CITY UNIVERSITY OF HONG KONG
Department of Chemistry

BSc in Applied Biology
Project Report

Optimization of 3' adapter ligation for labelling RNA

by
FENG Hengxin

May 2020

Optimization of 3' adapter ligation for labelling RNA

by

Name of Student: FENG Hengxin

Name of Supervisor: KWOK Chun Kit

Submitted in partial fulfilment of the

requirements for the degree of

BACHELOR OF SCIENCE IN APPLIED BIOLOGY

from

City University of Hong Kong

May 2020

Abstract

Fluorescent labelling of RNA is important for detecting and quantifying RNA. RNA with fluorophore added on either 5’ or 3’ end can be obtained by chemical synthesis but with a limitation of length. In this work, we demonstrate a labelling method based on ligation. A fluorescent labelled ssDNA adapter is ligated to the 3’ terminus of the RNA to give the fluorescent signal. We identify better conditions for the ligation. Factors including adapter types, PEG concentration, enzyme dosage, reaction time and RNA to adapter ratio were optimized. The ligation yield can reach to more than 60% under optimized conditions. Notably, this ligation based fluorescent labelling method would not significantly affect the biological properties of RNA, e.g. protein binding affinity, which provides a potential way for long RNA fluorescent labelling.

Table of Content

List of Figures

List of Tables

List of Abbreviations

A	Adenine
APP	Amyloid precursor protein
C	Cytosine
DNA	Deoxyribose nucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
FAM	Fluorescein amidites
G	Guanine
HCl	Hydrogen chloride
IR	Infrared
KCl	Potassium chloride
K_D	Equilibrium dissociation constant
LiCl	Lithium chloride
$MgCl_2$	Magnesium chloride
miRNA	microRNA
mRNA	Messenger RNA
MST	Microscale thermophoresis
ncRNA	Non-coding RNA
nt	Nucleotide
OH	Hydroxyl group
P	Phosphate group
PEG	Polyethylene glycol
piRNA	Piwi-interacting RNA
rG4	RNA G-quadruplex
RNA	Ribose nucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA

RTS	Transcriptase stalling assay
siRNA	Small interfering RNA
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
Tris	Trisaminomethane
tRNA	Transfer RNA
U	Uracil
WT	Wild type

ACKNOWLEDGMENTS

This project would not be accomplished without the help and support from Chem Department in City University of Hong Kong, professors, technical staff, my family and friends.

It was my fortune to have Dr. Chun Kit Kwok as my supervisor. You led me to a new world of science and research. I have learned so much from you. Maybe you would be surprised that I did not like molecular biology that much before being a member of Kwok Lab. You are the person who let me know the beauty of RNA and want to explore more in this field. I have been encouraged by you for a lot of time. In this year, I was faced with pressure from different aspects. Without your encouragements and considerations, I would not be able to handle all the things and enjoy my research and study under such a stressful environment. You let me know that I have the potential to be better and I am capable for more than I thought. Learning from you, I know how to manage time better and work more effectively. I have been encouraged by you to communicate more with others in our lab, which turned out to be so helpful with my research and study. Besides, I also learned from you that staying positive and working hard is the solution to any difficulties. Many thanks to you for leading me into the world of research.

Being a member of Kwok lab is the most correct choice I have made in my undergraduate study. There are so much help, support and encouragement I have received from our lab members.

I have to especially thank Ms. Jieyu Zhao and Ms. Xi Mou for mentoring and supporting me during the whole research project. You not only taught me how to do experiments but also offered me so much help when I was faced with poor results. Thank you so much for answering my question patiently and doing side-by-side experiments with me when I was confused by specific techniques.

I have to thank Ms. Kaixin Lyu, Ms. Danyang Ji, Mr. Mubarak Ishaq Umar for providing me advice when I faced with the difficulties in my experiments. Thank you so much for helping and supporting me. Ms. Hei Yuen Wong, Phoebe, you may not know how much support and encouragement I got from you. Your smile brought me relief in a busy day.

I would not forget our alumni members in Kwok lab. Ms. Pui Yan Yeung, you were the person who guided me from a total freshman to research. The two months that I worked with

you was an important period for building my interest in research and molecular biology. Mr. Chun Yin Chan, thank you for helping me with a lot of technical issues.

I should thank my friends, especially my roommates and my boyfriend for offering me help and support during this stressful and difficult period.

Last but not least, I have to thank my parents for understanding my decisions and supporting me with all their effort. I shortened my holiday and the time I stay with you during this year because of my research but you always said that you are with me and it is great to work in research. I have to thank my father, you are the person who told me about CityU and wrote the start of the story. I have to thank my mother, you always support me without any doubt.

1. Introduction

1.1 Background Knowledge and Biological Significance of RNA

Ribonucleic acid (RNA) is a polymeric molecule composed of nucleotide monomers. Each nucleotide consists of a phosphate group, a nitrogen-containing base, and a ribose, a five-carbon sugar. There are four types of nitrogenous bases in RNA, including adenine (A), guanine (G), cytosine (C) and uracil (U). Unlike the double helix structure of DNA, RNA usually exists in a single-stranded form. Different secondary structures were observed in RNA molecules, including hairpin, bulge, internal and multibranch loops[1] as well as RNA G-quadruplex (rG4)[2]. Secondary structures of RNA always involve in the interaction between RNA and other molecules such as proteins[3]. RNA can be sorted as encoding, i.e. messenger RNA (mRNA), and non-coding RNA. According to Central Dogma[4], mRNA is transcribed from DNA and encodes for proteins. A non-coding RNA (ncRNA) refers to an RNA with no or low potential of protein-coding, which can account for around 97% transcriptional output from the eukaryotic genome[5]. Non-coding RNAs include infrastructural RNAs involving in translation, such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs)[6], small RNAs (~18 to 30 nucleotides), e.g. microRNA (miRNA), small interfering RNA (siRNA), Piwi-interacting RNA (piRNA)[7], as well as long non-coding RNAs (>200 nucleotides). Many ncRNAs have been proved to play important roles in regulating the genetic information flowing from DNA to protein. Besides tRNA and rRNA that directly involve in translation, other ncRNAs participate in regulation of DNA replication[8], transcription[9], alternative splicing of pre-mRNA[10], post-transcriptional gene silencing[11] etc. To perform the proper function, an RNA may interact with other molecules such as DNAs, other RNAs and/or proteins. The interaction could be achieved by complementary base-pairing between nucleic acids or binding between RNA secondary structure and proteins[3]. The *in vivo* interaction between RNA and other molecules is complicated, therefore simplified *in vitro* models are always built up for detecting and analysing the interaction and studying the function of certain RNA.

