaction. Therefore, the few discoveries of miRNA-target RNA interactions might be due to the competition from other highly abundant interactions such as mRNA-snoRNA interaction (the most abundant interactions with several tens of thousands being detected), mRNA-mRNA interaction (nearly ten thousands) and mRNA-tRNA (nearly five thousands). If this is the only reason, then the miRNA-target RNA interactions can be enriched by simply applying an additional

step of enriching the Argonaute protein when this kind of interaction is the focus of study. However, this observation can also be explained in another way. In fact, MARIO uses the enrichment strategy to overcome the low abundance of chimeras in the total isolated RNA. This strategy also implies that the spatial hindrance from the protein (even though the protein is denatured) or the unexpected forms of intra- or inter-molecular secondary structures on the protruding RNA segments is still big enough for rejecting ligation (i.e. the ligation efficiency is still low). Especially for small RNAs, the spatial hindrance is extremely huge because of their relatively small size in the protein-RNA complexes. Furthermore, the entire sequences of miRNAs, siRNAs and piRNAs tend to be complementary to their targets in all these interactions involving small RNAs even though mismatches may be present in some cases. These secondary structures may further prevent or inhibit inter-molecular ligation. In contrast, these two effects are less serious for long RNA-long RNA interactions because their size relative to the protein is big and their length can carry the ends farther away from the protein to allow for ligation. These different hindrances may also contribute to the divergent proportions for different types of chimeric sequences identified by MARIO. If this is really the major cause of the bias, the anticipated ligation efficiency of miRNA-target site must be extremely poor in MARIO and enrichment of Argonaute may not help in discovery of the miRNA-target site interactions. In fact, the enrichment step substantially supported this anticipation.

Chimera enrichment is a strategy for overcoming the consumption of sequencing capacity by the major population of non-chimeras, but is not a good strategy. Obviously, the enrichment of chimeras still cannot overcome the representation bias of the library prepared by MARIO. This means that the small number of obtained chimeras is not enough for representing the entire population of interacting RNA pairs. From this viewpoint, MARIO is quite similar to CLASH. The enrichment may only provide limited advantage to

MARIO. Therefore, MARIO *may* not have a great advantage over CLASH in identifying miRNA interactome. Apparently, a better strategy for overcoming the undesirable consumption of sequencing capacity is to enhance the ligation efficiency because the number of representative chimeras can be elevated only if the ligation efficiency is enhanced.

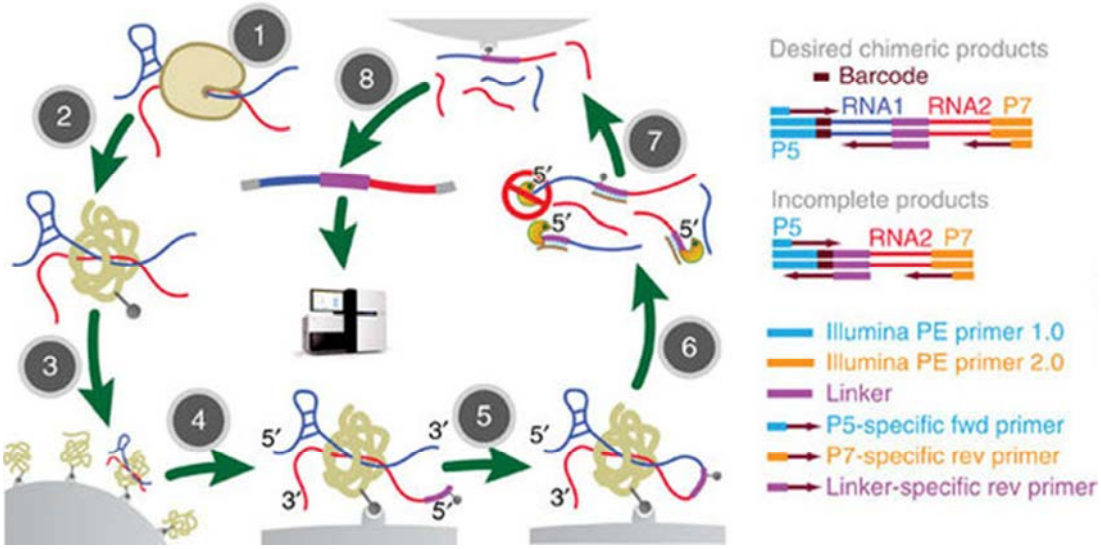


Figure 1.5.5 The experimental scheme of MARIO. See Section 1.5.6 for details. **(Adopted from Nature Communications 2016, 7:12023)**

1.5.7 PARIS

Similar to Mario, **P**soralen **A**nalysis of **R**NA **I**nteractions and **S**tructures (PARIS) is a method proposed to study interactions between RNAs in living cells by Chang's group in 2016 [303], the same year that MARIO was reported. Unlike Mario and CLASH, the preparation of chimeric sequence library in this method requires the process of chemical-assisted duplex stabilization. The chemical used in this method is 4'-aminomethyltrioxsalen (AMT), which is a psoralen derivative. Unlike the case with 4-thiouridine, which is used in PAR-CLIP, crosslinking mediated by AMT is photo-reversible. The crosslinking can be activated by irradiation with 365 nm UV and reversed by irradiation with 254 nm (Figure 1.5.6). Furthermore, 4-thiouridine is an analog of uridine. This compound can be utilized by cells and hence become incorporated into the RNAs that the cell produced (bound to RNA internally). On the other hand, AMT-mediated crosslinking works via the intercalation of AMT into the RNA duplexes and is subsequently activated by covalent crosslinking between AMT and the duplexes by irradiation with 365 nm UV (bound to RNA duplexes externally).

The experiment of PARIS begins with adding an appropriate concentration of AMT to the cultured cells and allows the culture to continue for a suitable time for AMT to be taken up by the cells. The cells are then crosslinked by irradiation with 365 nm UV for 30 minutes. Cell lysate is prepared by disrupting the cells in 3 volume of urea/SDS buffer after the crosslinking. For each 20 millions of cells, 200U of S1 nuclease (Thermo Scientific) is added for digesting the single-stranded portion of RNA into the length suitable for subsequent proximity ligation. S1 nuclease is an endonuclease that preferentially digests single-stranded nucleic acid, but also digests double-stranded nucleic acid when the concentration of the enzyme is high. This enzyme can be used for removing single-stranded termini protruding from double-stranded nucleic acid. After the digestion, proteinase K is

used to release all the RNAs from proteins to facilitate RNA purification by Trizol. Subsequently, ShortCut RNase III (4-8U, NEB) is used to digest each of 20 μg of Trizol-purified RNA in a 50-μl reaction. This RNase is a double-stranded RNase for reducing the size of the double-stranded portion of the RNA in PARIS. The sequential order of the above four steps (S1 nuclease-proteinase K-Trizol purification-ShortCut RNase III) is very critical to the success of PARIS. Actually, the RNA after AMT-mediated crosslinking is not favorable to general Trizol purification protocol because the molecular size of the RNAs becomes extremely big. These "giant RNA" molecules may further aggregate together and result in reduced solubility (this may be the reason why the crude sample is viscous). Thus, RNase treatment is necessary prior to Trizol purification. The digestion of RNA by RNase can efficiently reduce the molecular size of the "giant RNA". In addition, digesting the single-stranded portion of the "giant RNA" is more efficient in reducing the size of the "giant RNA" than digesting the double-stranded portion of the "giant RNA". It is because most of the RNAs are single-stranded. Accordingly, the "giant RNA" is mainly in single-stranded form as well. Therefore, we can expect that *S1 nuclease-proteinase K-Trizol purification-ShortCut RNase III* is better than *ShortCut RNase III-proteinase K-Trizol purification-S1 nuclease*. Furthermore, putting both RNase digestion steps in front of proteinase K-Trizol purification may not be very appropriate either. There are two reasons. First, the optimization can become very complicated because the efficiency and working condition of the two RNases are different. Second, the small size of the RNA causes reduced recovery of RNA by alcohol (ethanol or isopropanol) precipitation of RNA from aqueous phase that is obtained from Trizol extraction (the solubility of RNA in alcohol would increase when the size of the RNA is extremely small, e.g. ~20 nt RNA).

Next, special two-dimensional gel electrophoresis is used to isolate the crosslinked RNAs. This electrophoresis procedure consists of a *non-denaturing* gel electrophoresis for the first

dimension and a *denaturing* gel electrophoresis for the second dimension. The RNA sequences would migrate at different speed in the two dimensions if the RNA sequences contain uncrosslinked double-stranded structure. Only the RNA with crosslinked double-stranded structure or without secondary structure would have similar mobility in the two dimensions and results in locating these RNAs at the diagonal position of the gel after separation. Notably, the location is shifted slightly in the real situation because the mobility is slightly lower in denaturing gel than in non-denaturing gel. A control sample without AMT treatment or crosslinking can be utilized as a reference for localizing the RNAs of interest. The gel containing the crosslinked RNAs is then excavated and crushed for RNA recovery. This procedure effectively removes the majority of unwanted RNAs.

After the crosslinked RNA fragments have been purified, proximity ligation is used to conjugate the interacting RNAs (RNA in the same duplex with crosslinking) together. Here, proximity ligation is conducted under a very stringent condition (e.g. low concentration of RNA, limited ligase, limited ATP, etc.) so that ligation only occurs between neighboring RNAs (interacting RNAs), but not inter-RNA duplexes. Furthermore, the ligation here is also needed to prevent the circularization of the interested RNAs (both open ends are ligated). Then, the RNA sample has to be purified again by alcohol precipitation. In order to reverse the crosslinking and linearize the RNAs of interest, RNA sample is exposed to 254 nm UV for 15 minutes. Finally, a reverse transcription adaptor is ligated to the linearized RNAs. A primer complementary to the adapter is used to convert the linearized RNAs to cDNA library. The cDNA library can then be sequenced by next-generation sequencing.

Based on the underlying principle, PARIS can be used to identify the interactome of RNAs *in vivo* via the use of AMT to stabilize the interacting RNAs. However, this strategy is more favorable to RNA-RNA interactions that do not involve protein binding substantially. For

those protein-mediated RNA interactions, this strategy may not be suitable. Unlike 4-thiouridine, which is bound to RNA internally, AMT needs to bind to the interacting RNA externally. However, the protein that binds the interacting RNA may prevent the binding of AMT into the duplex and hence the crosslinking of RNA. For example, PARIS is not able to identify miRNA-target RNA interaction [303]. In fact, each of the interactions between proteins and specific region of their ligands usually requires multiple anchoring points for recognizing the specific site and stabilizing the binding. In other words, the space between a binding pocket of a protein and its corresponding ligand is extremely narrow and most likely not enough for an extra compound getting in between. This may explain why PARIS is not suitable for identifying protein-mediated RNA interactions.

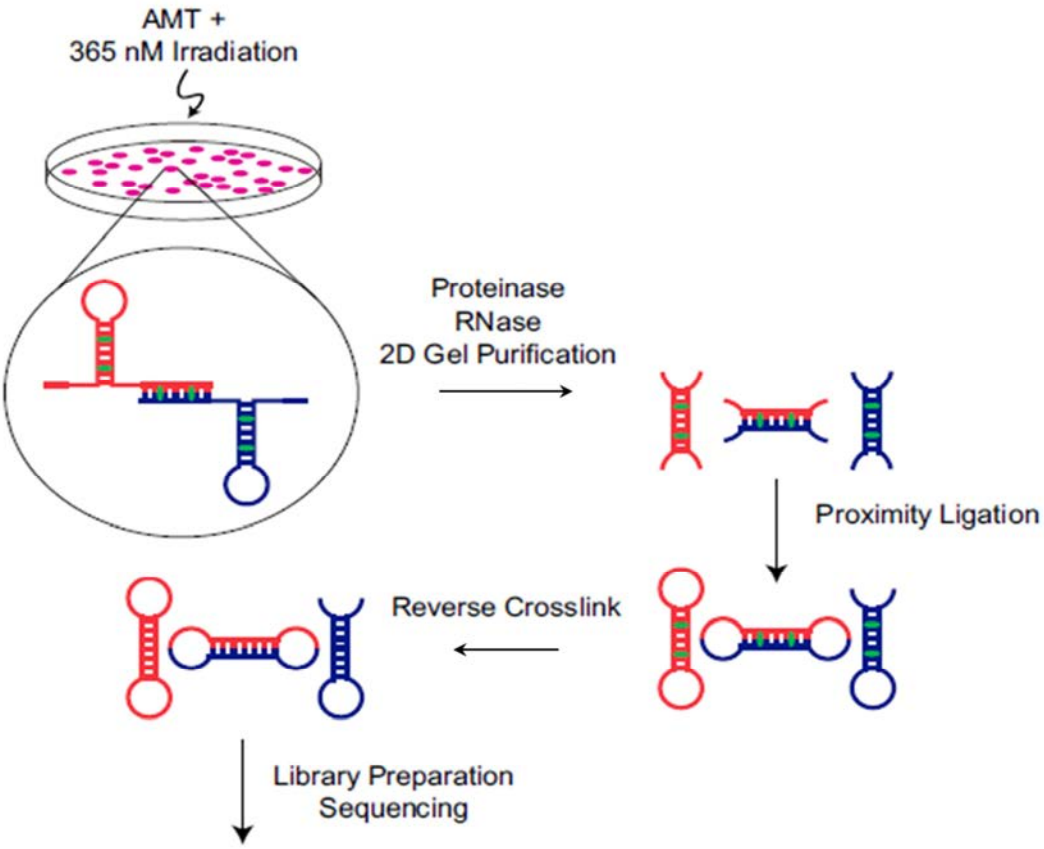


Figure 1.5.6 The experimental scheme of PARIS. See Section 1.5.7 for details. **(Adopted from Molecular Cell. 2016; 63(2): 186-189.)**

1.6 RNA Interactome Identification by Next Generation Sequencing (RIINGS)

With regard to the restrictions and limitations of the current experiment-based methods for identifying microRNA targets, this study primarily aimed at inventing a comprehensive method for identifying miRNA interactomes. Based on a different principle, the proposed method potentially enhances the generation of microRNA-target RNA chimeras from *all* interacting pairs during library preparation. In order to develop the current method, several critical techniques have been developed and evaluated.

First, a novel approach to conjugating the interacting miRNA-target RNA via a stem-loop adaptor has been developed and evaluated in a set of mock experiment. Second, a pair of "repairers" with enhanced ability to repair the mismatches at 5' and 3' ends of miRNA in the miRNA-interacting duplexes has been invented and evaluated in another set of mock experiment. Third, an approach to enhancing the ligation of an adaptor to the 3' end of target RNA has been evaluated. This approach not only is suitable for the "Stem-Loop Adaptor (RNA)" ligation in this method, but also can be used in other RNA interactome identification methods, e.g. HITS-CLIP and PAR-CLIP. However, the invention of "Library Enrichment Beads" (magnetic beads), which are capable of both capturing the interacting microRNA-target RNA duplexes and eluting the captured sequences under an extremely mild and non-denaturing condition, is still in progress. Currently, the capturing ability has been solved in general whereas the elution part is still undergoing testing. In fact, these beads can also be applied to other applications that require on-bead manipulations of nucleic acids if they are successfully developed. On the other hand, mass spectrometry analysis has revealed that the purity of immunoprecipitated Argonaute complexes is good enough for evaluating the current method directly. Furthermore, melting curve analysis has

found that the melting peak of miRNA-target RNA duplexes is about 40℃ and re-annealing of melted duplexes is probably impossible once they are melted. This evidence supports the feasibility of the current method. The conditions for immunoprecipitation, Stem-loop Adaptor ligation, Re-attaching Adaptor ligation, RNA fragmentation, UV crosslinking and proteinase K digestion have been well optimized.

All in all, the development of the current method is almost finished. Only a few optimizations of the step of "Library Enrichment Beads" are still pending for completion. This is a method potentially capable of identifying the complete interactome of microRNAs in a specific cell type or tissue, and revealing the interactome changes in different physiological states of cells. Furthermore, the method may also be used to identify other protein-mediated nucleic acid interactions. This new method is called **R**NA **I**nteractome **I**dentification by **N**ext **G**eneration **S**equencing (RIINGS).

1.6.1 Features and rationale of RIINGS

The RIINGS experiment starts with cells or tissues that have been disrupted into small pieces. Like the other RNA interactome identification methods, RIINGS also requires crosslinking, immunoprecipitation and RNA digestion.

Crosslinking can be achieved by 254 nm UV light or 365 nm UV light with the assistance of 4-thiouridine (4SU). In the latter crosslinking method, the treatment time and the 4SU concentration must be optimized for each type of cells. Tissue samples may not be suitable for crosslinking by the latter method because 4SU may not be taken up by the cells in tissue. After crosslinking, the cells or tissues are lysed in a non-denaturing buffer in order to keep the targeted RISCs intact (intact in native form). The action of this step must be mild enough to prevent the loss of complexes with only one of the two RNA strands being

crosslinked (see the discussion in Section 1.5.6). The targeted RISCs are then isolated from the crude lysate by immunoprecipitation by using anti-Argonaute antibody-conjugated magnetic beads. The antibody used in RIINGS is an antibody that particularly captures the non-denatured (native) Argonaute proteins. After immunoprecipitation, sequential ligation of miRNA 5'-mismatch repairer and then miRNA 3'-mismatch repairer is performed because both ends of the miRNA might stick out from the Argonaute protein. This ligation step can prevent the unwanted ligation of "Stem-Loop Adaptor (RNA)" and "Re-attaching Adaptor" to the miRNA ends protruding from the RISCs (though the probability is low) in the subsequent ligation steps. Then, RNA digestion is performed to trim the RNAs protruding from the RISCs to a length sufficient for adaptor ligation. In fact, crosslinking immunoprecipitation and protruding RNA trimming are based on previous methods, e.g. HITS-CLIP and PAR-CLIP. These steps are commonly used in all of the methods for studying protein-mediated RNA interactions. However, subsequent procedures are original and unique to RIINGS, including the use of stem-loop adaptor for connecting the interacting miRNA-target RNA (main idea of RIINGS).

"Re-attaching Adaptor" and "Stem-Loop Adaptor (RNA)" are ligated sequentially in this order to the 5' end and the 3' end of the target site RNA sequences respectively. The "Re-attaching Adaptor" is used to recover the RNA duplexes from proteinase-K-containing buffer in a subsequent step. By taking advantage of the fact that DNA base pairing is rapid and not susceptible to proteinase K digestion, the capture is based on DNA base pairing via the "Re-attaching Adaptor". The "Stem-Loop Adaptor (RNA)" is used to connect the miRNAs and their corresponding target site to generate the chimeric sequences that represent the miRNA-target site interactions. Subsequently, the Argonaute protein is removed by digestion using proteinase K at 4℃. This low temperature (4℃) keeps the RNA duplexes in duplex form. Furthermore, the activity of proteinase K at 4℃ is already good

enough for removing Argonaute protein from the duplexes. Form this point to the step in which chimeras are generated, all the reactions must be kept at low temperature to avoid the dissociation of the duplexes. After the hindrance from Argonaute protein is removed, the ligations on the duplex can be performed in a much more efficient way. This is one of the major improvements in RIINGS.

Then, RNA duplexes are captured by "Library Enrichment Beads". A second round of ligating miRNA 5'- and 3'-mismatch repairers is performed at the 5' and 3' ends of miRNAs respectively and sequentially. These two repairers generate new ends on the mismatched ends and these newly generated ends are potentially able to match their corresponding target sequence. The gap between the miRNA 5' end and the "Stem-Loop Adaptor (RNA)" 3' end is then repaired (filled) by the ligation of some short fillers (RNA tetramers). During this ligation process, the 3' end of the miRNAs is also extended, finally reaches and then passes the restriction enzyme cleavage site on the "Re-attaching Adaptor". After these ligation steps, the chimeric sequences are generated.

These chimeric sequences can be eluted by the restriction enzyme BanI because the "Re-attaching Adaptor" includes a BanI cutting site. Finally, the chimeras are recovered through reverse transcription, PCR and NGS. A quality control step for quantifying the chimera number is performed on the reverse transcribed product. All the details and evaluations of each of the above steps are discussed in Sections 4, 5 and 6.

1.6.2 Significances of RIINGS

Compared with the previously developed methods for identifying microRNA targets, the current method has the following advantages. First, the current method is able to identify the miRNAs-target site interactions *in exact correspondence* whereas HITS-CLIP, PAR-CLIP,

iCLIP and PARIS do not. Second, the current method potentially enhances the generation of microRNA-target RNA chimeras by removing the spatial hindrance from Argonaute proteins. This approach can most likely enhance the ligation efficiency in the ligation of adaptors and eventually improve the chimeric sequence generation. Third, the approach of generating chimeric sequence in RIINGS is most likely able to avoid the generation of fake chimeras because the 3' end of the target site (the end for generating chimeras) is protected by a stem-loop structure. Fourth, the increased efficiency of generating chimeras can result in a better representation of the entire population of the miRNA-target site interactions and generate a precise comprehensive map of miRNA interactions. Therefore, RIINGS is an important method with high value in identifying miRNA interactomes. Since miRNAs play a key role in gene expression regulation (mRNA gene expression and long non-coding RNA functioning), the understanding of miRNA interactomes can most probably explain many of phenomena in cells, such as differentiation, cell cycle regulation, cell metabolism and tumorigenesis. By comparing the differences of miRNA interactomes between different states of cells, e.g. cancer cells vs. normal counterpart, the detailed control and regulation of gene expression in response to the changes may probably be revealed. Because miRNAs can target many non-canonical target sites and target genes, e.g. 5' UTRs, introns, multiple types of non-coding RNAs and pseudogenes [134, 295, 297, 299, 302], some currently hidden functions of miRNAs may also be discovered. Therefore, a comprehensive method for identification of miRNA interactome is attractive in the biology field.

2. Aims and Objectives

The aim of the current study is to develop a novel technology (RIINGS) that can identify the *entire* miRNA-target RNA interactome *in exact correspondence* in a specific cell type or tissue. To achieve this aim, we have the following objectives:

1. **Selecting an appropriate model for the development of RIINGS** — Evaluate the expression level of AGO proteins.
2. **Establishing a suitable immunoprecipitation procedures for RIINGS** — Select and evaluate an antibody to evaluate different immunoprecipitation approaches.
3. **Establishing a novel approach to linking the interacting miRNA-target RNA** — Develop and evaluate the method for linking the interacting miRNA-target RNA via a stem-loop adaptor.
4. **Establishing the Re-attaching Adaptor for RIINGS** — Design the Re-attaching Adaptor and evaluate the ligation of the adaptor to the 5' end of the miRNA target site on *in vivo* isolated-AGO complex.
5. **Optimizing the RNA fragmentation procedure for RIINGS** — Optimize the working concentration of RNase A and RNase I on *in vivo* isolated-AGO complex.
6. **Identifying an optimal crosslinking condition for RIINGS** — Compare and evaluate the 254 nm UV light-assisted crosslinking method and photoactivatable ribonucleoside-enhanced crosslinking method.
7. **Investigating the approach to enhancing the efficiency of ligating adaptor to the 3' ends of RNase-treated RNA** — Evaluate the improvement to the ligation of adaptor

to 3’ ends of RNase-treated RNA (on *in vivo* isolated-AGO complex) by the assistance of T4 polynucleotide kinase.

8. **Establishing a pair of mismatch repairers for miRNAs** — Design a pair of miRNA mismatch repairers for repairing the mismatches on miRNA 5’ and 3’ ends, and evaluate the ligation of the repairers to the 5’ and 3’ ends of a mock miRNA.
9. **Assessing the purity of the isolated-AGO complex** — Evaluate whether the purity of the *in vivo* isolated-AGO complex is good and suitable enough for RIINGS.
10. **Measuring the thermostability of the interacting miRNA-target RNA** — Determine the melting temperature of *in vivo* isolated-miRNA-target RNA duplex.
11. **Quantifying the amount of *in vivo* isolated-miRNA-target RNA duplex** — Establish an approach to estimating the amount of yielded miRNA-target RNA duplex (before the two RNA sequences are linked) and an approach to quantifying the amount of chimeras in the RIINGS-prepared library (after the two RNA sequences are linked).
12. **Optimizing the digestion of proteinase K** — Investigate the digestion of Argonuate proteins by proteinase K at low temperature and evaluate whether proteinase K can release the miRNA-target RNA duplex from the *in vivo* isolated-AGO complex.

3. Schematic diagram of RIINGS

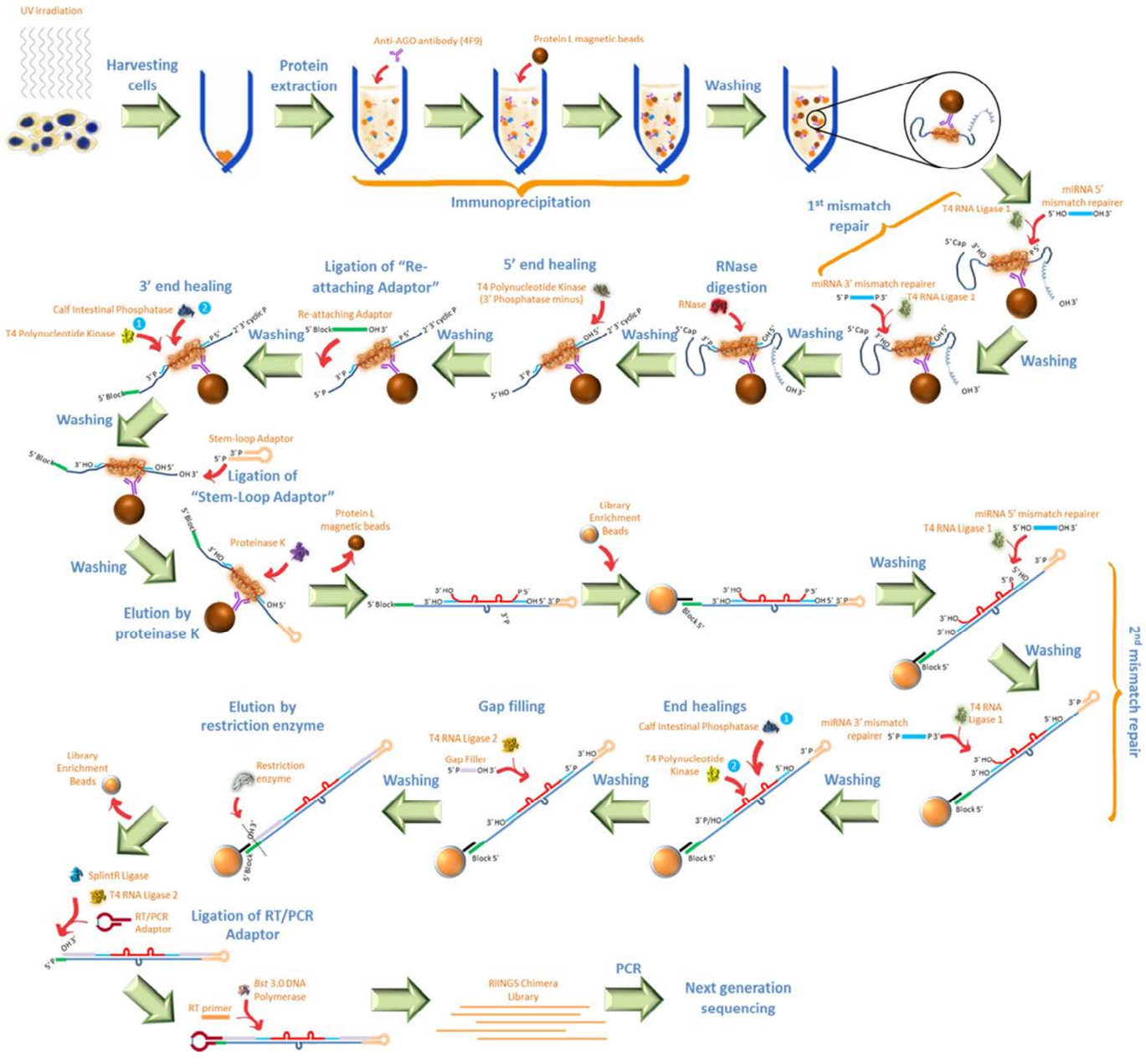


Figure 3.1 The workflow of RNA Interactome Identification by Next Generation Sequencing (RIINGS)

A brief description

The flow chart on the left indicates the working scheme of RNA Interactome Identification by Next Generation Sequencing (RIINGS). The workflow begins with UV irradiation to crosslink the *in vivo* effecting Argonaute complex, which includes an Argonaute protein, a miRNA and a target RNA. Both ends of miRNA may protrude from the Argonaute protein. The exposed ends have 3' hydroxyl and 5' phosphate, and are thus suitable for ligation. The procedure includes one extra mismatch repairing step (1st mismatch repair) before the ligation of "Re-attaching Adaptor" and "Stem-Loop Adaptor (RNA)" to prevent the unwanted ligation of these adaptors to the miRNA. After the 1st mismatch repair, the exposed ends of the miRNA would become 3' phosphate and 5' hydroxyl, and hence are unsuitable for ligation. On the other hand, the ends of miRNAs protected by Argonaute protein are still suitable for ligation (3' hydroxyl and 5' phosphate), but will be released only after protein digestion by proteinase K. Therefore, a second mismatch-repairing step (2nd mismatch repair) and a second round of ends healing are carried out after proteinase K treatment. As a result, all miRNA ends (exposed and protected) are ligated with their respective repairers, have 3' hydroxyl and 5' phosphate, and are now ready for the following gap filling step. Next come the generation of chimeric sequences by RT-PCR and the sequence analysis of the chimeric library by NGS.

4. Methods

4.1 Introduction

Generally speaking, miRNAs perform their functions via loading onto Argonaute proteins to form the RISCs (one miRNA loads onto one Argonaute protein). The effecting RISCs bind to their corresponding target site for initiating their functions through the guidance of each particular miRNA. The recognition of target site by the miRNA does not require fully complementary base-pairing in animals. Therefore, identification of miRNA target sites is a difficult task. This difficulty has slowed down our understanding of miRNA machinery and effects. Recently, two experiment-based methods (i.e. CLASH and MARIO) have been developed for identifying the exact interactions between miRNAs and their target sites. However, these two methods are still far from the ultimate goal of discovering the *entire* interactome in a specific cell type or tissue, which is caused by the inefficient preparation of interactome library.

The primary objective of the current study was to develop RIINGS – a comprehensive method for identifying miRNA interactomes. Through a *novel* strategy to improve the generation of chimeric sequences that represent the interacting miRNA-target site duplexes, RIINGS is potentially suitable for identifying the miRNA interactome in its entirety and in an exclusive manner. This method includes several important steps and techniques. First, basic procedures are used to stabilize and isolate the effecting RISCs by UV light-mediated RNA-protein crosslinking and immunoprecipitation respectively. Second, a *novel* approach is used to connect the pairs of interacting miRNA-target site by specifically ligating a “Stem-Loop Adaptor (RNA)” to the 3’ end of the target site. After it has been

ligated, this adaptor generates a gap between miRNA and its target site (between the 5' end of miRNA and the 3' end of the "Stem-Loop Adaptor (RNA)" to be exact; see Figure 3.1). The two interacting sequences can be physically linked together by filling the gap in between. Third, the removal of Argonaute proteins from the isolated RISCs by proteinase K has to be carried out at a low reaction temperature. The low reaction temperature serves to stabilize the interacting duplexes. Fourth, a pair of repairers (short RNA oligos) for repairing the mismatches at the 5' end and 3' end miRNAs must be ligated to the 5' and 3' ends of miRNAs before the filling of the gap between the interacting miRNAs and their target sites. Fifth, the "Re-attaching Adaptor" is ligated to the 5' end of the target sites. The "Re-attaching Adaptor" is designed to facilitate the recovery of the interacting duplexes after the step of protein digestion from the buffer containing proteinase K. Sixth, a unique reverse transcription reaction is developed for RNA templates containing 5-nitroindole (a universal base that can pair up with anyone of A, U, C and G). Seventh, "Library Enrichment beads" are used in RIINGS for the enrichment of the chimera sequences by capturing the interacting duplexes and eluting the captured sequences under an extremely mild and non-denaturing condition. The flowchart for demonstrating the complete workflow of RIINGS is shown in Chapter 3.

The method development is divided into multiple experiments for developing, testing and evaluating the above required procedures and techniques. The experimental methods and results of these experiments are described in Chapters 4 and 5 respectively.

4.2 Design of primers, adaptors and linker-loop

All the primers, adaptors and linkers were designed using with the primer analysis software Oligo (version 6.71; Molecular Biology Insights). Secondary structures of the designed oligos were further analyzed by both Oligo and Oligo Analyzer 3.1 (Integrated DNA Technologies (IDT)) when necessary. All the sequences of the primers, adaptors and linkers used in this study were ordered from IDT and the sequences are shown in Table 4.2.1.

Table 4.2.1 The sequences of primers, adaptors and linkers (Sequence format are according to IDT online ordering system.)

Oligo Name	Purification	Sequence (5' → 3')
Stem-Loop Adaptor (RNA)	RNase-free HPLC	rArArA rArArC rCrGrC rUrCrA rGrGrA rArGrG rCrGrA rArGrC rArCrC rGrArC rCrGrA rGrArC rArGrG rUrGrA rGrCrG rG
Re-attaching Adaptor (DNA-RNA)	RNase-free HPLC	GCTA GCT AGC TGA GCG CTA GGA AAG AGC TAG CTG ArGrA rGrCrU rG
Lib-A Primer A	Standard Desalting	CGT ATC GCC TCC CTC GCG CCA TC
Lib-A Primer B	Standard Desalting	CTA TGC GCC TTG CCA GCC CGC TC
Chimera Quanti probe	HPLC	/56-FAM/TCT CGG TCG GTG CTT CGC CTT C/3IABkFQ/
Chimera Quanti Synt Temp	Standard Desalting	CTA TGC GCC TTG CCA GCC CGC TCA GTC AAA AAC CGC TCA GGA AGG CGA AGC ACC GAC CGA GAC AGG TGA GCG GTT TTT GCA CGA CTG ATG GCG CGA GGG AGG CGA TAC G
Artificial Binding Site (RNA)	RNase-free HPLC	rArArC rUrArU rUrArC rCrUrG rArCrA rCrUrU rArUrC rUrUrU rArCrA rUrArC rCrUrA rCrUrU rArCrU rCrUrC rA
Artificial Mir1	RNase-free HPLC	/5Phos/rArArU rGrArA rArArG rArGrA rArCrU rGrUrC rArGrG rU
Artificial Mir2	RNase-free HPLC	/5Phos/rUrArU rGrArA rArArG rArUrA rArGrU rGrArC rArGrG rC
Artificial Mir3	RNase-free HPLC	/5Phos/ rArUrC rGrUrA rArArG rArGrA rArGrA rGrUrC rArGrC rC
Artificial Mir4	RNase-free HPLC	/5Phos/ rArUrU rGrUrC rArArG rArGrA rArCrU rGrUrC rArUrC rC
Random C/I Tetramer	Standard Desalting	[rC/rI] [rC/rI] [rC/rI] [rC/rI]
Tetramer N	Standard Desalting	rNrNrN rN
Tetramer C	RNase-free HPLC	rCrCrC rC
Tetramer (CCCI)	RNase-free HPLC	rCrCrC rI
Tetramer C-Phosphorylated	RNase-free HPLC	/5Phos/rCrCrC rC/3phos/
Tetramer-Phosphorylated (ICCC)	RNase-free HPLC	/5Phos/rIrCrC rC/3phos/
5' Repairer (GXXGA)	RNase-free HPLC	rG/i5NitInd//i5NitInd/rGrA
5' Repairer (GXAGA)	RNase-free HPLC	rG/i5NitInd/rA rGrA
3' Repairer (AAXXG)	RNase-free HPLC	/5Phos/rArA/i5NitInd//i5NitInd/rG/3phos/
3' Repairer (AXXXG)	RNase-free HPLC	/5Phos/rA/i5NitInd//i5NitInd//i5NitInd/rG/3phos/
Repaired 5' end Template	RNase-free HPLC	rUrUrA rUrCrU rUrUrA rCrArU rArCrC rUrArC rCrGrU rUrGrA rGrUrU rCrUrC rArArC rG
Repaired 5' end (GUXG-)	RNase-free HPLC	/5Phos/rGrU/i5NitInd/rG rGrUrA rArGrU rArArA rGrArU rArA/3SpC3/
Repaired 5' end (GXAG-)	RNase-free HPLC	/5Phos/rG/i5NitInd/rArG rGrUrA rArGrU rArArA rGrArU rArA/3SpC3/

Repaired 3' end Template	RNase-free HPLC	/5Phos/rCrGrU rArArA rGrUrU rCrUrU rUrArC rGrCrU rArUrU rArCrC rUrGrA rCrArC rUrUrA /3SpC3/
Repaired 3' end (-XXAG)	RNase-free HPLC	rUrArA rGrUrG rArCrA rGrGrU rA/i5NitInd//i5NitInd/rArG
Repaired 3' end (-XXXG)	RNase-free HPLC	rUrArA rGrUrG rArCrA rGrGrU rA/i5NitInd//i5NitInd//i5NitInd/rG
Artificial Mir5 (FAM)	RNase-free HPLC	/5Phos/rArUrG rGrArC rUrGrA rGrArA rGrArG rArArA rArGrU rArA/36-FAM/
Artificial Binding Site (RNA, HEX)	RNase-free HPLC	/5HEX/rArCrU rCrUrC rArUrU rCrArU rCrCrA rUrArC rArUrU rUrCrU rArUrU rCrArC rArGrU rCrCrA rUr(N:25252525)r(N) r(N)r(N)r(N) r(N)r(N)r(N) r(N)r(N)r(N) r(N)r(N)r(N) r(N)r(N)rG rCrCrU rCrUrC rArUrG rCrUrC rArCrG rArArU rUrUrU rGrArG rArGrG rC
5-P Tetramer N	Standard Desalting	/5Phos/rNrNrN rN
5-NitInd RT primer 1	Standard Desalting	TCT CAA CAT GAG AGA AAT GTG G
5-NitInd RT temp 1	RNase-free HPLC	rArUrC rCrArU rArCrA rU/i5NitInd//i5NitInd//i5NitInd/rUrA rUrUrC rArCrA rG/i5NitInd//i5NitInd/rCrA rUrUrA rUrCrA/i5NitInd/ rGrCrC rCrArC rArUrU rUrCrU rCrUrC rArUrG rUrUrG rArGrA rG
5-NitInd RT primer 2	Standard Desalting	GGA AAT GAA TGC TGA AT
5-NitInd RT temp 2	RNase-free HPLC	rArUrC rCrArU rArC/i5NitInd/rA rUrUrA rUrUrC rArCrA rCrCrA rUrArC rUrCrC rArU/i5NitInd/rA rUrCrU rArUrU rCrArG rCrArU rUrCrA rUrUrU rCrC

4.3 Selection of cell line and extraction buffer for the method development of RIINGS

Cell lines are the most suitable model for the method development of RIINGS because they are easily to obtain, not limited in the source, not involved with ethical issues if commercial cell lines are used, and mostly representative of cells or tissues or cells of their origins.

Argonaute 1-4 (AGO1-4) are involved in miRNA-mediated processes and are the effector proteins that bind to miRNAs [217, 218]. In order to reveal the binding partners of all functioning miRNAs under the physiological conditions of cells, a capturing method that can capture all these complexes should be chosen. Since AGO1-4 are the core proteins of miRNA-containing ribonucleoprotein (RNP) complexes, cell lines with higher expression levels of AGO proteins would be better suited for use in the capturing method optimization and the entire RIINGS development.

In addition, it is extremely important to keep the Argonaute-miRNA-target RNA complexes intact (i.e., in non-denatured form) in the protein lysate because the miRNA-target RNA duplex may be dissociated in damaged complexes. In order to maintain the native conformation of AGO proteins for the use in RINGS, non-denaturing extraction method and non-denaturing extraction buffer were tested in current investigations. Two potential non-denaturing extraction buffers, 1X Detergent-containing Lysis Buffer and 1X Detergent-free Lysis Buffer (Appendix A; from Abcam online immunoprecipitation protocol), were tested for the yield of the AGO proteins in this experiment [304].

MEC-1, K-562 and Raji cell lines were tested for the expression level of AGO proteins because they are readily available in our laboratory. MEC-1 was established from the peripheral blood of a 61-year-old Caucasian male with B-cell chronic lymphocytic leukemia

(B-CLL) in prolymphocytoid transformation to B-cell prolymphocytic leukemia (B-PLL) in 1993 [305]. K-562 is a cell line derived from a pleural effusion of a 53-year-old female patient suffering from chronic myelogenous leukemia (CML) for 4 years in 1970 [306]. K-562 is the first established myelogenous leukemic cell line with Philadelphia translocation [306]. Raji is a human B-lymphocyte cell line derived from an 11-year-old black male with Burkitt lymphoma in 1963 [307]. Moreover, MEC-1 and Raji cell lines carry the latent Epstein-Barr virus (EBV) genome and are positive for Epstein-Barr nuclear antigen (EBNA) [305, 307] while K-562 is free from EBV, herpes-like virus and mycoplasma [306]. Figure 4.3.1 shows the experimental scheme for selecting cell line and extraction buffer.

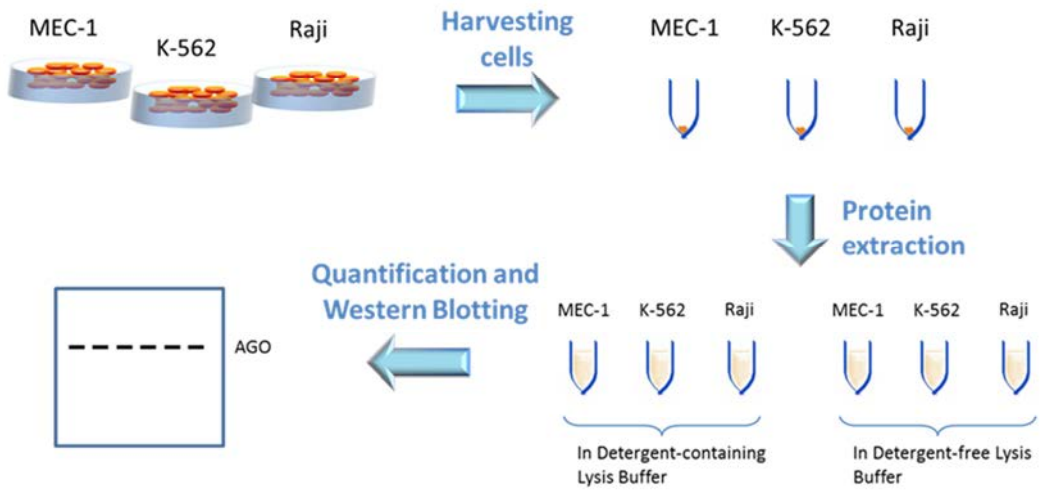


Figure 4.3.1 The experimental scheme for selecting cell line and extraction buffer. The cell line with the highest expression of AGO and the extraction buffer giving the highest yield of AGO were selected for use in the subsequent method development.

4.3.1 Cell culture

A cryotube containing around 5 million of cells in 2 ml cryoprotectant medium (culture medium with 20% fetal calf serum and 10% dimethyl sulfoxide [DMSO]) was taken from liquid nitrogen and immediately thawed in water bath at 37°C. Then, the cells were transferred to a 15-ml centrifuge tube and washed with 10 ml pre-warmed (37°C) culture medium (Iscove's Modified Dulbecco's Medium [IMDM] for MEC-1 and K-562 cells, and Roswell Park Memorial Institute [RPMI] 1640 medium for Raji cells). Subsequently, cells were collected by centrifugation at 125 g (800-1000 revolutions per minute or rpm) for 5 minutes. After centrifugation, the supernatant was discarded and the cell pellet resuspended in 10-ml pre-warmed (37°C) complete growth medium (IMDM with 10% fetal calf serum for MEC-1 and K-562 cells, and RPMI 1640 culture medium with 10% fetal calf serum for Raji cells). Finally, this 10-ml cell suspension was grown at 37°C in a T-25 culture flask in an incubator with 5% CO_2. To maintain the cell culture, cells were split into two aliquots and sub-cultured in T-75 or T-175 culture flask with subsequent resuspension at 3 – 5 × 10^5 cells/ml every 2 to 3 days when the cell density reached 2 – 3 × 10^6 cells/ml.

4.3.2 Protein extraction

Cells growing in exponential phase were collected by centrifugation at 125 × *g* (800-1000 rpm) for 5 minutes. The supernatant was discarded and the cell pellet washed with ice-cold phosphate-buffered saline (PBS) 3 times. Cell pellets of MEC-1, K-562 and Raji cells were stored at -70°C until the day for protein extraction. Each of the cell pellets of MEC-1, K-562 and Raji cells with around 5 × 10^7 cells was disrupted by a pestle in 500 µl of 1X Detergent-containing or 1X Detergent-free Lysis Buffer in a 1.5-ml centrifuge tube. Each lysate was then centrifuged at 13000 × *g* for 15 min at 4°C. Subsequently, the supernatant was transferred to a new 1.5-ml centrifuge tube and was ready for protein quantification.

4.3.3 Protein quantification

Protein concentrations were determined by means of Pierce® BCA Protein Assay Kit (Thermo Scientific) in accordance with the manufacturer's instructions. Absorbance at 562 nm was measured by Benchmark Plus Microplate Spectrophotometer System (Bio-Rad). The extracted cell lysates were stored at -20℃ until use.

4.3.4 Western blotting

Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad) was used to run sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transfer proteins from a gel to a polyvinylidene difluoride (PVDF) membrane. All the samples were mixed with 2X Laemmli Sample Buffer (Appendix A) in a 1:1 ratio and heat-denatured at 95℃ for 5 min before being loaded onto the gel. The loading amount and the loading volume of each sample were 30 µg and 30 µl respectively. The gels used in the SDS-PAGE were 8% separating and 4% stacking gels (acrylamide:bis ratio, 29:1, i.e. 3.3% crosslinker) (Appendix B), and had a dimension of 8.3 × 7.3 × 0.1 cm (width × length × thickness). Gels were allowed to polymerize for at least 2 hours. According to the instructions from the manufacturer (Bio-Rad), the SDS-PAGE was run under constant voltage at 200V until the dye front (bromophenol blue) reached the bottom of the gel in 1X SDS-PAGE Running Buffer (Appendix A) . The electric power was supplied by PowerPac™ Basic Power Supply (Bio-Rad). Proteins were then transferred to PVDF membrane (0.45 µm pore size; Millipore), which was activated in 100% methanol prior to use. The sequential order of the transfer sandwich from the cathode to the anode was fiber pad, 2 sheets of filter papers, gel, PVDF membrane, 2 sheets of filter papers and fiber pad. The transfer was run under constant voltage at 100V for 1 hour at 4℃ in 1X Transfer Buffer (Appendix A). During the transfer, the negatively charged proteins traveled from the gel onto the membrane according to the

electric field. In order to minimize the background signal from the non-specific binding of primary and secondary antibodies to the PVDF membrane, the membrane was blocked after transfer in 1X TBST buffer (Appendix A) with 5% (w/v) skim milk at room temperature for 1 hour with shaking on either rocking shaker or horizontal shaker. The primary antibody, eIF2C2 (4F9) (Santa Cruz), was diluted in 1X TBST with 1% (w/v) skim milk in a ratio of 1:500. The membrane was then incubated in this diluted antibody at 4℃ overnight with shaking. An intensive wash of the membrane was conducted after the incubation with the primary antibody. The membrane was washed in 1X TBST for 5 min each for 3 times and finally in 1X TBST for 10 min. The secondary antibody, Anti-mouse IgG HRP-linked Antibody (Cell Signaling), was diluted in 1X TBST with 1% (w/v) skim milk in a ratio of 1:2000. The membrane was incubated with shaking in this diluted secondary antibody for 1 hour at room temperature after the intensive wash. Subsequently, a second intensive wash was conducted. The washing procedures were the same as the first intensive wash. Chemiluminescence signal was developed by adding Clarity™ Western ECL Substrate (Bio-Rad) onto the washed membrane in accordance with the manufacturer's instructions. The image was then captured by ChemiDoc™ MP Imaging System (Bio-Rad) with Image Lab™ Software version 4.1. The signal intensities of the bands were calculated by Image Lab™ Software version 4.1.

4.4 Selection and evaluation of antibodies

Argonaute (AGO) family is considered to play a crucial role in miRNA-mediated regulation of gene expression. Among the AGO family in humans, 8 family members have been discovered [216]. They are under 2 subfamilies, the PIWI subfamily and eIF2C/AGO

subfamily [216]. Previous studies have found that all the members, AGO1 to AGO4, in the eIF2C/AGO subfamily are able to bind miRNAs in an indiscriminate manner and AGO2 is the only member responsible for RNA cleavage [217, 218].

As all the 4 members in eIF2C/AGO subfamily are capable of binding miRNA, the 4 AGO proteins may take part in miRNA-mediated regulations. In order to detect all the RNA binding partners of miRNAs under the physiological condition of cells, the following three anti-AGO antibodies all against AGO 1-4 were selected for testing their suitability for AGO capture in the current method:

1) Anti-pan Ago Antibody, clone 2A8 (Millipore) – an antibody originally used in HITS-CLIP, which was developed by Darnell's group [134]; a mouse monoclonal antibody raised against histidine-tagged human recombinant AGO protein;
2) eIF2C2 (4F9) (Santa Cruz), a mouse monoclonal antibody raised against human full-length recombinant elF2C2; and
3) eIF2C Antibody (B-3) (Santa Cruz), a mouse monoclonal antibody raised amino acids 552-851 at the C terminus of human elF2C2.

The binding ability and specificity of the candidate antibodies for the non-denaturing form of AGO protein were assessed by dot-blotting experiment. Figure 4.4.1 shows the experimental scheme for antibody selection.

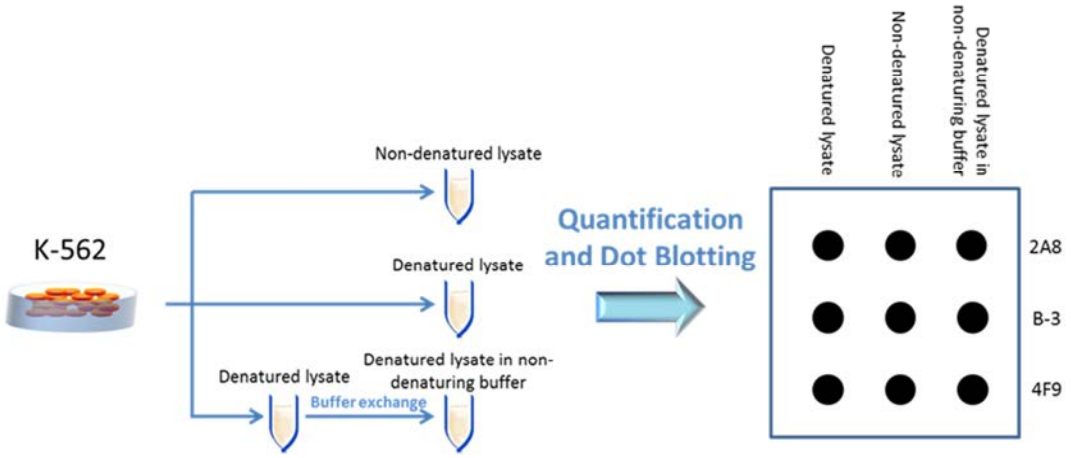


Figure 4.4.1 The experimental scheme for antibody selection. The antibody with the highest binding to AGO proteins would be selected as the capturing antibody to AGO proteins in the immunoprecipitation in RIINGS.

4.4.1 Cell growth, protein extraction and protein quantification

K562 cells were used in this experiment. Procedures of cell growth were the same as those described in Section 4.3.1. Three different extraction procedures with different extraction buffers were used for preparing the total cell lysate to generate three different denaturing states of total cell lysates, i.e. non-denatured lysate, denatured lysate and denatured lysate with buffer exchange to non-denaturing buffer. First, non-denatured lysate was made by extracting the cells in 1X Detergent-free Lysis Buffer (Appendix A). The procedures were the same as those described in Section 4.3.2. Second, denatured lysate was obtained by diluting the above non-denatured lysate with 2X Laemmli Sample Buffer (without dye) (Appendix A) in a ratio of 1:1 and then boiling the mixture at 95℃ for 5 minutes. Third, denatured lysate with buffer exchange to non-denaturing buffer was made by homogenizing the cell pellet (~5 × 10^7 cells) in Denaturing Lysis Buffer (Appendix A) [304], and then heating the lysate at 95℃ for 5 minutes. Then, the lysate was centrifuged at 8000 *g* for 15 minutes at 4℃. The supernatant was transferred to a new 15-ml centrifuge tube and topped up to 4 ml with 1X Detergent-free Lysis Buffer. Subsequently, the diluted lysate was transferred to the sample reservoir of Amicon Ultra-4 (30K) (Millipore) and centrifuged at 4000 *g* for 40 minutes. An aliquot of 3.5 ml of 1X Detergent-free Lysis Buffer was added to the sample reservoir after the centrifugation. The lysate was centrifuged again at 4000 *g* for 40 minutes. Finally, the remaining buffer-exchanged lysate was transferred to a new 1.5-ml centrifuge tube. Protein quantification for all the total cell lysates was carried out as previously described in Section 4.3.3.

4.4.2 Dot blotting

With Minifold I SRC 96 D Dot Blotter (Schleicher & Schuell), cell lysates were applied onto nitrocellulose membrane (0.45 µm pore size; Millipore), which was equilibrated in 1X

Detergent-free Lysis Buffer (Appendix A) prior to use. All the lysate samples from Section 4.4.1 were diluted with 1X Detergent-free Lysis Buffer to a concentration of 0.025 µg/µl before being loaded into the sample wells. The loading amount and the loading volume of each lysate sample were 5 µg and 200 µl respectively. Sample spotting was performed in accordance with the manufacturer's instructions. The nitrocellulose membrane was dried after the sample had been spotted onto it. The steps of blocking, antibody probing and signal development were carried out as previously described in Section 4.3.4. Three selected anti-AGO antibodies – Anti-pan Ago Antibody, clone 2A8 (Millipore), Anti-Ago2 monoclonal antibody 4F9 (Santa Cruz) and eIF2C Antibody (B-3) (Santa Cruz), all of which reacted with AGO 1-4 – were tested for their suitability for RIINGS. Anti-mouse IgG HRP-conjugated Antibody (Cell Signaling) was used as the secondary antibodies for detecting all three primary antibodies, which were raised from mouse. Chemiluminescence signal was developed by adding Clarity™ Western ECL Substrate (Bio-Rad) onto the washed membrane in accordance with the manufacturer's instructions. The image was then captured by ChemiDoc™ MP Imaging System (Bio-Rad) with Image Lab™ Software version 4.1. The signal intensities of the dots were calculated by Image Lab™ Software version 4.1.

4.5 Immunoprecipitation of AGO proteins

Immunoprecipitation is the first critical steps in RIINGS. Since the miRNA-target RNA interaction is a protein-mediated interaction, the interacting RNA duplexes can be captured and enriched by immunoprecipitation. AGO proteins are the core proteins in the

interacting miRNA complex. Therefore, anti-AGO antibody can be used to capture the interacting RNA duplexes.

Generally speaking, immunoprecipitation can be achieved in two different ways. One way is to immobilize the antibodies on the beads prior to capturing the target protein from the total cell lysate. Another way is to incubate the total cell lysate with the antibodies for a period of time and then capture the antibody-target protein complexes by adding the beads to the lysate-antibody mixture. The former strategy can usually reduce the non-specific binding. However, this strategy also reduces the yield of the target protein. The latter strategy can usually elevate the yield of the target protein; however, the purity may be relatively poorer than that of the former strategy. Both methods were tested in this section. Moreover, the input amount of the antibody was also investigated. Figure 4.5.1 shows the experimental scheme for evaluating immunoprecipitation methods.

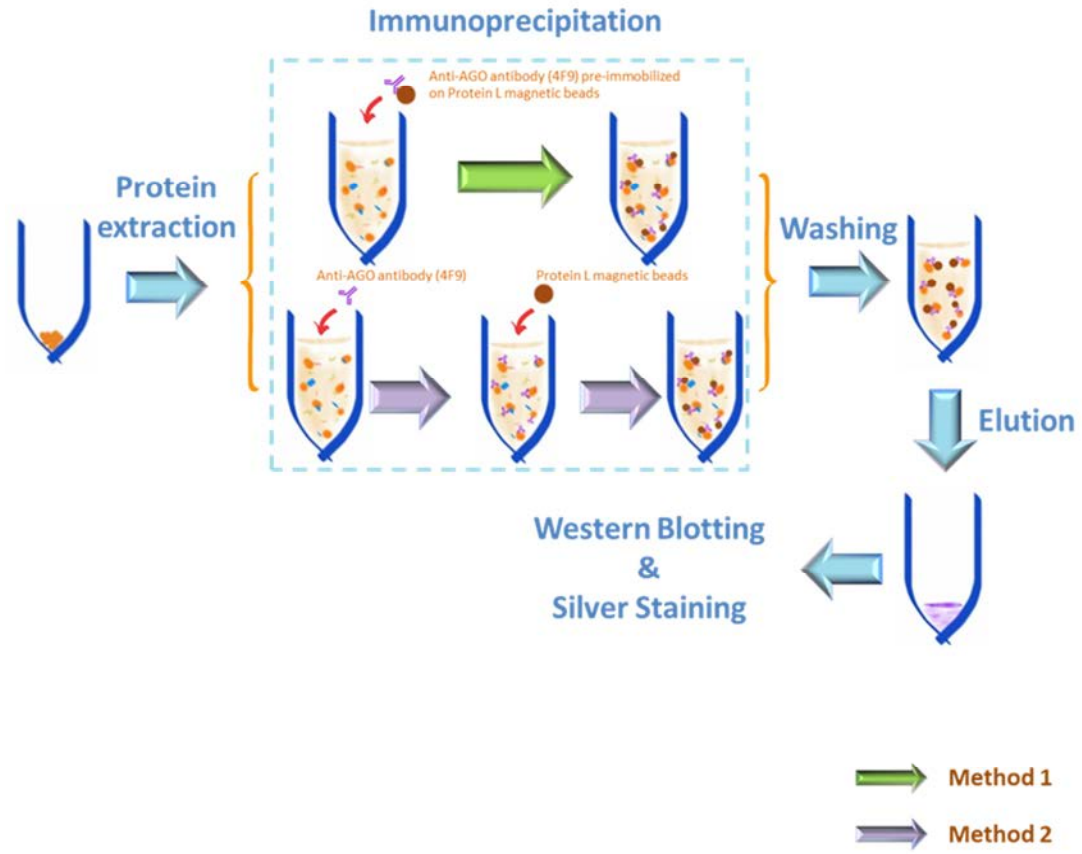


Figure 4.5.1 The experimental scheme for evaluating immunoprecipitation methods. Both methods were compared for the yield of AGO proteins by western blotting, and the purity of the yielded AGO proteins by silver staining. The appropriate input amount of anti-AGO antibody (4F9) was also evaluated.

4.5.1 Preparation and quantification of protein lysate

K562 cell extract was used in this experiment. Procedures for cell culture, protein extraction and quantification were the same as those described in sections 4.3.1 to 4.3.3 respectively. Total protein was extracted in 1X Detergent-free Lysis Buffer (Appendix A).

4.5.2 Preparation of beads for immunoprecipitation

Beads for Method 1 (see Figure 4.5.1)

The antibody eIF2C2 (4F9) (Santa Cruz), a mouse monoclonal antibody against the natural forms of AGO1-4 proteins [308], was selected for the immunoprecipitation step in the RIINGS. Pierce™ Protein L Magnetic Beads (Thermo Scientific) were resuspended by rotation for 5 minutes at room temperature. An aliquot of 150 µl of the bead suspension was transferred to a 1.5-ml tube. The tube containing the bead suspension was then placed on a magnetic stand for 2 minutes to attract the beads to side of the tube next to the magnet. The time for collecting the beads on the side wall of the tube was extended if necessary until the supernatant became clear. The supernatant was then discarded and the beads were washed three times each with 1X Detergent-free Lysis Buffer of the same volume as the original bead suspension. Subsequently, the washed beads were resuspended in 37.5 µl (7.5 µg) of Anti-Ago2 monoclonal antibody 4F9 (Santa Cruz) diluted in 1X Detergent-free Lysis Buffer to a final volume of 400 µl. By calculation, the binding capacity of the beads was 22-fold higher than the amount of the antibody used in the experiment. The mixture was rotated for 2 hours at room temperature to allow the antibody to immobilize on the beads. Finally, the supernatant was saved for SDS-PAGE and the proteins were visualized with silver staining to check whether the antibody was completely captured by the beads. The beads were washed three times each with 1X

Detergent-free Lysis Buffer of the same volume as the original bead suspension. The prepared beads could be left in last wash step if not used immediately.

Beads for Method 2 (see Figure 4.5.1)

Similarly, an aliquot of 150 µl of the bead suspension was used in the preparation. The preparation procedure was the same as described above for Method 1, but finished before adding antibody to the washed beads.

4.5.3 Immunoprecipitation

Method 1

For Method 1, each immunoprecipitation reaction used 2 mg of total protein extract as the starting material for the experiment. The 2 mg of total protein extract was topped up to 400 µl with 1X Detergent-free Lysis Buffer. Next, the diluted protein extract was transferred to the vial containing the beads prepared for Method 1 and incubated at 4℃ for 4 hours with rotation. The vial was placed on a magnetic stand and the supernatant was saved for western blotting to check whether all the targeted AGO proteins were captured by the beads. Then, the beads were washed thrice with 0.5 ml 1X Detergent-free Lysis Buffer. Elution was carried out by adding 60 µl of 1X Laemmli Sample Buffer to the vial containing the beads, and by heating the vial for 10 minutes at 70℃.

Method 2

Likewise, each immunoprecipitation reaction also used 2 mg of total protein extract as the starting material for Method 2. In order to compare with Method 1, an equivalent amount of Anti-Ago2 monoclonal antibody 4F9, i.e. 7.5 µg (37.5 µl), was employed. Additionally, a

comparable reaction with double amount of Anti-Ago2 monoclonal antibody 4F9, i.e. 15.0 µg (75.0 µl), was used to investigate the optimal antibody input amount. The antibody was added to the total protein extract and the reaction volume was brought to 400 µl with 1X Detergent-free Lysis Buffer. The antibody-protein extract mixture was then incubated at 4℃ for 2 hours with rotation. Afterward, the mixture was transferred to the vial containing the beads prepared for Method 2 and incubated at 4℃ for an additional 2 hours with rotation. In a way similar to the procedure for Method 1, the beads were collected with a magnetic stand and the supernatant was removed and saved for western blotting to check whether all the targeted AGO proteins were captured by the beads. Lastly, the beads were washed thrice with 0.5 ml 1X Detergent-free Lysis Buffer. The attached antibody was eluted by adding 60 µl of 1X Laemmli Sample Buffer to the vial containing the beads and by heating the vial for 10 minutes at 70℃.

4.5.4 Evaluation of the immunoprecipitation by western blotting

The procedures for western blotting were the same as previously described in Section 4.3.4 with some exceptions. The gels used in SDS-PAGE were 10% separating and 4% stacking gels (Appendix B) of 1 mm thick. An aliquot containing 20 µg of total proteins (total protein extract, the antibody, the supernatant or the eluate) was loaded for the SDS-PAGE. All the samples were mixed with 2X Laemmli Sample Buffer (Appendix A) and water to give a final concentration of 1X Laemmli Sample Buffer and a final volume of 20 µl. The transfer was run under constant voltage at 100V for 1.5 hour at 4℃ in 1X Transfer Buffer. An antibody raised from a non-mouse species was preferred to be the primary antibody in this western blotting because the eluate from the immunoprecipitation contained a mouse antibody (Anti-Ago2 monoclonal antibody 4F9). In order to prevent the signal from this mouse antibody, a rabbit antibody, Argonaute 2 (C34C6) Rabbit mAb (Cell Signaling), was selected

as the primary antibody here. The antibody was diluted in 1X TBST with 1% (w/v) skim milk in a ratio of 1:1000 for the working primary antibody. Similarly, the secondary antibody, Anti-rabbit IgG HRP-linked Antibody (Cell Signaling), was diluted in 1X TBST with 1% (w/v) skim milk in a ratio of 1:1000 for the working secondary antibody.

4.5.5 Evaluation of the immunoprecipitation by protein gel electrophoresis with silver staining

SDS-PAGE was use to resolve proteins by molecular weight. The gels used here for SDS-PAGE were 10% separating and 4% stacking gels (Appendix B) of 1 mm thick. Aliquots each containing 20 µg of total protein extract or the equivalent were loaded for SDS-PAGE. The samples included the Anti-Ago2 monoclonal antibody 4F9 (Santa Cruz), the supernatant of the antibody solution after conjugation to the beads from the bead preparation for immunoprecipitation Method 1, total protein extract of K562 cells, and the supernatants and the eluates of K562 cells' total protein extract after immunoprecipitation by Method 1 and Method 2. All the samples were mixed with 2X Laemmli Sample Buffer (Appendix A) and water to give a final concentration of 1X Laemmli Sample Buffer and a final volume of 20 µl. The samples were heat-denatured at 95℃ for 5 min before being loaded onto the gel. SDS-PAGE was run under constant voltage at 200V until the dye front reached the bottom of the gel in 1X SDS-PAGE Running Buffer. The gel was then fixed in Fixing Solution at room temperature for 1 hour. The gel was washed overnight in several changes of water at room temperature in order to remove all residual acetic acid. The gel was sensitized in 0.02% sodium thiosulfate for 1 minute at room temperature. Then, the gel was washed three times each in water for 20 seconds. Subsequently, the gel was incubated in cold 0.1% silver nitrate solution for 20 minutes at 4℃ with rocking. After staining, the gel was washed three times again each in water for 20 seconds. The gel was placed on a new tray and washed in

water for an additional 1 minute to further remove the residual silver nitrate so as to minimize the background staining. Development was achieved by incubating the gel in the Developer Solution with intensive shaking until the staining was sufficiently strong. Developer Solution was immediately changed when it turned yellow. The gel was washed in water for 20 seconds after it had been stained adequately. Staining was terminated by incubating the gel in 5% acetic acid for 5 minutes with rocking.

4.6 Ligation of a stem-loop adaptor

In the case of miRNA interaction, miRNA and its target form a duplex with the assistance of an Argonaute protein. The two interacting molecules are two different RNA sequences. They match together only through base pairing and the formation of hydrogen bonds between paired bases (A=U, C≡G and G=U). This interaction is easily destroyed by any one of the conventional extraction methods, including Trizol-based extraction method and column-based extraction method, owing to the weak bonding force of hydrogen bond. Accordingly, it is critical and important to join the miRNA and its target together before extraction or use any appropriate strategies to stabilize the duplex when the interaction between miRNA and its target is the focus.

A stem-loop adaptor has to be ligated to the 3' end of the miRNA target as shown in the scheme of RIINGS (Figure 3.1). This adaptor effectively extends the 3' end of the miRNA target and flips it over and backward, and positions the 3' end of the “miRNA target” and the 5' end of the miRNA in a head-to-tail manner with a gap in between. Moreover, the loop of the adaptor conceals a probe-binding site because it is in fact the binding site of a

specific probe. This binding site facilitates the quantification of valid cDNA by means of qPCR or digital PCR in a library prepared by RIINGS.

Ligation of the “Stem-Loop Adaptor (RNA)” to the 3’ end of an miRNA target is an on-bead procedure. Radioactivity was used to visualize the adaptor-ligated miRNA-target RNA-Argonaute complexes owing to the low abundance of these complexes. Three investigations into the ligation of stem-loop adaptor were conducted: (i) time-course analysis to determine the optimal time for the ligation of stem-loop adaptor, (ii) titration analysis to determine the optimal concentration of stem-loop adaptor in the ligation process, and (iii) titration analysis to study the optimal concentration of T4 RNA Ligase 1 (ssRNA Ligase, High Concentration; New England Biolabs (NEB)) in the ligation process. Figure 4.6.1 shows the experimental scheme for studying the ligation of a stem-loop adaptor to the 3’ end of a miRNA target.

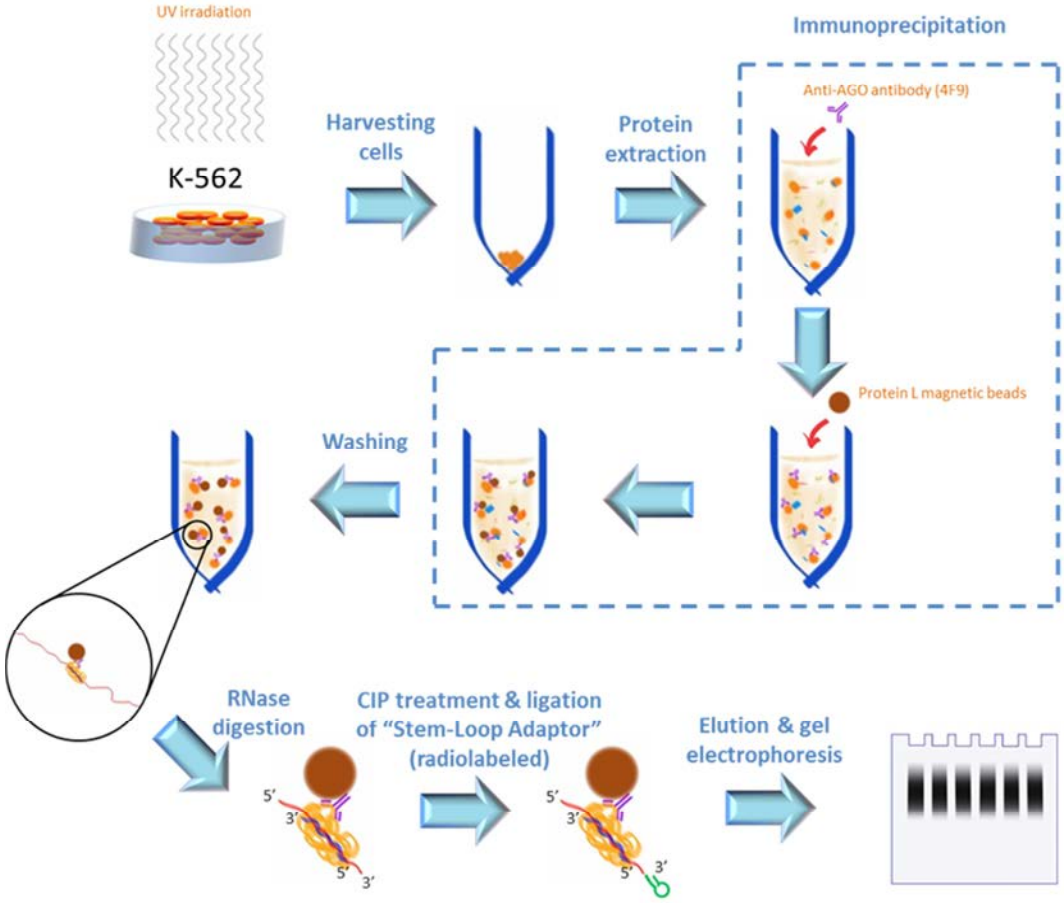


Figure 4.6.1 The experimental scheme for studying the ligation of a stem-loop adaptor to the 3' end of a miRNA target. A time course study was conducted to find the optimal ligation time. Titration analyses were also performed to find for the optimal concentrations of stem-loop adaptor and T4 RNA Ligase 1 (ssRNA Ligase).

4.6.1 Cell culture, UV irradiation, and preparation and quantification of protein lysate

K562 cell extract was used in this experiment. Procedures for cell culture were the same as previously described in Section 4.3.1. About 5×10^7 cells growing in exponential phase were collected by centrifugation at 125 *g* (800-1000 rpm) for 5 minutes. The supernatant was discarded and the cell pellet washed twice with 20 ml ice-cold 1X Hank's Balanced Salt Solution (HBSS) (Thermo Scientific). The cell pellet was resuspended in 2 ml of ice-cold 1X HBSS. This suspension was then transferred to a 10-cm dish and spread out on the dish uniformly. The dish of cells was irradiated twice under 254-nm UV light, once for 400 mJ/cm^2 and once for 200 mJ/cm^2, on a 4℃ pre-chilled ice pack in UVC 500 Ultraviolet Crosslinker (Amersham). The two irradiations were separated by 1 minute. The suspension was mixed before each irradiation. After the UV irradiation, an aliquot of 8 ml of ice-cold 1X HBSS was transferred to the dish to facilitate the collection of the cells from the dish. All 10 ml of the suspension was transferred to a 15-ml centrifuge tube, and the tube centrifuged at 125 *g* (800-1000 rpm) for 5 minutes at 4℃. The supernatant was discarded and then resuspended in 1ml of ice-cold 1X HBSS. The suspension was transferred to a 1.5-ml centrifuge tube and the cells were collected by centrifugation at 350 *g* (3000 rpm) for 5 minutes at 4℃. Total protein was extracted in 1X Detergent-free Lysis Buffer (Appendix A). The cell pellet was disrupted by a pestle in 500 µl of 1X Detergent-free Lysis Buffer supplemented with RNase Inhibitor (Murine; NEB) at a final concentration of 1 U/µl. Each total cell lysate was then centrifuged at 13000 *g* for 15 min at 4℃. Eventually, each supernatant was transferred to a new 1.5-ml centrifuge tube. Procedures for protein quantification were the same as described in Section 4.3.3.

