# Development of chemical tools for imaging RNA and studying RNA and protein interactions

Zhang, Tiancai

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
2022

[Link to publication in University of Groningen/UMCG research database](#)

# Development of chemical tools for imaging RNA and studying RNA and protein interactions

**Tiancai Zhang**

Development of chemical tools for imaging RNA and studying RNA and protein interactions


Tiancai Zhang
PhD thesis
University of Groningen

July 2022

Cover designed by: Tiancai Zhang

# Development of chemical tools for imaging RNA and studying RNA and protein interactions

**PhD thesis**

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Tuesday 30 August 2022    at 12.45 hours

by

**Tiancai Zhang**

born on 15 July 1987
in Gansu

**Supervisor**

Prof. A. Herrmann


**Co-supervisor**

Dr. P. van Rijn


**Assessment Committee**

Prof. R. Schirhagl
Prof. J.G. Roelfes
Prof. A. Marx

# Contents

# Chapter 1

# An introduction to technologies for RNA study

RNA plays a pivotal role in most of biological processes, especially in the process of protein translation. RNA was initially considered as a messenger for transferring genetic information in DNA to protein producing machine, we call this kind of RNA messenger RNA (mRNA). Cell activity is regulated by different kind of RNAs in different ways. Transfer RNAs (tRNAs) carrying amino acids recognize the corresponding mRNA template and help to fabricate proteins via ribosomal RNA (rRNA) synthesis. Besides aforementioned RNAs that take part in protein production, many other non-coding RNAs are also involved in this process, such as small nuclear RNA (snRNA) and micro-RNA (miRNA) participate in pre-mRNA processing and post-transcriptional regulation[1-5].

In consideration of the important functions of RNA in organisms, understanding the process of RNA from generation to decay in the whole cellular life and the mechanism of how its function is carried out is essential to study cell physiology. In order to comprehensively understand how RNA works, various technologies have been developed continuously[6-8]. We divide these technologies into two parts, direct detection and indirect investigation.

We here define RNA imaging technology as a direct method to observe RNA movement, RNA localization, and RNA exportation. While indirect methodologies refer to study target RNA through exploring the biomacromolecules it interacts with, such as DNA, RNA and RNA binding protein.

## 1.1 Direct method: RNA imaging

Several RNA imaging methods exist.

### 1.1.1 Fluorescence in situ hybridization (FISH)

FISH is a cytogenetic imaging technique involving a fluorescently labeled single-stranded DNA or RNA binding to the specific part of the target sequence[9-12]. The earliest hybridization based imaging method relied on a radiolabeled RNA or single-stranded DNA in the 1960s. Limited by several obvious drawbacks such as instability of the radioactive probes, low resolution and hazardous property, this technique was replaced by fluorescent labeled probes which were first applied in the 1980s by labeling RNA on its 3'end with fluorophore for DNA detection and localization[13]. There are mainly two methods to prepare the probes, one is enzymatically incorporating functional bases to oligonucleotide and the other one is chemically synthesizing a

fluorophore labeled single-stranded DNA. However, this probe is also limited by high background signal since the probes are also fluorescent when there is no target RNA binding with them. Since the oligonucleotide probes are negatively charged, cells need to be fixed and the lipids on the membrane need to be removed to enable the probes to enter cells, which means it is hard to use FISH probes in real time imaging. In order to reduce the background signal, unbound probes have to be washed away after hybridization. Another way to distinguish from background signal is using a mixture of multiple fluorescent labeled probes since only the target sequence is able to bind all those probes and induces enough fluorescence (**Scheme 1-1**).
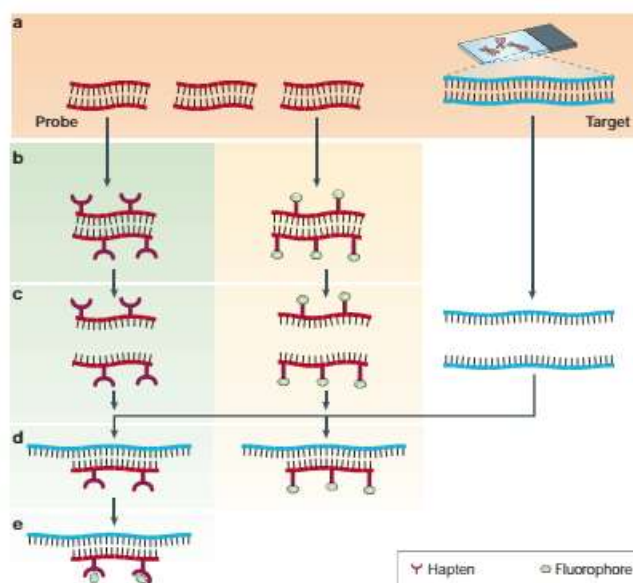


**Scheme 1-1 Illustration of fluorescence in situ hybridization. a**) A target double-stranded sequence and double stranded nucleotide oligomers for preparing FISH probes. **b**) Fluorophores are directly incorporated into the nucleotide oligomers (right panel) or haptens are labelled on the oligomers (left panel). **c**) Double-stranded target sequence and probes are denatured to form single-stranded sequences. **d**) Probes and target DNA/RNA hybridize on complementary sequences. **e**) The hapten labelled probe is visualized through immunological imaging.

### 1.1.2 MS2 imaging system

In order to break the temporal resolution limitation of FISH probes, MS2 imaging system was developed to visualize RNA in living cells[14]. MS2 imaging system is a classic RBP-FP (RNA binding protein-fluorescent protein) probe.

RBP-FP denotes probes consisting of fused RNA binding protein and fluorescent protein. Probe binds with target RNA with its binding domain and therewith labels the RNA with fluorescent protein. In MS2 system, the RNA binding protein is a bacterial

phage MS2 coat protein (MCP), which binds with target RNAs that contain several repeats of the MS2 binding sequence (MBS). To image mRNA with MS2 system, several units of MBS were incorporated into 3'-UTR (untranslated region) of target mRNA. Probes were prepared by fusion of MCP and fluorescent protein like GFP. The pre-treated mRNAs and probes were added in to cells, MCP will then bind with MBS and decorate GFPs on target mRNA. Comparing to the initially invented MS2 system with only 6 repeats of MBS inserted into target mRNA, scientists even incorporated 24 MBS repeats in a single mRNA for the purpose of enhancing the signal (**scheme 1-2a**)[15-17].

Furthermore, small molecule fluorophores were introduced to MS2 systems to engineer more versatile probes, since small molecule fluorophores have broader emission wavelength that can avoid the interference with cell auto-fluorescence. MS2-eDHFR (fusion of MS2 coat protein and Escherichia coli dihydrofolate reductase) and MS2-SNAP (fusion of MS2 coat protein and SNAP tags) systems were developed to decorate small molecule dyes on target mRNA. eDHFR can bind tightly to trimethoprim (TMP) and SNAP tag can form covalent bonds with benzylguanine (bG) or benzylchloropyrimidine (CP). Aaron A. Hoskins proved that MS2-eDHFR/fluorescein-TMP and MS2-SNAP/fluorescein-bG systems could be efficiently used to image mRNA in yeast. This technique has another obvious advantage that is the fluorescence control by adding small molecule fluorophore rather than depending on expression of probes in MS2-GFP system (**scheme 1-2b**)[18, 19].

However, the main drawback of the aforementioned MS2 based image systems is that the background signal cannot be eliminated, since the unbound MS2-FPs are still fluorescent. To overcome this limitation, bimolecular fluorescence complementation (BiFC) was introduced into MS2 systems (**scheme 1-2c, 2d**). In BiFC based system, fluorescent proteins are split into two parts: N-terminal fluorescent protein (N-FPs) and C-terminal fluorescent protein (C-FP), since separated N-FP or C-FP are non-fluorescent, the background signal could be eliminated. In this system, the two halves of fluorescent protein were respectively conjugated to two different RNA binding proteins: MCP and PCP (PP7 bacteriophage coat protein). Target RNA harboring MBS and PBS (PP7 binding sequence) binds with N-FP-MCP and C-FP-PCP, which enabled the N-FP and C-FP to get close to form a complete fluorescent protein, the target RNA is finally fluorescently labelled [20, 21]. In another BiFC systems, the binding mode of RBP and RNA is replaced by aptamer-protein interaction, leading to the fact that

aptamer RNAs are incorporated into target RNA to bind RBP-FP[22].

Although later developed BiFC system lower the background signal compared to traditional MS2 system, the interference is still not avoided since highly concentrated free N-FPs and C-FPs can fuse to form complete fluorescent protein in the absence of target RNAs. Moreover, fusion of these two parts takes time so that this method is not suitable for imaging short-lived RNAs.
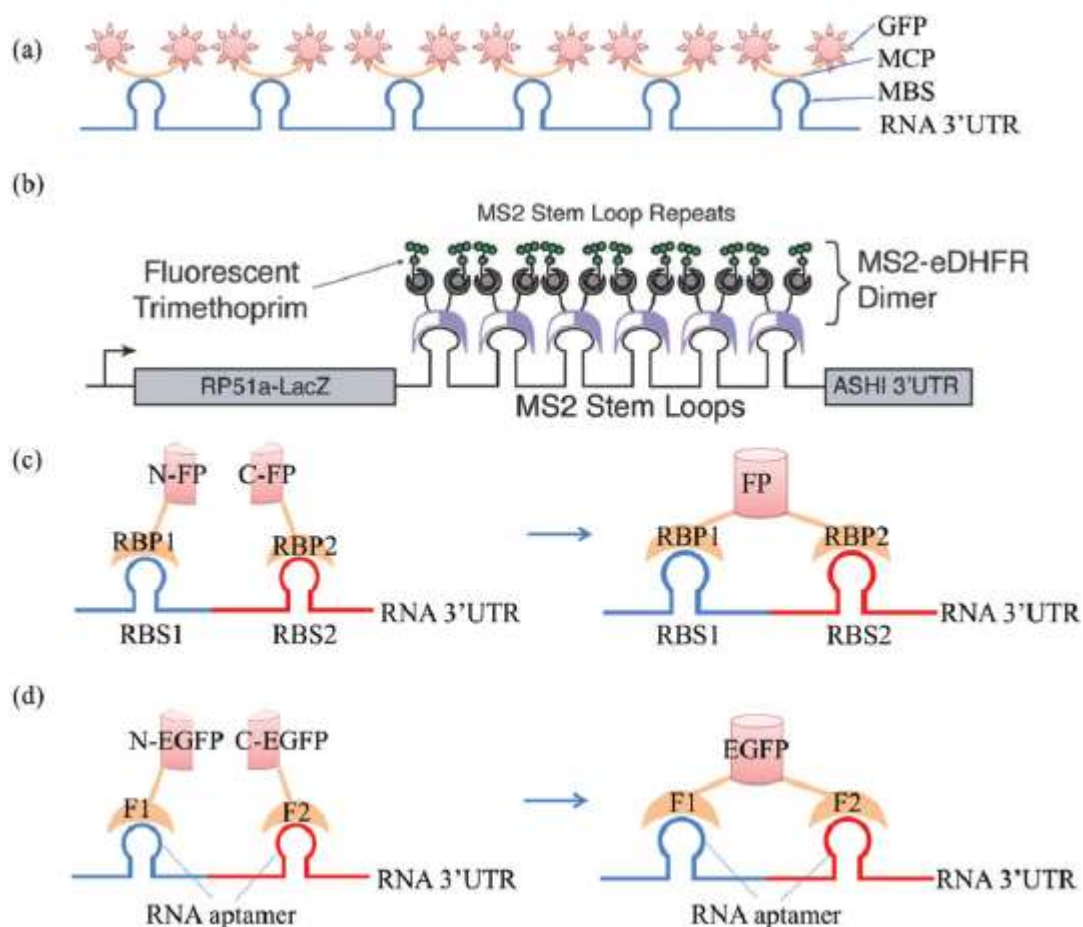


**Scheme 1-2 Different forms of RBP based probes. a)** Principle of MS2-GFP imaging system: MCPs carrying GFPs label target RNA through binding to six pre-inserted repeats of MBS. **b)** Principle of MS2-eDHFR imaging system: eDHFR is fused with MS2 coat protein and modified on target RNA by binding with MS2 binding sequence, then trimethoprim substituted fluorophores are added to illuminate the eDHFR labeled RNAs. **c)** Principle of BiFC system: two split parts of a fluorescent protein (N-FP and C-FP) are respectively fused with two RBPs and labeled on target RNA that was pre-treated with two corresponding RBP binding sequences (RBS). After the two RBPs binding to RBS1 and RBS2, the distance of N-FP and C-FP gets closer and enable them to fuse to form a complete fluorescent protein. **d)** Principle of aptamer based BiFC system: The strategy is almost the same as conventional BiFC. The difference is

the RBP binding sequences are exchanged by aptamer RNA.

**1.1.3 Pumilio imaging system**

Pumilio proteins are a group of RNA binding protein and function by regulating mRNA translation. The RNA binding domains in pumilio family protein are often known as pumilio homology domain (PUM-HD) in human pumilio1. Like BiFC system, two split fragments of enhanced GFP (EGFP) are fused to two PUM-HDs. The unique feature of this probe is that PUM-HD binds to RNA sequence through recognizing each single base with corresponding elements, respectively. Specifically, a PUM-HD comprises 8 elements and each element is able to recognize every single base of an RNA sequence UGUANAUA. To recognize an RNA, a sequence that is similar to UGUANAUA in 3'-UTR is targeted and corresponding elements of a PUM-HD are mutated to achieve high binding affinity. Another 8 base sequence is chosen near the former sequence, and the second PUM-HD is also pre-treated with the same strategy. Upon both PUM-HDs recognizing the two sequences of target RNA, the two split fragments of EGFP are fused close enough to recover fluorescence. Since two different PUM-HDs carrying 16 elements can specifically recognize 4*16 transcripts, it is sufficient to use this probe to label any target mRNA in a cell (**scheme 1-3**)[23-25].



**Scheme 1-3 Principle of pumilio imaging system.** Two PUM-HD mutants (mPUM-1 and mPUM-2) were fused to two split halves of EGFP (N-EGFP and C-EGFP), respectively. In the presence of target mRNA, mPUM-1 and mPUM-2 recognize corresponding binding sequences and enable the two EGFP parts to reconstitute to recover fluorescence.
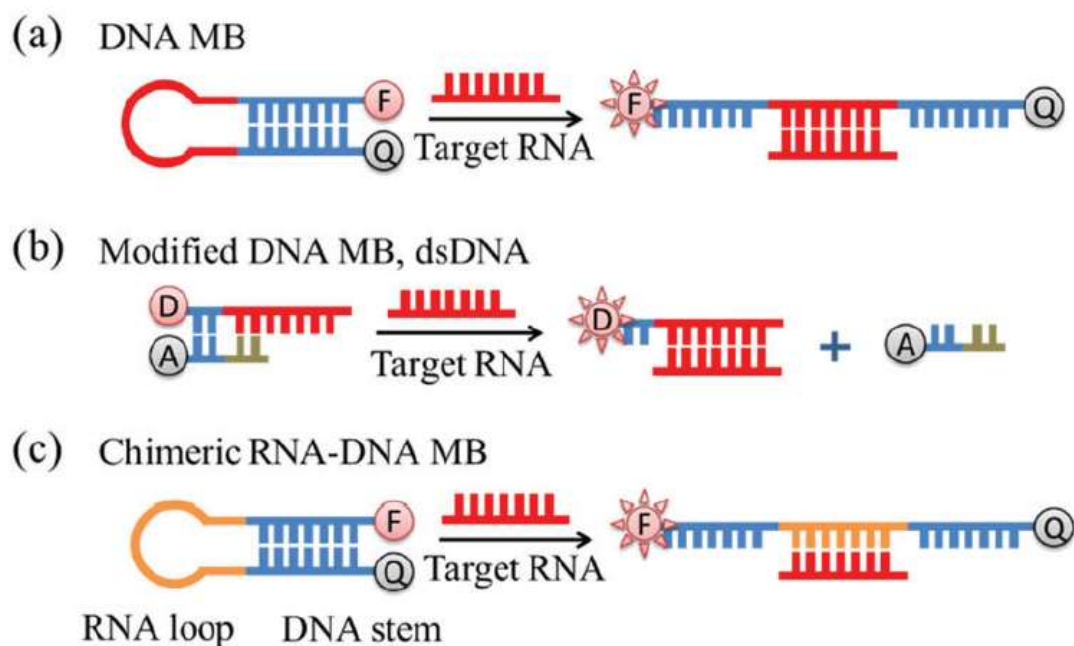
## 1.1.4 Molecular beacons (MBs)

Molecular beacon is a nucleic acid probe that has a fluorophore and a quencher labeled on the two ends of a hairpin shaped synthetic DNA. In this probe, DNA is stem-loop folded so that it can recognize the target RNA with its loop sequence while fluorophore and quencher pair are attached to the opposite ends of the stem sequence. In the absence of target sequence, stem-loop shaped probe enables closely positioned quencher and fluorophore, thus no fluorescence can be detected. While in the presence of target RNA that contains the complementary sequence, the loop part hybridizes to the analyte and causes unwinding of the stem part. As a result, fluorophore and quencher are separated and fluorescence is switched on (**Scheme 1-4a**)[26]. The secondary structure enables the MB probe to afford great selectivity, since only when the energy offered by target RNA is high enough, the probe-analyte hybrid will be formed. For this reason, this probe was used for analyzing single-nucleotide polymorphisms (SNPs). Modification of the backbone of MB was also pursued whereby the hairpin structure was replaced by double-stranded DNA, which consists of one recognition strand harboring a FRET donor and one strand harboring FRET acceptor as a quencher (**Scheme 1-4b**). Once target RNAs hybridize to the recognition strand, quencher strand is released and fluorescence is detected[27].