71.2 Fluorescent Labelling of RNA

To detect the interaction between RNA and other molecules, the prerequisite is to visualize the RNA of interest. Radiolabelling and fluorescent labelling are widely applied in visualizing nucleic acids *in vitro*.[12] Compared to radiolabelling, fluorescent labelling is less sensitive, but it can perform real-time detection with high resolution and lacks radiation hazards[13]. The non-specific nucleic acid fluorescent labelling can be performed chemically by attaching fluorophores to the nucleic acids during or after gel electrophoresis, i.e. gel staining. One fluorescent dye molecule is introduced to each RNA, which allows researchers to quantify RNA by intensity.[13] Some problems may be caused by the chemically staining gel, such as uneven staining, gel to gel staining variation and strong background noise, which could generate troubles in quantification[14]. Chemical synthesis of RNA provides alternative methods for RNA fluorescent labelling. Fluorophores such as fluorescein amidites (FAM) and cyanine can be added to 5' or 3' terminal of RNA by covalently bonding in oligonucleotide synthesis. It is commercially available and the sequence of RNA can be designed. Using the fluorescent labelling RNA can eliminate the problems caused by staining and is with higher sensitivity[15]. However, it is more expensive compared to the chemical fluorescent staining since the cost of chemical oligonucleotide synthesis counts on the number of bases and the addition of fluorophores has extra cost. Moreover, there is a limitation in length of chemically synthesized RNA, therefore it could not be applied to the study of long RNAs.

1.3 Detecting protein-nucleic acid interaction *in vitro*: EMSA and MST methods

As mentioned previously, *in vitro* model of RNA-protein interaction has been widely used for detecting RNA-protein binding and study the biological function of RNAs. The electrophoretic mobility shift assay (EMSA) is a widely applied method to detect protein-nucleic acid interaction, which has been originally described in 1981[16,17]. This method is based on the observation that the mobilities of protein-free nucleic acid and protein-nucleic acid complex during electrophoresis are different. The gel used in EMSA is the native gel. Different from the denaturing electrophoresis that denatures nucleic acids into linear form and fractionates them mainly by size, electrophoresis with native gel retains the structural features of the nucleic acids and protein-nucleic acid complex. The bands in native gel would be separated by size, electronegativity, and shape of nucleic acid and protein-nucleic acid

complex[18]. To visualize the nucleic acid used for EMSA, radiolabelling, fluorescent labelling or fluorescent staining after gel electrophoresis can be applied. However, there are limitations to EMSA method. One major limitation is that samples are not in the free solution during electrophoresis, which means that the complex are not at chemical equilibrium during the electrophoresis step and the interaction may differ a lot from the *in vivo* situation[18]. Besides, unexpected disturbance may occur during casting and handling the gel e.g. smear binding band and unclear gel, which causes difficulties in gel analysis.

Microscale thermophoresis (MST) is a technology for analysing the interaction between biomolecules that has been developed and commercially available in the past decade. As the name of this technology indicates, it is based on thermophoresis, which means the directed movement of molecules in a gradient of temperature. The temperature difference results in a depletion of the solvated biomolecules with the elevating temperature. The interface between the biomolecules and the solvent is the critical factor for the thermophoretic depletion, which means that the depletion is significantly different between the target-ligand binding complex and unbound target [19]. A fluorescent labelled target with a constant concentration, e.g. RNA, and a non-fluorescent labelled ligand, e.g. protein, with serially diluted concentration are required for performing MST. In a typical MST instrument, infrared (IR) laser is used for heating. The change of sample fluorescence intensity in the focal IR-laser zone is recorded from the initial state (before laser heating) to the steady state, which indicates the thermophoretic depletion. By comparing the change of fluorescence intensity with the ligand concentration, the binding affinity can be calculated. [20] Compared to EMSA, MST detects the interaction in a free solution and requires less handling time. However, the cost of performing MST is higher than EMSA.

1.4 T4 RNA Ligase

T4 RNA ligase is an enzyme forming a 3’ → 5’ phosphodiester bond between the 3’-hydroxyl and the 5’-phosphate oligoribonucleotide with the hydrolysis of ATP to AMP and PPi, which leads to the circulation of the oligoribonucleotide. In addition to the intramolecular ligation, T4 RNA ligase is able to ligate a donor ssRNA (with 5’-phosphate) to an acceptor ssRNA (with 3’-hydroxyl). The mechanism of the intermolecular ligation catalysed by T4 RNA

ligase includes three steps. The first step is the adenylation of enzyme, during which the AMP from ATP hydrolysis is added to the enzyme and PPi is released. The second step is the activation of donor molecule, in which the adenyl group is transferred from the adenylated enzyme to the 5' terminal of the donor molecule. The third step is the formation of 3' → 5' phosphodiester bond between the donor and the accepter, with the release of the AMP.[21] Moreover, T4 RNA ligase can be used for ssDNA intra/intermolecular ligation, although with less efficiency than using RNA substrate[22]. Being able to catalyse intermolecular ligation, T4 RNA ligase shows the potential of adding modifications to RNA by ligation.

1.5 Research objectives

In this research, we want to demonstrate that a synthetic fluorescent labelled ssDNA adapter with 5'-phosphate can be ligated to an RNA molecule with 3'-hydroxyl by T4 RNA ligase 1. With the ligation reaction, the fluorescent labelling can be added to the RNA (Fig. 1). Different factors are tested with a model RNA molecule for optimization of the ligation reaction. Factors to be tested include adapter types, PEG 8000 concentration, T4 RNA ligase 1 dosage, reaction time and RNA: adapter ratio. With the optimized conditions, we plan to perform the ligation reaction on an RNA molecule other than the model RNA to demonstrate that the ligated adapter would not affect its biological properties, e.g. binding affinity with proteins.