4.6.2 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads and the subsequent immunoprecipitation followed the Method 2 described in Sections 4.5.2 and 4.5.3 respectively. There were 3 sets of experiments. (i) Time-course analysis was carried for determining the optimal ligation time: 4 reactions were prepared for the time-course study. The time points were 0, 12, 24 and 36 hours of ligation. (ii) Titration analysis was performed to determine the optimal concentration of Stem-Loop Adaptor (RNA): 4 reactions were prepared to test 4 concentrations of adaptor (240 nM, 480 nM, 720 nM and 960 nM). (iii) Titration analysis was done to determine the optimal concentration of T4 RNA Ligase 1 (ssRNA Ligase, High Concentration; NEB): 10 reactions were prepared for testing the completeness of ligation under 2 different concentrations of ligase (1.5 U/µl and 3 U/µl) after 0, 6, 12, 24 and 36 hours of ligation. The amount of the antibody input for each reaction was 11.25 µg (56.25 µl). According to the calculation, the binding capacity of the Pierce™ Protein L Magnetic Beads was 16.5-fold higher than the amount of the antibody. The elution procedure in Section 4.5.3 was omitted and the reactions were continued with the steps in Section 4.6.4.

4.6.3 Preparation of stem-loop adaptor (radiolabeling)

The stem-loop adaptor, which was named Stem-Loop Adaptor (RNA), was ordered from IDT. The lyophilized RNA was dissolved in an appropriate volume of UltraPure™ Distilled Water (Life Technologies) to obtain the Stem-Loop Adaptor (RNA) with a concentration of 100 µM. Table 4.2.1 shows the sequence of the Stem-Loop Adaptor (RNA). The adaptor was labeled with ^{32}P-γ-ATP (PerkinElmer) by T4 Polynucleotide Kinase (NEB). A reaction mixture of 50 µl was set up for radiolabeling the adaptor with the following components: 2.4 µM Stem-Loop Adaptor RNA), 1X T4 Polynucleotide Kinase Reaction Buffer (70 mM Tris-HCL, 10 mM $MgCl_2$,

5 mM dithiothreitol [DTT], pH 7.6 at 25°C; NEB), 5 µCi/µl ^{32}P-γ-ATP (Perkin Elmer), 1 U/µl T4 Polynucleotide Kinase (NEB), 1 U/µl RNase Inhibitor, Murine (NEB), and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 50 µl (See Appendix C Table 1 for the preparation).

The reaction mixture was incubated at 37°C for 1 hour. Then, 2 µl of 1 mM ATP was added to the reaction and the mixture further incubated for another 5 minutes. The reaction was then inactivated at 65°C for 20 minutes. The final product was purified by illustra™ MicroSpin™ G-25 Columns (GE Healthcare) in accordance with the manufacturer's instructions. The purified product was stored at -20°C if not used immediately.

4.6.4 RNase treatment, ligation of stem-loop adaptor, and protein gel electrophoresis and visualization

Magnetic beads from Section 4.6.2, which had been washed with 1X Detergent-free Lysis Buffer, were resuspended in 150 µl of 1X Detergent-free Lysis Buffer (same volume as the original bead preparation). RNase digestion was carried out by adding an aliquot of 3 µl of 1:1000 diluted RNase One™ Ribonuclease (Promega) to each reaction (see Figure 4.6.1). The reactions were incubated on Eppendorf ThermoMixer® C (Eppendorf) at 4°C for 30 minutes with mixing at 1000 rpm for 15 seconds at 2-minute intervals. The reactions were then washed thrice with 150 µl of 1X Detergent-free Lysis Buffer. Before the on-bead calf intestinal phosphatase (CIP) treatment, the beads were equilibrated with 1X NEBuffer 1 (NEB) by washing twice with 150 µl of 1X NEBuffer 1. The CIP treatment mixture for each reaction consisted of 1x NEBuffer 1 (10mM Bis-Tris-Propane-HCl, 10mM $MgCl_2$, 1mM DTT, pH 7.0 at 25°C; NEB), 0.5 U/µl Alkaline Phosphatase, Calf Intestinal (NEB), 1 U/µl RNase

Inhibitor, Murine (NEB), and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 150 µl (See Appendix C Table 2 for the preparation)

The beads were resuspended in the above CIP treatment mixture (see Figure 4.6.1). The reaction mixture was incubated on Eppendorf ThermoMixer® C at 16℃ for 1 hour with mixing at 1000 rpm for 15 seconds at 2-minute intervals. The beads were washed with 0.5 ml of Phosphatase Wash Buffer twice after the CIP treatment. Next, the beads were equilibrated with 1X T4 RNA Ligase Reaction Buffer (without DTT) by washing again with 0.5 ml of 1X T4 RNA Ligase Reaction Buffer (without DTT) twice prior to the ligation of the radiolabeled stem-loop adaptor. During the last wash, the individual reactions in the same set of experiment were pooled together, thoroughly mixed and then re-divided again for eliminating the variations arising from the above preparation procedures.

Three different sets of experiments were conducted: (i) time-course analysis for the determination of optimal ligation time, (ii) titration analysis for determining the optimal concentration of Stem-Loop Adaptor (RNA), and (iii) titration analysis for determining the optimal concentration of T4 RNA Ligase 1 (ssRNA Ligase, High Concentration; NEB). For the first analysis, 480 nM of stem-loop adaptor and 1.5 U/µl of T4 RNA Ligase 1 were chosen for the experiment. Meanwhile, 4 different final concentrations (240 nM, 480 nM, 720 nM and 960 nM) of stem-loop adaptor together with a fixed concentration of T4 RNA Ligase 1 (1.5 U/µl) were tested in the second analysis. For the third analysis, 2 different final concentrations of T4 RNA Ligase 1 (1.5 U/µl and 3 U/µl) plus a fixed concentration of stem-loop adaptor (960 nM) in the radiolabeled stem-loop adaptor ligation mixture were investigated. The radiolabeled stem-loop adaptor ligation mixture for each reaction was made up of the following components: 1X NEBNext® Quick Ligation Reaction Buffer (66mM Tris-HCl, 10mM $MgCl_2$, 1mM DTT, 1mM ATP, 6% Polyethylene glycol (PEG 6000), pH 7.6 at

25°C; NEB), 1.5-3 U/µl T4 RNA Ligase 1 (ssRNA Ligase; NEB), 240-960 nM radiolabeled Stem-Loop Adaptor (RNA), 1 U/µl RNase Inhibitor, Murine (NEB), and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 100 µl (see Appendix C Table 3 for the preparation).

Each reaction of the beads was resuspended in 100 µl of corresponding radiolabeled stem-loop adaptor ligation mixture (see Figure 5.6.1) and was incubated on Eppendorf ThermoMixer® C at 16℃ with mixing at 1350 rpm for 15 seconds at 2-minute intervals. The 4 reactions for the time-course analysis of optimal ligation time were collected after 0, 12, 24 and 36 hours of incubation respectively. Meanwhile, all the reactions for the titration analysis of optimal Stem-Loop Adaptor (RNA) concentration were collected after 36 hours of incubation. Moreover, the 10 reactions for the titration analysis of optimal T4 RNA Ligase 1 (ssRNA Ligase, High Concentration; NEB) concentration were collected after 0, 6, 12, 24 and 36 hours of incubation respectively. The collected beads were washed 5 times with 0.5 ml of 1X Detergent-free Lysis Buffer (Proteinase Inhibitor-free) immediately after the ligation step. Elution was carried out by adding 25 µl of NuPAGE® LSD Sample Buffer (4X) (Thermo Fisher) and 75 µl UltraPure™ Distilled Water (Thermo Fisher) to the vial containing the washed beads, and by heating the vial for 10 minutes at 70℃. For each samples, a portion of 50 µl of the eluate was loaded onto NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well (Thermo Fisher), and the gel run under 175V at 4℃ for 90 minutes in 1X NuPAGE® MOPS SDS Running Buffer (Thermo Fisher). Lastly, the gel was wrapped by a Saran Wrap for avoiding any leakage and then exposed to an X-ray film for 24 hours. The film was developed by a film processor (FUJIFILM FPM 2800, FujiFilm). The signal intensities of the bands were analyzed by a gel electrophoresis image analysis software tool (GelAnalyzer 2010a; http://www.gelanalyzer.com/).

4.7 Ligation of re-attaching adaptor

After the ligation of the stem-loop adaptor in the experimental scheme of the RIINGS, the 3' end of the miRNA target extends via the stem-loop adaptor and in effect flips over and backward. This modification positions the 3' end of the miRNA target and the 5' end of the miRNA in a head-to-tail orientation and with a gap in between (Figure 3.1). Therefore, these two interacting RNAs can be joined together by filling up the gap in between the two ends. In the scheme of RIINGS, filling the gap by ligation was selected instead of filling the gap by synthesis. Under non-denaturing condition, however, both ends of the miRNA in the RNA duplex are actually shielded by the Argonaute protein or the even bigger protein complex, i.e. the RISC. Crystal structure studies of RNA-Argonaute protein complex have shown that both ends of miRNA are anchored to two different binding pockets [133, 214, 309]. The 5’ end of the miRNA anchors to the binding pocket located inside the MID domain of the Argonaute protein while the 3’ end of the miRNA anchors to the binding pocket located inside the PAZ domain of the protein [133, 214, 309]. Although the 3’ end of the miRNA is most likely released from the binding pocket in the PAZ domain during the process of the sequence-guided and -induced cleavage, the 3’ end of the miRNA is substantially anchored to the binding pocket under the gene translational repression processes [219]. Furthermore, the release of the end can only occur at the 3’ end of the miRNA and the released end may not be long enough or the spatial hindrance is too big for ligase to perform ligation.

Therefore, removal of the duplex-binding proteins, including the Argonaute proteins, is necessary to expose the duplex and particularly the gap to the ligase enzyme for its action in the RIINGS experiment. Here, proteinase K was selected for this purpose. However, this process also releases all the immobilized duplexes from the Pierce™ Protein L Magnetic

Beads (Figure 3.1). Therefore, it does not facilitate the subsequent steps. In addition, the proteinase K remaining in the reaction mixture also does not favor all following enzymatic reactions because it can digest proteins under a wide range of conditions.

In order to remove proteinase K after its action and hence facilitate subsequent steps, a "re-attaching adaptor" would be ligated to the 5' end of the target RNA *prior to* the proteinase K treatment. This "re-attaching adaptor" actually helps the miRNA-target RNA duplexes attach to new magnetic beads after the mild proteinase K treatment. In brief, these magnetic beads carry on their surface with a specific sequence. Therefore, these beads can capture target nucleic acid sequences via base pairing. These beads are one of the inventions in the current study and are named "Library Enrichment Beads". After the duplexes have been re-attached to magnetic beads, proteinase K can be washed away easily and the duplexes then processed for subsequent steps.

The ligation of the re-attaching adaptor is an on-bead procedure and is very similar to the ligation of the stem-loop adaptor described in Section 4.6. Instead of ligating an adaptor to the 3' end of the miRNA target, this ligation actually ligates an adaptor to the *5' end* of the miRNA target carried by the same Argonaute complex. The following experiments use a radioactive isotope to visualize the adaptor-ligated miRNA-target RNA-Argonaute complex owing to the low abundance of the complex. As the optimal adaptor concentration and ligase concentration had been determined in the ligation of stem-loop adaptor in Section 4.6, the optimization of these two parameters are omitted here. As the enzymatic kinetics is different between 5' end ligation and 3' end ligation, a time-course analysis was conducted to determine the optimal ligation time for the ligation of re-attaching adaptor. Four time points were evaluated: 0, 12, 24 and 36 hours. On the other hand, the results from Section 4.6 indicated that some non-specific binding did exist after

immunoprecipitation. Therefore, prior to the time-course analysis, two intensive washing procedures to minimize the non-specific binding were examined and compared with the original washing procedure. The washing procedure was placed in the step immediately after RNase treatment. The three washing procedures under evaluation are:

Washing procedure 1) Washing thrice with 1X Detergent-free Lysis Buffer (protease inhibitor-free) (the original washing procedure)

Washing procedure 2) Washing thrice with 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40)

Washing procedure 3) Washing twice with 5X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40) and twice with 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40)

Figure 4.7.1 shows the experimental scheme for studying the ligation of a re-attaching adaptor to the 5’ end of a miRNA target.

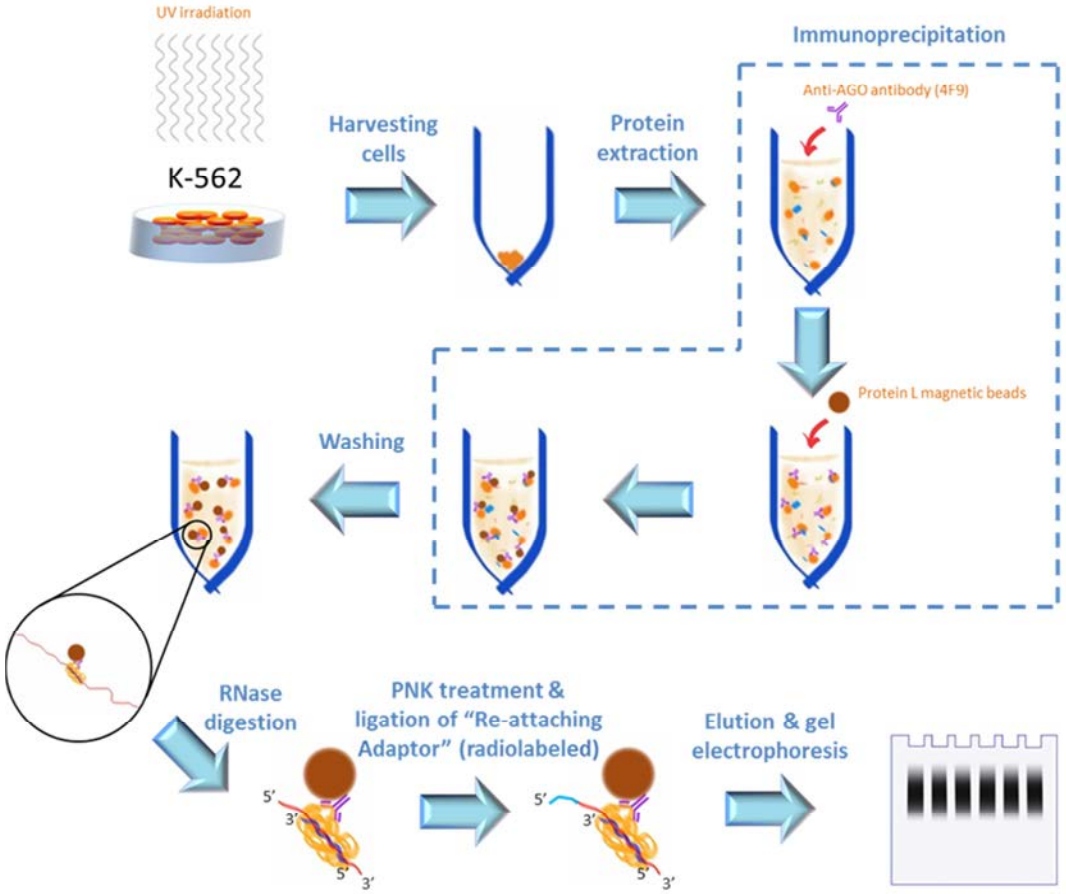


Figure 4.7.1 The experimental scheme for studying the ligation of a re-attaching adaptor to the 5' end of a miRNA target. In order to minimize non-specific binding, two intensive washing procedures were examined. In addition, a time-course study was conducted to determine the optimal ligation time.

4.7.1 Cell culture, UV irradiation, and preparation and quantification of protein lysate

K562 cell extract was used in this experiment. The procedures for cell culture, UV irradiation, and extraction and quantification of the protein lysate were the same as described in Section 4.6.1. In brief, 5×10^7 cells growing in exponential phase were collected. After being washed with 1X HBSS, cells were irradiated under 254 nm UV light, once for 400 mJ/cm^2 and once for 200 mJ/cm^2, on a 4℃ pre-chilled ice pack in UVC 500 Ultraviolet Crosslinker (Amersham). Total protein was extracted in 500 µl 1X Detergent-free Lysis Buffer (Appendix A). The concentration of the total protein lysate was determined by Pierce® BCA Protein Assay Kit (Thermo Scientific) in accordance with the manufacturer's instructions.

4.7.2 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads and the following immunoprecipitation were the same as the Method 2 described in Sections 4.5.2 and 4.5.3 respectively. The input amount of the antibody for each reaction was 11.25 µg (56.25 µl). According to the calculation, the binding capacity of the beads was 16.5-fold higher than the input amount of the antibody. The elution procedure in Section 4.5.3 was omitted and the reactions were continued with the steps in Section 4.7.4. Three reactions were set up for evaluating the 3 different washing procedures, and 4 reactions for investigating of the 4 time points (0, 12, 24 and 36 hours).

4.7.3 Preparation of re-attaching adaptor (radiolabeling)

The re-attaching adaptor, which was named Re-attaching Adaptor (DNA-RNA) (Table 4.2.1), was ordered from IDT. The lyophilized product was dissolved with an appropriate volume of UltraPure™ Distilled Water (Life Technologies) to obtain the Re-attaching Adaptor (DNA-RNA) with a concentration of 100 µM. The adaptor was labeled with [5'^{32}P] cytidine 3',5'-bis(phosphate) (pCp) (PerkinElmer) by T4 RNA Ligase 1, High Concentration (NEB). A final reaction volume of 50 µl was set up for radiolabeling the adaptor. The components in the reaction mixture were as follows: 2.4 µM Re-attaching Adaptor (DNA-RNA), 1X NEBNext® Quick Ligation Reaction Buffer (66mM Tris-HCl, 10mM $MgCl_2$, 1mM DTT, 1mM ATP, 6% Polyethylene glycol (PEG 6000), pH 7.6 at 25°C; NEB), 5 µCi/µl [5'-^{32}P] pCp (Perkin Elmer), 3 U/µl T4 RNA Ligase 1 (NEB), 1 U/µl RNase Inhibitor, Murine (NEB), and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 50 µl (see Appendix C Table 4 for the preparation).

The reaction mixture was incubated at 16℃ overnight (16-20 hours). Then, the ligation reaction was stopped by heat-inactivating T4 RNA Ligase 1 at 65 ℃ for 15 minutes. An additional 0.5 µl of Shrimp Alkaline Phosphatase (USB) was added to the reaction mixture and the mixture incubated at 37℃ for 1 hour. The reaction was stopped by heat inactivation again at 65℃ for 15 minutes. The reaction mixture was then topped up to 100 µl with UltraPure™ Distilled Water (Life Technologies). The final product was purified by illustra™ MicroSpin™ G-25 Columns (GE Healthcare) in accordance with the manufacturer's instructions. The purified product was stored at -20℃ if not used immediately.

4.7.4 RNase treatment, ligation of re-attaching adaptor, and protein gel electrophoresis and visualization

Magnetic beads from Section 4.7.2, which had been washed with 1X Detergent-free Lysis Buffer, were resuspended in 150 µl 1X Detergent-free Lysis Buffer (the same volume as the original bead preparation). RNase digestion was carried out by adding an aliquot of 3 µl, 1:1000 diluted RNase One™ Ribonuclease (Promega) to each reaction. The reactions were incubated on Eppendorf ThermoMixer® C (Eppendorf) at 4℃ for 30 minutes with mixing at 1000 rpm for 15 seconds at 2-minute intervals.

Evaluation of intensive washing procedures

The 3 reactions for evaluating the washing procedures were washed with the same volume (150 µl) of *one* of the following solutions: washing procedure (1) 1X Detergent-free Lysis Buffer (protease inhibitor-free) thrice; washing procedure (2) 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40) thrice; and washing procedure (3) 5X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40) twice plus 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40) twice. Subsequently, the reactions were equilibrated to 1X T4 Polynucleotide Kinase Reaction Buffer (NEB) by 2 additional washes with 150 µl 1X Polynucleotide Kinase Reaction Buffer. The polynucleotide kinase (PNK) treatment mixture for each reaction had a volume of 150 µl and contained 1X T4 Polynucleotide Kinase Reaction Buffer (70 mM Tris-HCL, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6 at 25℃; NEB), 1 mM ATP, 0.5 U/µl T4 Polynucleotide Kinase, 3’ phosphatase minus (NEB), 1 U/µl RNase Inhibitor, Murine (NEB) and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 150 µl (see Appendix C Table 5 for the preparation).

Evaluation of incubation time

Each of the 4 reactions for the time-course analysis was washed twice with 150 µl 5X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40), and twice with 150 µl 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40). Subsequently, the reactions were equilibrated to 1X T4 Polynucleotide Kinase Reaction Buffer (NEB) by 2 more washes with 150 µl 1X T4 Polynucleotide Kinase Reaction Buffer. The PNK treatment mixture was the same as that described in the last paragraph.

Each of the reactions was resuspended in 150 µl of the above PNK treatment mixture (same volume as the original bead preparation). The reactions were incubated on Eppendorf ThermoMixer® C at 16℃ for 1 hour with mixing at 1000 rpm for 15 seconds at 2-minute intervals.

Evaluation of intensive washing procedures

The Reaction 1 for evaluating the washing procedures was washed twice with 0.5 ml of 1X Detergent-free Lysis Buffer (protease inhibitor-free) after the PNK treatment. The Reactions 2 and 3 for evaluating the washing procedures were washed twice with 0.5 ml of 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40) after the PNK treatment. Next, all the 3 reactions were equilibrated to 1X T4 RNA Ligase Reaction Buffer (without DTT) by 2 additional washes with 0.5 ml of 1X T4 RNA Ligase Reaction Buffer (without DTT) and were processed for the ligation of radiolabeled re-attaching adaptor.

Evaluation of incubation time

The 4 reactions for the time-course analysis were washed twice with 0.5 ml of 1X Detergent-free Lysis Buffer (protease inhibitor-free, 0.5% NP40) after the PNK treatment. Next, all the 4 reactions were equilibrated to 1X T4 RNA Ligase Reaction Buffer (without

DTT) by 2 more washes with 0.5 ml of 1X T4 RNA Ligase Reaction Buffer (without DTT). During the last wash, individual reactions in the same set of experiment were pooled together and re-divided again to eliminate the variations from the above preparation procedures. The 4 reactions were processed for the ligation of radiolabeled re-attaching adaptor.

For both analyses, 480 nM of re-attaching adaptor and 1.5 U/µl of T4 RNA Ligase 1 were chosen for the experiment. The reaction mixture for the ligation of the radiolabeled re-attaching adaptor for each reaction had a volume of 100 µl and consisted of 1X NEBNext® Quick Ligation Reaction Buffer (NEB), 1.5 U/µl T4 RNA Ligase 1 (ssRNA Ligase, NEB), 120 nM *radiolabeled* Re-attaching Adaptor (DNA-RNA), 840 nM *unlabeled* Re-attaching Adaptor (DNA-RNA), 1 U/µl RNase Inhibitor, Murine (NEB), and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 100 µl (see Appendix C Table 6 for the preparation).

Each reaction of the beads (from Section 4.7.2) was resuspended in 100 µl of the above ligation mixture for the radiolabeled re-attaching adaptor, and incubated on Eppendorf ThermoMixer® C at 16℃ with mixing at 1350 rpm for 15 seconds at 2-minute intervals. The 3 reactions for evaluating the intensive washing procedures were collected after 36 hours of incubation. Meanwhile, the 4 reactions for the time-course analysis were collected after 0, 12, 24 and 36 hours of incubation. The collected beads were washed 5 times with 0.5 ml of 1X Detergent-free Lysis Buffer (Proteinase Inhibitor-free) immediately after the ligation reaction. Elution was carried out by adding 25 µl of NuPAGE® LSD Sample Buffer (4X) (Thermo Fisher) and 75 µl UltraPure™ Distilled Water (Thermo Fisher) to each vial of the washed beads and by heating the vial for 10 minutes at 70℃. For each sample, a portion (50 µl) of the eluate was loaded onto NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels, 1.0

mm, 10-well (Thermo Fisher) and the gel run under 175V at 4℃ for 90 minutes in 1X NuPAGE® MOPS SDS Running Buffer (Thermo Fisher). Lastly, the gel was wrapped by a Saran Wrap for avoiding any leakage and then exposed to an X-ray film for 24 hours. The film was developed by the film processor FUJIFILM FPM 2800 (FujiFilm). The signal intensities of the bands were analyzed by the gel electrophoresis image analysis software GelAnalyzer 2010a (http://www.gelanalyzer.com/).

4.8 RNA fragmentation

In most of the cases, the targets of miRNAs are very long RNA sequences such as mRNAs and long non-coding RNAs. When miRNA-Argonaute complex recognizes and attaches to these long sequences, two long flanking sequences, which are the 5' and the 3' regions of the target RNA strand, can be found on the final complex. It looks like two threads coming out from the final complex (Figure 4.8.1). These long flanking strands make it very difficult, if not impossible, to join the interacting miRNA and target RNA together by the methods used in RIINGS owing to the large gap between the two interacting sequences. Even if the long-spanning gap can be sealed, the length of the chimeric sequence is too long for sequencing. In addition, the chimeric sequences must not be fragmented before sequencing because this action will counteract all the efforts of directly linking the two interacting RNAs together. A very straightforward solution to this problem is to remove the long flanking sequences from the complexes, but retain both ends that are long enough for the ligation of adaptors. Notably, this strategy makes the gaps easier to seal, narrows down the length of the chimeric sequences to be sequenced, and most importantly targets the

sequenced region of the miRNA targets to the genuine binding sites of miRNA-Argonaute complexes.

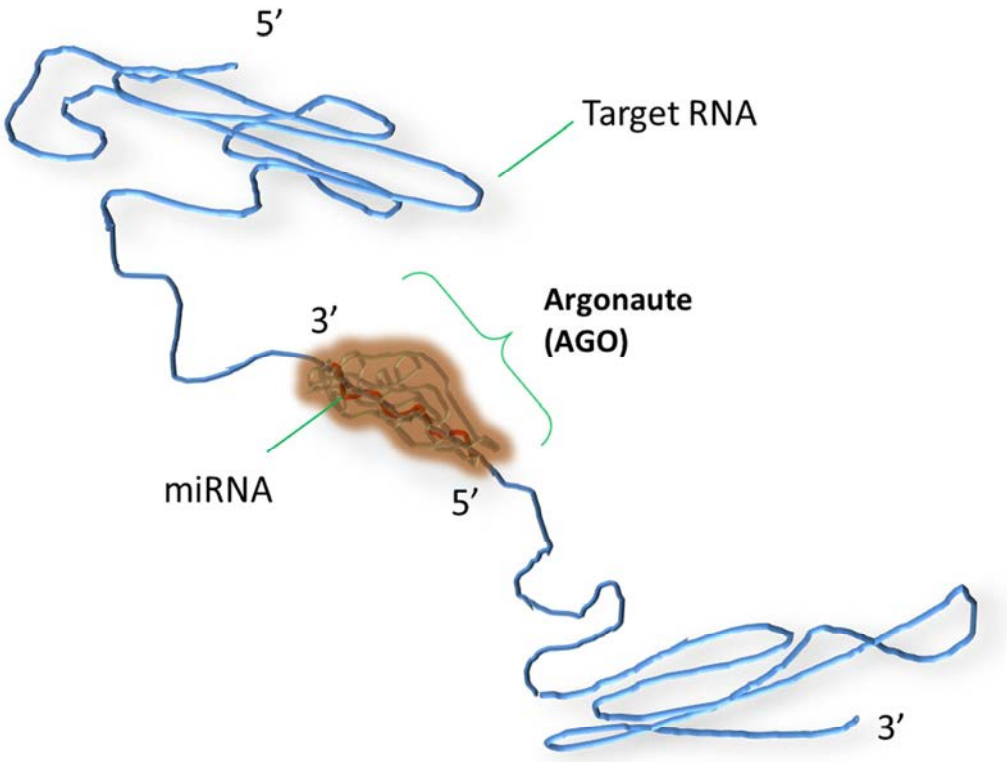


Figure 4.8.1 The microRNA-target RNA-Argonaute complex.

RNase A and RNase I are the candidate RNases selected for trimming the flanking RNA strands from the miRNA-target RNA-Argonaute complexes. RNase A is the endoribonuclease that was originally used in HITS-CLIP developed by Darnell's group [134]; however, this RNase preferentially cleaves the phosphodiester bond of RNA at the 3' end of pyrimidine residues (C and U) [310]. On the other hand, RNase I is an endoribonuclease that randomly cleaves RNA sequences endogenously without any sequence preference [311]. The completeness of RNA digestion by RNase depends on the enzyme concentration, the incubation time and the incubation temperature. In this experiment, 5 different dilutions were tested on both RNases: undiluted, 1/100 X, 1/1000 X, and 1/100000 X. The incubation temperature was fixed at 4℃ and the incubation time 30 minutes.

Agilent 2100 Bioanalyzer (Agilent) was used to determine the fragment size after digestion. In addition, protein gel electrophoresis with autoradiography to visualize the ligation of the radiolabeled adaptor was used to check whether the trimmed flanking RNA regions were still long enough for ligation of adaptors. Figure 4.8.2 shows the experimental scheme for determining the working concentration of RNase.

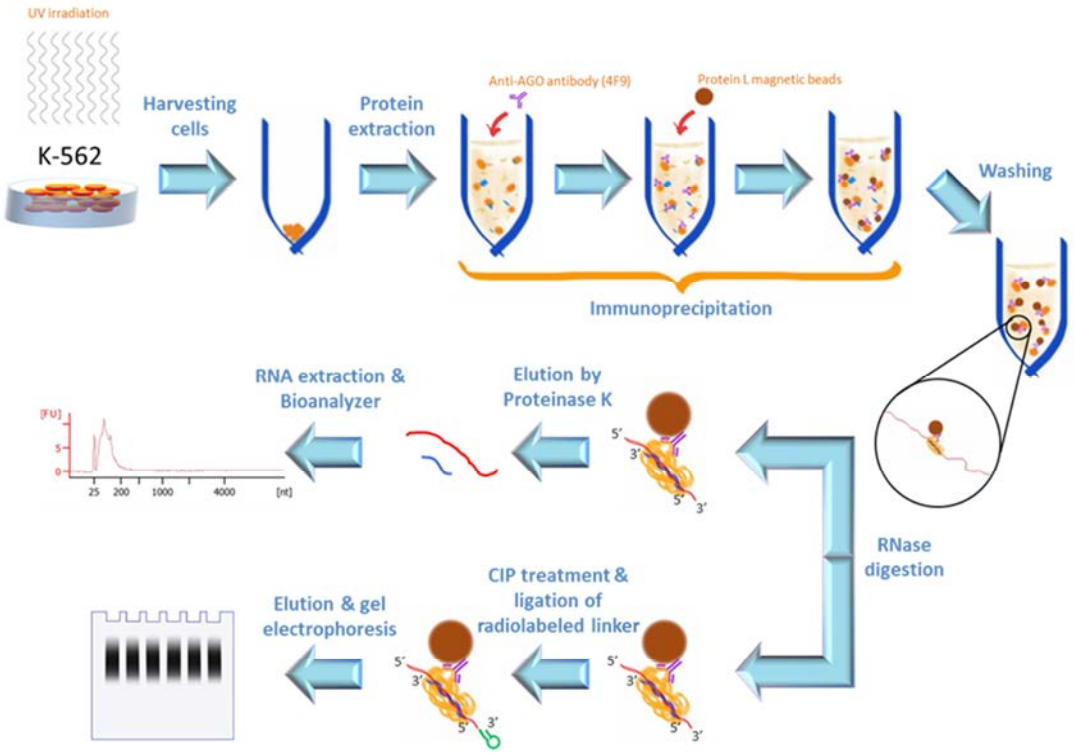


Figure 4.8.2 The experimental scheme for determining the working concentration of RNase. Titration analyses were conducted to determine the optimal concentrations of RNase A and RNase I. The fragment size was assessed by Agilent 2100 Bioanalyzer while the ligation of adaptor was assessed by protein gel electrophoresis.

4.8.1 Cell culture, UV irradiation, and preparation and quantification of protein lysate

K562 cell extract was used in this experiment. The procedures for cell culture, UV irradiation, and extraction and quantification of the protein lysate were the same as described in Section 4.6.1. In brief, 5×10^7 cells growing in exponential phase were collected. After being washed with 1X HBSS, cells were irradiated under 254 nm UV light, once for 400 mJ/cm^2 and once for 200 mJ/cm^2, on a 4℃ pre-chilled ice pack in UVC 500 Ultraviolet Crosslinker (Amersham). Total protein was extracted in 500 µl 1X Detergent-free Lysis Buffer (Appendix A). The concentration of the total protein lysate was determined by Pierce® BCA Protein Assay Kit (Thermo Scientific) in accordance with the manufacturer's instructions.

4.8.2 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads and the subsequent immunoprecipitation were the same as the Method 2 described in Sections 4.5.2 and 4.5.3 respectively. The input amount of the antibody for each reaction was 11.25 µg (56.25 µl). According to the calculation, the binding capacity of the beads was 16.5-fold higher than the input amount of the antibody. The elution procedure in Section 4.5.3 was omitted and the reactions were continued with the steps in Section 4.8.3. There were 5 reactions for evaluating fragment size and 4 reactions for testing the ligation of adaptor

4.8.3 RNase treatment

Both RNase A (Affymetrix) and RNase One™ Ribonuclease (Promega) were diluted to 1/100 X, 1/1000 X and, 1/100000 X with 1X Detergent-free Lysis Buffer. The original concentration

of the RNase A was 38.46 U/µl (Lot. 73715) while the original concentration of the RNase One™ Ribonuclease was 10 U/µl. Notably, the unit definition was not the same for these two RNases. All the vials of magnetic beads from Section 4.8.2, which had been washed with 1X Detergent-free Lysis Buffer, were resuspended in 150 µl 1X Detergent-free Lysis Buffer (same volume as the original bead preparation). All the reactions within a particular set of experiment were pooled together and the pooled mixture was re-divided again to eliminate the variation arising from the above preparation procedures. The RNase digestion for evaluating fragment size was carried out by adding an aliquot of 3 µl of undiluted, 1/100 X, 1/1000 X, 1/100000 X and 1/10000000 X diluted RNase A (Affymetrix) to the 5 reactions respectively while the RNase digestion for testing the ligation of adaptor was carried out by adding an aliquot of 3 µl of undiluted, 1/100 X, 1/1000 X and 1/100000 X diluted RNase A (Affymetrix) to the 4 reactions respectively. The same was repeated for RNase One™ Ribonuclease (Promega). All the reactions were incubated on Eppendorf ThermoMixer® C (Eppendorf) at 4°C for 30 minutes with mixing at 1000 rpm for 15 seconds at 2-minute intervals.

4.8.4 RNA extraction and evaluation of fragment size

After the RNase treatment, the reactions for evaluating fragment size were washed thrice with 150 µl 1X Detergent-free Lysis Buffer (protease inhibitor-free) (same volume as the original bead preparation). The washed beads were then resuspended in 200 µl of 1X Detergent-free Lysis Buffer (protease inhibitor-free). The elution was carried out by adding 10 µl of Proteinase K (Molecular Biology Grade, NEB) to the resuspended beads. The beads were then incubated at 56°C for 2 hours with mixing at 1000 rpm for 15 seconds at 2-minute intervals. Beneficially, all proteins associated with RNAs were also removed by this elution method. After the incubation, the beads were discarded and the RNA-containing

supernatant was saved for extraction. Subsequently, The RNA was extracted by miRNeasy Mini Kit (Qiagen) in accordance with the manufacturer's instructions. Then, RNA concentration was measured by NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies) and stored at -70°C until use.

The fragment size was analyzed using Agilent 2100 Bioanalyzer (Agilent) with Agilent RNA 6000 Pico Kit (Agilent). For each samples, around 2.5 ng of RNA was loaded for initialize the run. Electrophoresis was carried out in accordance with the manufacturer's instructions.

4.8.5 Ligation of adaptor, and protein gel electrophoresis and visualization

After the RNase treatment, the reactions for testing the ligation of adaptor were washed thrice 150 µl 1X Detergent-free Lysis Buffer (same volume as the original bead preparation). The adaptor used in this experiment and the radiolabeling procedures were the same as described in Section 4.6.3. The ligation of adaptor, protein gel electrophoresis and visualization followed the steps Section in 4.6.4, starting from the equilibration of the washed beads with 1X NEBuffer 1 (NEB). The RNase treatment and the subsequent washing steps were skipped. The concentration of the radiolabeled adaptor was 480 nM. The concentration of T4 RNA Ligase 1 (NEB) in the ligation master mix was 2.25 U/µl. The ligation time was 36 hours.

4.9 UV crosslinking between RNA and protein

In principle, RIINGS is designed for the identification of protein-mediated RNA-RNA interactions with exact pairwise correspondence *and* on a genomic scale. This method

requires immunoprecipitation to purify the ribonucleoprotein (RNP) complex of interest, ligation to link up the interacting RNA pairs, and finally high-throughput sequencing to sequence all the chimeric sequences. Therefore, stabilizing the RNP is critical to RIINGS. In order to prevent the disassembling of RNP or *in vitro* rearrangement after RNP extraction from cells, covalently crosslinking the interacting RNAs and RNA-binding protein inside the cells of interest (before extraction) is the commonly used technique. MicroRNA-target RNA interactions are mediated by RISC. The core protein of the RISC is Argonaute protein 1-4 for one of them. Argonaute-RNAs complexes are the target RNP when the miRNA-target RNA interactions are to be studied. Therefore, crosslinking of the interacting Argonaute-RNAs complexes is necessary.

Two crosslinking methods were separately used in HITS-CLIP and PAR-CLIP [134, 295]. These two crosslinking methods are the key methods utilized in RNA-protein crosslinking. The crosslinking method used in HITS-CLIP is achieved by exposing the living cells under ultraviolet light (UV light) of 254 nm. The RNAs and proteins in close proximity would be crosslinked together covalently under the 254 nm UV light. On the other hand, the crosslinking method used in PAR-CLIP is a chemical-assisted UV-light crosslinking method. The chemical used in PAR-CLIP crosslinking is 4-thiouridine, which is an analogue of uridine and a photoactivatable-ribonucleoside. The compound 4-thiouridine can be utilized by cells to synthesize RNAs. Thus, the RNAs would contain 4-thiouridine. Under the induction by 365 nm UV light, the 4-thioridine is efficiently crosslinked to the interacting proteins. Both approaches were tested in the current experiment. Figure 4.9.1 shows the experimental scheme for determining the optimal crosslinking condition.

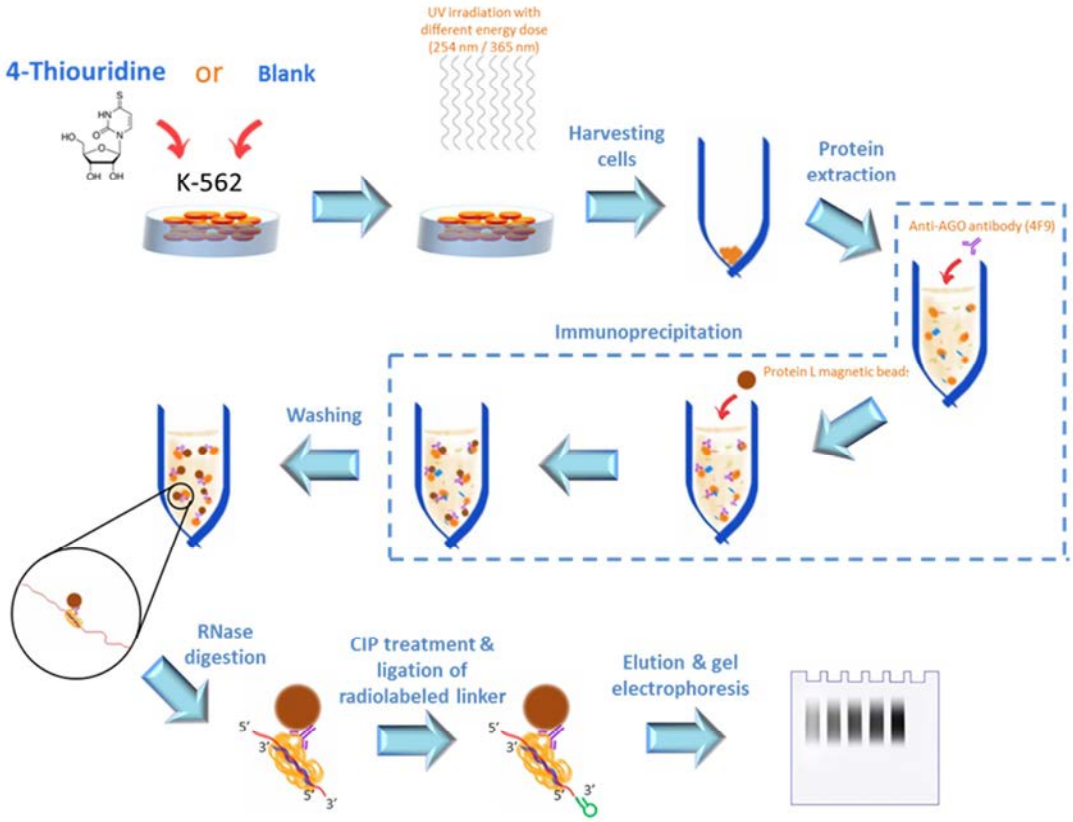


Figure 4.9.1 The experimental scheme for determining the optimal crosslinking condition. Both non-chemical assisted- and 4-thiouridine-assisted UV crosslinking methods were tested. The non-chemical-assisted method used 254 nm UV light for induction and an irradiation dose of 150 mJ/cm^2. The 4-thiouridine-assisted method used 365 nm UV light for induction and an irradiation dose of 600 mJ/cm^2.

4.9.1 Cell culture and 4-thiouridine treatment

K562 cell extract was used in this experiment. Both the non-chemical-assisted and 4-thiouridine-assisted UV crosslinking methods were tested in this experiment. Cells were seeded in four T-175 culture flasks with 43 ml cell suspension per flask at 5×10^5 cells/ml. Two flasks of cells were set up for the non-chemical-assisted UV crosslinking while the remaining two were set up for the 4-thiouridine-assisted UV crosslinking. The four flasks of cells were incubated at 37℃ for 48 hours in an incubator with 5% CO_2. An aliquot of 4.3 µl of 1 M 4-thiouridine in DMSO was added to the 2 flasks of cells for the 4-thiouridine-assisted UV crosslinking experiment to obtain a final 4-thiouridine concentration of 100 µM in the cell suspension. Similarly, an aliquot of 4.3 µl of DMSO was added to the 2 flasks of cells for the non-chemical-assisted UV crosslinking experiment. All 4 flasks of cells were incubated at 37℃ with 5% CO_2 for an additional 14 hours.

4.9.2 UV irradiation, and preparation and quantification of protein lysate

Four flasks of cells were used in this preparation. Two were 4-thiouridine-treated and two were control (treated with comparable volume of DMSO only). Each flask of cells was collected by centrifugation at 125 *g* (800-1000 rpm) for 5 minutes in 4 individual tubes. The supernatant in all the tubes were discarded. Each cell pellet was washed twice with 20 ml ice-cold 1X HBSS (Thermo Scientific). Cells with the same treatment were pooled together during the last wash for the treatment of crosslinking-assisting agent. After the pooling, the cells were pelleted again and the two cell pellets with different treatment of crosslinking-assisting agent were resuspended in 8 ml of ice-cold 1X HBSS individually. Each of the above suspensions was then transferred to four 10 cm dishes with 2 ml of suspension for each dish. The cell suspension was evenly spread out on each dish by gently shaking by hand. In other words, there were 4 dishes of cells for each kind of treatments (4 dishes for

4-thiouridine treated and 4 dishes for control). All dishes of cells were irradiated with 254 nm or 365 nm UV light for a certain dose (150 mJ/cm^2 or a combination of 400 and 200 mJ/cm^2) on a 4℃ pre-chilled ice pack in a UV crosslinker (UVC 500 Ultraviolet Crosslinker [Amersham] for 254 nm UV light, or CL-1000 Ultraviolet Crosslinker [UVP] for 365 nm UV light). Specifically, to both sets of the 4 dishes, 2 dishes of cells (one is 4-thiouridine treated and one is control) were irradiated under 254 nm UV light for 150 mJ/cm^2; 2 dishes of cells (one is 4-thiouridine treated and one is control) were irradiated twice under 254 nm UV light, once for 400 mJ/cm^2 and once for 200 mJ/cm^2; 2 dishes of cells (one is 4-thiouridine treated and one is control) were irradiated under 365 nm UV light for 150 mJ/cm^2; and 2 dishes of cells (one is 4-thiouridine treated and one is control) were irradiated twice under 365 nm UV light, once for 400 mJ/cm^2 and once for 200 mJ/cm^2. There was a 1-minute interval between the 2 irradiations for the dishes requiring a second time irradiation. The suspension was well mixed before each irradiation. Therefore, there were 8 different treatments of cells (Figure 4.9.2). Total protein was extracted in 1X Detergent-free Lysis Buffer (Appendix A). Protein extraction and quantification of the protein lysate were the same as described in Sections 4.3.2 and 4.3.3 respectively.

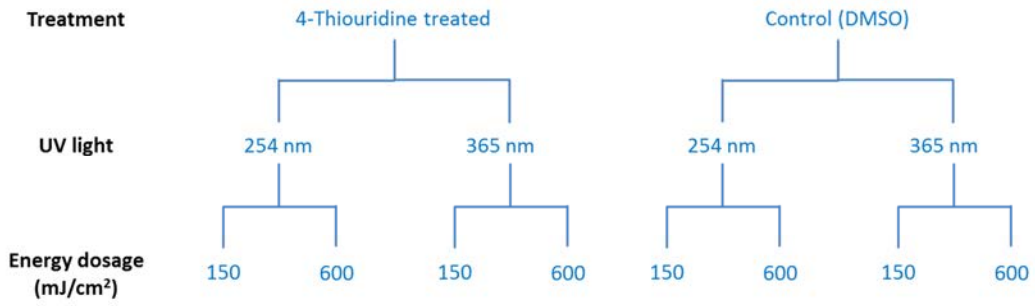


Figure 4.9.2 Summary of sample treatments for optimizing crosslinking condition.

4.9.3 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads and the subsequent immunoprecipitation was the same as the Method 2 described in sections 4.5.2 and 4.5.3 respectively. The amount of the antibody input for each reaction was 11.25 µg (56.25 µl). According to the calculation, the binding capacity of the beads was 16.5-fold higher than the amount of the antibody. The elution procedure in 4.5.3 was omitted and the reactions were continued with the steps in Section 4.9.4.

4.9.4 RNase treatment, ligation of stem-loop adaptor, and protein gel electrophoresis and visualization

The procedures for RNase treatment were the same as described in Section 4.6.4. Briefly, an aliquot of 3 µl, 1:1000 diluted RNase One™ Ribonuclease (Promega) was added to the beads for the RNase digestion at 4°C for 30 minutes. The adaptor used in this experiment and the radiolabeling procedures were the same as described in Section 4.6.3. The ligation of adaptor, and protein gel electrophoresis and visualization were carried out as described in Section 4.6.4, starting from the equilibration of the washed beads with 1X NEBuffer 1 (NEB). The concentration of the radiolabeled adaptor was 480 nM. The concentration of T4 RNA Ligase 1 (NEB) in the ligation master mix was 2.25 U/µl. The ligation time was 36 hours.

4.10 Conversion of 2',3'-cyclic monophosphate end

This is an important step that can drastically enhance the ligation of the stem-loop adaptor to the 3’ end of the target RNAs. In the RIINGS method, RNase fragmentation is required. Both RNase A and RNase I were investigated in the current study. These two RNases are

endogenous ribonuclease that could potentially be used in RIINGS. They catalyze the cleavage of RNA dinucleotide bonds endogenously, and leave a 5′ hydroxyl end and 2′,3′-cyclic monophosphate ends on the product RNAs or on the product monoribonucleotides. The 2′,3′-cyclic monophosphate end is an intermediate product. This end is further converted by the RNases to the corresponding 3' monophosphate via hydrolysis; however, the hydrolysis is very slow (Figure 4.10.1) [312, 313]. As the result, the cleaved RNAs mainly have 5′ hydroxyl end and 2′,3′-cyclic monophosphate ends at their two termini.

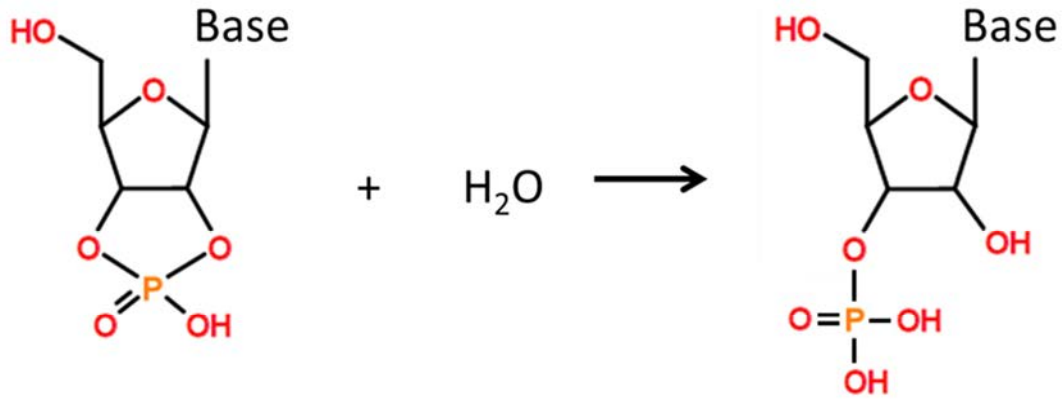


Figure 4.10.1 The hydrolysis of 2',3'-cyclic monophosphate nucleotide to 3' monophosphate ribonucleotide.

Notably, all these ends are not suitable for ligation because they are not the substrates of ligases [314]. These ends should be converted to the corresponding 5’ monophosphate end and 3’ hydroxyl end before ligation. T4 Polynucleotide kinase (T4 PNK) is a robust enzyme that can convert the 5’ hydroxyl end to 5’ monophosphate end via its phosphorylation activity. On the other hand, the 3’ monophosphate end can be converted to the 3’ hydroxyl end rapidly by many of the phosphatases. Examples for these phosphatases are Antarctic phosphatase (AP), shrimp alkaline phosphatase (SAP) and calf intestinal alkaline phosphatase (CIP). Nevertheless, these phosphatases are not able to remove the phosphate group from the 2',3'-cyclic monophosphate end.

In fact, T4 PNK has two enzymatic activities: addition of 5’ phosphate to oligonucleotides (5’ kinase activity) and removal of 3’ phosphoryl groups (3’ phosphatase activity) [315]. Besides, T4 PNK is able to convert the 2',3'-cyclic monophosphate end to 3’ hydroxyl end via either one of two possible catalytic processes suggested by Das et. al. in 2013 [315]. The ordinary phosphatase activity and 2',3'-cyclic monophosphate hydrolytic activities of the T4 PNK are illustrated in Figure 4.10.2. Therefore, T4 PNK is the enzyme particularly suitable for converting the 2',3'-cyclic monophosphate end to 3’ hydroxyl end that can be used in the subsequent ligation. The present experiment tested whether the additional PNK treatment could improve the efficiency of ligating the stem-loop adaptor via the conversion of both 3’ monophosphate end and 2',3'-cyclic monophosphate end to 3’ hydroxyl end. This evaluation was performed using both RNaseA-fragmented RNA and RNaseI-fragmented RNA. Figure 4.10.3 shows the experimental scheme for investigating whether additional T4 PNK treatment could improve the efficiency of ligating stem-loop adaptor ligation.

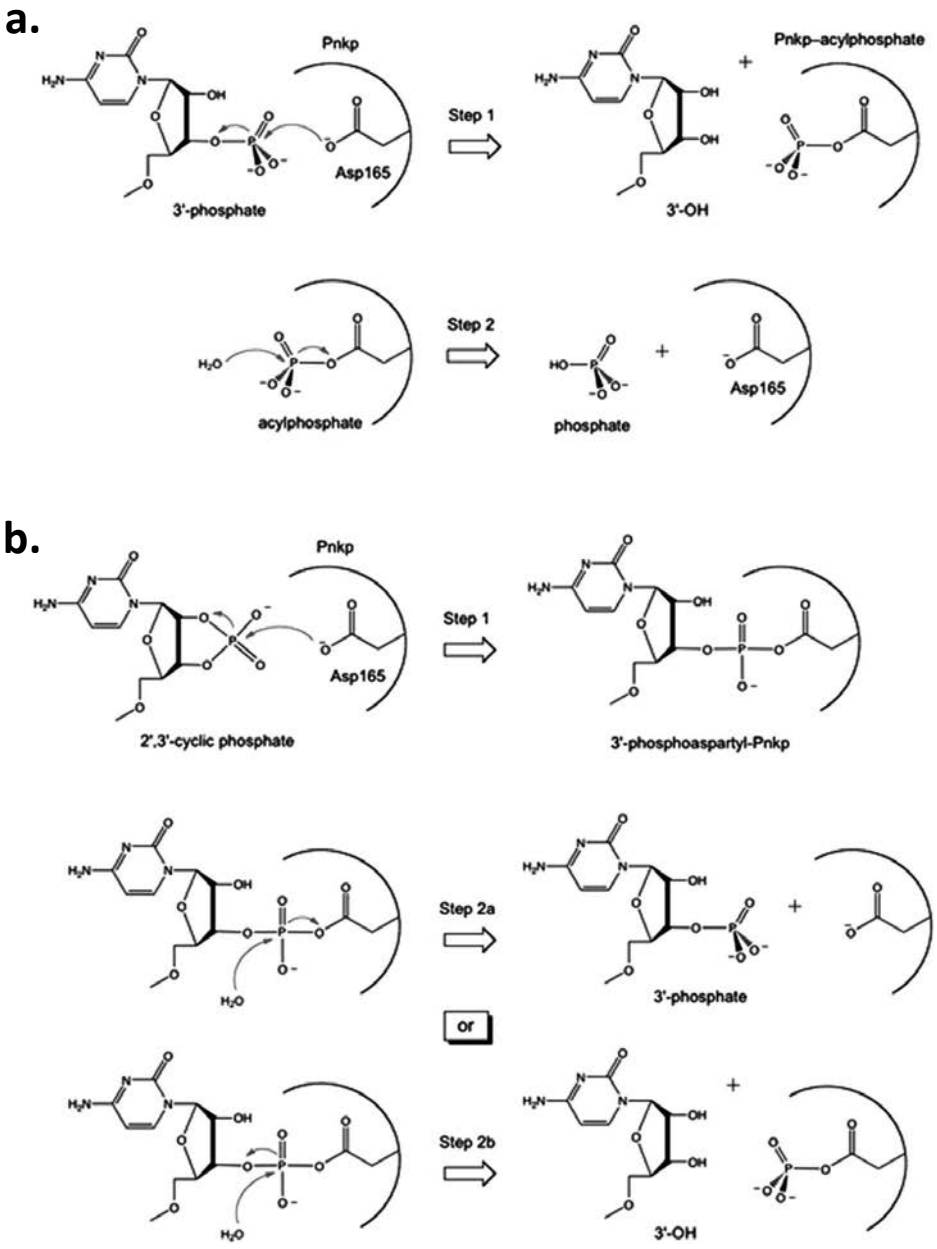


Figure 4.10.2 Pathways of RNA 3′ end-healing by T4 polynucleotide kinase. (a) A 2-step mechanism of removing 3′-phosphate via the (Step 1) formation and (Step 2) hydrolysis of a covalent phosphoaspartyl-enzyme (P-Asp-Enz) intermediate. **(b)** Presented here are two possible pathways for removing a 2′,3′-cyclic phosphate. Involved in a common 1st step is the attack by the aspartate nucleophile on the 2',3'-cyclic monophosphate end of the RNA to (Step 1) form and then (Step 2) hydrolyze a covalent RNA-(P-Asp-Enz) intermediate. In the Step 1 reaction presented, the O2′ of the ribose is the leaving group while the 3′-phosphate of the RNA is linked to the enzyme. Two versions of the hydrolytic reaction (Steps 2a or 2b) are shown here. In Step 2a, the enzymatic aspartate is the leaving group while the 3′ monophosphate end of the RNA is the product that can enter the reaction described in (a) to be subsequently converted to 3′ hydroxyl end. In Step 2b, the O3′ of the ribose is the leaving group and a phosphoaspartyl-enzyme remains. **(Adopted from Nucleic Acids Research. 2013; 41: 355-365)**

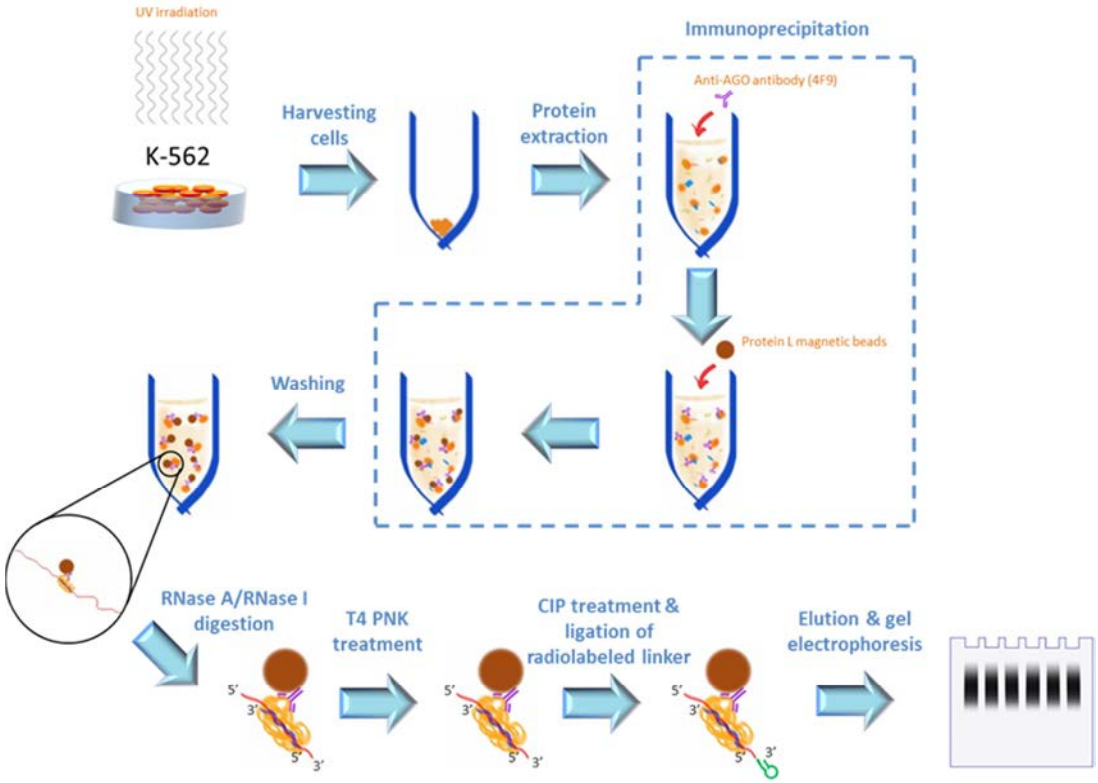


Figure 4.10.3 The experimental scheme for investigating whether additional T4 PNK treatment could improve the efficiency of ligating stem-loop adaptor ligation. The franking/protruding RNAs on the Argonaute complex were trimmed by either RNase I or RNase A. Both the 3' ends generated by these two RNase were investigated. T4 PNK treatment was additionally added to the procedure at the point that immediately after the RNase digestion.

4.10.1 Cell culture, UV irradiation, and preparation and quantification of protein lysate

K562 cell extract was used in this experiment. The procedures for cell culture, UV irradiation, and extraction and quantification of the protein lysate were the same as described in Section 4.6.1. In brief, cells were irradiated under 254 nm UV light 2 times, once for 400 mJ/cm^2 and once for 200 mJ/cm^2, on a 4℃ pre-chilled ice pack in UVC 500 Ultraviolet Crosslinker (Amersham). Total protein was extracted in 1X Detergent-free Lysis Buffer (Appendix A).

4.10.2 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads preparation and the succeeding immunoprecipitation were the same as the Method 2 described in Sections 4.5.2 and 4.5.3 respectively. The amount of the antibody input for each reaction was 11.25 µg (56.25 µl). According to the calculation, the binding capacity of the beads was 16.5-fold higher than the amount of the antibody. The elution procedure of Section 4.5.3 was omitted and the reactions were continued with the steps in Section 4.10.3. There were 2 reactions (control and with an additional T4 PNK treatment) for investigating the possible improvement in the ligation of stem-loop adaptor with an additional T4 PNK treatment after RNase I fragmentation. Similarly, there were also 2 such reactions (control and with an additional T4 PNK treatment) after RNase A fragmentation.

4.10.3 RNase treatment

The procedures for RNase treatment followed those described in Section 4.6.4 with some differences (see below). For the 2 reactions used to investigate the RNA fragmentation by

using RNase I, an aliquot of 3 µl, 1:1000 diluted RNase One™ Ribonuclease (Promega) was added to each vial of the beads for digestion at 4℃ for 30 minutes. For the other 2 reactions used to investigate the RNA fragmentation by using RNase A, an aliquot of 3 µl, 1:1000 diluted RNase A (Affymetrix) was added to each vial of the beads for digestion at 4℃ for 30 minutes.

4.10.4 T4 PNK treatment

The reactions were washed thrice with 150 µl 1X Detergent-free Lysis Buffer (same volume as the original bead preparation) after the RNase treatment. Then, the beads were equilibrated to 1X PNK Buffer (without DTT) by 2 additional washes with 150 µl 1X PNK Buffer (without DTT). The T4 PNK treatment mixture for each 150-µl reaction had the following components: 1X Polynucleotide Kinase Reaction Buffer (70 mM Tris-HCL, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6 at 25℃; NEB), 1 U/µl T4 Polynucleotide Kinase (NEB), 1 U/µl RNase Inhibitor, Murine (NEB), and UltraPure™ Distilled Water (Life Technologies) to make up the volume to 150 µl (see Appendix C Table 7 for the preparation).

The beads for the experiment were resuspended in 150 µl T4 PNK treatment mixture (see above; same volume as the original bead preparation). The beads for the control were resuspended in 150 µl of the above T4 PNK treatment mixture *without* T4 Polynucleotide Kinase. The reaction mixtures were incubated on Eppendorf ThermoMixer® C at 16℃ for 30 minutes with mixing at 1000 rpm for 15 seconds at 2-minute intervals. The beads were washed once with 150 µl 1X Detergent-free Lysis Buffer after T4 PNK treatment. Next, the beads were equilibrated to 1X NEBuffer 1 (NEB) by 2 additional washes with 150 µl 1X NEBuffer 1 (10mM Bis-Tris-Propane-HCl, 10mM $MgCl_2$, 1mM DTT, pH 7.0 at 25℃; NEB) and then processed for the steps in Section 4.10.5.

4.10.5 Ligation of stem-loop adaptor, and protein gel electrophoresis and visualization

The adaptor used in this experiment and the radiolabeling procedures were the same as described in Section 4.6.3. The ligation of adaptor, and protein gel electrophoresis and visualization followed those in Section 4.6.4 with the procedures starting from the CIP treatment. The concentration of the radiolabeled adaptor was 480 nM. The concentration of T4 RNA Ligase 1 (NEB) in the ligation master mix was 2.25 U/µl. The ligation time was 36 hours.

4.11 Analysis of AGO complexes isolated by immunoprecipitation

Obtaining stabilized Argonaute complexes of high purity, in which each complex has miRNA, target RNA and Argonaute protein, is one of the most important parts and one of the core techniques in RIINGS. The complexes, particularly the RNA duplex inside the Argonaute protein, must be in their native forms because the RNA duplex cannot be ligated or the level of RNA-duplex *in vitro* dissociation/reassociation can be elevated if the RNA duplex is denatured (the two RNA sequences separated from each other).

There are three important questions to address. First, does the isolated Argonaute complexes pure enough for identifying the Argonaute-mediated miRNA-target RNA interactions? Second, are the miRNA-target RNA duplexes stable enough to get through the processes of physically joining the two individual sequences together (what is the melting temperature of the duplexes)? Third, does the quantity of the harvested duplexes good enough for representing the profile of the entire interactome? All these questions are

equally important. Accordingly, three sets of experiment were designed to examine these issues.

4.11.1 Assessment of AGO complexes for purity by mass-spectrometry

Prior to evaluating the purity of Argonaute complexes by mass spectrometry, the purity had been roughly assessed by silver staining of SDS-PAGE as described in Section 4.5.5. However, the silver staining method is not capable of identifying the proteins and not sensitive enough to visualize some proteins present in very trace amounts. In order to further assess the purity of the purified Argonaute complexes and identify protein contaminants if any, TripleTOF™ 6600 (AB SCIEX) was used for these analyses. This mass spectrometer is equipped with NANOSpray® III Source (AB SCIEX) and ekspert™ nanoLC 415 (eksigent®), and thus combines liquid chromatography with quadrupole/time-of-flight (QqTOF) hybrid mass spectrometry with ion source provided by electrospray ionization (i.e. a Nano LC-ESI-Q-TripleTOF-MS). Figure 4.11.1 shows the experimental scheme for assessing the purity of purified Argonaute complexes.

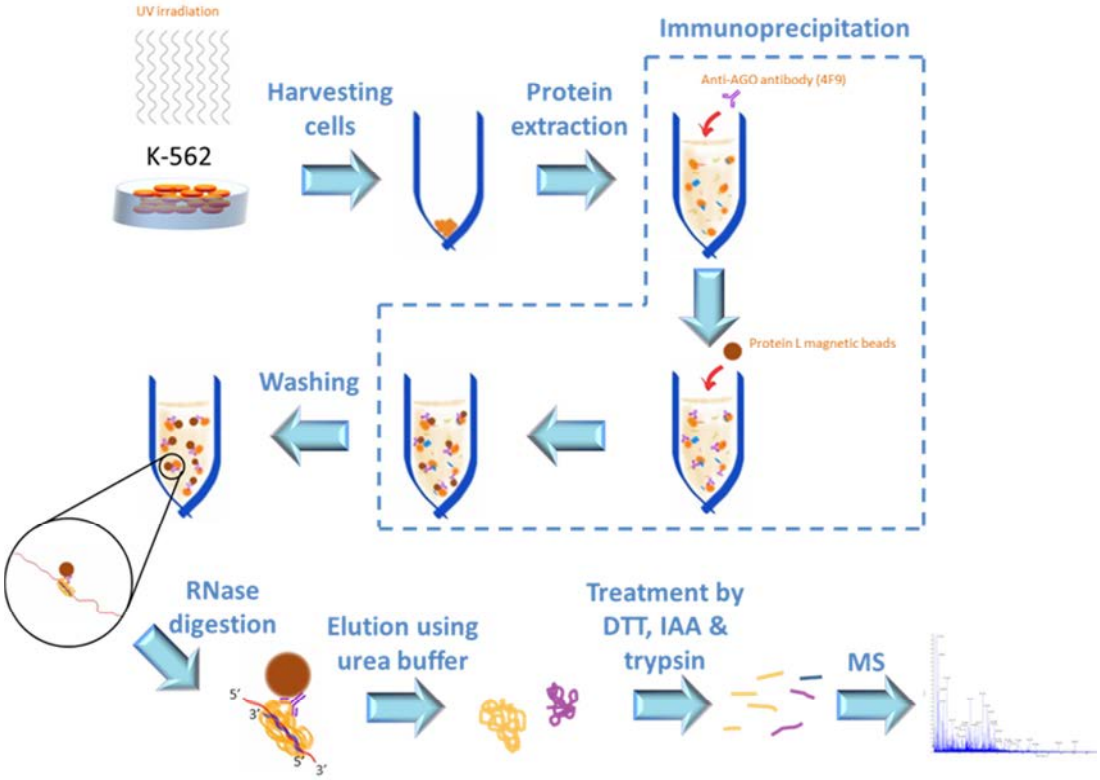


Figure 4.11.1 The experimental scheme for assessing the purity of purified Argonaute complexes. The purified proteins from a preparation of AGO complexes (immunoprecipitation of AGO proteins in RIINGS) were eluted under a denaturing condition to make sure all the captured proteins could be eluted from the magnetic beads. Then, the eluate went through ordinary pretreatments for analysis by mass spectrometry. Mass spectrometry analysis was carried out on TripleTOF™ 6600 (AB SCIEX), which is equipped with NANOSpray® III Source (AB SCIEX) and ekspert™ nanoLC 415 (eksigent®).

4.11.1.1 Cell culture, UV irradiation, and preparation and quantification of protein lysate

K562 cell extract was used in this experiment. The procedures for cell culture, UV irradiation, and protein lysate preparation and quantification were the same as described in Section 4.6.1. In brief, cells were irradiated under 254 nm UV light 2 times, once for 400 mJ/cm^2 and once for 200 mJ/cm^2, on a 4℃ pre-chilled ice pack in UVC 500 Ultraviolet Crosslinker (Amersham). Total protein was extracted in 1X Detergent-free Lysis Buffer (Appendix A).

4.11.1.2 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads and the ensuing immunoprecipitation were the same as the Method 2 described in Sections 4.5.2 and 4.5.3 respectively. The amount of the antibody input for each reaction was 11.25 µg (56.25 µl). According to the calculation, the binding capacity of the beads was 16.5-fold higher than the amount of the antibody. The elution procedure in Section 4.5.3 was omitted and the reactions were continued with the steps in Section 4.11.1.3.

4.11.1.3 RNase treatment

The procedures for RNase treatment were similar to those described in Section 4.6.4. Briefly, an aliquot of 3 µl, undiluted RNase One™ Ribonuclease (Promega) was added to the beads for RNase digestion at 4℃ for 30 minutes. The beads were washed thrice with 150 µl 1X Detergent-free Lysis Buffer (protease inhibitor-free) (Appendix A).

4.11.1.4 Elution and trypsinization

The beads were washed once with 500 µl Pre-Urea Wash Buffer and then the purified protein was eluted with 30 µl of Urea Elution Buffer with incubation on Eppendorf ThermoMixer® C at 37℃ for 30 minutes with mixing at 1000 rpm for 15 seconds at 2-minute intervals [304]. The supernatant was collected and transferred to a new 1.5-ml centrifuge tube. An aliquot of 30 µl of 20 mM DTT was added to the eluate to obtain a final concentration of 10 mM DTT in the mixture. The mixture was incubated on Eppendorf ThermoMixer® C at 25℃ for 30 minutes with mixing at 1000 rpm for 15 seconds at 2-minute intervals. Next, an aliquot of freshly prepared 30 µl of 75 mM iodoacetamide (IAA) was added to the eluate to give a final concentration of 25 mM IAA in the mixture. The mixture was incubated on Eppendorf ThermoMixer® C at 25℃ for 1 hour with mixing at 1000 rpm for 15 seconds at 2-minute intervals. For trypsinization, 228 µl of UltraPure™ Distilled Water (Life Technologies) and 36 µl of 10X Trypsin Reaction Buffer were added to the DTT- and IAA-treated mixture sequentially. Then, an aliquot of 6 µl of 0.4 µg/µl Trypsin (Promega) was added to the eluate. At this point, the urea concentration in the mixture was reduced down to 0.5 M. This is the urea concentration that the trypsin can tolerate. The mixture was incubated on Eppendorf ThermoMixer® C at 37℃ for 16 hours with mixing at 1200 rpm for 15 seconds at 2-minute intervals. The reaction was stopped by adding 36 µl of 1% formic acid to give a final concentration of ~0.1% formic acid in the trypsinized mixture. The trypsinized mixture was stored at -20℃ until being analyzed by mass spectrometry [316].