Another widely applied backbone modification of this probe is chimeric RNA-DNA MB (**Scheme 1-4c**)[28, 29]. In order to reduce background signal, the hybridizing affinity of target RNA and recognition strand needs to be enhanced. Comparing to typical DNA harpin MB, the DNA stem and RNA loop comprising chimeric RNA-DNA MB have higher specificity with target RNA, since thermodynamic stability of RNA-RNA strand is higher than RNA-DNA strand. Enhanced specificity enables faster hybridization which reduces dynamic opening of hairpin structure that attributes to background fluorescence.

(a) DNA MB

(b) Modified DNA MB, dsDNA

(c) Chimeric RNA-DNA MB

RNA loop    DNA stem

**Scheme 1-4 Structure of molecular beacons (MBs). a**) The classic DNA hairpin MB; **F** and **Q** represent fluorophore and quencher. **b**) Modified structure of MB; **D** and **A** denote FRET donor and acceptor. **c**) Chimeric RNA-DNA MB consists of RNA loop and DNA stem.

### 1.1.5 Aptamer based RNA imaging

The functional group of green fluorescent protein (GFP) is 4-hydroxybenzilidiene imidazolinone (HBI), which is formed by intra-cyclization of three residues of GFP (Ser65-Tyr66-Gly67). However, chemically synthesized HBI is non-fluorescent, which is due to free rotation as a way to dissipate energy when excited. Therefore, it is concluded that HBI is only fluorescent when bound in the protein, which restricts the molecular rotation. Inspired by this mechanism, researchers designed a series of RNA aptamers of HBI, which function as GFP. When HBI binds to the aptamer, fluorescence is detected. A series of HBI derivatives were synthesized to improve the behavior of the fluorophore; finally 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) was targeted as the extraordinary candidate, whose quantum yield is even higher than in eGFP. The RNA aptamer of DFHBI was evolved by (Systematic Evolution of Ligands by Exponential Enrichment) SELEX and named SPINACH[30-32] (**Scheme 1-5**).
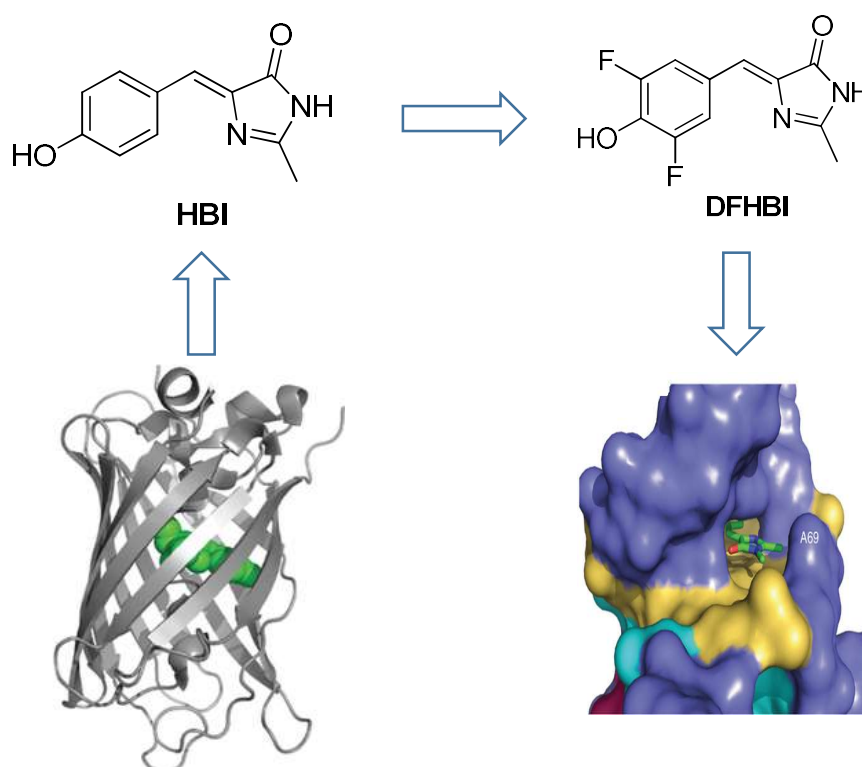
**Scheme 1-5 The evolutionary process of GFP mimicing imaging system.** The compound HBI was found as the fluorophore of GFP that was formed through cyclization of protein residues. The key of the photoluminescence mechanism of GFP is that HBI is restricted in a special site of the protein. This light up process was applied in RNA imaging, which mimics the restriction of the fluorophore motion in the protein by a RNA aptamer. DFHBI was derived from HBI and the corresponding aptamer SPINACH library was screened to obtain a brighter fluorescence system.

With the development of aptamer/fluorophore imaging system, a series of fluorophores such as DFHBI-1T, PFP-DFHBI and TO1 (Thiazole Orange) were developed to image RNA in a wider spectral range. To improve the specificity of the imaging system, aptamers like SPINACH2 and Broccoli were developed to increase the binding affinity of fluorophore and RNA[33, 34]. The most common form of aptamer/fluorophore system is that aptamer is engineered into 3'-UTR of target RNA, fluorescence will be detected from target RNA after fluorophores bind to aptamers (**Scheme 1-6a**). To reduce the background signals, fluorophore-quencher methods were introduced into aptamer/fluorophore system (**Scheme 1-6b**). Aptamer of fluorophore or quencher is pre-inserted into 3'-UTR of target RNA, after the binding of fluorophore or quencher

to its aptamer, fluorophore is separated from the quencher and fluorescence is recovered. Aptamer/fluorophore imaging systems have two advantages over RBP-FP and BiFC imaging systems. First, comparing to RBP-FP probe wherein RBP-FP binds to target RNA and BiFC probe wherein two split parts of fluorescent protein fuse together, aptamer RNA binds to fluorophore and illuminates target RNA much faster, therefore, this probe is more suitable for real-time RNA imaging. Second, unbound fluorophores like DFHBI have very low fluorescence but brighter than EGFP when binding to the aptamer. Hence, the signal-to-noise ratio of this system is higher than in RBP-FP and BiFC probes[35, 36].
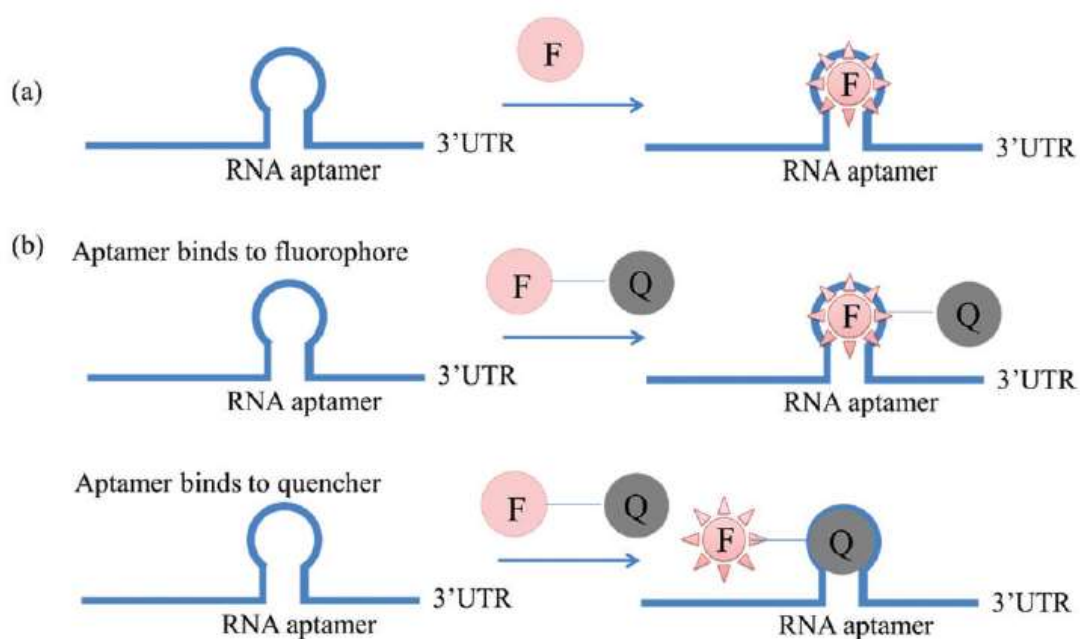


**Scheme 1-6 Two different forms of aptamer/fluorophore probe. a**) RNA aptamer is fused to target RNA. After fluorophore (F) binds to aptamer, the target RNA is illuminated. **b**) Fluorophore-quencher system is combined with fluorophore/aptamer probe. The aptamer of fluorophore (F) or quencher (Q) is incorporated into target RNA. After binding of fluorophore or quencher with aptamer, F and Q is separated and fluorescence is emitted.

## 1.2 Indirect method: RNA/RNA or RNA/protein crosslinking

While in the sections above RNA imaging probes were discussed, in this paragraph strategies will be discussed that elucidate the function RNA. Many noncoding RNAs regulate transcription, Pre-mRNA maturation and translation through direct interaction,

like micro-RNA regulating gene expression by hybridizing to target mRNA. The interaction of these RNAs could be captured by crosslinking of hybridized RNAs with chemical crosslinkers such as formaldehyde, psoralen, and disuccinimidyl glutarate. Most of long noncoding RNAs (lncRNA) interact with target mRNA through one or several protein intermediates. In order to study the function of these RNAs, the complexes of lncRNA and protein intermediate need to be captured, which is realized by RNA and protein crosslinking.

**RNA-RNA crosslinking: RAP-RNA**

**1.2.1 RAP-RNA (AMT)**

Direct RNA-RNA interactions are usually studied by AMT-based RNA crosslinking method. 4'-aminomethyltrioxalen (AMT) is a derivative of psoralen, which is applied as a crosslinker between two uridines upon photo activation. AMT crosslinks two RNA molecules through reacting with two opposite positioned uridines in the base pairing fragments. (**Scheme 1-7**)[37]
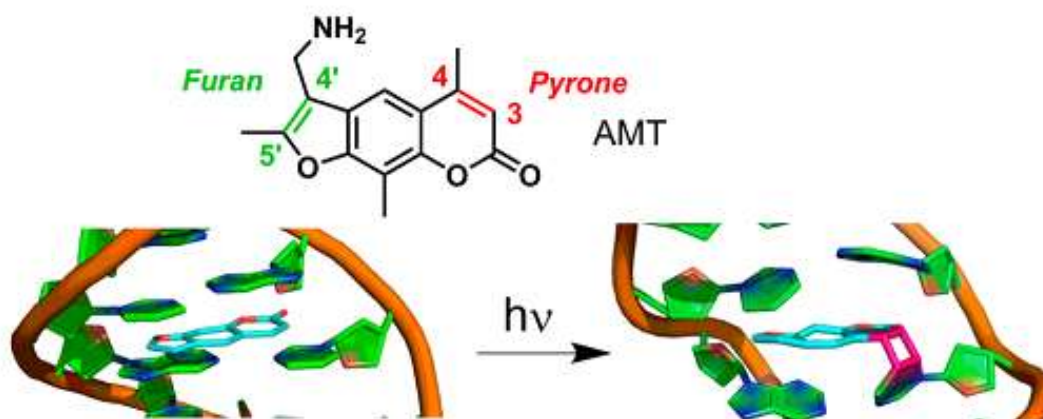


**Scheme 1-7 The principle of UV induced crosslinking between AMT and double stranded nucleic acid.** The photoactive bonds of AMT are highlighted in green and red. The plane structure enables AMT to intercalate into the groove of double stranded nucleic acid, and upon irradiation with short wavelength light (365 nm), the double bond of pyrone and neighboring uridine undergoes 2+2 cycloaddition. The second cycloaddition takes place between the double bonds of furan and uridine as long as the stagger base of another strand is uridine.

Purification of crosslinked product is a critical step in studying of RNA-RNA interactions. An important technique has been developed for enrichment of the

conjugate of target RNA and its interaction object. The target RNA is captured by pools of probes composed of biotinylated single-stranded DNA, which is complimentary to a certain sequence of target RNA. This purification technique is normally called RNA antisense purification (RAP).

The main procedure of RAP-RNA (AMT) starts with incubating cells in AMT solution, then crosslinking is triggered by irradiation with 365 nm light, after lysis of the cells, protein and DNA are digested by proteinase K and DNase. Target RNA is enriched by RAP, washed, eluted and fragmented. A 3'adapter is ligated to the RNA fragment followed by RT-PCR. A 5'adapter is then introduced and the cDNA library is finally constructed for high through-put sequencing (**Scheme 1-8**)[38].
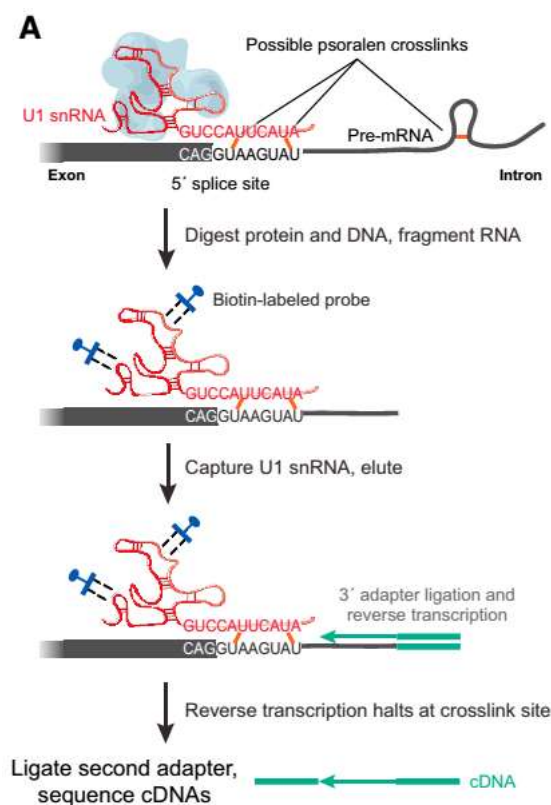


**Scheme 1-8 The RNA-RNA interaction research method based on AMT crosslinking.** Two interacting RNAs (U1 snRNA and target pre-mRNA) are captured by a psoralen derivative (AMT) and crosslinking upon UV irradiation. Proteins (pale blue) and DNAs are digested to exclusively obtain the crosslinked RNAs. The crosslinked RNAs are then pulled down with biotin labeled probes, which are complementary to certain sequence of U1 snRNA. The collected RNAs are fragmented so that the RNA-AMT-RNA crosslink sites could be precisely recognized. The RNA fragments are then ligated with a 3' adapter and reverse-transcribed to

cDNAs. The reverse transcription is normally hindered by covalent crosslinked AMT, so that the terminal of cDNA is near the crosslink sites.

## 1.2.2 RAP-RNA (FA and DSG)

Besides the AMT based crosslinking strategy, another chemical reagent is more often used in RNA-RNA interaction research due to higher crosslinking efficiency. As a ubiquitous crosslinking reagent, formaldehyde is widely used for studying protein-protein interactions, DNA-protein interactions and RNA-protein interactions.

Comparing to psoralen that reacts with uridine upon UV irradiation, formaldehyde can form a covalent bond with RNA by reacting with bases, and proteins by reacting with amino groups of amino acids. This feature enables the technique not only to capture indirect RNA-RNA interaction (interaction through protein intermediate) by crosslinking target RNAs and their protein intermediate, but also to capture direct RNA-RNA interaction (interaction through hybridization) through crosslinking target RNAs and proteins that have interaction with RNAs.