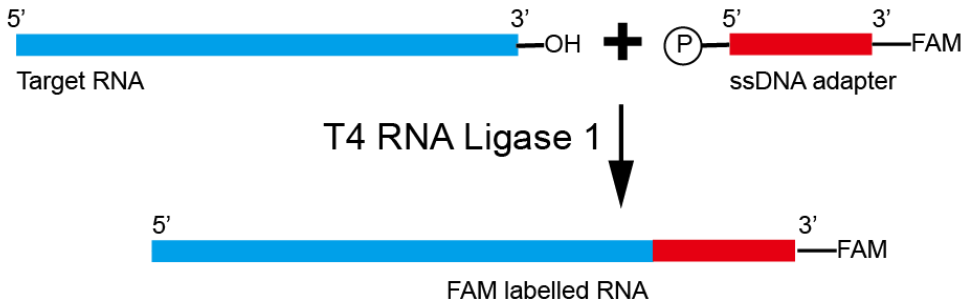

Figure 1. Schematic of 3' adapter ligation for labelling RNA. The ssDNA adapter (with 5' phosphate and 3' FAM group) is ligated to the 3' end of the target RNA (with 3' hydroxyl group) by T4 RNA Ligase 1.

2. Materials and Methods

2.1 Optimization of 3' adapter ligation using T4 RNA ligase 1

For optimization, reactions were performed with a total volume of 10 μL each. The standard reaction mixture consisted of 0.1 μM N_{40}-3'OH RNA, 3' FAM labelled adapter, 1x T4 Ligase Reaction Buffer (NEB, B0216L; provided a final concentration of 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT)), 1mM ATP, PEG8000, high concentration T4 RNA Ligase 1 (NEB M0437M,). Reaction was set up on ice and performed at 25 °C for 60 minutes following by 65 °C for 3 minutes in the thermocycler machine.

The optimizations of different factors were performed following the order of adapter selection (see Figure 2), PEG percentage optimization (see Figure 3), enzyme dosage optimization (see Figure 4), reaction time optimization (Figure 5), RNA to adapter ratio optimization (see Figure 6). Each reaction was performed with at least three replicates.

Adapter selection

Two different 20 nt 5' phosphoryl-terminated and 3' FAM-labelled DNA adapters, (5'p-N4A16-3' FAM and 5'p-A20-3' FAM), were designed for 3' adapter ligation with N_{40}-3'OH RNA. To compare the ligation rate of the two adapters, based on the standard condition, the final concentration of 5'p-N4A16-3' FAM or 5'p-A20-3'FAM adapters were 0.1, 0.25, 0.5 or 1 μM (each concentration performed for both adapters). N_{40}-3'OH RNA was at 0.1 μM final concentration. The final concentration of PEG8000 was 20% and the enzyme dosage for high concentration T4 RNA Ligase 1 (NEB, M0437M) was 1 μL (15 U final).

The 5'p-N4A16-3' FAM adapter showed higher ligation efficiency than 5'p-A20-3'FAM adapter. 1 μM 5'p-N4A16-3' FAM adapter was selected for further optimization since it showed the highest ligation rate in the comparison.

PEG percentage optimization

To optimize the PEG8000 concentration, reactions with PEG8000 final concentration (w/v) of 15%, 17.5%, 20%, 22.5%, 25% were performed. Reactions were performed based on the conditions and result of adapter selection.

The PEG8000 with a final concentration of 20% gave the highest ligation efficiency and was chosen for further optimization. Comparison of PEG8000, PEG6000 and PEG4000 was performed, and no significant difference was observed, therefore PEG8000 was used in subsequent reactions.

Enzyme dosage optimization

To optimize enzyme dosage is to find the dosage for both ligation and cost efficiency. 0.5, 1, 1.5, 2 μL of high concentration T4 RNA Ligase 1 (NEB M0437M), i. e. 15, 30, 45, 60 units of enzyme, was added. Reactions were performed based on the conditions and result of PEG percentage optimization mentioned above.

The highest ligation efficiency was shown in the reaction with 0.5 μL enzyme (15 units/10 μL reaction), which was also the most cost-effective one. The 0.5 μL enzyme dosage was selected for later reactions.

Reaction time optimization

To maximize the ligation efficiency, based on the conditions and result in enzyme dosage optimization, reactions with the incubation time of 30, 60, 90, 120 minutes at 25 °C were performed.

The difference between 90- and 120-minute reaction time was not significant, and 90 minutes was therefore used for further optimization.

RNA to adapter ratio optimization

To obtain higher ligation yield, the adapter concentration was further increased. Based on the conditions and result of reaction time optimization, reactions were performed with 1, 1.5, 2, 2.5 μM final concentration of 5'p-N4A16-3' FAM adapter, with fixed N_{40}-3'OH RNA at 0.1 μM final concentration. The highest adapter concentration showed the highest ligation yield and was selected for subsequent experiments.

All the listed reactions were terminated by adding equal volume of 2X formamide orange G dye (94% deionized formamide, orange G dye, 20 mM Tris pH 7.5, 20 mM EDTA). The 8M urea 10% polyacrylamide denaturing gel was used for electrophoresis of samples from each reaction. Samples were heated for 3 minutes at 95 °C and held at 60 °C before loading to the

pre-heated gel (300V for 15-20 minutes). 2 μL of each sample was loaded and the electrophoresis was performed at a constant voltage of 300V for 20 minutes. SYBR Gold (Thermo, S11494, 1:10000 dilution) was used to stain the gel. The stained gel was scanned by Bio-rad ChemiDoc ™ Touch Imaging System for further analysis.

2.2 Scaled up ligation and gel purification of APP rG4 motif + 11 nt

Ligation of the RNA, WT APP rG4 motif + 11 nt (abbreviated as APP below, see Table S1), was with the optimized conditions tested by the N_{40}-3'OH RNA. The reaction volume was scaled up to 100 μL with an input of 50 pmol APP RNA and 125 pmol adapter. The dye addition, electrophoresis and gel staining were the same as stated previously, with the only change of sample loading volume from 2 μL to 20 μL in each well.