4.11.1.5 Mass spectrometry analysis

Mass spectrometry analysis was carried out on TripleTOF™ 6600 (AB SCIEX. A portion of 0.5 µg of peptide sample was used to initiate the run of Information Dependent Acquisition on

TripleTOF™ 6600 (AB SCIEX). The sample was loaded onto NanoLC Trap with the phase of ChromXP C18CL (3 µm particle diameter, 300 Å pore size) and with the column size at 350 µm inner diameter × 0.5 mm long (eksigent). Then, the sample was separated by NanoLC column with the phase of ChromXP C18CL (3 µm particle diameter, 12 nm pore size) and with the column size at 75 µm inner diameter x 15 cm long (eksigent). Sample was run using a protocol of 30-minute gradient from 95% solvent A (MS) and 5% solvent B (MS) to 65% solvent A for MS and 35% solvent B for MS at a flow rate of 300 nl/min. Data were acquired under the conditions of an ion spray voltage of 2300 V, nebulizer gas at 15 pounds per square inch and curtain gas at 30 pounds per square inch. The total run time was 70 minutes. The protocol of Information Dependent Acquisition was set up with the MS survey range between 0.350 and 1.500 kilodalton (kDa) in positive mode with accumulation time at 250 ms followed by dependent tandem MS scan with a mass range between 100 and 1800 Da in positive mode with accumulation time at 100 ms. The 40 most intense ions with charge state at 2+ to 4+ under the high-sensitivity mode were monitored in each detection cycle. Exclusion of former target ions was set for a period of 18 seconds after each occurrence and at a mass tolerance of 50 parts per million. The raw data were analyzed by ProteinPilot™ Software 5.0. The data were searched against the proteomic data of *Homo sapiens* in the database of UniProt Proteomes database (16 October 2016, 20123 reviewed proteins). The result was filtered by false discovery rate in less than 1% at both the peptide level and the protein levels.

4.11.2 Thermostability of miRNA-target site duplexes

The thermostability of miRNA-target RNA duplexes is extremely critical to RIINGS. If the duplexes can easily be dissociated at low temperature, the idea of the whole method would fall apart. The two interacting RNAs in the duplex can no longer be ligated together

under the principle used in RIINGS. Furthermore, the melting temperature of the miRNA-target RNA duplexes also restricts the reaction temperature in the steps after the duplexes have been extracted. The reaction temperatures should be lower than the melting temperature of the duplexes; otherwise, duplexes of interest would be dissociated from each other. Here, the melting curve analysis was used to investigate the melting temperature of miRNA-target RNA target site duplexes extracted from K562 cells. Two different amounts of proteinase K (8 U and 24 U) and two different digestion temperatures (4℃ and 16℃) were tested for the preparation of miRNA-target RNA duplexes. The measurement was carried out on LightCycler® 480 (Roche). Figure 4.11.2 shows the experimental scheme for investigating the melting temperature of the miRNA-target RNA target site duplexes.

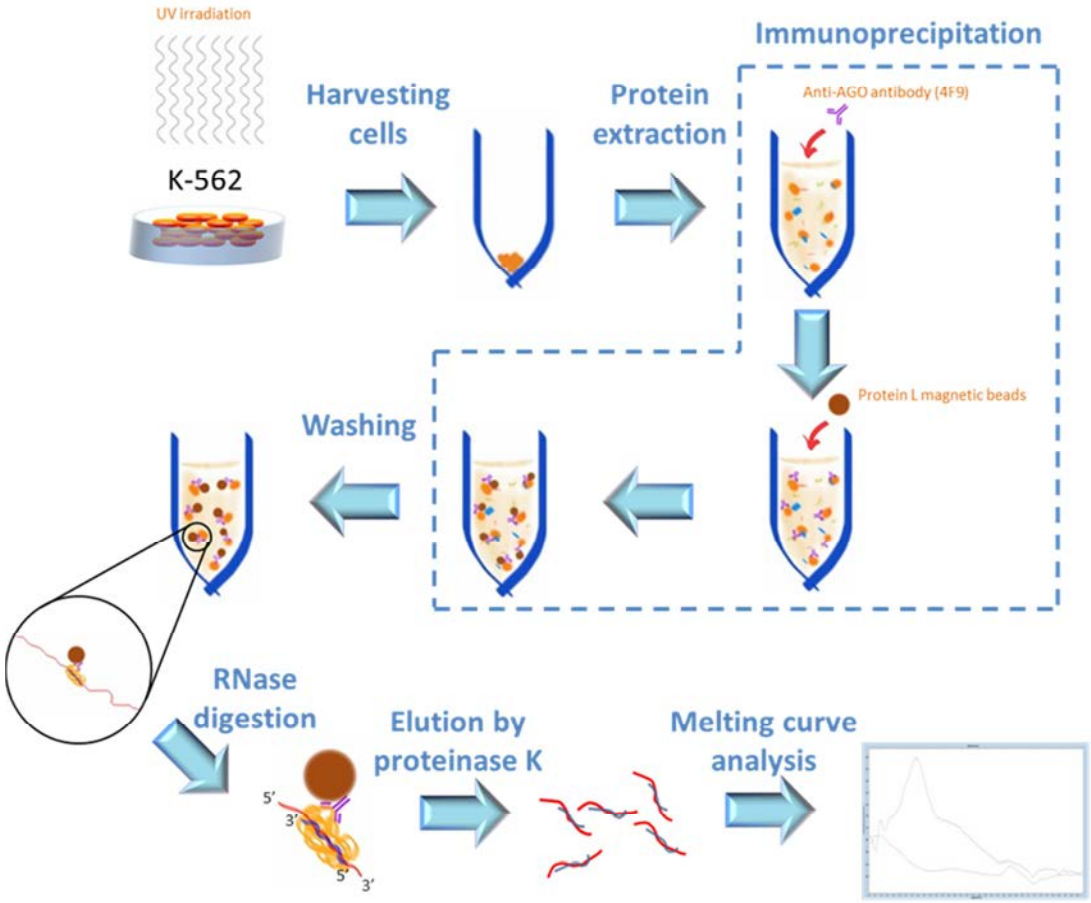


Figure 4.11.2 The experimental scheme for investigating the melting temperature of the miRNA-target RNA target site duplexes. This experiment was set up for assessing the melting curve of the miRNA-target RNA target site duplexes. In examining the suitable condition for releasing the duplexes by proteinase K, two different proteinase K concentrations and two different reaction temperatures for proteinase K digestion were examined.

4.11.2.1 Cell culture, UV irradiation, and preparation and quantification of protein lysate

K562 cell extract was used in this experiment. The procedures for cell culture, UV irradiation, and protein lysate preparation and quantification were the same as described in Section 4.6.1. In brief, cells were irradiated under 254 nm UV light 2 times, once for 400 mJ/cm^2 and once for 200 mJ/cm^2, on a 4℃ pre-chilled ice pack in UVC 500 Ultraviolet Crosslinker (Amersham). Total protein was extracted in 1X Detergent-free Lysis Buffer (Appendix A).

4.11.2.2 Preparation of beads and immunoprecipitation

The procedures for preparing Pierce™ Protein L Magnetic Beads and the subsequent immunoprecipitation were the same as the Method 2 described in Sections 4.5.2 and 4.5.3 respectively with some minor changes. There were 4 reactions set up to investigate the optimum condition for proteinase K to release the RNA duplexes from the protein complexes. The reactions in this experiment were scaled up 2-fold: 4 mg of total protein extract was used as the starting material for the immunoprecipitation reaction. The amount of the antibody input for each reaction was 22.5 µg (112.5 µl). The input beads for initializing the reaction were 300 µl. According to the calculation, the binding capacity of the beads was 16.5-fold higher than the amount of the antibody. The elution procedure in Section 4.5.3 was omitted and the reactions were continued with the steps in Section 4.11.2.3.

4.11.2.3 RNase treatment

The procedures for RNase treatment were similar to those described in Section 4.6.4 with some minor changes. An aliquot of 6 µl undiluted RNase One™ Ribonuclease (Promega) was added to the beads for RNase digestion at 4℃ for 30 minutes. The beads were washed thrice with 1 ml of 1X Detergent-free Lysis Buffer (protease inhibitor-free) (Appendix A).

4.11.2.4 Elution and melting curve analysis of RNA duplex

Each vial of beads was resuspended in 200 µl 1X Detergent-free Lysis Buffer (protease inhibitor-free). Proteinase K (NEB) was added to the vials to digest the proteins: 8 U (10 µl) to two vials and 24 U (30 µl) to the remaining two vials. This resulted in eluting and releasing the miRNA-target RNA target site duplexes from the protein complexes. One of the reactions with 8 U of proteinase K and one of the reactions with 24 U of proteinase K were incubated on Eppendorf ThermoMixer® C at 4℃ for 24 hours with mixing at 1250 rpm for 15 seconds at 2-minute intervals. The remaining two vials were incubated on Eppendorf ThermoMixer® C at 16℃ for 24 hours with mixing at 1250 rpm for 15 seconds at 2-minute intervals. After digestion, the supernatant from each vial was collected and transferred to a new centrifuge tube. All the samples were kept on ice immediately to prevent dissociation of the duplexes.

Melting curve analysis was carried out using LightCycler® 480 (Roche) immediately after the samples had been obtained. Melting curves were examined under 9 different conditions with 3 different Syto 9 (Molecular Probes) concentrations (1.5, 2.0 and 2.5 µM) and 3 different amounts of RNA duplexes (7.5, 10.0 and 12.5 µl) each in a reaction volume of 20 µl. A parallel set of 9 reaction mixtures for blanks was also set up with the respective samples being replaced by equivalent volumes of blank that only had the corresponding

amount of proteinase K in 1X Detergent-free Lysis Buffer (protease inhibitor-free). Details of how to prepare these reaction mixtures can be found in Appendix C Table 8. The reaction mixtures were prepared in a 96-well plate (LightCycler® 480 Multiwell Plate 96, white; Roche) placed on a pre-chilled PCR-Cooler 0.2 ml (Eppendorf). Master Mixes containing different volumes of Duplex Buffer (100 mM potassium acetate; 30 mM HEPES, pH 7.5; IDT) and Syto 9 (50 µM, diluted in water) were prepared and loaded into the plate prior to loading the sample. Melting analysis was performed using the following thermal cycles: rapid ramping (4.4℃/seconds) to 20℃ and holding at 20℃ for 30 seconds, slowing ramping (0.02℃/seconds) to 90℃ with 25 signal acquisitions per ℃ during this slow ramping phase, rapid cooling (2.2℃/second) to 40℃ and holding at 40℃ for 10 seconds for protecting the machine.

4.11.3 Quantification of miRNA- target RNA duplexes

The quantity of the yielded miRNA-target RNA duplexes is another important issue that we have to concern. The more the number of the chimeric sequences obtained from the library preparation, the more representative the whole interactome. In other words, the read depth of the subsequent next-generation sequencing is inherently increased if the number of the chimeric sequences in the sample is increased. This quantity gives a good estimate of how representative the prepared library is before sequencing.

However, it is difficult to count the absolute number of chimeric sequences directly. Only a few approaches can give us a rough idea to the answer. First, the results from the Bioanalyzer 2100 (Agilent) can be used to calculate an approximate molar amount of the yielded RNA duplexes because the results indicate the size range of the fragmented RNAs in the experiment for determining the RNase working concentration (Section 4.8, Section

5.6.1 and Section 5.6.2). In this experiment, all the flanking (protruding) RNAs from the Argonaute-containing complexes had been trimmed and washed away. Therefore, the RNAs input to the Bioanalyzer 2100 (Agilent) were actually the miRNAs and their corresponding target RNA target sites. On the other hand, the molecular mass of adenosine monophosphate, guanosine monophosphate, uridine monophosphate and cytidine monophosphate are 347.22 g/mol, 363.22 g/mol, 324.18 g/mol and 323.20 g/mol respectively. They are the building blocks of RNA and their average molecular mass is 339.46 g/mol. Accordingly, the molecular mass of the input RNA duplexes can be estimated because we can approximately determine the average RNA length based on the results from Bioanalyzer 2100 (Agilent). The molecular mass of RNA can be calculated using Equation 1 as follows:

$$\boldsymbol{Molecular\ Mass}\ (\boldsymbol{g/mol}) = (\boldsymbol{n} \times \boldsymbol{M}) - (\mathbf{18.02} \times (\boldsymbol{n} - \mathbf{1})) \quad \text{(Equation 1)}$$

where **n** is the length of the RNA, **M** is the averaged molecular mass of the ribonucleotide monophosphate (**M** = 339.46 g/mol). The loss of a water molecule during each condensation reaction has been taken into account con in this formula.

Together with the molecular mass, the volume and the concentration (w/v), the molar amount can be calculated using Equation 2. The concentration (w/v) can be measured by either Bioanalyzer 2100 (Agilent) or NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies).

$$\boldsymbol{Mole\ (mol)} = \frac{\boldsymbol{Concentration\ (g/ul)\ \times\ Sample\ Volume\ (ul)}}{\boldsymbol{Molecular\ Mass\ (g/mol)}} \quad \text{(Equation 2)}$$

The molecule number can then be calculated using Equation 3.

$$\boldsymbol{Molecule\ number = Mole\ (mol)\ \times N_A} \quad \text{(Equation 3)}$$

Where $\boldsymbol{N_A}$ is the Avogadro constant (6.022 X 10^{23} mol^{-1}).

Second, quantitative real-time polymerase chain reaction (qPCR) can also be used to quantify the chimeric sequences in the library prepared by RIINGS. During the library preparation, a stem-loop adaptor sequence is eventually inserted in between each pair of miRNA and its corresponding target RNA target site in the chimeric sequence. Furthermore, two sequencing adaptors are added to the two ends of the chimeric sequence. Therefore, a hydrolysis probe (Chimera Quanti Probe) can be designed to be complementary to the loop region and a set of primers (Lib-A Primer A and Lib-A Primer B) complementary to the two sequencing adaptors. This is actually a hydrolysis-probe-based qPCR assay. The PCR efficiency and the working range of the assay would be tested with a synthetic template (Chimera Quanti Synt Temp). All the sequences of the primers, hydrolysis probe and

synthetic template using in this experiment can be found in Table 4.2.1. Figure 4.11.3 shows the design of the Chimeric Sequence qPCR Assay.

With these two methods, we can monitor the quantity of the miRNA-target RNA duplex at two different levels during the RIINGS experiment. The first approach can monitor the duplex at the point after purification. The second approach can monitor the duplex at the point after the chimeric RNA has been constructed from the two interacting RNAs.

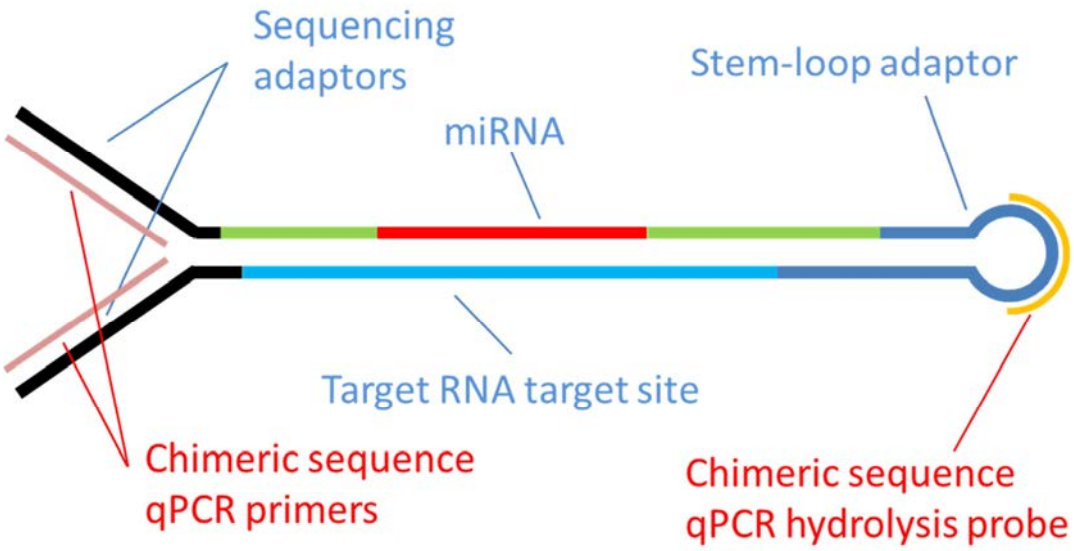


Figure 4.11.3 The design of the Chimeric Sequence qPCR Assay. A hydrolysis probe (Chimera Quanti Probe) was designed to be complementary to the loop region while two primers (Lib-A Primer A and Lib-A Primer B) were designed to be complementary to the two sequencing adaptors.

4.11.3.1 Validation of Chimeric Sequence Quantification Assay

A 10-fold serial dilution of the synthetic template (Chimera Quanti Synt Temp; Table 4.2.1) was prepared to give the following concentrations: 100 nM, 10 nM, 1 nM, 100 pM, 10 pM, 1 pM, 100 fM, 10 fM, 1 fM, 100 aM, 10 aM and 0 M. All dilutions were assayed in triplicate. The qPCR was carried out on LightCycler® 480 (Roche) in 96-well format. Dispensed in a 96-well plate (LightCycler® 480 Multiwell Plate 96, white; Roche), the 20-µl reaction mixtures each contained 1 µl diluted synthetic template (Chimera Quanti Synt Temp; see above), 10 µl TaqMan® Universal PCR Master Mix II, no UNG (ABI), 250 nM hydrolysis probe (Chimera Quanti Probe), 500 nM each of the two primers (Lib-A Primer A and Lib-A Primer B). The final volume was adjusted to 20 µl by adding appropriate volume of water. The amplification cycles of the Chimeric Sequence qPCR Assay were 95℃ 10 minutes followed by 50 cycles of 95℃/15 seconds and 60℃/1 minute.

4.12 Extraction of interacting duplexes by Proteinase K

In order to release the miRNA-target RNA target site duplexes from the Argonaute protein complexes, proteinase K was selected for the digestion of the attached Argonaute protein. Proteinase K is a subtilisin-related serine proteinase [317] and is able to work in a broad range of conditions, including different temperatures and buffers [318]. This enzyme is capable of hydrolyzing nearly all proteins (including keratin) to their corresponding amino acid residues.

The digestion condition is essential in this digestion process. While the protein attaching to the miRNA-target RNA target site duplex must be removed, the two interacting RNAs must

be maintained in duplex form. Therefore, the temperature for proteinase K digestion must not be too high because dissociation of the RNA duplexes will increase with increasing temperature. Moreover, the reaction buffer must not contain any nucleic acid denaturing agent.

With the regard to these issues, a mock experiment was first used to roughly check the proteinase K digestion at low temperature (4-16℃). *Bovine serum albumin* (BSA) was chosen as the substrate in this pilot test. A time-course investigation into the proteinase K digestion was conducted. Each of the time-course experiments was incubated at 4 different digestion temperatures. The digestion temperatures were 4℃, 8℃, 12℃ and 16℃. The experiment was performed in 1X Detergent-free Lysis Buffer (Appendix A). The completeness of the digestion was visualized by SDS-PAGE with Coomassie blue staining. Figure 4.12.1 shows the experimental scheme for a quick check of proteinase K digestion at low temperature (4-16℃).

Moreover, L-Amino Acid Quantification Kit (Sigma-Aldrich), which can be used to quantify the L-amino acid concentration, was used to monitor the increment of the amino acid level during proteinase K digestion in the real situation (the substrate was the *immuno-precipitated Argonaute complexes*). The kinetic curve of plotting amino acid concentration against time could suggest an optimal digestion time and digestion temperature in the real situation in RIINGS. Figure 4.12.2 shows the experimental scheme for monitoring the increment of the amino acid level during proteinase K digestion of immunoprecipitated Argonaute complexes.

Furthermore, melting curve analysis, which was used to investigate the melting temperature of miRNA-target RNA target site duplexes extracted from K562 cells in Section

4.11.2, was also used to further confirm the optimal digestion time and digestion temperature in the real situation in RIINGS. As the melting peak height is proportional to the duplex concentration, the melting peak height can be used as a reference for comparing the relative duplex concentration in the extract. The duplexes extracted by proteinase K under different digestion times and different digestion temperatures were investigated. The figure of the experimental scheme and the experimental procedures are the same as those in Section 4.11.2.

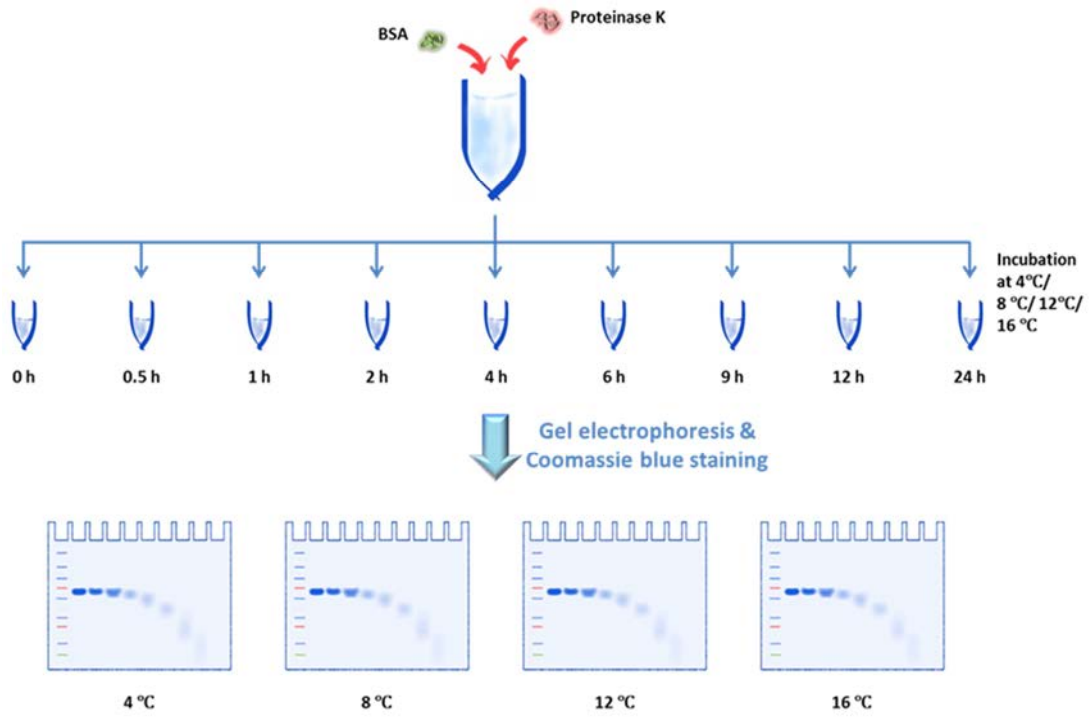


Figure 4.12.1 The experimental scheme for a quick check of proteinase K digestion at low temperature (4-16℃). A time course experiment has been conducted to have a quick look at the digestion progress of proteinase K. 4 different digestion temperatures have been investigated. BSA has been used as the substrate in this mock experiment.

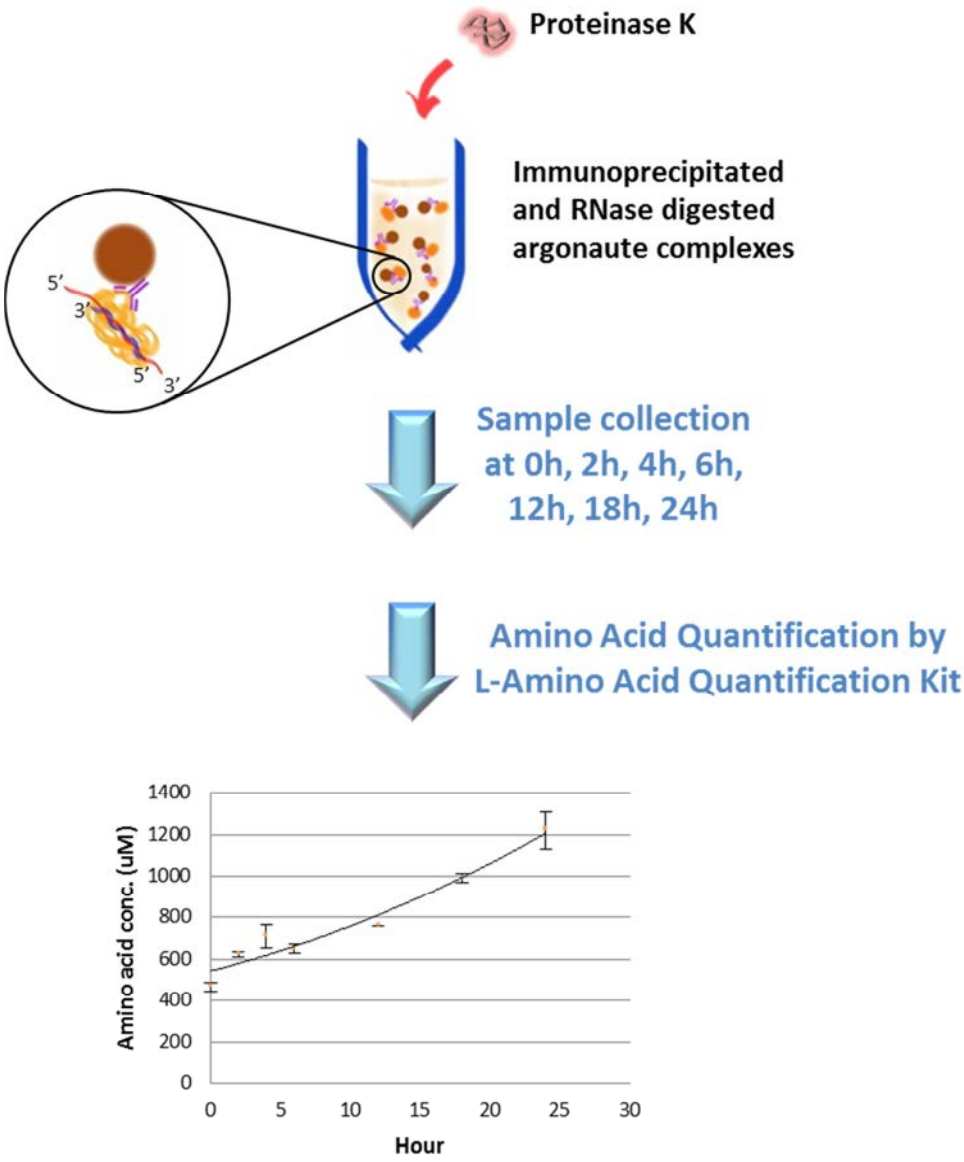


Figure 4.12.2 The experimental scheme for monitoring the increment of the amino acid level during the proteinase K digestion of immunoprecipitated Argonaute complexes. The Argonaute complexes were first prepared by immunoprecipitation. The flanking RNAs were then trimmed away. Proteinase K was used to digest the extracted Argonaute complexes. The amino acid concentration in the supernatant at different time points was finally quantified by L-Amino Acid Quantification Kit (Sigma-Aldrich).

4.12.1 Mock experiment for checking proteinase K digestion at low temperature (4-16 ℃)

BSA was selected as the digestion substrate in this mock experiment. The reactions were set up in final volume of 25 µl. A master mix was first prepared and then dispensed into individual PCR tubes. Each reaction mixture contained 1X Detergent-free Lysis Buffer (Appendix A), 2 µg/µl BSA (Thermo Scientific), 1U proteinase K (NEB) and sterile water to make up to 25 µl. The reactions were incubated on Veriti® 96-well Thermal Cycler (ABI). There were 4 sets of the time-course experiment prepared for investigating 4 different digestion temperatures. The digestion temperatures were 4℃, 8℃, 12℃ and 16℃. Each set of the time-course experiment contained 9 reactions for 9 time points: 0, 0.5, 1, 2, 4, 6, 9, 12 and 24 hours. The reactions were terminated by heat inactivation at 95℃ for 10 minutes.

SDS-PAGE was used to visualize the digested BSA. All the samples were mixed with 2X Laemmli Sample Buffer (Appendix A) in a 1:1 ratio and heat denatured at 95℃ for 5 minutes before being loaded onto the gel. The loading volume of each sample was 20 µl (i.e. 10 µl sample + 10 µl 2X Laemmli Sample Buffer). The SDS-PAGE was carried out on Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad). The gels used in the SDS-PAGE were 15% separating and 4% stacking gels (Appendix B) of 1 mm thick. Gels were allowed to polymerize for at least 2 hours. According to the instructions from the manufacturer, the SDS-PAGE was run under constant voltage at 200V until the dye front reached the bottom of the gel in 1X SDS-PAGE Running Buffer. The electric power was supplied by PowerPac™ Basic Power Supply (Bio-Rad). The gels were soaked in Coomassie Brilliant Blue R-250 Staining Solution with shaking for 1 hour after electrophoresis. Then, the protein bands on the gels were differentiated by soaking in Coomassie Brilliant Blue Destaining Solution with shaking. The Coomassie Brilliant Blue Destaining Solution was changed when the gel color

and the solution color appeared the same. The gels were destained until the background became clear.

4.12.2 Measurement of amino acid level during proteinase K digestion

K562 cell extract was used in this experiment for the preparation of the miRNA-target site-Argonaute complexes. The procedures for preparing miRNA-target RNA-Argonaute complexes were the same as those for the thermostability of miRNA-target site duplexes described in Section 4.11.2. First, the procedures for cell culture, UV irradiation, and preparation and quantification of protein lysate strictly followed those in Section 4.11.2.1. Second, the procedures for preparing bead and the ensuing immunoprecipitation were exactly the same as those in Section 4.11.2.2. There were 4 reactions set up to examine the kinetics for monitoring the increment of the amino acid level during the proteinase K digestion in the real situation. Third, the RNase treatment followed the steps described in Section 4.11.2.3.

After the preparation of the miRNA-target site-Argonaute complexes, the beads with the complexes were resuspended in 200 µl 1X Detergent-free Lysis Buffer (protease inhibitor-free). Proteinase K was added to the beads to digest proteins. This resulted in releasing amino acids to the supernatant and the concentration of the amino acid increased with increasing incubation time. Two of the reactions were added with 8 U of proteinase K (NEB) while the remaining two reactions were added with 24 U of proteinase K (NEB) for digestion. One of the reactions with 8 U of proteinase K and one of the reactions with 24 U of proteinase K were incubated on Eppendorf ThermoMixer® C at 4℃ for 24 hours with mixing at 1250 rpm for 15 seconds at 2-minute intervals. The remaining two reactions were incubated on Eppendorf ThermoMixer® C at 16℃ for 24 hours with mixing at 1250 rpm for

15 seconds at 2-minute intervals. A portion of 15 µl bead suspension was collected and transferred to a new centrifuge tube at 0, 2, 4, 6, 12, 18, 24 hours after the digestion. The beads were collected by placing the vial on a magnetic stand for 30 seconds. The supernatant was transferred to a PCR tube and proteinase K inactivated at 95℃ for 10 minutes immediately. All the samples were kept in a freezer at -20℃ until the amino acid concentration measurement.

The amino acid concentration of supernatant samples was quantified by L-Amino Acid Quantification Kit (Sigma-Aldrich) in accordance with the manufacturer's instructions for fluorometric detection. An aliquot of 5 µl sample was used to initialize the quantification reaction for each sample. The reactions for the measurement were prepared on ice and a reaction mixture of 30 µl was pipetted into a 96-well plate (LightCycler® 480 Multiwell Plate 96, white; Roche) on a pre-chilled PCR-Cooler 0.2 ml (Eppendorf). The fluorescence intensity was measured by LightCycler® 480 (Roche) in 96-well format with the filter combination of λ_{Ex} = 523 nm / λ_{Em} = 610 nm.

4.13 Mismatch repair of miRNA-target RNA heteroduplexes at miRNA 5' and 3' ends

During miRNA targeting, miRNA incorporated in the Argonaute protein binds to the target site via partial base pairing. This imperfect pairing results in some mismatches and/or bulging structures in the miRNA-target heteroduplex. In real situations, the mismatches may also be at either end or both ends of the miRNA and lift up the 5' and/or 3' ends of the miRNA (Figure 4.13.1a). However, any mismatched end is unfavorable to RIINGS in the gap

repairing (filling) step because no ligase is able to ligate a nick with a mismatch at either side of the nicking site.

In order to repair the mismatches with the miRNA target at the 5’ and the 3’ ends of miRNA, ligation of a pair of mismatch repairing adaptors (short RNA oligos; simply called repairers hereafter) is needed. This pair of repairers extends the 5' and the 3' mismatched ends to new extended ends that each perfectly match to their corresponding target sequence (Figure 4.13.1b). However, there are several remarkable challenges in this task. First, the sequence of the target RNA is not known in advance in the real situation. Second, a repairer with a particular sequence is unable to ligate to a particular end specifically even if the required sequence is known. Third, the requisite length of the repairer is not easy to determine because the minimal number of bases needed to reform a stable complimentary end is unknown. Fourth, the rate of successful repair would be extremely low if the repairer is composed of degenerate bases; for example, a tetramer N (NNNN) would theoretically generate a complimentary end with all four bases matched with a probability of $(1/4)^4$ or 1/256.

There are two possible ways to resolve the difficulties in the mismatch repair. One approach is the use of inosine in the repairers while the other approach is the use of 5-nitroindole in the repairers (see Section 6.8 for discussion). Two different sets of mock experiments were designed for evaluating the ligation of all these repairers. One set was used for testing the ligation of miRNA 5’ repairers (inosine-containing 5’ repairer and 5-nitroindole-containing 5’ repairer) to the 5’ end of miRNAs. This set of mock tests were carried out on four different sets of mock miRNA-target RNA duplexes that were designed to mimic the 5’ end of the miRNAs with no mismatch, 1 mismatch, 2 mismatches or 3 mismatches. The other set was used for testing the ligation of miRNA 3’ repairers (inosine-

containing 3’ repairer and 5-nitroindole-containing 3’ repairer) to the 3’ end of miRNAs. This set of mock tests was carried out on the same sets of mock duplexes used in testing the ligation of miRNA 5’ repairers. Figure 4.13.2 shows the four duplexes used in the mock tests. Table 4.2.1 shows the sequences of these RNA oligos. Since the shortest commercially available oligo is 4-mer oligo, the mock test was started with tetramer repairers.

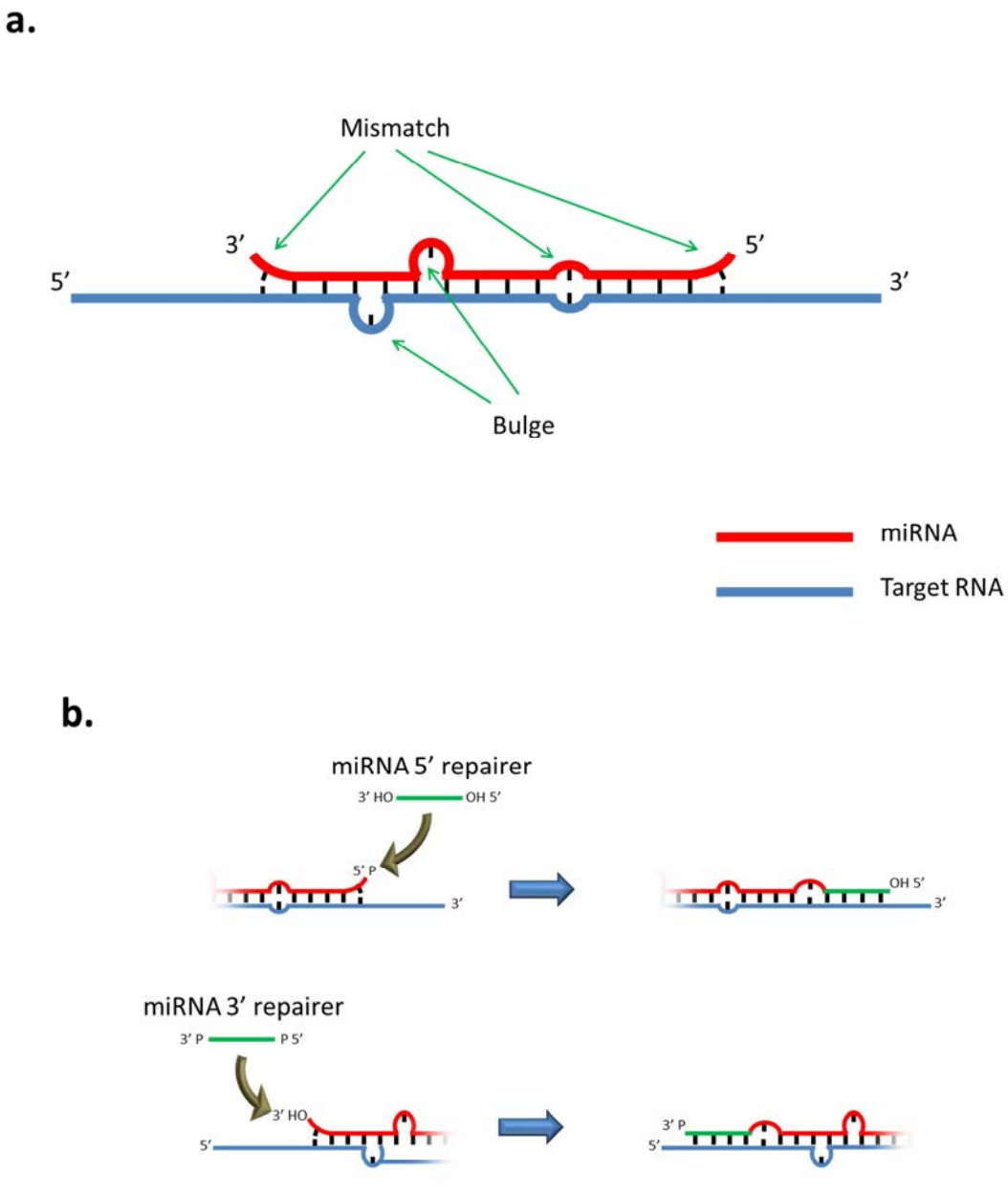


Figure 4.13.1 An example of miRNA-target RNA duplex and the strategy for repairing the mismatches at the both ends of the miRNA. (a) The figure shows examples of bulging and mismatched structures in the miRNA-target RNA duplex. (b) The figure illustrates the repair of miRNA 5′ mismatch (the upper panel) and the repair of miRNA 3′ mismatch (the lower panel).

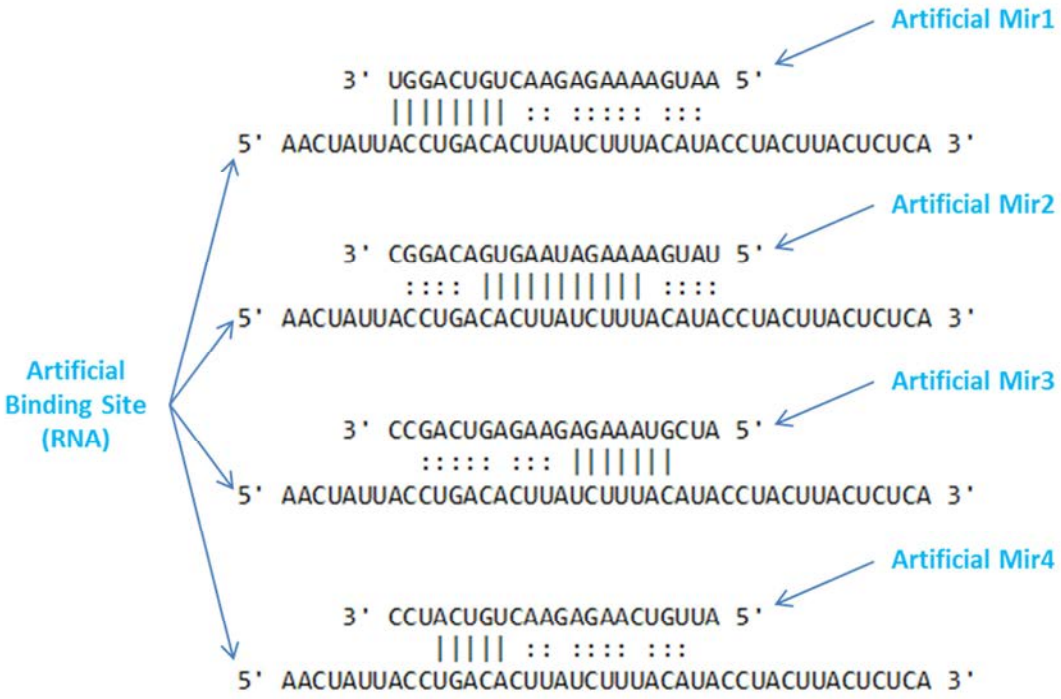


Figure 4.13.2 Four mock duplexes for mimicking miRNA duplexes with mismatch(es) at the 5' and 3' ends of miRNAs. From top to bottom, the first mock duplex mimics the miRNAs that have 1 mismatch at the 5' end and no mismatch at the 3' end. The second mock duplex mimics the miRNAs that have no mismatch at the 5' end and 1 mismatch at the 3' end. The third mock duplex mimics the miRNAs that have 3 mismatches at the 5' end and 2 mismatches at the 3' end. The fourth mock duplex mimics the miRNAs that have 3 mismatches at the 5' end and 2 mismatches at the 3' end.

4.13.1 Random cytosine/inosine tetramer

Inosine (I) is a natural base found in RNAs. It can pair up with adenosine (A), uracil (U) and cytosine (C), but not guanosine (G) in RNA [319]. Accordingly, inosine cannot be used alone in the repairers. It must be used together with cytosine in the repairers to compensate for the inosine-guanosine mismatch. Therefore, tetramer of degenerate bases composed of cytosine and inosine was the design chosen in this approach. Tetramer with phosphate groups on both ends was the miRNA 3' end repairer while tetramer with hydroxyl groups on both ends was the miRNA 5' end repairer.

4.13.1.1 Evaluation of inosine-containing miRNA 5' repairer

First, "Random C/I Tetramer" (Table 4.2.1) with hydroxyl groups on both ends (miRNA 5' repairer) was tested in a set of mock experiments mimicking the ligation of 5' repairer to the 5' end of the miRNAs with 2 mismatches or 3 mismatches. Second, tetramer N (N stands for A, U, C or G) (Table 4.2.1) with hydroxyl groups on both ends were also tested in the same set of mock experiments. Third, tetramer C (Table 4.2.1) and tetramer (CCCI) with hydroxyl groups on both ends (Table 4.2.1) were tested in the mock experiment mimicking the ligation of the repairer to the 5' end of the miRNAs with 3 mismatches. Figure 4.13.3 shows the experimental scheme.

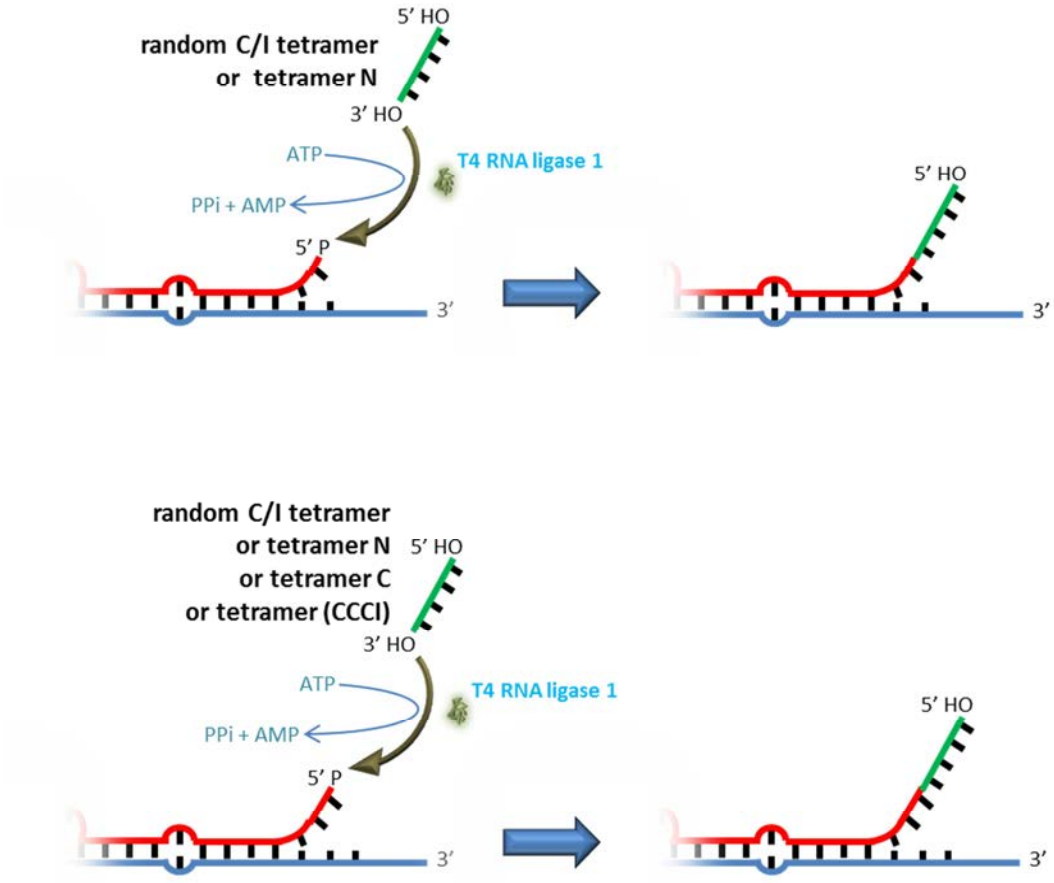


Figure 4.13.3 The scheme of the mock experiments mimicking the ligation of miRNA 5′ repairers to miRNAs with different numbers of 5′ mismatches with the target RNA. From top to bottom, the first panel shows the ligation of 5′ repairers (green) to the 5′ end of the mock miRNA (red) whose 5′ end has 2 mismatches with the target RNA (blue). The second panel shows the ligation of 5′ repairers to the 5′ end of the mock miRNA whose 5′ end has 3 mismatches with the target RNA. Note that the miRNA 5′ repairers have hydroxyl groups at both 3′ and 5′ ends.

Preparation of the mock duplexes

Appendix C Table 9 shows the preparation of annealing reaction mixtures. Briefly, an aliquot of 1.5 µl Artificial Binding Site (RNA) (100 µM) and an aliquot of 0.15 µl of either Artificial Mir3 (100 µM) or Artificial Mir4 (100 µM) (see Figure 4.13.2) were mixed in a PCR tube for each reaction. The annealing of the mock duplexes was carried out on Veriti® 96-well Thermal Cycler (ABI): 95℃ 1 minute and 80℃ 1 minute followed by gradual cooling down to 25℃ at a ramping rate of 1% of the full ramping speed. The reaction was then further cooled down to 4℃ and kept at this temperature until use.

Ligation of miRNA 5' end repairers

Appendix C Table 10 shows the preparation of miRNA 5' repairer ligation mixtures. Briefly, the setup started with preparing a master mix that did not contain the annealed duplex in a clean tube. For each reaction of this master mix, an aliquot of 6 µl of water was first added to the tube. Then, 1 µl of 10X T4 RNA Ligase Reaction Buffer (NEB), 0.1 µl of ATP (100 mM) (NEB), 0.75 µl of miRNA 5' mismatch repairer (1000 µM) and 0.5 µl of T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30U/ µl, NEB) were transferred into the tube sequentially. The master mix was mixed and spun down briefly. Next, the master mix was chilled on ice until the temperature was equilibrated (preventing denaturation of the duplex at room temperature). The 1.65-µl annealed duplex was then added to the 8.35-µl master mix to obtain the 10-µl reaction mixture (The annealed duplexes and master mixes could be scaled up proportionally if multiple reactions were to perform in parallel).

The mixture was mixed and spun down briefly. Finally, the tubes were immediately incubated in thermal block of Veriti® 96-well Thermal Cycler (ABI).

A time-course study was conducted for evaluating the ligation of "Random C/I Tetramer" to 5' end with 2 mismatches (Artificial Binding Site (RNA) + Artificial Mir4) and to 5' end with 3 mismatches (Artificial Binding Site (RNA) + Artificial Mir3). The incubation conditions were 16℃ for 0, 2, 4, 8, 16 and 24 hours. All the reactions were inactivated at 65℃ for 15 minutes. Similarly, a time-course study was also carried out for evaluating the ligation of tetramer N to 5' end with 2 mismatches and to 5' end with 3 mismatches, and tetramer C or tetramer (CCCI) to 5' end with 3 mismatches. The incubation conditions were 16℃ for 0, 0.5, 1, 1.5, and 2 hours. All the reactions were inactivated at 65℃ for 15 minutes.

RNA denaturing gel electrophoresis, staining and visualization

Hoefer Electrophoresis Unit SE600 (Hoefer) equipped with PolyScience® Digital Temperature Controller (PolyScience) was used to run the denaturing/urea polyacrylamide gel electrophoresis (Urea-PAGE). All the samples were mixed with 2X Urea-PAGE Loading Buffer (Appendix A) in a 1:1 ratio and heat-denatured at 90℃ for 5 minutes. The samples were then chilled on ice immediately for at least 1 minute. The loading volume of each sample was 10 µl (5 µl sample + 5 µl 2X Urea-PAGE Loading Buffer). The urea polyacrylamide gels used in this Urea-PAGE were 8M urea, 10% TBE-polyacrylamide gel (Appendix B). Gels were set up in SE6102 Glass Plates (PR) (18 cm width X 16 cm length, Hoefer) with 1 mm spacers according to the instructions from the manufacturer. Gels were allowed to polymerize for at least 2 hours. Electrophoresis was pre-run under constant voltage

at 150 V for 30 minutes at 50℃ in 1X TBE Buffer. Then, the samples were loaded onto the wells of the gel and run under constant voltage at 150V for 2.5 hours at 50℃ in 1X TBE Buffer. The electric power was supplied by PowerPac™ Basic Power Supply (Bio-Rad). After electrophoresis, the gel was post-stained for 5 to 10 minutes in 0.5X SYBR® Gold nucleic acid gel stain (Life Technologies) that was diluted in 1X TBE Buffer. The image was then captured by ChemiDoc™ MP Imaging System (Bio-Rad) with Image Lab™ Software version 4.1.

4.13.1.2 Evaluation of inosine-containing miRNA 3' repairer

Two miRNA 3' repairers, "Tetramer C-Phosphorylated" (Table 4.2.1) and "Tetramer-Phosphorylated (ICCC)" (Table 4.2.1), with phosphate groups on both ends were tested in a set of mock experiments mimicking the ligation of 3' repairer to the 3' end of the miRNAs with 2 mismatches or 3 mismatches. Figure 4.13.4 shows the experimental scheme.

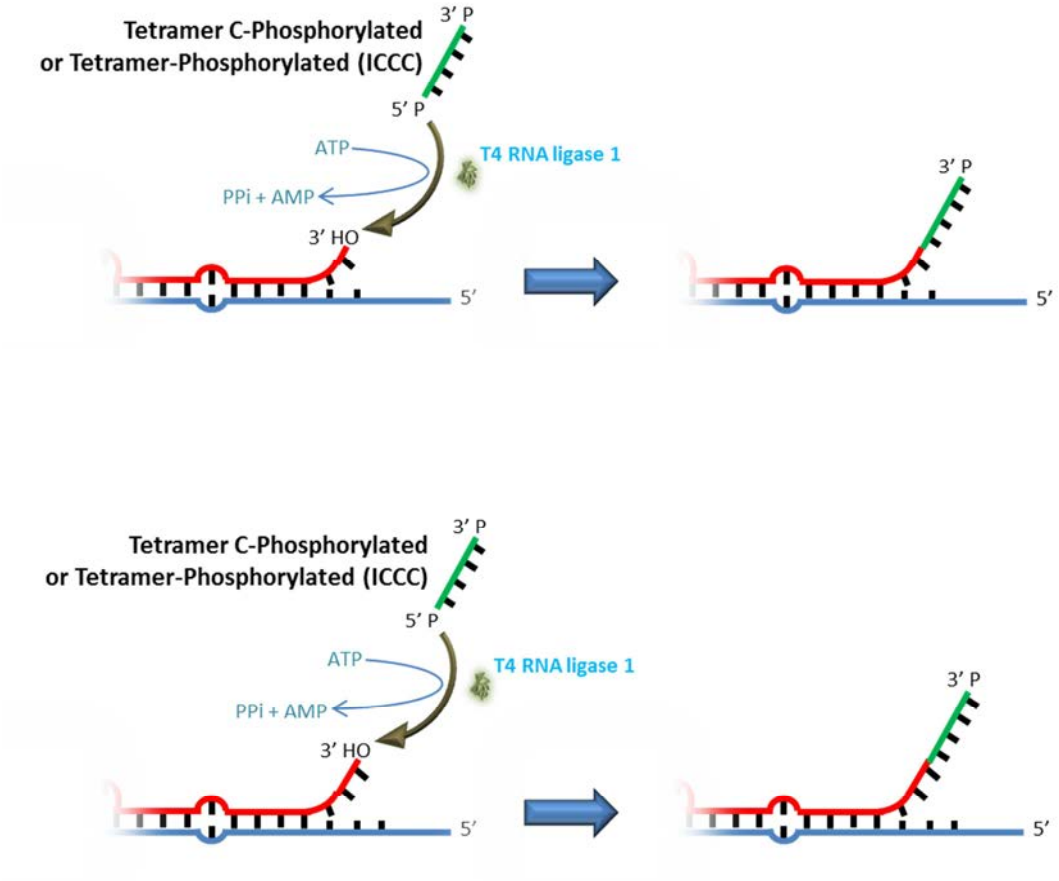


Figure 4.13.4 The scheme of the mock experiments mimicking the ligation of miRNA 3′ repairers to miRNAs with different numbers of 3′ mismatches with the target RNA. From top to bottom, the first panel shows the ligation of 3′ repairers (green) to the 3′ end of the mock miRNA (red) whose 3′ end has 2 mismatches with the target RNA (blue). The second panel shows the ligation of 3′ repairers to the 3′ end of the mock miRNA whose 3′ end has 3 mismatches with the target RNA. Note that the miRNAs 3′ repairers have phosphate groups at both 3′ and 5′ ends.

5' Dephosphorylation of mock miRNAs

The mock miRNAs for evaluating the ligation of miRNA 3' repairers to their 3' end needed to have the phosphate group on their 5' end removed prior to the evaluation experiment. Appendix C Table 11 shows the preparation of dephosphorylation reaction mixtures. Briefly, an aliquot of 8 µl of water, 2 µl of 10X Shrimp Alkaline Phosphatase Reaction Buffer (USB), 5 µl of Shrimp Alkaline Phosphatase (1 U/µl, USB) and 5 µl of either Artificial Mir 3 (100 µM) or Artificial Mir 4 (100 µM) (see Figure 4.13.2) were mixed in a PCR tube for each 20-µl dephosphorylation reaction. The tubes were incubated in the thermal block of Veriti® 96-well Thermal Cycler (ABI): 37℃ for 24 hours and then 65℃ for 15 minutes (inactivation). The dephosphorylated mock miRNAs were ready for use in the assessment of 3' end repairer ligation.

Preparation of the mock duplexes, ligation of miRNA 3' end repairers and RNA denaturing gel electrophoresis

The procedures for the preparation of mock duplexes, ligation of miRNA 3' repairers, and Urea-PAGE were the same as those described in Section 4.13.1.1. Dephosphorylated "Artificial Mir3" and "Artificial Mir4" were annealed to "Artificial Binding Site (RNA)" individually for mimicking the 3' end of miRNAs with 2 mismatches and 3 mismatches respectively. Two miRNA 3' repairers were evaluated for their ligation to both of the mock duplexes prepared above. These two miRNA 3' repairers were "Tetramer C-Phosphorylated" and "Tetramer-Phosphorylated (ICCC)". Since the concentration of mock miRNAs were reduced 4 times after dephosphorylation, the volume of dephosphorylated mock miRNA

added to the annealing mixture (Appendix C Table 12) and the ligation mixture (Appendix C Table 13) were increased 4 times accordingly. A time-course study was conducted for evaluating the ligation of these 3' repairers to 3' end with 2 mismatches and to 3' end with 3 mismatches. The incubation conditions were 16℃ for 0, 2, 4, 8, 16, 24 and 36 hours.

4.13.2 5-Nitroindole-containing repairers

5-Nitroindole is an artificial base. Unlike ordinary bases, 5-nitroindole is a genuine universal base and can pair up with all canonical natural DNA bases, i.e. A, T, C and G [320, 321]. Accordingly, the repairer is much more favorable in mismatch repairing if it is a 5-nitroindole oligomer. However, the 5-nitroindole base *may* not be a suitable substrate for T4 RNA Ligase 1 and other enzymes because this base is an artificial base. Moreover, RNA substrates of T4 RNA Ligase 1 have to fulfil the following requirements: the 5' end with 5' phosphate group (*donor end* that provides electrons) should have at least one base, and the 3' end with 3' hydroxyl group (*acceptor end* that receives electrons) should have at least three bases [314]. With regard to the above suspicion and the substrate requirement of T4 RNA Ligase 1, two 5-nitroindole-containing miRNA 5' repairers and two 5-nitroindole-containing miRNA 3' repairers were designed and evaluated.

4.13.2.1 Evaluation of 5-nitroindole-containing miRNA 5' repairer

The two miRNA 5' repairers were "5' Repairer (GXXGA)" and "5' Repairer (GXAGA)" (Table 4.2.1). Both their 5' and 3' ends had hydroxyl groups and their RNA sequences were GXXGA and GXAGA respectively (X represents 5-nitroindole). During the ligation of 5' repairer, the 5' end of miRNA (donor end) attacked the 3' end of the repairer (acceptor end) and formed a phosphodiester bond at the junction. Therefore, the *acceptor end* of "5' Repairer

(GXXGA)" in fact contained one 5-nitroindole (XGA) whereas "5' Repairer (GXAGA)" did not contain 5-nitroindole (AGA). The experimental scheme of this evaluation was identical to Figure 4.13.3 except the repairers.

Annealing of the mock duplexes

The procedure for annealing the mock duplexes for this investigation was the same as that described in Section 4.13.1.1. An aliquot of 1.5 µl Artificial Binding Site (RNA) (100 µM) and anyone of an aliquot of 0.15 µl Artificial Mir1 (100 µM), 0.15 µl Artificial Mir2 (100 µM), 0.15 µl Artificial Mir3 (100 µM) or Artificial Mir4 (100 µM) (see Figure 4.13.2) were mixed to generate the duplexes.

Ligation of miRNA 5' repairers

First, the "5' Repairer (GXXGA)" and "5' Repairer (GXAGA)" were evaluated using mock duplexes in which the 5' end of the mock miRNAs had 2 mismatches (Artificial Binding Site (RNA) + Artificial Mir4) or 3 mismatches (Artificial Binding Site (RNA) + Artificial Mir3) (see Figure 4.13.2). The experimental procedures were the same as those described in Section 4.13.1.1.

Second, evaluation of "5' Repairer (GXAGA)" was repeated in another buffer system (NEBNext® Quick Ligation Reaction Buffer) for duplexes with the 5' end of the mock miRNAs that had no mismatches (Artificial Binding Site (RNA) + Artificial Mir2), 1 mismatch (Artificial Binding Site (RNA) + Artificial Mir1), 2 mismatches (Artificial Binding Site (RNA) + Artificial Mir4) or 3 mismatches (Artificial Binding Site (RNA) + Artificial Mir3). Appendix C Table 14 shows the preparation of the miRNA 5' repairer ligation mixtures. Briefly, the setup of the 10-µl ligation reaction

was generally the same as that described in Section 4.13.1.1 with some exceptions. For each reaction of the master mix, only an aliquot of 5.1 µl of water was first added to the tube. Then, 2 µl of 5X NEBNext® Quick Ligation Reaction Buffer (NEB), 0.75 µl of miRNA 5' mismatch repairer (1000 µM) and 0.5 µl of T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30U/µl, NEB) were transferred into the tube sequentially. The master mix was mixed and spun down briefly. Then, the prepared master mix was used in the subsequent procedures mentioned in Section 4.13.1.1.

Time-course studies were conducted for evaluating the 5' repairers with reaction mixtures prepared using two different buffers. The incubation conditions were 16℃ for 0, 0.5, 1, 1.5, and 2 hours for the first evaluation, and 16℃ for 0, 0.5, 1, 2, and 4 hours for the second evaluation. All the reactions were inactivated at 65℃ for 15 minutes.

RNA denaturing gel electrophoresis, staining and visualization

The procedures for gel electrophoresis, staining and visualization were the same as those described in Section 4.13.1.1.

4.13.2.2 Evaluation of 5-nitroindole-containing miRNA 3' repairer

The two miRNA 3' repairers were "3' Repairer (AAXXG)" and "3' Repairer (AXXXG)" (Table 4.2.1). Both their 5' and 3' ends were phosphorylated and their RNA sequences were AAXXG and AXXXG respectively. In contrast to the 5' repairer, the 5' end of the miRNA 3' repairer (*donor end*) was ligated to the 3' end of miRNA (*acceptor end*) during the ligation of 3' repairer. Therefore, the donor (5') end of both "3' Repairer (AAXXG)" and "3' Repairer (AXXXG)" was an adenosine. The donor adenosine was next to an adenosine in the former

repairer, but immediately next to 5-nitroindole in the latter repairer. The experimental scheme of this evaluation was identical to Figure 4.13.4 except the repairers.

Annealing of the mock duplexes

The procedure for annealing the mock duplexes for this investigation was the same as that described in Section 4.13.1.1. An aliquot of 1.5 µl Artificial Binding Site (RNA) (100 µM) and anyone of an aliquot of 0.15 µl Artificial Mir1 (100 µM), 0.15 µl Artificial Mir2 (100 µM), 0.15 µl Artificial Mir3 (100 µM) or Artificial Mir4 (100 µM) (see Figure 4.13.2) were mixed to generate the duplexes.

Ligation of miRNA 3' repairers

First, the "3' Repairer (AAXXG)" was evaluated using mock duplexes in which the 5' end of the mock miRNAs had 2 mismatches (Artificial Binding Site (RNA) + Artificial Mir3) or 3 mismatches (Artificial Binding Site (RNA) + Artificial Mir4) (see Figure 4.13.2). The experimental procedures were the same as those described in Section 4.13.1.1.

Second, evaluation of "3' Repairer (AXXXG)" was evaluated in NEBNext® Quick Ligation Reaction Buffer (NEB) for duplexes with the 3' end of the mock miRNAs that had no mismatches (Artificial Binding Site (RNA) + Artificial Mir1), 1 mismatch (Artificial Binding Site (RNA) + Artificial Mir2), 2 mismatches (Artificial Binding Site (RNA) + Artificial Mir3) or 3 mismatches (Artificial Binding Site (RNA) + Artificial Mir4) (see Figure 4.13.2). The experimental procedures were the same as those described for the second evaluation in Section 4.13.2.1. The miRNA 3' repairer

ligation mixtures were prepared as indicated in Appendix C Table 14 except that the 3' repairer (AXXXG) was used instead.

Time-course studies were conducted for these two evaluation experiments. The incubation conditions were 16℃ for 0, 2, 4, 8, 16, 24 and 36 hours for the first evaluation, and 16℃ for 0, 0.5, 1, 2, and 4 hours for the second evaluation. All the reactions were inactivated at 65℃ for 15 minutes.

RNA denaturing gel electrophoresis, staining and visualization

The procedures for gel electrophoresis, staining and visualization were the same as those described in Section 4.13.1.1.

4.14 Evaluation of the repair of the nick for joining the interacting miRNA-target RNA sequences and the potential of extension from the miRNA 3' end after mismatch repair

This is an extension of investigations into the sequences of the miRNA 5' and 3'repairers.

After the miRNA 5' repairer has been ligated to the 5' end of the miRNAs and the mismatch at the 5' end of the miRNAs successfully repaired (after the second mismatch repair as seen in Figure 3.1), a gap is consequently formed between the miRNA sequence and the "Stem-Loop Adaptor (RNA)". Note that the "Stem-Loop Adaptor (RNA)" forms an intra-strand duplex stem-loop with the 3' end recessed on the stem structure and facing the 5' end of the miRNA and the 5' end of the linked miRNA 5' repairer (see Figure 3.1). This gap has to be repaired or filled in order to physically link the interacting miRNA and target RNA

together. After the gap is filled, a *nick* between the 5’ end of the miRNA 5’ repairer and the 3’ end of the “Stem-Loop Adaptor (RNA)” also has to be sealed (Figure 4.14.1a); note that a nick refers to a break in a double-stranded nucleic acid duplex where adjacent nucleotides of one strand are not linked by a phosphodiester bond. Likewise, extension from the 3’ end of the miRNA 3’ repairer to the restriction enzyme site on the “Re-attaching Adaptor” by repeated ligation of some short RNA oligos is required after the miRNA 3’ repairer has been ligated to the 3’ end of the miRNAs and the mismatch at the 3’ end of the miRNAs successfully repaired (Figure 4.14.1b). This task can be achieved by T4 RNA ligase 2 (NEB) – a ligase that actively catalyzes the joining of nicks on *double-stranded* RNA.

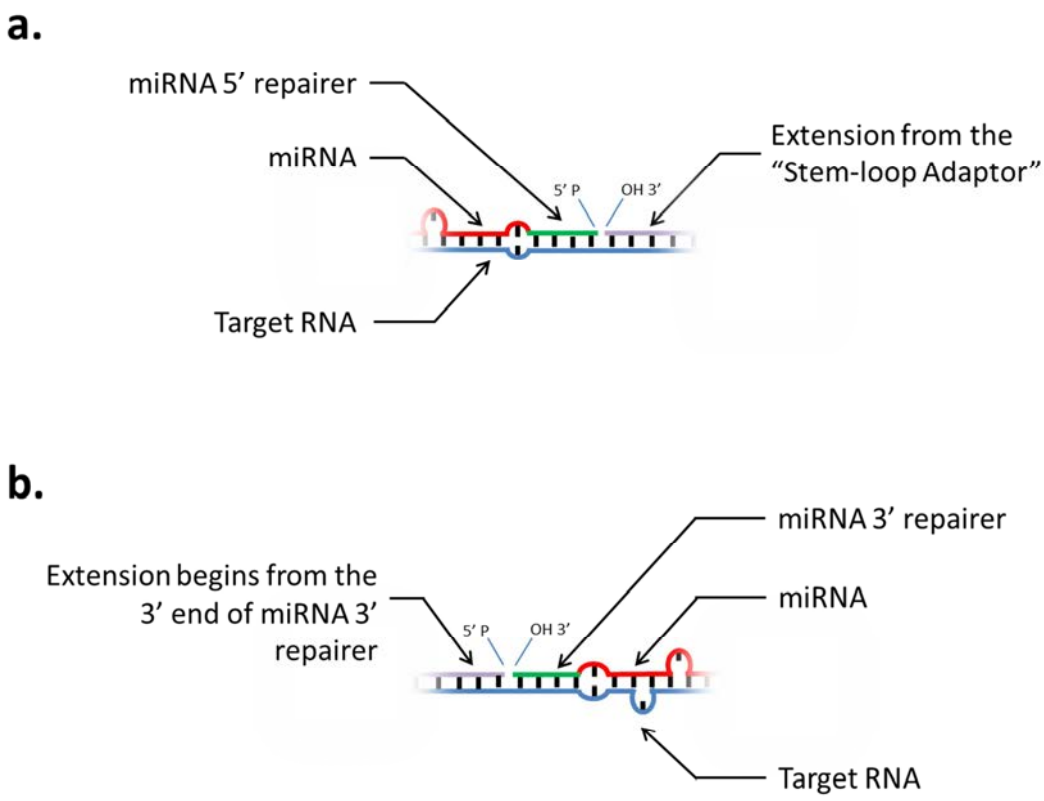


Figure 4.14.1 Examples of two special nicks that require sealing. A nick refers to a break in a double-stranded nucleic acid duplex where adjacent nucleotides of one strand are not linked by a phosphodiester bond. (a) The figure illustrates the nick between the 5’ end (with phosphate) of the miRNA 5’ repairer (green) and the 3’ end (with hydroxyl) of the extension from the “Stem-Loop Adaptor (RNA)” (purple). (b) The figure illustrates the nick between the 3’ end (with hydroxyl) of the miRNA 3’ repairer (green) and the 5’ end (with phosphate) the short RNA oligo (purple) for extending the 3’ end of the miRNA 3’ repairer.

One issue found in the ligation of miRNA repairers to the 5’ and 3’ ends of the miRNA also occurs in the ligation of these two special nicks as illustrated in Figure 4.14.1. This is whether the 5-nitroindole is a suitable substrate for T4 RNA Ligase 2. Two sets of mock experiments were designed for evaluating this issue.

First, the repair of the nick as illustrated in Figure 4.14.1a was evaluated with two sets of mock duplexes. An RNA oligo with stem-loop structure and recessed 3’ end was used as a template (Repaired 5’ end Template, Table 4.2.1 and Figure 4.14.2a). Two other RNA oligos with 5-nitroindole close to the 5’ end, which could anneal to the 5’ protruding single-strand part of the “Repaired 5’ end Template”, were used to mimic the repaired 5’ end of miRNA (Figure 4.14.2a). One of these 5-nitroindole-containing oligo was “Repaired 5’ end (GUXG-)”, which contained a 5-nitroindole base (X) at the third position from its 5’ end (Table 4.2.1). The other 5-nitroindole-containing oligo was “Repaired 5’ end (GXAG-)”, which contained a 5-nitroindole base (X) at the second position from its 5’ end (Table 4.2.1). These three oligos could form two different kinds of nicks mimicking the special nick shown in Figure 4.14.1a. One of these had a 5-nitroindole base at the third position from the *donor (5’) end* while the other had a 5-nitroindole base at the second position from the *donor (5’) end*. Ligation was evaluated for these two different nicks (Figure 4.14.2a).

Second, the repair of the nick as illustrated in Figure 4.14.1b was also evaluated with two sets of mock duplexes. An RNA oligo with stem-loop structure and recessed 5’ end was used as a template (Repaired 3’ end Template, Table 4.2.1 and Figure 4.14.2b). Two other RNA oligos with 5-nitroindole close to the 3’ end, which could anneal to the 3’ protruding single-strand part of the “Repaired 3’ end Template”, were used to mimic the repaired 3’ end of miRNA (Figure 4.14.2b). One of these 5-nitroindole-containing oligo was “Repaired 3’ end (-XXAG)”, which contained two 5-nitroindole bases (XX) at the third and fourth

positions from its 3’ end (Table 4.2.1). The other 5-nitroindole-containing oligo was “Repaired 3’ end (-XXXG)”, which contained three 5-nitroindole bases (XXX) at the second to fourth positions from its 3’ end (Table 4.2.1). These three oligos could form two different kinds of nicks mimicking the special nick shown in Figure 4.14.1b. One of these had two 5-nitroindole bases at the third and fourth positions from the *acceptor (3’) end* while the other had three 5-nitroindole bases at the second to fourth positions from the *acceptor (3’) end*. Ligation was evaluated for these two different nicks (Figure 4.14.2b).

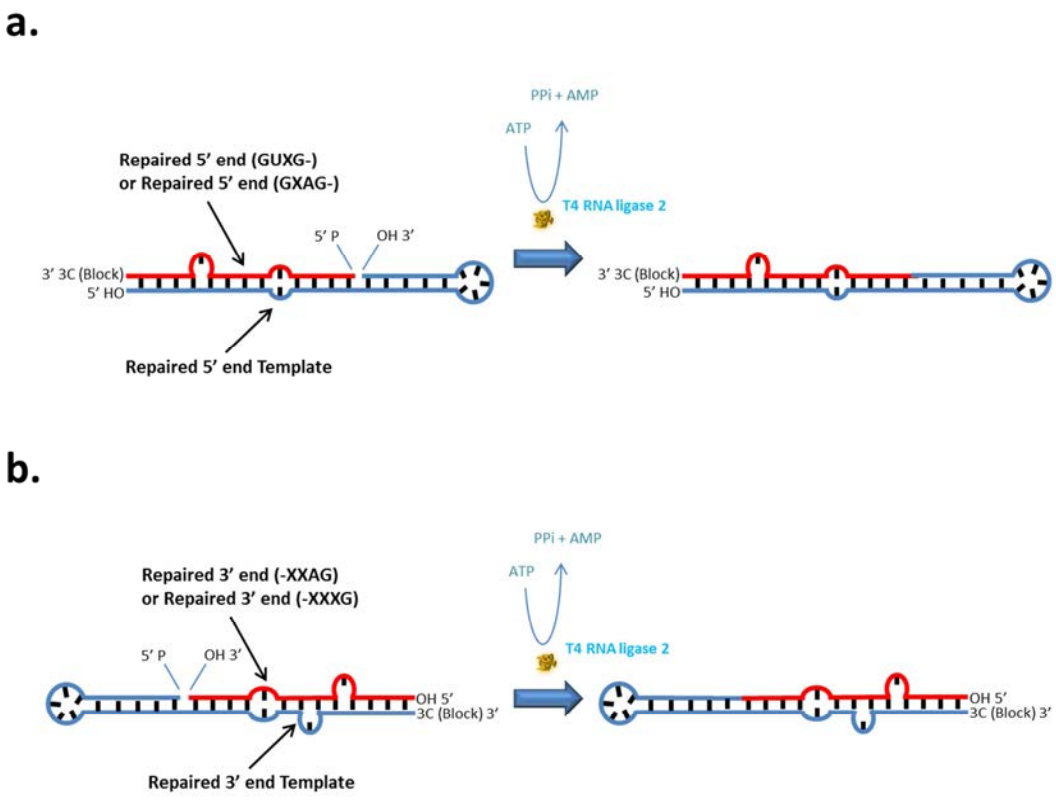


Figure 4.14.2 The scheme of the mock experiments mimicking the ligation of two different kinds of nicks that are generated after the repair of the 5′ and 3′ ends of mismatched miRNA. (a) The figure shows the ligation of the nick between the 5′ end (with phosphate) of the miRNA 5′ repairer (red) and the recessed 3′ end (with hydroxyl) of the "Stem-Loop Adaptor (RNA)" (blue). (b) The figure shows the ligation of the nick between the 3′ end (with hydroxyl) of the miRNA 3′ repairer (red) and the 5′ end (with phosphate) of the short RNA oligo (blue) for extending the 3′ end of the miRNA 3′ repairer.