However, RAP-FA and RAP-AMT can only capture zero-distance contacting RNA/RNA or RNA/protein，for some RNAs that interact indirectly through multiple protein intermediates，another crosslinking reagent disuccinimidyl glutarate (DSG) is needed. As a stronger protein crosslinker, DSG and FA are often used together to capture both direct and indirect RNA-RNA interactions. Comparing to RAP-AMT and RAP-FA protocol that relies on strongly fragmenting the crosslinked RNAs to precisely map the binding sites, the FA-DSG protocol integrates RNA before capture to obtain interacting target RNAs (**Scheme 1-9**)[38].
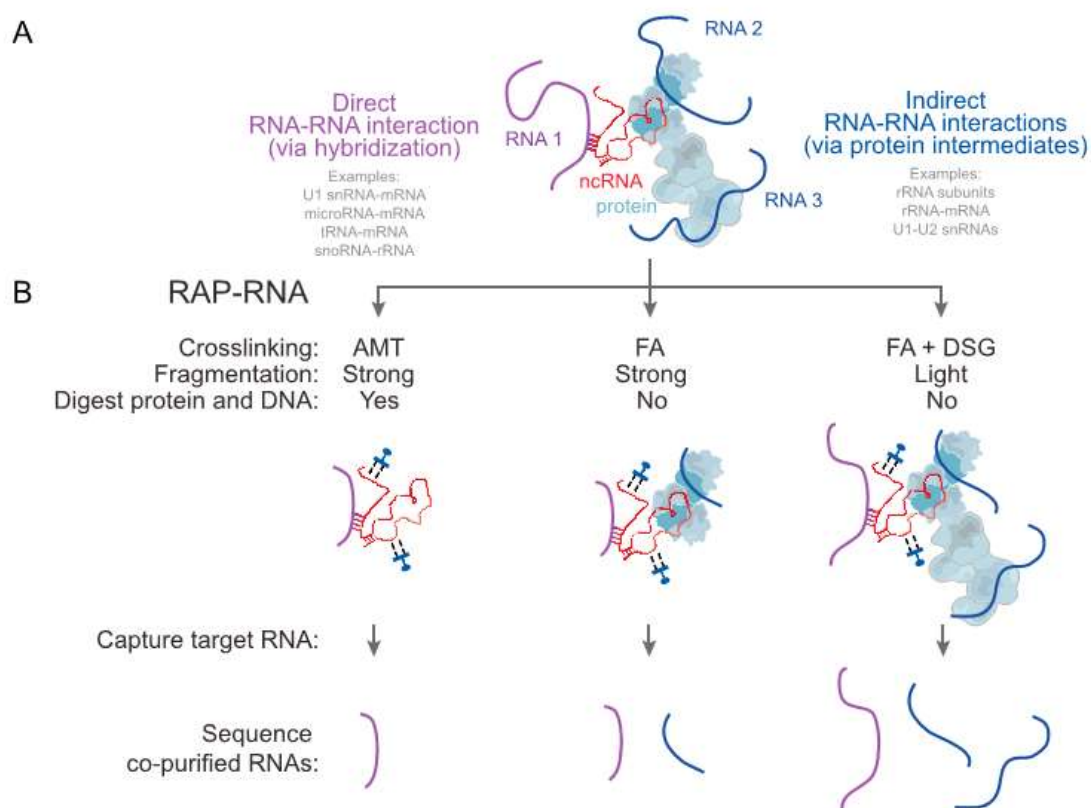
**Scheme 1-9 comparison of three RAP-RNA protocols based on AMT，FA and DSG.**

Crosslinking: AMT is used to crosslink directly interacting RNA (ncRNA/RNA1). The interaction occurs through RNA/RNA hybridization and the opposing uridines contained in the base paring sites could be crosslinked by AMT. FA is able to crosslink RNA/RNA or protein/protein interactions. Hence, it is applied to capture RNA/RNA interactions through crosslinking target RNA (ncRNA) and proteins that mediate the interaction of ncRNA and RNA2. FA is also used to crosslink target RNA (ncRNA) and proteins that wrap the interacting RNAs (RNA2). DSG is a strong protein/protein crosslinker, the combination of DSG and FA can capture RNA and RNA interactions that are mediated by multiple proteins.

### 1.2.3 RNA and protein crosslinking: CLIP

The most fundamental method for studying RNA binding proteins is pulling down the RNA and protein complex without crosslinking, but only through immunoprecipitation, the proteins of the complex are digested and RNAs are reverse transcribed to obtain a cDNA library, which is analyzed by microarray or high throughput sequencing. This technique was defined as RIP-Chip[39, 40] and RIP-seq[41]. However, the main drawback of this technique is the weak binding force between the target RBP and the associated

RNAs, which leads to missing of a great part of RNAs in the process of cell lysis, immunoprecipitation and bead washing. Compared to the weak hydrogen bonding and van der Waals forces in RIP-seq method, covalent crosslinking between RBP and RNA is necessary to capture more comprehensively protein and RNA interactions.

Photo-reactivity of uridine and its analogues was discovered to be very useful for exploring RNA participating in cellular activity, especially for studying RNA binding proteins. Uridine is able to react with some amino acids like tyrosine, phenylalanine and tryptophan upon irradiation with UV light of 254 nm through a mechanism of free radical induced addition. Its derivatives, such as 4-SU, crosslinks with aromatic amino acids when exposed to 365 nm light (**Scheme 1-10**)[42].
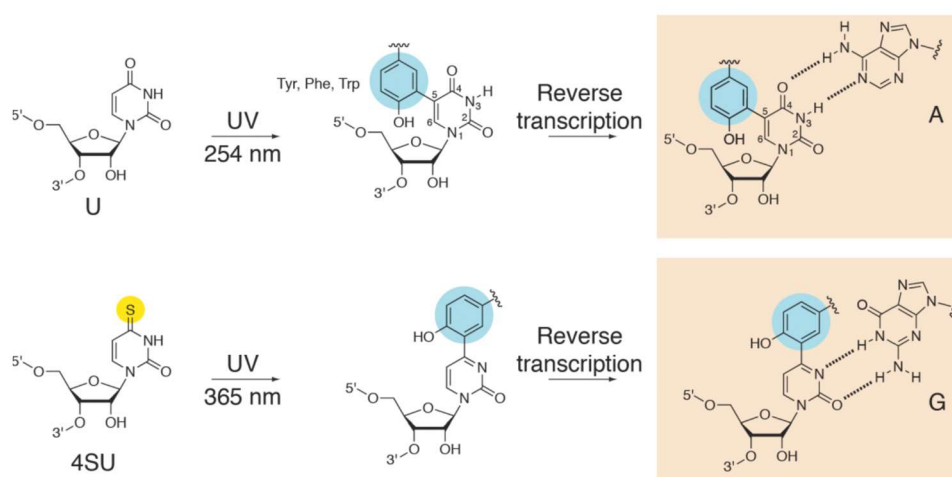


**Scheme 1-10 Principle of protein and RNA crosslinking based on photo-activity of uridine or 4-thiouridine.** The top panel shows the crosslinking of protein and RNA occurring between uridines and amino acids. Uridines and photoactivatable amino acids (Tyr, Phe and Trp) are crosslinked upon 254 nm UV light irradiation, the crosslinking sites hinder the reverse transcription and enable to map RNA and protein interaction sites precisely. The bottom panel shows the process of crosslinking occurring between 4-thiouridine (4-SU) and amino acids. 4-SU and amino acids (Tyr, Phe and Trp) adducts are formed by irradiation with 365 nm UV light, unlike the photo-adduct of uridine, the 4-SU adduct pairs with guanosine after reverse transcription, which leads to T to C transition at the crosslinking site.

The development of CLIP (Crosslinking-immunoprecipitation) techniques is based on photo-induced crosslinking of RNA and the RBP and the crosslinked products are captured by immunoprecipitation. High throughput sequencing is exploited to recognize the crosslinked RNA sequences and interaction sites, which is called CLIP-

seq or HiTS-CLIP. To improve crosslinking activity of RNAs, photoactivatable ribonucleoside analogs such as 4-SU, 5-IU and 6-SG were introduced into organisms. This technique is called PAR-CLIP (Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation). To recognize RBP binding sites in single nucleotide resolution, iCLIP (Individual-nucleotide resolution) is developed as a refinement of CLIP[43-46].

### 1.2.3 RNA and protein crosslinking: HiTS-CLIP

A combination of high throughput sequencing and crosslinking-immunoprecipitation (HiTS-CLIP) is widely used to capture RBP binding RNAs and identification of the binding sites[47, 48]. This method can not only genome-wide detect RBP binding RNAs, but also precisely identify the binding sites in 30-50 bases resolution.

Target protein and RNAs are crosslinked upon UV light irradiation within cells or tissues. In the lysate of the cells, total RNA is partially digested by RNase and the resulting ribonucleoproteins are immunoprecipitated by antibody covered beads. The enriched RNAs are ligated with an adapter on the 3'end, which is enabling reverse transcription and cDNA preparation. The RNA-protein complexes are further purified by radio-labelling with phosphorus isotope [$\gamma$-$^{32}$P]-ATP and electroelution with SDS-PAGE, which are then transferred to nitrocellulose membrane and isolated by autoradiography. The isolated RNA-protein complexes are digested with proteinase K to cleave the peptide bonds to generate RNAs with amino acids on crosslinking sites. 5' linker are introduced to the resulting RNAs, which are reverse transcribed and PCR amplified to generate cDNA, which is finally read out by high through-put sequencing. (**Scheme 1-11**).
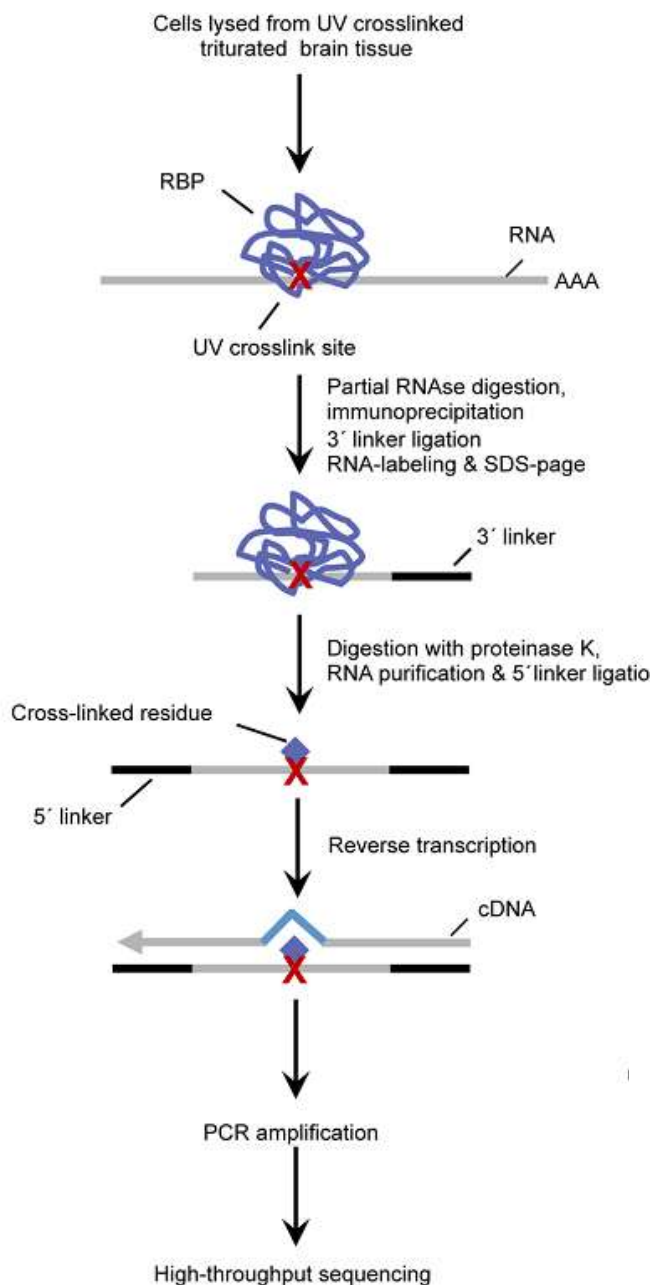
**Scheme 1-11 Scheme of HITS-CLIP procedure.** Brain tissues were irradiated with UV light for protein and RNA crosslinking. The tissues were triturated and the collected cells were lysed. The lysate was treated with RNase and purified by immunoprecipitation. The resulting RNA-protein complex was modified with a 3' adapter which was later used for reverse transcription. The complexes were then radio labeled with $\gamma$-$^{32}$P on 5' end and purified with SDS-PAGE electrophoresis. The complexes were transferred from gel to nitrocellulose membrane and visualized upon exposing to X-ray film. These regions of the membrane were excised to obtain further purified complexes. The resulting RNA-protein complexes were treated with proteinase K to digest RNA binding proteins and only leave amino acid residues on crosslinking site. 5'

linker was introduced to the resulting RNAs as the complimentary sequence of DNA primers used in RT-PCR. The CLIP RNA tags were finally amplified to generated cDNA and analyzed by high through-put sequencing.

## 1.2.4 RNA and protein crosslinking: iCLIP

High-resolution localization of the binding sites of RBP and target RNA is prequisite to achieve precise understanding of protein and RNA interactions. As mentioned above, HiTS-CLIP enables to identify the binding sites in a resolution of 30-50 nt, which is not enough to precisely recognize the real binding sites. To deal with this drawback, a modified CLIP technique was developed to realize single nucleotide resolution, which was defined as iCLIP (individual crosslinking and immunoprecipitation)[49-51]. The iCLIP method has another key advantage when comparing with traditional CLIP methods. In cDNA preparation procedure, a large proportion of reverse transcription is prevented by the crosslinked amino acid residues and these cDNAs are not amplified in the following steps, which causes information missing in sequencing and genome mapping. While in iCLIP method the truncated cDNAs are reserved in PCR step through linearization and restriction enzyme cleavage to introduce 3' and 5' adapter to the cDNAs, which are used for later amplification.

In iCLIP method, protein and RNA binding sites are fixed by 254 nm UV light induced crosslinking. The resulting conjugates are enriched by immunoprecipitation, followed by 3'end adaptor ligation and 5'end radio-labelling. The ligated mixture is size-purified by SDS-PAGE and transferred to nitrocellulose membrane. The target bands are cut from membrane and eluted to afford the purified ribonucleoprotein complexes, which are digested by proteinase K in the following procedure. RNAs containing amino acids on crosslinking sites are then reverse transcribed with a primer that harbours an endonuclease site and a random barcode. Due to the blocking of the residues, a portion of cDNA is truncated near binding sites, which is discarded in HiTS-CLIP for lack of 5'end adaptor. While in iCLIP, the read through and truncated cDNAs are circularized with DNA ligase and linearized with restriction enzyme to form cDNAs ending with restriction sites, which are complementary to the primers used for PCR amplification. The random barcode introduced in reverse transcription is used to discriminate the unique protein binding RNAs from PCR duplicates, and the crosslinking site is theoretically the nucleotide adjacent to the barcode. (**Scheme 1-12**)
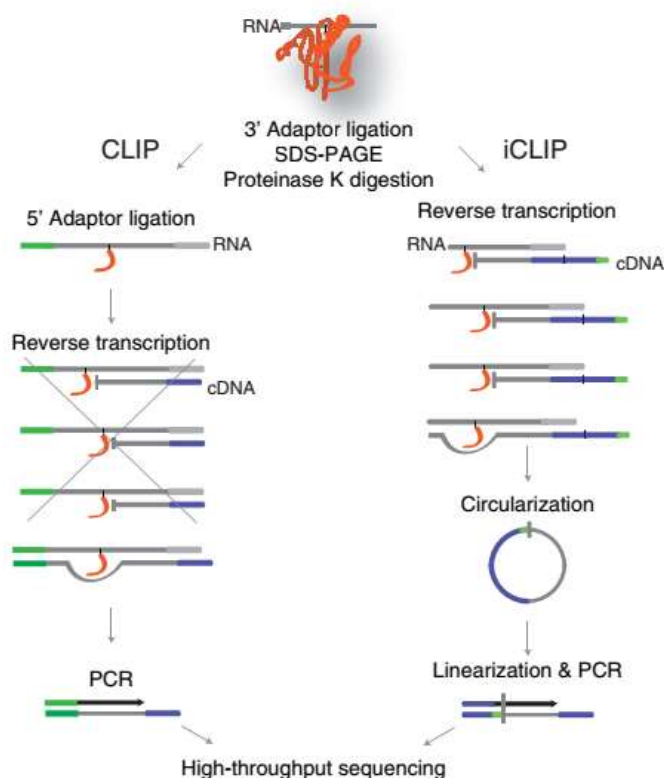
**Scheme 1-12 Comparison of iCLIP with traditional CLIP.** Both methods crosslink target protein to corresponding RNAs through 254 nm UV light irradiation, the protein and RNA complexes are purified by immunoprecipitation and further purified through radio labelling, SDS-PAGE separation and membrane transfer. The resulting RNA and protein complexes are digested with proteinase K to afford RNAs with amino acids residues on crosslinking sites. The main difference of the two methods is the preparation of cDNA. In CLIP and iCLIP，a large proportion of cDNAs are truncated because of the amino acids residues preventing reverse transcriptase reading through. In CLIP method, the truncated cDNAs can not be amplified for lack of a 5' adapter, which is complementary to a PCR primer. As a result, these cDNAs are missed in high throughput sequencing. While in iCLIP, a cleavable (enzyme-cut) adapter is introduced to overcome this problem. Through circularization and linearization, two adapters are formed at the cleavage site, which represents the restriction enzyme site. As a result, the truncated cDNAs are amplified with primers complementary to the cleavage sites. A barcode (green bar) is also contained in this adapter for distinguishing the individual cDNA from PCR duplicates.