The gel stained by SYBR Gold was viewed with a blue light transilluminator and the ligation product band was cut out from the gel with a razor blade. The gel slice was then transferred to a 2 mL nuclease-free microcentrifuge tube and crushed into small pieces. 320 μL of TE-Li-800 (containing 1X TE buffer, pH 7.4 and 0.8M LiCl) was added to the tube. The mixture was incubated in a thermoshaker at 4 °C overnight at 1300 rpm, without light exposure. After the overnight incubation, the liquid in the mixture was transferred to a centrifuge filter tube and centrifuged at 13, 500 rpm for 1 minute. RNA Clean & Concentrator ™ (Zymo Research) was used for purification of the flow-through. Column purification was performed by following manufacturer's protocol. The RNA was eluted with 15 μL nuclease-free water. The concentration of the ligated RNA was quantified by using RNA ScreenTape on the 4200 Tapestation (Agilent Technologies). The procedure followed the manufacturer's protocol.

2.3 EMSA for unligated and ligated APP RNA

Ten 10 μL reactions with peptide gradient were performed for EMSA. 11 μL of 40 nM APP RNA was heated at 75 °C for 3 minutes with the thermocycler and cooled to ambient temperature after heating to reform the RNA secondary structure. Serial dilution (1:2) was performed to the peptide, RHAU53, with the highest concentration of 2500 nM and lowest concentration of 9.77 nM. The volume of peptide and water was 5 μL in total. One in the ten reactions contained water only, served as a negative control, i.e. peptide free. 2 μL of 5x binding

buffer (containing 750 mM KCl, 5mM $MgCl_2$ and 125 mM Tris-HCl, pH=7.6) and 2 μL of 40% sucrose were added to each reaction respectively. 1 μL of the cooling down RNA was added to each reaction. The reaction mixture was mixed well by pipetting and centrifuged briefly to the bottom of the tube. The reaction mixtures were then incubated at 37 °C in the thermocycler for one hour. The 10% polyacrylamide native gel was used for electrophoresis, which was casted with 0.5x TBE, the running buffer for electrophoresis. 10 μL of sample was loaded to each well. Electrophoresis was performed with a constant current of 10 mA at 4 °C for 40 minutes without exposure to light. The gel after electrophoresis was stained by SYBR Gold (1:10000 dilution) and then scanned by Bio-rad ChemiDoc ™ Touch Imaging System for data analysis.

EMSA and electrophoresis procedure for ligated APP RNA was same as the one for unligated RNA, excepting using 12% polyacrylamide native gel rather than 10% gel. Instead of staining with SYBR Gold, the gel was scanned by FLA-9000 Starion (FujiFilm) directly.

2.4 MST for ligated APP RNA

17 μL of 200 nM ligated APP RNA was heated at 75 °C for 3 minutes with the thermocycler and cooled to ambient temperature after heating to reform the RNA secondary structure. Fifteen times of serial dilution (1:2) were performed to RHAU53 with the highest concentration of 7000 nM. 2 μL of 5x binding buffer (containing 750 mM KCl, 5mM $MgCl_2$ and 125 mM Tris-HCl, pH=7.6) was added to each reaction. 1 μL of the cooling down ligated APP RNA was added to each reaction, giving a final concentration of 20 nM RNA. The reaction mixtures were then incubated at 37 °C in the thermocycler for one hour. Each binding sample was loaded to a capillary (NanoTemper) and the machine, Monolith NT. 115 (NanoTemper) was used for analysis.

2.5 Data processing and analysis

The background was corrected for the unligated N_{40} RNA band (U) and ligated product band (L). The analysis was done by Image Lab ™ 6.0.1. The equation below is used for calculating the percentage ligation yield:

$$\text{Ligation yield} = \frac{U}{U + L} \times 100\%$$

The standard error of the mean value of at least 3 independent replicates is shown as the

error bar in the histograms. It is calculated from the equation below:

$$\text{Standard Error} = \frac{\text{Standard deviation}}{\sqrt{\text{Total number of samples}}}$$

The intensity of bound and unbound RNA in EMSA was analysed by ImageJ. The fraction of RNA bound was calculated from the intensity using the equation below:

$$\text{RNA fraction bound} = \text{Bound}/(\text{Bound} + \text{Unbound})$$

The RNA fraction bound and corresponding RHAU53 concentration were imported to Graph Pad Prism to estimate the equilibrium dissociation constant (K_D) with non-linear regression.

In MST, K_D was generated automatically by Monolith Analysis software.

3. Result and Discussion

3.1 Selection of adapter

The synthetic RNA oligonucleotide, N_{40} -3' OH (see Table S1), was used as a model RNA molecule for optimization of ligation reaction conditions. Two different ssDNA adapters, 5'p-N4A16-3' FAM and 5'p-A20-3' FAM, were designed for ligation reaction. The ligation efficiency of each adapter was tested at four different RNA: adapter ratio ranging from 1: 1 to 1: 10 under the same condition (see Methods). Generally, at each RNA: adapter ratio, 5'p-N4A16-3' FAM showed a higher ligation efficiency than 5'p-A20-3' FAM (Fig. 2B). Increasing ligation efficiency was observed with the decreasing RNA: adapter ratio for both 5'p-N4A16-3' FAM and 5'p-A20-3' FAM adapters (Fig. 2). We found that for 5'p-A20-3'FAM, the highest ligation efficiency was observed at the RNA: adapter ratio of 1: 10, which yielded 18±1% (Fig. 2A lane 5). For 5'p-N4A16-3'FAM, similarly, the highest ligation efficiency also occurred at the RNA: adapter ratio of 1: 10, while the ligation yield is 44±2%, which is higer than the 5'p-A20-3'FAM adapter. (Fig. 2A lane 10). Thus, we identified 5'p-N4A16-3'FAM showed better ligation efficiency than 5'p-A20-3'FAM with 1:10 of RNA: adapter ratio.