4.14.1 Preparation of the mock duplexes

Appendix C Table 15 shows the preparation of annealing reaction mixtures. Briefly, an aliquot of 0.2 µl "Repaired 5' end Template" (100 µM) and an aliquot of 0.2 µl of either "Repaired 5' end (GUXG-)" (100 µM) or "Repaired 5' end (GXAG-)" (100 µM) were mixed in a PCR tube for each reaction for mimicking the first kind of nick illustrated in Figure 4.14.1a. Similarly, an aliquot of 0.2 µl "Repaired 3' end Template" (100 µM) and an aliquot of 0.2 µl of either "Repaired 3' end (-XXAG)" (100 µM) or "Repaired 3' end (-XXXG)" (100 µM) were mixed in a PCR tube for each reaction for mimicking the second kind of nick illustrated in Figure 4.14.1b.

The annealing of the mock duplexes was carried out on Veriti® 96-well Thermal Cycler (ABI): 95℃ for 1 minute and 80℃ for 1 minute followed by gradual cooling down to 25℃ at a ramping rate of 1% of the full ramping speed. The reaction was then further cooled down to 4℃ and kept at this temperature until use.

4.14.2 Nick repair

Appendix C Table 16 shows the preparation of ligation mixtures for evaluating the repair of the nicks. Briefly, the setup started with preparing a master mix that did not contain the annealed duplex in a clean tube. For each reaction of this master mix, an aliquot of 7.1 µl of water was first added to the tube. Then, 2 µl of 5X NEBNext® Quick Ligation Reaction Buffer (NEB) and 0.5 µl of T4 RNA Ligase 2 (dsRNA Ligase) (10U/ µl, NEB) were transferred into the tube sequentially. The master mix was mixed and spun down briefly. Next, the master mix was chilled on ice until the temperature was equilibrated (preventing denaturation of the duplex at room temperature). The 0.4-µl annealed duplex was then added to the 9.6-µl master mix to obtain the 10-µl reaction mixture (The annealed duplexes and master mixes

could be scaled up proportionally if multiple reactions were needed in parallel). The mixture was mixed and spun down briefly. Finally, the tubes were immediately incubated in the thermal block of Veriti® 96-well Thermal Cycler (ABI).

A time-course study was conducted for evaluating the repair to all the nicks under study. The incubation conditions were 4℃ for 0, 2, 4, 8, 16, 24 and 36 hours. All the reactions were inactivated at 65℃ for 20 minutes.

4.14.3 RNA denaturing gel electrophoresis, staining and visualization

Hoefer Electrophoresis Unit SE600 (Hoefer) equipped with PolyScience® Digital Temperature Controller (PolyScience) was used to run the denaturing/urea-formamide polyacrylamide gel electrophoresis (Urea-formamide-PAGE). All the samples were mixed with 2X Urea-PAGE Loading Buffer (Appendix A) in a 1:1 ratio and heat-denatured at 90℃ for 5 minutes. The samples were then chilled on ice immediately for at least 1 minute. The loading volume of each sample was 10 µl (5 µl sample + 5 µl 2X Urea-PAGE Loading Buffer). The urea-formamide polyacrylamide gels used in this Urea-formamide-PAGE were 7.5 M urea, 30% formamide, 10% TBE-polyacrylamide gel (Appendix B). Gels were setup in SE6102 Glass Plates (PR) (18 cm width X 16 cm length, Hoefer) with 1 mm spacers according to the instructions from the manufacturer. Gels were allowed to polymerize for at least 2 hours. Electrophoresis was pre-run under constant voltage at 150 V for 30 minutes at 50℃ in 1X TBE Buffer. Then, the samples were loaded into the wells of the gel and run under constant voltage at 170V for 2.5 hours at 50℃ in 1X TBE Buffer. The electric power was supplied by PowerPac™ Basic Power Supply (Bio-Rad). After electrophoresis, the gel was post-stained for 5 to 10 minutes in 0.5X SYBR® Gold nucleic acid gel stain (Life

Technologies) diluted in 1X TBE Buffer. The image was then captured by ChemiDoc™ MP Imaging System (Bio-Rad) with Image Lab™ Software version 4.1.

4.15 The repair of the gap for joining the interacting miRNA and its target RNA together

After the ligation of the "Stem-Loop Adaptor" to the 3' end of the miRNA target, the 5' end of the miRNA and the 3' end of the miRNA target in the same duplex are aligned in a head-to-tail position with a gap in between. This gap has to be repaired for joining the two interacting RNAs together.

Two possible approaches are potentially able to fill this gap for joining the two interacting RNAs. The first approach is "gap repair by synthesis" while the second approach is "gap repair by ligation" (these two approaches will further be discussed in Section 6.9). The current study adopted the approach involving ligation of short RNA oligos for filling up the gap (gap repair by ligation, Figure 4.15.1). As the sequence in the gap is not known, the filler RNA oligo must be of random sequence. Currently, T4 RNA ligase 2 (NEB) is used in this approach to ligate RNA random tetramers (5' phosphate group-NNNN-hydroxyl group 3') to both of the RNA ends in the gap repeatedly until the gap is completely filled or remains as a three-base gap, two-base gap or one-base gap. This is because a three-base gap, two-base gap and one-base gap cannot be filled by a tetramer oligo (Figure 4.15.1). The T4 RNA ligase 2 (NEB) used here preferentially catalyzes the joining of double-stranded RNA with blunt end or cohesive end, and the repair of nicks in double-stranded RNA. On the other hand, tetramer RNA oligo is the shortest RNA oligo that is commercially available at IDT. Therefore, RNA random tetramer has been chosen as the filler oligo.

Furthermore, T4 DNA Ligase (NEB) and nucleoside monophosphates (NMPs — a mixture of adenosine 5'-monophosphate, uridine 5'-monophosphate, cytidine 5'-monophosphate and guanosine 5'-monophosphate in a ratio of 1:1:1:1), which substituted T4 RNA Ligase 2 and RNA random tetramer in the mock experiment respectively, were tested as well.

A mock experiment was designed to evaluate the approach of “gap repair by ligation”. Two RNA oligos, one mimicking miRNA (labeled with FAM at its 3’ end) and another mimicking target RNA after ligation of stem-loop adaptor to its 3’ end (labeled with HEX at its 5’ end), were first annealed to generate a gap for investigating the gap repair. The two RNA oligos were Artificial Mir 5 (FAM) (23 nt) and Artificial Binding Site (RNA, HEX) (82 nt) (Table 4.2.1). The gap between the two annealed sequences was a 16-base gap with random sequence (Figure 4.15.2). The gap could be filled by four RNA tetramers in stoichiometric manner. After the gap was fixed, a joined sequence (121 nt) with both fluorescent labels (FAM and HEX) was generated.

All the fluorescent-labeled sequences were different in length, and could be separated by gel electrophoresis and detected by any gel-imaging system supporting fluorescence detection. Figure 4.15.2 shows the experimental scheme for evaluating the approach of “gap repair by ligation”.

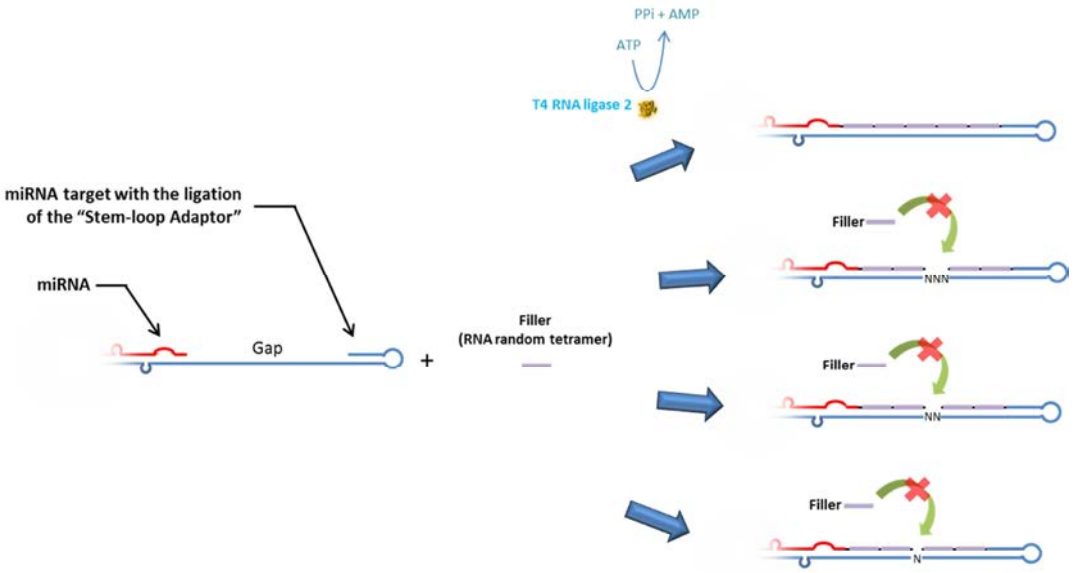


Figure 4.15.1 An example of repairing the gap between the 5' end of an miRNA and the 3' end of the miRNA target in the same duplex by the ligation approach. The length of the gap varies since the digestion of RNA by RNase is a random process. Since RNA random tetramer is selected as the filler in the current approach, four possibilities can occur. First, the gap can be completely filled (the top result at the right-hand-side of the figure). Second, a three-base gap remains and this gap cannot be filled by RNA tetramer (the second result at the right-hand-side of the figure). Third, a two-base gap remains and this gap cannot be filled by RNA tetramer (the third result at the right-hand-side of the figure). Fourth, a one-base gap remains and this gap cannot be filled by RNA tetramer (the bottom result at the right-hand-side of the figure).

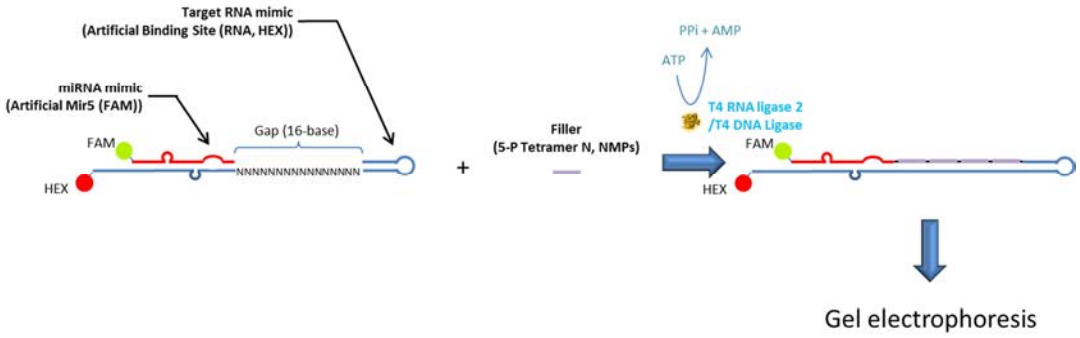


Figure 4.15.2 The scheme of the mock experiments mimicking the repair of the gap between the 5' end of an miRNA and the 3' end of the miRNA target in the same duplex by the ligation approach. A 16-base gap with random sequence was generated for the evaluation. This gap was constructed by a duplex composed of an miRNA mimic (23 nt and labeled with FAM) and a target RNA mimic (82 nt and labeled with HEX). RNA random tetramer was used to repair this gap and generate a 121-nt RNA sequence with both FAM and HEX labels. The ligation was evaluated under different conditions.

4.15.1 Preparation of the gap

Appendix C Table 17 shows the preparation of annealing reaction mixtures. Briefly, an aliquot of 0.1 µl Artificial Mir 5 (FAM) (23 nt, 100 µM) and an aliquot of 0.1 µl of Artificial Binding Site (RNA, HEX) (82 nt, 100 µM) were mixed in a PCR tube for each reaction for mimicking the 16-base gap illustrated in Figure 4.15.2. Annealing of the mock duplexes was carried out on Veriti® 96-well Thermal Cycler (ABI): 95℃ 1 minute and 80℃ 1 minute followed by gradual cooling down to 25℃ at a ramping rate of 1% of the full ramping speed. The reaction was then further cooled down to 4℃ and kept at this temperature until use.

4.15.2 Gap repair

Seven ligation conditions were evaluated in a pilot study. All the tested conditions were conducted in 1X NEBNext® Quick Ligation Reaction Buffer (66mM Tris-HCl, 10mM $MgCl_2$, 1mM DTT, 1mM ATP, 6% Polyethylene glycol (PEG 6000), pH 7.6 at 25°C; NEB). In the 1st condition, the filler for repairing the gap was 100 µM 5-P Tetramer N (5' phosphate group-NNNN-hydroxyl group 3') and the ligation was achieved by 5 U of T4 RNA Ligase 2 (dsRNA Ligase) (NEB). The concentration of the gap was 2 µM (2 µM Artificial Mir 5 (FAM) + 2 µM Artificial Binding Site (RNA, HEX)). Thus, the ratio of the number of the filler required for filling up the gap (each 16-base gap required 4 RNA tetramer molecules) to the filler available was 1:12.5. In the 2nd condition, the filler for repairing the gap was 200 µM 5-P Tetramer N and the ligation was achieved by 5 U of T4 RNA Ligase 2 (dsRNA Ligase). The concentration of the gap was 1 µM (1 µM Artificial Mir 5 (FAM) + 1 µM Artificial Binding Site (RNA, HEX)). Thus, the ratio of the number of the filler required for filling up the gap to the filler available was 1:50. In the 3rd condition, the filler for repairing the gap was 200 µM 5-P Tetramer N and the ligation was achieved by 1000 U of T4 DNA Ligase (NEB). The

concentration of the gap was 1 µM. The ratio of the number of the filler required for filling up the gap to the filler available was 1:50. In the 4th condition, the filler for repairing the gap was 1 mM NMPs (nucleoside monophosphates — a mixture of adenosine 5'-monophosphate, uridine 5'-monophosphate, cytidine 5'-monophosphate and guanosine 5'-monophosphate in a ratio of 1:1:1:1, 1 mM of each NMP) and the ligation was achieved by 5 U of T4 RNA Ligase 2 (dsRNA Ligase). The use of NMPs as the filler in this examined condition was used to study whether the T4 RNA Ligase 2 could utilize monomer N (ribonucleoside monomers) to repair the gap. The concentration of the gap was 1 µM. In the 5th condition, the filler for repairing the gap were 200 µM 5-P Tetramer N and 1 mM NMPs and the ligation was achieved by 5 U of T4 RNA Ligase 2 (dsRNA Ligase). The use of 5-P Tetramer N and NMPs as the filler in this examined condition was used to study whether the repair could be enhanced. The concentration of the gap was 1 µM. In the 6th condition, the filler for repairing the gap was 1 mM NMPs and the ligation was achieved by 1000 U of T4 DNA Ligase. The concentration of the gap was 1 µM. In the 7th condition, the filler for repairing the gap were 200 µM 5-P Tetramer N and 1 mM NMPs and the ligation was achieved by 1000 U of T4 DNA Ligase. The concentration of the gap was 1 µM.

Appendix C Table 18 shows the preparation of reaction mixtures of the above conditions for evaluating the gap repair. Briefly, the setup started with preparing a master mix that did not contain the duplex with gap, and ligase in a clean tube. For each reaction of this master mix, an aliquot of 4.3-6.1 µl of water was first added to the tube. Then, 2 µl of 5X NEBNext® Quick Ligation Reaction Buffer (NEB), 0-2 µl of 5-P Tetramer N (1000 µM) and 0-1 µl of NMPs (10 mM) were added to the vial sequentially. The master mix was mixed well and spun down briefly. Next, the master mix was chilled on ice until the temperature was equilibrated (preventing denaturation of the duplex at room temperature). An aliquot of 0.2-0.4-µl annealed duplex, which was prepared as described in Section 4.15.1 was then

added to the 9.1-9.3-µl master mix to obtain a 9.5-µl reaction mixture (The annealed duplexes and master mixes could be scaled up proportionally if multiple reactions were needed in parallel). The mixture was mixed and spun down briefly. Then, the tubes were immediately incubated in the thermal block of a Veriti® 96-well Thermal Cycler (ABI) at 4℃ for 30 minutes to allow the filler to hybridize to the template within the gap for some time. Finally, an aliquot of 0.5 µl of ligase (T4 RNA Ligase 2 (10 U/µl, NEB) or T4 DNA Ligase (2000 U/µl, NEB)) was added to the 9.5-µl master mix to obtain the 10-µl final reaction mixture. The mixture was mixed and spun down briefly again. Then, the tubes were immediately incubated in the thermal block of a Veriti® 96-well Thermal Cycler (ABI). The incubation condition was 16℃ for 36 hours. All the reactions were stopped by inactivating the enzyme at 65℃ for 20 minutes.

In addition, a time-course study for further investigating the optimal ligation time in the selected ligation condition was conducted. Five time points were investigated: 0, 3, 6, 12 and 18 hours. The experiment was repeated at 4℃ and 16℃.

Appendix C Table 19 shows the preparation of reaction mixtures for the above time-course study. Briefly, the setup started with preparing a master mix that did not contain the duplex with gap, and ligase in a clean tube. For each reaction of this master mix, an aliquot of 4.3 µl of water was first added to the tube. Then, 2 µl of 5X NEBNext® Quick Ligation Reaction Buffer (NEB) and 2 µl of 5-P Tetramer N (1000 µM) were added to the vial sequentially. The master mix was mixed and spun down briefly. Next, the master mix was chilled on ice until the temperature was equilibrated (preventing denaturation of the duplex at room temperature). An aliquot of 0.2-µl annealed duplex, which was prepared as described in Section 4.15.1 was then added to the 8.3-µl master mix to obtain an 8.5-µl reaction mixture (The annealed duplexes and master mixes could be scaled up

proportionally if multiple reactions were needed in parallel). The mixture was mixed and spun down briefly. Then, the reactions for the time points were immediately incubated in the thermal block of a Veriti® 96-well Thermal Cycler (ABI) at 4°C for 30 minutes to allow the filler to hybridize to the template within the gap for some time. Finally, an aliquot of 1.5 µl of 1/3X diluted T4 RNA Ligase 2 (10 U/µl, NEB), which was equivalent to 5 U of T4 RNA Ligase 2, was added to each of the 8.5-µl master mix to obtain the 10-µl final reaction mixture. The diluted T4 RNA Ligase 2 allowed much more accurate pipetting. The mixture was mixed and spun down briefly again. Then, the tubes were immediately incubated in the thermal block of a Veriti® 96-well Thermal Cycler (ABI). The incubation conditions were 16°C for 0, 3, 6, 18 and 18 hours, and 4°C for 0, 3, 6, 18 and 18 hours. All the reactions were stopped by inactivating the enzyme at 65°C for 20 minutes.

4.15.3 RNA denaturing gel electrophoresis, staining and visualization

The RNA denaturing gel electrophoresis procedure was the same as described in Section 4.14.3 with the exception of using 2X Urea-PAGE Loading Buffer (Bromophenol Blue-free)(Appendix A) instead of 2X Urea-PAGE Loading Buffer for the sample loading buffer. The absence of the bromophenol blue in the sample loading buffer can prevent the interference of the dye in reading the result. After electrophoresis, the gel image was scanned by Typhoon 9400 Variable Mode Imager (Amersham Biosciences) with Typhoon Scanner Control version 5.0 (Amersham Biosciences). The image was extracted by ImageQuant™ TL (Amersham Biosciences).

The gel was then counterstained by SYBR® Gold nucleic acid gel stain (Life Technologies). The gel was post-stained for 5 to 10 minutes in 0.5X SYBR® Gold nucleic acid gel stain

diluted in 1X TBE Buffer. The image was then captured by ChemiDoc™ MP Imaging System (Bio-Rad) with Image Lab™ Software version 4.1.

4.16 Reverse transcription of RNA templates with 5-nitroindole base

Reverse transcription is the last step of the library preparation in RIINGS. This step actually converts the RNA chimeras, which represent the miRNA-target RNA interacting pairs, into cDNA. The product cDNA library can be sequenced by next-generation sequencing and the miRNA interactome can be finally revealed. However, an important issue in the conversion must be considered.

A pair of mismatch repairers is actually ligated to the 5' and the 3' ends of the miRNA for repairing the mismatches in those two ends (see Section 4.13 and Figure 4.13.1). The results of the ligation of the two repairers indicate that they can be efficiently ligated to the 5' and the 3' ends of miRNA with 0 to 3 mismatches (see the results in Section 5.11.2). Indeed, these two repairers are 5-nitroindole base-containing short RNA oligos. Accordingly, the RNA chimeras also contain 5-nitroindole bases (upstream of miRNA 5' end and downstream of miRNA 3' end). The 5-nitroindole base is one of the artificial bases that can pair up with all four natural bases non-selectively (see Section 6.8 for more information of the use of 5-nitroindole in RIINGS). However, we do not know whether reverse transcriptases (i.e. RNA-dependent DNA polymerases) are able to use 5-nitroindole base as the substrate to synthesize cDNA.

Two mock experiments were designed for investigating the above concern. Two 5-nitroindole base-containing RNA templates were designed. One was used to evaluate

whether the polymerase could deal with successive 5-nitroindole bases. From 3’ to 5’, the template contained 5-nitroindole bases at position 26, positions 35-36 and positions 46-48 (see Figure 4.16.1). The full length of this sequence was 52 nt and was named 5-NitInd RT temp 1 (Table 4.2.1). This template was used to test the ability of the polymerase in dealing with one, two successive and three successive 5-nitroindole bases. After the confirmation of only one 5-nitroindole base substitution in the template RNA could be utilized by *Bst* 3.0 DNA polymerase in the above experiment (see result in Section 5.14.1), an additional mock experiment was performed to further study the real situation. A second RNA template, which contained 5-nitroindole bases at positions 22 and 44 (3’ to 5’) (see Figure 4.16.1), was designed for this purpose. This was a template actually mimicking two single 5-nitroindole bases included in the 5’ repairer and the 3’ repairer. The full length of this sequence was 50 nt and was named 5-NitInd RT temp 2 (Table 4.2.1). Ten polymerases with RNA-dependent DNA polymerase activity from different companies were tested. Figure 4.16.1 shows the experimental scheme for evaluating whether the polymerases could utilize 5-nitroindole base as the template for synthesizing cDNA.

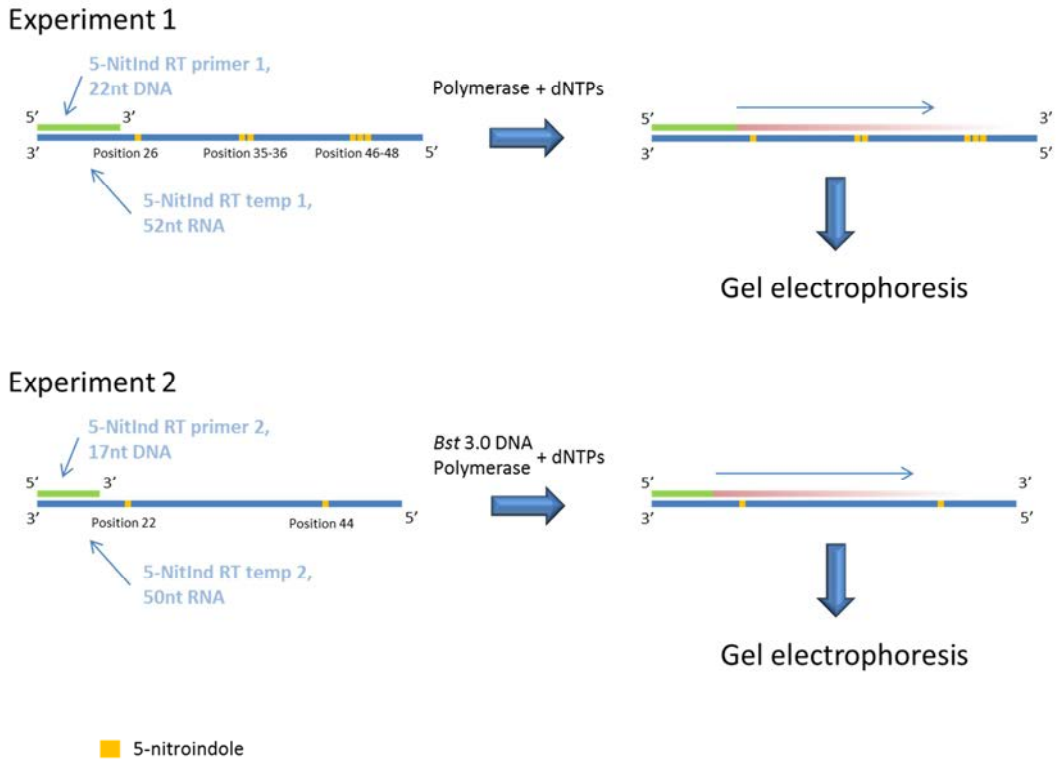


Figure 4.16.1 The scheme of the mock experiments for evaluating whether the polymerases can utilize 5-nitroindole base as the template for synthesizing cDNA. In experiment 1, 5-NitInd RT temp 1 was used as the template to evaluate whether the candidate polymerases could deal with a template that contained successive 5-nitroindole bases. One to three successive 5-nitroindole bases were evaluated. In experiment 2, 5-NitInd RT temp 2 was used as a template to mimic the real situation. *Bst* 3.0 DNA polymerase was tested in this experiment.

4.16.1 Reverse transcription with 5-nitroindole-containing template

4.16.1.1 Experiment 1 (for evaluating whether the candidate polymerases could deal with successive 5-nitroindole)

Ten polymerases with RNA-dependent DNA polymerase activity from different companies were tested. These polymerases were **(i)** *Tth* DNA Polymerase (5U/µl, Promega, M2101), **(ii)** MMLV (200U/µl, Promega, M1705), **(iii)** MMLV RNaseH-, Point Mutant (200U/µl, Promega, M3681), **(iv)** MMLV (200U/µl, Invitrogen, 28025-021), **(v)** SuperScript II (200U/µl, Invitrogen, 18064-014), **(vi)** SuperScript III (200U/µl, Invitrogen, 18080-044), **(vii)** Transcriptor High Fidelity Reverse Transcriptase (20U/µl, supplied with Transcriptor High Fidelity cDNA Synthesis Kit, Roche, 05081963001), **(viii)** Cloned AMV Reverse Transcriptase (15U/µl, Invitrogen, 12328-019), **(ix)** AMV Reverse Transcriptase (25U/µl, USB, 70041Y), **(x)** *Bst* 3.0 DNA Polymerase (8U/µl, NEB, M0374S). Basically, all the candidate polymerases were tested in both 5X NEBNext® Quick Ligation Reaction Buffer (66mM Tris-HCl, 10mM $MgCl_2$, 1mM DTT, 1mM ATP, 6% Polyethylene glycol (PEG 6000), pH 7.6 at 25℃; NEB) and CutSmart® Buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/ml BSA, pH 7.9 at 25℃; NEB). *Tth* DNA Polymerase was additionally tested in the RT 10X Buffer (10 mM Tris-HCl, 90 mM KCl, pH 8.3 at 25℃; Promega) supplied with the enzyme. *Bst* 3.0 DNA Polymerase was additionally tested in the 10X Isothermal Amplification Buffer II (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 150 mM KCl, 2 mM $MgSO_4$, 0.1% Tween® 20, pH 8.8 at 25℃; NEB) supplied with the enzyme. Two concentrations of dNTPs were evaluated: 0.5 and 2 mM (each dNTP). Five concentrations of $MnCl_2$ were evaluated: 0, 1, 2, 3 and 5 mM.

Appendix C Table 20 shows the preparation of the reverse transcription reaction mixtures. For each reaction, briefly, an aliquot of 5.27-9.02 µl of water was first added to a clean tube. Then, 0-0.625 µl of $MnCl_2$ (100 mM), 0.625-2.5 µl of dNTPs, 2.5 µl of 5X buffer or 1.25 µl of 10X buffer (the buffer was one of the four buffers tested: **(1)** 5X NEBNext® Quick Ligation Reaction Buffer, **(2)** CutSmart® Buffer (10X), **(3)** RT 10X Buffer (tested with *Tth* DNA Polymerase only), **(4)** 10X Isothermal Amplification Buffer II (tested with *Bst* 3.0 DNA Polymerase only)), 0.3 µl of 5-NitInd RT temp 1 and 0.3 µl of 5-NitInd RT primer 1 (Table 4.2.1) were added to the vial sequentially. The master mix was mixed and spun down briefly. An aliquot of 1 µl polymerase was finally added. The polymerase was one of the ten polymerases tested (see above for details. The mixture was mixed and spun down again. All the reactions were incubated in the thermal block of a Veriti® 96-well Thermal Cycler (ABI).

The incubation conditions were 25℃ for 10 minutes followed by **(a)** 55℃ for 1, 2, 4, 8, 16, 24 and 36 hours (for polymerase **(i)**, **(iii)**, **(vii)**, **(viii)** and **(x)** only), **(b)** 42℃ for 1, 2, 4, 8, 16, 24 and 36 hours (for polymerase **(ii)**, **(iv)**, **(v)** and **(ix)** only), and **(c)** 50℃ for 1, 2, 4, 8, 16, 24 and 36 hours (for polymerase **(vi)** only). All the reactions were stopped by heat inactivation at 85℃ for 5 minutes (polymerase **(i)** cannot be heat-inactivated and reactions with this polymerase were frozen at -20℃ immediately). Finally, all the samples were stored at -20℃ until gel electrophoresis. An extra reaction for time 0 (without incubation but with heat inactivation) was performed as well.

4.16.1.1.1 Experiment 2 (for further studying the real situation)

As only the *Bst* 3.0 DNA Polymerase (NEB, M0374S) had the ability in dealing with 5-nitroindole base (single substitution, but not successive substitutions), experiment 2 tested this polymerase only.

Appendix C Table 21 shows the preparation of the reverse transcription reaction mixtures. For each reaction, briefly, an aliquot of 4.65 µl of water was first added to a clean tube. Then, 1.25 µl of $MnCl_2$ (100 mM), 2.5 µl of dNTPs, 2.5 µl of 5X NEBNext® Quick Ligation Reaction Buffer (NEB), 0.3 µl of 5-NitInd RT temp 2 and 0.3 µl of 5-NitInd RT primer 2 (Table 4.2.1) were added to the vial sequentially. The master mix was mixed and spun down briefly. An aliquot of 1 µl *Bst* 3.0 DNA Polymerase (8U/µl, NEB, M0374S) was finally added. The mixture was mixed and spun down again. All the reactions were incubated in the thermal block of a Veriti® 96-well Thermal Cycler (ABI).

The incubation conditions were 25℃ for 10 minutes followed by 50℃ for 1, 2, 4, 8, 16 and 24 hours. All the reactions were stopped by inactivation at 80℃ for 5 minutes and stored at -20℃ until gel electrophoresis. An extra reation for time 0 (without incubation but with heat inactivation) was performed as well.

4.16.2 RNA denaturing gel electrophoresis, staining and visualization

The procedures for gel electrophoresis, staining and visualization were the same as those described in Section 4.14.3.

5. Results

5.1 Comparison of AGO expression level in cell-lines and the yield of AGO in different extraction buffers

Two different non-denaturing protein extraction buffer were tested for the yield of AGO proteins (see Figure 4.3.1). The yield was tested with three different cell-lines. The three cell-lines were MEC-1, K562 and Raji. Equal amounts of total protein lysates were loaded for SDS-PAGE. Figure 5.1.1 summarizes the results from the series of experiments.

The western blotting result showed that the level of AGO proteins in the total protein lysate extracted by 1X Detergent-free Lysis Buffer was significantly higher than that extracted by 1X Detergent-containing Lysis Buffer (with 1% NP-40). This result was reproduced in all three different cell-lines. There was 49.20% (p = 0.046) increment in MEC-1 total protein lysate, 50.09% (p = 0.027) increment in K562 total protein lysate and 58.32% (p = 0.039) increment in Raji total protein lysate when the total protein lysate was extracted by 1X Detergent-free Lysis Buffer. These results indicated that the 1X Detergent-free Lysis Buffer yielded higher amount of AGO proteins in total protein extraction. Figure 5.1.1a shows the western blotting result and Figure 5.1.1b to Figure 5.1.1d show the corresponding intensity charts of the western blotting result for MEC-1, K562 and Raji cells respectively. Thus, 1X Detergent-free Lysis Buffer was selected for protein extraction in RIINGS.

On the other hand, the comparison of the expression of AGO proteins among MEC-1, K562 and Raji cells showed that K562 cells expressed the highest amount of AGO proteins, MEC-1 cells were the second and Raji cells expressed the lowest amount of AGO proteins (Figure

5.1.1a). The expression of AGO protein in K562 cells was 1.30-fold higher than in MEC-1 cells (p = 0.036) and 2.91-fold higher than in Raji cells (p = 0.00024) when the total protein lysates were prepared in 1X Detergent-free Lysis Buffer. Compared to the MEC-1 cells and Raji cells, K562 cells were the most suitable for the current method development. Figure 5.1.1e shows the level of AGO proteins in MEC-1, K562 and Raji cell lysates.

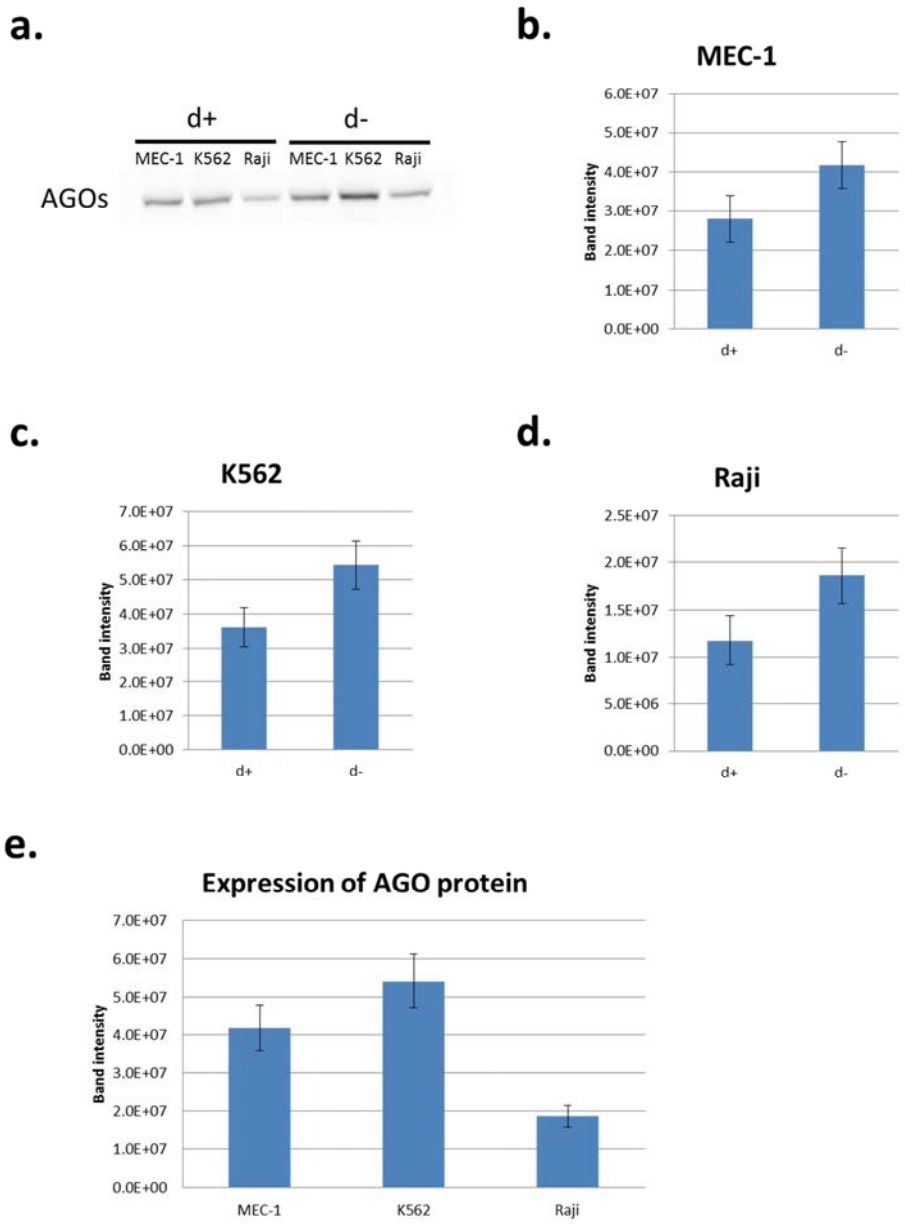


Figure 5.1.1 The AGO protein levels in total protein lysates extracted from three different cell lines with two selected non-denaturing extraction buffers. (a) The AGO proteins (~100 kDa) were examined by western blotting. This was the western blotting result of MEC-1, K562 and Raji cell lysates that were prepared in either 1X Detergent-containing Lysis Buffer (d+) or 1X Detergent-free Lysis Buffer (d-). (b) to (d) The comparison of the level of AGO proteins in cell lysates extracted from (b) MEC-1, (c) K562 and (d) Raji cell with either 1X Detergent-containing Lysis Buffer (d+) or 1X Detergent-free Lysis Buffer (d-). (e) The expression level of AGO proteins in three different cell-lines. The total protein lysates were prepared in 1X Detergent-free Lysis Buffer. All the bar charts were plotted based on the mean value ± SD (n = 3). The intensity of each band was quantified (in arbitrary unit) by Image Lab™ Software version 4.1.

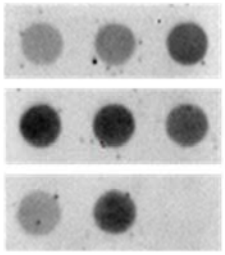
5.2 Comparison of three anti-AGO antibodies in AGO binding ability

Three different anti-AGO antibodies were tested for their ability and specificity of binding the AGO proteins in three different K562 total protein lysates each with a different denaturing status (see Figure 4.4.1). The three tested antibodies were Anti-pan Ago Antibody, clone 2A8 (Millipore), eIF2C Antibody (B-3) (Santa Cruz) and eIF2C2 (4F9) (Santa Cruz). The three tested protein lysates were denatured protein lysate, non-denatured protein lysate, and denatured protein lysate with buffer exchange to non-denaturing buffer. Dot blotting analysis was used to evaluate which antibody specifically bound the non-denatured AGO proteins and also which antibody had the highest binding ability for binding the non-denatured AGO proteins. The dot blotting result showed the ability of the three antibodies to bind AGO proteins of different denaturing status. Figure 5.2.1 summarizes the results of this experiment.

The result revealed that the ability of Anti-pan Ago Antibody, clone 2A8 to bind the non-denatured AGO proteins was significantly lower than that of eIF2C Antibody (B-3) by 101.70% (p = 0.036) and also significantly lower than that of eIF2C2 (4F9) by 79.43% (p = 0.0076). On the other hand, eIF2C2 (4F9) specifically bound non-denatured AGO proteins. The ability of eIF2C2 (4F9) to bind the AGO proteins in non-denatured protein lysate was significantly higher than that to bind the AGO proteins in denatured protein lysate by 2.37-fold (p = 0.00017). In addition, this antibody did not bind the AGO proteins in denatured protein lysate with buffer exchange to non-denaturing buffer. However, eIF2C Antibody (B-3) did not show any preference in binding the three forms of AGO proteins. There was no significant difference in the binding among the AGO proteins in denatured protein lysate, the AGO proteins in non-denatured protein lysate, and the AGO proteins in denatured protein lysate with buffer exchange to non-denaturing buffer.

Consequently, eIF2C2 (4F9) was the only antibody fulfilling the requirements for RIINGS among the three tested antibodies. This antibody strongly and specifically bound non-denatured AGO proteins.

a.



b.

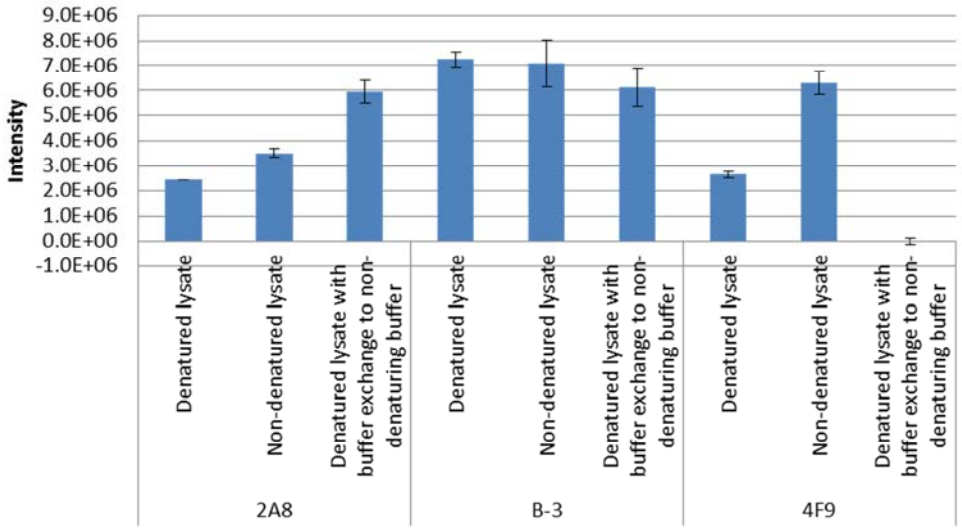


Figure 5.2.1 Dot blotting analysis for evaluating the binding ability and specificity of anti-AGO antibodies against the non-denatured form of AGO proteins. Three different anti-AGO antibodies were tested for their binding to three protein lysates with different denaturing status. (a) The original dot blotting result showed the ability of the three antibodies to bind AGO proteins of different denaturing status. From top to bottom row, they were Anti-pan Ago Antibody, clone 2A8 (Millipore), eIF2C Antibody (B-3) (Santa Cruz) and eIF2C2 (4F9) (Santa Cruz). From left to right column, they were denatured lysate, non-denatured lysate and denatured lysate with buffer exchange to non-denaturing buffer. (b) The bar chart showed the intensities of each dot. The bar chart was plotted based on the mean value ± SD (n = 3). The intensity (in arbitrary unit) of each dot was quantified by Image Lab™ Software version 4.1.

5.3 Immunoprecipitation

The yielding ability of the IP methods was investigated by western blotting while the purity of the yielded AGO proteins was examined by SDS-PAGE with silver staining.

5.3.1 Evaluation of the immunoprecipitation result by western blotting

Two different IP procedures were evaluated. In Method 1, the antibody was first immobilized on the beads and then these beads were used to capture the target protein from the total protein lysate (see Figure 4.5.1). In Method 2, the antibody and the total protein lysate were first incubated together for enough of time to allow the target protein and the antibody to bind each other. Then, beads were used to capture the target protein-antibody complex (see Figure 4.5.1). Western blotting was used to evaluate the yielding ability of these two different methods of immunoprecipitating the AGO proteins. The antibody eIF2C2 (4F9) (Santa Cruz) was used in this experiment (see Section 5.2). Figure 5.3.1 summarizes the results of western blotting.

The result showed that Method 1 and Method 2 yielded 86.97% ± 8.76% and 106.88% ± 15.26% of AGO proteins from the total protein lysates respectively. The yielding ability of Method 2 was significantly higher than that of Method 1 by 1.23–fold (p = 0.022). According to this result, therefore, the IP method 2 was selected for use in RIINGS. Actually, the yield of the AGO complex is very important to RIINGS because the abundance of the AGO complex is expected to be low. We generally know that the abundance of miRNAs (comprising less than 1% of total RNA) and a particular protein (a particular protein vs. total protein) is low. Therefore, an approach that can yield a higher amount of AGO complex is preferred in RIINGS if the purity of the isolated AGO complex is good enough. In fact, the

purity of the isolated AGO complex is pure enough for RIINGS by the approach currently used (this will be found in Sections 5.3.2 and 5.9.1).

Moreover, the input amount of the antibody for IP was also investigated. Two different amounts (7.5 μg and 15 μg) of the antibody were used to initialize the IP experiment. A lesser amount (7.5 μg) of the antibody was used in the pilot study of the current IP experiment. This amount showed consistent yield in the current IP with many replicates (see Figure 5.3.1a; other data not shown). Western blotting was used to evaluate the yielding ability of these two different input amounts of eIF2C2 (4F9) (Santa Cruz). The result showed that the yielding ability of the two initial amounts of antibody did not show any significant difference. With the consideration of the varying amounts of AGO complex may among cells or tissues and in order to ensure that IP could capture the AGO complex as much as possible, the optimal input amount of the eIF2C2 (4F9) was chosen at 11.25 μg per reaction (the average of the tested amounts). This decision also considered the cost involved.

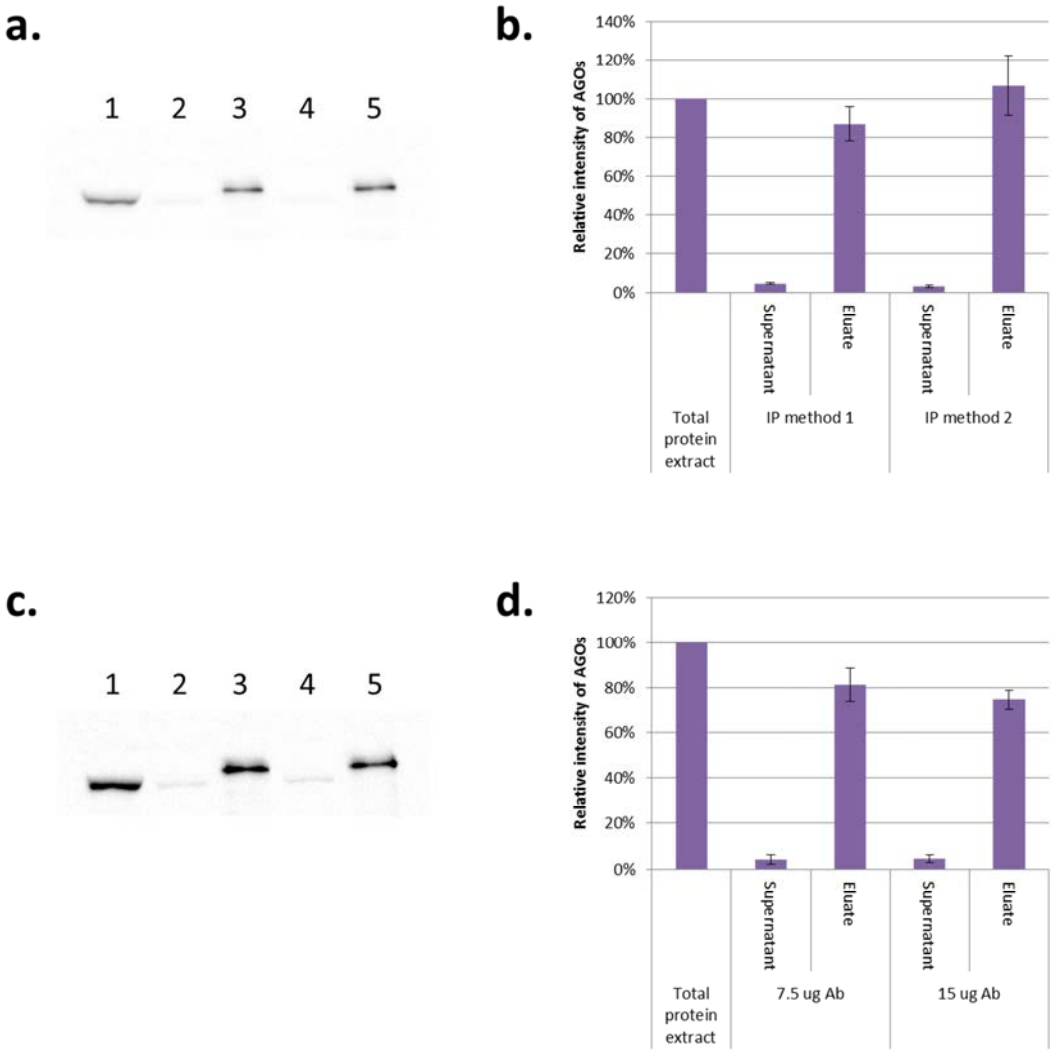


Figure 5.3.1 Western blotting analysis for evaluating immunoprecipitation performed with different modifications in the procedure. (a) Western blotting was used to evaluate the yield of AGO proteins by two different immunoprecipitation (IP) methods. From lane 1 to 5 were total protein extract, supernatant after IP Method 1, eluate from IP Method 1, supernatant after IP Method 2, and eluate from IP Method 2 respectively. (b) A bar chart indicated the corresponding intensity of the protein bands in (a). (c) Western blotting was used to evaluate the yield of AGO proteins by two different input amounts of antibody for IP. IP Method 2 was used in this evaluation. From lane 1 to 5 were total protein extract, supernatant after IP with 7.5 µg antibody, eluate from IP with 7.5 µg antibody, supernatant after IP with 15 µg antibody, and eluate from IP with 15 µg antibody respectively. (d) A bar chart indicated the corresponding intensity of the protein bands in (c). All the bar charts were plotted based on the mean value ± SD (n = 4). The intensity of each band was quantified by Image Lab™ Software version 4.1.

5.3.2 Evaluation of the purity of the yielded AGO proteins by SDS-PAGE with silver staining

The purity of the yielded AGO proteins was first evaluated by SDS-PAGE with silver staining. This method was only able to visualize the proteins, which included the target AGO proteins and other co-purified proteins. However, it was unable to identify the co-purified proteins. Therefore, an in-depth evaluation by using mass-spectrometry was also conducted for identifying the co-purified proteins (results shown in Section 5.9.1).

The proteins in the eluate from the IP were first separated by the SDS-PAGE. Silver staining was then used to visualize protein bands on the gel. Besides assessing the purity of the yielded AGO proteins, the original eIF2C2 (4F9) (Santa Cruz) and the supernatant of the antibody solution after conjugation to the beads from the bead preparation for IP Method 1 were also loaded for the gel electrophoresis. The original antibody was used as a reference for the antibody bands. The supernatant was used to evaluate for the completeness of the antibody conjugation in the bead preparation for IP Method 1.

Figure 5.3.2 revealed the completeness of the bead-antibody conjugation in the bead preparation for IP Method 1 and the purity of the eluates purified by either Method 1 or Method 2. Lane 1 was the protein ladder. Lane 2 was eIF2C2 (4F9) (an IgA antibody). The molecular weight of the two major bands was quite similar to the light chain and the heavy chain (α chain) of an IgA antibody respectively. The expected molecular weight of the light chain and the heavy chain of an IgA antibody was ~25 kDa and ~55 kDa respectively [322]. Lane 3 was the supernatant of the antibody solution after conjugation to the beads from the bead preparation for IP Method 1. The gel picture showed that no heavy chain and light chain of eIF2C2 (4F9) could be detected. Thus, the antibody conjugation in the bead preparation for IP Method 1 was complete. Lane 4 was the total protein lysate of K562 cells.

Lanes 5 and 6 were the supernatant and the eluate of K562 cells' total protein lysate after IP (Method 1)**.** The result showed that the majority of the proteins from the total protein lysate were retained in the supernatant. Nearly all the antibody *and* some other proteins were detected in the eluate. There were several bands close to the molecular weight of AGO proteins (~100 kDa). Therefore, the IP (Method 1) successfully removed the majority of the proteins from the crude lysate, but still co-purified some other proteins with the AGO proteins, and some of these contaminants had a similar molecular weight to AGO proteins. Lanes 7 and 8 were the supernatant and the eluate of K562 cells' total protein lysate after IP (Method 2). Likewise, the result showed that the majority of the proteins from the total protein lysate were retained in the supernatant. Nearly all the antibody and some other proteins were detected in the eluate. There was no big difference between the pattern of the results from Method 1 and Method 2.

As the result, IP Method 2 was selected in RIINGS with the consideration of the ability of harvesting the AGO proteins and the clearance of contaminants. Furthermore, an amount of 11.25 µg eIF2C2 (4F9) was chosen as the initial amount for the IP in order to ensure the yield of AGO proteins (see Section 5.3.1).

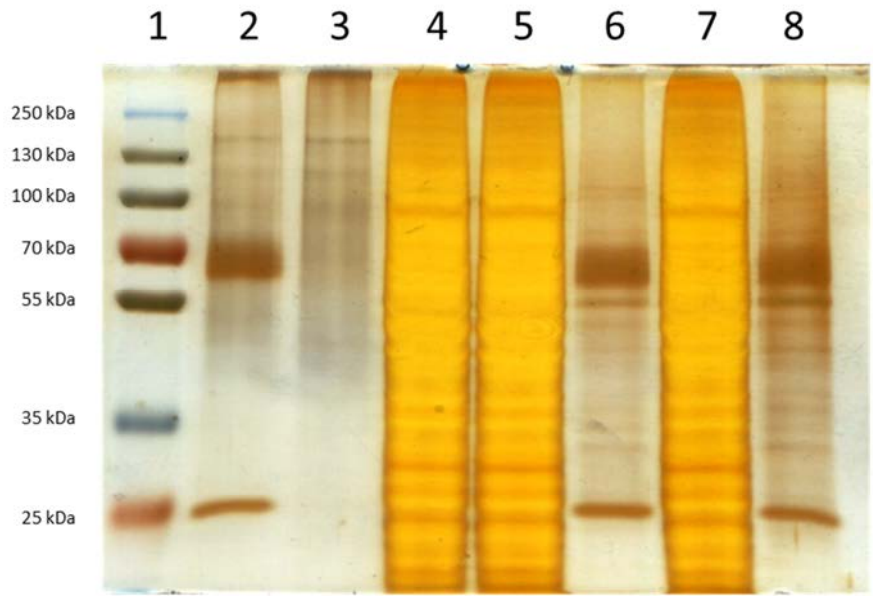


Figure 5.3.2 Silver staining of SDS-polyacrylamide gel for evaluating the efficiency of various steps in two immunoprecipitation methods. Samples loaded in the gel are as follows: lane 1, the protein ladder; lane 2, eIF2C2 (4F9); lane 3, the supernatant of the antibody solution after conjugation to the beads from the bead preparation for immunoprecipitation Method 1; lane 4, total protein extract of K562 cells; lanes 5 and 6, the supernatant and the eluate of K562 cells' total protein extract after immunoprecipitation (Method 1) respectively; and lanes 7 and 8, the supernatant and the eluate of K562 cells' total protein extract after immunoprecipitation (Method 2) respectively. Aliquots each containing 20 µg total protein extract (lane 4) and proteins with equivalent portion to the 20 µg total protein extract were loaded for the SDS-PAGE.

5.4 Conditions for the ligation of stem-loop adaptor

Three ligation parameters were optimized for the ligation of stem-loop adaptor, including ligation time, adaptor concentration and input amount of the ligase. All these three conditions were investigated by protein gel electrophoresis with autoradiography to visualize the results (see Figure 4.6.1). In brief, the stem-loop adaptor was first radiolabeled with ^{32}P-γ-ATP. After the adaptor had been ligated to the 3' end of the miRNA target, the sample was resolved by protein gel electrophoresis with NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well (Thermo Fisher). Electrophoresis was run under 175V at 4°C for 90 minutes in 1X NuPAGE® MOPS SDS Running Buffer (Thermo Fisher). The gel image was finally visualized by autoradiography on X-ray film and analyzed by GelAnalyzer 2010a.

5.4.1 Optimal ligation time for the ligation of stem-loop adaptor

A time-course study was conducted to identify the optimal ligation time for the ligation of the stem-loop adaptor to the 3' end of the miRNA target. Figure 5.4.1a shows the progress of the ligation reaction at different time points; 0, 12, 24 and 36 hours for lanes 1 to 4 respectively. A radioactive smear of about 220 kDa in molecular mass was detected with 12 to 36 hours of ligation. This smear was probably the AGO complexes that each consisted of an AGO protein, a miRNA and a target RNA. The expected molecular weight of all the isoforms of the AGO protein were about 100 kDa [323]. The shift in the molecular weight was due to the associated RNAs and the stem-loop adaptor. Figure 5.4.1b shows the plot of the corresponding radioactive signals of the AGO complex smears against time. Ligation progressed rapidly from 0 to 24 hours. This trend started to plateau after 24 hours. The increment between 24 hours and 36 hours was 9.60% only ($p = 0.0014$). Therefore, the optimal ligation time for the ligation of the stem-loop adaptor could be 24 to 36 hours.

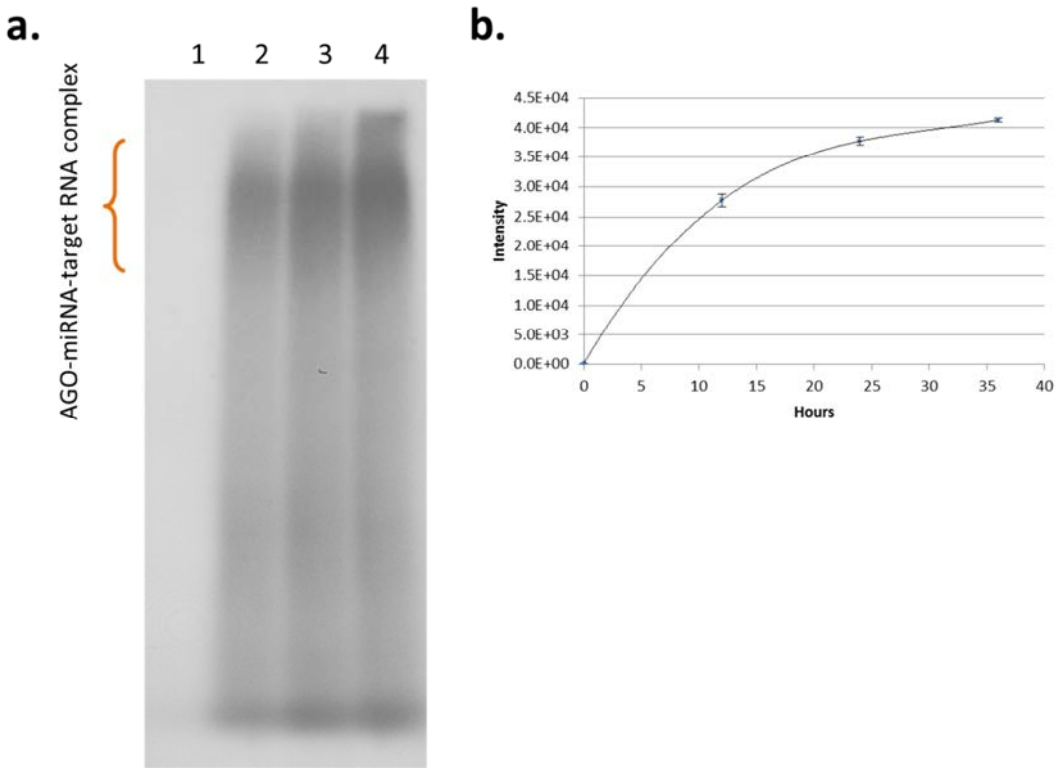


Figure 5.4.1 Time-course analysis for studying the ligation of a stem-loop adaptor to the 3' end of the miRNA target. (a) The ligation of the stem-loop adaptor was visualized by autoradiography. From lane 1 to 4, the ligation time was 0, 12, 24 and 36 hours respectively. b) The corresponding plot of intensity against time illustrated the completeness of the ligation during the time course studied. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.4.2 Optimal adaptor concentration for the stem-loop adaptor ligation

A titration study was conducted to optimize the optimal stem-loop adaptor concentration for its ligation to the 3' end of the miRNA target. The tested stem-loop adaptor concentrations were 240 nM, 480 nM, 720 nM and 960 nM. The actual amounts of the stem-loop adaptor in the reactions were 24 pmole, 48 pmole, 72 pmole and 96 pmole respectively. Figure 5.4.2a showed the ligation reactions with different concentrations of the stem-loop adaptor and with the same concentration of T4 RNA Ligase 1 at 1.5 U/µl. A radioactive smear of about 220 kDa in molecular mass was detected in all reactions. As has been explained above (Section 5.4.1) this smear was most likely the AGO complex composed of an AGO protein, a miRNA and a target RNA ligated with the radiolabeled stem-loop adaptor. The shift in the molecular weight of the AGO protein was due to the intercalated miRNA (~7 kDa), target RNA (depended on the length of the sequence) and the stem-loop adaptor (~17.5 kDa). From Figure 5.4.2b, it is obvious that the amount of AGO complex with ligated stem-loop adaptor increased with increasing concentrations of the input radiolabeled stem-loop adaptor, and that the increase steadily slowed down. To take into account such factors as the varying amounts of AGO complex in different cells or tissues, completeness of the ligation reaction and the cost involved, the optimal stem-loop adaptor concentration was chosen to be 2000 nM.

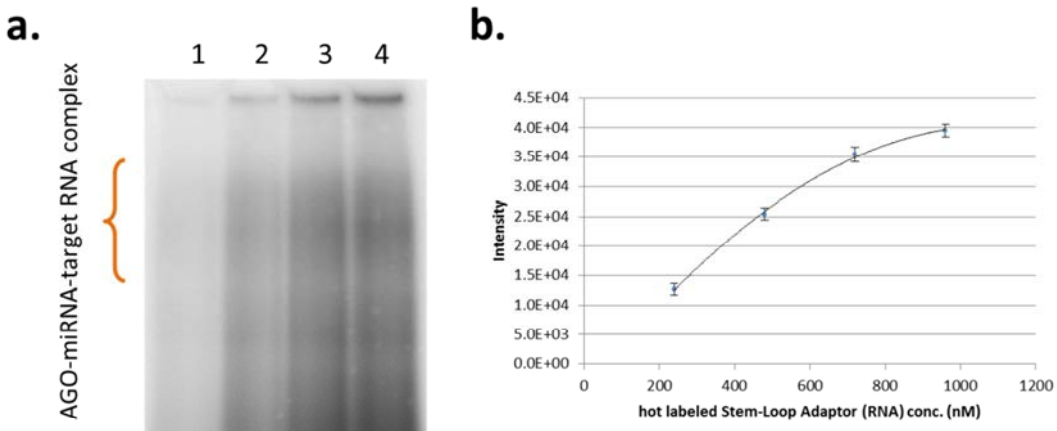


Figure 5.4.2 Titration study for determining the optimal concentration of the stem-loop adaptor in the ligation of the adaptor to the 3' end of the miRNA target. (a) The ligation of the stem-loop adaptor was visualized by autoradiography. From lane 1 to 4, the concentration of the stem-loop adaptor was 240, 480, 720 and 960 µM respectively. (b) The radioactive signals (intensities of the smears) from AGO-miRNA-target RNA complex were plotted against the adaptor concentrations. The higher the intensity, the more the adaptor ligated to the 3' end of the miRNA target. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.4.3 Optimal concentration of T4 RNA Ligase 1 for the ligation of the stem-loop adaptor

A titration study was conducted to optimize the concentration of the T4 RNA Ligase 1 (NEB) for the ligation of the stem-loop adaptor. The tested concentrations of the T4 RNA Ligase 1 in the ligation reaction were 1.5 U/µl (lanes 1-5) and 3 U/µl (lanes 6-19) as shown in Figure 5.4.3a. The concentration of the stem-loop adaptor in all the samples was fixed at 960 nM. A radioactive smear of about 220 kDa (the AGO complex; see Section 5.4.1 and Section 5.4.2) was detected in all reactions except those for 0-hour time points. As shown in Figure 5.4.3b, there was no big different between the reactions with 1.5 U/µl and 3 U/µl of T4 RNA Ligase 1 during the time course. To consider again such factors as the varying amounts of AGO complex in different cells or tissues, the completeness of the ligation reaction, and the cost involved, the optimal working concentration of the T4 RNA Ligase 1 was chosen to be 2.25 U/µl (the average of the two tested concentrations).

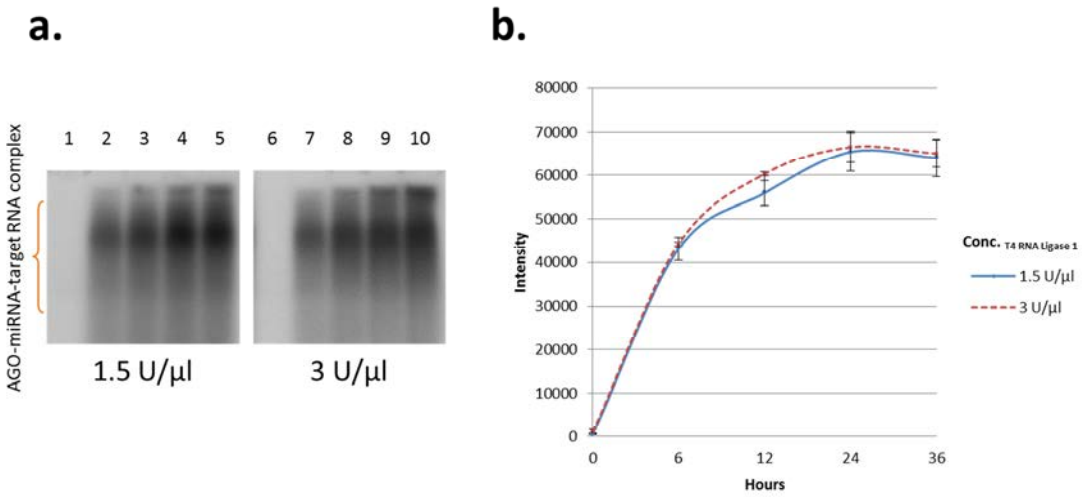


Figure 5.4.3 Titration study for determining the optimal concentration of T4 RNA Ligase 1 in the ligation of the adaptor to the 3' end of the miRNA target. (a) The ligation of the stem-loop adaptor was visualized by autoradiography. From lane 1 to 5 (ligation times of 0, 6, 12, 24 and 36 hours respectively), the concentration of the T4 RNA Ligase 1 in the reaction was 1.5 U/µl. From lane 6 to 10 (ligation times of 0, 6, 12, 24 and 36 hours respectively), the concentration of the T4 RNA Ligase 1 in the reaction was 3 U/µl. (b) The radioactive signals from AGO-miRNA-target RNA complex were plotted against time for the ligase concentrations tested. The higher the intensity, the more the adaptor ligated to the 3' end of the target RNA. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.5 Conditions for the ligation of re-attaching adaptor

With the ligation results of Section 5.4 taken in consideration, two intensive washing procedures for minimizing non-specific binding were examined in this section for the ligation of the re-attaching adaptor (see Figure 4.7.1). In addition, time-course analysis was also conducted to determine the optimal ligation time for the ligation of re-attaching adaptor (see Figure 4.7.1). Four time points were evaluated: 0, 12, 24 and 36 hours. All these investigations were assessed by protein gel electrophoresis with autoradiography to visualize the results. The re-attaching adaptor was first radiolabeled with [5'-^{32}P] pCp (PerkinElmer). After the adaptor had been ligated to the 5' end of the miRNA target RNA, the sample was resolved by protein gel electrophoresis with NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well (Thermo Fisher). The gel electrophoresis was run under 175V at 4℃ for 90 minutes in 1X NuPAGE® MOPS SDS Running Buffer (Thermo Fisher). The gel image was finally visualized by autoradiography on X-ray film and analyzed by GelAnalyzer 2010a.

5.5.1 Intensive washing procedures evaluation

In order to eliminate the non-specific binding occurring during IP, two intensive washing procedures were examined and compared with the original washing procedure. The washing procedure under evaluation was placed in the step immediately after RNase treatment (see Figure 4.7.1). Figure 5.5.1 shows the evaluation result of the three washing procedures. Washing procedure 2 did not show any significant difference when compared with the original washing procedure (Washing procedure 1). However, Washing procedure 3 removed a lot of non-specific binding from the immunoprecipitated fraction when compared with the original washing procedure (Washing procedure 1). This result indicated

that Washing procedure 3 actually could help minimize non-specific binding in the immunoprecipitated fraction.

Moreover, the pattern of the non-specific bands in Figure 5.5.1 was different from those for the ligation of stem-loop adaptor in Section 5.4, in which the ligation was performed on the 3' end of the miRNA target. Notably, the non-specific bands found for the ligation of the re-attaching adaptor to the 5' end of the miRNA target were more and clearer than those for the ligation of the stem-loop adaptor. This result might indicate that the ligation on 5' end was much robust than that on 3' end.

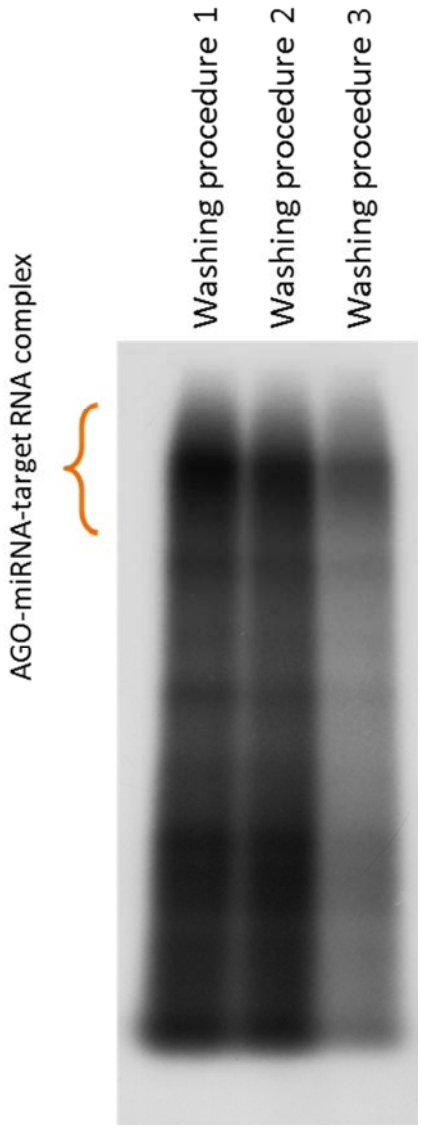


Figure 5.5.1 Evaluation of three washing procedures in eliminating non-specific binding occurring during immunoprecipitation. From the left to the right, the lanes showed the preparation of the samples washed with the original washing procedure (Washing procedure 1), the Washing procedure 2 and the Washing procedure 3, respectively, to eliminate the non-specific binding during immunoprecipitation.

5.5.2 Optimal ligation time for the ligation of re-attaching adaptor

A time-course study was conducted to determine the optimal ligation time for ligating the re-attaching adaptor to the 5' end of the miRNA target. The time points were 0, 12, 24 and 36 hours (Figure 5.5.2). A radioactive smear of about 220 kDa in molecular mass was detected from lane 2 to lane 4 (Figure 5.5.2a). In a way similar to that for the ligation of stem-loop adaptor to the 3’ end of the miRNA target (see Section 5.4), this smear was probably the AGO complex. As shown Figure 5.5.2b, the ligation reaction progressed rapidly and reached a plateau by 12 hours. One-way ANOVA indicated that there was no significant difference among the last 3 time points (12, 24 and 36 hours). Accordingly, the optimal ligation time for ligating the re-attaching adaptor to the 5’ end of the miRNA target was chosen to be 12 hours.