## 1.2.5 RNA and protein crosslinking: PAR-CLIP

Photoactivatable ribonucleotide-enhanced crosslinking and immunoprecipitation (PAR-CLIP) is a long wavelength UV (365 nm) induced crosslinking method for studying proteins binding to RNA[42, 52, 53]. Unlike traditional CLIP that is induced by short wavelength UV, photoactive nucleotides like 4-thiouridine (4SU), 4-bromouridine, 5-iodouridine, 5-iodocytosine and 6-thioguanasine are incorporated into cells to substitute the endogenous uridine of RNA (**Figure 1-13**). This improvement not only dramatically increases the crosslinking efficiency by 100-1000 folds, but also introduces the mutation of nucleotides (T to C) on crosslinking sites, which enables identification of the protein-RNA interaction site with single nucleotide resolution.
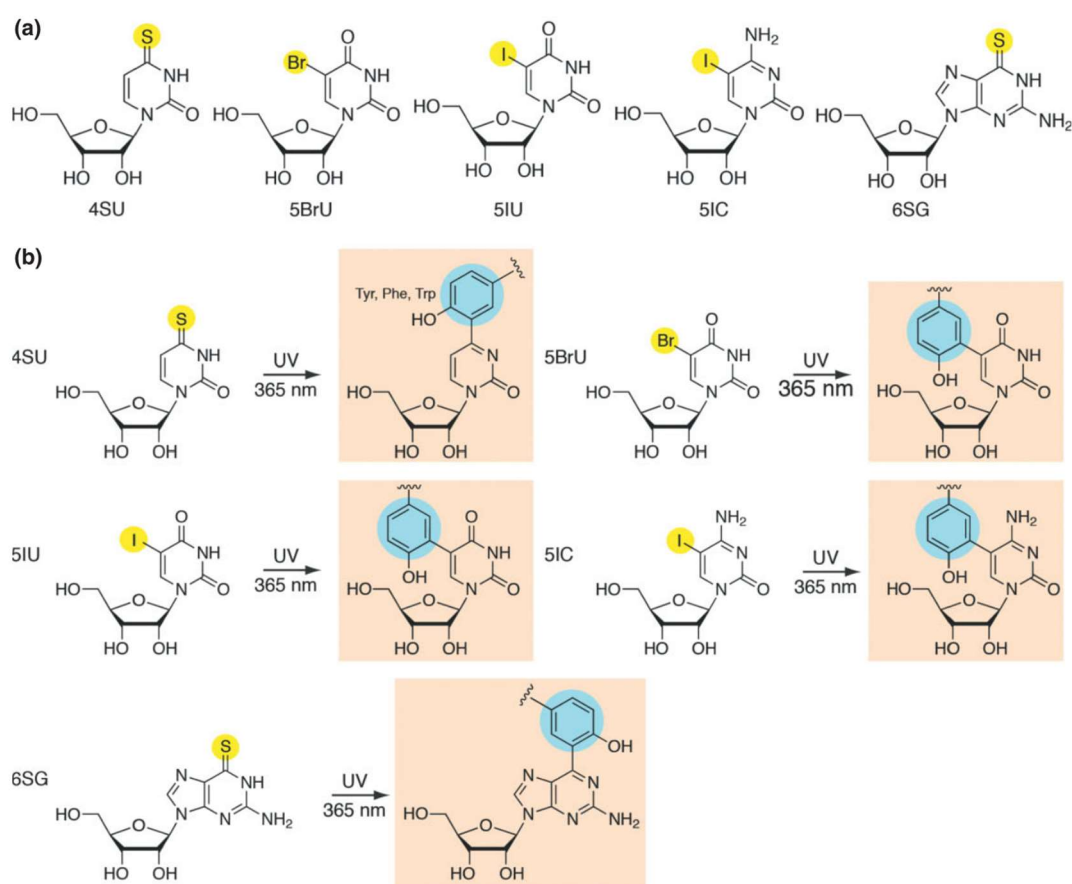


**Figure 1-13 Photoactivatable ribonucleotide analogous applied in PAR-CLIP. a)** Structure of photoactivatable nucleosides. **b)** The 365 nm UV light induced reaction between photoactivatable ribonucleotide analogous and aromatic amino acids.

As the most efficient protein and RNA crosslinker among above mentioned photoactivatable ribonucleotide analogues, 4SU is frequently used in PAR-CLIP. Chemically synthesized 4SU is added to the cells to substitute endogenous uridine of

RNA prior triggering crosslinking with 365 nm light. The fixed ribonucleoproteins are then immuno-precipitated, membrane transferred, size-fractionated and digested as described in the traditional CLIP protocol. Resulting RNAs are 5' ligated, reverse-transcribed and PCR amplified to afford a cDNA library. Sequence data is aligned to associated genome to locate the mutated nucleotide, which represents the RNA and protein binding site (**Scheme 1-14**).
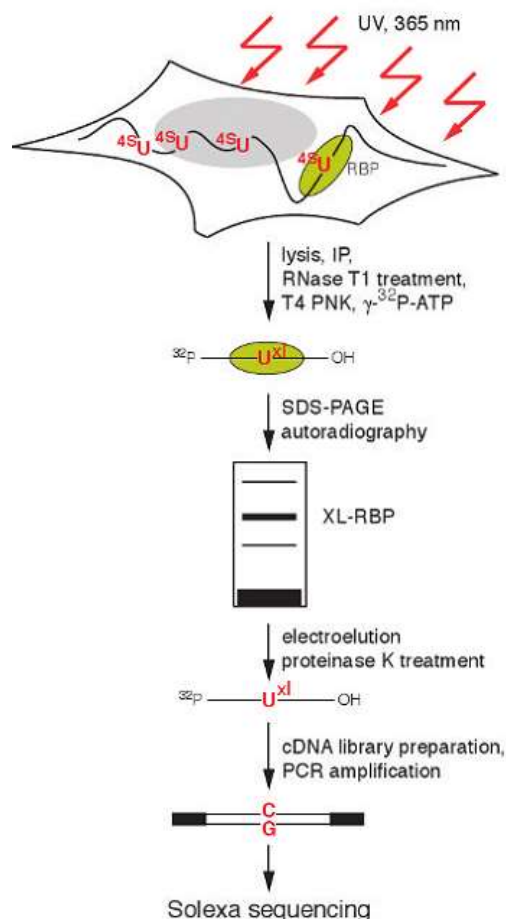


**Scheme 1-14 Illustration of PAR-CLIP.** 4SU-labeled transcripts were crosslinked to RBPs under 365 nm UV light irradiation, cells were lysed and treated with RNase T1 to partially digest RNA. The digested ribonucleoprotein complexes were immune-purified and treated with T4 polynucleotide kinase (PNK) and $\gamma$-$^{32}$P-ATP for radiolabeling of the RNA on 5'end. The mixture was purified by SDS-PAGE and transferred to nitrocellulose membrane, followed by autoradiography and electro-elution. Purified ribonucleoprotein was digested with proteinase K to yield a RNA pool. cDNA library was constructed by reverse transcription and PCR amplification, and sequenced by Solexa. The red letters indicate the crosslinking site, which is concluded from the nucleotide mutation (T-C).

## 1.3 Motivation and thesis overview

The overall goal of the work in this thesis was the development of tools for investigating how RNA is modulating gene expression in living cells. The thesis is divided into two parts dealing with the design of a RNA imaging architecture and a RNA crosslinking system. Background signal is always an important barrier against precise imaging and localization of target RNA. To overcome this, an RNA aptamer/fluorophore probe was developed, which reduced background signal through improving the specificity of target RNA and fluorophore. For RNA crosslinking systems, traditional crosslinkers only capture short range RNA-protein interactions but miss a large portion of long range interactions. A length tunable crosslinker was developed that can both pull down short range and long range interacting RNAs and proteins. In addition, in order to eliminate the crosslinker causing interference with cells, the target protein was incorporated with a photoactivatable unnatural amino acid, which might be used as an endogenous RNA/protein crosslinker.

In **chapter 2**, a powerful aptamer based probe is introduced to image RNA in living cells. In order to improve the specificity of fluorophore and target RNA, a ligand/aptamer1 pair, which has very high binding affinity, is introduced to a reported fluorophore/aptamer2 probe. The molecularly combined new probe (fluorophore-ligand/aptamer1-aptamer2) should have a higher binding affinity than the reported one due to the presence of two binding sites. Experiments in vitro and in vivo finally proved that the specificity and affinity of the new probe apparently improved.

In **chapter 3**, an RNA/RNA and an RNA/protein crosslinker is introduced as a useful tool for studying of RNA-RNA interactions or RNA-protein interactions. The tools described in this chapter have a common character: a length tunable linker, which is designed for capturing both short range and long range interaction. It is proved in vitro that the crosslinkers are able to effectively crosslink RNA/RNA or RNA/protein contacts.

In **chapter 4**, the RNA/protein crosslinker described in **chapter 3** is applied to study RNA and protein interactions in living cells. The main purpose of this chapter is to testify if the new tool has some advantages over traditional methods. The RNAs captured by the crosslinker in vivo are compared with reported data, which are obtained

by a traditional method. The result tells that a combination of the new crosslinker and traditional crosslinker should be an ideal way to study RNA-protein interactions.

In **chapter 5**, a protein is modified with a photoactivatable unnatural amino acid, which could be used as a crosslinking site to study RNA-protein interactions. The unnatural amino acids is incorporated into the protein through expanding the genetic code of bacteria. After three rounds of positive selection and two rounds of negative selection, the unnatural amino acids is site specifically incorporated into the protein.

**References**

1. Battle, A.; Khan, Z.; Wang, S. H.; Mitrano, A.; Ford, M. J.; Pritchard, J. K.; Gilad, Y., Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **2015,** *347* (6222), 664-7.

2. Albert, F. W.; Muzzey, D.; Weissman, J. S.; Kruglyak, L., Genetic influences on translation in yeast. *PLoS Genet* **2014,** *10* (10), e1004692.

3. Wilson, R. C.; Doudna, J. A., Molecular mechanisms of RNA interference. *Annu Rev Biophys* **2013,** *42*, 217-39.

4. Cabili, M. N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.;t Regev, A.; Rinn, J. L., Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **2011,** *25* (18), 1915-27.

5. Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M. C.; Maeda, N.; Oyama, R.; et al. The transcriptional landscape of the mammalian genome. *Science* **2005,** *309* (5740), 1559-63.

6. Tutucci, E.; Livingston, N. M.; Singer, R. H.; Wu, B., Imaging mRNA In Vivo, from Birth to Death. *Annu Rev Biophys* **2018,** *47*, 85-106.

7. Urbanek, M. O.; Galka-Marciniak, P.; Olejniczak, M.; Krzyzosiak, W. J., RNA imaging in living cells - methods and applications. *RNA Biol* **2014,** *11* (8), 1083-95.

8. Xia, Y.; Zhang, R.; Wang, Z.; Tian, J.; Chen, X., Recent advances in high-performance fluorescent and bioluminescent RNA imaging probes. *Chem Soc Rev* **2017,** *46* (10), 2824-2843.

9. Levsky, J. M.; Singer, R. H., Fluorescence in situ hybridization: past, present and future. *J Cell Sci* **2003,** *116* (Pt 14), 2833-8.

10. Bishop, R., Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. *Bioscience Horizons* **2010,** *3* (1), 85-95.

11. Femino, A. M.; Fay, F. S.; Fogarty, K.; Singer, R. H., Visualization of single RNA transcripts in situ. *Science* **1998,** *280* (5363), 585-90.

12. Femino, A. M.; Fogarty, K.; Lifshitz, L. M.; Carrington, W.; Singer, R. H., Visualization of single molecules of mRNA in situ. *Methods Enzymol* **2003,** *361*, 245-304.

13. Bauman, J. G. J.; Wiegant, J.; Borst, P.; van Duijn, P., A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Experimental Cell Research* **1980,** *128* (2), 485-490.

14. Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M., Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell* **1998,** *2* (4), 437-445.

15. Lionnet, T.; Czaplinski, K.; Darzacq, X.; Shav-Tal, Y.; Wells, A. L.; Chao, J. A.; Park, H. Y.; de Turris, V.; Lopez-Jones, M.; Singer, R. H., A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* **2011,** *8* (2), 165-70.

16. Park, H. Y.; Lim, H.; Yoon, Y. J.; Follenzi, A.; Nwokafor, C.; Lopez-Jones, M.; Meng, X.; Singer, R. H., Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* **2014,** *343* (6169), 422-4.

17. Krzywinski, M.; Altman, N., Visualizing samples with box plots. *Nat Methods* **2014,** *11* (2), 119-20.

18. Carrocci, T. J.; Hoskins, A. A., Imaging of RNAs in live cells with spectrally diverse small molecule fluorophores. *Analyst* **2014,** *139* (1), 44-7.

19. Wu, B.; Chao, J. A.; Singer, R. H., Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophys J* **2012,** *102* (12), 2936-44.

20. Hu, C.-D.; Chinenov, Y.; Kerppola, T. K., Visualization of Interactions among bZIP and Rel Family Proteins in Living Cells Using Bimolecular Fluorescence Complementation. *Molecular Cell* **2002,** *9* (4), 789-798.

21. Dictenberg, J., Genetic encoding of fluorescent RNA ensures a bright future for visualizing nucleic acid dynamics. *Trends Biotechnol* **2012,** *30* (12), 621-6.

22. Yiu, H.-W.; Demidov, V. V.; Toran, P.; Cantor, C. R.; Broude, N. E., RNA Detection in Live Bacterial Cells Using Fluorescent Protein Complementation Triggered by Interaction of Two RNA Aptamers with Two RNA-Binding Peptides. *Pharmaceuticals* **2011,** *4* (3), 494-508.

23. Cheong, C. G.; Hall, T. M., Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A* **2006,** *103* (37), 13635-9.

24. Tilsner, J., Pumilio-based RNA in vivo imaging. *Methods Mol Biol* **2015,** *1217*, 295-328.

25. Edwards, T. A.; Pyle, S. E.; Wharton, R. P.; Aggarwal, A. K., Structure of Pumilio Reveals Similarity between RNA and Peptide Binding Motifs. *Cell* **2001,** *105* (2), 281-289.

26. Tyagi, S.; Kramer, F. R., Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **1996,** *14* (3), 303-8.

27. Kim, E.; Yang, J.; Park, J.; Kim, S.; Kim, N. H.; Yook, J. I.; Suh, J. S.; Haam, S.; Huh, Y. M., Consecutive targetable smart nanoprobe for molecular recognition of cytoplasmic microRNA in metastatic breast cancer. *ACS Nano* **2012,** *6* (10), 8525-35.

28. El-Yazbi, A. F.; Loppnow, G. R., Chimeric RNA-DNA molecular beacons for quantification of nucleic acids, single nucleotide polymophisms, and nucleic acid damage. *Anal Chem* **2013,** *85* (9), 4321-7.

29. Nakano, S.; Kanzaki, T.; Sugimoto, N., Influences of ribonucleotide on a duplex conformation and its thermal stability: study with the chimeric RNA-DNA strands. *J Am Chem Soc* **2004,** *126* (4), 1088-95.

30. Paige, J. S.; Wu, K. Y.; Jaffrey, S. R., RNA mimics of green fluorescent protein. *Science* **2011,** *333* (6042), 642-6.

31. Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferre-D'Amare, A. R., Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat Struct Mol Biol* **2014,** *21* (8), 658-63.

32. Strack, R. L.; Disney, M. D.; Jaffrey, S. R., A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat Methods* **2013,** *10* (12), 1219-24.

33. Li, X.; Kim, H.; Litke, J. L.; Wu, J.; Jaffrey, S. R., Fluorophore-Promoted RNA Folding and Photostability Enables Imaging of Single Broccoli-Tagged mRNAs in Live Mammalian Cells. *Angew Chem Int Ed Engl* **2020,** *59* (11), 4511-4518.

34. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J Am Chem Soc* **2014,** *136* (46), 16299-308.

35. Sunbul, M.; Jaschke, A., Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. *Angew Chem Int Ed Engl* **2013,** *52* (50), 13401-4.

36. Arora, A.; Sunbul, M.; Jaschke, A., Dual-colour imaging of RNAs using quencher- and fluorophore-binding aptamers. *Nucleic Acids Res* **2015,** *43* (21), e144.

37. Frobel, S.; Reiffers, A.; Torres Ziegenbein, C.; Gilch, P., DNA Intercalated Psoralen Undergoes Efficient Photoinduced Electron Transfer. *J Phys Chem Lett* **2015,** *6* (7), 1260-4.

38. Engreitz, J. M.; Sirokman, K.; McDonel, P.; Shishkin, A. A.; Surka, C.; Russell, P.; Grossman, S. R.; Chow, A. Y.; Guttman, M.; Lander, E. S., RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **2014,** *159* (1), 188-199.

39. Keene, J. D.; Komisarow, J. M.; Friedersdorf, M. B., RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **2006,** *1* (1), 302-7.

40. Tenenbaum, S. A.; Carson, C. C.; Lager, P. J.; Keene, J. D., Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A* **2000,** *97* (26), 14085-90.

41. Zhao, J.; Ohsumi, T. K.; Kung, J. T.; Ogawa, Y.; Grau, D. J.; Sarma, K.; Song, J. J.; Kingston, R. E.; Borowsky, M.; Lee, J. T., Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **2010,** *40* (6), 939-53.

42. Ascano, M.; Hafner, M.; Cekan, P.; Gerstberger, S.; Tuschl, T., Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* **2012,** *3* (2), 159-77.

43. Li, X.; Song, J.; Yi, C., Genome-wide mapping of cellular protein-RNA interactions enabled by chemical crosslinking. *Genomics Proteomics Bioinformatics* **2014,** *12* (2), 72-8.

44. Lee, F. C. Y.; Ule, J., Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell* **2018,** *69* (3), 354-369.

45. Ule, J.; Jensen, K. B.; Ruggiu, M.; Mele, A.; Ule, A.; Darnell, R. B., CLIP identifies Nova-regulated RNA networks in the brain. *Science* **2003,** *302* (5648), 1212-5.

46. Ule, J.; Jensen, K.; Mele, A.; Darnell, R. B., CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **2005,** *37* (4), 376-86.

47. Licatalosi, D. D.; Mele, A.; Fak, J. J.; Ule, J.; Kayikci, M.; Chi, S. W.; Clark, T. A.; Schweitzer, A. C.; Blume, J. E.; Wang, X.; Darnell, J. C.; Darnell, R. B., HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **2008,** *456* (7221), 464-9.