The ssDNA adapters were designed to consist mainly of adenine deoxyribonucleotide, which is less likely to form secondary structures and affect the interaction between the original

RNA and other molecules. The result indicated that the adapter with four random nucleotides at the 5' terminal performed better ligation, which may due to the reducing ligation bias by the random nucleotides as reported in previous study[23].

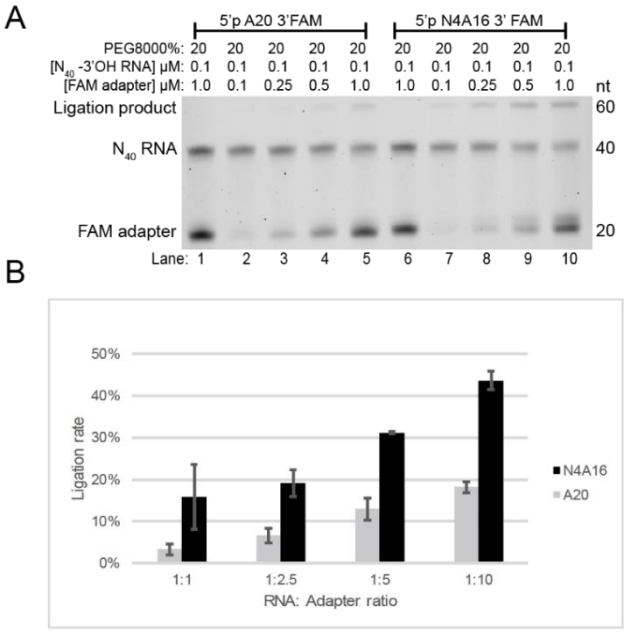

Figure 2. The effect of different adapters on the 3' adapter ligation. A) Ligations with RNA: adapter ratio from 1:1 to 1:10 were performed between N40-3'OH and 5' p-A20-3' FAM (lanes 1-5 or 5' p-N4A16-3' FAM (lanes 6-10) adapters. Negative controls were performed by adding nuclease-free water instead of enzyme (lane 1 and 6). For both adapters, an increasing ligation rate was observed with a higher adapter concentration. At each RNA: adapter ratio, a higher ligation efficiency using 5' p-N4A16-3' FAM adapter than 5' p-A20-3' FAM was observed. B) Histogram showing the gel result in A). The error bars represent the standard error of at least 3 independent replicates.

3.2 Optimization of PEG concentration

Previous study[24] has reported that PEG 8000 could enhance the ligation efficiency of T4 RNA ligase, therefore we reasoned that there might be an optimal concentration of PEG 8000 for ligation reaction. Here we assessed 2 factors including PEG concentration and PEG type. In a pre-test (Supplementary Fig. 1), we found that without PEG 8000 presence, no ligation was observed. The ligation product occurred from 10% PEG 8000. An increasing trend was shown from 10% to 20% PEG 8000, yet no peak of ligation efficiency was observed. Based on the pre-test, the PEG 8000 concentrations between 15% to 25% were tested with an interval of

2.5% using the same condition (see Method). Our result showed the ligation yield gradiently increased with more PEG 8000 addition from 15%-20% of concentration range, then decreased after 20%. The highest ligation efficiency was 46±2% with 20% PEG8000 addition (Fig. 3A lane 4). Then the effects of different PEG types including PEG 4000, PEG 6000 and PEG 8000 on ligation efficiency were tested with 20% of concentration (Supplementary Fig. 1). No significant difference was observed between PEG 8000 (46±2%) and PEG 6000 (45±3%), while lower ligation efficiency was shown with PEG 4000 (38±3%) compared to the other two PEGs. Therefore, we found that PEG 8000 was most efficient in the ligation reaction with 20% of concentration.

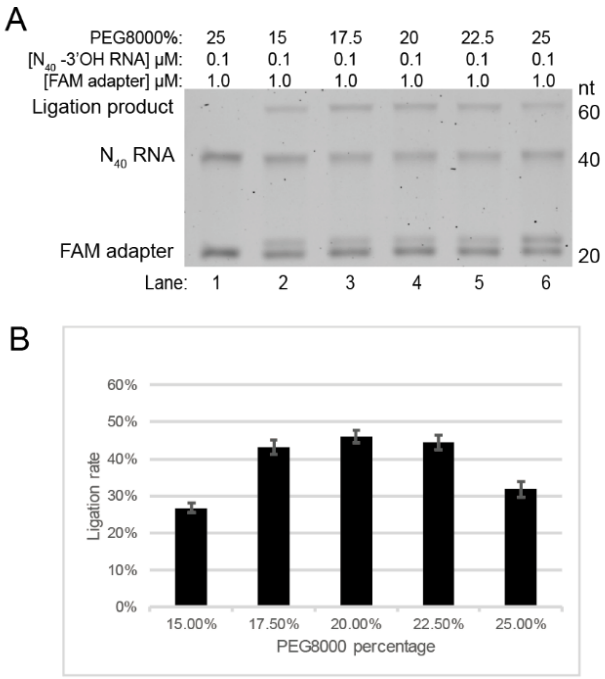

Figure 3. The effect of different PEG8000 percentage on the 3' adapter ligation. A) Ligation reactions were performed with 15, 17.5, 20, 22.5, and 25% PEG 8000. Negative control was performed by adding nuclease-free water instead of enzyme (Lane 1). An increasing ligation yield occurred from 15 to 20% PEG8000 (lanes 2-4), while from 20 to 25% PEG 8000 (lanes 4-6), the ligation yield dropped. B) Histogram showing the gel results in A). The error bars represent the standard error of at least 3 independent replicates.

3.3 Optimization of T4 RNA ligase 1 dosage

T4 RNA Ligase 1 was used in this study which has been reported to catalyze the ligation of a 5' phosphoryl-terminated adapter to a 3' hydroxyl-terminated RNA through the

formation of a 3’ → 5’ phosphodiester bond[22]. Here, different enzyme dosages were tested to explore the highest ligation efficiency (Fig. 4A). In reactions with the final volume of 10 µL, 0.5, 1, 1.5 and 2 µL (15, 30, 45 and 60 U final) T4 RNA ligase 1 was added in the reaction respectively (see Method). The result showed that the decrease of ligation efficiency was observed with increase of enzyme dosage (Fig. 4B). Highest ligation efficiency occurred with 0.5 µL (15 U final) of T4 RNA ligase 1 (47±5%), which was also the most cost-efficient dosage.