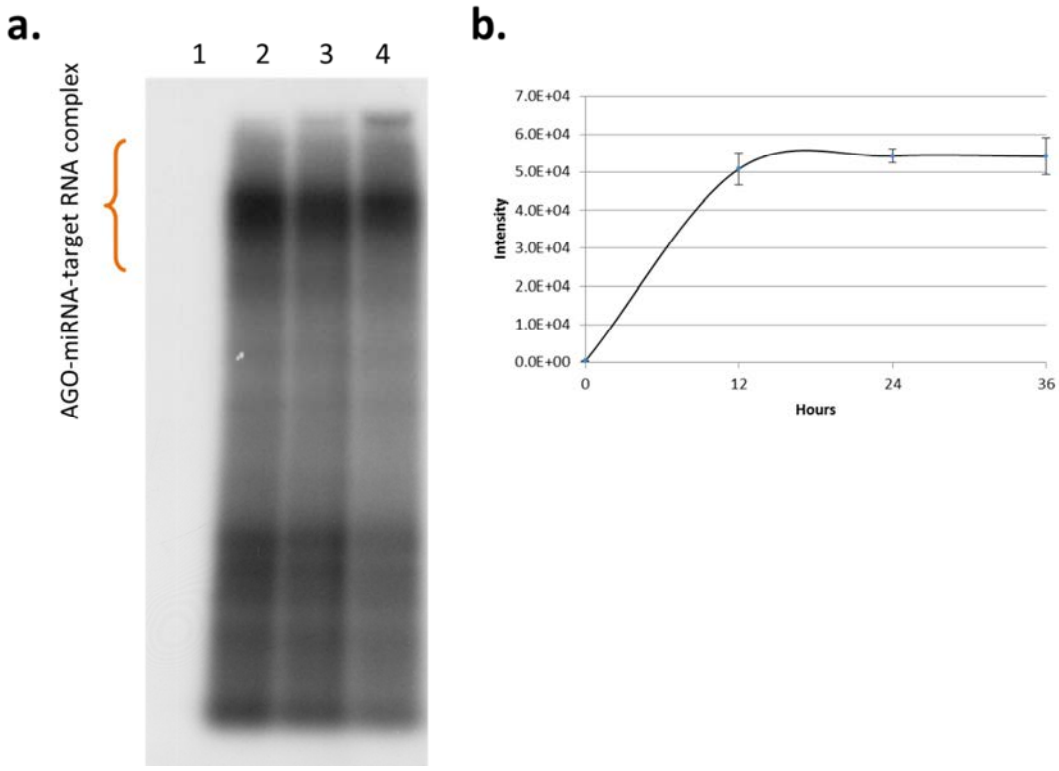


Figure 5.5.2 Time-course study for assessing the ligation of the re-attaching adaptor to the 3' end of the miRNA target. (a) The ligation of the re-attaching adaptor was visualized by autoradiography. From lane 1 to 4, the ligation time was 0, 12, 24 and 36 hours respectively. (b) The radioactive signals of the AGO-miRNA-target RNA complex were plotted against time. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.6 Determination of RNases working concentration

RNase A (Affymetrix) and RNase One™ Ribonuclease (Promega) were tested for trimming the flanking RNAs on the immunoprecipitated AGO complex, and their optimal concentrations were determined (see Figure 4.8.2). The original concentration of the RNase A was 38.46 U/µl (Lot. 73715) while the original concentration of the RNase One™ Ribonuclease was 10 U/µl. Of note was the different unit definition for these two RNases. Four different dilutions of each RNase were tested: undiluted, 1:100 diluted, 1:1000 diluted and 1:100000 diluted. An additional reaction without RNase treatment was used as the undigested control. After each of the RNA treatment, the effects were evaluated in terms of (a) the fragment size of AGO-associated RNAs and (b) the accessibility of the digested AGO complex for the ligation of a stem-loop adaptor. The fragment size was evaluated by Bioanalyzer 2100 (Agilent) while the accessibility for adaptor ligation was evaluated by protein gel electrophoresis with autoradiography to visualize the ligation of the radiolabeled adaptor.

5.6.1 Evaluation of the fragment size after RNase A treatment

A titration study was conducted to determine the optimal RNase A concentration for the procedure of RNA fragmentation in RIINGS. Without digestion (Figure 5.6.1a), the extracted RNAs (undigested control) varied greatly in size from ~40 nt to ~4200 nt with many fragments showing prominent peaks. With digestion by undiluted RNase A (Figure 5.6.1b), the majority of the extracted RNAs packed between ~40 nt and ~200 nt with 2 distinct peaks at ~100 nt. With digestion by 1:100-diluted RNase A (Figure 5.6.1c), the extracted RNAs distributed in size from ~40 nt to ~2000 nt with a skew toward smaller size and multiple distinct peaks at ~100 nt. In fact, the majority of the RNAs were less than 500 nt. With digestion by 1:1000-diluted RNase A (Figure 5.6.1d), the extracted RNAs varied in size

from ~40 nt to ~2600 nt with multiple distinctive peaks between ~100 nt and ~200 nt. The majority of the RNAs were less than 1000 nt. With digestion by 1:100000-diluted RNase A (Figure 5.6.1e and f), the distribution and the banding patterns of the extracted RNAs closely resembled those of the undigested control (Figure 5.6.1a). This indicated that the RNase activity was negligible at this dilution. Table 5.6.1 lists the details of the upper border, lower border and average length of the sequence length distribution of the extracted RNAs after digestion by RNase A.

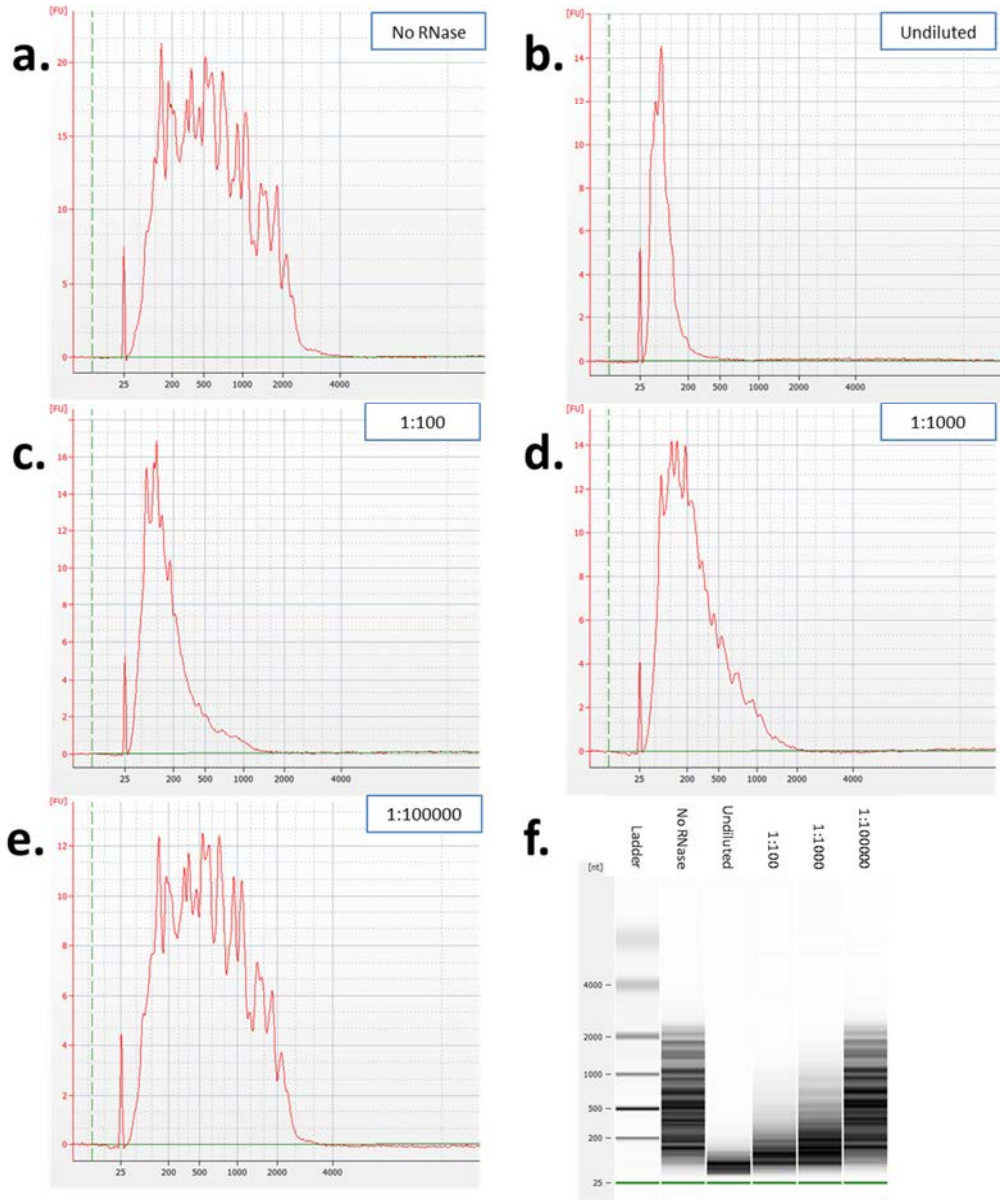


Figure 5.6.1 Analysis of the sequence length distribution of the miRNA target site after RNA fragmentation with different amounts of RNase A. The RNA samples (undigested or digested) were analyzed using Bioanalyzer 2011 to give (a)-(e) the electrophoretograms or (f) the corresponding gel picture. RNAs extracted from the AGO complex were mixed with 3 µl of (a) water (i.e. undigested control) or (b)-(e) different dilutions of RNase A (Affymetrix) in a 150-µl fragmentation reaction. The RNase A added were (b) undiluted, (c) 1:100 diluted, (d) 1:1000 diluted, and (e) 1:100000 diluted. For the electrophoretograms (a)-(e), the x-axis shows the size of the RNAs in nucleotides (nt) while y-axis indicates the amount of RNAs in arbitrary fluorescent unit (FU). (f) The gel picture converted from the above electrophoretograms to mimic a gel electrophoresis picture. The 25-nt peaks seen in (a)-(e) and the 25-nt green bands seen in (f) represent the marker fragment (RNA 6000 Pico Marker, Agilent) that migrates to a position equivalent to a 25-nt RNA transcript.

Table 5.6.1 The fragment size distribution and the average length of the RNA fragments extracted from AGO complex after RNase A treatment.

	Fragment distribution (nt)				Average length (nt)	
	Lower	CV	Upper	CV	Length	CV
no RNase	41.67	5.00%	4170.00	11.00%	637.33	21.52%
undiluted	37.33	5.58%	499.33	9.15%	103.33	1.48%
1:100	37.33	8.18%	1994.33	35.31%	164.33	6.68%
1:1000	37.00	7.15%	2633.67	17.96%	236.00	16.84%
1:100000	40.00	4.33%	4198.33	14.27%	648.00	21.30%

5.6.2 Evaluation of the fragment size after RNase One™ Ribonuclease treatment

A titration study was conducted to determine the optimal RNase One™ Ribonuclease concentration for the procedure of RNA fragmentation in RIINGS. Without digesting (Figure 5.6.2a) , the extracted RNAs distributed in size from ~50 nt to ~4000 nt with many fragments showing distinct peaks. With digestion by undiluted RNase One™ Ribonuclease (Figure 5.6.2b), the extracted RNAs packed between ~40 nt and ~200 nt with 3 distinct peaks centering around ~100 nt. With digestion by 1:100-diluted RNase One™ Ribonuclease (Figure 5.6.2c), the extracted RNAs varied in size from ~40 nt to ~1900 nt with major peaks close to ~200 nt. In effect, the majority of the RNAs were less than 500 nt. With digestion by 1:1000-diluted RNase One™ Ribonuclease (Figure 5.6.2d), the extracted RNAs showed a much broader distribution from ~40 nt to ~2900 nt with many prominent peaks in between. With digestion by 1:100000-diluted RNase One™ Ribonuclease (Figure 5.6.2e and f) , the distribution and the banding patterns of the extracted RNAs were very similar to those of the undigested control (Figure 5.6.2a). This indicated that the RNase activity was insignificant at this dilution. Table 5.6.2 lists the details of the upper border, lower border and average length of the sequence length distribution of the extracted RNAs after digestion by RNase One™ Ribonuclease.

All in all, digestion by either RNase A or RNase One™ Ribonuclease gave very similar size distributions of the RNAs extracted from the immunoprecipitated AGO complex at the corresponding dilutions. However, a slight difference existed between RNase A or RNase One™ Ribonuclease. The ribonuclease activity of RNase One™ Ribonuclease was slightly less than that of RNase A. This could be observed from 1:100 dilution and 1:1000 dilution (Figure 5.6.1c & d vs Figure 5.6.2c & d, and Table 5.6.1 vs Table 5.6.2). At 1:100 dilution, the distribution of RNA fragments was shifted to the right slightly when the RNAs were treated

with RNase One™ Ribonuclease. This shift was more extensive when the dilution was 1:1000. These right shifts indicated the comparatively longer fragments, and also implied the weaker ribonuclease activity of RNase One™ Ribonuclease.

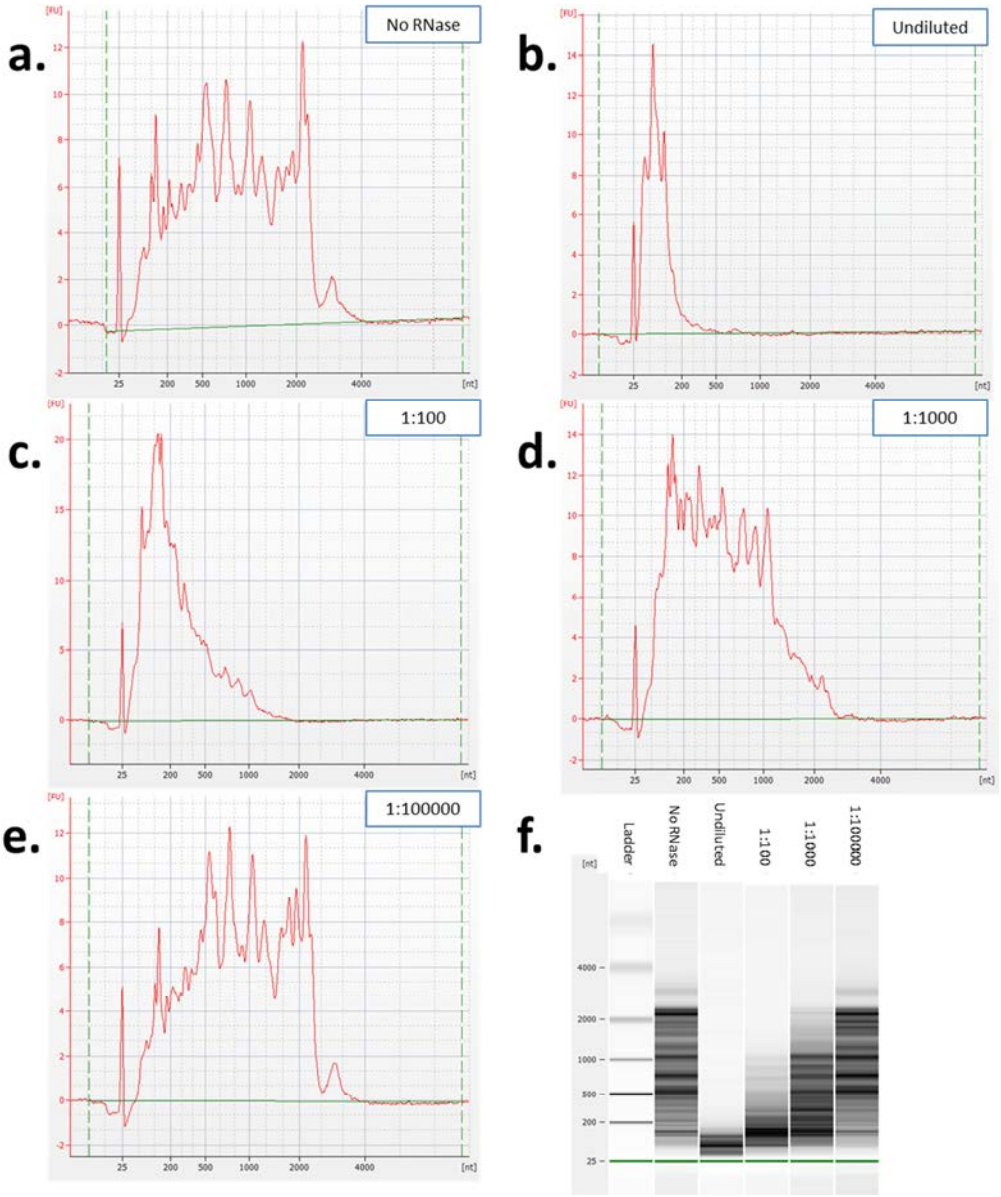


Figure 5.6.2 Analysis of the sequence length distribution of the miRNA target site after RNA fragmentation with different amount of RNase One™ Ribonuclease. The RNA samples (undigested or digested) were analyzed using Bioanalyzer 2011 to give (a)-(e) the electrophoretograms or (f) the corresponding gel picture. RNAs extracted from the AGO complex were mixed with 3 µl of (a) water (i.e. undigested control) or (b)-(e) different dilutions of RNase One™ Ribonuclease (Promega) in a 150-µl fragmentation reaction. The RNase One™ Ribonuclease added were (b) undiluted, (c) 1:100 diluted, (d) 1:1000 diluted, and (e) 1:100000 diluted. For the electrophoretograms (a)-(e), the x-axis shows the size of the RNAs in nucleotides (nt) while y-axis indicates the amount of RNAs in arbitrary fluorescent unit (FU). (f) The gel picture converted from the above electrophoretograms to mimic a gel electrophoresis. The 25-nt peaks seen in (a)-(e) and the 25-nt green bands seen in (f) represent the marker fragment (RNA 6000 Pico Marker, Agilent) that migrates to a position equivalent to a 25-nt RNA transcript.

Table 5.6.2 The fragment size distribution and the average length of the RNA fragments extracted from AGO complex after RNase One™ Ribonuclease treatment.

	Fragment distribution (nt)				Average length (nt)	
	Lower	CV	Upper	CV	Length	CV
no RNase	48.67	18.98%	4000.33	0.58%	892.00	16.99%
undiluted	35.00	2.86%	447.00	26.46%	105.33	3.33%
1:100	37.00	4.68%	1876.00	26.96%	195.33	12.27%
1:1000	38.67	3.95%	2869.00	17.12%	407.33	12.13%
1:100000	48.00	4.17%	3801.67	8.75%	894.33	11.49%

5.6.3 *Evaluation of the accessibility of the immunoprecipitated AGO complex for ligation after RNase A treatment*

In order to check whether the RNA ends on the immunoprecipitated AGO complex after the trimming by RNase A were long enough for adaptor ligation, an experiment was used to evaluate this accessibility. A radiolabeled adaptor was first ligated to the trimmed RNA 3' end on the immunoprecipitated AGO complex. Then, protein gel electrophoresis was used to resolve the complex. Finally, the smear representing the immunoprecipitated AGO complex, which was successfully ligated with the radiolabeled adaptor, was visualized by autoradiography. Figure 5.6.3 summarizes the results of this experiment based on RNase A.

The immunoprecipitated AGO complex consisted of an AGO protein (~100 kDa), a miRNA (~7 kDa) and a target RNA (depended on the length of the sequence). Therefore, the location of AGO complex could be determined by the band shift and appeared as a smear that was indicated by a vertical blue line on the left hand side of each lane in Figure 5.6.3a. These smears gradually increased in molecular mass and spread wider with *decreasing* concentration of RNase A used. This pattern agreed with the finding in Section 5.6.1, in which the sequence length (or the molecular mass) and the length distribution of the RNAs from the immunoprecipitated AGO complex gradually increased with *decreasing* concentration of RNase A. Therefore, the increasing molecular mass of the smears was most probably due to the increasing sequence length of the RNAs from the immunoprecipitated AGO complex. This result also implied that this smear most likely carried an RNA binding protein because the molecular mass of this smear was sensitive to RNase treatment.

From Figure 5.6.3b, it is obvious that the ligation of radiolabeled stem-loop adaptor increased with decreasing concentration of RNase used. The increment in ligation between

the sample treated with undiluted RNase A and the sample treated with 1:100 diluted RNase A was not significant. However, the increment in ligation was 2.35-fold (p = 0.0026) between the sample treated with 1:100 diluted RNase A and the sample treated with 1:1000 diluted RNase A, and 1.86-fold (p = 0.0038) between the sample treated with 1:1000 diluted RNase A and the sample treated with 1:100000 diluted RNase A. The spatial hindrance was expected to be large enough to block the ligation reaction dramatically when the protruding RNAs were too short. Indeed, the result matched the expectation.

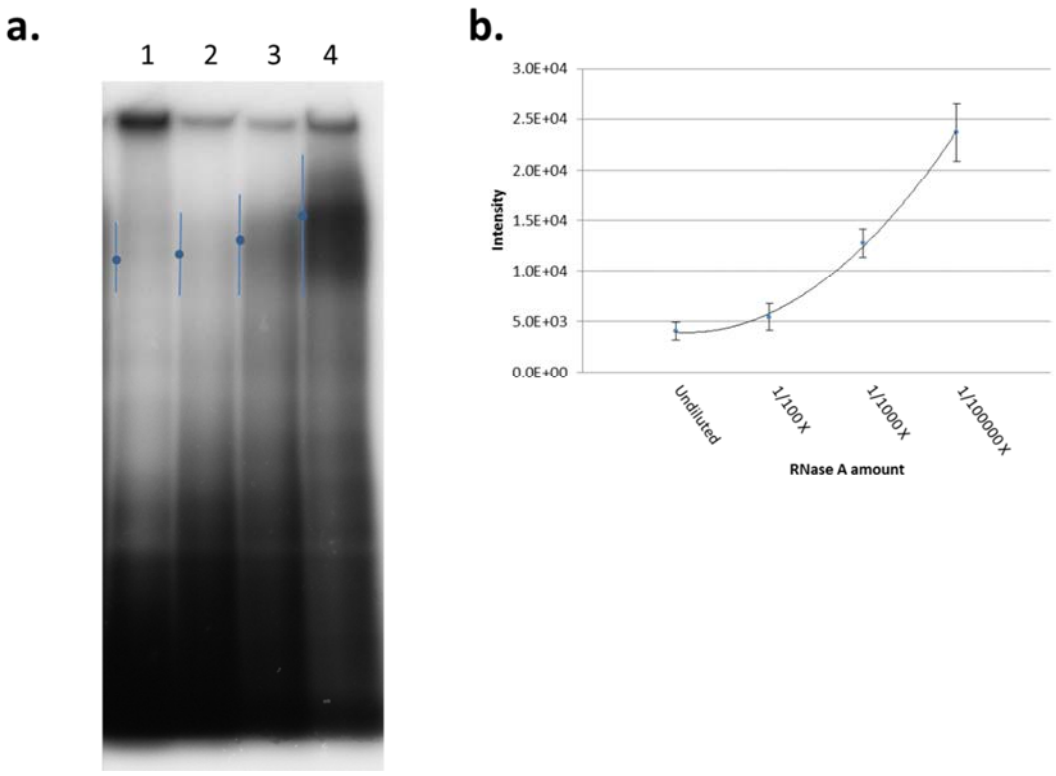


Figure 5.6.3 Evaluation of the accessibility of the immunoprecipitated AGO complex for ligation of adaptor to the RNA 3' end on the AGO complex after RNase A treatment. (a) The ligation of the radiolabeled stem-loop adaptor was visualized by autoradiography. The vertical blue line on the left hand side of each lane indicated the location of the smear representing the immunoprecipitated AGO complex ligated with the radiolabeled stem-loop adaptor while the blue dot showed the peak of the smear. From lanes 1 to 4, the marked smears were the immunoprecipitated AGO complex after digestion of the protruding RNAs with (1) undiluted, (2) 1:100-diluted, (3) 1:1000-diluted and (4) 1:100000-diluted RNase A in a 150-µl fragmentation reaction. (b) The radioactive signals (intensities of the bands) from the AGO complex with ligated radiolabeled stem-loop adaptor were plotted against the concentration of RNase A used. The higher the intensity, the more the adaptor ligated to the 3' end of the miRNA target. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.6.4 Evaluation of the accessibility of the immunoprecipitated AGO complex for ligation after RNase One™ Ribonuclease treatment

The whole experiment assessing the accessibility of the immunoprecipitated AGO complex for ligation was repeated with RNase One™ Ribonuclease in place of RNase A. Figure 5.6.4 summarizes the results of this experiment based on RNase One™ Ribonuclease.

The smears representing the AGO complex with ligated radiolabeled stem-loop adaptor gradually increased in molecular mass and spread wider with *decreasing* concentration of RNase One™ Ribonuclease used (Figure 5.6.4a). This pattern agreed with the finding in Section 5.6.2, in which the sequence length (or the molecular mass) and the length distribution of the RNAs from the immunoprecipitated AGO complex gradually increased *decreasing* concentration of RNase One™ Ribonuclease. Therefore, the increasing molecular mass of the smears was most probably due to the increasing sequence length of the RNAs from the immunoprecipitated AGO complex. This result also implied that this smear most likely contained an RNA binding protein because the molecular mass of this smear was sensitive to RNase treatment.

When the concentration of RNase One™ Ribonuclease was *reduced* from undiluted to 1:100-diluted, the extent of ligation increased 2.63-fold (p = 0.0032) (Figure 5.6.4b). However, the extent of ligation did not increase further with decreasing concentration of RNase One™ Ribonuclease lower than a dilution of 1:100. This result also matched the expectation based on spatial hindrance.

On the other hand, the increasing trend in the extent of ligation was obviously very different between these two RNases under evaluation (Figure 5.6.3b vs Figure 5.6.4b). For RNase A, it increased gradually with decreasing RNase concentration (Figure 5.6.3b).

However, it reached a plateau with 1:100-diluted RNase One™ Ribonuclease (Figure 5.6.4b). This difference might be due to the fragment length differences between the two RNase treatments, which in turn might be due to the repair of 2'3' cyclic monophosphate end (see a detailed discussion in Section 6.5).

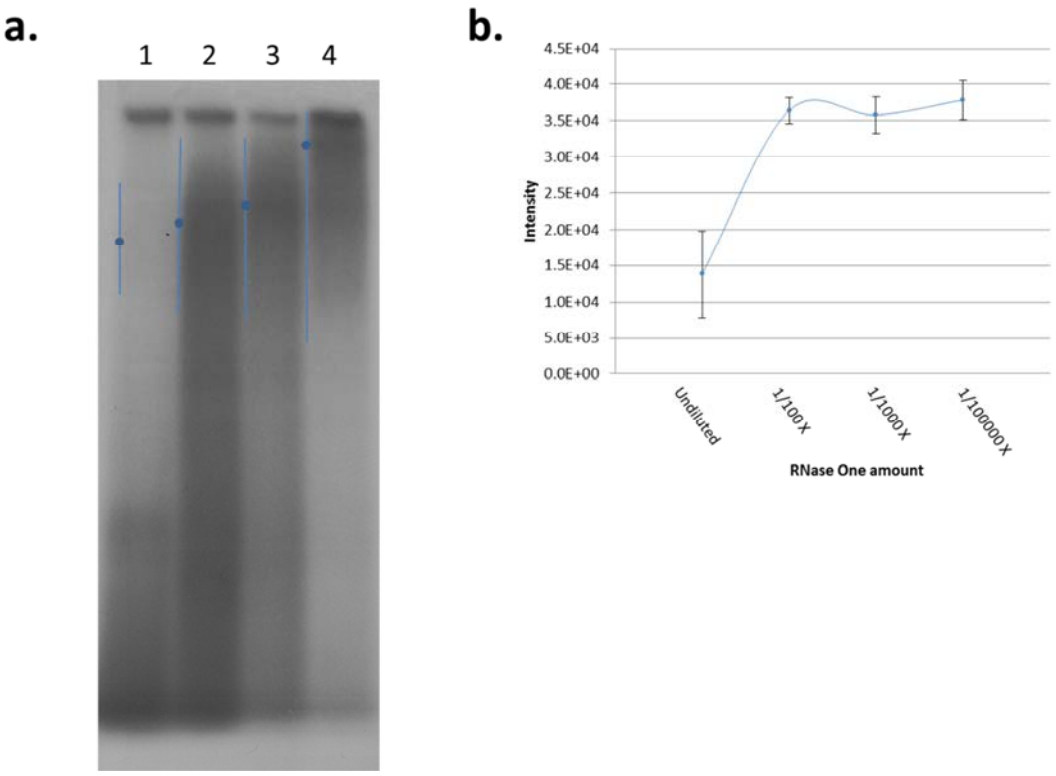


Figure 5.6.4 Evaluation of the accessibility of the immunoprecipitated AGO complex for ligation of adaptor to the RNA 3' end on the AGO complex after RNase One™ Ribonuclease treatment. (a) The ligation of the radiolabeled stem-loop adaptor was visualized by autoradiography. The vertical blue line on the left hand side of each lane indicated the location of the smear representing the immunoprecipitated AGO complex ligated with the radiolabeled stem-loop adaptor while the blue dots located the peaks of the smears. From lanes 1 to 4, the marked smears were the immunoprecipitated AGO complex after digestion of the protruding RNAs with (1) undiluted, (2) 1:100-diluted, (3) 1:1000-diluted and (4) 1:100000-diluted RNase One™ Ribonuclease in a 150-µl fragmentation reaction. (b) The radioactive signals (intensities of the bands) from the AGO complex with ligated radiolabeled stem-loop adaptor were plotted against the concentration of RNase One™ Ribonuclease used. The higher the intensity, the more the adaptor ligated to the 3' end of the miRNA target. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.7 Optimization of crosslinking condition

The two crosslinking methods used in HITS-CLIP and PAR-CLIP were evaluated in the current study to determine an optimal crosslinking condition for K562 cells. The comparison of the crosslinking methods was conducted by irradiating the K562 cells either with or without 4-thiouridine treatment and with either 254 nm UV light or 365 nm UV light. Total protein of the cells was then extracted and the target Argonaute complex purified by IP. After the ligation of a radiolabeled adaptor to the purified Argonaute complex, the result was visualized by protein gel electrophoresis with autoradiography on X-ray film and was analyzed by GelAnalyzer 2010a (see Figure 4.9.1).

Figure 5.7.1 summarizes the results of comparing the two crosslinking methods for K562 cells. Levene's test was used to test the homogeneity of variances across different groups of samples prepared by different crosslinking conditions, and did not show significant difference (p = 0.451). This indicated that the variances were equal across groups. One-way ANOVA comparing these eight groups of samples showed significant differences among these groups (p = 0.006). Post-hoc test (Tukey HSD) indicated that significant results found by ANOVA could be ascribed to significantly higher yield by the group treated with 4-thiouridine and crosslinked by 600 mJ, 254 nm UV light when compared to any one of the following four groups (pairwise comparisons): (i) the group treated with 4-thiouridine and crosslinked by 150 mJ, 365 nm UV light (1.37-fold; p = 0.024); (ii) the group without 4-thiouridine treatment and crosslinked by 150 mJ, 254 nm UV light (1.41-fold; p = 0.012); (iii) the group without 4-thiouridine treatment and crosslinked by 150 mJ, 365 nm UV light (1.42-fold; p = 0.011); and (iv) the group without 4-thiouridine treatment and crosslinked by 600 mJ, 365 nm UV light (1.36-fold; p = 0.027). However, these differences were not big differences (only up to 36% – 42% higher in the group the group treated with 4-thiouridine

and crosslinked by 600 mJ, 254 nm UV light). Therefore, all the eight crosslinking conditions could be used for the crosslinking in RIINGS according to the above findings. However, crosslinking would be much better with lower-energy UV light (i.e. 356 nm) than with higher-energy UV light (i.e. 254 nm) because the higher-energy of UV light may damage the target AGO complex and even to the RNA duplex structure to a greater extent.

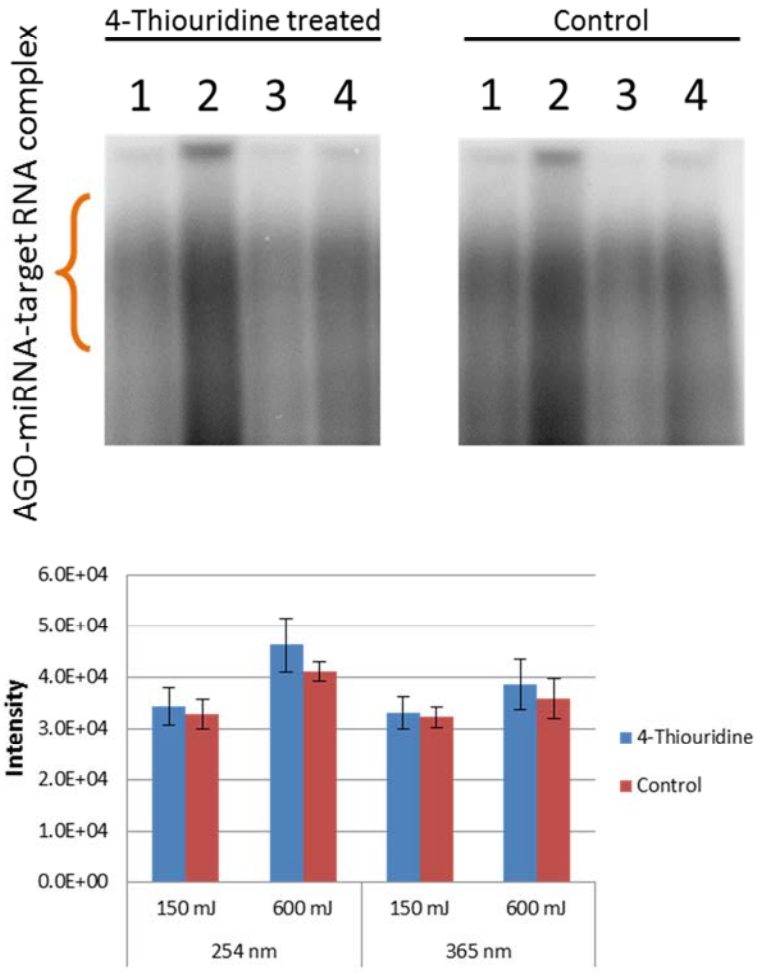


Figure 5.7.1 Comparison of the efficiency of crosslinking by different strategies. (a) The target AGO complex prepared by different crosslinking methods and labeled via 3' adaptor ligation to the 3' end of the target RNA were separated by protein gel electrophoresis and visualized by autoradiography. The gel photo on the left-hand side shows 4-thiouridine-treated samples whereas the photo on the right-hand side shows the corresponding control samples (without 4-thiouridind treatment). Lane 1 is the sample crosslinked by 150 mJ, 254 nm UV light. Lane 2 is the sample crosslinked by 600 mJ, 254 nm UV light. Lane 3 is the sample crosslinked by 150 mJ, 365 nm UV light. Lane 4 is the sample crosslinked by 600 mJ, 365 nm UV light. (b) The corresponding bar chart shows the signal intensities of the smears representing AGO complex prepared by different crosslinking methods. The stronger the signal intensity, the more the AGO complex yielded in the sample. Signal intensities represent mean values ± SD (n = 3).

5.8 Conversion of 2´, 3´-cyclic monophosphate end

T4 polynucleotide kinase (T4 PNK) was used to convert the 2´,3´-cyclic monophosphate end of the fragmented RNA, which was generated by either RNase A or RNase I during the RNA fragmentation process, to the corresponding 3’ hydroxyl end to facilitate the ligation of the stem-loop adaptor. A side-by-side comparison was conducted to determine whether additional T4 PNK treatment could improve the efficiency of the ligation of the stem-loop adaptor to the 3’ end of miRNA target via the conversion of 3’ monophosphate end and 2',3'-cyclic monophosphate end to 3’ hydroxyl end. The additional T4 PNK treatment was placed in the step immediately after the RNase treatment (see Figure 4.10.3). This additional treatment was carried out on Eppendorf ThermoMixer® C at 16℃ for 30 minutes with mixing at 1000 rpm for 15 seconds at each 2-minute intervals. The efficiency of adaptor ligation was evaluated by protein gel electrophoresis followed by autoradiography to visualize the ligation of the radiolabeled adaptor to the 3’ end of the RNA that was bound to the Argonaute complex.

5.8.1 Evaluation of the effect of T4 PNK treatment to the stem-loop adaptor ligation (RNA was fragmented by RNase A)

A comparison between with and without additional T4 PNK treatment in the ligation of the stem-loop adaptor was conducted to evaluate the effect of additional T4 PNK treatment on the adaptor ligation. Here, RNA fragmentation was achieved by the use of RNase A (Affymetrix). Figure 5.8.1 summarizes the results of this comparison with the RNAs being digested by RNase A. The signal intensity of AGO complex with ligated radiolabeled adaptor was 2.31 folds higher (p = 0.002) in the reaction with additional T4 PNK treatment than in the reaction without additional T4 PNK treatment. Obviously, T4 PNK treatment was beneficial to the ligation of stem-loop adaptor in RIINGS when RNA fragmentation was

conducted by RNase A. This was most likely because 2´,3´-cyclic monophosphate end on the RNA generated by RNase A digestion was successfully repaired and converted to 3’ hydroxyl end by the T4 PNK.

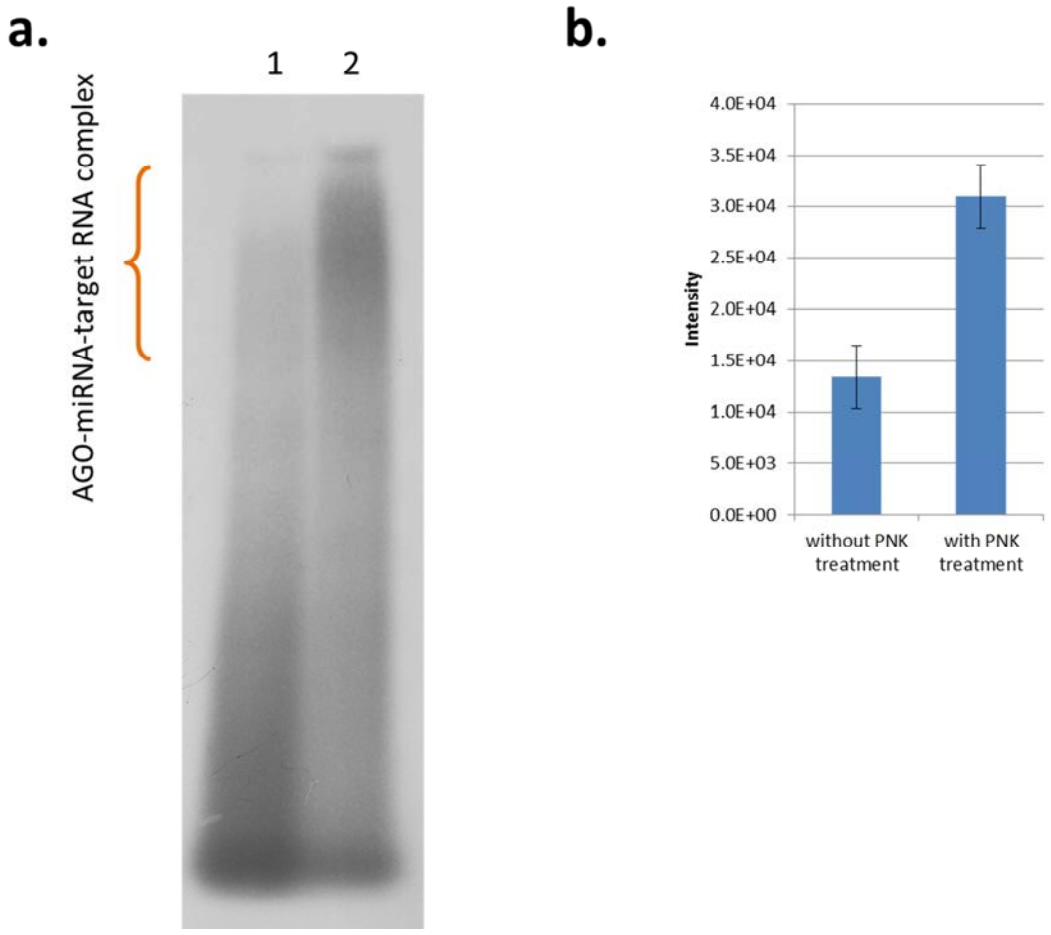


Figure 5.8.1 Comparison of the efficiency of ligating a stem-loop adaptor to the 3' end of RNase A-digested RNAs with and without additional T4 PNK treatment. (a) The ligation of the stem-loop adaptor was visualized by autoradiography. Lane 1 was the control (the reaction without additional T4 PNK treatment). Lane 2 was the reaction with additional T4 PNK treatment. (b) The corresponding bar chart showed the signal intensities of the smears representing AGO complex with the ligated radiolabeled adaptor between the two studied conditions. The higher the intensity, the more the adaptor ligated to the 3' end of the miRNA target. Signal intensities were based on the mean value ± SD (n = 3).

5.8.2 Evaluation of the effect of T4 PNK treatment to the stem-loop adaptor ligation (RNA was fragmented by RNase I)

A comparison between with and without additional T4 PNK treatment in the ligation of the stem-loop adaptor was carried out to assess the effect of the additional T4 PNK treatment on the adaptor ligation. Here, the RNA fragmentation was achieved by the use of RNase I (Promega). Figure 5.8.2 summarizes the results of this comparison with the RNAs being digested with RNase I. The signal intensity of AGO complex with ligated radiolabeled adaptor was 2.44 folds higher (p = 0.00009) in the reaction with additional T4 PNK treatment than in the reaction without additional T4 PNK treatment. Clearly, the T4 PNK treatment was beneficial to the ligation of stem-loop adaptor in RIINGS when RNA was digested by RNase I. This was most likely because 2´,3´-cyclic monophosphate end on the RNA produced by RNase I digestion was successfully converted to 3’ hydroxyl end by the T4 PNK.

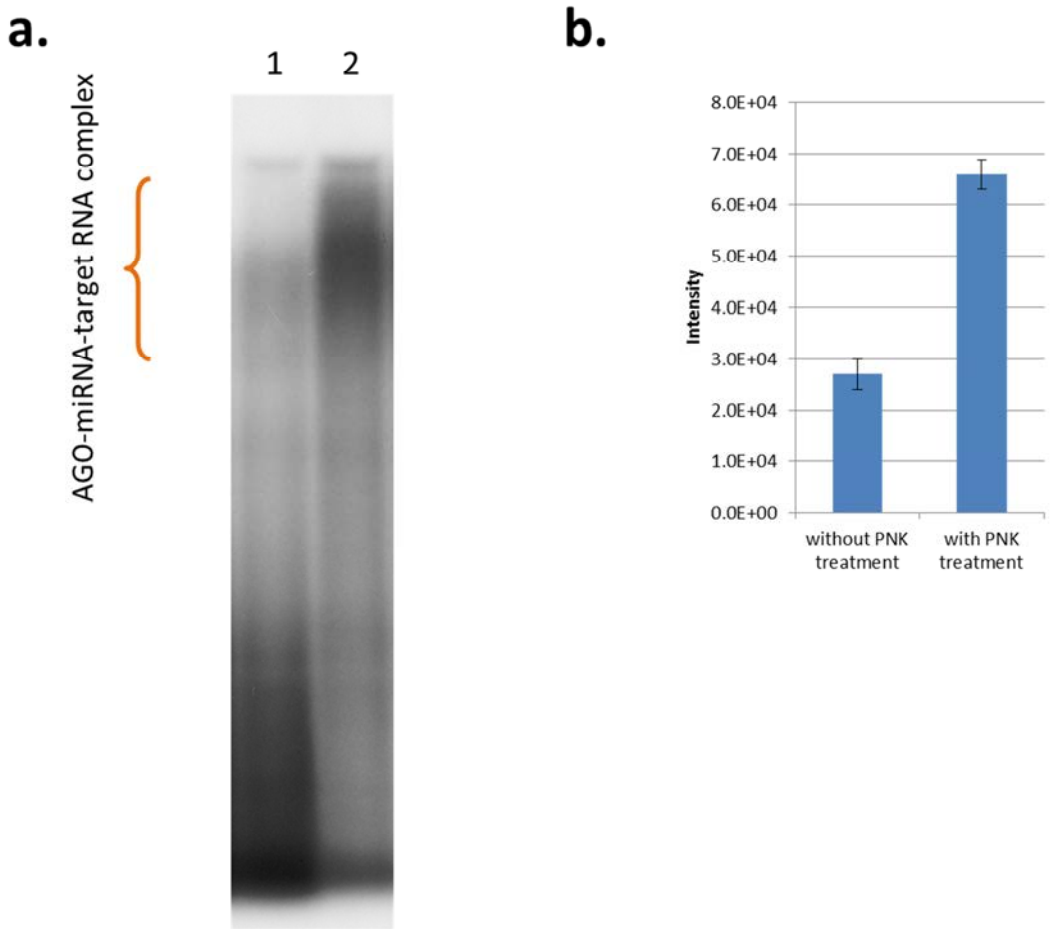


Figure 5.8.2 Comparison of the efficiency of ligating a stem-loop adaptor to the 3' end of RNase I-digested RNAs with and without additional T4 PNK treatment. (a) The ligation of the stem-loop adaptor was visualized by autoradiography. Lane 1 was the control (the reaction without additional T4 PNK treatment). Lane 2 was the reaction with additional T4 PNK treatment. (b) The corresponding bar chart showed the signal intensities of the smears representing AGO complex with the ligated radiolabeled adaptor between the two studied conditions. The higher the intensity, the more the adaptor ligated to the 3' end of the miRNA target. The signal intensities were based on the mean value ± SD (n = 3).

5.9 The properties of the isolated AGO complex

Three important properties of AGO complex for RIINGS experiment were assessed: the purity of the IP-purified AGO complex, the stability of the miRNA-target RNA duplex, and the quantity of the miRNA-target RNA duplex. The results of these three investigated properties are reported in the following three sections respectively.

5.9.1 Evaluation of the purity of the immunoprecipitation purified AGO complex by mass spectrometry

The purity of the IP-purified AGO complex was first evaluated by SDS-PAGE with silver staining. The result can be found in Section 5.3.2. In addition, the purity was also assessed by TripleTOF™ 6600 (AB SCIEX) (see Figure 4.11.1) The MS data were analyzed by ProteinPilot™ Software 5.0. Table 5.9.1 lists the proteins that were identified in the IP-purified protein eluate. The proteins in the list were filtered by a false discovery rate of less than 1% at both the peptide level and the protein level, and listed in a descending order by the confidence score of protein. The first column from the left hand side (Order) was the order number of the specified protein relative to the proteins in the list. It was ranked by the confidence score of the protein. The second column (Unused ProtScore (Conf)) was a transformation of the percent confidence of a protein. The transformation actually made the percent confidence more easily to compare and read. The calculation was followed Equation 4.

$$ProtScore = -log\left(1 - \frac{Precent\ confidence}{100}\right) \qquad \text{(Equation 4)}$$

For example, if the percent confidence of a protein was 99%, the ProtScore would be 2. Therefore, the higher the score in the second column, the higher the confidence of

identifying the protein concerned. The third column (%Cov(95)) was the percentage of amino acids matched to the full length of the protein sequence. The identified peptide sequences used in this alignment should have confidence of not less than 95%. The forth column was the full names of the proteins. The fifth column was the species to which the specified protein belonged. The sixth column was the number of peptides that covered the specified protein. The seventh column was the gene symbols of the proteins.

In Table 5.9.1, there were 148 proteins that were identified with confidence greater than 99% by TripleTOF™ 6600. AGO1 and AGO2 were included in these 148 identities. AGO1 was covered by 16 peptides with confidence of not less than 95% and with coverage of 20.65% of the full length of the protein. AGO2 was covered by 27 peptides with confidence of not less than 95% and with coverage of 30.38% of the full length of the protein. On the other hand, AGO3 and AGO4 were identified with confidence in 0% and 61.98% respectively. They were covered by 11 and 10 peptides with confidence of not less than 95% respectively. The coverage of the full length of AGO3 and AGO4 were 11.51% and 11.38% respectively. The differences of the percent confidence in calling the protein may partially be due to the amount of the corresponding AGO protein in the eluate. It implied that AGO2 and AGO1 may be the dominant AGO proteins among the four AGO proteins in K562 cells.

Besides the targeted AGO1-4, a large number of other proteins were co-purified by the current IP process. About 40% of these co-purified proteins were nucleic acid binding proteins and were mainly RNA binding protein among the list in Table 5.9.1. Moreover, 40% of these nucleic acid binding proteins were ribosomal proteins and were related to translation process in living cells. According to their functions, the other nucleic acid binding proteins were related to mRNA transcription, maturation, transportation, storage and translation. Only two proteins, TNRC6A (order 103) and TNRC6B (order 6), were related

to miRNA-mediated cellular process. These two proteins actually interacted with the RISC and contributed to AGO-mediated cellular process. Thus, this is not surprising that they were co-purified by the current IP process.

All in all, although there were a large number of proteins co-purified by the current IP process, it was not necessary to further purify the purified AGO1-4 to exclude other co-purified proteins. It is because these co-purified proteins did not introduce any miRNA-target RNA duplex other than AGO-introduced duplex to the current investigation. Furthermore, the co-purified proteins would be digested by proteinase K in the following step. On the other hand, co-purified RNAs would not establish reads in the next generation sequencing easily.

Table 5.9.1 The list of the current immunoprecipitation process purified proteins. Analysis was conducted by TripleTOF™ 6600 (AB SCIEX). The AGO1-4 proteins are shown in **bold** letters. Because the confidence for AGO3 and AGO4 was than 99%, they are highlighted in yellow.

Order	Unused ProtScore (Conf)	%Cov(95)	Name	Species	Peptides(95%)	Gene Name
1	54.68	70.72	Tubulin beta chain	HUMAN	50	TUBB
2	**49.09**	**30.38**	**Protein argonaute-2**	**HUMAN**	**27**	**AGO2**
3	38.44	26.80	Heat shock protein HSP 90-beta	HUMAN	19	HSP90AB1
4	38.21	48.56	Tubulin alpha-1B chain	HUMAN	30	TUBA1B
5	27.82	55.80	Pyrroline-5-carboxylate reductase 1, mitochondrial	HUMAN	18	PYCR1
6	26.68	10.15	Trinucleotide repeat-containing gene 6B protein	HUMAN	16	TNRC6B
7	26.02	53.09	40S ribosomal protein S3	HUMAN	13	RPS3
8	25.45	20.42	Collagen alpha-1(I) chain	HUMAN	32	COL1A1
9	22.69	32.53	Actin, cytoplasmic 2	HUMAN	13	ACTG1
10	21.9	18.89	Heat shock cognate 71 kDa protein	HUMAN	11	HSPA8
11	21.71	14.38	Heterogeneous nuclear ribonucleoprotein M	HUMAN	8	HNRNPM
12	21.62	47.81	Pyrroline-5-carboxylate reductase 2	HUMAN	16	PYCR2
13	19.66	40.06	60S acidic ribosomal protein P0	HUMAN	9	RPLP0
14	19.05	57.93	40S ribosomal protein S19	HUMAN	10	RPS19
15	18.31	28.67	40S ribosomal protein S2	HUMAN	9	RPS2
16	17.48	26.84	Putative elongation factor 1-alpha-like 3	HUMAN	13	EEF1A1P5
17	16.56	25.24	Receptor of activated protein C kinase 1	HUMAN	8	RACK1
18	16.04	17.28	Alpha-enolase	HUMAN	8	ENO1
19	15.45	31.29	Nucleophosmin	HUMAN	9	NPM1
20	13.87	19.86	Serpin H1	HUMAN	7	SERPINH1
21	13.84	25.00	Poly(rC)-binding protein 1	HUMAN	7	PCBP1
22	13.07	18.72	Heat shock 70 kDa protein 1B	HUMAN	9	HSPA1B
23	12.82	11.62	Transcription intermediary factor 1-beta	HUMAN	6	TRIM28
24	**12.79**	**20.65**	**Protein argonaute-1**	**HUMAN**	**16**	**AGO1**
25	12.3	29.27	Heat shock protein beta-1	HUMAN	7	HSPB1
26	12.22	20.19	Testin	HUMAN	8	TES
27	12	6.81	Collagen alpha-2(I) chain	HUMAN	10	COL1A2
28	11.25	3.40	Golgin subfamily A member 3	HUMAN	4	GOLGA3
29	10.97	23.41	Ras-related protein Rab-1A	HUMAN	5	RAB1A
30	10.96	14.83	RNA-binding motif protein, X chromosome	HUMAN	6	RBMX
31	10.8	19.95	Heat shock protein HSP 90-alpha	HUMAN	14	HSP90AA1
32	10.76	11.07	D-3-phosphoglycerate dehydrogenase	HUMAN	5	PHGDH
33	10.61	19.76	Heterogeneous nuclear ribonucleoprotein F	HUMAN	7	HNRNPF
34	10.49	41.54	40S ribosomal protein S15a	HUMAN	7	RPS15A
35	10.26	16.29	40S ribosomal protein S3a	HUMAN	5	RPS3A
36	10.23	28.08	40S ribosomal protein S16	HUMAN	5	RPS16
37	9.41	11.15	T-complex protein 1 subunit alpha	HUMAN	5	TCP1
38	9.32	9.29	Heterogeneous nuclear ribonucleoprotein K	HUMAN	4	HNRNPK
39	9.22	8.59	Poly(U)-binding-splicing factor PUF60	HUMAN	5	PUF60
40	8.75	10.84	Serum albumin	HUMAN	8	ALB
41	8.71	24.85	40S ribosomal protein S10	HUMAN	4	RPS10
42	8.48	35.92	Histone H4	HUMAN	4	HIST1H4A
43	8.44	22.50	Four and a half LIM domains protein 3	HUMAN	5	FHL3
44	8.34	8.70	Keratin, type II cytoskeletal 8	HUMAN	4	KRT8
45	8.31	10.70	78 kDa glucose-regulated protein	HUMAN	6	HSPA5
46	8.23	4.31	Elongation factor 2	HUMAN	4	EEF2
47	8.2	6.59	Calnexin	HUMAN	4	CANX
48	8.16	11.44	Heterogeneous nuclear ribonucleoproteins C1/C2	HUMAN	4	HNRNPC
49	8.04	24.51	40S ribosomal protein S5	HUMAN	5	RPS5
50	7.93	6.97	Endoplasmin	HUMAN	6	HSP90B1
51	7.64	22.52	40S ribosomal protein S14	HUMAN	5	RPS14
52	7.29	17.41	14-3-3 protein gamma	HUMAN	3	YWHAG
53	7.19	2.94	Rho guanine nucleotide exchange factor 2	HUMAN	3	ARHGEF2
54	6.86	24.24	60S ribosomal protein L12	HUMAN	3	RPL12
55	6.8	20.22	60S ribosomal protein L11	HUMAN	4	RPL11
56	6.66	7.28	Insulin-like growth factor 2 mRNA-binding protein 1	HUMAN	3	IGF2BP1
57	6.47	7.87	Zyxin	HUMAN	4	ZYX
58	6.32	27.14	60S ribosomal protein L23	HUMAN	3	RPL23
59	6.27	25.77	40S ribosomal protein S7	HUMAN	4	RPS7
60	6.15	13.73	Glyceraldehyde-3-phosphate dehydrogenase	HUMAN	3	GAPDH
61	6.06	9.45	Leucine-rich repeat-containing protein 59	HUMAN	3	LRRC59
62	6.05	13.36	Trypsin-1	HUMAN	7	PRSS1
63	6	64.49	Tubulin beta-4B chain	HUMAN	50	TUBB4B
64	6	44.89	Tubulin alpha-3C/D chain	HUMAN	26	TUBA3C
65	6	46.38	40S ribosomal protein S28	HUMAN	3	RPS28
66	5.66	34.88	Signal recognition particle 9 kDa protein	HUMAN	3	SRP9
67	5.51	3.43	Enhancer of mRNA-decapping protein 4	HUMAN	3	EDC4
68	5.42	62.61	60S acidic ribosomal protein P2	HUMAN	4	RPLP2
69	5.42	6.72	Thyroid receptor-interacting protein 6	HUMAN	3	TRIP6
70	5.33	3.09	Probable ATP-dependent RNA helicase DDX5	HUMAN	2	DDX5
71	5.28	1.12	CAD protein	HUMAN	2	CAD
72	5.13	2.51	Bifunctional glutamate/proline--tRNA ligase	HUMAN	4	EPRS
73	4.93	24.69	Nuclease-sensitive element-binding protein 1	HUMAN	3	YBX1

74	4.83	2.16	Exportin-2	HUMAN	2	CSE1L
75	4.77	22.40	60S ribosomal protein L9	HUMAN	4	RPL9
76	4.65	5.18	Alpha-2-HS-glycoprotein	HUMAN	2	AHSG
77	4.46	8.46	Heterogeneous nuclear ribonucleoprotein H	HUMAN	5	HNRNPH1
78	4.45	3.59	Aspartate--tRNA ligase, cytoplasmic	HUMAN	2	DARS
79	4.4	3.76	Protein disulfide-isomerase A3	HUMAN	2	PDIA3
80	4.4	3.61	TRAF-type zinc finger domain-containing protein 1	HUMAN	2	TRAFD1
81	4.38	2.04	Protein diaphanous homolog 1	HUMAN	2	DIAPH1
82	4.28	26.56	Ubiquitin-60S ribosomal protein L40	HUMAN	3	UBA52
83	4.28	2.40	Importin subunit beta-1	HUMAN	2	KPNB1
84	4.27	9.76	Cyclin-dependent kinase 1	HUMAN	3	CDK1
85	4.24	15.65	Hemoglobin subunit gamma-2	HUMAN	2	HBG2
86	4.22	5.07	Polyadenylate-binding protein 3	HUMAN	3	PABPC3
87	4.17	22.69	40S ribosomal protein S20	HUMAN	3	RPS20
88	4.16	22.62	40S ribosomal protein S27	HUMAN	2	RPS27
89	4.16	4.02	T-complex protein 1 subunit theta	HUMAN	2	CCT8
90	4.14	18.75	60S ribosomal protein L22	HUMAN	2	RPL22
91	4.12	18.31	Hemoglobin subunit zeta	HUMAN	2	HBZ
92	4.12	4.97	RuvB-like 2	HUMAN	2	RUVBL2
93	4.08	7.12	40S ribosomal protein SA	HUMAN	2	RPSA
94	4.02	39.29	40S ribosomal protein S29	HUMAN	3	RPS29
95	4	17.53	Poly(rC)-binding protein 2	HUMAN	5	PCBP2
96	3.98	3.85	T-complex protein 1 subunit gamma	HUMAN	2	CCT3
97	3.92	5.44	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	HUMAN	3	RPN1
98	3.88	4.11	TAR DNA-binding protein 43	HUMAN	2	TARDBP
99	3.87	6.16	Eukaryotic initiation factor 4A-I	HUMAN	2	EIF4A1
100	3.72	3.36	PDZ and LIM domain protein 5	HUMAN	2	PDLIM5
101	3.64	10.09	Ras-related protein Rab-11B	HUMAN	2	RAB11B
102	3.62	3.63	YTH domain-containing family protein 2	HUMAN	2	YTHDF2
103	3.57	1.83	Trinucleotide repeat-containing gene 6A protein	HUMAN	3	TNRC6A
104	3.46	3.03	Heterogeneous nuclear ribonucleoprotein U	HUMAN	2	HNRNPU
105	3.42	2.10	Fermitin family homolog 3	HUMAN	1	FERMT3
106	3.29	4.07	Nicotinamide phosphoribosyltransferase	HUMAN	2	NAMPT
107	3.29	5.04	RuvB-like 1	HUMAN	2	RUVBL1
108	3.26	2.95	T-complex protein 1 subunit eta	HUMAN	2	CCT7
109	3.1	0.41	eIF-2-alpha kinase activator GCN1	HUMAN	1	GCN1
110	3.1	8.43	LIM and SH3 domain protein 1	HUMAN	2	LASP1
111	3.01	0.89	Trifunctional purine biosynthetic protein adenosine-3	HUMAN	1	GART
112	2.99	3.14	60 kDa heat shock protein, mitochondrial	HUMAN	2	HSPD1
113	2.94	1.15	Alpha-2-macroglobulin	HUMAN	2	A2M
114	2.92	3.44	Polypyrimidine tract-binding protein 3	HUMAN	2	PTBP3
115	2.89	2.44	Methionine--tRNA ligase, cytoplasmic	HUMAN	2	MARS
116	2.86	4.85	Putative RNA-binding protein Luc7-like 2	HUMAN	2	LUC7L2
117	2.85	10.43	60S ribosomal protein L30	HUMAN	1	RPL30
118	2.82	13.23	THO complex subunit 4	HUMAN	2	ALYREF
119	2.69	2.83	Heterogeneous nuclear ribonucleoproteins A2/B1	HUMAN	1	HNRNPA2B1
120	2.63	5.03	Peroxiredoxin-1	HUMAN	1	PRDX1
121	2.62	1.92	Nucleolar RNA helicase 2	HUMAN	1	DDX21
122	2.59	3.32	ATP-dependent RNA helicase DDX3X	HUMAN	2	DDX3X
123	2.59	0.94	Leucine--tRNA ligase, cytoplasmic	HUMAN	1	LARS
124	2.59	2.20	Paxillin	HUMAN	1	PXN
125	2.58	6.32	Eukaryotic translation elongation factor 1 epsilon-1	HUMAN	1	EEF1E1
126	2.54	2.37	116 kDa U5 small nuclear ribonucleoprotein component	HUMAN	2	EFTUD2
127	2.5	1.06	Large proline-rich protein BAG6	HUMAN	1	BAG6
128	2.37	1.09	Reticulon-4	HUMAN	1	RTN4
129	2.36	5.44	Nuclear migration protein nudC	HUMAN	2	NUDC
130	2.35	1.13	Cytoskeleton-associated protein 5	HUMAN	2	CKAP5
131	2.34	2.72	Voltage-dependent anion-selective channel protein 2	HUMAN	1	VDAC2
132	2.33	3.78	X-ray repair cross-complementing protein 6	HUMAN	2	XRCC6
133	2.33	1.69	Lactotransferrin	HUMAN	1	LTF
134	2.3	8.63	14-3-3 protein epsilon	HUMAN	2	YWHAE
135	2.28	2.84	ATP-dependent RNA helicase DDX1	HUMAN	1	DDX1
136	2.24	8.00	40S ribosomal protein S25	HUMAN	1	RPS25
137	2.22	12.43	Cellular nucleic acid-binding protein	HUMAN	2	CNBP
138	2.18	1.88	Coatomer subunit beta'	HUMAN	2	COPB2
139	2.14	4.43	Catechol O-methyltransferase	HUMAN	1	COMT
140	2.13	2.72	60S ribosomal protein L8	HUMAN	2	RPL8
141	2.11	38.84	Tubulin alpha-4A chain	HUMAN	25	TUBA4A
142	2.11	7.39	60S ribosomal protein L18a	HUMAN	1	RPL18A
143	2.1	2.62	Protein-glutamine gamma-glutamyltransferase 2	HUMAN	2	TGM2
144	2.07	2.70	Ras GTPase-activating protein-binding protein 2	HUMAN	1	G3BP2
145	2.05	2.83	Muscleblind-like protein 1	HUMAN	1	MBNL1
146	2.02	1.00	Extended synaptotagmin-1	HUMAN	1	ESYT1
147	2.01	1.22	Transportin-1	HUMAN	1	TNPO1
148	2.01	2.81	Eukaryotic translation initiation factor 3 subunit G	HUMAN	1	EIF3G
233	**0.42**	**11.38**	**Protein argonaute-4**	**HUMAN**	**10**	**AGO4**
684	**0**	**11.51**	**Protein argonaute-3**	**HUMAN**	**11**	**AGO3**

5.9.2 Evaluation of the thermostability of the miRNA-target RNA duplex by melting curve analysis

Melting curve analysis was used to assess the thermostability of the miRNA-target RNA duplexes. The analysis was conducted by the use of genuine miRNA-target site interacting duplexes that was obtained from K562 cells (see Figure 4.11.2). The release of the interacting duplexes was conducted by proteinase K digestion in the concentration of 0.1 U/µl of proteinase K at 4℃. A serial of digestion times were investigated, i.e. 1, 2, 4, 6, 9, 12, 15, 18, 24, 36 hours through three different batch of experiments. The first 4 time points were in one batch. The three successive time points following by the first four time points were in the second batch. The last 3 time points were in the last batch. Nine melting conditions (see Appendix C Table 8) were performed on each sample collected from different time points to obtain the most regular melting peak.

Figure 5.9.1 shows the melting curves of the interacting duplexes that were released by digestion with proteinase K for 10 different lengths of time. For these 10 different times of digestion (1 hour to 36 hours), the patterns of the melting peaks did not show any obvious differences. This suggested that the miRNA-target RNA duplexes could be released from the AGO complex even with one hour of proteinase K digestion, and were still stable even after a prolonged time in 1X Detergent-free Lysis Buffer at 4℃ (the experiment was tested up to 36 hours). The latter observation was supported by re-melting analysis of the samples, in which the RNA duplexes were completely melted before a repeat melting analysis was performed again. As expected, the originally identified melting peak disappeared in the re-melting experiments (Figure 5.9.2). The disappearance of the duplexes in the re-melting analysis implied that they could not be reformed once completely melted (see discussion in Section 6.3). Therefore, melting peaks would not be observed if the duplexes were not

stable in the prolonged incubation. The average melting temperatures of the *in vivo* extracted miRNA-target RNA duplexes were 36.74℃, 38.01℃ and 38.82℃ when the volumes of the sample added were 7.5 µl, 10 µl and 12.5 µl respectively.

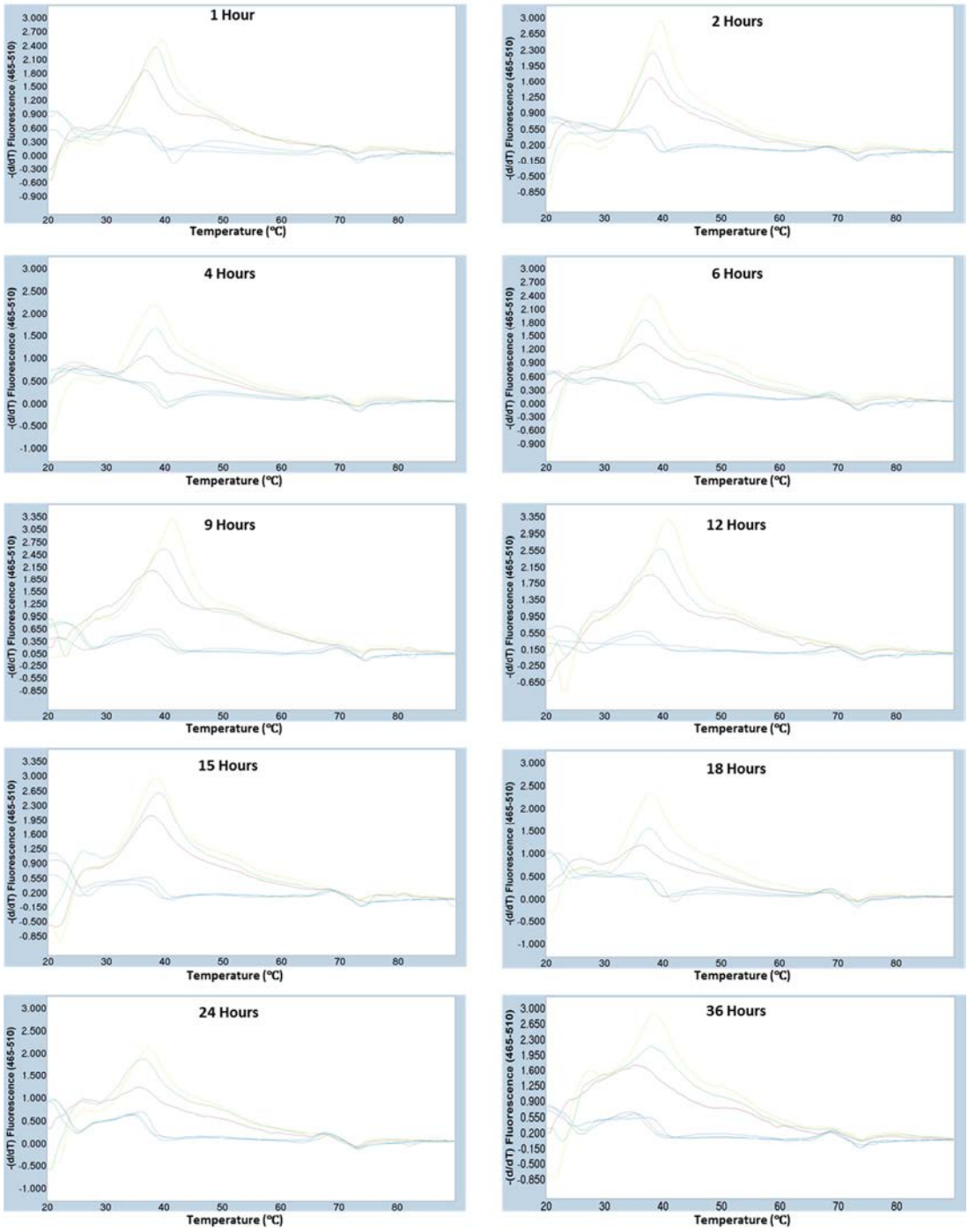


Figure 5.9.1 Melting curve analysis of miRNA-target RNA duplexes extracted from K562 cells for investigating the thermostability of this duplex mixture. Digestion by proteinase K for releasing the interacting RNA duplex from AGO complex was carried out for 10 different durations and evaluated by melting curve analysis of the extracted RNA duplexes . The samples collected from these time points (shown on each plot) were analyzed individually. The mixtures for melting analysis contained 1.5 µM Syto 9 and different volumes of samples (7.5 µl, red curves; 10 µl, green curves; and 12.5 µl, yellow curves). The blue lines in each plot are the three corresponding control samples (the same setup with the sample, but without duplex).

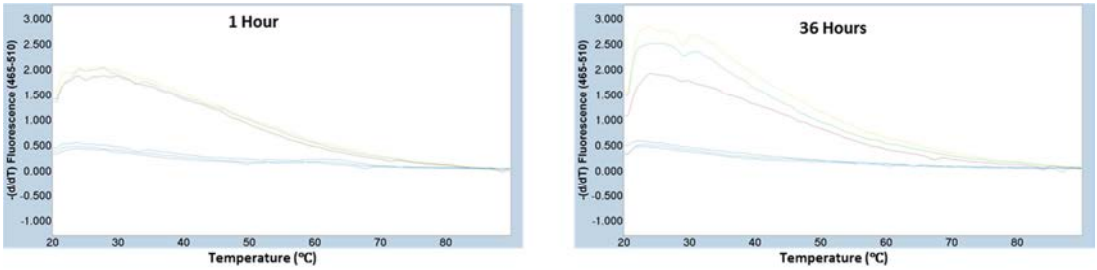


Figure 5.9.2 Re-melting analysis of miRNA-target RNA duplexes extracted from K562 cells. Two sets of typical re-melting curve are shown in here. These two samples have been collected after 1 hour (the first time point) and 36 hours (the last time point) of proteinase K digestion. The samples collected from these two time points (shown on each plot) were analyzed individually. The mixtures for melting analysis contained 1.5 µM Syto 9 and different volumes of samples (7.5 µl, red curves; 10 µl, green curves; and 12.5 µl, yellow curves). The blue lines in each plot are the three corresponding control samples (the same setup with the sample, but without duplex).

5.9.3 The quantity of the isolated miRNA-target RNA duplex

Two approaches were used to investigate the quantity of the yielded miRNA-target RNA duplex. First, the quantity was determined by Bioanalyzer 2100 (Agilent). Second, the quantity was determined by Chimeric Sequence qPCR assay that was developed in this study.

Quantification by Bioanalyzer 2100

Since the molecular mass of the isolated miRNA-target RNA duplex could roughly be determined by Bioanalyzer 2100, the molar amount of the RNA molecules was calculated according the equations shown in Section 4.11.3. The concentration of the isolated duplex was measured by either Bioanalyzer 2100 with Agilent RNA 6000 Pico Kit (Agilent) or NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies). The volume of the isolated duplex was equal to the elution volume (30 µl) in RNA extraction by miRNeasy Mini Kit (Qiagen). The results indicated that the molar amounts of the yielded RNA molecules from the effecting AGO complex were quite similar after digestion by different concentrations of RNase A Ribonuclease (Table 5.9.2). Their amounts were in the order of several picomoles (pmol). Similarly, the molar amounts of the yielded RNA molecules from the effecting AGO complex were also similar after digestion by different concentrations of RNase One™ Ribonuclease, and were in the order of several picomoles (Table 5.9.3). These results agreed with the expectation because the yielded molar amount of effecting Ago complex was the same between the reactions when the amounts of protein lysate and anti-Ago antibody-conjugated beads for initiating each experiment were the same.