48. Zhang, C.; Darnell, R. B., Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **2011,** *29* (7), 607-14.

49. Sugimoto, Y.; Konig, J.; Hussain, S.; Zupan, B.; Curk, T.; Frye, M.; Ule, J., Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome biology* **2012,** *13* (8), R67.

50. Haberman, N.; Huppertz, I.; Attig, J.; Konig, J.; Wang, Z.; Hauer, C.; Hentze, M. W.; Kulozik, A. E.; Le Hir, H.; Curk, T.; Sibley, C. R.; Zarnack, K.; Ule, J., Insights into the design and interpretation of iCLIP experiments. *Genome biology* **2017,** *18* (1), 7.

51. Huppertz, I.; Attig, J.; D'Ambrogio, A.; Easton, L. E.; Sibley, C. R.; Sugimoto, Y.; Tajnik, M.; Konig, J.; Ule, J., iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **2014,** *65* (3), 274-87.

52. Hafner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M., Jr.; Jungkamp, A. C.; Munschauer, M.; Ulrich, A.; Wardle, G. S.; Dewell, S.; Zavolan, M.; Tuschl, T., Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **2010,** *141* (1), 129-41.

53. Hafner, M.; Lianoglou, S.; Tuschl, T.; Betel, D., Genome-wide identification of miRNA targets by PAR-CLIP. *Methods* **2012,** *58* (2), 94-105.

# Chapter 2

# Development of a high specificity probe for RNA imaging in mammalian cell

## 2.1 Introduction

RNA imaging is an important method to explore the localization and migration of RNAs in living cells. Even more, it confers a direct insight into RNA function and regulation patterns in the process of transcription and translation. Traditional RNA probes basically recruit small molecule dyes or fuse green fluorescence protein (GFP) to specific RNA binding proteins.[1] Background signal elimination is the most fundamental problem for all imaging techniques. FISH (fluorescence in situ hybridization) as the most frequently used RNA imaging method, can only be used to label fixed cells, besides, the relatively short oligo barcode cannot guarantee 100% specific hybridization with the target sequence.[2, 3] The development of molecular beacons overcomes this drawback to a certain extent, which enables switching on of fluorescence only when probes hybridize to target RNA, while the poor reliability caused by incomplete hybridization limits its application.[4, 5] Although incorporating GFP into target RNA, such as MS2 or PUMILIO01 systems allow spatial and temporal RNA imaging, the bulky modification however may interfere with the function of the target RNA.[6, 7]

To further improve RNA imaging technique in living cells, an RNA aptamer and GFP mimicking probe have been introduced in 2011.[8] An organic molecule (HBI) was designed according to the structure of the chromophore of GFP, which only emits fluorescence when wrapped by the protein environment. A reasonable mechanism was proposed that light illuminated DFHBI is free to rotate to dissipate energy from exited state to ground state. RNA aptamer in this technique plays a role like GFP to constrain the rotation of the chromophore, and as a result, the excited molecule emits fluorescence to release the energy, which means probes are brightened upon binding to the RNA aptamer. The high binding affinity between chromophore molecule and aptamer enabled more specific RNA imaging than traditional methods.

SPINACH is an aptamer generated by systematic evolution of ligands by exponential enrichment (SELEX) for DFHBI, probe is brightened when DFHBI binding to SPINACH. Afterwards, an optimized aptamer was designed to obtain a brighter and more stable RNA probe, which was termed Broccoli (**Figure 2-1d**).[9] To image RNA in vivo, the RNA of interest needs to be fused to the aptamer and transfected into cells which are later treated with DFHBI. Unlike protein and DNA, the abundance of most RNAs is very low in living cells, especially for some noncoding RNAs like microRNA.

Although the Broccoli/DFHBI pair aimed to improve the sensitivity, there is still plenty of space to enhance the specificity.
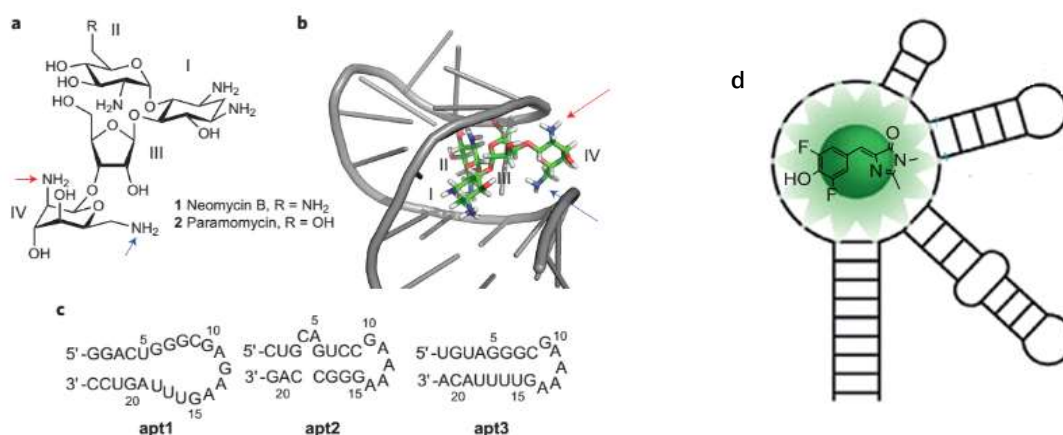


**Figure 2-1 Structure of ligands and aptamers. a** Structure of paromomycin, the red arrow indicates the amine that was modified by DFHBI. **b** The binding mode of the antibiotic with the aptamer, ring I, II and III is wrapped in the groove of the RNA aptamer, and the two amino groups marked by red and blue arrows are exposed, which means modification on these two sites will not significantly affect the binding affinity. **c** The sequence of three aptamers evolved by SELEX against neomycin B. **d** The motif of DFHBI-SPINACH RNA aptamer forms a cavity to suppress the subtle movement of DFHBI. As a result, the non-fluorescent DFHBI is brightened when irradiated.

Here we describe a new RNA probe that is assembled by a joined aptamer and conjugated ligands. To improve the specificity of the new probe compared to previous systems, the binding affinity of ligand and aptamer needs to be enhanced. We introduced paromomycin (**Figure 2-1a-c**)[10] as a second ligand of DFHBI and an aptamer scaffold composed of Broccoli and paromomycin's aptamer (Apt-P). The reasons of choosing paromomycin as a molecular chaperone are listed as follows. **I** The binding affinity of paromomycin/Apt-p is moderate (10 μM)[11] when comparing with DFHBI/SPINACH (537 nM) or DFHBI-1T/Broccoli (360 nM), while the binding affinity of DFHBI-paromomycin (D-P) with the aptamer scaffold (Apt-DP) should be higher than each single pair. **II** Hydrophilicity is very important for the application of small organic molecules in cytobiology, thus the incorporation of paromomycin makes the probe more hydrophilic. **III** Paromomycin contains five positive charged amino groups, which in theory, should enables D-P to penetrate cells more easily (**Table 2-1**).

**Table 2-1 Comparison of Paromomycin and DFHBI in several basic properties**

|  | Paromomycin | DFHBI |
| --- | --- | --- |
| Fluorescence | no | good |
| Cytotoxity | no in low conc. | no |
| Cell penetration | good | moderate |
| Hydrophilicity | good | poor |
| Specificity with aptamer | good | moderate |

We regard this improvement as lifting a bucket (RNA) with two hands (DFHBI and paromomycin), which is stronger than lifting with only one hand (DFHBI) (**Scheme 2-1**).
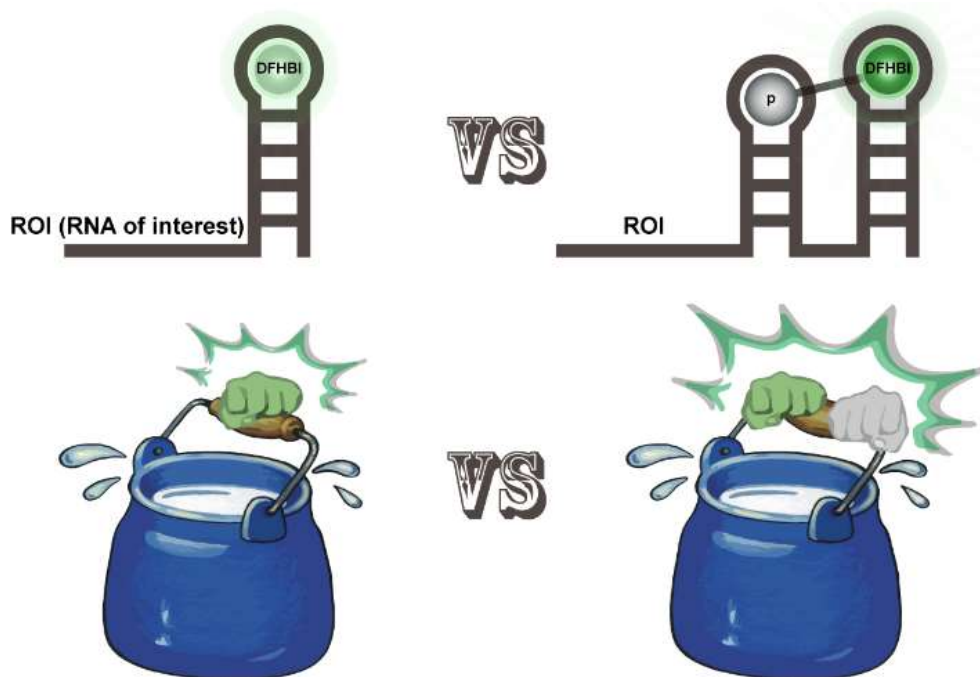


**Scheme 2-1 A metaphor for comparison of single DFHBI and D-P in terms of binding strength with their RNA aptamers**. The grey and green ball indicate paromomycin (P) and DFHBI, which bind to their aptamers, respectively. We metaphor the binding of chromophore and aptamer as grasping and lifting a bucket with one hand or two hands. Comparing to DFHBI/Broccoli (one hand lifting), D-P/Apt-DP (two hands lifting) has a higher binding affinity.

## 2.2 Result and discussion

We synthesized conjugate D-P starting from 2,6-difluorophenol. The construction of the five-membered cyclic ring of compound **2** was realized as previously reported.[12] A functional tail was introduced by attacking the lactone with N-Boc-ethylenediamine. After removal of the Boc group with trifluoroacetic acid (TFA), a N-hydroxysuccinimide (NHS) contained linker was ligated to compound **4**. The obtained compound **5** was then hydrolyzed and activated by NHS, and finally reacted with paromomycin to afford the target molecule D-P. The distinguished pKa of one amino group in paromomycin enables regioselective modification of the bottom amino group under neutral conditions without any need for a protecting manipulation. (**Scheme 2-2**)
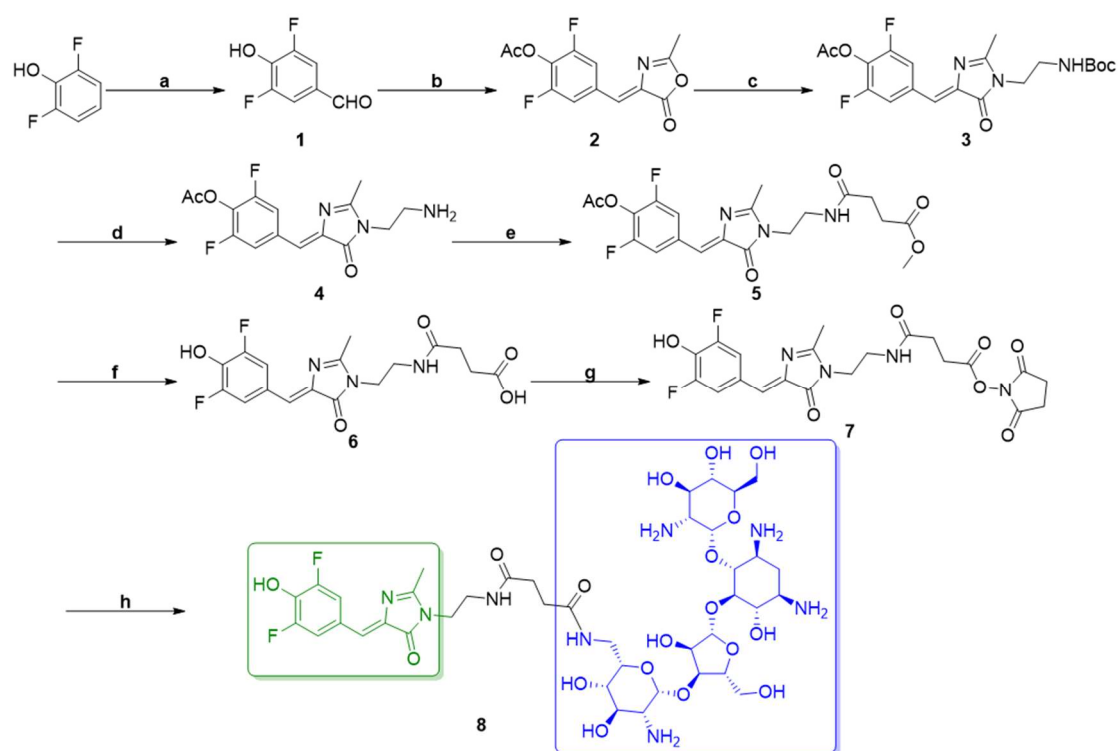


**Scheme 2-2 Synthesis of DFHBI-Paromomycin (D-P).** Conditions: **a)** hexamethylenetetramine, trifluoroacetic acid, rt; **b)** N-acetylglycine, anhydrous Ac₂O, anhydrous NaOAc, ethanol, 0 °C to 100 °C (60%); **c)** N-Boc-ethylenediamine, K₂CO₃, ethanol, rt to reflux (73%); **d)** CF₃COOH, CH₂Cl₂, rt (85%); **e)** 9, Et₃N, dimethylformamide (DMF), rt (80%); **f)** 1M NaOH (aq), MeOH, rt (90%); **g)** N-hydroxy succinimide, EDC·HCl, DMF, rt; **h)** paromomycin, DMF, H₂O (65%).

In consideration of fabricating Apt-DP and Broccoli in a proper way to facilitate their distance within reach of compound D-P, we attached the two aptamers respectively on different arms of a 3 way junction tRNA scaffold (**3WJ**), which was reported to be helpful for assembling and stabilizing aptamer structures.[13] The size and angle of the tRNA arms enables the ligated Apt-P and Broccoli to locate in a proper distance. Besides, the spatial independent arms of tRNA make it possible to assemble two aptamers without mutual affection of RNA folding (**Scheme 2-3**). We synthesized a gene consisting of Apt-P, Broccoli and tRNA scaffold by polymerase chain assembly (PCA), and incorporated it in between T7 promoter and T7 terminator of plasmid pET22b. After PCR amplification of the target gene, the Apt-DP was transcribed in vitro with T7 RNA polymerase. Target RNA was purified with TRIzol and characterized by agarose gel electrophoresis.(**Figure 2-6**)
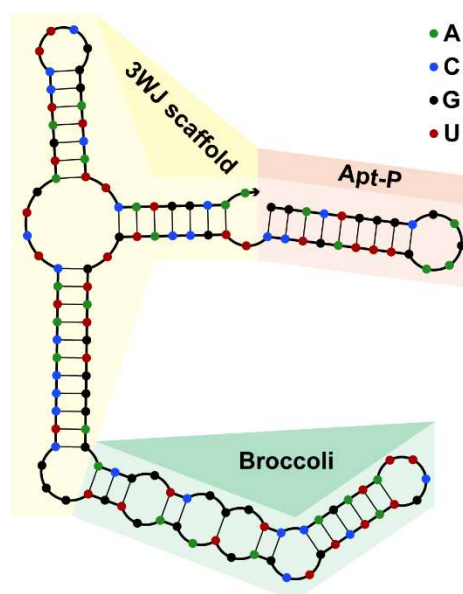


**Scheme 2-3 The NUPACK predicted structure of aptamer scaffold.** The pink and green shadows indicate the aptamer of paromomycin (Apt-P) and Broccoli respectively. Apt-P and Broccoli were separately modified on the two arms of the 3 way junction RNA scaffold (yellow shadow). Apt-P is able to freely swing because of the single stranded joint of Apt-P and tRNA, which enables the two aptamers adjust their mutual distance more flexible.

We measured the absorbance and fluorescence of D-P before and after aptamer Apt-DP binding. The maximum absorbance was shifted from 424 nm to 454 nm upon aptamer binding which is consistent with Broccoli/DFHBI and pair. (**Figure 2-2a** and

**Table 2-2**) The fluorescence can only be detected upon aptamer binding, which make it a fluorescence turn-on pair. As expected, compared with the Broccoli/DFHBI pair, the fluorescence signal of the Apt-DP/D-P pair was significantly enhanced by about 4-fold. (**Figure 2-2b**).   To measure the binding affinity of Broccoli/DFHBI and Apt-DP/D-P pairs, varied concentration (10 nM-10 μM) DFHBI or D-P was incubated with 1 μM Broccoli or Apt-DP, the fluorescence of the complex pair was obtained. (**Figure 2-2c**) The maximum fluorescence at 503 nm was used to calculate the $K_d$. The relative fluorescence at 503 nm and the concentration of fluorophore was fitting using Hill slope. As for Apt-DP/D-P pair, the $K_d$ is 80 nM, which is about 7-fold significant improvement compared to Broccoli/DFHBI pair (537 nM). (**Figure 2-2d** and **Table 2-2**).