One reason for the observation that the ligation efficiency decreased with more T4 RNA ligase 1 dosage may be that the amount of RNA (acceptor) was relatively poor compared to the amount of enzyme. When the enzyme could not ligate the adenylated adapter (donor) to the RNA due to the low concentration of the RNA, the activated adapter can dissociate from the enzyme.[21]

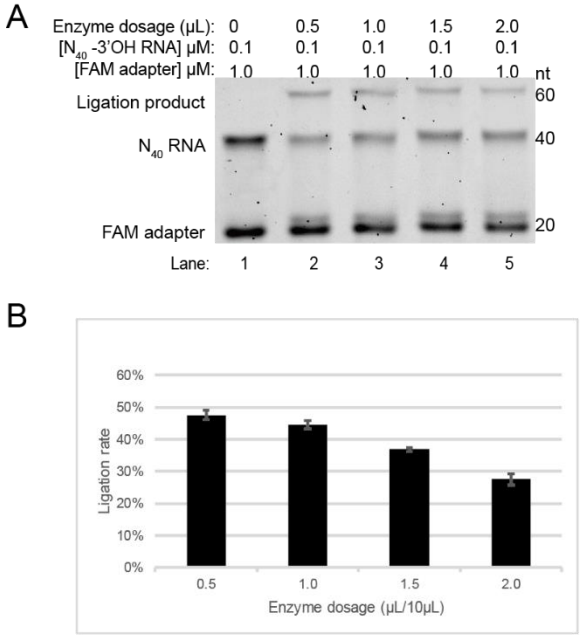

Figure 4. The effect of different T4 RNA Ligase dosages on 3’ adapter ligation. A) Ligation reactions were performed by adding 0.5, 1, 1.5 and 2 µL T4 RNA Ligase 1 (lane 1-5), i.e. 15, 30, 45, 60 units of enzyme. Reactions with no enzyme served as negative control (lane 1). The ligation efficiency decreased with the increase of enzyme dosage. B) Histogram showing the gel results in A. Error bars represent standard error of at least three independent replicates.

3.4 Optimization of reaction time

We anticipated that the ligation efficiency would increase with the longer reaction time. To study this, we tested the ligation efficiency with the different reaction time of 30, 60, 90 and 120 minutes respectively (see Method). Our result suggested that the production yield increases with the prolonged reaction time (Fig. 5). The reaction time of 30, 60, 90 and 120 minutes yielded 41 ± 1%, 52 ± 1%, 57 ± 1% and 58 ± 1% ligation efficiency (Fig 5A lanes 3-6), respectively, furthermore, the difference of ligation efficiency between 90 and 120 minute was not significant. Negative controls were performed with 0-minute (Fig. 5 lane 1) and 120-minute (Fig. 5 lane 7) incubation at 25 °C, confirming that no RNA degradation occurred during the 120-minute incubation at 25 °C. In sum, 120 min of ligation time was chosen to be the optimal condition.

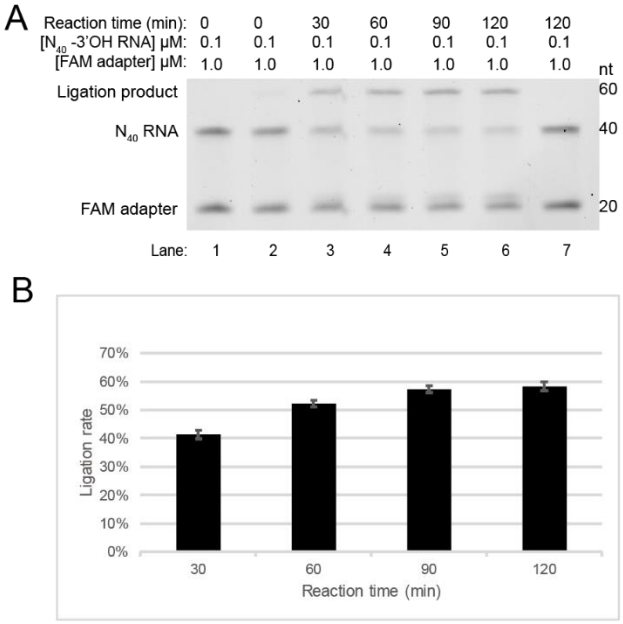

Figure 5. The effect of reaction time on 3' adapter ligation. A) Ligation reactions were performed by incubating at 25 °C for 30, 60, 90 and 120 minutes in thermocycler. Negative controls were performed by adding nuclease-free water instead of enzyme, without incubation at 25 °C (lane 1) or with 120-minute incubation at 25 °C (lane 7). Degradation was not observed after 120-minute incubation at 25 °C. A 0-minute reaction (lane 2), i.e. without incubation at 25 °C in the thermocycler machine and was heat inactivated right after adding enzyme, was set up to exclude interference of ligation occur before incubation at 25 °C. The ligation efficiency increased with the prolonged reaction time. The ligation rate of 90- (lane 5) and 120-minute (lane 6) reaction was similar. B) Histogram of gel result in A. Error bars represent standard error of at least 3 independent replicates.

3.5 Optimization of RNA to adapter ratio.

In adapter selection section (Figure 2) we observed that the ligation yield increased with the lower RNA: adapter ratio. We reasoned that the further decreasing RNA: adapter ratio, i.e. higher to RNA ratio, would lead to higher ligation yield. To verify this, ligation reactions with RNA: adapter ratios ranging from 1: 10 to 1: 25 were performed with the same condition (see Method). From our result, the ligation yields at the RNA: adapter ratio of 1: 10, 1: 15, 1: 20 and 1: 25 were 46 ± 3%, 53 ± 2%, 60 ± 3% and 63 ± 3% respectively (Fig. 6 lanes 2-5). Generally, production yield increased with the decreasing RNA: adapter ratio, so we choose 1:25 as the optimal condition of RNA: adapter ratio.