Table 5.9.2 The quantity of the extracted RNA fragments from effecting AGO complex after RNase A Ribonuclease treatment (RNA length and concentration were each shown as an average of 3 replicates).

	RNA Average length			RNA concentration measured by Bioanalyzer 2100			RNA concentration measured by NanoDrop ND-1000		
	length	CV (%)	Calculated molecular mass	Average concentration (ng/µl)	Volume (µl)	pmol (RNA molecules)	Average concentration (ng/µl)	Volume (µl)	pmol (RNA molecules)
no RNase	637.33	21.52%	204882.45	25.21	30	3.69	22.28	30	3.26
undiluted	103.33	1.48%	33233.49	6.58	30	5.94	8.57	30	7.73
1:100	164.33	6.68%	52841.33	12.55	30	7.12	13.20	30	7.50
1:1000	236.00	16.84%	75877.86	16.72	30	6.61	15.87	30	6.27
1:100000	648.00	21.30%	208311.14	18.39	30	2.65	19.77	30	2.85

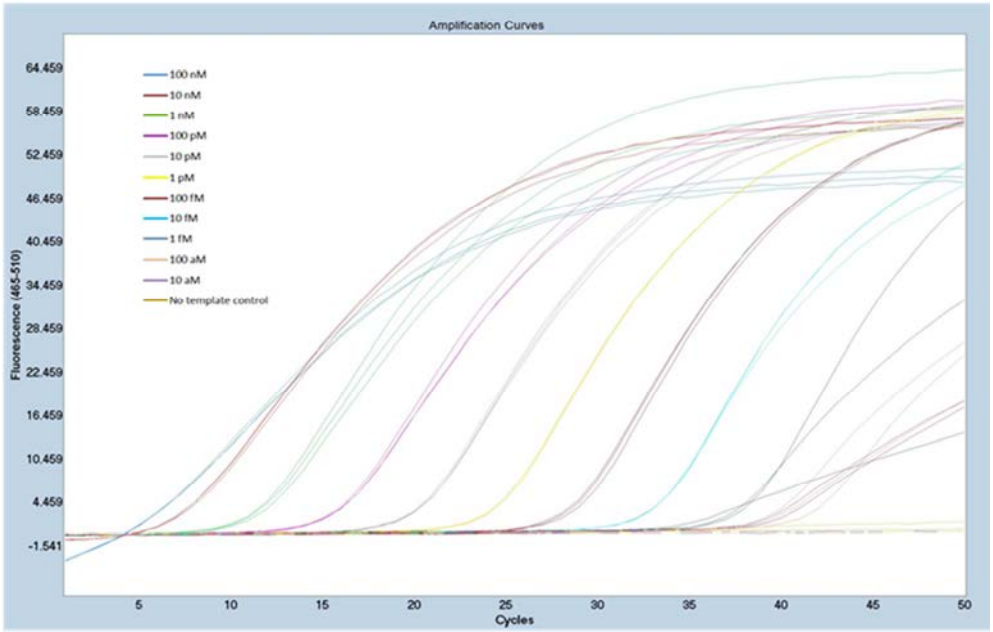
Table 5.9.3 The quantity of the extracted RNA fragments from effecting AGO complex after RNase One™ Ribonuclease treatment (RNA length and concentration were each shown as an average of 3 replicates).

	RNA Average length			RNA concentration measured by Bioanalyzer 2100			RNA concentration measured by NanoDrop ND-1000		
	length	SD (%)	Calculated molecular mass	Average concentration (ng/µl)	Volume (µl)	pmol (RNA molecules)	Average concentration (ng/µl)	Volume (µl)	pmol (RNA molecules)
no RNase	892.00	16.99%	286742.50	24.08	30	2.52	26.13	30	2.73
undiluted	105.33	3.33%	33876.37	7.41	30	6.56	11.40	30	10.10
1:100	195.33	12.27%	62805.97	17.14	30	8.19	19.37	30	9.25
1:1000	407.33	12.13%	130951.25	22.05	30	5.05	23.68	30	5.43
1:100000	894.33	11.49%	287492.53	20.46	30	2.14	24.68	30	2.58

Quantification by Chimeric Sequence qPCR assay

The Chimeric Sequence Quantification Assay was evaluated by measuring serial dilutions of a synthetic template with known concentration. The measurement was carried out on LightCycler® 480 (Roche) in 96-well format. The regular amplification curves of sample concentrations from 10 nM to 1 fM indicated that the working range of the current quantification assay was between 10 nM and 1 fM (Figure 5.9.3a). The efficiency of the current quantification assay was 1.7074, indicating that the PCR was amplified by 1.7074 folds in each cycle (Figure 5.9.3b). The R^2 of this evaluation was 1 (Figure 5.9.3b).

a.



b.

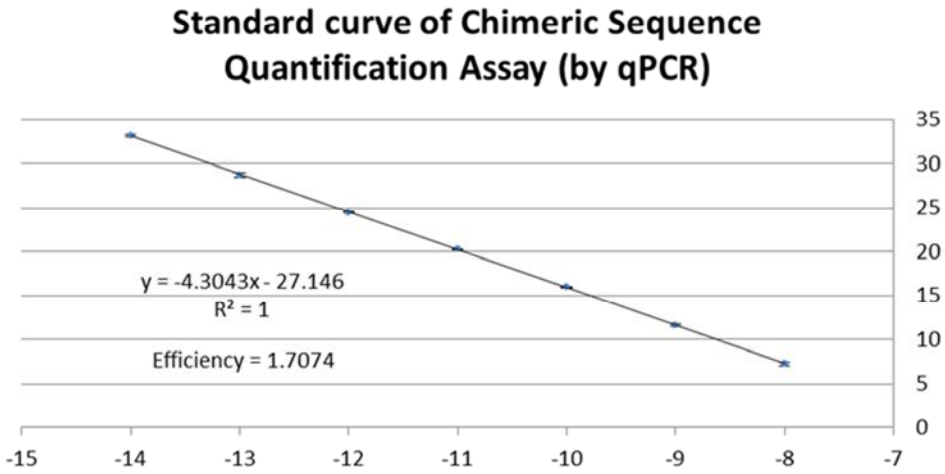


Figure 5.9.3 qPCR results for determining the working range and PCR efficiency of the Chimeric Sequence Quantification Assay. (a) The figure shows the amplification curves of serial dilutions of the synthetic template. (b) From the corresponding standard curve, the PCR amplification efficiency was found to be 1.7074 and the working range of the current quantification assay was between 10 nM and 1 fM. The standard curve was constructed by plotting the log value of the concentration on x-axis and the crossing point (Cp-value) on the y-axis. Each data point was the mean value (± SD) for 3 replicate measurements.

5.10 Determination of optimal condition for digestion by proteinase K

Two different experiments were used to investigate the digestion of proteinase K under different conditions. First, a mock experiment based on digestion of BSA was used to demonstrate the digestion of proteinase K under different temperatures. Second, an experiment monitoring the concentration of released amino acids was used to further evaluate the digestion of proteinase K under different temperatures.

5.10.1 Separation of proteinase K digested BSA by gel electrophoresis

Four time-course mock experiments of digesting BSA at a concentration of 1 mg/ml in 1X Detergent-free Lysis Buffer by 0.04U/µl of proteinase K at 4℃, 8℃, 12℃ and 16℃ were conducted to demonstrate the digestion completeness of BSA by proteinase K. The time points were 0, 0.5, 1, 2, 4, 6, 9, 12 and 24 hours and the volume of each individual reaction was 25 µl. There was no obvious difference among the four digestion temperatures (Figure 5.10.1). At any one temperature being tested, proteinase K efficiently digested the BSA substrate: no intact BSA remained in the reactions even if the digestion time was as short as 0.5 hours. On the other hand, small amounts of small protein fragments were detected in each reaction (from 0 hour to 24 hours digestion). The fragments were less than 15 kDa in size. This observation implied that the majority of BSA was digested into fragments that were too small to be retained in the current electrophoresis. The relative less amount of these small protein fragments at time 0 was most likely owing to some degradation in the original BSA protein.

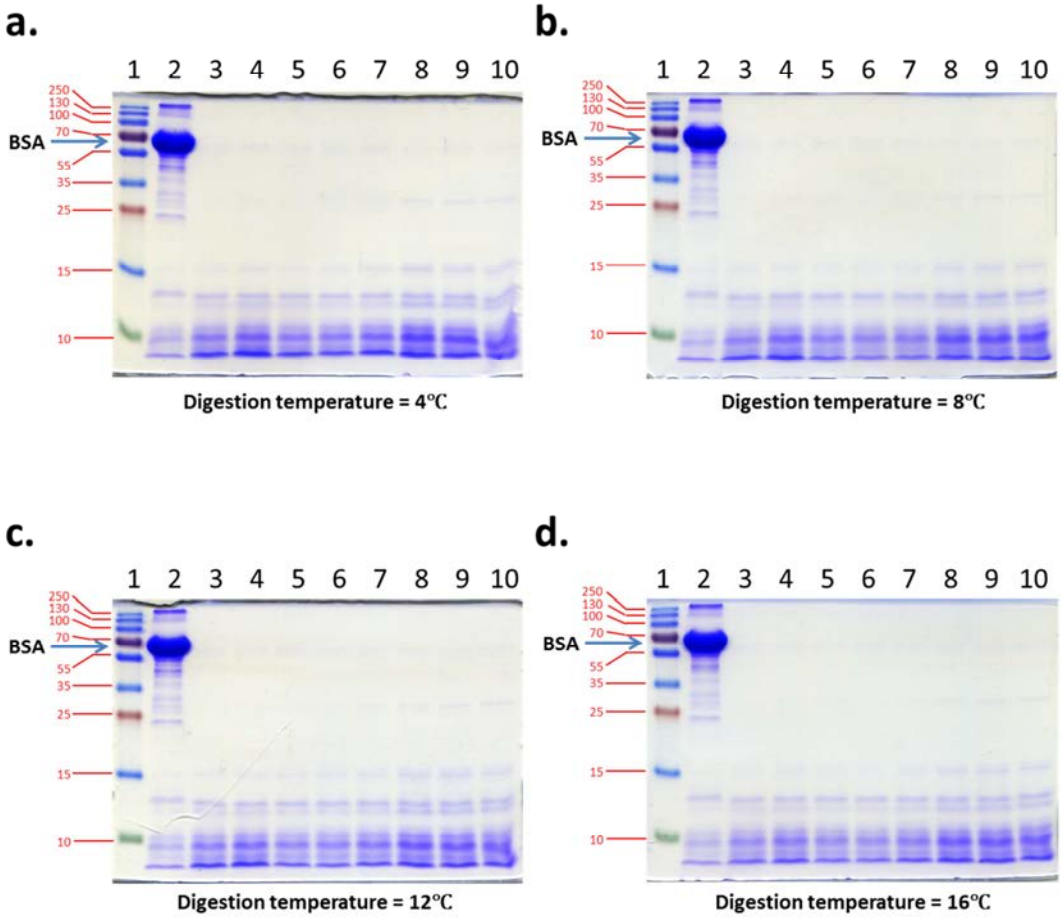


Figure 5.10.1 SDS-PAGE of BSA and its breakdown products after digestion by proteinase K. Four sets of time-course study were conducted for evaluating the digestion completeness of proteinase K at 4℃, 8℃, 12℃ and 16℃. From lane 1 to 10, the samples were PageRuler Plus Prestained Protein Ladder (Thermo Scientific), BSA digested for 0, 0.5, 1, 2, 4, 6, 9, 12 and 24 hours by proteinase K respectively.

5.10.2 Monitoring of amino acid concentration during digestion of isolated RISCs by proteinase K

L-Amino Acid Quantification Kit (Sigma-Aldrich) measures amino acid concentration, and was used to further investigate the digestion by proteinase K (NEB) of isolated AGO complex at low temperature (see Figure 4.12.2). Two temperatures for proteinase K digestion were investigated: 4°C and 16°C. In addition, two different concentrations of proteinase K in the digestion reaction were also investigated at each of the above digestion temperatures: 0.038 U/µl and 0.1 U/µl. Therefore, four different digestion conditions were studied accordingly. For each individual condition, a time-course study was performed and the time points were 0, 2, 4, 6, 12, 18, 24 hours.

The digestion of isolated AGO complex with 0.038 U/µl of proteinase K revealed that the net concentrations of the released amino acid increased with increasing digestion time and reached ~200 µM after 24 hours digestion both at 4°C and 16°C (Figure 5.10.2a and b, graphs on the right). Likewise, the digestion of isolated AGO complex with 0.1 U/µl of proteinase K revealed the net concentrations of the released amino acid also increased with increasing digestion time and reached ~400 µM after 24 hours digestion both at 4°C and 16°C (Figure 5.10.2c and d, graphs on the right). Notably, an increased amount of input proteinase K in the digestion increased the net concentration of released amino acid by about 2-fold. However, the digestion was not affected by temperature.

The R^2 values of some plots were not very high (0.6 – 0.7) probably because of low concentration of amino acids in the mixture and/or small differences of amino acids between the experimental and the control mixtures, particularly for the first few time points (their time interval was too close). However, the trend of increasing concentrations of released amino acid over time was still very obvious.

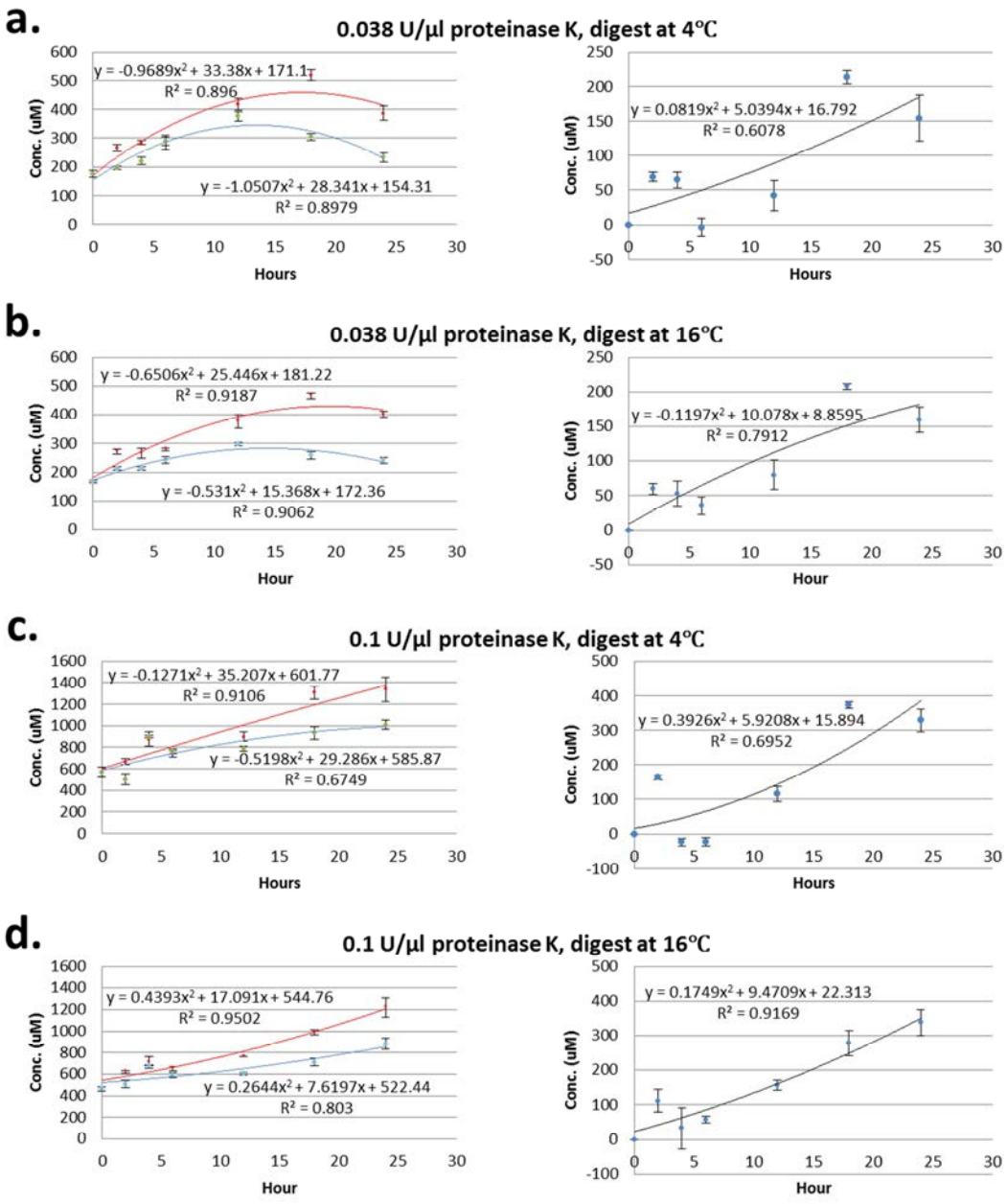


Figure 5.10.2 The result in monitoring the released amino acid concentration during the digestion of isolated AGO complex by proteinase K. Digestion of isolated AGO complex was performed using proteinase K at a concentration of (a) 0.038 U/µl at 4°C, (b) 0.038 U/µl at 16°C, (c) 0.1 U/µl at 4°C, and (d) 0.1 U/µl at 16°C. All graphs are plots of amino acid concentration against time of digestion. On the left-hand side are the plots for digestion mixtures with AGO complex added (red lines, experimental curves) or without AGO complex added (blue lines, control curves – background digestion). On the right-hand side are plots of net concentration of released amino acid in the samples. The amino acid concentrations show mean values ± SD (n = 3).

5.11 MicroRNA mismatch repairs

Both inosine and 5-nitroindole can be used to reduce the complexity of the miRNA 5' repairer and miRNA 3' repairer. In order to evaluate the feasibility of including either nucleoside in the two repairers, two different sets of mock experiments were designed for evaluating the ligation of these repairers to the 5' end and the 3' end of miRNAs.

5.11.1 Random cytosine/inosine tetramer

Inosine is a natural base found in RNAs. This base can pair up with adenosine (A), uracil (U) and cytosine (C), but not guanosine (G) in RNA [319]. Thus, inosine is potentially able to reduce the complexity of the repairers. Assuming that a natural base is compatible for T4 RNA Ligase 1, 4-base repairers with random cytosine/inosine were first evaluated. Note that, as a single-stranded RNA ligase, T4 RNA ligase 1 joins two single-stranded RNA oligos rather than nicks on double-stranded RNAs as long as the 5' end has a phosphate group and the 3' end a hydroxyl group.

5.11.1.1 5' end repair

The miRNA 5' end repairer "Random C/I Tetramer" (Table 4.2.1) with hydroxyl groups on both ends was first evaluated (see Figure 4.13.3). The results showed that the ligation of "Random C/I Tetramer" was not efficient in both mock experiments for mimicking the ligation of 5' repairer to the 5' end of the miRNAs with 3 mismatches (Figure 5.11.1a) or 2 mismatches (Figure 5.11.1b). In the above evaluation, six ligation times (0, 2, 4, 8, 16 24 and 36 hours) were investigated. The two investigated mock miRNAs were ligated with the "Random C/I Tetramer" very slowly and the ligation of the mock miRNA with 3 mismatches (Artificial Mir3) was better than that with 2 mismatches (Artificial Mir4) (Figure 5.11.1a vs.

Figure 5.11.1b). As the ligation of "Random C/I Tetramer" was inefficient, it was suspected that inosine was not a good substrate for T4 RNA Ligase 1.

"Tetramer N" (Table 4.2.1) was then used to confirm whether the ligation of a 5' repairer to the 5' end of mock miRNA was repressed by the inclusion of inosine in the 5' repairer. With the replacement of "Random C/I Tetramer" by "Tetramer N", the ligation was improved tremendously. Both the ligation of "Tetramer N" to the 5' end of the mock miRNAs with 3 mismatches (Artificial Mir3; Figure 5.11.1c) and to the 5' end of the mock miRNAs with 2 mismatches (Artificial Mir4; Figure 5.11.1d) were almost complete in 2 hours. The ligation of "Tetramer N" was slightly faster for the mock miRNA with 3 mismatches than for that with 2 mismatches. These results demonstrated that the ligation efficiency of the 5' repairer could be improved if the inosine was removed from the repairer.

Two other 5' repairers, "Tetramer (CCCI)" and "Tetramer C" (Table 4.2.1), were used to demonstrate more clearly the effect of inosine in the 5' repairer on ligation. The 3' base was an inosine for the former repairer and a cytosine for the latter repairer. Both the repairers were tested for ligation to the 5' end of a mock miRNA with 3 mismatches (Artificial Mir3) for up to 2 hours. The ligation of "Tetramer (CCCI)" to the 5' end of the mock miRNAs with 3 mismatches was much less efficient than the ligation of "Tetramer C" to the same set of mock miRNA-target RNA duplex (Figure 5.11.1e vs. Figure 5.11.1f). These results confirmed that the inosine was not a good substrate for T4 RNA Ligase 1.

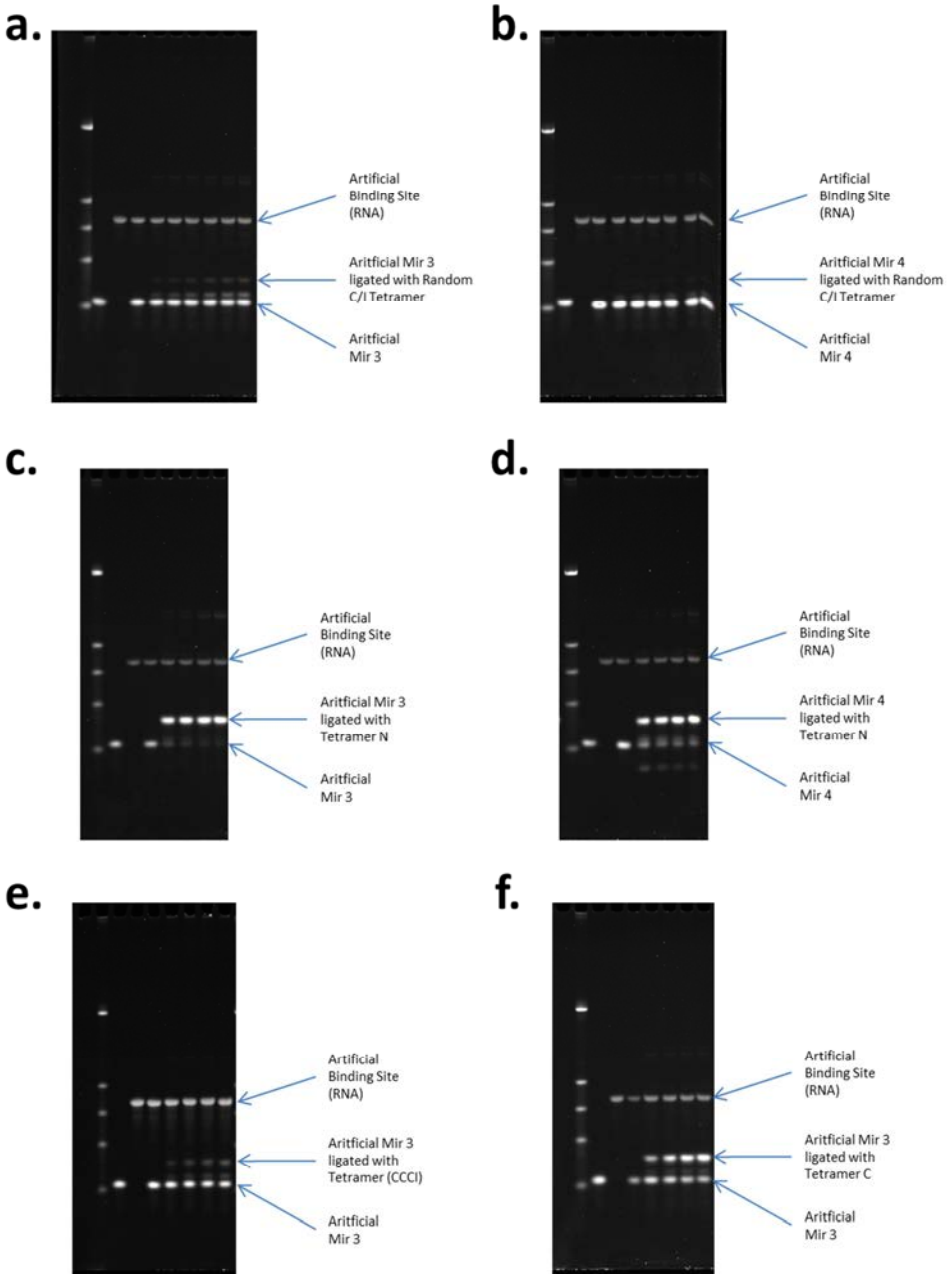


Figure 5.11.1 Mock experiment for evaluating the ligation of inosine-containing 5' repairer to the 5' end of mock miRNAs. Random C/I Tetramer (a & b) or Tetramer N (c & d) was ligated to the 5' end of the mock miRNAs that had 3 mismatches (Artificial Mir3; a & c) or 2 mismatches (Artificial Mir4; b & d) with the 43-nt Artificial Binding site (RNA). In addition, Tetramer (CCCI) (e) or Tetramer C (f) was ligated to the 5' end of the mock miRNA that had 3 mismatches (Artificial Mir3). All tested 5' repairers (Random C/I Tetramer and Tetramer N) were 4 nt long while the mock miRNAs were 22 nt long (see Figure 4.13.2). Therefore, the ligation products were 26 nt long. In (a) to (f), lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20nt from top to bottom), lane 2 was the mock miRNA, lane 3 was the mock target RNA, and the successive lanes were the time points studied. In (a) and (b), the time points were 0, 2, 4, 8, 16 24 and 36 hours. In (c) to (f), the time points were 0, 0.5, 1, 1.5 and 2 hours.

5.11.1.2 3' end repair

Two miRNA 3' end repairers with phosphate groups on both ends were evaluated: "Tetramer-Phosphorylated (ICCC)" and "Tetramer C-Phosphorylated" (Table 4.2.1). There was only one difference between these two repairers with the 5' end being an inosine for "Tetramer-Phosphorylated (ICCC)" and a cytosine for "Tetramer C-Phosphorylated". These two miRNA 3' end repairers were used to evaluate the effect of inosine in the miRNA 3' end repairers on their ligation to the 3' end of miRNA. In the evaluation of "Tetramer-Phosphorylated (ICCC)", the efficiency was very poor for the ligation of this 3' repairer to the mock miRNAs with 2 mismatches (Artificial Mir3) or 3 mismatches (Artificial Mir4) at the 3' end of the mock miRNAs (Figure 5.11.2a and Figure 5.11.2b respectively). The ligation was evaluated for up to 36 hours. In contrast, the repeat experiments with "Tetramer C-Phosphorylated" showed that the ligation efficiency was much better. Figure 5.11.2c shows the ligation result of "Tetramer C-Phosphorylated" to the mock miRNAs with 2 mismatches at the 3' end of the mock miRNA. Figure 5.11.2d shows the ligation result of "Tetramer C-Phosphorylated" to the mock miRNAs with 3 mismatches at the 3' end of the mock miRNA.

These results indicated that inosine in the miRNA 3' repairer drastically impaired the ligation of the repairer to the 3' end of the mock miRNAs and that inosine was not a good substrate for T4 RNA Ligase no matter in the donor end or acceptor end. Remarkably, the ligation of a repairer to the 5' end of the mock miRNA was much more efficient than that to the 3' end of the mock miRNA. For example, the ligation of "Tetramer C" (a tested 5' repairer) to the 5' end of the mock miRNA with 3 mismatches yielded more than half of the expected ligation product in about 2 hours (Figure 5.11.1f). In comparison, the ligation of "Tetramer C-Phosphorylated" (a tested 3' repairer) to the 3' end of the mock miRNA with 3

mismatches yielded only about half of ligation product about 16 hours of incubation and the result hardly changed even after prolonged incubation (Figure 5.11.2d).

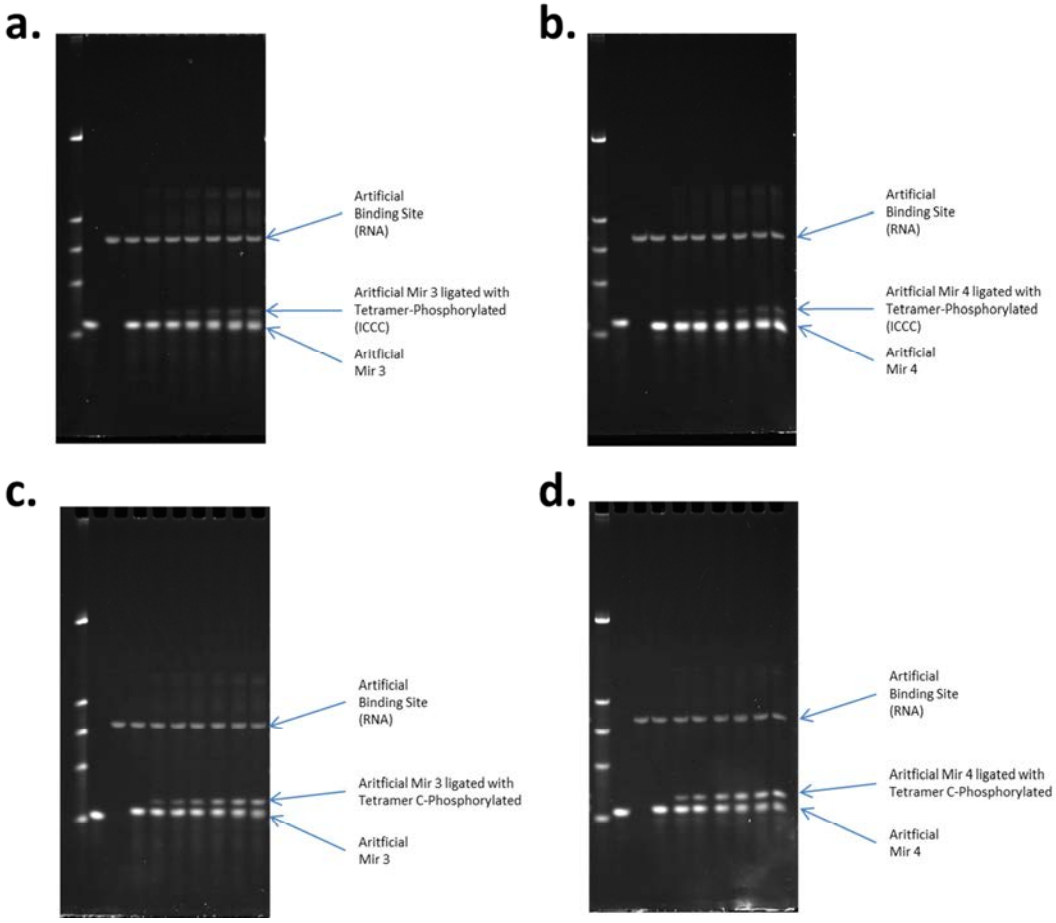


Figure 5.11.2 Mock experiment for evaluating the ligation of inosine-containing 3' repairer to the 3' end of mock miRNAs. Tetramer-Phosphorylated (ICCC) (a & b) or Tetramer C-Phosphorylated (c & d) was ligated to the 3' end of the mock miRNAs that had 2 mismatches (Artificial Mir3; a & c) or 3 mismatches (Artificial Mir4; b & d) with the 43-nt Artificial Binding site (RNA). All tested 3' repairers (Tetramer-Phosphorylated (ICCC) and Tetramer C-Phosphorylated) were 4 nt long while the mock miRNAs were 22 nt long (see Figure 4.13.2). Therefore, the ligation products were 26 nt long. In (a) to (d), lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20nt from top to bottom), lane 2 was the mock miRNA, lane 3 was the mock target RNA, and the successive lanes were the time points studied (0, 2, 4, 8, 16, 24 and 36 hours).

5.11.2 5-nitroindole-containing repairer

5-Nitroindole is a universal base that can pair up with all canonical DNA bases, i.e. A, T, C and G, non-selectively [320, 321]. As 5-Nitroindole (X) is an artificial base, it *may* not be a suitable substrate for T4 RNA Ligase 1 and other enzymes. The results in Section 5.11.2.1 and 5.11.2.2 show (i) whether the 5-nitroindole is a suitable substrate for T4 RNA Ligase 1 when included in constructing the repairers, and (ii) how close the base can be placed to the donor and the acceptor ends.

5.11.2.1 5' end repair

The miRNA 5' end repairer "5' Repairer (GXXGA)" (Table 4.2.1) with hydroxyl groups on both ends was first evaluated. The results showed that the ligation of "5' Repairer (GXXGA)" was not efficient in both mock experiments for mimicking the ligation of 5' repairer to the 5' end of the miRNAs with 3 mismatches (Figure 5.11.3a) or 2 mismatches (Figure 5.11.3b). In the above evaluation, five ligation times (0, 0.5, 1, 1.5 and 2 hours) were investigated. The two investigated mock miRNAs were ligated to the "5' Repairer (GXXGA)" very slowly and the ligation of the mock miRNA with 3 mismatches was better than that with 2 mismatches (Figure 5.11.3a vs. Figure 5.11.3b). As the ligation of "5' Repairer (GXXGA)" was inefficient, it was suspected that 5-nitroindole was not a good substrate for T4 RNA Ligase 1.

Another miRNA 5' repairer "5' Repairer (GXAGA)" (Table 4.2.1) with hydroxyl groups on both ends and with a replacement of "X" (5-nitroindole) by "A" at the third position from the acceptor (3') end was used to evaluate whether the ligation could occur when the "X" was placed at the fourth position from the acceptor end (one more base farther away when compared with "5' Repairer (GXXGA)"). With the replacement of "5' Repairer (GXXGA)" by "5' Repairer (GXAGA)", the ligation of the 5' repairer to the 5' end of mock

miRNAs was improved dramatically. The ligation of “5’ Repairer (GXAGA)” to the 5’ end of the mock miRNAs with 3 mismatches (Artificial Mir3) was almost complete in 2 hours (Figure 5.11.3c) whereas the ligation of the same repairer to the 5’ end of the mock miRNAs with 2 mismatches (Artificial Mir4) was about 50% complete in 2 hours (Figure 5.11.3d). These results suggested that the ligation efficiency of the 5’ repairer would not be inhibited if 5-nitroindole was placed at the fourth position from the acceptor (3’) end of the miRNA 5’ repairer.

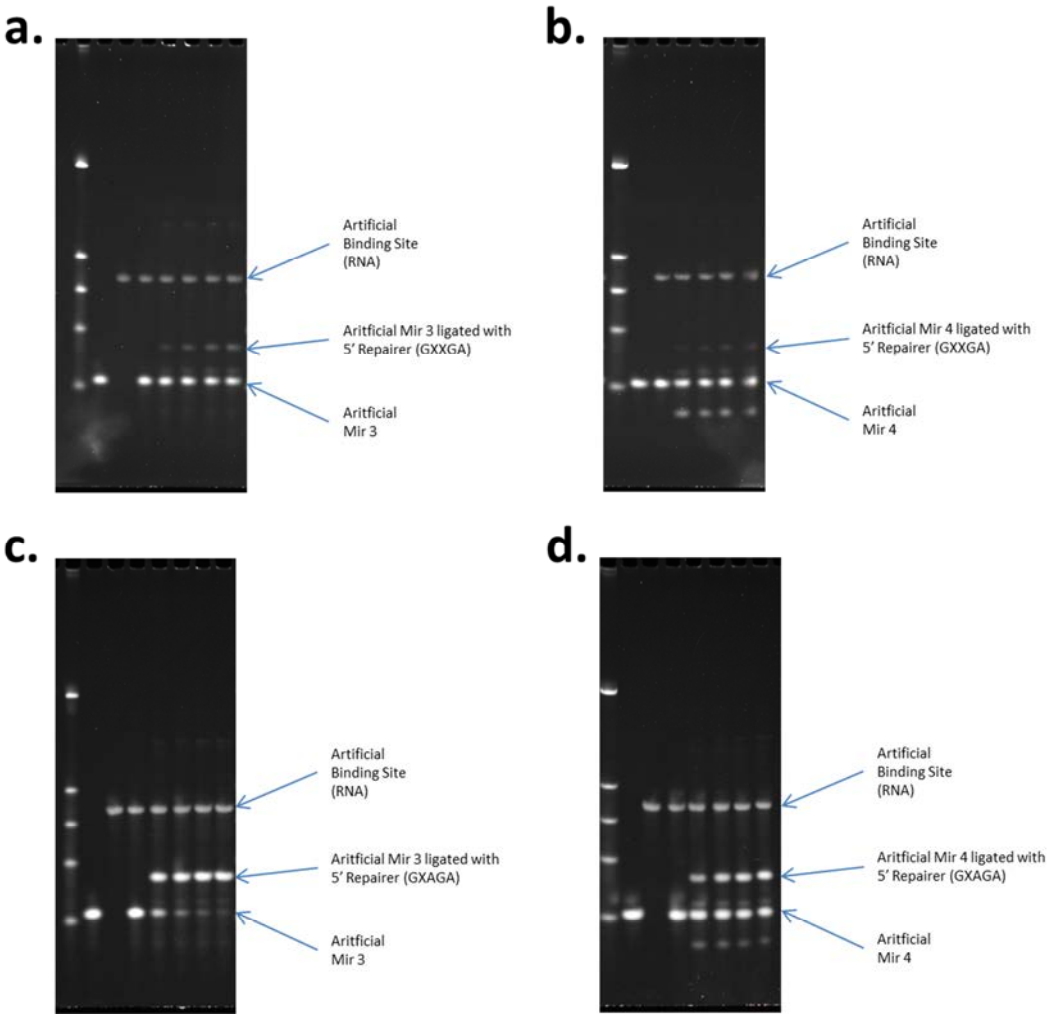


Figure 5.11.3 Mock experiment for evaluating the ligation of 5-nitroindole-containing 5' repairer to the 5' end of mock miRNAs. The 5' Repairer (GXXGA) (a & b) or 5' Repairer (GXAGA) (c & d) was ligated to the 5' end of the mock miRNAs that had 3 mismatches (Artificial Mir3; a & c) or 2 mismatches (Artificial Mir4; b & d) with the 43-nt Artificial Binding Site (RNA). All tested 5' repairers (5' Repairer (GXXGA) and 5' Repairer (GXAGA)) were 5-nt long while the mock miRNAs 22-nt long (see Figure 4.13.2). Therefore, the ligation products were 27-nt long. In (a) to (d), lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20nt from top to bottom), lane 2 was the mock miRNA, lane 3 was the mock target RNA (this lane was missing in (b) by mistake), and the successive lanes were the time points studied (0, 0.5, 1, 1.5, and 2 hours).

The ligation of “5’ Repairer (GXAGA)” was then evaluated again in NEBNext® Quick Ligation Reaction Buffer (NEB) to find reaction conditions for a faster ligation process. The evaluation of ligation was expanded to mock miRNAs with no mismatch and one mismatch as well.

With a change in the reaction buffer system (from T4 RNA Ligase Reaction Buffer (NEB) to NEBNext® Quick Ligation Reaction Buffer (NEB)), ligation of “5’ Repairer (GXAGA)” to the 5’ end of the mock miRNAs with 3 mismatches and 2 mismatches were both enhanced. With this change in the reaction buffer system, the ligation was complete in 0.5 hour (Figure 5.11.4d) rather than 2 hours (Figure 5.11.3c) for mock miRNA with 3 mismatches, and complete in 2 hours (Figure 5.11.4c) rather than half complete in 2 hours (Figure 5.11.3d) for mock miRNA with 2 mismatches. In the presence of NEBNext® Quick Ligation Reaction Buffer (NEB) again, the ligation of “5’ Repairer (GXAGA)” to the 5’ end of the mock miRNAs with one mismatch and no mismatch were complete in 2 hours (Figure 5.11.4b) and 4 hours (Figure 5.11.4a) respectively.

These results indicated that NEBNext® Quick Ligation Reaction Buffer (NEB) was able to speed up the ligation of short RNA (repairer), and the repairer containing 5-nitroindole four bases from the acceptor (3’) end could be efficiently ligated to 5’ end of mock miRNAs with 0 – 3 mismatches.

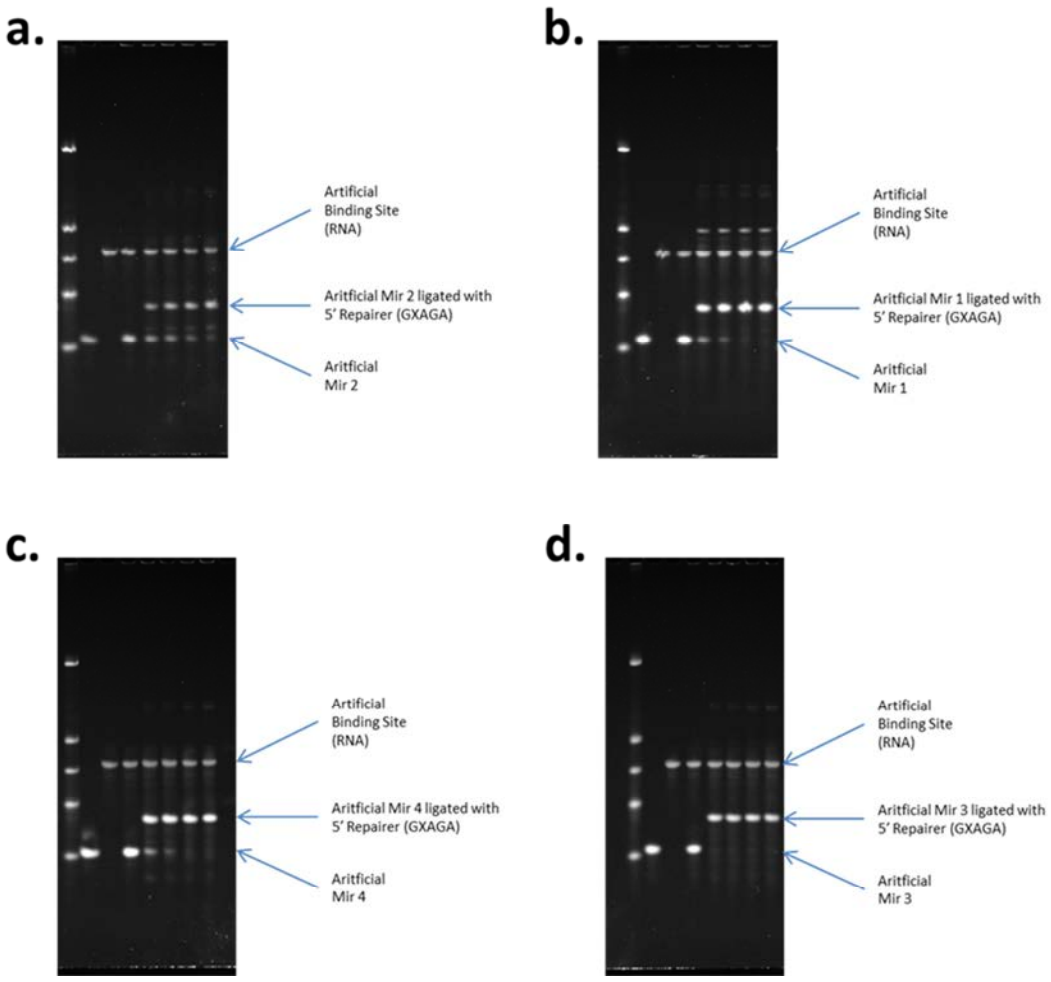


Figure 5.11.4 Mock experiment for evaluating the ligation of 5-nitroindole-containing 5' repairer to the 5' end of mock miRNAs in NEBNext® Quick Ligation Reaction Buffer (NEB). The 5' Repairer (GXAGA) was ligated to the 5' end of the mock miRNAs that had 0 mismatch (Artificial Mir2; a), 1 mismatch (Artificial Mir1; b), 2 mismatches (Artificial Mir4; c) or 3 mismatches (Artificial Mir3; d) with the 43-nt Artificial Binding Site (RNA). The tested 5' repairer (5' Repairer (GXAGA)) was 5-nt long while the mock miRNAs (Artificial Mir1 to Mir4) were 22-nt long. Therefore, the ligation products were 27-nt long. In (a) to (d), lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20nt from top to bottom), lane 2 was the mock miRNA, lane 3 was the mock target RNA, and the successive lanes were the time points studied (0, 0.5, 1, 2, and 4 hours).

5.11.2.2 3' end repair

In a similar way, the miRNA 3' end repairer "3' Repairer (AAXXG)" (Table 4.2.1) with hydroxyl groups on both ends was first evaluated using T4 RNA Ligase Reaction Buffer (NEB) in the ligation reaction. The results showed that the ligation of "3' Repairer (AAXXG)" was efficient in both mock experiments for mimicking the ligation of 3' repairer to the 3' end of the miRNAs with 2 mismatches (Figure 5.11.5a) or 3 mismatches (Figure 5.11.5b). In the above evaluation, seven ligation times (0, 2, 4, 8, 16, 24 and 36 hours) were investigated. The two investigated mock miRNAs were ligated to the "3' Repairer (AAXXG)" with complete ligation occurring in 16-24 hours of incubation, and the ligation to both kinds of 3' ends (with 2 and 3 mismatches) was comparable (Figure 5.11.5a vs. Figure 5.11.5b).

The evaluation of ligation of 5-nitroindole-containing miRNA 3' repairer was then expanded to mock miRNAs with no mismatch and one mismatch as well. The evaluated 3' repairer and reaction buffer were changed to "3' Repairer (AXXXG)" (one base closer to the 5' end when compared with "3' Repairer (AAXXG)") and NEBNext® Quick Ligation Reaction Buffer (NEB) respectively in this expanded evaluation. The ligation of this repairer to the 3' end of mock miRNAs with no mismatch, 1 mismatch, 2 mismatches and 3 mismatches were completed in 0.5 hours (Figure 5.11.5c), 0.5 hours (Figure 5.11.5d), 1 hour (Figure 5.11.5e) and 2 hours (Figure 5.11.5f) respectively.

These results suggested that the ligation of 5-nitroindole-containing miRNA 3' repairer to the 3' end of mock miRNA with 0 - 3 mismatches could be achieved efficiently even if the artificial base was located at the second position from the donor (5') end. The NEBNext® Quick Ligation Reaction Buffer (NEB) was also able to speed up these ligation reactions because the ligation of "3' Repairer (AAXXG)" in T4 RNA Ligase Reaction Buffer (NEB) took 16-24 hours for completion while the ligation of 3' Repairer (AXXXG)" in NEBNext® Quick

Ligation Reaction Buffer (NEB) took only 0.5-2 hours for completion (Figure 5.11.5a and b vs. c to f).

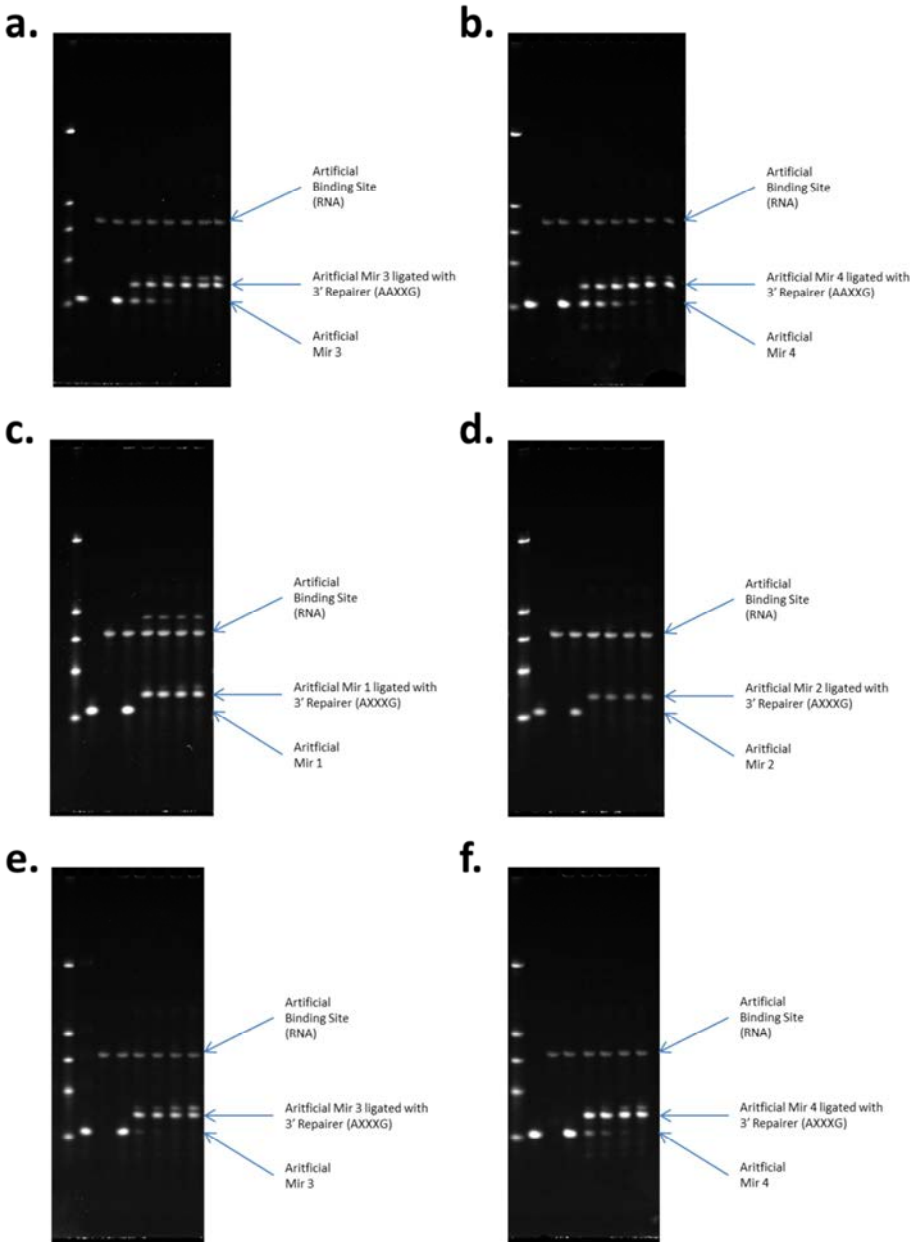


Figure 5.11.5 Mock experiment for evaluating the ligation of 5-nitroindole-containing 3' repairer to the 3' end of mock miRNAs. The 3' Repairer (AAXXG) was ligated to the 3' end of the mock miRNAs that had 2 mismatches (Artificial Mir3; a) or 3 mismatches (Artificial Mir4; b) with the 43-nt Artificial Binding Site (RNA). On the other hand, 3' Repairer (AXXXG) was ligated to the 3' end of the mock miRNAs that had 0, 1, 2 or 3 mIsmatches (Artificial Mir1 to Mir4; c to f, respectively) with the Artificial Binding Site (RNA). The tested 3' repairers (3' Repairer (AAXXG) and 3' Repairer (AXXXG)) were 5-nt long while the mock miRNAs (Artificial Mir1 to Mir4) were 22-nt long. Therefore, the ligation products were 27-nt long. The reaction buffer was T4 RNA Ligase Reaction Buffer (NEB) in (a) and (b), but was changed to NEBNext® Quick Ligation Reaction Buffer (NEB) in (c) to (f). In (a) to (f), lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20nt from top to bottom), lane 2 was the mock miRNA, lane 3 was the mock target RNA, and the successive lanes were the time points studied (0, 2, 4, 8, 16, 24 and 36 hours in (a) and (b); 0, 0.5, 1, 2 and 4 hours in (c) to (f)).

5.12 Nick repairing for joining the interacting miRNA-target RNA sequences and for extending the 3' end of the miRNA 3' repairer

As the same issue found in the ligation of miRNA repairers to the 5' and 3' ends of the miRNA also occurs in the joining of two special nicks (see Figure 4.14.1) by T4 RNA Ligase 2, two sets of mock experiments were conducted for evaluating these nick repairs. This issue is whether the 5-nitroindole is a suitable substrate for T4 RNA Ligase 2.

In the evaluation of the repair of the first kind of nick as illustrated in Figure 4.14.1a, the results indicated that the T4 RNA Ligase 2 was able to join very efficiently both the nick containing a 5-nitroindole base at the third position and the nick containing a 5-nitroindole base at the second position from the donor (5') end (Figure 5.12.1a and b). All the nick repairs were completed in 2 hours.

In the evaluation of the repair of the second kind of nick as illustrated in Figure 4.14.1b, the results indicated that the T4 RNA Ligase 2 was able to join very efficiently the nick containing two 5-nitroindole bases at the third and fourth positions from the acceptor (3') end (Figure 5.12.1c). The repair of this nick was completed in 2 hours. In addition, the ligase was also able to join the nick containing three 5-nitroindole bases at the second to fourth positions from the acceptor (3') end (one base closer to the acceptor end when compared with the former mock example) (Figure 5.12.1d). The repair of this nick was completed in about 16 hours.

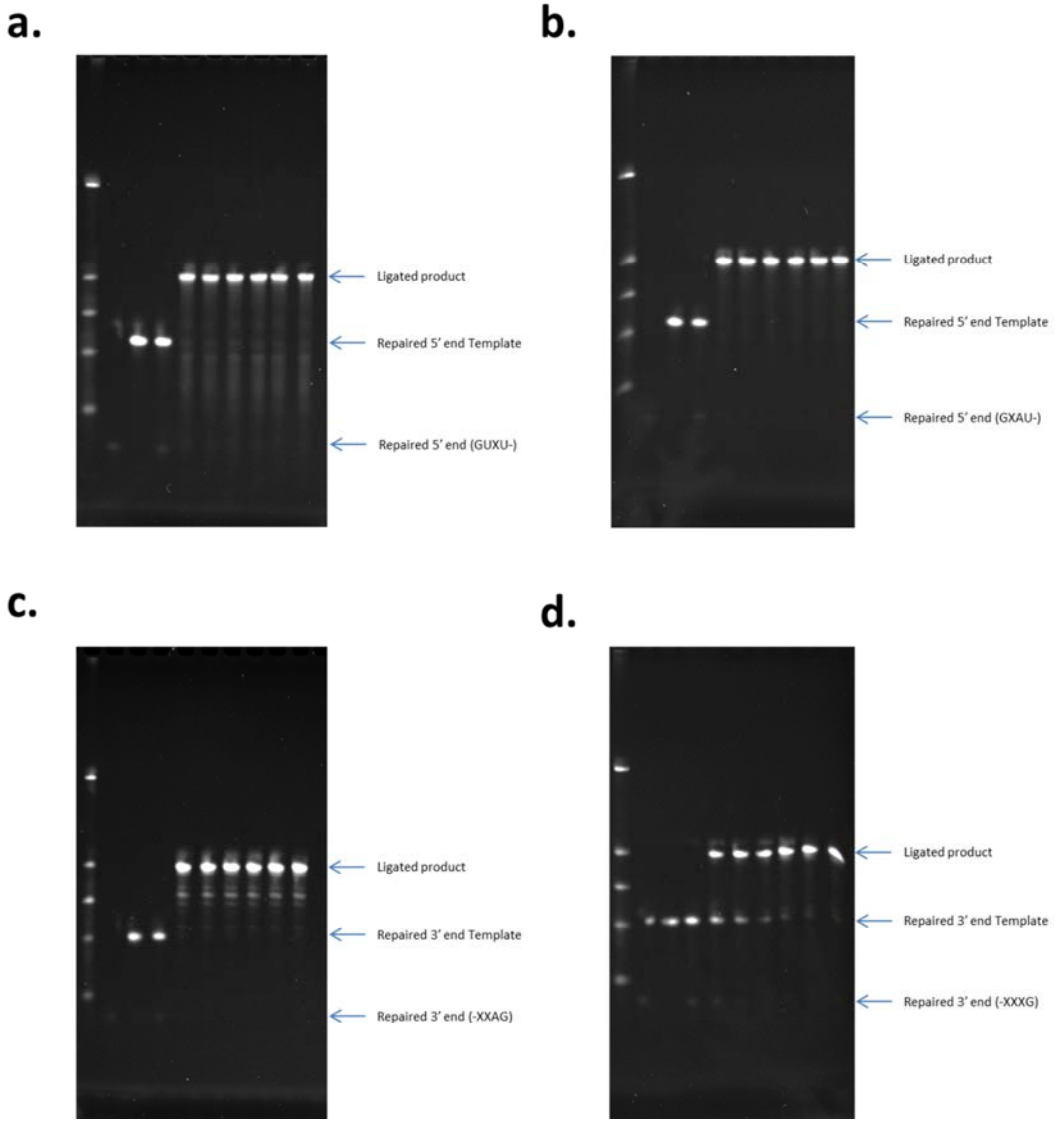


Figure 5.12.1 Mock experiment for evaluating the joining of two different kinds of nicks generated after the repair of the 5' and the 3' ends of mismatched miRNA. The joining of the nick occurred between (a & b) the 5' end of the miRNA 5' repairer and the 3' end of the "Stem-Loop Adaptor (RNA)" or (c & d) the 3' end of the miRNA 3' repairer and the short RNA oligo for extending the 3' end of the miRNA 3' repairer (see Figure 4.14.1). Short RNA oligos mimicking the repaired ends of miRNAs had a 5-nitroindole base (X) at the (a) 3rd or (b) 2nd position from the donor (5') end forming the nick, (c) XX at the 3rd and 4th positions from the acceptor (3') end forming the nick, or (d) XXX at the 2nd to 4th positions from the acceptor (3') end forming the nick (see Figure 4.14.2). Short RNA oligos were 18 nt (Repaired 5' end (GUXG-) and Repaired 5' end (GXAG-)) or 17 nt (Repaired 3' end (-XXAG) and Repaired 3' end (-XXXG)) long while the templates were 34 nt (Repaired 5' end Template) or 33 nt (Repaired 3' end Template) long. Therefore, the ligation products were 52 nt long in (a & b) and 50 nt long in (c & d). In (a) to (d), lane 1 was Small RNA Marker (Abnova: 100 nt, 50 nt, 40 nt, 30 nt and 20nt from top to bottom), lane 2 was the oligo for mimicking the repaired ends of miRNA, lane 3 was the template for constructing the nicks, and the successive lanes were the time points studied (0, 2, 4, 8, 16, 24 and 36 hours).

5.13 Gap repair by ligation

For generating the chimeric RNA molecule that represents the interacting miRNA and its target RNA, the gap between the 5' end of miRNA and the 3' end of target RNA (with the Stem-Loop Adaptor (RNA) already ligated) has to be repaired in the idea of RIINGS. A mock experiment was set up for investigating the ligation approach for repairing the gap. First, seven ligation conditions were evaluated. Second, a time-course study for identifying the optimal ligation time in the selected working condition was conducted.

5.13.1 Selection of working condition for the approach of "gap repair by ligation"

The conjugation of the two interacting mock RNAs only occurred in the enzymatic reactions that were commonly catalyzed by T4 RNA ligase 2 (dsRNA Ligase)(NEB) (see the yellow bands in lane 1, 2, 4 and 5 in the right panel in Figure 5.13.1). Except for the reaction shown in lane 4, the other three reactions gave the expected product (121-nt conjugated RNA) (Figure 5.13.2). In lane 1 and 2, the ligation conditions were quite similar. The filler in these two conditions was 5-P Tetramer N. The only difference between these two conditions was the ratio of the number of the filler required for filling up the gap (each 16-base gap required 4 RNA tetramer molecules to repair) to the filler available. The ratio was 1:12.5 in lane 1, and 1:50 in lane 2. This difference did not affect the yield of the conjugated RNA (Figure 5.13.1), but a larger amount of the filler (5-P Tetramer RNA) was consumed in single-strand ligation in the reaction in lane 2 (detected by counterstain with SYBR® Gold nucleic acid gel stain, see Figure 5.13.2). This result indicated that the increased input of the filler did not increase the yield of the conjugated RNA. On the other hand, the reaction condition of lane 5 was the same as lane 2 plus 1 mM NMPs. The addition of 1 mM NMPs served to test whether NMPs could act as supplementary filler in the gap filling reaction. However, the supplementary filler did not provide any contribution in the gap repair as the

yield of the expected gap-repaired product (121 nt) in lane 5 was closely similar to those in lane 1 and lane2. Interestingly, the exceptional lane, lane 4, also had a yellow band indicating the occurrence of conjugation of the two mock RNAs. However, the product length of the yellow band was distinct from those in lane 1, 2 and 5. With this observed product length, this product was probably the result *of direct ligation* between Artificial Mir 5 (FAM) (23 nt) and Artificial Binding Site (RNA, HEX) (82 nt) because the observed length of this product was close to the length of the directly ligated product (105 nt). This result might be due to a combination of reasons. First, "the presence of NMPs in the reaction"— NMPs may be able to destabilize the stability of the duplex through the competition in base pairing. Second, "the absence of 5-P Tetramer N"— when the 5-P Tetramer N was ligated to the Artificial Mir 5 (FAM), the length of the duplex prolonged and the stability of the duplex could be increased accordingly. Thus, with the loss of the stabilizing effect from the 5-P Tetramer N and the presence of destabilizing effect from NMPs, the duplex was dissociated and the two mock sequences were ligated together. Furthermore, the result in lane 4 also indicated that the NMPs could not be the substrate of T4 RNA ligase 2.

According to all the above observations, the condition of lane 1 was selected for performing the gap repair.

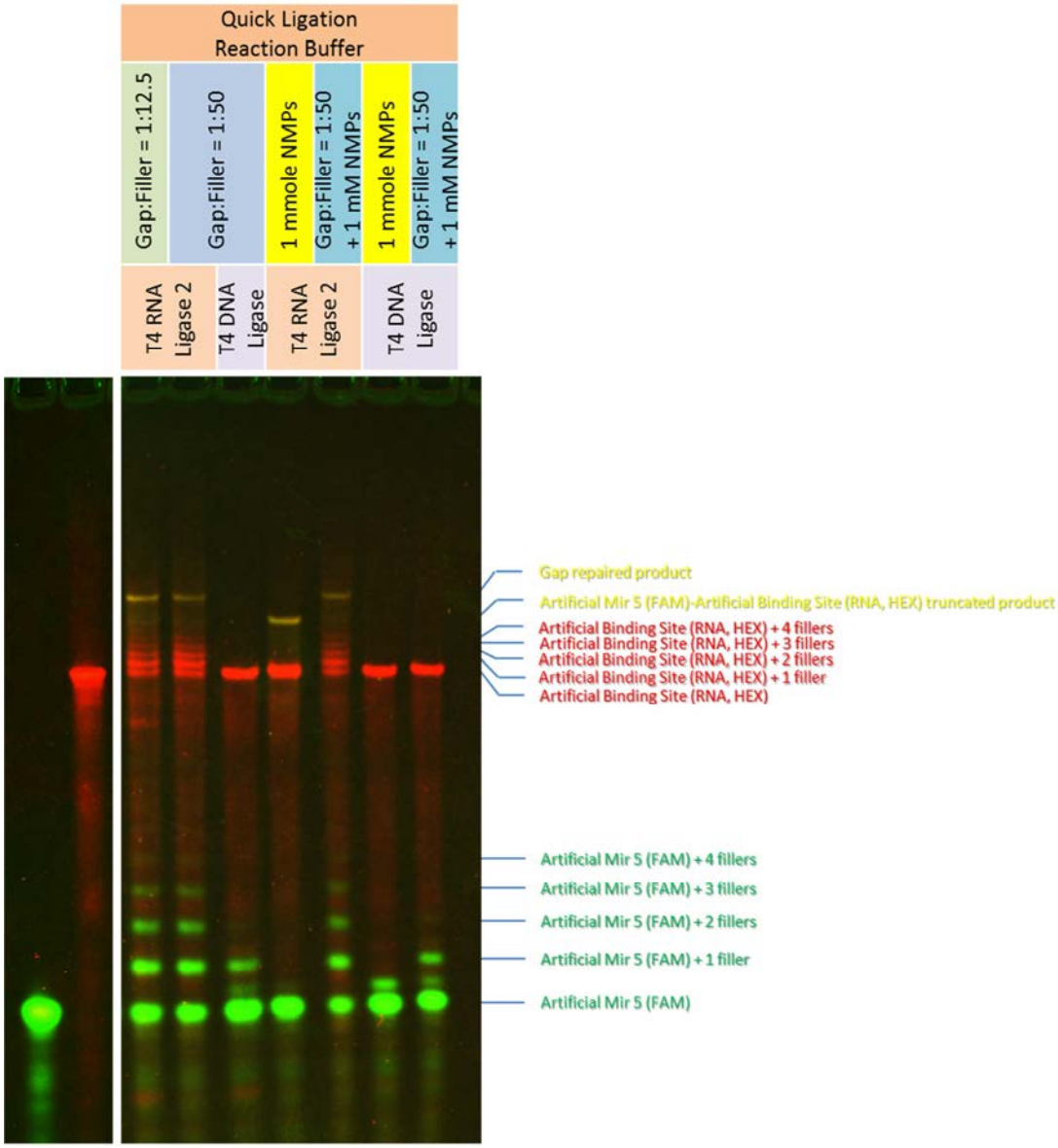


Figure 5.13.1 The result of the seven conditions for conjugating the two interacting RNAs by the approach of "gap repair by ligation". From left to right in the left panel, lane 1 is Artificial Mir 5 (FAM), and lane 2 Artificial Binding Site (RNA, HEX). From left to right in the right panel, lanes 1 to 7 are the samples prepared under condition 1 to 7 respectively. The details of these conditions are available in Section 4.15.1 and are shown on the top of the panel.

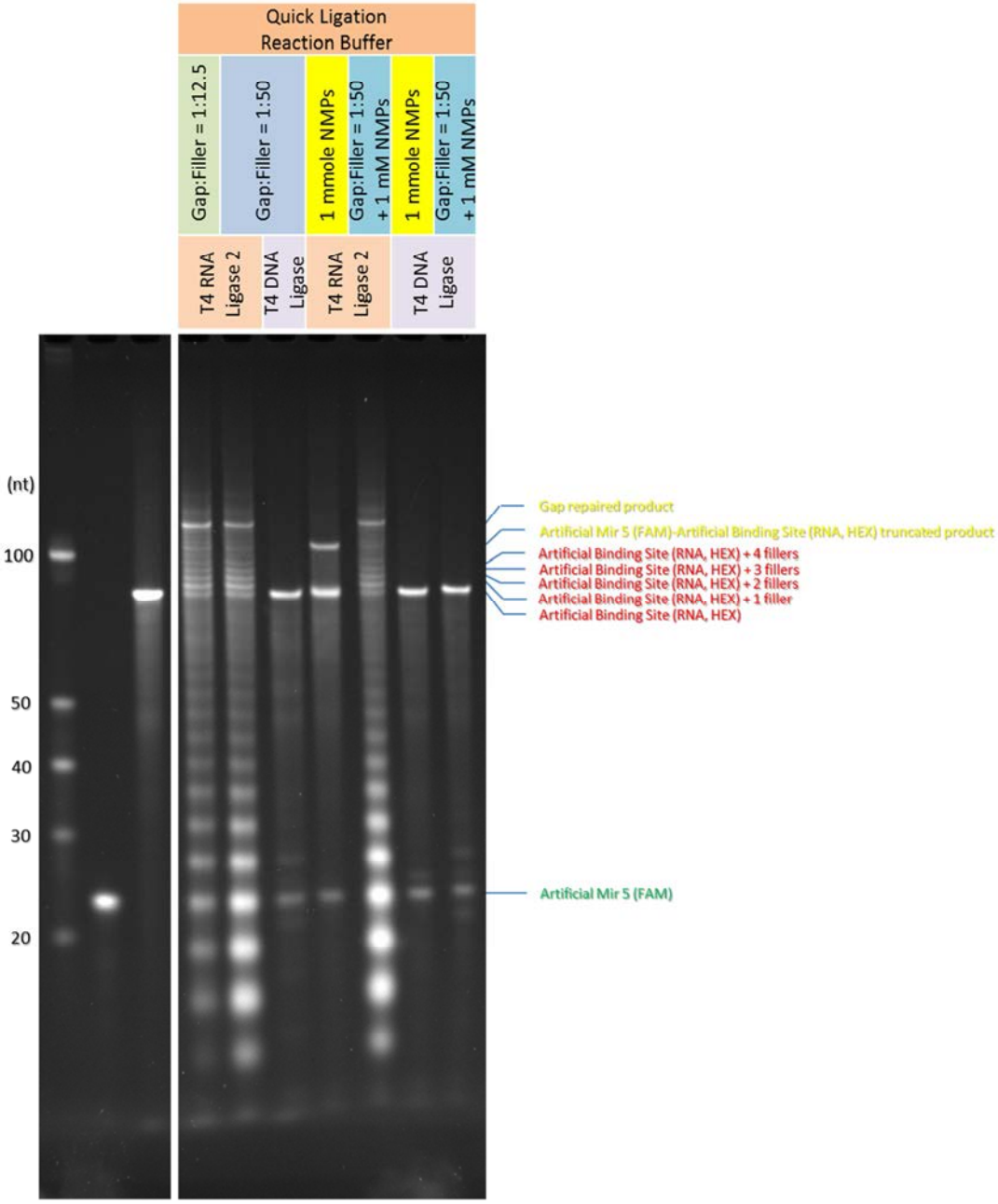


Figure 5.13.2 Counterstaining by SYBR® Gold nucleic acid gel stain of the gel shown in Figure 5.13.1. The additional leftmost lane in the left panel is Small RNA Marker (Abnova). The expected length of Artificial Mir 5 (FAM) and Artificial Binding Site (RNA, HEX) were 23 nt and 82 nt respectively. The expected length of the directly ligated product of Artificial Mir 5 (FAM) and Artificial Binding Site (RNA, HEX) was 105 nt. The expected length of the gap-repaired product was 121 nt. The bands had a ladder-like pattern in lanes 1, 2 and 5 in the right panel were the truncated products of 5-P Tetramer N. The expected length of 5-P Tetramer N and the truncated product of itself were multiple of 4 nt. They were 4 nt, 8 nt, 12 nt, 16 nt, 20 nt, 24 nt 28 nt and so on.

5.13.2 Optimal treatment time for conjugating the two interacting RNAs

After the working condition for the "gap repair by ligation" was defined, a time-course study for identifying the optimal treatment time was then performed. Incubation at both 4℃ and 16℃ was evaluated. The results in Figure 5.13.3 indicated that ligation occurred in both ends during incubation, regardless of the incubation temperature (The gap filling starts at both 3' and 5' ends. In other words, the following products can be identified in the gel: Artificial Mir 5 (FAM) plus 1 filer, 2 filers, 3 filers and 4 filers, and Artificial Binding Site (RNA, HEX) plus 1 filer, 2 filers, 3 filers and 4 filers). Nevertheless, a product band with both labels only appeared in the set of experiments that was incubated at 16℃. This indicated that ligation at 16℃ was more efficient than that at 4℃.

On the other hand, Figure 5.13.4 shows the increasing amount of conjugated product with expected size (121 nt and represented by the fluorescence intensity) with increasing incubation time at 16℃. Levene's test was used to test the homogeneity of variances across different groups of samples prepared by different crosslinking conditions, and did not show significant difference ($p = 0.117$). This indicated that equal variance across groups could be assumed. One-way ANOVA comparing these five groups of samples showed significant differences among these groups ($p = 0.00000027$). Post-hoc test (Tukey HSD) indicated that increment from time 0 to 3-hour-incubation was not statistically significant (5.47-fold; $p = 0.462$); increment from 3-hour-incubation to 6-hour-incubation was statistically significant (2.91-fold; $p = 0.016$); increment from 6-hour-incubation to 12-hour-incubation was statistically significant (1.81-fold; $p = 0.0038$); increment from 12-hour-incubation to 18-hour-incubation was statistically significant (1.29-fold; $p = 0.048$). However, the increasing trend slowed down during the time course and started to plateau

between 12-hour-incubation and 18-hour-incubation. Therefore, the optimal incubation time for the gap repair was 12 hours at 16°C.