Subsequently, we tested whether the introduction of paromomycin will disturb the fluorescence quantum yield of D-P.(**Figure 2-7**) compared to the reported quantum yield of DFHBI/Broccoli (0.72), the relative quantum yield of D-P/Apt-DP is 0.79, which indicates that modification of DFHBI by paromomycin did not cause negative effects in terms of fluorescent quantum yield (**Table 2-2**), Above all, the D-P/Apt-DP pair we developed shows no diminished optical properties with enhanced 5-fold binding affinity.
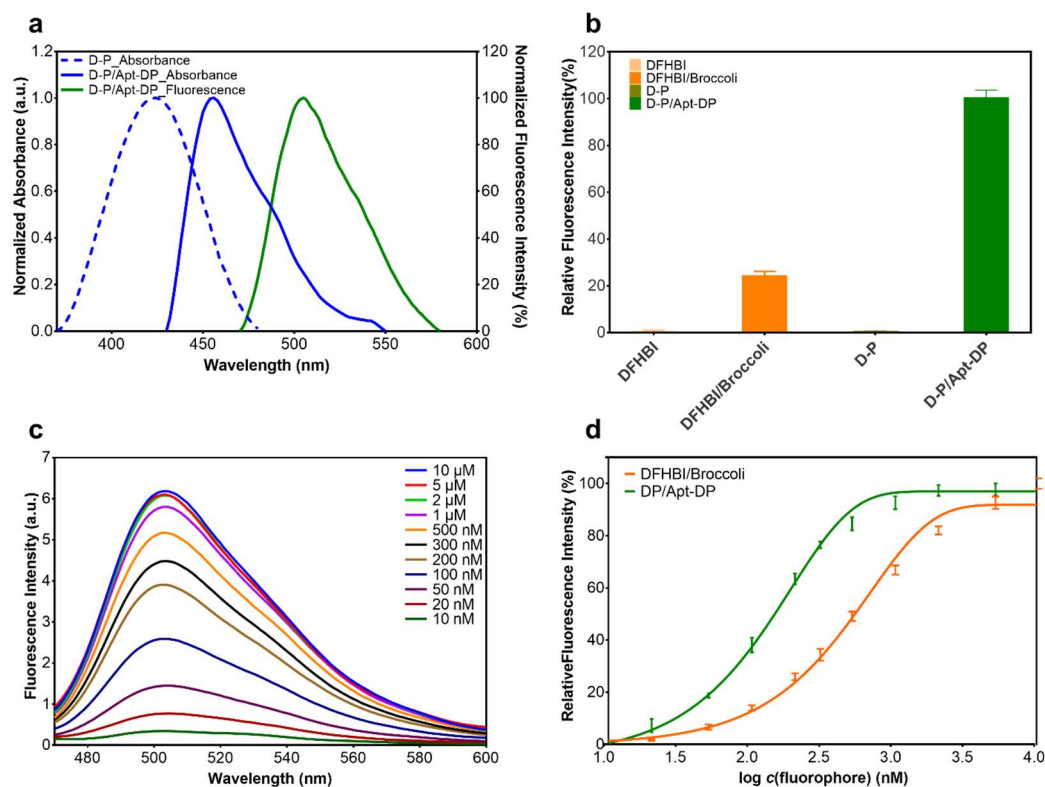
**Figure 2-2. Characterization of the D-P probe in terms of its UV-visible absorption spectra, fluorescence intensity, and binding affinity. a**) The absorption spectra of 1 μM D-P (Blue dashed line) or 1 μM D-P with 5 μM Apt-DP aptamer (Blue solid line), fluorescence spectra of 5 μM D-P with 1 μM Apt-DP (Green solid line) were measured in HEPES buffer containing 40 mM HEPES (pH=7.4), 100 mM KCl and 5 mM MgCl₂. All spectra were normalized to fraction or percentage according to the maximum absorbance or fluorescence. **b**) The relative fluorescence intensity was measured by using 10 nM DFHBI, 10 nM D-P, 10 nM DFHBI with 100 nM Broccoli or 10 nM D-P with 100 nM Apt-DP in HEPES buffer (pH 7.4). **c**) The binding affinity of D-P and Apt-DP was determined by measuring the fluorescence intensity in the presence of 100 nM Apt-DP with varied concentration of D-P (from 10 nM~10 μM) in HEPES buffer (pH 7.4). **d**) Dissociation constant was calculated by fitting the maximum fluorescence at 503 nm with the Hill slope equation in Graphpad Prism.

**Table 2-2 A comparison of D-P/Apt-DP with reported DFHBI probes**

| fluorophore | aptamer | $\lambda ex$ / nm | $\lambda ex$ / nm | $\varPhi$ [a] | Kd / nM |
|---|---|---|---|---|---|
| D-P | Apt-DP | 456 | 503 | 0.79 | 132 |
| DFHBI | Broccoli | 469 | 501 | 0.72 | 660 |

[a] quantum yield.

To calculate the fluorescence quantum yield of D-P/Apt-DP, respective titration of different concentration of DFHBI or D-P was performed in the presence of 5 μM Broccoli or Apt-DP. The resulting data was substituted into the formula $\Phi0/\Phi1=F0A1/F1A0$ to generate the relative fluorescence quantum yield

To examine whether Apt-DP can be observed in living cells, we constructed a plasmid for expressing Apt-DP in mammalian cell. Apt-DP was expressed in HEK 293T and monitored in the presence of DFHBI or D-P. Fluorescence was not detected in DFHBI or D-P treated cells in the absence of Apt-DP expression. When Apt-DP expressing cells were treated with DFHBI, weak fluorescence was observed, but the combination of D-P and Apt-DP afforded apparently stronger fluorescence, as expected (**Figure 2-3**).
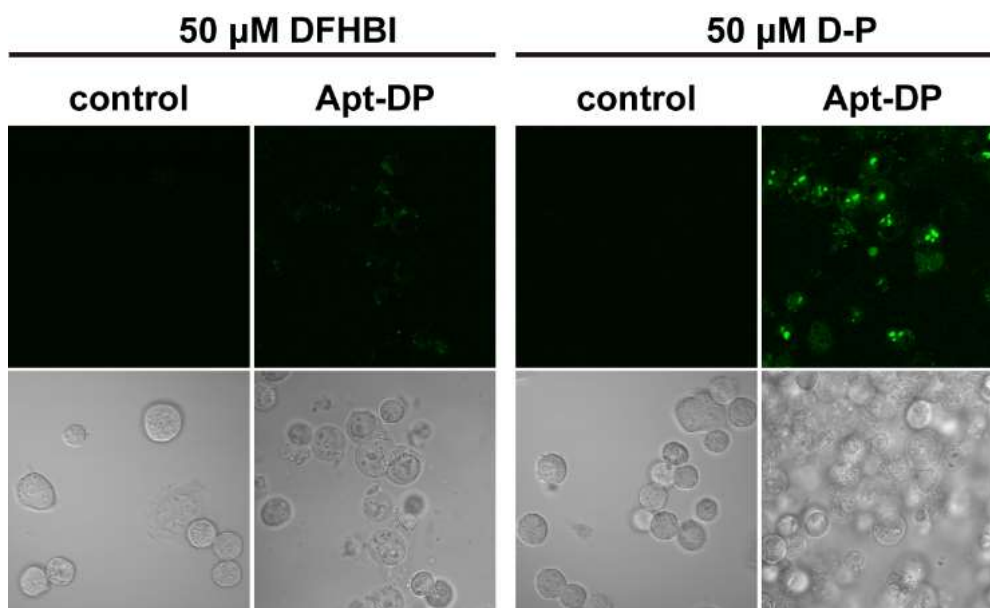
**Figure 2-3**. **RNA aptamer imaging in living cell.** Cells were transfected with target plasmids 36 hours before imaging. Then transfected and mock-treated cells were transferred to imaging buffer and incubated with 50 μM DFHBI or 50 μM D-P 30 minutes before imaging. Fluorescence was visualized upon excitation of 488 nm laser with confocal microscopy.

To further examine whether Apt-DP has superior properties for RNA imaging, we have used a 5S rRNA fusion system to express Broccoli and Apt-DP in living cells.[14] Broccoli or Apt-DP was inserted at the end of the 5S rRNA coding sequence. The plasmids bearing 5S rRNA, 5S_Broccoli and 5S_Apt-DP were transfected into HEK 293T cells and cultured for 2 days. After that, the cells were incubated with 50 μM DFHBI or D-P at 37℃ for 30 min and visualized on confocal microscopy by exciting with 488 nm laser. (**Figure 2-4A**) Fluorescence was not observed either in DFHBI or D-P treated 5S rRNA control groups, while Both DFHBI/Broccoli and D-P/Apt-DP pair shows turn-on fluorescence, and the Apt-DP/D-P treated cells are apparently brighter. To quantify the significance of the difference, we calculated the average fluorescence intensities, the brightness of Apt-DP/D-P is 3-fold higher than Broccoli/DFHBI. Besides, the fluorescence intensity of Broccoli/DFHBI treated cell are not as uniform as Apt-DP/D-P (**Figure 2-4B**). We speculate this may attribute to low level expression of RNA aptamer in some cells, and the binding affinity of Broccoli/DFHBI is not high enough to brighten these cells.
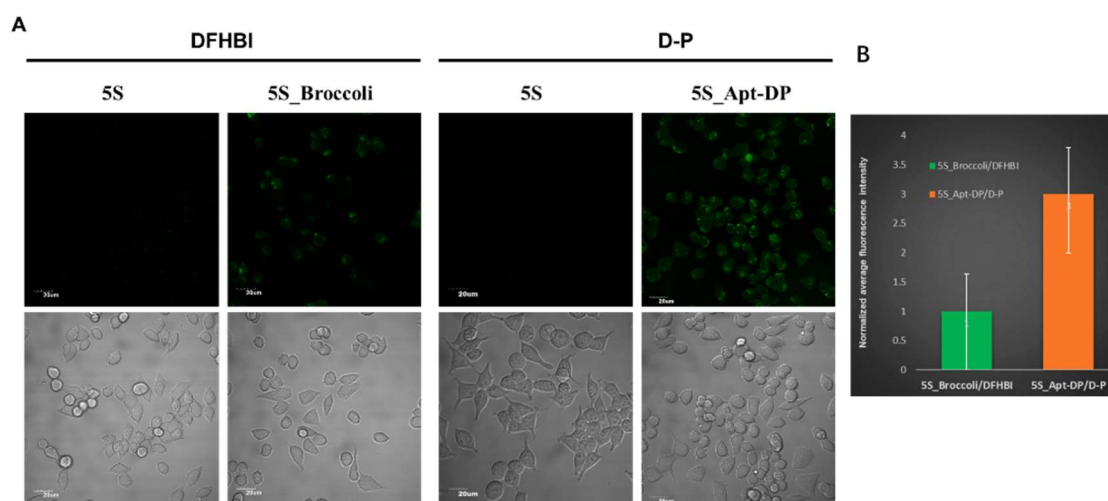


**Figure 2-4. Aptamer fused 5S rRNA imaging in mammalian cells. A**) HEK 293T cells expressing 5S rRNA or 5S_Broccoli were treated with 50 μM DFHBI, expressing 5S rRNA or 5S_Apt-DP were treated with 50 μM D-P. After 30 min incubation, cells were visualized upon

exciting with 488 nm laser on confocal microscopy (upper); **B)** Quantification of average fluorescence intensites of 5S_Broccoli/DFHBI or 5S_Apt-DP/D-P treated cells with ImageJ.

## 2.3 Summary and conclusion

In summary, we have synthesized a new RNA probe based on previously reported aptamers. The chromophore composed of DFHBI and paromomycin enables switching on of fluorescence when binding with the associated aptamer, which reduces background signal to a large extent. Moreover, the incorporation of paromomycin improves the biocompatibility of the probe, and most importantly, enhances the binding affinity with RNA aptamer. We engineered the aptamers of DFHBI and paromomycin on two adjacent arms of a tRNA, which is not only beneficial to the stability of the aptamer structure but also favourable to shorten the distance of the two aptamers without interfering with RNA folding. We lastly realized imaging of the RNA aptamer in mammalian cells, and the fluorescence obtained from newly constructed probe was apparently brighter than the one of the DFHBI probe. In consideration of low abundance of some RNA in living cells, improving the sensitivity and SNR (signal to noise ratio) of a RNA probe is very important. In RNA aptamer based imaging techniques, enhancing the binding affinity between ligand and aptamer is the most direct strategy. Thus the probe we described here should be an ideal way for imaging RNA in living cell.

## 2.4 Experimental section

**Material and methods**

Primers were synthesized by Sangon Biotech (Shanghai) Co., Ltd. Molecular biology enzymes were purchased from New England BioLabs Inc. or Thermo Fisher Scientific. NTPs were purchased from Sangon Biotech (Shanghai) Co., Ltd. RNase inhibitor and TRIzol reagent were purchased from Thermo Fisher Scientific. Other chemicals were purchased from Sigma-Aldrich without further purification. Fluorescence spectra were measured on Varioskan Flash spectral scanning multimode reader (Thermo Scientific) and absorbance spectra were measured on NanoDrop 2000 spectrophotometer. NMR spectra were recorded on Varian Unity Inova (500 MHz). LC/MS (MS: ESI+) was measured on Waters ACQUITY UPLC system connected to a QDa Detector.

**Synthesis of DFHBI-paromomycin (D-P)**

**(Z)-2,6-difluoro-4-((2-methyl-5-oxooxazol-4(5H)-ylidene)methyl)phenyl acetate 2**

Hexamethylenetetramine (1 g, 7.1 mmol) was added to a solution of 2, 6-difluorophenol (5 g, 38.4 mmol) and trifluoroacetic acid (5 ml). After the reaction was stirred at room temperature for overnight, saturated $NaHCO_3$ was added and the solution was extracted with $CH_2Cl_2$. Organic phase was washed with brine, dried with $Na_2SO_4$ and evaporated to give white solid **1**, which was used for the following step without further purification. To a solution of **1** (1.00 g, 6.3 mmol), anhydrous acetic anhydride（5 ml）and N-acetylglycine (0.77 g, 6.3 mmol), anhydrous sodium acetate (0.52 g, 6.3 mmol) was added and stirred at 100 °C for 2 h. Cold ethanol (15 ml) was added after allowing the mixture cooling to room temperature with stirring. The reaction was left stirring overnight at 4 °C. Then the precipitate was washed with cold ethanol, hot water, hexane and dried to give a pale yellow solid (1.05 g, 60%).

$^1$H-NMR (CDCl$_3$, 400 MHz) δ, 2.39 (s, 3H), 2.42 (s, 3H), 6.96 (s, 1H), 7.75 (s, 1H), 7.77 (s, 1H). $^{13}$C NMR (101 MHz, CDCl$_3$) δ 167.71, 166.96, 156.27 (d, $J$ = 4.7 Hz), 153.82 (d, $J$ = 4.8 Hz), 134.47, 131.72 (t, $J$ = 9.5 Hz), 129.03 (t, $J$ = 16.5 Hz), 127.46 (t, $J$ = 3.0 Hz), 115.47 (d, $J$ = 5.6 Hz), 115.24 (d, $J$ = 5.6 Hz), 20.05, 15.71. LC/MS (LC: gradient 10-90% MeOH [0.1% HCO$_2$H] over 15.0 min, 1.2 ml/min flow rate, retention time, 11.29 min; MS (ESI$^+$) ($m/z$): found 281.97 [M+H]$^+$, 314.03 [M+Na]$^+$; calculated, 282.05, [M+H]$^+$, 314.04 [M+Na]$^+$.

**(Z)-tert-butyl(2-(4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-5-oxo-4,5-dihydro-1H-imidazol-1-yl)ethyl)carbamate 3**

N-Boc-ethylenediamine (0.53 g, 3.31 mmol) and potassium carbonate (0.90 g, 6.52 mmol) were added to a solution of compound **1** (0.60 g, 2.14 mmol) in ethanol (10 ml), the reaction was refluxed for 6 h. After cooling to room temperature, the solvent was evaporated and the residual material was re-dissolved in a 1:1 mixture of ethyl acetate and sodium acetate (500 mM, pH 3.0). Afterwards, the organic layer was separated, dried and evaporated. The resulting crude product was purified by silica gel column chromatography and eluted with ethyl acetate: hexane (4:1) to afford **3** as yellow solid (0.60 g, 73%).

$^1$H-NMR (CD$_3$OD, 400 MHz) δ, 1.38 (s, 9H), 2.41 (s, 3H), 3.29 (t, $J$ = 5.6 Hz, 2H), 3.71 (t, $J$ = 5.6 Hz, 2H), 6.87 (s, 1H), 7.74 (s, 1H), 7.77 (s, 1H). $^{13}$C NMR (101 MHz, MeOD) δ 171.89, 165.19, 158.39, 154.69 (d, $J$ = 7.3 Hz), 152.35 (d, $J$ = 7.3 Hz), 138.71,

137.73 (t, $J$ = 16.5 Hz), 126.02 (t, $J$ = 9.3 Hz), 116.37 (d, $J$ = 7.5 Hz), 116.14 (d, $J$ = 7.5 Hz), 80.33, 54.75, 41.85, 39.79, 28.68. LC/MS (LC: gradient 10-90% MeOH [0.1% $HCO_2H$] over 15.0 min, 1.2 ml/min flow rate, retention time, 10.43 min; MS (ESI$^+$) MS (EI$^+$) ($m/z$): found 382.12 [M+H]$^+$, 404.11 [M+Na]$^+$; calculated, 382.15, [M+H]$^+$, 404.14 [M+Na]$^+$.