In sum, we found the T4 RNA ligase 1 can efficiently ligate the 5'p-N4A16-3' FAM adapter to the synthetic model RNA, N_{40} -3' OH, with the PEG 8000 concentration of 20%, enzyme dosage of 0.5 μL (15 U final) per 10 μL reaction, and 1:25 RNA: adapter ratio, with production yield higher than 60% within 90 minutes at 25 °C.

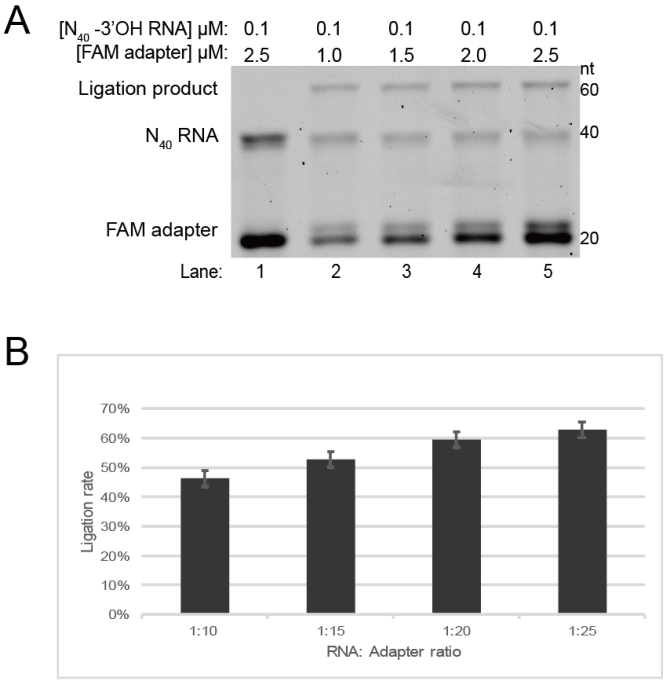

Figure 6. The effect of RNA to adapter ratio on 3' adapter ligation. A) Ligations with RNA: adapter ratio from 1: 10 to 1: 25 were performed with 5' p-N4A16-3' FAM adapter (lanes 2-5). The ligation rate increased with the increasing concentration of FAM labelled adapter. Negative control was performed by adding nuclease-free water instead of enzyme (lane 1) B) Histogram showing the gel results in A. Error bars represent standard error of at least three independent replicates.

3.6 EMSA of unligated and ligated APP RNA

To test whether the ligated adapter would affect the binding affinity between RNA and protein, EMSA was performed for both unligated and ligated APP RNA (see Methods). Bindings under the same conditions were performed to unligated and ligated APP RNAs. EMSA result of unligated APP RNA (**Figure 7**) was analysed using ImageJ and Prism, giving the estimated K_D of 278.4±45 nM. As for the EMSA of ligated RNA (**Supplementary Figure S3**), however, the background was too strong to proceed with the intensity measurement and data analysis.

Problems were observed during EMSA, which may lead to the non-accurate estimation of K_D. From **Figure 7** and **Supplementary Figure S3** we can see that the background noise was strong and the bands of binding complex were smear, which resulted in difficulties in intensity measurement and K_D estimation. The EMSA protocol needs further optimization for lowering down the background noise as well as avoiding smear binding bands. However, due to the time limitation, we were not able to further optimize the EMSA protocol nor perform more replicates for validation of K_D. We hope this can be achieved in future experiments.

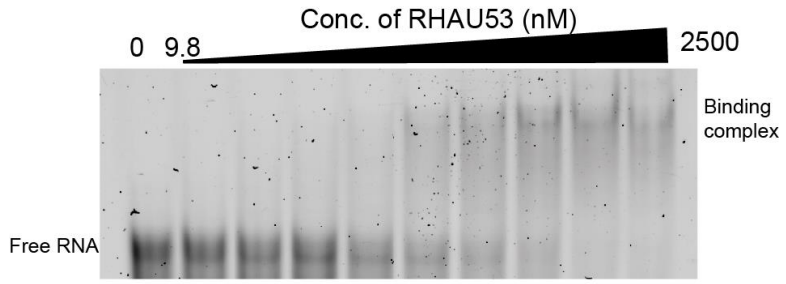

Figure 7 Binding complex was formed between unligated APP RNA and RHAU53. A constant 100 nM concentration of unligated APP RNA and a variable RHAU53 concentration ranging from 0 to 2500 nM were used for EMSA. Concentration gradient was generated by serial dilution (1:2) from 2500nM and a negative control without RHAU53, e.g. 0 nM, was generated. The estimated Kd was 278.4±45 nM.

3.7 MST of ligated APP RNA

MST was performed to ligated APP RNA for estimation of K_D instead of EMSA (**Figure 8**). The K_D estimated by MST was 200±80 nM, which was not significantly different from the value obtained from EMSA of unligated APP RNA, i.e. 278.4±45 nM. However, due to the limitation of time, only one replication of MST was performed, which is not enough to lead to the conclusion of the effect of ligated adapter. Future experiments are needed to provide more replicates for validation of K_D and whether the ligated adapter would affect the biological properties of the RNA.

Compared to EMSA, MST has some advantages such as short handling time and can detect the binding in free solutions as mentioned previously[18]. There is no consideration of the gel casting or smear binding band. However, MST requires fluorescent labelled targets for detection while radiolabelled target or gel staining can be applied for visualization in EMSA. In other words, labelling RNA with the ligation method can provides more choice of technologies for detecting the binding between RNA and other molecules.