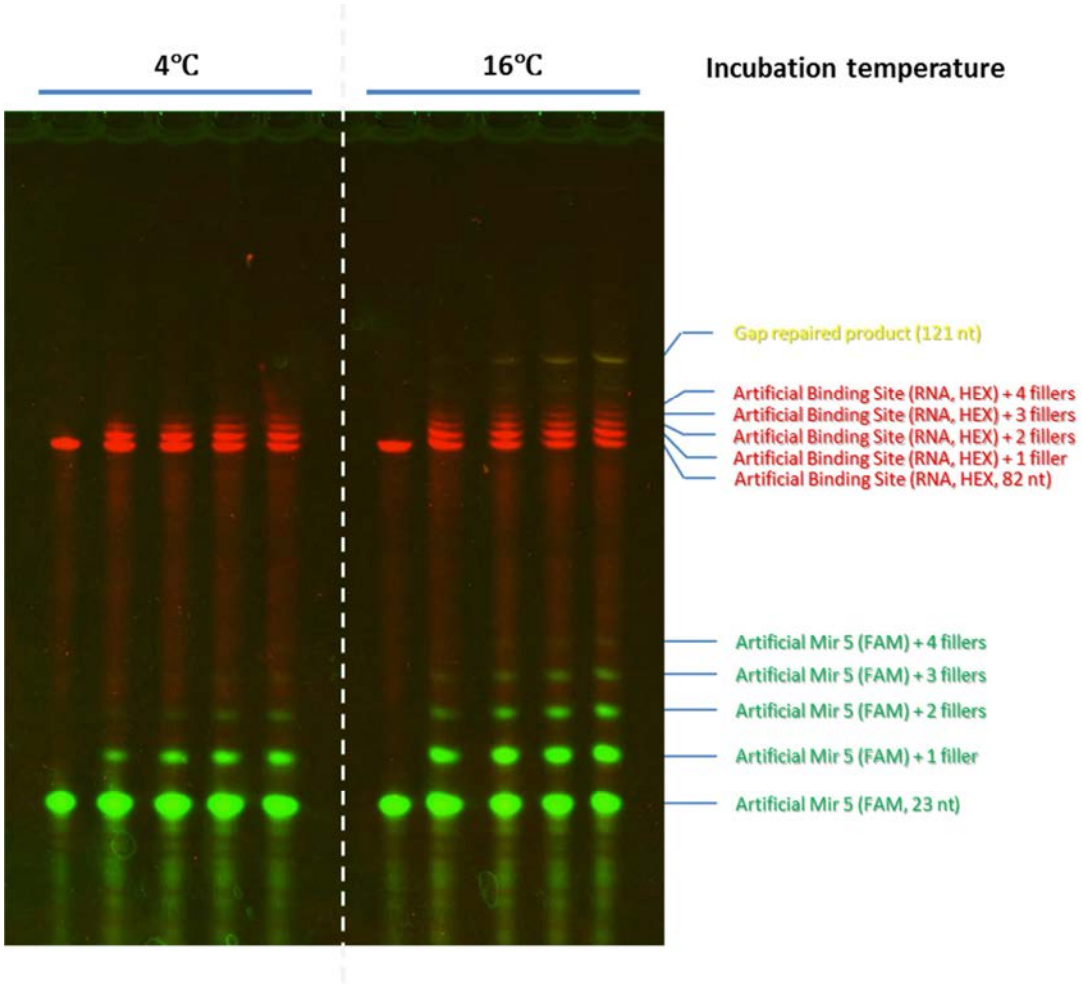


Figure 5.13.3 Time-course analysis for studying the conjugation of the two interacting RNAs by the approach of "gap repair by ligation" at 4℃ and 16℃. From left to right in the left panel, the lanes are the samples incubated for 0, 3, 6, 12 and 18 hours respectively. The incubation temperature was 4℃. From left to right in the right panel, the lanes are the samples incubated for 0, 3, 6, 12 and 18 hours respectively. The incubation temperature was 16℃. Artificial Mir 5 (FAM) was represented in green color; Artificial Binding Site (RNA, HEX) was represented in red color; conjugated product with both FAM and HEX labels was represented in yellow color.

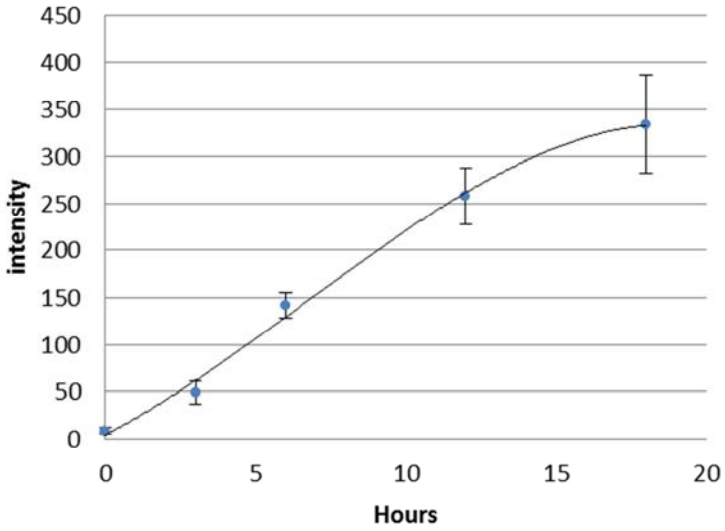


Figure 5.13.4 The fluorescence intensity of the conjugated products (yellow bands seen in Figure 5.13.3) during the time course studied for "gap repair by ligation". The corresponding plot of intensity against time illustrates the amount of the conjugated products during the time course studied. The signal intensities (in arbitrary unit) were based on the mean value ± SD (n = 3).

5.14 Reverse transcription of the chimeric sequences – nitroindole base-containing template

Two experiments were performed to study whether there were reverse transcriptases (specifically the RNA-dependent DNA polymerase activity) able to convert RNA templates with 5-nitroindole base substitution(s) to cDNA. First, ten candidate polymerases were tested for their ability to convert RNA templates with one, two successive and three successive 5-nitroindole bases to cDNA. Second, the selected polymerase was then further investigated in a mock experiment mimicking the real situation.

5.14.1 Selection of polymerase for converting RNA templates with 5-nitroindole base substitution(s) to cDNA

The experiment (Experiment 1) only identified that *Bst* 3.0 DNA Polymerase (NEB, M0374S) had some ability to convert 5-nitroindole-containing RNA templates to cDNA. Other tested polymerases did not show any similar conversion when dealing with 5-nitroindole base. The conversions initiated by these nine polymerases were completely blocked by the first 5-nitroindole base-containing site at position 26 (see Figure 4.16.1 for the sequence structure of the template).

The extension mediated by *Bst* 3.0 DNA Polymerase was partially blocked by the first 5-nitroindole base-containing site at position 26 (from template 3' end to 5' end) and was blocked by the second 5-nitroindole base-containing site at positions 35-36 (from template 3' to 5') (Figure 5.14.1). The reactions were conducted under the following conditions: 2.4 µM of 5-NitInd RT temp 1, 2.4 µM of 5-NitInd RT primer 1, 5 mM of $MnCl_2$, 2 mM of dNTPs, 1X NEBNext® Quick Ligation Reaction Buffer and 0.64U/µl of *Bst* 3.0 DNA Polymerase. The final volume of each reaction was 12.5 µl. The incubation conditions were 25℃ for 10

minutes followed by 55℃ for 1, 2, 4, 8, 16, 24 and 36 hours. All the reactions were stopped by inactivation at 85℃ for 5 minutes. During the time course under investigation, the extension gradually escaped from the first 5-nitroindole base-containing site to the second 5-nitroindole base-containing site. Figure 5.14.1 shows the bands, which were the extension product blocked by the first 5-nitroindole base-containing site (position 26, from template 3’ end to 5’ end), gradually decreased in their intensity during the time. Meanwhile, the bands, which were relavent to the extension of the sequence blocked by the second 5-nitroindole base-containing site (positions 35-36, from template 3’ end to 5’ end), gradually increased in their intensity in the course of time. Notably, there was no obvious band that could represent the product of the sequence blocked by the third 5-nitroindole base-containing site (positions 46-48, from template 3’ end to 5’ end). This implied that the extension could be fully blocked by the second 5-nitroindole base-containing site, which contained two successive 5-nitroindole bases.

Therefore, this experiment found that *Bst* 3.0 DNA Polymerase was the only polymerase for converting 5-nitroindole base-containing RNA template to cDNA. This polymerase could deal with RNA templates with single 5-nitroindole base substitution but not successive substitutions. This polymerase was selected for converting the chimeric RNA library to cDNA library in RIINGS.

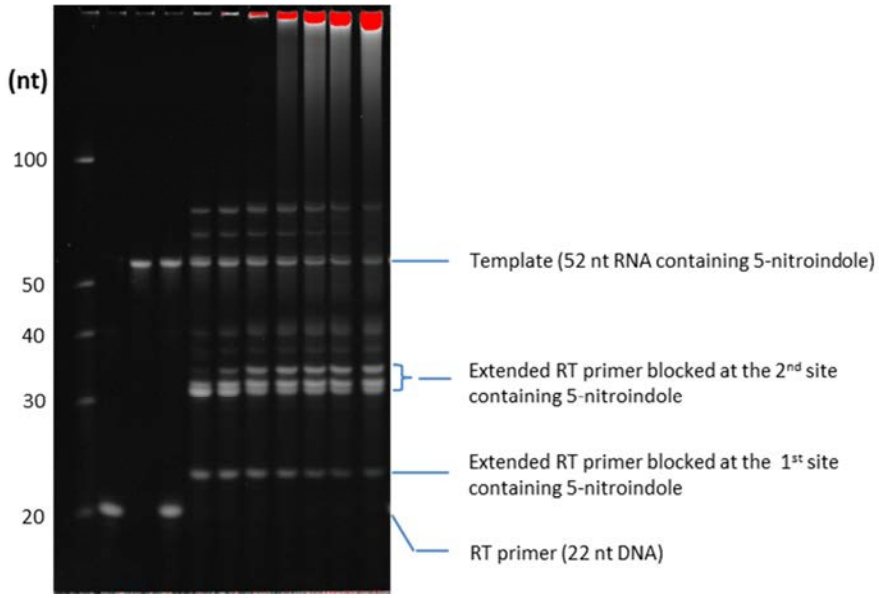


Figure 5.14.1 Reverse transcription of 5-nitroindole base-containing RNA templates by Bst 3.0 DNA Polymerase. 5-NitInd RT primer 1 (22 nt DNA) was used as the primer to convert the 5-NitInd RT temp 1 (52 nt) to cDNA. Three 5-nitroindole substitution sites were found in the RNA template at position 26, positions 35-36 and positions 46-48 (from template 3' end to 5' end). From left to right, lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20 nt from top to bottom), lane 2 was 5-NitInd RT primer 1, lane 3 was 5-NitInd RT temp 1, and the successive lanes were the time points studied (0, 1, 2, 4, 8, 16 24 and 36 hours).

5.14.2 The reaction completeness of Bst 3.0 DNA Polymerase in converting a mock template mimicking the real situation

The experiment (Experiment 2) revealed that *Bst* 3.0 DNA Polymerase (NEB, M0374S) was able to deal with the RNA template with two separated 5-nitroindole base substitutions. 5-NitInd RT temp 2 was the RNA template used in this experiment — a template mimicking the chimeras that had two separated 5-nitroindole base substitutions originating from the 5' and 3' repairers (each repairer had one substitution). The result showed that the extension of the 5-NitInd RT primer 2 was blocked by the first 5-nitroindole substitution at position 22 (from template 3' end to 5' end) only very slightly and was slightly blocked by the second 5-nitroindole substitution at position 44 (from template 3' end to 5' end) (Figure 5.14.2). The majority of 5-NitInd RT primer 2 was extended to the end of the template (Figure 5.14.2). This result actually supported the usefulness of *Bst* 3.0 DNA Polymerase in RIINGS for converting the chimeric RNA library to cDNA library.

Notably, the working condition of these experiments had some minor changes that could help minimize non-specific elongation. These included the increase of $MnCl_2$ concentration from 5 mM in Experiment 1 to 10 mM in Experiment 2 and the incubation temperature reduced from 55 ℃ in Experiment 1 to 50℃ in Experiment 2.

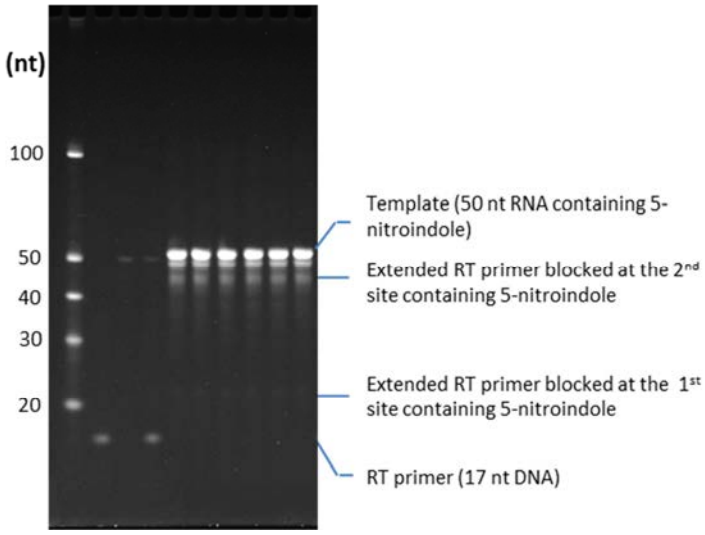


Figure 5.14.2 Reverse transcription of a mock RNA template mimicking the real RNA chimeras prepared by RIINGS. 5-NitInd RT primer 2 (17 nt DNA) was used as the primer to convert the 5-NitInd RT temp 2 (50 nt) to cDNA. Two single 5-nitroindole substitutions in the template RNA were at positions 22 and 44 (from template 3' end to 5' end). From left to right, lane 1 was Small RNA Marker (Abnova; 100 nt, 50 nt, 40 nt, 30 nt and 20 nt from top to bottom), lane 2 was 5-NitInd RT primer 2, lane 3 was 5-NitInd RT temp 2, and the successive lanes were the time points studied (0, 1, 2, 4, 8, 16 and 24 hours).

6. Discussion

6.1 Antibody for immunoprecipitation (IP)

There are many anti-AGO antibodies that are commercially available. However, the selection of an appropriate antibody is very critical. This is not only for RIINGS, but also for other IP-related experiments because the selection of antibody directly determines the success of IP. Faulty or inappropriate selection can cause a waste of time and money. The selection of an antibody for IP starts with noting and understanding the properties of the antibodies from the information provided by the manufacturers. This does not mean that we need to entirely rely on the recommendations (e.g. suitability for use with which type of experiments) and neglect other details. Several important issues have to be considered.

First, what is the immunogen used for raising the antibody? In general, many antibodies are raised by using epitopes (small fragments of a protein) as the immunogens while others are raised by intact proteins. An antibody raised by an epitope only recognizes the corresponding epitope. If the epitope is not fully exposed on the protein surface in the native form of the protein, the antibody should not be used in co-IP (a technique for identifying the interacting partner of the target protein). This kind of antibodies can be used to capture the target protein when the target protein is denatured.

Second, is the antibody a monoclonal antibody or a polyclonal antibody? A monoclonal antibody is produced from a single clone of B-cells while a polyclonal antibody represents a collection of antibodies produced from many different B-cell clones. The monoclonal antibody only recognizes a particular site or epitope on the target protein (one antibody targets on one site) while the polyclonal antibody may target multiple sites. The polyclonal

antibody has the advantage that the antibody has a bigger chance of capturing the target. Like a shot from a shotgun, some bullets may hit the target while others may not. For the polyclonal antibody, some of the antibodies from some clones may be able to capture the target protein efficiently while others may not be. In brief, not all antibodies in the collection are able to capture the target. On the other hand, for a monoclonal antibody able to capture the target, the capturing efficiency is expected to be much better because all the applied antibody molecules are able to capture the target.

Third, what is the origin of the antibody (raised from which animal)? If western blotting, ELISA or other related experiment is required in the future, the origin of the antibody may be critical. It is because the antibody used in the IP might be eluted together with the target protein. When this eluate is used for running a western blotting as an example, the IP antibody would also appear on the membrane. This antibody can generate two very strong bands (one is the heavy chain and the other is the light chain) on the membrane if a second antibody that can recognize this IP-antibody is used. These two additional bands may affect the interpretation of the results.

The antibody selected for RIINGS was eIF2C2 (4F9) (Santa Cruz) (see Section 6.2). This antibody (4F9) is most likely able to recognize human AGO1-4 through a common epitope on AGO1-4 protein [308]. Sequential overlapping peptides have been used for mapping the epitope on AGO2 that is exactly recognized by eIF2C2 (4F9) by Ikeda et al [308]. However, all the tested peptides gave negative results. This suggests that eIF2C2 (4F9) may target a conformational epitope. If it is the case, it can explain why the antibody only recognize the natural form of AGO proteins, but not the denatured form – a recognition bias confirmed by Ikeda et al. and the dot blotting result in current study (Figure 5.2.1). Therefore, eIF2C2 (4F9) is particularly suitable for capturing biologically functioning miRNAs because

denatured AGO complex, which may also contain damaged structure of the interacting RNA duplex, are excluded by the recognition bias.

6.2 Immunoprecipitation

To start an immunoprecipitation (IP) experiment, protein lysate must be well prepared in an appropriate buffer. Lysate preparation is the first critical step in IP. Two major kinds of lysis buffers can be used for extracting proteins from cells.

The first kind of lysis buffer is the non-denaturing lysis buffer. This kind of buffer does not contain any denaturing agents and is particularly suitable for sample preparation in co-IP. There are two subtypes of non-denaturing lysis buffers. One subtype does not contain any detergent while the other subtype contains detergent. The key function of the detergent is to make the insoluble proteins become soluble in aqueous-based buffer. Usually non-ionic detergents, such as NP-40 and Triton X, are much preferred in detergent-containing non-denaturing buffer because they are not that active in physically interacting with charged groups on proteins, which are usually important in maintaining the 3-dimensional structure of proteins. On the other hand, ionic detergents, such as sodium dodecyl sulfate and lithium dodecyl sulfate, can also be used as assisting agents to make the insoluble proteins become soluble. However, ionic detergents are much stronger than non-ionic detergents and may denature proteins. For extracting soluble proteins, detergent-free non-denaturing buffer is good enough for the extraction and lysate preparation. Notably, a mechanical lysing method for disrupting cells and releasing the soluble proteins is required. In addition, heat generation during the disruption step must be prevented to avoid denaturing proteins.

The second kind of lysis buffer is the denaturing lysis buffer. This kind of buffer contains denaturing agents and is not preferred in co-IP experiments and in recovering the isolated protein for studying the protein properties that require the native structure of the protein, e.g. protein activity and protein conformation. However, for epitopes hiding inside the native form of the protein, denaturing lysis buffers are capable of exposing this kind of epitopes to the surface for the corresponding antibodies to access and recognize. Moreover, this kind of buffers can also be used for insoluble proteins that cannot be extracted from cells with non-ionic detergents. Heating and reducing agents can facilitate the extraction of these proteins by denaturing them and breaking the disulfide bonds inside the protein respectively.

There are two approaches to performing the IP experiment. One of these approaches is to immobilize the antibodies on the beads prior to the capture of the target protein from the total lysate. The other approach is to incubate the total lysate with the antibodies first, and then capture the antibody-target protein complexes by adding the beads to the lysate-antibody mixture. The first strategy can usually reduce non-specific binding. However, this strategy also reduces the yield of the target protein. The second strategy can usually elevate the yield of the target protein; however, the purity may be relatively poorer than that of the first strategy (see Figure 4.5.1). These differences can perhaps be explained by the freedom of movement of the molecules. In the second strategy, the freedom of movement of the antibody and the target protein (including the proteins in non-specific bindings) is relatively higher than that in the first strategy, and hence results in a higher opportunity for the antibody meeting the target protein at any moment. In contrast, the freedom of movement of the antibody is relatively lower in the first strategy than in the second strategy. This would result in lowering the opportunity of capturing the target protein and eventually reducing the yield of the target protein.

Furthermore, after the conjugation of antibody onto the proper beads in the first strategy, the antibody and the beads can be further covalently crosslinked by some chemicals. This crosslinking step is beneficial in preventing the IP-antibody from getting into the eluate. In reality, the immobilization of the antibody on the beads is usually achieved via the binding between antibody and protein A or G, which is pre-crosslinked on beads. This is not a covalent bonding and can be destroyed during elution. The chemicals that are used for further covalently crosslinking the antibody and protein A or G, such as disuccinimidyl suberate, have two functional groups each on one end of a long spacer. Each of the functional groups can react with an amine group on protein and form covalent linkage. Therefore, when the two functional groups on the same molecule form covalent linkages with the antibody and the protein A or G simultaneously, the antibody would covalently immobilize on the relevant beads.

The elution of the isolated target protein is usually achieved by adding elution buffer to the target protein-carrier beads. Based on different elution strategies, the elution buffers can be classified into two major types – denaturing buffer and non-denaturing buffer. Denaturing buffers, such as Laemmli sample buffer (Appendix A) often used as sample loading buffer in SDS-PAGE, usually contain denaturing agent in the buffer for disrupting the binding between the isolated target protein and the antibody and hence enable the isolated target protein to be released from the beads. In fact, this is exactly the function of the strong ionic-detergent SDS in Laemmli buffer. Non-denaturing elution buffers elute the isolated target protein by means of transiently denaturing the proteins (including target protein and the antibody) or disrupting the charge interactions between the isolated target protein and the antibody. Both strategies can eventually disrupt the binding between the isolated target protein and the antibody. Low-pH buffers (e.g. 0.1 M Tris-glycine buffer, pH 2.7) and high-salt buffers are the two kinds of buffers used in these denaturing and non-

denaturing strategies respectively. Neutralization of pH and dialysis for removing the high concentration of salt are necessary after the above elution strategies respectively. Denaturing elution has the advantage of higher efficiency in harvesting the isolated target protein, but the disadvantage of disrupting the native conformation and activity(ies) of the target protein. In contrast, non-denaturing elution has just the opposite effects. The selection of elution buffer is highly constrained by the requirements of subsequent steps.

In RIINGS, the antibody eIF2C2 (4F9) (Santa Cruz) was selected for isolating the AGO complex. This is an IgA antibody that has no free fragment-crystallizable region (Fc region) on the antibody for the binding of protein A or G. Instead, protein L magnetic beads were used to immobilize the eIF2C2 (4F9) antibody for the IP experiment, in which the "L" in the name means light chain. Protein L is capable of binding the light chain of antibody. Because the assessed purity of the isolated AGO complex was good (see Sections 5.3.2 and 5.9.1), the IP experiment was performed using the approach of allowing the antibody to capture the target protein in the crude lysate first, and then using protein L magnetic beads to immobilize the target protein-antibody interacting complex for maximizing the yield of AGO complex. The elution was then carried out by adding proteinase K to the beads with AGO complex because the RIINGS target of interest is miRNA-target RNA duplex, which is not susceptible to digestion by proteinase K. All the proteins were digested accordingly under a low-temperature (4℃) condition. This elution method has the advantages of being efficient and mild. The AGO protein on each RNA duplex and together with other co-purified proteins can be removed by proteinase K digestion. As the result, the miRNA-target RNA duplex of interest can be efficiently released under an extremely mild condition.

6.3 Generation of the chimeric sequence by the use of stem-loop adaptor

The use of stem-loop adaptor in assisting the physical linkage between the two strands of the interacting miRNA-target RNA duplex is one of the key, strategic designs in RIINGS. The ligation of the “Stem-Loop Adaptor (RNA)” to the 3' end of the trimmed miRNA target (i.e. target RNA that interacts with miRNA) has three effects. It (i) extends the 3' end of the target RNA target site, (ii) flips it over and backward, and (iii) positions the 3' end of the “miRNA target” and the 5' end of the miRNA in a head-to-tail manner with a gap in between (see Figure 3.1). By repairing the gap in between, chimeric sequence that represents the interacting miRNA-target RNA duplex is generated. Importantly, the Argonaute protein on the RNA duplex has to be removed before the gap is repaired because the Argonaute protein on the interacting RNA duplex in effect prevents the miRNA from being engaged in the ligation process, which is unfavorable to the gap repairing process. Similarly, this problem also exists in the CLASH [299] and MARIO [302] methods. However, the inventors of CLASH and MARIO have not made any effort to resolve this issue. Unlike the strategy used in CLASH and MARIO for generating the chimeric sequence (for the details of the strategy, please refer to Sections 1.5.5 and 1.5.6), the current strategy efficiently and physically joins the interacting miRNA-target RNA duplex even if the miRNA is shielded by the Argonaute protein. In the generation of the chimeric sequence, the current strategy can avoid the bias of only ligating the interacting miRNA-target RNA duplexes that do not have spatial hindrance. In fact, this bias is an inherent drawback in CLASH and MARIO (see Sections 1.5.5 and 1.5.6).

However, the removal of the Argonaute protein may not be beneficial because the Argonaute protein on the interacting miRNA-target RNA duplex may also function as a

stabilizer for preventing the interacting duplex from melting. The removal of the Argonaute protein may destabilize the duplex and dissociate the duplex. Therefore, it is critical for the development of RIINGS that the thermostability of the interacting miRNA-target RNA duplex is assessed after the removal of the Argonaute protein. Melting curve analysis revealed that the melting temperature of the interacting miRNA-target RNA duplex was about 40℃ (see results in Section 5.9.2). This indicates that the duplex is stable at low temperature like 4-16℃, and strongly supports the current strategy for generating the chimeric sequence.

Critically, the analysis of re-melting to the duplex samples, in which the RNA duplex was completely melted before a repeat melting analysis, revealed that the miRNA-target RNA duplex could not reform once the duplex was melted. Normally, repeated melting is possible and can always be performed in experiments involving melting curve analysis such as high-resolution melting for genotyping and dissociation curve analysis in SYBR-Green I-based qPCR assay because the two interacting nucleic acid sequences can re-anneal after each melting. However, the two interacting RNAs (miRNA and target RNA) in each duplex of the Argonuate complex are unique in the current situation, i.e. the diversity of the sequence of these duplexes is high. In fact, there is a pool of duplexes with different sequences in the sample extracted *in vivo*, in which the duplexes are constructed by different miRNAs and their corresponding target RNAs via imperfect base pairing. Therefore, the interacting RNAs are difficult to pair up again with their original partners once the duplexes are melted because the chance of identifying the original partner or a comparable partner is extremely difficult under “a very dilute condition” because the abundance is very low indeed for each particular miRNA sequence and their corresponding target RNA in the pool. Here, the idea of “a very dilute condition” does not mean that the

concentration of the RNA is really low, but means that an RNA with a particular sequence is very few in the RNA pool.

After the thermostability of the interacting miRNA-target RNA duplex was confirmed, another critical issue demanded immediate attention for the development of RIINGS. This issue was whether the proteinase K was able to remove the Argonaute protein from the Argonaute complex at low temperature (e.g. 4-16℃). The results in Section 5.10.1 revealed that proteinase K was very efficient in digesting BSA at 4-16℃ (no difference among different digestion temperatures 4℃, 8℃, 12℃ and 16℃), indicating that the activity of proteinase K was high even at 4-16℃. Measuring the concentrations of amino acids released during digestion of isolated Argonaute complex by proteinase K showed that the efficiency of proteinase K digestion depended on the concentration of proteinase K rather than the digestion temperature (see Section 5.10.2).

In the time-course study conducted to determine the optimal ligation time for ligating the stem-loop adaptor to the 3' end of the miRNA target, a radioactive smear of about 220 kDa was detected in lane 2 to lane 4 in Figure 5.4.1a (Section 5.4.1). This smear was most likely the AGO complex, which consisted of an AGO protein, a miRNA and a target RNA. The shift in molecular weight was attributed to the intercalated RNAs and the stem-loop adaptor. The molecular weight of the miRNA (~22 nt) was about 7 kDa while the stem-loop adaptor (46 nt) was about 17.5 kDa. The molecular weight of the target RNA would depend on the length (and hence size) of the target RNA remaining after RNA digestion. Normally, the target RNA showed a *distribution of size* with an average of about 400 nt after the RNA was digested by 1:1000 diluted RNase One™ Ribonuclease (Promega) (see Section 5.6.2 and Figure 5.6.2d). Thus, the molecular weight of the target RNA was about 140 kDa. Furthermore, this also explained why the radioactive band was a *smear*. Overall, the total

increased molecular weight should be about 164.5 kDa in this case. However, the actual band shift was less than this calculation because the RNA was negatively charged under the gel electrophoresis condition and the conformation of the flanking/protruding RNA may also shrink into a relatively small size with less spatial hindrance.

6.4 Rescue of the released interacting miRNA-target RNA duplex from proteinase-K-containing buffer by the "Re-attaching Adaptor"

Since the interacting miRNA-target RNA duplex is released from the magnetic beads by proteinase K, the obtained solution with the released RNA duplex also contains proteinase K. Proteinase K strongly inhibits all enzymatic reactions by its robust proteolytic activity. Therefore, proteinase K must be removed before proceeding to the steps of the second mismatch repair and gap filling in RIINGS. Moreover, the interacting miRNA-target RNA duplex has not yet been physically linked together at this point. Accordingly, designing a novel method for recovering the interacting miRNA-target RNA duplex from the proteinase-K-containing buffer is a critical problem in RIINGS. First, this method should not require a harsh working condition. The working condition must be mild enough to keep the interacting miRNA-target RNA in duplexed form. Therefore, inactivation of proteinase K by heat denaturation (95℃ for 10 minutes) is not applicable in the current situation. Second, isolation via capturing biotinylated oligos by streptavidin-conjugated beads, which is widely used for recovering target molecules, is not applicable here either because streptavidin (a protein) is easily degraded by proteinase K already present in the buffer. Therefore, ligating a biotinylated adaptor to the interacting miRNA-target RNA duplex for recovering the

duplex would not work in RIINGS. Third, the method should be able to recover the RNA duplex rapidly and efficiently. The longer is the time of keeping the duplex in the intermediate stage, the higher is the risk of dissociating the duplex. In addition, RNA is not a very stable molecule for long-term storage. Therefore, shortening the recovery time is beneficial to the whole process of RIINGS.

A very good approach that met all the above demanding requirements for recovering the released miRNA-target RNA duplex was developed in the current study. By taking the advantages of being rapid, specific and non-susceptible to proteinase K digestion, a potential design based on base pairing of nucleic acids has been chosen for developing the recovery method. First, an adaptor called “Re-attaching Adaptor” is ligated to the 5’ end of the target RNA on the isolated AGO complex (Figure 3.1). Actually, it is a DNA adaptor carrying (from 5’ end) a specific sequence, a restriction enzyme site and a few bases of RNA at its 3’ end. The specific sequence can pair up with an antisense DNA oligo conjugated on “Library Enrichment beads”. As the result, the “Library Enrichment beads” recover the miRNA-target RNA duplex rapidly and efficiently through annealing. Cleavage of the restriction enzyme site by the corresponding restriction enzyme finally releases the chimeric sequence product from the “Library Enrichment beads” (Figure 3.1). The released chimeric sequence is in a stem-loop structure with a sticky end at the end of the stem. This sticky end facilitates the ligation of the sequencing adaptor to the chimeric sequence. On the other hand, the few bases of *RNA* at the 3’ end of the adaptor enables the ligation of this adaptor to the 5’ end of the target *RNA* by the T4 RNA Ligase 1. T4 RNA Ligase 1 is an enzyme that catalyzes the joining of two single-stranded *RNA* sequences through the formation of a phosphodiester bond in between. The required phosphate group is provided by the 5’ phosphate end (donor end) while the 3’ hydroxyl end (acceptor end) is the end to receive the electrons from the phosphate group in the formation of a phosphodiester bond.

This enzyme does not catalyze the ligation of a 5’-phosphorylated RNA to DNA with 3’ hydroxyl end, but to RNA with 3’ hydroxyl end. Thus, the addition a few bases of RNA at the 3’ end of the adaptor actually enables the ligation of the “Re-attaching Adaptor” to the 5’ end of the target RNA.

6.5 Trimming of flanking RNA

Ribonuclease (abbreviated RNase) catalyzes the degradation of RNA. There are two classes of ribonucleases based on the positon of degradation: exoribonuclease and endoribonuclease. Based on the mechanisms of degradation, RNases can be divided into phosphorolytic RNase (EC 2.7) and hydrolytic RNase (EC 3.1). Phosphorolytic RNases break down RNA by cleaving the phosphodiester bond between two ribonucleosides via the attack of an inorganic phosphate and this reaction releases diphosphate nucleotides. The hydrolytic RNases cleave RNA by hydrolyzing the phosphodiester bond between two ribonucleosides. On the other hand, exoribonucleases degrade RNA by cleaving the RNA molecule from the 5’ or 3’ end of the RNA molecule while endoribonucleases degrade RNA by cleaving internal phosphodiester bonds of RNA molecule [324].

Both RNase A and RNase One Ribonuclease belong to EC 3.1.27.5 and are endoribonucleases [324]. The enzymes of this family are able to cleave RNA sequences endogenously. In fact, They catalyze the cleavage of RNA dinucleotide bonds endogenously, and leave a 5′ hydroxyl end and 2′,3′-cyclic monophosphate ends on the product RNAs or on the product monoribonucleotides. The 2′,3′-cyclic monophosphate end is an intermediate product. This end is further converted by the RNases to the corresponding 3' monophosphate via hydrolysis; however, the hydrolysis is very slow (Figure 4.10.1) [309,

310]. As the result, the cleaved RNAs mainly have 5´ hydroxyl end and 2´,3´-cyclic monophosphate end at their two termini. However, there is an intrinsic difference between these two RNases. RNase A preferentially cleaves the phosphodiester bond of RNA at the 3' end of pyrimidine residues (C and U) [310]. However, RNase I randomly cleaves RNA sequences endogenously without any sequence preference [311].

To facilitate the generation of the chimeric sequence through the approach used in RIINGS (see Section 6.3 or Figure 3.1), it is necessary to trim the two long flanking sequences, which are the 5' and the 3' regions of the target RNA strand, and protrude from the Argonaute complex. This task can be achieved by RNase. However, the digestion process should be under control because both retained ends must be kept long enough for the ligation of adaptors (see Figure 3.1). In fact, the completeness of RNA digestion by RNase depends on the enzyme concentration, the incubation time and the incubation temperature.

In the current situation, the incubation temperature should not be too high in order to avoid dissociation of the miRNA-target RNA complexes. Therefore, the incubation temperature to be tested was fixed at 4°C.

Two issues were considered for the incubation time. First, in order to shorten the total processing time for RIINGS, the incubation time for each individual step should not be too long. Second, variation of the fragment size of the product RNAs would be expected to be great if the incubation time is too short. For example, if the incubation time is set to 5 minutes and you need 3 minutes to finish the process for 10 samples, a variation of 3 minutes is equivalent to adding an additional 60% of incubation time to the original

incubation. Considering manipulation convenience, the incubation time was therefore fixed at 30 minutes. It would be an appropriate incubation time for this step.

When the digestion temperature and time are fixed, the concentration of the input RNase is the only remaining parameter that requires optimization. The results of evaluation of fragment size after RNase treatment indicated that the fragment size was negatively correlated with the input concentration of the both RNases (see Sections 5.6.1 and 5.6.2). The results of evaluation of the accessibility of the immunoprecipitated AGO complex for ligation after RNase treatment indicated that treating with 1:100-diluted and 1:1000-diluted RNase A gave significantly different results (see Section 5.6.3). This result also implies that an increase of RNase A concentration between these dilutions can significantly decrease the accessibility of the retained ends for adaptor ligation. Likewise, treating with undiluted and 1:100-diluted RNase One™ Ribonuclease also gave significantly different results (see Section 5.6.4). This in turn implies that an increase of RNase One™ Ribonuclease concentration between these dilutions can significantly decrease the accessibility of the retained ends for adaptor ligation. Considering all these results (fragment size and accessibility), 1:100-diluted and 1:1000-diluted RNase A, and undiluted and 1:100-diluted RNase One™ Ribonuclease were the chosen concentrations of these RNases for use in the pilot run of RIINGS.

6.6 The stability of the isolated miRNA-target RNA duplexes

Not surprisingly, the measured melting temperature of the isolated miRNA-target RNA duplexes from living cells (K562) is lower than that of a long duplex about ~20 bp. As we know, the melting temperature of primers (~20 bp) is usually about 50-60℃. Furthermore, the stability of duplex is known to be in the order of double-stranded RNA > double-

stranded DNA [325]. However, why is the melting temperature of the duplexes that are extracted from living cells (K562) only 40℃ or less (see results in Section 5.9.2)? It is likely due to the existence of mismatch(es) and bulging(s) in the miRNA-target RNA duplexes. A previous study has found that the change of melting temperature (ΔT_m) in double-stranded DNA with 0.55% to 7.2% base pair mismatch corresponds to a reduction of 0.28-4.17℃ in melting temperature [326]. A DNA duplex with 5% base pair mismatch (equivalent to 1 mismatch in a 20-bp duplex) corresponds to a reduction of ~3℃ in melting temperature [326]. In fact, a pairing between miRNA and target RNA usually contains multiple mismatches and even with bulging. Therefore, the low melting temperature of the miRNA-target RNA duplexes obtained from a living cell can be explained.

6.7 Why is the purity of the Argonaute complexes important and how pure should the complexes be?

When Argonaute-mediated miRNA-target RNA interactions are to be studied, the Argonaute-miRNA-target RNA complexes are the complexes that we need to focus on. Each of these complexes includes an RNA duplex inside. However, the two individual RNA sequences are not covalently linked together. The two RNA sequences are hybridized together only via some relatively weak forces, e.g. hydrogen bonds. Accordingly, the two interacting RNA sequences have to be linked together before sequencing via ligating a stem-loop adaptor and then filling the gap between the two individual RNA sequences in the design of RIINGS. In order to achieve ligation, the protein part (Argonaute) of the complex must be removed in advance. It is because the Argonaute protein most likely

shields the ends of the miRNA as has been shown in previous crystal structure investigations of the Argonaute complexes [133, 214, 219, 309].

Obviously, purified miRNA-target RNA duplexes without any contaminants are the required input for the chimeric sequence preparation in RIINGS. Why do the Argonaute complexes require purification? First, the chimeric sequences prepared by RIINGS have to be determined by next-generation sequencing (NGS) in RIINGS. If the Argonaute complexes are contaminated with some other RNA binding proteins (RBPs) during the library preparation, the RNAs from these unknown RBPs can also be sequenced by NGS. As a result, these junk reads would consume the sequencing capacity and reduce the read depths of the genuine chimeric sequences. The more are the junk reads, the less are the chimeric reads that represent the interacting miRNAs and their corresponding target RNAs (the meaningful reads). Therefore, depleting the intrinsic RBPs is one of the aims of the purification step. Second, proteinase K is very robust and efficient in digesting nearly all the proteins to their corresponding residues (amino acids), including keratin. Thus, proteinase K is the proteinase selected for removing the Argonaute protein in RIINGS. To this end, the purity of the isolated AGO complex is still not guaranteed because both the AGO protein on the interacting miRNA-target RNA duplex and other impurities (co-purified proteins) can be digested by proteinase K. Nevertheless, the kinetics of proteinase K follows the Michaelis-Menten equation as proteinase K is an enzyme. The reaction velocity would reach a maximum value when the substrate concentration is high enough and the enzyme concentration remains unchanged. In other words, the enzymatic reaction takes longer time to reach the chemical equilibrium if the substrate concentration is high. Therefore, minimizing the input protein amount can shorten the reaction time for digesting the Argonaute protein in the process of RIINGS. The purer the Argonaute complex is obtained, the more efficient the proteinase K digests.

6.8 Repair of mismatches at microRNA ends

Mismatches can usually be identified in miRNA targeting because the binding of miRNAs to their targets does not require full complementarity. In other words, partial complementarity is sufficient for the miRNA to function. These mismatches often come to the 5' and 3' ends of miRNAs too. Therefore, mismatch repairing at both ends of miRNAs before the conduction of gap filling is critical because there is no ligase that is able to ligate a nick with a mismatch at either side of the nicking site. However, there are several remarkable difficulties in this task. First, the sequence of the target RNA is not known in advance in the real situation. Second, a repairer with a particular sequence is unable to be ligated to a particular end specifically even if the required sequence is known. Third, the requisite length of the repairer is not easy to determine because the minimum number of bases needed to reform a stable complimentary end is unknown. Fourth, the rate of successful repair would be extremely low if the repairer is made up of degenerate bases. For example, a tetramer N (NNNN) would generate a complimentary end with all four bases matched with a probability of $(1/4)^4$ or 1/256. Even if wobble base pairing (G:U pairing) [319] is considered in the design of repairers, the calculated chance for four bases matched is $(1/3)^4$ or 1/81. (The calculation can be explained below. U, C and G are enough for pairing up with all possible canonical RNA bases with wobble base pairing because G can replace the function of A in pairing. In these three remaining bases, U pairs up with A with a chance of 1/4, C pairs up with G with a chance of 1/4 and G pairs up with C and U with a chance of 1/2. Therefore, the average chance of matching for a degenerate base that only contains U, C and G is 1/3. For four degenerate bases, the chance is reduced to

1/81. However, I eventually found that this calculation was wrong. The correct calculation is discussed in Section 6.8.6).

6.8.1 Contribution of universal base in the mismatch repairers

In order to solve the problems raised above, reducing the complexity of the two repairers is a straightforward solution to the difficulties in designing a set of suitable repairers for current mismatch repairs. Two kinds of bases can be chosen for the above purpose: natural universal base and artificial universal base. Inosine is a well-known natural universal base. It commonly exists in tRNAs and other RNAs [327, 328]. The advantage of inosine is that it may be utilized by T4 RNA Ligase 1 (a ligase used in the ligation of repairers) and other enzymes because it is a natural base. However, inosine is not truly universal in base pairing. It only pairs up with A, U and C in RNA [319] (and with A, T, C and G in DNA [329]), and preferentially pairs up with C in DNA, but not the others bases [329]. On the other hand, 5-nitroindole is an artificial base also considered in current construction of repairers. The advantage of 5-nitroindole is that it can non-selectively pair up with all canonical natural DNA bases, i.e. A, T, C and G because it is a true universal base [320, 321]. However, the commercially available form of 5-nitroindole is deoxyribonucleoside 5-nitroindole. Together with its artificial nature, 5-nitroindole base *may* not be a good substrate for T4 RNA Ligase 1 and other enzymes.

6.8.2 Ligation of miRNA 5' and 3' repairers by T4 RNA Ligase 1

Ligation of repairers (both miRNA 5' and 3' repairers) involves the ligation of *single-stranded* repairer (RNA) to *single-stranded* miRNA (here single-stranded because of mismatches with the miRNA target), i.e. the joining of two *single-stranded* RNA molecules. This kind of ligation can be carried out by T4 RNA Ligase 1 (NEB) — an RNA ligase that

catalyzes the formation of phosphodiester bond between a 5’ phosphate group and a 3’ hydroxyl group through a condensation reaction [314, 330]. This enzyme *actively* catalyzes the joining of two single-stranded RNAs or one single-stranded RNA (self-ligation), but it also catalyzes the joining of nicks on double-stranded RNA. The minimal substrates of T4 RNA Ligase 1 in the joining of two single-stranded RNAs are a nucleoside 3’,5’-bisphosphate donor (pNp) and a trinucleoside diphosphate acceptor (NpNpN) [314].

Both inosine and 5-nitroindole were evaluated in the construction of 5’ and 3’ repairers (see Section 5.11). The results indicated that inosine did not show any advantages for being a substrate of T4 RNA Ligase 1. Both inosine and 5-nitrindole should not be included in the minimal substrates of T4 RNA Ligase 1. In other words, the 5’ end of the RNA for joining and the first 3 bases at the 3’ end of the RNA for joining should not be inosine or 5-nitroindole. If inosine or 5-nitroindole is included in these particular positions, the ligation of repairers to either sides of miRNA can be strongly impaired. These requirements suggest that the first 3 bases at the 3’ end (donor end) of the miRNA 5’ repairer and the 5’ end (acceptor end) of the miRNA 3’ repairer should be canonical bases. To maintain the ability in repairing the mismatches, random canonical bases are employed in these particular sites. On the other hand, 5-nitroindole is much more appropriate to be placed in other positions of the repairers because inosine does not show any advantages for being a substrate of T4 RNA Ligase 1 and 5-nitroindole is a true universal base in pairing up with all canonical bases. Thus, 5-nitroindole is selected in constructing the two mismatch repairers in RIINGS.

6.8.3 Repair of 5-nitroindole-containing nicks by T4 RNA Ligase 2

On the other hand, two essential processes for generating the chimeric sequence of interest in RIINGS is needed after the ligation of the repairers for repairing the mismatches

at the 5’ and the 3’ ends of miRNAs. One of them is the repair of the gap between the 5’ end of the miRNA 5’ repairer and the 3’ end of the “Stem-Loop Adaptor (RNA)”. The other one is the extension of the 3’ end of the miRNA 3’ repairer for allowing this end to reach and form duplex at the restriction enzyme site that is located at the “Re-attaching Adaptor”. Both of these processes are achieved by the repeated ligation of the “Random Tetramer (5’ phosphorylated)” in RIINGS. This is a strategy of using random tetramer to fill up and repair the gap between the 5’ end of the miRNA 5’ repairer and the 3’ end of the “Stem-Loop Adaptor (RNA)” sequentially, and concurrently extend the 3’ end of the miRNA 3’ repairer (see Figure 3.1). In fact, this ligation is a process of continuous nick repair in double-stranded manner and can be carried out by T4 RNA Ligase 2. In contrast to T4 RNA Ligase 1, T4 RNA Ligase 2 is an enzyme that *actively* catalyzes the joining of nicks on *double-stranded* RNA although it also catalyzes the joining of two single-stranded RNAs or a single-stranded RNA (RNA self-ligation).

Since miRNA 5’ and 3’ repairers are 5-nitroindole-containing sequences, the nicks involving these two repairers are different from nicks not involving them. In reality, there is no information at all about whether the T4 RNA Ligase 2 can deal with these two kinds of nicks (One kind contains 5-nitroindole close to the donor (3’) end while the other kind contains 5-nitroindole close to the acceptor (3’) end). Therefore, a set of additional mock experiments was conducted for evaluating the joining of these two kinds of nicks. The results indicated that the joining of the two kinds of nicks was still allowed even when the second position of the donor (5’) end or the acceptor (3’) end was a 5-nitroindole base (see Section 5.12). However, additional experiments have not yet been conducted for investigating whether ligation can still occur if the 5-nitroindole base is placed at the first position of the donor end or the acceptor end. It is because the binding of 5-nitroindole to any of the four canonical bases is weaker than C≡G pairing and A=U pairing [320, 331]. If

this less stable base pairing is introduced to the first position of the new extended ends (the 5' end of the 5' repairer and the 3' end of the 3' repairer), the stability of these two repairers may reduce.

6.8.4 Reverse transcription of 5-nitroindole-containing RNA template

As the 5-nitroindole-containing miRNA repairers are used in the construction of the chimeric sequence, the final chimeric sequence product also contains 5-nitroindole. This product is in fact a pool of 5-nitroindole-containing RNA sequences. These chimeric sequences need to be converted to DNA sequences before high-throughput sequencing. In the current investigation, *Bst* 3.0 DNA Polymerase (NEB) is the only polymerase that can reverse-transcribe the 5-nitroindole-containing RNA to cDNA. *Bst* 3.0 DNA Polymerase is a DNA/RNA-dependent DNA polymerase. This enzyme can choose either DNA or RNA sequence as the template to synthesize a complementary DNA sequence from a primer. In the evaluation of *Bst* 3.0 DNA Polymerase, two continuous 5-nitroindole bases at the template RNA can obviously impair the extension of the complementary sequence and three continuous 5-nitroindole bases can completely block the extension of the complementary sequence. The polymerase can only deal with a single substitution of 5-nitroindole in RNA template.

6.8.5 The setup of the miRNA 5' and 3' repairers

According to the above findings, the first three bases at the 3' end of the miRNA 5' repairer should be canonical bases. If tetramer repairer is desired, the fourth position from the 3' end (or the first position from the 5' end) should be canonical base as well because 5-nitroindole base is not preferred at the first position of the new extended end (see Section 6.8.3 for the reason). Alternatively, the fourth position can be 5-nitroindole base if a

pentamer repairer is desired because a canonical base can be placed at the fifth position so that the 5-nitroindole base will not be located at the first position of the new extended end. In this design, the probability of the complementary of the entire pentamer repairer to a particular sequence is the matching probability of each canonical base raised to the power of four ($[P_{canonical\ base}]^4$), the same probability of a tetramer repairer. However, if only the first four bases at the new extended end (5' end of the 5' repairer), the probability of complementarity would increase to $[P_{canonical\ base}]^3$. Moreover, the existence of 5-nitroindole in the 5' repairer can indeed help to determine the 5' end of the miRNA because 5' heterogeneity is commonly present in miRNAs. The 5-nitroindole base would be converted into canonical bases during reverse transcription. As the result, an induced variation can be detected in the sequencing result and regarded as a mark. Thus, the pentamer design is selected for constructing the miRNA 5' end repairer in RIINGS.

On the other hand, the first position at the 5' end of the miRNA 3' repairer must be a canonical base according to the results in Section 5.11.2.2. Meanwhile, the assignment of a canonical base at the first position of the 3' end is also preferable in this repairer because 5-nitroindole is not desirable at the first position of the new extended end. Therefore, only the two remaining positions in the middle are suitable for the 5-nitroindole base. As continuous 5-nitroindole bases at RNA template can strongly inhibit the reverse transcription, only one 5-nitroindole base can be included in the 3' repairer. In the pilot design, the 5-nitroindole base is placed at the third position from the 5' end. This design might have the advantage of stabilizing the binding of the 3' repairer. Because the 5' end of the 3' repairer is the end directly connected to the mismatched 3' end of the miRNA and this end may thus be very unstable in annealing, the placement of 5-nitroindole away from this end may have some positive effect in stabilizing the binding of the 3' repairer. Likewise,

the existence of 5-nitroindole in the 3' repairer can indeed help to determine the 3' end of the miRNA as 3' heterogeneity is commonly present in miRNAs.

6.8.6 The use of canonical bases in the repairers and the calculation of the matching probability

Canonical RNA bases have to be included in both miRNA 5' and 3' repairers. There are 4 canonical RNA bases, i.e. A, U, C and G. Unlike the base pairing in DNA bases, G:U pairing is also allowed in RNA base pairing. The G:U pairing is one of the wobble base pairings [319]. With this additional base pairing, the pairing of all the four canonical bases can indeed be covered by only two bases because G can pair up with C and U, and U can pair up with A and G. Therefore, random G/U base is good enough for matching up with all four possible canonical bases.

Using random G/U for constructing the miRNA 5' and 3' repairer can reduce the complexity of these repairers when compared with random A/U/C/G. Accordingly, the matching probability of the 5' repairer should be $(1/2)^4 = 1/16$ if random G/U is included in all four canonical base positions in the 5' repairer. The matching probability would increase to $(1/2)^3 = 1/8$ if only the first four bases are considered at the new extended end of this repairer (since the second position from the 5' end of the 5' repairer is a 5-nitroindole, the power in the above calculation should be 3 instead of 4). On the other hand, the matching probability of the 3' repairer should be $(1/2)^3 = 1/8$ as it has 3 canonical base positions. Consequently, the probability of repairing both ends of each miRNA is $(1/8)^2 = 1/64$.

6.9 The repair of the gap for joining the two interacting RNAs

After the “Stem-loop Adaptor (RNA)” is ligated to the 3’ end of target RNA, a gap is generated between the 5’ end of miRNA and 3’ end of the stem-loop adaptor (see Figure 3.1 and Section 6.3). Two possible approaches can be used to fill this gap for joining the two interacting RNAs together. The first approach is “gap repair by synthesis”. The second approach is “gap repair by ligation”.

In the first approach, the gap is first filled by an appropriate polymerase. After that, the nick between the newly synthesized sequence and the miRNA has to be sealed by a ligase (Figure 6.9.1, gap repair by synthesis). In fact, this approach is quite complicated. First, the polymerase must be able to synthesize a complementary strand for filling up the gap with the given RNA template and such polymerase should not have strand displacement activity. RNA-dependent DNA polymerase and RNA-dependent RNA polymerase are the two kinds of polymerases that can utilize RNA as templates for synthesizing a complementary strand. Polymerases of the former category are used to assemble complementary DNA strand while polymerases of the latter category are used to assemble complementary RNA strand. Second, the nick between the 5’ end of miRNA and the 3’ end of the newly synthesized sequence (Figure 6.9.1, gap repair by synthesis) needs to be sealed. However, no ligase can deal with the nick that is formed between the 5’ end of RNA and the 3’ end of DNA on a template RNA (Figure 6.9.1, left column of gap repair by synthesis) [312]. Therefore, gap filling by RNA-dependent DNA polymerase is not applicable here and only the RNA-dependent RNA polymerase is potentially usable in the process of “gap repair by synthesis”. However, the use of RNA-dependent RNA polymerase in laboratory applications is lagging far behind when compared with other kinds of polymerases. Therefore, this kind of polymerases lacks commercial value and commercial sources. Phi6 RNA Replicase is an

RNA-dependent RNA polymerase supplied by Thermo Scientific. This polymerase is a primer-independent polymerase, which initiates the synthesis of complementary RNA from the 3' end of a single-stranded nucleic acid template (DNA or RNA). According to the product description, this polymerase may have template specificity, which means it requires a specific recognition sequence for initiating the synthesis. Furthermore, the provided protocol indicates that this enzyme has strand displacement activity on RNA/DNA hybrid. However, we do not know whether this enzyme also has strand displacement activity on double-stranded RNA, which is detrimental to RIINGS. Considering all the above issues, the Phi6 RNA Replicase is not a suitable polymerase for filling the gap in RIINGS. Moreover, our knowledge of the properties of RNA-dependent RNA polymerases is lacking. Therefore, we are unable to find a suitable RNA-dependent RNA polymerase for evaluating the possibility of "gap repair by synthesis" up to the moment of submitting this thesis.

In the second approach, the gap has to be filled up by short RNA oligos, and the reaction is catalyzed by a ligase (Figure 6.9.1, gap repair by ligation). T4 RNA ligase 2 (NEB) is a ligase that preferentially ligates double-stranded RNA or nick on double-stranded RNA. The properties of this enzyme indeed meet the requirements of "gap repair by ligation". On the other hand, the short oligo for filling up the gap should have random sequence because the sequence of the miRNA target within the gap is unknown and variable. Moreover, one great disadvantage of this gap-filling strategy is that only the gap with a length equal to a multiple of the length of the short oligo can be repaired. For example, a tetramer RNA oligo only can fix the gap with a length equal to a multiple of four (4, 8, 12, 16, etc.). Tetramer RNA oligo cannot fix 10-base gap, 15-base gap, 17-base gap, etc. This particular setting also means that only 25% of the interacting miRNA-target RNA duplexes can be conjugated if a tetramer oligo is used. Actually, the shorter is the filler RNA used in the gap filling, the higher portion the duplexes can be conjugated, e.g. monomer can fill up 100% of gap,

dimer can fill up 50% of gap, trimer can fill up 33% of gap, tetramer can fill up 25% of gap and pentamer can fill up 20% of gap. Currently, the shortest RNA oligo available from IDT is tetramer. Therefore, random tetramer RNA oligo has been selected and evaluated in “gap repair by ligation” (see Section 5.13 for the results). Additionally, RNA monomer (NMPs) was also evaluated; however, the monomer is not a suitable substrate of T4 RNA ligase 2 (see Section 5.13.1). Thus, only the tetramer RNA oligo can be used as the filler currently and the maximum effectiveness of the current gap repair approach is 25%.

Notably, a large amount of mock duplex still cannot be ligated even the incubation has been performed at 16℃ and extended to 18 hours (see Section 5.13.2 and Figure 5.13.3). This indicates that the strategy of “gap repair by ligation” is working, but the efficiency is not too high. This is possibly because the filler (5-P Tetramer RNA) has been consumed in single-strand ligation (see Figure 5.13.2). However, because the process of “gap repair by ligation” is an on-bead process in the real situation, we can wash away the used ligation mixture and replace with new ligation mixture repeatedly until the gap is repaired.

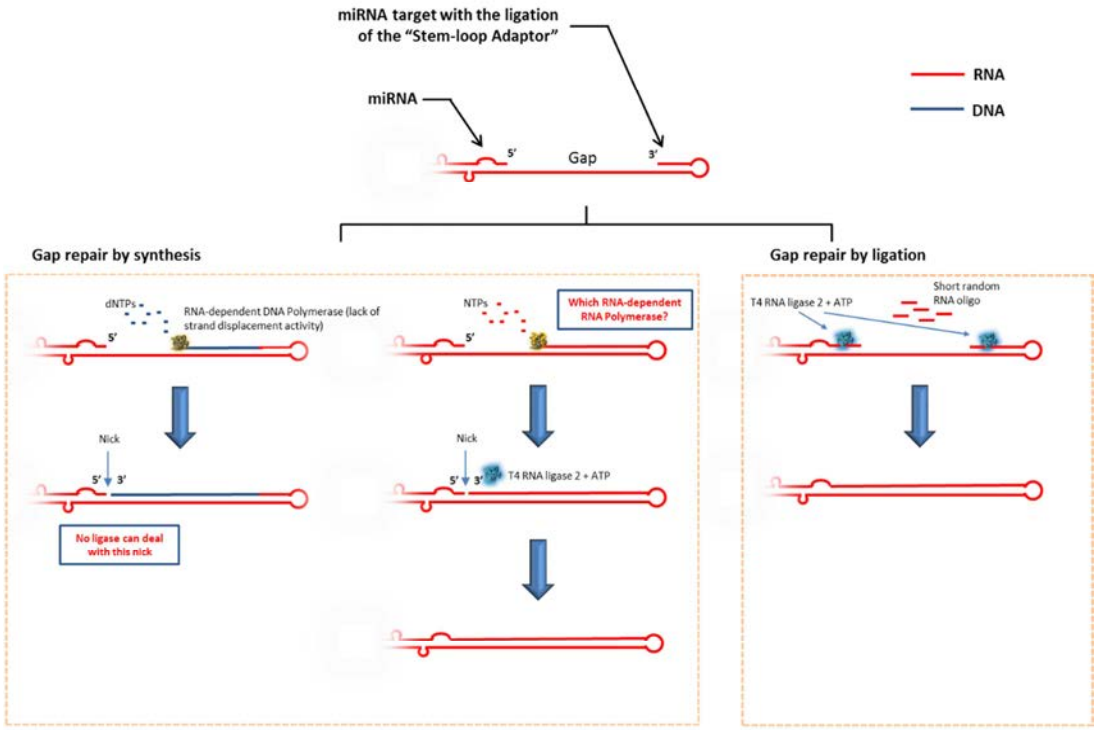


Figure 6.9.1 Two gap-repair strategies: "gap repair by synthesis" and "gap repair by ligation".

6.10 Advantages and applications of RIINGS

Compared with previously developed experimental methods for identifying microRNA targets, the idea of the current method has the following advantages in theory. First, the current method is able to identify the miRNA-target site interactions in exact correspondence (not all the developed methods do, see Section 1.5). Second, the current method potentially enhances the generation of microRNA-target RNA chimeras by removing the spatial hindrance from Argonaute proteins. In addition, the removal of Argonaute proteins before generating chimeras can also avoid the bias in the generation of chimeras because only both miRNA and target RNA simultaneously protruding from the *same* AGO complex can be ligated and then generate the valid chimeras. Third, the approach to generating chimeric sequences in RIINGS is most likely able to avoid the generation of fake chimeras because the 3' end of the target site (the end for generating chimeras) is protected by a stem-loop structure. Fourth, the increased efficiency of generating chimeras can result in a much better representation of the *entire* population of the miRNA-target site interactions. Therefore, RIINGS is an important method with high value in identifying miRNA interactomes.

Since miRNAs play a key role in gene expression regulation (mRNA gene expression and long non-coding RNA functioning), the understanding of miRNA interactomes may explain many of phenomena in cells, such as differentiation, cell cycle regulation, cell metabolism and tumorigenesis. By the comparison of the differences of miRNA interactomes between different states of cells (e.g. cancer cells vs. normal counterpart), the detailed control and regulation of gene expression in response to the changes may be revealed. Because miRNAs can target many non-canonical target sites and target genes, e.g. 5' UTRs, introns, multiple types of non-coding RNAs and pseudogenes [134, 295, 297, 299, 302], some

currently hidden functions of miRNAs may also be discovered. Therefore, a comprehensive method for identifying miRNA interactome is attractive in biology.

6.11 Future work

The first and most important work is to continue the construction of Library Enrichment Beads. These magnetic beads capture the interacting miRNA-target RNA duplexes released from Argonaute complex, and then allow the subsequent on-bead enzymatic reactions to be performed on the captured miRNA-target RNA duplex for generating the chimeric sequence. The current problem of the developed beads is that the captured mock targets are not efficiently released from the beads, i.e. the recovery rate of the captured sequence is low. This is most likely due to either or both of two possibilities. First, the spatial hindrance around the restriction enzyme site is great. That means that the restriction enzyme site for releasing the current mock target sequence by BanI is too close to the beads so that the BanI restriction enzyme is unable to cleave the cutting site and release the mock target. Second, the PEG coating may still be not good enough for covering the beads entirely. Originally, the beads are amine-modified magnetic beads. Each functional group has a positive charge. These positively charged amine groups on the beads are able to bind to negatively charged nucleic acid and cause non-specific enrichment of nucleic acids. This effect can be extended to the released mock target in the current evaluation and results in the released mock target re-attaching on the beads. To overcome the above two issues, a sandwich-like setting is worth exploring.

The RIINGS method can potentially solve the mystery of miRNA interactomes. This Pandora box has been an unsolvable challenge in miRNA study and post-transcriptional regulation of

gene expression for about two decades. CLASH and MARIO are two experimental methods that are able to investigate the miRNA-target RNA interaction. However, their performance is not good enough. The miRNA interactions identified by CLASH were only supported by less than 2% of chimeric reads and more than 10% of these chimeric reads were fake chimeras as reported by the research team. MARIO used by the original group discovered only two hundred miRNA-target RNA interactions in the test cell model (including 100 miRNA-mRNA interactions and 100 miRNA-snoRNA interactions). With RIINGS, we may able to unlock this unreachable mystery. This new idea may also open a new and invaluable research approach other than gene expression study and genomic study, which have already generated many fruitful results in the past. This new approach is the interactomic study. It has a tremendous potential in making breakthroughs to the current biological studies. If the genomic studies were the methods to identify the parts of a machine and the gene expression studies were the methods to count the number of piece of each kind of screw, gear, electronic component, etc. The interactomic study would be the method to investigate the blueprint of the connections between the parts and how the parts working with each other.

7. Conclusions

RNA Interaction Identification by Next Generation Sequencing (RIINGS) is a method for investigating the exact corresponding interactions between microRNAs and their target RNAs on the scale of entire interactome.

In the evaluation of the expression of AGO proteins in three different cell lines, the expression of AGO proteins in K562 cells is higher than in MEC-1 cells and Raji cells.

The antibody eIF2C2 (4F9) (Santa Cruz) is selected for the immunoprecipitation step in RIINGS. Immunoprecipitation method that first incubates the total cell lysate with the antibodies for a period of time and then captures the antibody-target protein complexes by adding the beads to the lysate-antibody mixture has been selected in RIINGS.

During the method development, a novel approach to linking the interacting miRNA-target RNA duplex via the stem-loop adaptor has been developed. This approach circumvents the restrictions imposed by ligase when it is used to ligate the interacting nucleic acid partners. This approach can be applied to the linking of other interacting nucleic acids with some minor modifications.

Re-attaching Adaptor for RIINGS has been designed and the ligation condition of this adaptor to the 5’ end of the miRNA target site on *in vivo* isolated-AGO complex has been well optimized.

The condition for RNase digestion has been well optimized for both RNase A and RNase I.

Eight different crosslinking conditions have been evaluated. However, there is no big difference between these conditions in the amounts of yielded AGO complex. Thus, all the eight crosslinking conditions can be used for the crosslinking in RIINGS.

T4 polynucleotide kinase can be used to convert the 2′,3′-cyclic monophosphate end of the fragmented RNA to 3’ hydroxyl end. This treatment can improve the ligation of adaptor to 3’ end of RNase-treated RNA.

During the method development, a pair of mismatch repairers was invented. This pair of repairers each has particular sequence and length. The design minimizes the complexity of the two repairers and results in greater probability in repairing the 5' mismatch and the 3' mismatch on the miRNA to 1/16 and 1/8 respectively.

The purity of the current immunoprecipitated AGO complex is good enough for RIINGS. No further purification is required for the AGO complex.

Melting curve analysis has found that the melting peak of miRNA-target RNA duplexes is about 40℃ and the re-annealing of the melted duplexes is unlikely to occur. These evidences support the feasibility of RIINGS.

Two approaches for estimating the amount of *in vivo* isolated-miRNA-target RNA duplex has been developed. One approach is used to estimate the amount of yielded miRNA-target RNA duplex (before the two RNA sequences are linked) and another approach is used to quantify the amount of chimeras in the RIINGS-prepared library (after the two RNA sequences are linked).

Proteinase K is able to digest proteins at low temperature (4-16°C). It is able to release the interacting miRNA-target RNA duplex from the *in vivo* isolated-AGO complex at 4°C.

Appendices

Appendix A: Preparation of buffers and stock solutions

1. 0.5 M Tris-HCl, pH 8.0

Reagent	Amount to add	Final concentration
Tris	30.29 g	0.5 M
Milli-Q Water	400 ml	--

Adjust the pH to 8.0 with HCl. Then adjust the volume to 500 ml with Milli-Q Water and sterilize by autoclave. Store the buffer at 4℃.

2. 0.5 M Tris-HCl, pH 6.8

Reagent	Amount to add	Final concentration
Tris	30.29 g	0.5 M
Milli-Q Water	400 ml	--

Adjust the pH to 6.8 with HCl. Then adjust the volume to 500 ml with Milli-Q Water and sterilize by autoclave. Store the buffer at 4℃.

3. 1.5 M Tris-HCl, pH 8.8

Reagent	Amount to add	Final concentration
Tris	90.86 g	0.5 M
Milli-Q Water	400 ml	--

Adjust the pH to 8.8 with HCl. Then adjust the volume to 500 ml with Milli-Q Water and sterilize by autoclave. Store the buffer at 4℃.

4. 0.5 M EDTA, pH 8.0

Reagent	Amount to add	Final concentration
EDTA disodium salt dihydrate	93.06 g	0.5 M
Milli-Q Water	400 ml	--

Adjust the pH to 8.0 with NaOH. Then adjust the volume to 500 ml with Milli-Q Water and sterilize by autoclave. Store the buffer at room temperature.

5. 10% (v/v) Nonidet P-40 (NP-40)

Reagent	Amount to add	Final concentration
NP-40	5 ml	10% (v/v)

Adjust the volume to 50 ml with UltraPure™ Distilled Water (Life Technologies). Store the solution at room temperature.

6. 10% (v/v) Tween® 20

Reagent	Amount to add	Final concentration
Tween 20	5 ml	10% (v/v)

Adjust the volume to 50 ml with UltraPure™ Distilled Water (Life Technologies). Store the solution at room temperature.

7. 10% (w/v) Sodium dodecyl sulfate (SDS)

Reagent	Amount to add	Final concentration
SDS	5 g	10% (w/v)

Dissolve in 40 ml UltraPure™ distilled water (Life Technologies). Then adjust the volume to 50 ml. Store the solution at room temperature.

8. 30% (w/v) Acrylamide (29:1)

Reagent	Amount to add	Final concentration
Acrylamide	72.5 g	29% (w/v)
N,N'-methylenebisacrylamide	2.5 g	1% (w/v)

Dissolve in 200 ml Milli-Q water. Then adjust the volume to 250 ml. Store the solution in dark at 4℃.

9. 10% (w/v) Ammonium persulfate

Reagent	Amount to add	Final concentration
Ammonium persulfate	1 g	10% (w/v)

Dissolve in 8 ml Milli-Q water. Then adjust the volume to 10 ml. This solution only can be stored at 4℃ for no more than 2-3 weeks because ammonium persulfate slowly decay in aqueous form.

10. 10X SDS-PAGE Running Buffer

Reagent	Amount to add	Final concentration
Tris	30.3 g	250 mM
Glycine	144 g	1.92 M
SDS	10 g	1% (w/v)

Dissolve in 800 ml Milli-Q water. Then adjust the volume to 1 L. The pH of the buffer should be around 8.3 and pH calibration is not needed. Store the solution at room temperature.

11. 1X SDS-PAGE Running Buffer

Reagent	Amount to add	Final concentration
10X SDS-PAGE Running Buffer	100 ml	1X
Milli-Q water	900 ml	--

Store the solution at room temperature.

12. 1X Transfer Buffer

Reagent	Amount to add	Final concentration
Tris	6.06 g	25 mM
Glycine	28.8 g	192 mM
Methanol	400 ml	20% (v/v)

Dissolve in 1200 ml Milli-Q water. Then adjust the volume to 2 L. The pH of the buffer should be around 8.3 and pH calibration is not needed. Store the solution at 4℃.

13. 10X TBST

Reagent	Amount to add	Final concentration
Tris	24.23 g	200 mM
NaCl	87.66 g	1.5 M
Tween 20	10 ml	1% (v/v)
Milli-Q Water	800 ml	--

Adjust the pH to 7.5 with HCl. Then adjust the volume to 1 L with Milli-Q Water. Store the solution at room temperature.

14. 1X TBST

Reagent	Amount to add	Final concentration
10X TBST	100 ml	1X
Milli-Q water	900 ml	--

Store the solution at room temperature.

15. 1X Detergent-contained Lysis Buffer

Reagent	Amount to add	Final concentration
0.5 M Tris-HCl, pH 8.0	2 ml	20 mM
5 M NaCl (Life Technologies)	1.37 ml	137 mM
10% NP-40	5 ml	1%
0.5 M EDTA, pH 8.0	200 µl	2 mM
cOmplete ULTRA Tablets (Roche)	1 tablet	1X

Adjust the volume to 50 ml with UltraPure™ Distilled Water (Life Technologies). Store the buffer at -20°C.

16. 1X Detergent-free Lysis Buffer

Reagent	Amount to add	Final concentration
10X PBS (Life Technologies)	5 ml	1X
0.5 M EDTA, pH 8.0	0.5 ml	5 mM
cOmplete ULTRA Tablets (Roche)	1 tablet	1X

Adjust the volume to 50 ml with UltraPure™ Distilled Water (Life Technologies). Store the buffer at -20℃.

17. 1X Detergent-free Lysis Buffer (protease inhibitor-free)

Reagent	Amount to add	Final concentration
10X PBS (Life Technologies)	5 ml	1X
0.5 M EDTA, pH 8.0	0.5 ml	5 mM

Adjust the volume to 50 ml with UltraPure™ Distilled Water (Life Technologies). Store the buffer at -20℃.

18. Denaturing lysis Buffer

Reagent	Amount to add	Final concentration
0.5 M EDTA, pH 8.0	0.5 ml	5 mM
10% (w/v) SDS	5 ml	1%
cOmplete ULTRA Tablets (Roche)	1 tablet	1X

Adjust the volume to 50 ml with UltraPure™ Distilled Water (Life Technologies). Add 100 µl of 0.1 M dithiothreitol to 900 µl of the above buffer to obtain a final concentration of 10 mM dithiothreitol before use. Store the buffer at RT up to 1 week.

19. 2X Laemmli Sample Buffer

Reagent	Amount to add	Final concentration
0.5 M Tris-HCl, pH 6.8	1.25 ml	0.0625M
10% (w/v) SDS	2 ml	2%
Glycerol	2.5 ml	25%
0.5% (w/v) bromophenol blue	0.2 ml	0.01%
Deionized water	3.55 ml	--

Add 50 µl of β-mercaptoethanol to 950 µl of the above buffer to obtain a final concentration of 5% β-mercaptoethanol before use. Store the buffer at 4℃.

20. 1X Laemmli Sample Buffer

Reagent	Amount to add	Final concentration
0.5 M Tris-HCl, pH 6.8	0.625 ml	0.03125M
10% (w/v) SDS	1 ml	1%
Glycerol	1.25 ml	12.5%
0.5% (w/v) bromophenol blue	0.1 ml	0.005%
Deionized water	7 ml	--

Add 25 µl of β-mercaptoethanol to 975 µl of the above buffer to obtain a final concentration of 5% β-mercaptoethanol before use. Store the buffer at 4℃.

21. 2X Laemmli Sample Buffer (without dye)

Reagent	Amount to add	Final concentration
0.5 M Tris-HCl, pH 6.8	1.25 ml	0.0625M
10% (w/v) SDS	2 ml	2%
Glycerol	2.5 ml	25%
Deionized water	3.75 ml	--

Add 50 µl of β-mercaptoethanol to 950 µl of the above buffer to obtain a final concentration of 5% β-mercaptoethanol before use. Store the buffer at 4°C.