**(Z)-1-(2-aminoethyl)-4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-1H-imidazol-5(4H)-one 4**

Compound **2** (0.60 g, 1.57 mmol) was dissolved in a solution of trifluoroacetic acid (5 ml) in dichloromethane (5 ml). After stirring at room temperature for 2 h, solvent was evaporated and the mixture was purified by silica gel column chromatography and eluted with chloroform : methanol (10 :1)to afford **4** as yellow solid (0.37 g, 85%).

$^1$H-NMR (CD$_3$OD, 400 MHz) δ, 2.43 (s, 3H), 3.25 (t, $J$ = 6.0 Hz, 2H), 3.96 (t, $J$ = 6.0 Hz, 2H), 6.91 (s, 1H), 7.77 (s, 1H), 7.79 (s, 1H). $^{13}$C NMR (101 MHz, d-DMSO) δ 169.23, 164.14, 153.16 (d, $J$ = 7.4 Hz), 150.68 (d, $J$ = 7.4 Hz), 136.06 (t, $J$ = 17.6 Hz), 124.18 (t, $J$ = 9.2 Hz), 123.42, 115.36 (d, $J$ = 7.3 Hz), 115.14 (d, $J$ = 7.3 Hz), 38.08, 36.95, 15.27. LC/MS (LC: gradient 10-90% MeOH [0.1% $HCO_2H$] over 15.0 min, 1.2 ml/min flow rate, retention time, 6.50 min; MS (ESI$^+$) ($m/z$): found 282.05 [M+H]$^+$; calculated, 282.10, [M+H]$^+$.

**(Z)-methyl-4-((2-(4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-5-oxo-4,5-dihydro-1H-imidazol-1-yl)ethyl)amino)-4-oxobutanoate 5**

Compound **9** (0.62 g, 2.66 mmol) was added to a solution of compound **3** (0.37 g, 1.33 mmol) and triethylamine (0.92 ml, 6.65 mmol) in dimethylformamide (3 ml). The reaction was stirred at room temperature for 4 h. Then 1 M HCl (aq) was slowly added dropwise until pH = 2. The mixture was extracted by dichloromethane and the organic layer was washed with water and brine at least for 4 times. After drying with sodium sulfate, the solvent was removed under reduced pressure. The crude product was further purified by silica gel column chromatography and eluted with ethyl acetate : hexane (4:1) to afford **5**as yellow solid (0.42 g, 80%).

$^1$H-NMR (d-DMSO, 500 MHz) δ, 2.30 (t, $J$ = 7.0 Hz, 2H), 2.33 (s, 3H), 2.46 (t, $J$ = 7.0 Hz, 2H), 3.24 (t, $J$ = 5.9 Hz, 2H), 3.55 (s, 3H), 3.58 (s, $J$ = 5.9 Hz, 2H), 6.88 (s, 1H), 7.91 (s, 1H), 7.99 (s, 1H). $^{13}$C NMR (101 MHz, d-DMSO) δ 172.66, 171.05, 169.71, 163.79, 152.98 (d, $J$ = 7.5 Hz), 150.59 (d, $J$ = 7.3 Hz), 138.12, 135.62 (t, $J$ = 16.9 Hz),

124.74 (d, $J = 9.3$ Hz), 122.67, 115.17 (d, $J = 7.3$ Hz), 114.94 (d, $J = 7.3$ Hz), 51.20, 37.19, 29.73, 28.58, 15.29. LC/MS (LC: gradient 10-90% MeOH [0.1% $HCO_2H$] over 15.0 min, 1.2 ml/min flow rate, retention time, 9.11 min; MS (ESI$^+$) (*m/z*): found 396.09 [M+H]$^+$, 418.08 [M+Na]$^+$; calculated, 396.13, [M+H]$^+$, 418.12 [M+Na]$^+$.

**(Z)-4-((2-(4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-5-oxo-4,5-dihydro-1H-imidazol-1-yl)ethyl)amino)-4-oxobutanoic acid 6**

Compound **4** (0.20 g, 0.51 mmol) was dissolved in a 2:1 mixture (v/v) of methanol and 1M NaOH (aq) and stirred at room temperature for 2 h. Then 1 M HCl (aq) was slowly added dropwise until pH = 1. The reaction mixture was extracted by dichloromethane and the organic layer was washed by brine. The solvent was dried by sodium sulfate and evaporated to afford **5** as yellow solid (173 mg, 90 %).
$^1$H-NMR (d-DMSO, 500 MHz) δ, 2.25 (t, J = 7.0 Hz, 2H), 2.37 (t, J = 7.0 Hz, 2H), 2.39 (s, 3H), 3.24 (t, 5.8 Hz, 2H), 3.61 (t, 5.8 Hz, 2H), 6.94 (s, 1H), 7.92 (s, 1H), 7.94 (s, 1H). $^{13}$C NMR (101 MHz, d-DMSO) δ 173.76, 171.45, 169.76, 164.00, 153.12 (d, $J = 7.7$ Hz), 150.73 (d, $J = 7.7$ Hz), 138.14, 135.68 (d, $J = 18.8$ Hz), 124.79 (d, $J = 9.5$ Hz), 122.77, 115.27 (d, $J = 7.2$ Hz), 115.04 (d, $J = 7.2$ Hz), 37.30, 30.02, 29.00, 15.38. LC/MS (LC: gradient 10-90% MeOH [0.1% $HCO_2H$] over 15.0 min, 1.2 ml/min flow rate, retention time, 8.68 min; MS (ESI$^+$) (*m/z*): found 382.12 [M+H]$^+$, 404.11 [M+Na]$^+$; calculated, 382.11, [M+H]$^+$, 404.10 [M+Na]$^+$.

**Synthesis of DFHBI-Paromomycin 8**

Compound **6** (45 mg, 0.12 mmol), N-hydroxy succinimide (16 mg, 0.14 mmol) and 1-ethyl-3-(3-dimethyllaminopropyl) carbodiimide hydrochloride (27 mg, 0.14 mmol) were dissolved in dimethylformamide (200 μl) and stirred at room temperature overnight. The mixture was extracted by dichloromethane and the organic layer was alternately washed with water and brine (2 times). Solvent was dried by sodium sulfate and evaporated under reduced pressure to afford a yellow solid. The resulting product was re-dissolved in a solution of water (2 ml), and was then added dropwise to a solution of paromomycin (738 mg, 1.2 mmol) in dimethylfomamide (1 ml). The reaction was stirred at room temperature overnight. After evaporating the solvent under vacuum, the crude product was purified by column chromatography and eluted by DCM : MeOH : ammonium (2:2:1) to afford the product as an orange solid. $^1$H-NMR ($D_2O$, 400 MHz) δ, 1.87-1.91 (m, 1H), 2.30 (s, 3H), 2.34-2.50 (m, 4H), 3.03-3,27 (m,

6H), 3.28-3.35 (m, 2H), 3.40-3.50 (m, 4H), 3.53-3.60 (m, 2H), 3.61-3.69 (m, 2H), 3.70-3.93 (m, 7H), 4.03-4.20 (m, 4H), 5.02 (s, 1H), 5.26 (s, 1H), 5.57 (s, 1H), 6.87 (s, 1H), 7.46-7.60 (m, 2H). $^{13}$C NMR (101 MHz, D$_2$O) δ 179.30, 174.74, 172.64, 171.26, 161.30, 156.46, 154.10, 149.59, 132.94, 115.73, 109.44, 96.63, 84.88, 81.84, 81.20, 76.45, 73.96, 73.69, 73.39, 72.84, 70.01, 69.57, 68.17, 66.76, 60.57, 54.46, 51.41, 50.42, 49.34, 40.43, 39.82, 37.67, 32.73, 31.73, 30.78, 29.21, 24.98, 14.38.
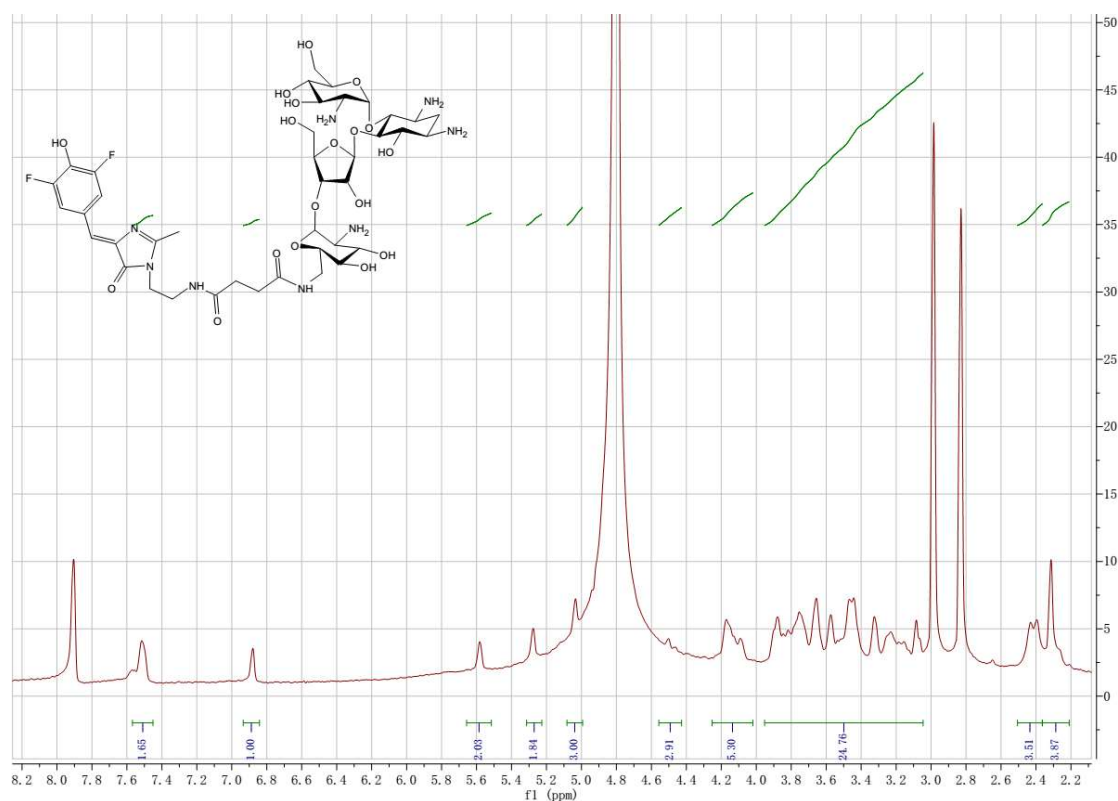


**Figure 2-4 $^1$H-NMR (d-H$_2$O, 400 MHz) spectrum of DFHBI-Paromomycin**

**Construction of RNA aptamer expression plasmids**

To obtain the aptamers in vitro by DNA transcription with T7 RNA polymerase, the two DNA templates 3WJ-Broccoli and Apt-DP were constructed into pET22b vector flanked by T7 promoter and T7 terminator. To express Apt-DP in mammalian cells, the plasmid was constructed by introducing the DNA template that was flanked by U6 promoter and poly-U terminator (**Figure 2-5**). All the sequences were confirmed by DNA sequencing.

**A. pET22b_T7_3WJ-Broccoli**

```
TAATACGACTCACTATAGGGTTGCCATGTGTATGTGGGAGACGGTCGGGTCCAGATATTCGTATCTGTCGAGTAGAGTGTGGGCTC
├─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┤
ATTATGCTGAGTGATATCCCAACGGTACACATACACCCTCTGCCAGCCCAGGTCTATAAGCATAGACAGCTCATCTCACACCCGAG
```
T7 promoter | 3WJ scaffold | Broccoli

```
CCACATACTCTGATGATCCTTCGGGATCATTCATGGCAAGCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTT
├┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┤
GGTGTATGAGACTACTAGGAAGCCCTAGTAAGTACCGTTCGATCGTATTGGGGAACCCCGGAGATTTGCCCAGAACTCCCCAAAAA
```
3WJ scaffold | T7 terminator

```
TG
├─┤
AC
□
```

**B. pET22b_T7_Apt-DP**

```
TAATACGACTCACTATAGGGGGGACTGGGCGAGAAGTTTAGTCCGTCATTGCCATGTGTATGTGGGAGACGGTCGGGTCCAGATATT
├─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┤
ATTATGCTGAGTGATATCCCCCTGACCCGCTCTTCAAATCAGGCAGTAACGGTACACATACACCCTCTGCCAGCCCAGGTCTATAA
```
T7 promoter | Apt-P | 3WJ scaffold

```
CGTATCTGTCGAGTAGAGTGTGGGCTCCCACATACTCTGATGATCCTTCGGGATCATTCATGGCAAGCTAGCATAACCCCTTGGGG
├┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┤
GCATAGACAGCTCATCTCACACCCGAGGGTGTATGAGACTACTAGGAAGCCCTAGTAAGTACCGTTCGATCGTATTGGGGAACCCC
```
Broccoli | 3WJ scaffold

```
CCTCTAAACGGGTCTTGAGGGGTTTTTG
├┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┤
GGAGATTTGCCCAGAACTCCCCAAAAAAC
```
T7 terminator

**C. pU6_Apt-DP**

```
GAGGGCC....TGGAAAGGACGAAACACCGGACTGGGCGAGAAGTTTAGTCCGTCATTGCCATGTGTATGTGGGAGACGGTCGGGT
├─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┤
CTCCCGG....ACCTTTCCTGCTTTGTGGCCTGACCCGCTCTTCAAATCAGGCAGTAACGGTACACATACACCCTCTGCCAGCCCA
```
U6 promoter | Apt-P | 3WJ scaffold

```
CCAGATATTCGTATCTGTCGAGTAGAGTGTGGGCTCCCACATACTCTGATGATCCTTCGGGATCATTCATGGCAATTTTTT
├┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┼─┼─┼─┼┼─┼─┤
GGTCTATAAGCATAGACAGCTCATCTCACACCCGAGGGTGTATGAGACTACTAGGAAGCCCTAGTAAGTACCGTTAAAAAA
```
Broccoli | 3WJ scaffold | poly-U terminator

**Figure 2-5 Gene sequences of constructed aptamers for in vitro (A), in E. coli. (B) and in mammalian cell (C) transcription.**

**In vitro transcription of RNA aptamers**

Template DNAs for RNA aptamer transcription were obtained by PCR, separated by Omega gel extraction kit and eluted by nuclease-free water. The concentrations were measured by a NanoDrop spectrophotometer according to 260 nm absorbance.

T7 RNA polymerase was previously purified as a lab reserve with test and verification. All the water or buffer used in RNA transcription was DEPC treated. The conditions of RNA aptamer transcription are listed in the following. The reaction mixture was incubated at 37°C for 4 h.

| Component | 500 µL Reaction | Final Concentration |
|---|---|---|
| 500 mM HEPES (pH=7.5) | 80 µL | 80 mM |
| 200 mM MgCl$_2$ | 50 µL | 20 mM |
| 100 mM spermidine | 5 µL | 1 mM |
| 200 mM DDT | 25 µL | 10 mM |
| 20 mM NTPs | 100 µL | 4 mM |
| DNA template | varied volume | pmol |
| T7 RNA polymerase | 50 µL | - |
| RNase inhibitor | 0.5 µL | 40 U/ml |
| Nuclease-free water | to 500 µL | |

Template DNAs were digested by adding 25 µL DNase I at 37°C for 30 min. RNA aptamers were extracted by TRIzol reagent following the manufacture's protocol. The purified RNA was dissolved in nuclease-free water and stored at -20°C before use (**Figure 2-6**).
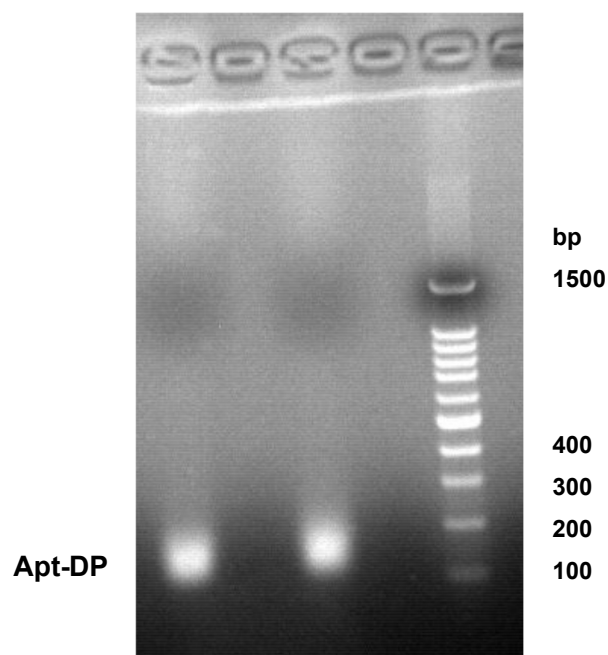


**Figure 2-6 Agarose gel electrophoresis of in vitro transcribed Apt-DP**

**Fluorescence measurement in vitro**

The PCR amplified template DNA was in vitro transcribed to generate aptamer Apt-DP and Broccoli. 100 nM Apt-DP or Broccoli was treated with 1µM D-P or Broccoli

respectively in 25 mM HEPES buffer (pH 7.4), which was then transferred to a cuvette for fluorescence measurement.