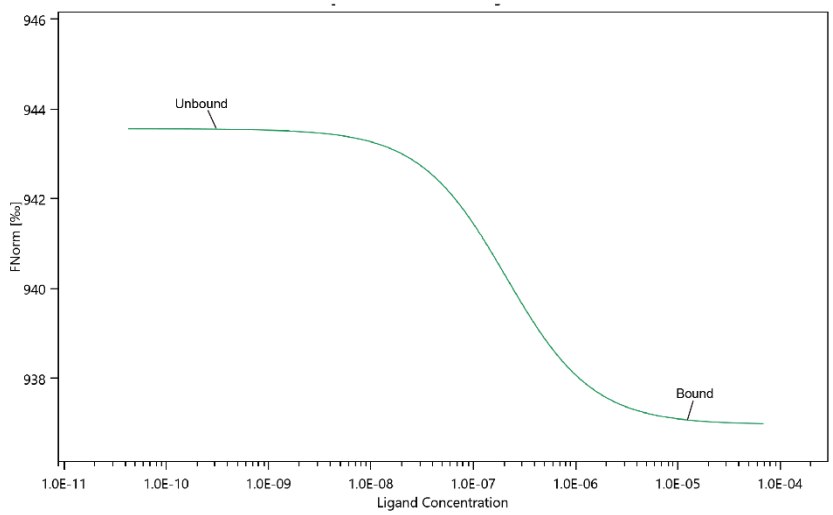

Figure 8. Binding between ligated APP RNA and RHAU53 was detected by MST. A constant 20 nM concentration of ligated APP RNA and a variable RHAU53 concentration generated by serial dilution (1:2) from 7000 nM were used for MST. The estimated Kd of binding between ligated APP RNA and RHAU53 was 200±80 nM.

4. Future directions

In future experiments, we will perform more replicates of EMSA and MST for APP RNA

to validate the effect of ligated adapter. We also plan to test the ligation between the adapter and longer RNAs and analyse the effect of the ligated adapter on the interaction between longer RNAs and other molecules.

Moreover, we anticipate that the FAM labelled adapter can be used for primer hybridization in reverse transcriptase stalling assay (RTS) or RNA structure probing. Experiments will be performed to test this hypothesis.

5. Conclusion

In this project, we have reported a method to fluorescent label RNA by ligation of a ssDNA 3' adapter. We demonstrated that with the optimized protocol of the 3' adapter ligation, the ligation yield can be higher than 60%. The ligated adapter did not show significant effect on the biological properties, e.g. binding affinity with other molecules, which can be applied for technologies such as EMSA and MST.

6. References

1 Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 7287-7292, doi:10.1073/pnas.0401799101 (2004).

2 Huppert, J. L. Structure, location and interactions of G-quadruplexes. *The FEBS Journal* **277**, 3452-3458, doi:10.1111/j.1742-4658.2010.07758.x (2010).

3 Sanchez de Groot, N. *et al.* RNA structure drives interaction with proteins. *Nature Communications* **10**, 3246, doi:10.1038/s41467-019-10923-5 (2019).

4 Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561-563, doi:10.1038/227561a0 (1970).

5 Mattick, J. S. & Gagen, M. J. The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms. *Molecular Biology and Evolution* **18**, 1611-1630, doi:10.1093/oxfordjournals.molbev.a003951 (2001).

6 Li, J. & Liu, C. Coding or Noncoding, the Converging Concepts of RNAs. *Frontiers in Genetics* **10**, doi:10.3389/fgene.2019.00496 (2019).

7 Zhang, C. Novel functions for small RNA molecules. *Curr Opin Mol Ther* **11**, 641-651 (2009).

8 Christov, C. P., Gardiner, T. J., Szüts, D. & Krude, T. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Mol Cell Biol* **26**, 6993-7004, doi:10.1128/MCB.01060-06 (2006).

9 Natoli, G. & Andrau, J.-C. Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annual Review of Genetics* **46**, 1-19, doi:10.1146/annurev-genet-110711-155459 (2012).

10 Kishore, S. & Stamm, S. The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C. *Science* **311**, 230, doi:10.1126/science.1118265 (2006).

11 Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281-297, doi:https://doi.org/10.1016/S0092-8674(04)00045-5 (2004).

12 Hermanson, G. T. *Bioconjugate Techniques*. (Elsevier Science & Technology, 2013).

13 Proudnikov, D. & Mirzabekov, A. Chemical Methods of DNA and RNA Fluorescent Labeling. *Nucleic Acids Research* **24**, 4535-4542, doi:10.1093/nar/24.22.4535 (1996).

14 Ferrari, V. C. & Hollibaugh, J. T. Distribution of microbial assemblages in the Central Arctic Ocean Basin studied by PCR/DGGE: analysis of a large data set. *Hydrobiologia* **401**, 55-68, doi:10.1023/A:1003773907789 (1999).

15 Neufeld, J. D. & Mohn, W. W. Fluorophore-labeled primers improve the sensitivity, versatility, and normalization of denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **71**, 4893-4896, doi:10.1128/AEM.71.8.4893-4896.2005 (2005).

16 Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic acids research* **9**, 3047-3060, doi:10.1093/nar/9.13.3047 (1981).

17 Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic acids research* **9**, 6505-6525, doi:10.1093/nar/9.23.6505 (1981).

18 Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* **2**, 1849-1861, doi:10.1038/nprot.2007.249 (2007).

19 Seidel, S. A. I. *et al.* Label-Free Microscale Thermophoresis Discriminates Sites and Affinity of Protein–Ligand Binding. *Angewandte Chemie International Edition* **51**, 10656-10659, doi:10.1002/anie.201204268 (2012).

20 Asmari, M., Ratih, R., Alhazmi, H. A. & El Deeb, S. Thermophoresis for characterizing biomolecular interaction. *Methods* **146**, 107-119, doi:https://doi.org/10.1016/j.ymeth.2018.02.003 (2018).

21 Uhlenbeck, O. C. & Gumport, R. I. in *The Enzymes* Vol. 15 (ed Paul D. Boyer) 31-58 (Academic Press, 1982).

22 Brennan, C. A., Manthey, A. E. & Gumport, R. I. in *Methods in Enzymology* Vol. 100 38-52 (Academic Press, 1983).

23 Yeung, P. Y. *et al.* Systematic evaluation and optimization of the experimental steps in RNA G-quadruplex structure sequencing. *Scientific Reports* **9**, 8091, doi:10.1038/s41598-019-44541-4 (2019).

24 Harrison, B. & Zimmerman, S. B. Polymer-stimulated ligation: enhanced ligation of oligo- and polynucleotides by T4 RNA ligase in polymer solutions. *Nucleic acids research* **12**, 8235, doi:10.1093/nar/12.21.8235 (1984).