Appendix B: Preparation of polyacrylamide gels

SDS polyacrylamide gel

Stacking gel

	4% Gel
Milli-Q water	3 ml
30% Acrylamide	0.67 ml
0.5 M Tris-HCl, pH 6.8	1.25 ml
10% SDS	50 µl
10% Ammonium persulfate	50 µl
TEMED (Bio-Rad)	5 µl
Total Volume	**5 ml**

The volume of 5 ml in total is sufficient for casting one stacking gel for Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad).

Separating gel

	8% Gel	10% Gel	15% Gel
Milli-Q water	3.7 ml	3.2 ml	1.8 ml
30% Acrylamide	2.13 ml	2.67 ml	4 ml
1.5 M Tris-HCl, pH 8.8	2 ml	2 ml	2 ml
10% SDS	80 µl	80 µl	80 µl
10% Ammonium persulfate	80 µl	80 µl	80 µl
TEMED (Bio-Rad)	8 µl	8 µl	8 µl
Total Volume	**8 ml**	**8 ml**	**8 ml**

The volume of 8 ml in total is sufficient for casting one separating gel for Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad).

Appendix C: Preparation of reaction mixtures

Table 1 Composition of the radiolabeling reaction for the Stem-Loop Adaptor (RNA).

Reagent	Amount to add	Final concentration
Stem-Loop Adaptor (RNA) (100 µM)	1.2 µl	2.4 µM
10X T4 Polynucleotide Kinase Reaction Buffer (NEB)	5 µl	1X
^{32}P-γ-ATP (10 µCi/µl, 3.3 µM, PerkinElmer)	25 µl	5 µCi/µl
T4 Polynucleotide Kinase (10 U/µl, NEB)	5 µl	1 U/µl
RNase Inhibitor, Murine (40 u/µl, NEB)	1.25 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	12.55 µl	--
Total	**50 µl**	--

Table 2 Composition of the CIP treatment mixture.

Reagent	Amount to add	Final concentration
10X NEBuffer 1 (NEB)	15 µl	1X
Alkaline Phosphatase, Calf Intestinal (10 U/µl, NEB)	7.5 µl	0.5 U/µl
RNase Inhibitor, Murine (40 u/µl, NEB)	3.75 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	123.75 µl	--
Total	**150 µl**	--

Table 3 Composition of the reaction mixture for the ligation of radiolabeled stem-loop adaptor.

Reagent	Amount to add	Final concentration
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	20 µl	1X
T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30 U/µl, NEB)	5-10 µl	1.5-3 U/µl
Radiolabeled Stem-Loop Adaptor (RNA)(2.4 µM)	10-40 µl	240-960 nM
RNase Inhibitor, Murine (40 u/µl, NEB)	2.5 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	32.5-62.5 µl	--
Total	**100 µl**	--

Table 4 Composition of the reaction mixture for radiolabeling the Re-attaching Adaptor (DNA-RNA).

Reagent	Amount to add	Final concentration
Re-attaching Adaptor (DNA-RNA) (100 µM)	1.2 µl	2.4 µM
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	10 µl	1X
[5'-^{32}P] pCp (10 µCi/µl, 3.3 µM, PerkinElmer)	25 µl	5 µCi/µl
T4 RNA Ligase 1, High Concentration (30 U/µl, NEB)	5 µl	3 U/µl
RNase Inhibitor, Murine (40 U/µl, NEB)	1.25 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	7.55 µl	--
Total	**50 µl**	--

Table 5 Composition of PNK treatment mixture with ATP.

Reagent	Amount to add	Final concentration
10X T4 Polynucleotide Kinase Reaction Buffer (NEB)	15 µl	1X
10 mM ATP	15 µl	1 mM
T4 Polynucleotide Kinase, 3' phosphatase minus (10 U/µl, NEB)	7.5 µl	0.5 U/µl
RNase Inhibitor, Murine (40 u/µl, NEB)	3.75 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	108.75 µl	--
Total	**150 µl**	--

Table 6 Composition of the reaction mixture for the ligation of radiolabeled re-attaching adaptor.

Reagent	Amount to add	Final concentration
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	20 µl	1X
T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30 U/µl, NEB)	5 µl	1.5 U/µl
Radiolabeled Re-attaching Adaptor (DNA-RNA)(1.2 µM)	10 µl	120 nM
Unlabeled Re-attaching Adaptor (DNA-RNA)(10 µM)	8.4 µl	840 nM
RNase Inhibitor, Murine (40 U/µl, NEB)	2.5 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	54.1 µl	--
Total	**100 µl**	--

Table 7 Composition of T4 PNK treatment mixture without ATP.

Reagent	Amount to add	Final concentration
10X T4 Polynucleotide Kinase Reaction Buffer (NEB)	15 µl	1X
T4 Polynucleotide Kinase (10 U/µl, NEB)	15 µl	1 U/µl
RNase Inhibitor, Murine (40 U/µl, NEB)	3.75 µl	1 U/µl
UltraPure™ Distilled Water (Life Technologies)	116.25 µl	--
Total	**150** µl	--

Table 8 Preparation of reaction mixtures for melting curve analysis pf RNA duplex.

Condition	1.5 µM Syto9, 7.5 µl sample	1.5 µM Syto9, 10 µl sample	1.5 µM Syto9, 12.5 µl sample	2 µM Syto9, 7.5 µl sample	2 µM Syto9, 10 µl sample	2 µM Syto9, 12.5 µl sample	2.5 µM Syto9, 7.5 µl sample	2.5 µM Syto9, 10 µl sample	2.5 µM Syto9, 12.5 µl sample
Duplex Buffer (IDT)	11.90	9.40	6.90	11.70	9.20	6.70	11.50	9.00	6.50
Syto 9 (50 µM, diluted in water)	0.6	0.6	0.6	0.8	0.8	0.8	1	1	1
Sample (Extract/Blank*)	7.50	10.00	12.50	7.50	10.00	12.50	7.50	10.00	12.50
Total volume	20	20	20	20	20	20	20	20	20

*The blank was 200 µl of 1X Detergent-free Lysis Buffer (protease inhibitor-free) plus corresponding amount of proteinase K.

Table 9 Composition of annealing reaction mixtures for evaluating miRNA 5' end repairer.

Reagent	Amount to add	Final concentration
Artificial Binding Site (RNA)(100 µM)	1.5 µl	90.91 µM
Artificial Mir 3 or Artificial Mir 4 (100 µM)	0.15 µl	9.09 µM
Total	**1.65** µl	--

Table 10 Composition of miRNA 5' repairer ligation mixtures for evaluating miRNA 5' end repairer.

Reagent	Amount to add	Final concentration
Annealing reaction mixture from Table 9 miRNA 5' mismatch repairer (Random C/I Tetramer, tetramer N, tetramer C or tetramer (CCCI))(1000 µM)	1.65 µl 0.75 µl	Artificial Binding Site (RNA): 15 µM Artificial Mir3/4: 1.5 µM 75 µM (50 times of artificial miRNA)
10X T4 RNA Ligase Reaction Buffer (NEB)	1 µl	1X
ATP (100 mM)(NEB)	0.1 µl	1 mM
T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30U/ µl, NEB)	0.5 µl	1.5 U/ µl
UltraPure™ Distilled Water (Life Technologies)	6 µl	--
Total	**10** µl	--

Table 11 Composition of dephosphorylation reaction mixtures

Reagent	Amount to add	Final concentration
Artificial Mir 3 or Artificial Mir 4 (100 µM)	5 µl	25 µM
10X Shrimp Alkaline Phosphatase Reaction Buffer (USB)	2 µl	1X
Shrimp Alkaline Phosphatase (1 U/µl, USB)	5 µl	0.25 U/µl
UltraPure™ Distilled Water (Life Technologies)	8 µl	--
Total	**20** µl	--

Table 12 Composition of annealing reaction mixtures for evaluating miRNA 3' end repairer.

Reagent	Amount to add	Final concentration
Artificial Binding Site (RNA)(100 µM)	1.5 µl	90.91 µM
Dephosphorylated Artificial Mir 3 or dephosphorylated Artificial Mir 4 (25 µM)	0.6 µl	9.09 µM
Total	**2.1** µl	--

Table 13 Composition of miRNA 3’ repairer ligation mixtures for evaluating miRNA 3’ end repairer.

Reagent	Amount to add	Final concentration
Annealing mixture from Table 12	2.1 µl	Artificial Binding Site (RNA): 15 µM Artificial Mir3/4: 1.5 µM
miRNA 3’ mismatch repairer (Tetramer C-Phosphorylated or Tetramer-Phosphorylated (ICCC))(1000 µM)	0.75 µl	75 µM (50 times of artificial miRNA)
10X T4 RNA Ligase Reaction Buffer (NEB)	1 µl	1X
ATP (100 mM)(NEB)	0.1 µl	1 mM
T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30U/ µl, NEB)	0.5 µl	1.5 U/ µl
UltraPure™ Distilled Water (Life Technologies)	5.55 µl	--
Total	**10 µl**	--

Table 14 Table for the setup of miRNA 5’ repairer ligation mixtures in NEBNext® Quick Ligation Reaction Buffer (for miRNA 5’ end repairer evaluation).

Reagent	Amount to add	Final concentration
Annealing mixture from Table 9	1.65 µl	Artificial Binding Site (RNA): 15 µM Artificial Mir3/4: 1.5 µM
miRNA 5’ mismatch repairer (5’ Repairer (GXAGA))(1000 µM)	0.75 µl	75 µM (50 times of artificial miRNA)
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	2 µl	1X
T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (30U/ µl, NEB)	0.5 µl	1.5 U/ µl
UltraPure™ Distilled Water (Life Technologies)	5.1 µl	--
Total	**10 µl**	--

Table 15 Composition of annealing reaction mixtures for evaluating the ligation of two different kinds of nicks that are generated after the repair of the 5' and 3' ends of mismatched miRNA.

Reagent	Amount to add	Final concentration
Repaired 5' end Template or Repaired 3' end Template (100 µM)	0.2 µl	50 µM
Repaired 5' end (GUXG-) or Repaired 5' end (GXAG-) or Repaired 3' end (-XXAG) or Repaired 3' end (-XXXG) (100 µM)	0.2 µl	50 µM
Total	**0.4 µl**	--

Table 16 Composition of ligation mixtures for evaluating the ligation of two different kinds of nicks that are generated after the repair of the 5' and 3' ends of mismatched miRNA.

Reagent	Amount to add	Final concentration
Annealing reaction mixture from Table 15	0.4 µl	2 µM (each oligo)
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	2 µl	1X
T4 RNA Ligase 2 (dsRNA Ligase) (10U/ µl, NEB)	0.5 µl	1.5 U/ µl
UltraPure™ Distilled Water (Life Technologies)	7.1 µl	--
Total	**10 µl**	--

Table 17 Composition of annealing reaction mixtures for evaluating the repair of gap by the ligation approach.

Reagent	Amount to add	Final concentration
Artificial Binding Site (RNA, HEX)(100 µM)	0.2 µl	50 µM
Artificial Mir 5 (FAM) (100 µM)	0.2 µl	50 µM
Total	**0.4 µl**	--

Table 18 Composition of ligation mixtures for evaluating the repair of gap by the ligation approach under the seven conditions. The subtotal 1 should be chilled on ice until the temperature is equilibrated before the next step. The subtotal 2 should be incubated at 4℃ for 30 minutes before the next step. (See Section 4.15.2 for the details of the procedure)

Reagent	**Amount to add**	**Final concentration in the 10 µl reaction**
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	2 µl	1X
5-P Tetramer N (1000 µM)	0-2 µl	0-200 µM
NMPs (10 mM)	0-1 µl	0-1 mM
UltraPure™ Distilled Water (Life Technologies)	4.3-6.1 µl	--
Subtotal 1	**9.1-9.3 µl**	--
Subtotal 1	9.1-9.3 µl	--
Annealing reaction mixture from Table 17	0.2-0.4 µl	1-2 µM (each oligo)
Subtotal 2	**9.5 µl**	--
Subtotal 2	9.5 µl	--
T4 RNA Ligase 2 (10 U/µl, NEB) or T4 DNA Ligase (2000 U/µl, NEB)	0.5 µl	0.5 U/µl (T4 RNA Ligase 2), 10 U/µl (T4 DNA Ligase)
Total	**10 µl**	--

Table 19 Composition of ligation mixtures for investigating the optimal ligation time in the repair of gap by the ligation approach. The subtotal 1 should be chilled on ice until the temperature is equilibrated before the next step. The subtotal 2 should be incubated at 4℃ for 30 minutes before the next step. (See Section 4.15.2 for the details of the procedure)

Reagent	Amount to add	Final concentration in the 10 µl reaction
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	2 µl	1X
5-P Tetramer N (1000 µM)	2 µl	200 µM
UltraPure™ Distilled Water (Life Technologies)	4.3 µl	--
Subtotal 1	**8.3 µl**	--
Subtotal 1	8.3 µl	--
Annealing reaction mixture from Table 17	0.2 µl	1 µM (each oligo)
Subtotal 2	**8.5 µl**	--
Subtotal 2	8.5 µl	--
T4 RNA Ligase 2 (10 U/µl, NEB)(1/3X diluted)	1.5 µl	0.5 U/µl (T4 RNA Ligase 2)
Total	**10 µl**	--

Table 20 Composition of reverse transcription reaction mixtures for evaluating whether candidate polymerases can deal with successive 5-nitroindole.

Reagent	Amount to add	Final concentration
5-NitInd RT temp 1	0.3 µl	2.4 µM
5-NitInd RT primer 1	0.3 µl	2.4 µM
RT 10X Buffer (supplied with *Tth* DNA Polymerase, Promega), 10X Isothermal Amplification Buffer II (supplied with *Bst* 3.0 DNA Polymerase, NEB), 5X NEBNext® Quick Ligation Reaction Buffer (NEB), CutSmart® Buffer (10X, NEB)	2.5 µl (for 5X buffers) 1.25 µl (for 10X buffers)	1X
dNTPs (10 mM)	0.625-2.5 µl	0.5-2 mM
$MnCl_2$ (100 mM)	0-0.625 µl	0-5 mM
Tth DNA Polymerase (5U/ µl, Promega, M2101) MMLV (200U/ µl, Promega, M1705) MMLV RNaseH-, Point Mutant (200U/ µl, Promega, M3681) MMLV (200U/ µl, Invitrogen, 28025-021) SuperScript II (200U/ µl, Invitrogen, 18064-014) SuperScript III (200U/ µl, Invitrogen, 18080-044) Transcriptor High Fidelity Reverse Transcriptase (20U/ µl, supplied with Transcriptor High Fidelity cDNA Synthesis Kit, Roche, 05081963001) Cloned AMV Reverse Transcriptase (15U/ µl, Invitrogen, 12328-019) AMV Reverse Transcriptase (25U/ µl, USB, 70041Y) *Bst* 3.0 DNA Polymerase (8U/ µl, NEB, M0374S)	1 µl	0.4-16U/µl
UltraPure™ Distilled Water (Life Technologies)	5.27-9.02 µl	--
Total	**12.5 µl**	--

Table 21 Composition of reverse transcription reaction mixtures for further investigating the real situation in converting 5-nitroindole containing RNA template to cDNA.

Reagent	Amount to add	Final concentration
5-NitInd RT temp 2	0.3 µl	2.4 µM
5-NitInd RT primer 2	0.3 µl	2.4 µM
5X NEBNext® Quick Ligation Reaction Buffer (NEB)	2.5 µl	1X
dNTPs (10 mM)	2.5 µl	2 mM
$MnCl_2$ (100 mM)	1.25 µl	10 mM
Bst 3.0 DNA Polymerase (8U/ µl, NEB, M0374S)	1 µl	0.64U/µl
UltraPure™ Distilled Water (Life Technologies)	4.65 µl	--
Total	**12.5 µl**	--

References

1. Waddington CH: **The Epigenotype**. *Endeavour* 1942, **1**:18-20.

2. Riggs AD, Russo VE, Martienssen RA: **Epigenetic mechanisms of gene regulation**. New York: Cold Spring Harbor Laboratory Press; 1996.

3. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A: **An operational definition of epigenetics**. *Genes & Development* 2009, **23**(7):781-783.

4. Holliday R: **Epigenetics comes of age in the twentyfirst century**. *Journal of Genetics* 2002, **81**(1):1-4.

5. Finch JT, Klug A: **Solenoidal model for superstructure in chromatin**. *Proceedings of the National Academy of Sciences* 1976, **73**(6):1897-1901.

6. Woodcock CL, Frado LL, Rattner JB: **The higher-order structure of chromatin: evidence for a helical ribbon arrangement**. *The Journal of Cell Biology* 1984, **99**(1):42-52.

7. Ghirlando R, Felsenfeld G: **Hydrodynamic Studies on Defined Heterochromatin Fragments Support a 30-nm Fiber Having Six Nucleosomes per Turn**. *Journal of Molecular Biology* 2008, **376**(5):1417-1425.

8. Mateos-Langerak J, Bohn M, de Leeuw W, Giromus O, Manders EMM, Verschure PJ, Indemans MHG, Gierman HJ, Heermann DW, van Driel R *et al*: **Spatially confined folding of chromatin in the interphase nucleus**. *Proceedings of the National Academy of Sciences* 2009, **106**(10):3812-3817.

9. Nekrasov M, Amrichova J, Parker BJ, Soboleva TA, Jack C, Williams R, Huttley GA, Tremethick DJ: **Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics**. *Nat Struct Mol Biol* 2012, **19**(11):1076-1083.

10. Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J: **Organization of the Mitotic Chromosome**. *Science* 2013, **342**(6161):948.

11. Holliday R: **Epigenetics: An overview**. *Developmental Genetics* 1994, **15**(6):453-457.

12. Bird A: **Perceptions of epigenetics**. *Nature* 2007, **447**(7143):396-398.

13. **NIH Roadmap Epigenomics Project** [http://www.roadmapepigenomics.org/overview]

14. Liu D, Diorio J, Tannenbaum B, Caldji C, Francis D, Freedman A, Sharma S, Pearson D, Plotsky PM, Meaney MJ: **Maternal Care, Hippocampal Glucocorticoid Receptors,**

and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science* 1997, **277**(5332):1659.

15. Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, Meaney MJ: **Epigenetic programming by maternal behavior**. *Nat Neurosci* 2004, **7**(8):847-854.

16. de Kloet ER, Joels M, Holsboer F: **Stress and the brain: from adaptation to disease**. *Nat Rev Neurosci* 2005, **6**(6):463-475.

17. Ridder S, Chourbaji S, Hellweg R, Urani A, Zacher C, Schmid W, Zink M, H 繹 rtnagl H, Flor H, Henn FA *et al*: **Mice with Genetically Altered Glucocorticoid Receptor Expression Show Altered Sensitivity for Stress-Induced Depressive Reactions**. *The Journal of Neuroscience* 2005, **25**(26):6243.

18. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M, Turecki G, Meaney MJ: **Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse**. *Nat Neurosci* 2009, **12**(3):342-348.

19. Vanyushin BF, Nemirovsky LE, Klimenko VV, Vasiliev VK, Belozersky AN: **The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents**. *Gerontology* 1973, **19**(3):138-152.

20. Wilson VL, Smith RA, Ma S, Cutler RG: **Genomic 5-methyldeoxycytidine decreases with age**. *Journal of Biological Chemistry* 1987, **262**(21):9948-9951.

21. Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, Miura K, Miyazaki T, Ogura C, Okazaki Y, Jinno Y: **Age Related Changes in 5-methylcytosine Content in Human Peripheral Leukocytes and Placentas: an HPLC-based Study**. *Annals of Human Genetics* 2004, **68**(3):196-204.

22. Bjornsson HT, Sigurdsson MI, Fallin M, et al.: **Intra-individual change over time in dna methylation with familial clustering**. *JAMA* 2008, **299**(24):2877-2883.

23. Jung M, Pfeifer GP: **Aging and DNA methylation**. *BMC Biology* 2015, **13**:7.

24. Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Bürkle A, Caiafa P: **Reconfiguration of DNA methylation in aging**. *Mechanisms of Ageing and Development* 2015, **151**:60-70.

25. Waki T, Tamura G, Sato M, Motoyama T: **Age-related methylation of tumor suppressor and tumor-related genes: an analysis of autopsy samples**. *Oncogene* 2003, **22**(26):4128-4133.

26. So K, Tamura G, Honda T, Homma N, Waki T, Togawa N, Nishizuka S, Motoyama T: **Multiple tumor suppressor genes are increasingly methylated with age in non-neoplastic gastric epithelia**. *Cancer Science* 2006, **97**(11):1155-1158.

27. Esteller M: **Epigenetics in Cancer**. *New England Journal of Medicine* 2008, **358**(11):1148-1159.

28. Simó-Riudalbas L, Esteller M: **Cancer genomics identifies disrupted epigenetic genes**. *Human Genetics* 2014, **133**(6):713-725.

29. Gurdon JB: **The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles**. *Journal of Embryology and Experimental Morphology* 1962, **10**(4):622.

30. Gurdon JB, Uehlinger V: **"Fertile" Intestine Nuclei**. *Nature* 1966, **210**(5042):1240-1241.

31. Gurdon JB: **The Egg and the Nucleus: A Battle for Supremacy**. *Rambam Maimonides Medical Journal* 2015, **6**(3):e0023.

32. Jurkowska RZ, Jurkowski TP, Jeltsch A: **Structure and Function of Mammalian DNA Methyltransferases**. *ChemBioChem* 2010, **12**(2):206-222.

33. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M *et al*: **Human DNA methylomes at base resolution show widespread epigenomic differences**. *Nature* 2009, **462**(7271):315-322.

34. Meehan RR, Lewis JD, Bird AP: **Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA**. *Nucleic Acids Research* 1992, **20**(19):5085-5092.

35. Valinluck V, Tsai H-H, Rogstad DK, Burdzy A, Bird A, Sowers LC: **Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2)**. *Nucleic Acids Research* 2004, **32**(14):4100-4108.

36. Dawson MA, Kouzarides T: **Cancer Epigenetics: From Mechanism to Therapy**. *Cell* 2012, **150**(1):12-27.

37. Li B, Carey M, Workman JL: **The Role of Chromatin during Transcription**. *Cell* 2007, **128**(4):707-719.

38. Hargreaves DC, Crabtree GR: **ATP-dependent chromatin remodeling: genetics, genomics and mechanisms**. *Cell Res* 2011, **21**(3):396-420.

39. Zaret KS, Carroll JS: **Pioneer transcription factors: establishing competence for gene expression**. *Genes & Development* 2011, **25**(21):2227-2241.

40. Flynn RA, Chang HY: **Active chromatin and noncoding RNAs: an intimate relationship**. *Current Opinion in Genetics & Development* 2012, **22**(2):172-178.

41. Krol J, Loedige I, Filipowicz W: **The widespread regulation of microRNA biogenesis, function and decay**. *Nat Rev Genet* 2010, **11**(9):597-610.

42. Black DL: **Mechanisms of Alternative Pre-Messenger RNA Splicing**. *Annual Review of Biochemistry* 2003, **72**(1):291-336.

43. Liu J, Jia G: **Methylation modifications in eukaryotic messenger RNA**. *Journal of Genetics and Genomics* 2014, **41**(1):21-33.

44. Aravind L, Burroughs AM, Zhang D, Iyer LM: **Protein and DNA modifications: Evolutionary imprints of bacterial biochemical diversification and geochemistry on the provenance of eukaryotic epigenetics**. *Cold Spring Harbor Perspectives in Biology* 2014, **6**(7).

45. Saluz HP, Jiricny J, Jost JP: **Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis**. *Proceedings of the National Academy of Sciences of the United States of America* 1986, **83**(19):7167-7171.

46. Brenner S, Miller JH: **Encyclopedia of Genetics**, vol. 2: Academic Press; 2002.

47. Eddy SR: **Non-coding RNA genes and the modern RNA world**. *Nat Rev Genet* 2001, **2**(12):919-929.

48. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: **Molecular biology of the cell**, 4 edn. New York: Garland Science; 2002.

49. Turner P, McLennan A, Bates A, White M: **Molecular biology**, 3 edn. Liverpool: Taylor & Francis Group; 2005.

50. Russell J, Zomerdijk JCBM: **The RNA polymerase I transcription machinery**. *Biochemical Society symposium* 2006(73):203-216.

51. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN: **MicroRNA genes are transcribed by RNA polymerase II**. *The EMBO Journal* 2004, **23**(20):4051-4060.

52. Sims III RJ, Mandal SS, Reinberg D: **Recent highlights of RNA-polymerase-II-mediated transcription**. *Current Opinion in Cell Biology* 2004, **16**(3):263-271.

53. Willis IM: **RNA polymerase III**. *European Journal of Biochemistry* 1993, **212**(1):1-11.

54. Borchert GM, Lanier W, Davidson BL: **RNA polymerase III transcribes human microRNAs**. *Nature Structural Molecular Biology* 2006, **13**(12):1097-1101.

55. Anamika K, Gyenis À, Tora L: **How to stop: The mysterious links among RNA polymerase II occupancy 3' of genes, mRNA 3' processing and termination**. *Transcription* 2013, **4**(1):7-12.

56. Aguzzi A: **Unraveling prion strains with cell biology and organic chemistry**. *Proceedings of the National Academy of Sciences* 2008, **105**(1):11-12.

57. Hofmann J, Vorberg I: **Life cycle of cytosolic prions**. *Prion* 2013, **7**(5):369-377.

58. Dobson CM: **Protein folding and misfolding**. *Nature* 2003, **426**(6968):884-890.

59. Apirion D, Miczak A: **RNA processing in prokaryotic cells**. *BioEssays* 1993, **15**(2):113-120.

60. Clancy S: **RNA splicing: introns, exons and spliceosome**. In., vol. 1: Nature Education; 2008: 31.

61. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T: **Epigenetics in Alternative Pre-mRNA Splicing**. *Cell* 2011, **144**(1):16-26.

62. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigó R: **Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs**. *Genome Research* 2012, **22**(9):1616-1625.

63. Matera AG, Wang Z: **A day in the life of the spliceosome**. *Nat Rev Mol Cell Biol* 2014, **15**(2):108-121.

64. Cech TR: **Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena**. *Bioscience Reports* 1990, **10**(3):239.

65. Paushkin SV, Patel M, Furia BS, Peltz SW, Trotta CR: **Identification of a human endonuclease complex reveals a link between tRNA splicing and Pre-mRNA 3' end formation**. *Cell* 2004, **117**(3):311-321.

66. Sun W-J, Li J-H, Liu S, Wu J, Zhou H, Qu L-H, Yang J-H: **RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data**. *Nucleic Acids Research* 2015, **44**(D1):D259-D265.

67. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM *et al*: **MODOMICS: a database of RNA modification pathways--2013 update**. *Nucleic Acids Research* 2013, **41**(D1):D262-D267.

68. Shatkin AJ: **Capping of eucaryotic mRNAs**. *Cell* 1976, **9**(4):645-653.

69. Visa N, Izaurralde E, Ferreira J, Daneholt B, Mattaj IW: **A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the**

ribonucleoprotein particle during nuclear export. *The Journal of Cell Biology* 1996, **133**(1):5.

70. Konarska MM, Padgett RA, Sharp PA: **Recognition of cap structure in splicing in vitro of mRNA precursors**. *Cell* 1984, **38**(3):731-736.

71. Karijolich J, Yu Y-T: **Converting nonsense codons into sense codons by targeted pseudouridylation**. *Nature* 2011, **474**(7351):395-398.

72. Zhao X, Yang Y, Sun B-F, Shi Y, Yang X, Xiao W, Hao Y-J, Ping X-L, Chen Y-S, Wang W-J *et al*: **FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis**. *Cell Res* 2014, **24**(12):1403-1419.

73. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G *et al*: **N6-methyladenosine-dependent regulation of messenger RNA stability**. *Nature* 2014, **505**(7481):117-120.

74. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C: **N6-methyladenosine modulates messenger RNA translation efficiency**. *Cell* 2015, **161**(6):1388-1399.

75. Alarcon CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF: **N6-methyladenosine marks primary microRNAs for processing**. *Nature* 2015, **519**(7544):482-485.

76. Agris PF, Waldo EC, Klvle M: **The Importance of Being Modified: Roles of Modified Nucleosides and Mg2+ in RNA Structure and Function**. In: *Progress in Nucleic Acid Research and Molecular Biology.* vol. Volume 53: Academic Press; 1996: 79-129.

77. Basti MM, Stuart JW, Lam AT, Guenther R, Agris PF: **Design, biological activity and NMR-solution structure of a DNA analogue of yeast tRNAPhe anticodon domain**. *Nat Struct Mol Biol* 1996, **3**(1):38-44.

78. Yarian CS, Basti MM, Cain RJ, Ansari G, Guenther RH, Sochacka E, Czerwinska G, Malkiewicz A, Agris PF: **Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39**. *Nucleic Acids Research* 1999, **27**(17):3543-3549.

79. Khoury GA, Baliban RC, Floudas CA: **Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database**. *Scientific Reports* 2011, **1**:90.

80. Varki A: **Essentials of glycobiology**, 2 edn: Cold Spring Harbor Laboratory Press; 2009.

81. Crispin M, Doores KJ: **Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design**. *Current opinion in virology* 2015, **11**:63-69.

82. Maston GA, Evans SK, Green MR: **Transcriptional Regulatory Elements in the Human Genome**. *Annual Review of Genomics and Human Genetics* 2006, **7**(1):29-59.

83. van Kouwenhove M, Kedde M, Agami R: **MicroRNA regulation by RNA-binding proteins and its implications for cancer**. *Nat Rev Cancer* 2011, **11**(9):644-656.

84. Lee RC, Feinbaum RL, Ambros V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14**. *Cell* 1993, **75**(5):843-854.

85. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: **The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans**. *Nature* 2000, **403**(6772):901-906.

86. Chen C-Y, Gherzi R, Ong S-E, Chan EL, Raijmakers R, Pruijn GJM, Stoecklin G, Moroni C, Mann M, Karin M: **AU Binding Proteins Recruit the Exosome to Degrade ARE-Containing mRNAs**. *Cell* 2001, **107**(4):451-464.

87. Isken O, Maquat LE: **Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function**. *Genes & Development* 2007, **21**(15):1833-3856.

88. Łabno A, Tomecki R, Dziembowski A: **Cytoplasmic RNA decay pathways - Enzymes and mechanisms**. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2016, **1863**(12):3125-3147.

89. Wilkinson KD: **The discovery of ubiquitin-dependent proteolysis**. *Proceedings of the National Academy of Sciences* 2005, **102**(43):15280-15282.

90. Pimentel-Muiños FX, Boada-Romero E: **Selective autophagy against membranous compartments**. *Autophagy* 2014, **10**(3):397-407.

91. Xu H, Ren D: **Lysosomal Physiology**. *Annual Review of Physiology* 2015, **77**(1):57-80.

92. Ambros V: **microRNAs: Tiny Regulators with Great Potential**. *Cell* 2001, **107**(7):823-826.

93. He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation**. *Nature Review Genetics* 2004, **5**(7):522-531.

94. Vasudevan S, Tong Y, Steitz JA: **Switching from repression to activation: microRNAs can up-regulate translation**. *Science* 2007, **318**(5858):1931-1934.

95. Wightman B, Ha I, Ruvkun G: **Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans**. *Cell* 1993, **75**(5):855-862.

96. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P *et al*: **Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA**. *Nature* 2000, **408**(6808):86-89.

97. Bernstein E, Caudy AA, Hammond SM, Hannon GJ: **Role for a bidentate ribonuclease in the initiation step of RNA interference**. *Nature* 2001, **409**(6818):363-366.

98. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD: **A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA**. *Science* 2001, **293**(5531):834-838.

99. Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J: **Rapid and simple method for purification of nucleic acids**. *Journal of Clinical Microbiology* 1990, **28**(3):495-503.

100. Elbashir SM, Lendeckel W, Tuschl T: **RNA interference is mediated by 21- and 22-nucleotide RNAs**. *Genes & Development* 2001, **15**(2):188-200.

101. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: **Identification of Novel Genes Coding for Small Expressed RNAs**. *Science* 2001, **294**(5543):853.

102. Lau NC, Lim LP, Weinstein EG, Bartel DP: **An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans**. *Science* 2001, **294**(5543):858.

103. Lee RC, Ambros V: **An extensive class of small RNAs in Caenorhabditis elegans**. *Science* 2001, **294**(5543):862-864.

104. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E *et al*: **Identification of hundreds of conserved and nonconserved human microRNAs**. *Nat Genet* 2005, **37**(7):766-770.

105. Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA, Sjoblom T, Barad O, Bentwich Z, Szafranska AE, Labourier E *et al*: **The colorectal microRNAome**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(10):3687-3692.

106. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-L, Zhao Y, McDonald H, Zeng T, Hirst M *et al*: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells**. *Genome Research* 2008, **18**(4):610-621.

107. Stark MS, Tyagi S, Nancarrow DJ, Boyle GM, Cook AL, Whiteman DC, Parsons PG, Schmidt C, Sturm RA, Hayward NK: **Characterization of the melanoma miRNAome by deep sequencing**. *PLoS ONE* 2010, **5**(3):e9685.

108. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M *et al*: **A mammalian microRNA expression atlas based on small RNA library sequencing**. *Cell* 2007, **129**(7):1401-1414.

109. Lui W-O, Pourmand N, Patterson BK, Fire A: **Patterns of known and novel small RNAs in human cervical cancer**. *Cancer Research* 2007, **67**(13):6031.

110. Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, Loman N, Jonsson G, Naya H, Hoglund M *et al*: **Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene**. *Cancer Research* 2011, **71**(1):78.

111. Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P, Kaessmann H: **Birth and expression evolution of mammalian microRNA genes**. *Genome Research* 2013, **23**(1):34-45.

112. Salvi A, Abeni E, Portolani N, Barlati S, De Petro G: **Human hepatocellular carcinoma cell-specific miRNAs reveal the differential expression of miR-24 and miR-27a in cirrhotic/non-cirrhotic HCC**. *International Journal of Oncology* 2013, **42**(2):391-402.

113. Wang H-j, Zhang P-j, Chen W-j, Jie D, Dan F, Jia Y-h, Xie L-x: **Characterization and identification of novel serum microRNAs in sepsis patients with different outcomes**. *Shock* 2013, **39**(6):480-487.

114. Pfeffer S, Zavolan M, Grässer FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C *et al*: **Identification of virus-encoded microRNAs**. *Science* 2004, **304**(5671):734.

115. Amen MA, Griffiths A: **Identification and Expression Analysis of Herpes B Virus-Encoded Small RNAs**. *Journal of Virology* 2011, **85**(14):7296-7311.

116. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR: **Rfam: an RNA family database**. *Nucleic Acids Research* 2003, **31**(1):439-441.

117. Griffiths-Jones S: **The microRNA Registry**. *Nucleic Acids Research* 2004, **32**(suppl 1):D109-D111.

118. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature**. *Nucleic Acids Research* 2006, **34**(suppl 1):D140-D144.

119. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics**. *Nucleic Acids Research* 2008, **36**(suppl 1):D154-D158.

120. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data**. *Nucleic Acids Research* 2011, **39**(suppl_1):D152-D157.

121. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence microRNAs using deep sequencing data**. *Nucleic Acids Research* 2014, **42**(D1):D68-D73.

122. Lee Y, Jeon K, Lee JT, Kim S, Kim VN: **MicroRNA maturation: stepwise processing and subcellular localization**. *The EMBO Journal* 2002, **21**(17):4663-4670.

123. Cai X, Hagedorn CH, Cullen BR: **Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs**. *RNA* 2004, **10**(12):1957-1966.

124. Carthew RW, Sontheimer EJ: **Origins and mechanisms of miRNAs and siRNAs**. *Cell* 2009, **136**(4):642-655.

125. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S *et al*: **The nuclear RNase III Drosha initiates microRNA processing**. *Nature* 2003, **425**(6956):415-419.

126. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U: **Nuclear Export of MicroRNA Precursors**. *Science* 2004, **303**(5654):95-98.

127. Yi R, Qin Y, Macara IG, Cullen BR: **Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs**. *Genes & Development* 2003, **17**(24):3011-3016.

128. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ: **Argonaute2, a link between genetic and biochemical analyses of RNAi**. *Science* 2001, **293**(5532):1146.

129. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G: **miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs**. *Genes & Development* 2002, **16**(6):720-728.

130. Diederichs S, Haber DA: **Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression**. *Cell* 2007, **131**(6):1097-1108.

131. Yoda M, Kawamata T, Paroo Z, Ye X, Iwasaki S, Liu Q, Tomari Y: **ATP-dependent human RISC assembly pathways**. *Nat Struct Mol Biol* 2010, **17**(1):17-23.

132. Lingel A, Simon B, Izaurralde E, Sattler M: **Nucleic acid 3[prime]-end recognition by the Argonaute2 PAZ domain**. *Nat Struct Mol Biol* 2004, **11**(6):576-577.

133. Elkayam E, Kuhn C-D, Tocilj A, Haase AD, Greene EM, Hannon GJ, Joshua-Tor L: **The structure of human argonaute-2 in complex with miR-20a**. *Cell* 2012, **150**(1):100-110.

134. Chi SW, Zang JB, Mele A, Darnell RB: **Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps**. *Nature* 2009, **460**(7254):479-486.

135. Friedman RC, Farh KK-H, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs**. *Genome Research* 2009, **19**(1):92-105.

136. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V: **Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation**. *Genome Biology* 2004, **5**(3):R13.

137. Kloosterman WP, Wienholds E, de Bruijn E, Kauppinen S, Plasterk RHA: **In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes**. *Nat Meth* 2006, **3**(1):27-29.

138. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, Schmittgen TD: **Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors**. *RNA* 2008, **14**(1):35-42.

139. Saini HK, Griffiths-Jones S, Enright AJ: **Genomic analysis of human microRNA transcripts**. *Proceedings of the National Academy of Sciences* 2007, **104**(45):17719-17724.

140. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T: **New microRNAs from mouse and human**. *RNA* 2003, **9**.

141. Baskerville S, Bartel DP: **Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes**. *RNA* 2005, **11**(3):241-247.

142. Stormo GD: **Gene-Finding Approaches for Eukaryotes**. *Genome Research* 2000, **10**(4):394-397.

143. Eilers M, Eisenman RN: **Myc's broad reach**. *Genes & Development* 2008, **22**(20):2755-2766.

144. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT: **c-Myc-regulated microRNAs modulate E2F1 expression**. *Nature* 2005, **435**(7043):839-843.

145. Tagawa H, Karube K, Tsuzuki S, Ohshima K, Seto M: **Synergistic action of the microRNA-17 polycistron and Myc in aggressive cancer development**. *Cancer Science* 2007, **98**(9):1482-1490.

146. Chang T-C, Yu D, Lee Y-S, Wentzel EA, Arking DE, West KM, Dang CV, Thomas-Tikhonenko A, Mendell JT: **Widespread microRNA repression by Myc contributes to tumorigenesis**. *Nat Genet* 2008, **40**(1):43-50.

147. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D *et al*: **A microRNA component of the p53 tumour suppressor network**. *Nature* 2007, **447**(7148):1130-1134.

148. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: **Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs**. *Nature* 2005, **433**(7027):769-773.

149. Conaco C, Otto S, Han J-J, Mandel G: **Reciprocal actions of REST and a microRNA promote neuronal identity**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(7):2422-2427.

150. Davis-Dusenbery BN, Hata A: **Mechanisms of control of microRNA biogenesis**. *The Journal of Biochemistry* 2010, **148**(4):381-392.

151. Lujambio A, Esteller M: **How epigenetics can explain human metastasis: A new role for microRNAs**. *Cell Cycle* 2009, **8**(3):377-382.

152. Lujambio A, Calin GA, Villanueva A, Ropero S, Sánchez-Céspedes M, Blanco D, Montuenga LM, Rossi S, Nicoloso MS, Faller WJ *et al*: **A microRNA DNA methylation signature for human cancer metastasis**. *Proceedings of the National Academy of Sciences* 2008, **105**(36):13556-13561.

153. Scott GK, Mattie MD, Berger CE, Benz SC, Benz CC: **Rapid Alteration of MicroRNA Levels by Histone Deacetylase Inhibition**. *Cancer Research* 2006, **66**(3):1277.

154. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, Dohner K, Bentz M, Lichter P: **Genomic aberrations and survival in chronic lymphocytic leukemia**. *New England Journal of Medicine* 2000, **343**(26):1910-1916.

155. Glassman AB, Hayes KJ: **The value of fluorescence in situ hybridization in the diagnosis and prognosis of chronic lymphocytic leukemia**. *Cancer Genetics and Cytogenetics* 2005, **158**(1):88-91.

156. Xu W, Li JY, Pan JL, Qiu HR, Shen YF, Li L, Wu YF, Xue YQ: **Interphase fluorescence in situ hybridization detection of cytogenetic abnormalities in B-cell chronic lymphocytic leukemia**. *International Journal of Hematology* 2007, **85**(5):430-436.

157. Qiu HX, Xu W, Cao XS, Zhou M, Shen YF, Xu YL, Sun XM, Liu Q, Wang R, Qiu HR *et al*: **Cytogenetic characterisation in Chinese patients with chronic lymphocytic leukemia: A prospective, multicenter study on 143 cases analysed with interphase fluorescence in situ hybridisation**. *Leukemia and Lymphoma* 2008, **49**(10):1887-1892.

158. Lai Y, Huang X: **Cytogenetic characteristics of B cell chronic lymphocytic leukemia in 275 Chinese patients by fluorescence in situ hybridization: a multicenter study.** *Chinese Medical Journal* 2011, **124**(16):2417-2422.

159. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K *et al*: **Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia**. *Proceedings of the National Academy of Sciences* 2002, **99**(24):15524-15529.

160. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, Wojcik SE, Aqeilan RI, Zupo S, Dono M *et al*: **miR-15 and miR-16 induce apoptosis by targeting BCL2**. *Proceedings of the National Academy of Sciences* 2005, **102**(39):13944-13949.

161. Ouillette P, Erba H, Kujawski L, Kaminski M, Shedden K, Malek SN: **Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14**. *Cancer Research* 2008, **68**(4):1012-1021.

162. Zhu DX, Miao KR, Fang C, Fan L, Zhu W, Zhu HY, Zhuang Y, Hong M, Liu P, Xu W *et al*: **Aberrant microRNA expression in Chinese patients with chronic lymphocytic leukemia**. *Leukemia Research* 2011, **35**(6):730-734.

163. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio MV, Visone R, Sever NI, Fabbri M *et al*: **A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia**. *New England Journal of Medicine* 2005, **353**(17):1793-1801.

164. Raveche ES, Salerno E, Scaglione BJ, Manohar V, Abbasi F, Lin Y-C, Fredrickson T, Landgraf P, Ramachandra S, Huppi K *et al*: **Abnormal microRNA-16 locus with synteny to human 13q14 linked to CLL in NZB mice**. *Blood* 2007, **109**(12):5079-5086.

165. Zamore PD: **Thirty-Three Years Later, a Glimpse at the Ribonuclease III Active Site**. *Molecular Cell* 2001, **8**(6):1158-1160.

166. Filippov V, Solovyev V, Filippova M, Gill SS: **A novel type of RNase III family proteins in eukaryotes**. *Gene* 2000, **245**(1):213-221.

167. Fortin KR, Nicholson RH, Nicholson AW: **Mouse ribonuclease III. cDNA structure, expression analysis, and chromosomal location**. *BMC Genomics* 2002, **3**(1):26.

168. Han J, Lee Y, Yeom K-H, Kim Y-K, Jin H, Kim VN: **The Drosha-DGCR8 complex in primary microRNA processing**. *Genes & Development* 2004, **18**(24):3016-3027.

169. Nguyen TA, Jo MH, Choi Y-G, Park J, Kwon SC, Hohng S, Kim VN, Woo J-S: **Functional anatomy of the human Microprocessor**. *Cell* 2015, **161**(6):1374-1387.

170. Kwon SC, Nguyen Tuan A, Choi Y-G, Jo Myung H, Hohng S, Kim VN, Woo J-S: **Structure of human DROSHA**. *Cell* 2016, **164**(1):81-90.

171. Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ: **Processing of primary microRNAs by the Microprocessor complex**. *Nature* 2004, **432**(7014):231-235.

172. Gregory RI, Yan K-p, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R: **The Microprocessor complex mediates the genesis of microRNAs**. *Nature* 2004, **432**(7014):235-240.

173. Landthaler M, Yalcin A, Tuschl T: **The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis**. *Current Biology* 2004, **14**(23):2162-2167.

174. Han J, Pedersen JS, Kwon SC, Belair CD, Kim Y-K, Yeom K-H, Yang W-Y, Haussler D, Blelloch R, Kim VN: **Posttranscriptional crossregulation between Drosha and DGCR8**. *Cell* 2009, **136**(1):75-84.

175. Triboulet R, Chang H-M, LaPierre RJ, Gregory RI: **Post-transcriptional control of DGCR8 expression by the Microprocessor**. *RNA* 2009, **15**(6):1005-1011.

176. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: **Many roads to maturity: microRNA biogenesis pathways and their regulation**. *Nat Cell Biol* 2009, **11**(3):228-234.

177. Viswanathan SR, Daley GQ, Gregory RI: **Selective blockade of microRNA processing by Lin28**. *Science* 2008, **320**(5872):97.

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

179. Roush S, Slack FJ: **The let-7 family of microRNAs**. *Trends in Cell Biology* 2008, **18**(10):505-516.

180. Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, Wulczyn FG: **A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment**. *Nat Cell Biol* 2008, **10**(8):987-993.

181. Attisano L, Wrana JL: **Signal transduction by the TGF-β superfamily**. *Science* 2002, **296**(5573):1646.

182. Davis BN, Hilyard AC, Nguyen PH, Lagna G, Hata A: **Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha**. *Molecular Cell* 2010, **39**(3):373-384.

183. Davis BN, Hilyard AC, Lagna G, Hata A: **SMAD proteins control DROSHA-mediated microRNA maturation**. *Nature* 2008, **454**(7200):56-61.

184. Wada T, Kikuchi J, Furukawa Y: **Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8**. *EMBO reports* 2012, **13**(2):142.

185. Tang X, Li M, Tucker L, Ramratnam B: **Glycogen synthase kinase 3 beta (GSK3β) phosphorylates the RNAase III enzyme Drosha at S300 and S302**. *PLoS ONE* 2011, **6**(6):e20391.

186. Herbert KM, Pimienta G, DeGregorio SJ, Alexandrov A, Steitz JA: **Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile**. *Cell Reports* 2013, **5**(4):1070-1081.

187. Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W: **Single processing center models for human Dicer and bacterial RNase III**. *Cell* 2004, **118**(1):57-68.

188. Tian Y, Simanshu DK, Ma J-B, Park J-E, Heo I, Kim VN, Patel DJ: **A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer**. *Molecular Cell* 2014, **53**(4):606-616.

189. MacRae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA: **Structural Basis for Double-Stranded RNA Processing by Dicer**. *Science* 2006, **311**(5758):195.

190. Lau P-W, Guiley KZ, De N, Potter CS, Carragher B, MacRae IJ: **The molecular architecture of human Dicer**. *Nat Struct Mol Biol* 2012, **19**(4):436-440.

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

192. Tsutsumi A, Kawamata T, Izumi N, Seitz H, Tomari Y: **Recognition of the pre-miRNA structure by Drosophila Dicer-1**. *Nat Struct Mol Biol* 2011, **18**(10):1153-1158.

193. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R: **TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing**. *Nature* 2005, **436**(7051):740-744.

194. Haase AD, Jaskiewicz L, Zhang H, Lainé S, Sack R, Gatignol A, Filipowicz W: **TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing**. *EMBO reports* 2005, **6**(10):961-967.

195. Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN: **The role of PACT in the RNA silencing pathway**. *The EMBO Journal* 2006, **25**(3):522.

196. Lau P-W, Potter CS, Carragher B, MacRae IJ: **Structure of the human Dicer-TRBP complex by electron microscopy**. *Structure* 2009, **17**(10):1326-1332.

197. Wilson Ross C, Tambe A, Kidwell Mary A, Noland Cameron L, Schneider Catherine P, Doudna Jennifer A: **Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis**. *Molecular Cell* 2015, **57**(3):397-407.

198. Saito K, Ishizuka A, Siomi H, Siomi MC: **Processing of pre-microRNAs by the Dicer-1–Loquacious complex in Drosophila cells**. *PLoS Biology* 2005, **3**(7):e235.

199. Förstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD: **Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein**. *PLoS Biology* 2005, **3**(7):e236.

200. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R: **Human RISC couples microRNA biogenesis and posttranscriptional gene silencing**. *Cell* 2005, **123**(4):631-640.

201. Kim Y, Yeo J, Lee Jung H, Cho J, Seo D, Kim J-S, Kim VN: **Deletion of human tarbp2 reveals cellular microRNA targets and cell-cycle function of TRBP**. *Cell Reports* 2014, **9**(3):1061-1074.

202. Betancur JG, Tomari Y: **Dicer is dispensable for asymmetric RISC loading in mammals**. *RNA* 2012, **18**(1):24-30.

203. Fareh M, Yeom K-H, Haagsma AC, Chauhan S, Heo I, Joo C: **TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments**. *Nature Communications* 2016, **7**:13694.

204. Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD: **Asymmetry in the assembly of the RNAi enzyme complex**. *Cell* 2003, **115**(2):199-208.

205. Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD: **A protein sensor for siRNA asymmetry**. *Science* 2004, **306**(5700):1377.

206. Khvorova A, Reynolds A, Jayasena SD: **Functional siRNAs and miRNAs exhibit strand bias**. *Cell* 2003, **115**(2):209-216.

207. Makarova KS, Wolf YI, van der Oost J, Koonin EV: **Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements**. *Biology Direct* 2009, **4**(1):29.

208. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA: **Bacterial Argonaute samples the transcriptome to identify foreign DNA**. *Molecular Cell* 2013, **51**(5):594-605.

209. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJJ *et al*: **DNA-guided DNA interference by a prokaryotic Argonaute**. *Nature* 2014, **507**(7491):258-261.

210. Song J-J, Smith SK, Hannon GJ, Joshua-Tor L: **Crystal structure of Argonaute and its implications for RISC slicer activity**. *Science* 2004, **305**(5689):1434.

211. Yuan Y-R, Pei Y, Ma J-B, Kuryavyi V, Zhadina M, Meister G, Chen H-Y, Dauter Z, Tuschl T, Patel DJ: **Crystal structure of A. aeolicus Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage**. *Molecular Cell* 2005, **19**(3):405-419.

212. Ma J-B, Yuan Y-R, Meister G, Pei Y, Tuschl T, Patel DJ: **Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein**. *Nature* 2005, **434**(7033):666-670.

213. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ: **Structure of the guide-strand-containing argonaute silencing complex**. *Nature* 2008, **456**(7219):209-213.

214. Schirle NT, MacRae IJ: **The Crystal Structure of Human Argonaute2**. *Science* 2012, **336**(6084):1037-1040.

215. Frank F, Sonenberg N, Nagar B: **Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2**. *Nature* 2010, **465**(7299):818-822.

216. Sasaki T, Shiohama A, Minoshima S, Shimizu N: **Identification of eight members of the Argonaute family in the human genome**. *Genomics* 2003, **82**(3):323-330.

217. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T: **Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs**. *Molecular Cell* 2004, **15**(2):185-197.

218. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song J-J, Hammond SM, Joshua-Tor L, Hannon GJ: **Argonaute2 Is the Catalytic Engine of Mammalian RNAi**. *Science* 2004, **305**(5689):1437-1441.

219. Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ: **Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes**. *Nature* 2009, **461**(7265):754-761.

220. Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ: **Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex**. *Nature* 2008, **456**(7224):921-926.

221. Bartel DP: **MicroRNAs**. *Cell* 2004, **116**(2):281-297.

222. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets**. *Cell* 2005, **120**(1):15-20.

223. Liu X, Jiang F, Kalidas S, Smith D, Liu Q: **Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes**. *RNA* 2006, **12**(8):1514-1520.

224. Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, Wang X: **R2D2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway**. *Science* 2003, **301**(5641):1921.

225. Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q: **Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila**. *Genes & Development* 2005, **19**(14):1674-1679.

226. Kawamata T, Seitz H, Tomari Y: **Structural determinants of miRNAs for RISC loading and slicer-independent unwinding**. *Nat Struct Mol Biol* 2009, **16**(9):953-960.

227. Martinez J, Patkaniowska A, Urlaub H, Lührmann R, Tuschl T: **Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi**. *Cell* 2002, **110**(5):563-574.

228. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y: **Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes**. *Molecular Cell* 2010, **39**(2):292-299.

229. Kawamata T, Tomari Y: **Making RISC**. *Trends in Biochemical Sciences* 2010, **35**(7):368-376.

230. Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M: **In vitro assembly of plant RNA-Induced silencing complexes facilitated by molecular chaperone HSP90**. *Molecular Cell* 2010, **39**(2):282-291.

231. Landthaler M: **Chaperones get RISC loaded**. *Molecular Cell* 2010, **39**(2):161-162.

232. Okamura K, Ishizuka A, Siomi H, Siomi MC: **Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways**. *Genes & Development* 2004, **18**(14):1655-1666.

233. Leuschner PJF, Ameres SL, Kueng S, Martinez J: **Cleavage of the siRNA passenger strand during RISC assembly in human cells**. *EMBO reports* 2006, **7**(3):314.

234. Gu S, Jin L, Zhang F, Huang Y, Grimm D, Rossi JJ, Kay MA: **Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes**. *Proceedings of the National Academy of Sciences* 2011, **108**(22):9208-9213.

235. Kwak PB, Tomari Y: **The N domain of Argonaute drives duplex unwinding during RISC assembly**. *Nat Struct Mol Biol* 2012, **19**(2):145-151.

236. Schirle NT, Sheu-Gruttadauria J, MacRae IJ: **Structural basis for microRNA targeting**. *Science* 2014, **346**(6209):608.

237. Park JH, Shin C: **Slicer-independent mechanism drives small-RNA strand separation during human RISC assembly**. *Nucleic Acids Research* 2015, **43**(19):9418-9433.

238. Gu S, Jin L, Huang Y, Zhang F, Kay MA: **Slicing-independent RISC activation requires the Argonaute PAZ domain**. *Current Biology* 2012, **22**(16):1536-1542.

239. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H: **Drosophila endogenous small RNAs bind to Argonaute2 in somatic cells**. *Nature* 2008, **453**(7196):793-797.

240. Miyoshi K, Miyoshi T, Hartig JV, Siomi H, Siomi MC: **Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila**. *RNA* 2010, **16**(3):506-515.

241. Eulalio A, Huntzinger E, Izaurralde E: **GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay**. *Nat Struct Mol Biol* 2008, **15**(4):346-353.

242. Eystathioy T, Chan EKL, Tenenbaum SA, Keene JD, Griffith K, Fritzler MJ: **A Phosphorylated Cytoplasmic Autoantigen, GW182, Associates with a Unique Population of Human mRNAs within Novel Cytoplasmic Speckles**. *Molecular Biology of the Cell* 2002, **13**(4):1338-1351.

243. Schneider MD, Najand N, Chaker S, Pare JM, Haskins J, Hughes SC, Hobman TC, Locke J, Simmonds AJ: **Gawky is a component of cytoplasmic mRNA processing bodies required for early Drosophila development**. *The Journal of Cell Biology* 2006, **174**(3):349.

244. Pfaff J, Meister G: **Argonaute and GW182 proteins: an effective alliance in gene silencing**. *Biochemical Society Transactions* 2013, **41**(4):855.

245. Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E: **mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes**. *Genes & Development* 2006, **20**(14):1885-1898.

246. Yang Z, Jakymiw A, Wood MR, Eystathioy T, Rubin RL, Fritzler MJ, Chan EKL: **GW182 is critical for the stability of GW bodies expressed during the cell cycle and cell proliferation**. *Journal of Cell Science* 2004, **117**(23):5567.

247. Gibbings DJ, Ciaudo C, Erhardt M, Voinnet O: **Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity**. *Nat Cell Biol* 2009, **11**(9):1143-1149.

248. Hilbert M, Karow Anne R, Klostermeier D: **The mechanism of ATP-dependent RNA unwinding by DEAD box proteins**. *Biological Chemistry* 2009, **390**:1237-1250.

249. Chu C-y, Rana TM: **Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54**. *PLOS Biology* 2006, **4**(7):e210.

250. Robb GB, Rana TM: **RNA Helicase A interacts with RISC in human cells and functions in RISC loading**. *Molecular Cell* 2007, **26**(4):523-537.

251. Meister G, Landthaler M, Peters L, Chen PY, Urlaub H, Lührmann R, Tuschl T: **Identification of novel Argonaute-associated proteins**. *Current Biology* 2005, **15**(23):2149-2155.

252. Hammell CM, Lubin I, Boag PR, Blackwell TK, Ambros V: **nhl-2 Modulates MicroRNA Activity in Caenorhabditis elegans**. *Cell* 2009, **136**(5):926-938.

253. Neumuller RA, Betschinger J, Fischer A, Bushati N, Poernbacher I, Mechtler K, Cohen SM, Knoblich JA: **Mei-P26 regulates microRNAs and cell growth in the Drosophila ovarian stem cell lineage**. *Nature* 2008, **454**(7201):241-245.

254. Schwamborn JC, Berezikov E, Knoblich JA: **The TRIM-NHL protein TRIM32 activates microRNAs and prevents self-renewal in mouse neural progenitors**. *Cell* 2009, **136**(5):913-925.

255. Weinmann L, Höck J, Ivacevic T, Ohrt T, Mütze J, Schwille P, Kremmer E, Benes V, Urlaub H, Meister G: **Importin 8 is a gene silencing factor that targets Argonaute proteins to distinct mRNAs**. *Cell* 2009, **136**(3):496-507.

256. van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN: **Control of stress-dependent cardiac growth and gene expression by a microRNA**. *Science* 2007, **316**(5824):575.

257. Gatfield D, Le Martelot G, Vejnar CE, Gerlach D, Schaad O, Fleury-Olela F, Ruskeepää A-L, Oresic M, Esau CC, Zdobnov EM *et al*: **Integration of microRNA miR-122 in hepatic circadian gene expression**. *Genes & Development* 2009, **23**(11):1313-1326.

258. Krol J, Busskamp V, Markiewicz I, Stadler MB, Ribi S, Richter J, Duebel J, Bicker S, Fehling HJ, Schübeler D *et al*: **Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs**. *Cell* 2010, **141**(4):618-631.

259. Hwang H-W, Wentzel EA, Mendell JT: **A hexanucleotide element directs microRNA nuclear import**. *Science* 2007, **315**(5808):97.

260. Buck AH, Perot J, Chisholm MA, Kumar DS, Tuddenham L, Cognat V, Marcinowski L, Dölken L, Pfeffer S: **Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection**. *RNA* 2010, **16**(2):307-315.

261. Ibrahim F, Rymarquis LA, Kim E-J, Becker J, Balassa E, Green PJ, Cerutti H: **Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas**. *Proceedings of the National Academy of Sciences* 2010, **107**(8):3906-3911.

262. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, Wolf DA, Mizgerd JP: **Zcchc11-dependent uridylation of microRNA directs cytokine expression**. *Nat Cell Biol* 2009, **11**(9):1157-1163.

263. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S-i, Baba T, Suzuki T: **Selective stabilization of mammalian microRNAs by 3??adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2**. *Genes & Development* 2009, **23**(4):433-438.

264. Lu S, Sun Y-H, Chiang VL: **Adenylation of plant miRNAs**. *Nucleic Acids Research* 2009, **37**(6):1878-1885.

265. Li J, Yang Z, Yu B, Liu J, Chen X: **Methylation protects miRNAs and siRNAs from a 3' end uridylation activity in Arabidopsis**. *Current Biology* 2005, **15**(16):1501-1507.

266. Chatterjee S, Groszhans H: **Active turnover modulates mature microRNA activity in Caenorhabditis elegans**. *Nature* 2009, **461**(7263):546-549.

267. Ramachandran V, Chen X: **Degradation of microRNAs by a family of exoribonucleases in Arabidopsis**. *Science* 2008, **321**(5895):1490.

268. Chiorazzi N, Rai KR, Ferrarini M: **Chronic lymphocytic leukemia**. *New England Journal of Medicine* 2005, **352**(8):804-815.

269. Corcoran MM, Hammarsund M, Zhu C, Lerner M, Kapanadze B, Wilson B, Larsson C, Forsberg L, Ibbotson RE, Einhorn S *et al*: **DLEU2 encodes an antisense RNA for the putative bicistronic RFP2/LEU5 gene in humans and mouse**. *Genes, Chromosomes and Cancer* 2004, **40**(4):285-297.

270. Kalachikov S, Migliazza A, Cayanis E, Fracchiolla NS, Bonaldo MF, Lawton L, Jelenc P, Ye X, Qu X, Chien M *et al*: **Cloning and gene mapping of the chromosome 13q14 region deleted in chronic lymphocytic leukemia**. *Genomics* 1997, **42**(3):369-377.

271. Liu Y, Corcoran M, Rasool O, Ivanova G, Ibbotson R, Grandér D, Iyengar A, Baranova A, Kashuba V, Merup M *et al*: **Cloning of two candidate tumor suppressor genes within a 10 kb region on chromosome 13q14, frequently deleted in chronic lymphocytic leukemia**. *Oncogene* 1997, **15**(20):2463-2473.

272. Corcoran MM, Rasool O, Liu Y, Iyengar A, Grander D, Ibbotson RE, Merup M, Wu X, Brodyansky V, Gardiner AC *et al*: **Detailed molecular delineation of 13q14.3 loss in B-cell chronic lymphocytic leukemia**. *Blood* 1998, **91**(4):1382-1390.

273. Stilgenbauer S, Nickolenko J, Wilhelm J, Wolf S, Weitz S, Döhner K, Boehm T, Döhner H, Lichter P: **Expressed sequences as candidates for a novel tumor suppressor gene at band 13q14 in B-cell chronic lymphocytic leukemia and mantle cell lymphoma**. *Oncogene* 1998, **16**(14):1891-1897.

274. Bullrich F, Fujii H, Calin G, Mabuchi H, Negrini M, Pekarsky Y, Rassenti L, Alder H, Reed JC, Keating MJ *et al*: **Characterization of the 13q14 tumor suppressor locus in CLL: identification of ALT1, an alternative splice variant of the LEU2 gene**. *Cancer Research* 2001, **61**(18):6640-6648.

275. Migliazza A, Bosch F, Komatsu H, Cayanis E, Martinotti S, Toniato E, Guccione E, Qu X, Chien M, Murty VVV *et al*: **Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia**. *Blood* 2001, **97**(7):2098-2104.

276. Fabbri M, Bottoni A, Shimizu M, Spizzo R, Nicoloso MS, Rossi S, Barbarotto E, Cimmino A, Adair B, Wojcik SE *et al*: **Association of a microRNA/TP53 feedback circuitry with pathogenesis and outcome of B-cell chronic lymphocytic leukemia**. *JAMA* 2011, **305**(1):59-67.

277. Fulci V, Chiaretti S, Goldoni M, Azzalin G, Carucci N, Tavolaro S, Castellano L, Magrelli A, Citarella F, Messina M *et al*: **Quantitative technologies establish a novel microRNA profile of chronic lymphocytic leukemia**. *Blood* 2007, **109**(11):4944-4951.

278. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, Ambesi-Impiombato A, Califano A, Migliazza A, Bhagat G *et al*: **The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia**. *Cancer Cell* 2010, **17**(1):28-40.

279. Hanlon K, Rudin CE, Harries LW: **Investigating the targets of MIR-15a and MIR-16-1 in patients with chronic lymphocytic leukemia (CLL)**. *PLoS ONE* 2009, **4**(9):e7169.

280. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW: **Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans**. *Nat Struct Mol Biol* 2010, **17**(2):173-179.

281. Bracht JR, Van Wynsberghe PM, Mondol V, Pasquinelli AE: **Regulation of lin-4 miRNA expression, organismal growth and development by a conserved RNA binding protein in C. elegans**. *Developmental biology* 2010, **348**(2):210-221.

282. Lewis BP, Shih Ih, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets**. *Cell* 2003, **115**(7):787-798.

283. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M *et al*: **Combinatorial microRNA target predictions**. *Nat Genet* 2005, **37**(5):495-500.

284. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing**. *Molecular Cell* 2007, **27**(1):91-105.

285. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP: **Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs**. *Nat Struct Mol Biol* 2011, **18**(10):1139-1146.

286. Nam J-W, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, Yildirim MA, Rodriguez A, Bartel DP: **Global analyses of the effect of different cellular contexts on microRNA targeting**. *Molecular Cell* 2014, **53**(6):1031-1043.

287. Agarwal V, Bell GW, Nam J-W, Bartel DP: **Predicting effective microRNA target sites in mammalian mRNAs**. *eLife* 2015, **4**:e05005.

288. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: **Human microRNA targets**. *PLOS Biology* 2004, **2**(11):e363.

289. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: **The role of site accessibility in microRNA target recognition**. *Nat Genet* 2007, **39**(10):1278-1284.

290. Miranda KC, Huynh T, Tay Y, Ang Y-S, Tam W-L, Thomson AM, Lim B, Rigoutsos I: **A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes**. *Cell* 2006, **126**(6):1203-1217.

291. Calin GA, Cimmino A, Fabbri M, Ferracin M, Wojcik SE, Shimizu M, Taccioli C, Zanesi N, Garzon R, Aqeilan RI *et al*: **MiR-15a and miR-16-1 cluster functions in human leukemia**. *Proceedings of the National Academy of Sciences* 2008, **105**(13):5166-5171.

292. Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, Chowdhury D, Dykxhoorn DM, Tsai P, Hofmann O *et al*: **miR-24 Inhibits Cell Proliferation by Targeting E2F2, MYC, and Other Cell-Cycle Genes via Binding to ?倪eedless??3?涸TR MicroRNA Recognition Elements**. *Molecular Cell* 2009, **35**(5):610-625.

293. Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR, Johnson JM, Cummins JM, Raymond CK, Dai H *et al*: **Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression**. *Molecular and Cellular Biology* 2007, **27**(6):2240-2252.

294. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.

295. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano Jr M, Jungkamp A-C, Munschauer M *et al*: **Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP**. *Cell* 2010, **141**(1):129-141.

296. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M *et al*: **PAR-CliP - A Method to Identify Transcriptome-wide the Binding Sites of RNA Binding Proteins**. 2010(41):e2034.

297. König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J: **iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution**. *Nat Struct Mol Biol* 2010, **17**(7):909-915.

298. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J: **iCLIP - Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution**. *JoVE* 2011(50):e2638.

299. Helwak A, Kudla G, Dudnakova T, Tollervey D: **Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding**. *Cell* 2013, **153**(3):654-665.

300. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D: **Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast**. *Proceedings of the National Academy of Sciences* 2011, **108**(24):10010-10015.

301. Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing Chromosome Conformation**. *Science* 2002, **295**(5558):1306.

302. Nguyen TC, Cao X, Yu P, Xiao S, Lu J, Biase FH, Sridhar B, Huang N, Zhang K, Zhong S: **Mapping RNA-RNA interactome and RNA structure in vivo by MARIO**. *Nature Communications* 2016, **7**:12023.

303. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, Davidovich C, Gooding AR, Goodrich KJ, Mattick JS *et al*: **RNA duplex map in living cells reveals higher-order transcriptome structure**. *Cell* 2016, **165**(5):1267-1279.

304. **Immunoprecipitation (IP) Protocol**. In.: abcam.

305. Stacchini A, Aragno M, Vallario A, Alfarano A, Circosta P, Gottardi D, Faldella A, Rege-Cambrin G, Thunberg U, Nilsson K *et al*: **MEC1 and MEC2: two new cell lines derived from B-chronic lymphocytic leukaemia in prolymphocytoid transformation**. *Leukemia Research* 1999, **23**(2):127-136.

306. Lozzio CB, Lozzio BB: **Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome**. *Blood* 1975, **45**(3):321-334.

307. Pulvertaft RJV: **Cytology of burkitt's tumour (african lymphoma)**. *The Lancet* 1964, **283**(7327):238-240.

308. Ikeda K, Satoh M, Pauley KM, Fritzler MJ, Reeves WH, Chan EKL: **Detection of the argonaute protein Ago2 and microRNAs in the RNA induced silencing complex (RISC) using a monoclonal antibody**. *Journal of Immunological Methods* 2006, **317**(1-2):38-44.

309. Ma J-B, Ye K, Patel DJ: **Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain**. *Nature* 2004, **429**(6989):318-322.

310. Rushizky GW, Knight CA, Sober HA: **Studies on the Preferential Specificity of Pancreatic Ribonuclease as Deduced from Partial Digests**. *Journal of Biological Chemistry* 1961, **236**(10):2732-2737.

311. Spahr PF, Hollingworth BR: **Purification and Mechanism of Action of Ribonuclease from Escherichia coli Ribosomes**. *Journal of Biological Chemistry* 1961, **236**(3):823-831.

312. Boyer PD: **The Enzymes**, vol. XV, 3 edn: ELSEVIER; 1982.

313. Cuchillo CM, Nogués MV, Raines RT: **Bovine Pancreatic Ribonuclease: 50 Years of the First Enzymatic Reaction Mechanism()**. *Biochemistry* 2011, **50**(37):7835-7841.

314. England TE, Uhlenbeck OC: **Enzymic oligoribonucleotide synthesis with T4 RNA ligase**. *Biochemistry* 1978, **17**(11):2069-2076.

315. Das U, Shuman S: **Mechanism of RNA 2', 3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase**. *Nucleic Acids Research* 2013, **41**(1):355-365.

316. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, Van Eyk JE: **Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a**

Bottom-Up Proteomics Workflow. In: *Current Protocols in Molecular Biology.* John Wiley & Sons, Inc.; 2001.

317. Pähler A, Banerjee A, Dattagupta JK, Fujiwara T, Lindner K, Pal GP, Suck D, Weber G, Saenger W: **Three-dimensional structure of fungal proteinase K reveals similarity to bacterial subtilisin**. *The EMBO Journal* 1984, **3**(6):1311-1314.

318. Bajorath J, Saenger W, Pal GP: **Autolysis and inhibition of proteinase K, a subtilisin-related serine proteinase isolated from the fungus Tritirachium album Limber**. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1988, **954**:176-182.

319. Murphy FV, Ramakrishnan V: **Structure of a purine-purine wobble base pair in the decoding center of the ribosome**. *Nat Struct Mol Biol* 2004, **11**(12):1251-1252.

320. Loakes D, Brown DM: **5-Nitroindole as an universal base analogue**. *Nucleic Acids Research* 1994, **22**(20):4039-4043.

321. Loakes D, Brown DM, Linde S, Hill F: **3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR**. *Nucleic Acids Research* 1995, **23**(13):2361-2366.

322. Harlow E, Lane D: **Using Antibodies - A Laboratory Manual**. New York: Cold Spring Harbor Laboratory Press; 1999.

323. Carmell MA, Xuan Z, Zhang MQ, Hannon GJ: **The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis**. *Genes & Development* 2002, **16**(21):2733-2742.

324. Boyer PD: **The Enzymes**, vol. 14, 3rd edn: Academic Press; 1982.

325. Lesnik EA, Freier SM: **Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure**. *Biochemistry* 1995, **34**(34):10807-10815.

326. Caccone A, DeSalle R, Powell JR: **Calibration of the change in thermal stability of DNA duplexes and degree of base pair mismatch**. *Journal of Molecular Evolution* 1988, **27**(3):212-216.

327. Holley RW, Everett GA, Madison JT, Zamir A: **Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid**. *Journal of Biological Chemistry* 1965, **240**(5):2122-2128.

328. Gerber AP, Keller W: **An Adenosine Deaminase that Generates Inosine at the Wobble Position of tRNAs**. *Science* 1999, **286**(5442):1146.

329. Watkins JNE, SantaLucia JJ: **Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes**. *Nucleic Acids Research* 2005, **33**(19):6258-6267.

330. Silber R, Malathi VG, Hurwitz J: **Purification and Properties of Bacteriophage T4-Induced RNA Ligase**. *Proceedings of the National Academy of Sciences of the United States of America* 1972, **69**(10):3009-3013.

331. Bergstrom DE, Zhang P, Johnson WT: **Comparison of the base pairing properties of a series of nitroazole nucleobase analogs in the oligodeoxyribonucleotide sequence 5'-d(CGCXAATTYGCG)-3'**. *Nucleic Acids Research* 1997, **25**(10):1935-1942.