**Relative quantum yield measurement**

To calculate fluorescent quantum yield of Apt-DP/D-P, respective titration of different concentration of DFHBI or D-P was performed in the presence of 5 μM Broccoli or Apt-DP (**Figure 2-7**). The resulted data was substituted into the formula $\Phi_0/\Phi_1 = F_0 A_1 / F_1 A_0$ to generate the relative fluorescent quantum yield.
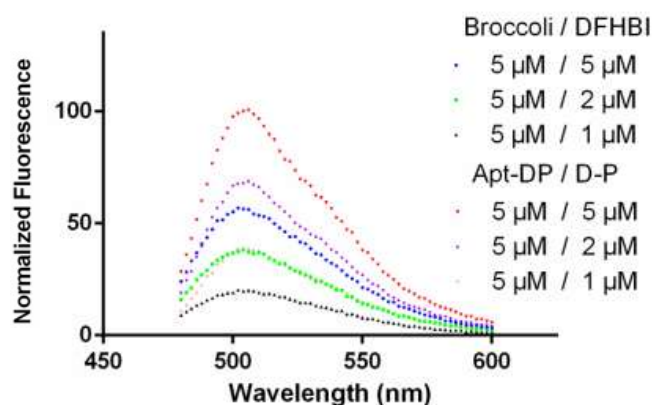


**Figure 2-7 Titration of different concentration of DFHBI or D-P in the presence of 5 μM Broccoli or Apt-DP.**

**Binding affinity measurement**

The dissociation constant (Kd) of aptamer and chromophore was obtained by measuring the fluorescence of the complex in the presence of fixed concentration (100 nM) of aptamer and increasing concentration (1 nM-10 μM) of the chromophore. Curves were determined by using nonlinear regression analysis and fitted by Hill slope equation on GraphPad Prism software.

**Apt-DP imaging in mammalian cells**

HEK 293T cells were cultured in DMEM supplemented with 10% FBS (fetal bovine serum) at 37°C, 5%(v/v) $CO_2$ for 16-24 h till ~80% confluency. Afterwards, the cells were transfected with pU6_Apt-DP by using Lipofectamine 2000 following the manufacturer's protocol. 24 h prior to imaging, the cells were trypsin digested and passaged to confocal dishes. The culture medium was replaced by imaging buffer (10 mM HEPES, pH 7.4, 125 mM NaCl, 5 mM $MgCl_2$, 1.5 mM $CaCl_2$) supplemented with

50 µM DFHBI or 50 µM D-P or vehicle 30 min before imaging. The living cell imaging was performed on Olympus FV1000 confocal microscope and excited by a 488 nm laser. Fluorescence images were taken through 60x oil immersion objective mounted on Olympus Inverted IX81 microscope and analyzed with FV10-ASW2.0 software.

**Construction of pcDNA3.1_5S, pcDNA3.1_5S_F30-Broccoli and pcDNA3.1_5S_Apt-DP**

The 5S and its fusion genes, 5S_F30-Broccoli and 5S_Apt-DP were synthesized by Nanjing Genescript and cloned into a pcDNA3.1 vector with MfeI and BglII restriction enzymes. All the sequences were confirmed by DNA sequencing. The sequence information was listed below.

**Gene name: 5S**

**Vector name: pcDNA3.1(+);**

**Cloning site: MfeI/BglII**

**Sequence:**

CAATTGCCCCGGGCTGGCGGTGTCGGCTGCAATCCGGCGGGCACGGCCGGGCCGGGCTGGGCTCTTGGGGCAGCCAGGCGCCTCCTTCAGCGTCTACGGCCATACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGTCGGGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAGGCGTCGACTCTAGAGCGGACTTCGGTCCGCTTTTTTAGATCT

**Gene name: 5S_F30-Broccoli**

**Vector name: pcDNA3.1(+);**

**Cloning site: MfeI/BglII**

**Sequence:**

CAATTGCCCCGGGCTGGCGGTGTCGGCTGCAATCCGGCGGGCACGGCCGGGCCGGGCTGGGCTCTTGGGGCAGCCAGGCGCCTCCTTCAGCGTCTACGGCCATACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGGGTCGGGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCGGGTGCTGTAGGCGTCGACTTGCCATGTGTATGTGGGAGACGGTCGGGTCCAGATATTCGTATCTGTCGAGTAGAGTGTGGGCTCCCACATACTCTGATGATCCTTCGGGATCATTCATGGCAATCTAGAGCGGACTTCGGTCCGCTTTTTTAGATCT

**Gene name: 5S_Apt-DP**

**Vector name: pcDNA3.1(+)**

**Cloning site: MfeI/BglII**

**Sequence:**

CAATTGCCCCGGGCTGGCGGTGTCGGCTGCAATCCGGCGGGCACGGCCGG
GCCGGGCTGGGCTCTTGGGGCAGCCAGGCGCCTCCTTCAGCGTCTACGGC
CATACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGG
GTCGGGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCGGGTGC
TGTAGGCGTCGACGGACTGGGCGAGAAGTTTAGTCCGTCATTGCCATGTGT
ATGTGGGAGACGGTCGGGTCCAGATATTCGTATCTGTCGAGTAGAGTGTGG
GCTCCCACATACTCTGATGATCCTTCGGGATCATTCATGGCAATCTAGAGCG
GACTTCGGTCCGCTTTTTTAGATCT

Gray: 5S RNA promoter; Green: 5S rRNA; Blue: Broccoli; Magenta: Apt-DP.

**Detection of 5S rRNA-fused aptamers transcription in mammalian cells by reverse transcription-polymerase chain reaction (RT-PCR)**

HEK293T cells were cultured to a confluence of 90% in DMEM medium supplemented with 10%(v/v) FBS. The plasmids pcDNA3.1_5S, pcDNA3.1_5S_Broccoli and pcDNA3.1_5S_Apt-DP were transfected into the cells respectively and 24 hours later total RNAs of these cells were extracted by TRIZol reagent (Ambion, #15596018). In order to exclude plasmid contaminants from RNA samples, the samples were digested by the RQ1 RNase-Free DNase (Promega, #M6101) at 37°C for 30 min and then extracted by TRIZol reagent again. Reverse transcription of the RNAs samples was carried out by PrimeScript RT-PCR Kit (Takara, ##RR014A) following the manufacturer's protocol. The RT products were further amplified by PCR with primers listed below. 2%(w/v) agarose gel electrophoresis was used to identify the PCR products.(**Figure 2-8**)

**Primers:**

**5S_RNA_RT F**: GATCTCGTCTGATCTCGGAAGCTAAGC

**5S_RNA_RT R**: CGGACCGAAGTCCGCTCTAG
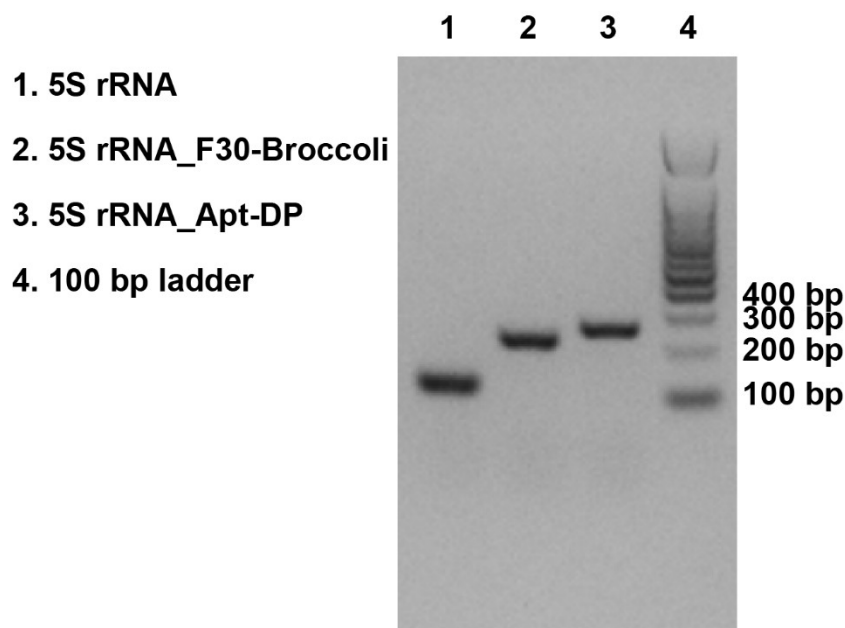
**Figure 2-9**. **Amplification of 5S_Broccoli and 5S_Apt-DP by RT-PCR.** RT-PCR was performed with the total RNAs extracted from HEK293T cells transfected with pcDNA3.1_5S (Lane 1), pcDNA3.1_5S_F30-Broccoli (Lane 2) and pcDNA3.1_5S_Apt-DP (Lane 3) plasmids.

**5S rRNA imaging in mammalian cells**

HEK 293T cells were cultured in DMEM supplemented with 10% FBS (fetal bovine serum) at 37°C, 5% (v/v) $CO_2$ for 16-24 h till ~80% confluency. Afterward, the cells were transfected with pU6_Apt-DP by using Lipofectamine 2000 following the manufacturer's protocol. 24 h prior to imaging, the cells were trypsin digested and passaged to confocal dishes. The culture medium was replaced by imaging buffer (10 mM HEPES, pH 7.4, 125 mM NaCl, 5 mM $MgCl_2$, 1.5 mM $CaCl_2$) supplemented with 50 μM DFHBI or 50 μM D-P or vehicle 30 min before imaging. The living cell imaging was performed on Olympus FV1000 confocal microscopy and excited by 488 nm laser. Fluorescence images were taken through a 60x oil objective and analyzed with FV10-ASW2.0 software.

**Quantification of average fluorescence intensity**

The living cell 5S rRNA imaging fluorescence intensity was quantified by manually selecting every cell in bright field and measured their fluorescence in imageJ.

## References

1.    Urbanek, M. O.;   Galka-Marciniak, P.;   Olejniczak, M.; Krzyzosiak, W. J., RNA imaging in living cells - methods and applications. *RNA Biol* **2014,** *11* (8), 1083-95.

2.    Singer, R. H.; Ward, D. C., Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *Proceedings of the National Academy of Sciences* **1982,** *79* (23), 7331-7335.

3.    Levsky, J. M.; Singer, R. H., Fluorescence in situ hybridization: past, present and future. *J Cell Sci* **2003,** *116* (Pt 14), 2833-8.

4.    Kolpashchikov, D. M., An elegant biosensor molecular beacon probe: challenges and recent solutions. *Scientifica (Cairo)* **2012,** *2012*, 928783.

5.    Sokol, D. L.;   Zhang, X.;   Lu, P.; Gewirtz, A. M., Real time detection of DNA.RNA hybridization in living cells. *Proc Natl Acad Sci U S A* **1998,** *95* (20), 11538-43.

6.    Bertrand, E.;   Chartrand, P.;   Schaefer, M.;   Shenoy, S. M.;   Singer, R. H.; Long, R. M., Localization of ASH1 mRNA particles in living yeast. *Mol Cell* **1998,** *2* (4), 437-45.

7.    Ozawa, T.;   Natori, Y.;   Sato, M.; Umezawa, Y., Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* **2007,** *4* (5), 413-9.

8.    Paige, J. S.;   Wu, K. Y.; Jaffrey, S. R., RNA mimics of green fluorescent protein. *Science* **2011,** *333* (6042), 642-6.

9.    Filonov, G. S.;   Moon, J. D.;   Svensen, N.; Jaffrey, S. R., Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J Am Chem Soc* **2014,** *136* (46), 16299-308.

10.    Bastian, A. A.;   Rodriguez-Pulido, A.;   Gruszka, A.;   Gerasimov, J. Y.; Herrmann, A., Probing the shielding properties of aptameric protective groups. *Chem Asian J* **2014,** *9* (8), 2225-31.

11.    Osborne, S. E.; Ellington, A. D., Nucleic Acid Selection and the Challenge of Combinatorial Chemistry. *Chem Rev* **1997,** *97* (2), 349-370.

12.    Song, W.;   Strack, R. L.;   Svensen, N.; Jaffrey, S. R., Plug-and-play fluorophores extend the spectral properties of Spinach. *J Am Chem Soc* **2014,** *136* (4), 1198-201.

13.    Shu, D.;   Khisamutdinov, E. F.;   Zhang, L.; Guo, P., Programmable folding of fusion RNA in vivo and in vitro driven by pRNA 3WJ motif of phi29 DNA packaging motor. *Nucleic Acids Res* **2014,** *42* (2), e10.

14.    Paul, C. P.;   Good, P. D.;   Li, S. X.;   Kleihauer, A.;   Rossi, J. J.; Engelke, D. R., Localized expression of small RNA inhibitors in human cells. *Mol Ther* **2003,** *7* (2), 237-47.

# Chapter 3

## Psoralen based photoactive crosslinkers development for in vitro crosslinking of RNA & RNA or RNA & protein

## 3.1 Introduction

Psoralen is well known as a drug for treating psoriasis, eczema, vitiligo, and cutaneous T-cell lymphoma in combination with ultraviolet radiation, which is named PUVA (psoralen + UVA)[1]. Inspired by the photo-activity of psoralen, a class of psoralen derivatives were developed for investigation of bio-macromolecular interactions [2-5]. Psoralen was reported to have a high reactivity with nucleic acids upon 365 nm UV light irradiation, especially with uridine and thymidine. The mechanism of how psoralen react with nucleic acids is well understood[6]. The photo-induced reaction is realized through a three step process[7]. Firstly, a planar psoralen intercalates the groove of DNA or double stranded RNA. In the second step, the double bond of the furan ring of psoralen and double bond of pyrimidine undergoes a 2+2-cycloaddition upon irradiation of UV light (365 nm) to form a monoadduct. Lastly, a similar 2 + 2 cycloaddition occurs between pyrone ring of psoralen and another pyrimidine to form a diadduct (**Scheme 3-1**). In addition, the photo induced crosslinking can be completely reversed by irradiation of short wavelength UV light (254 nm).
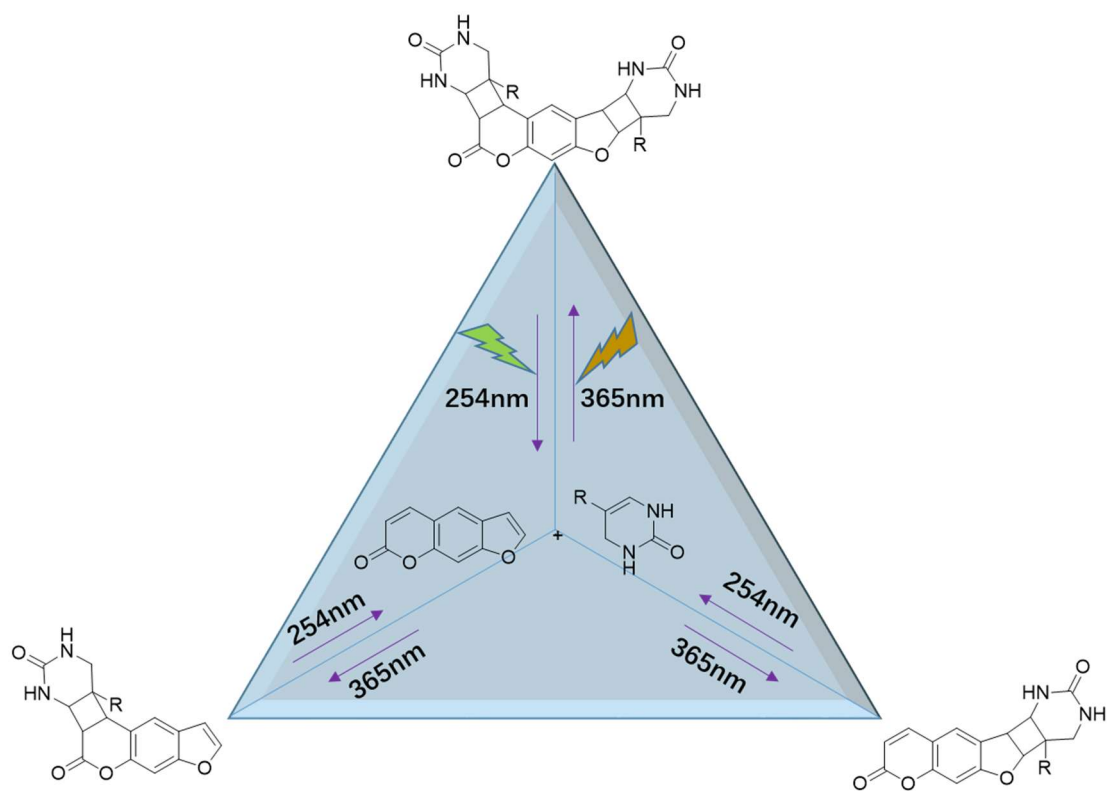
**Scheme 3-1 Photo induced reaction between psoralen and pyrimidine.** The cycloaddition occurs between furan side and pyrimidine or pyrone side and pyrimidine when triggered by 365 nm UV light. Monoadduct forms as long as a psoralen locates close to a pyrimidine, while the diadduct forms on the premise that a psoralen embeds into two staggered pyrimidines. R = H or $NH_2$, indicating uridine or thymidine. Crosslinking can be reversed by 254 nm UV light irradiation.

The exploration of a variety of photoreactive psoralen derivatives enabled wide application of photocrosslinking methods in study of nucleic acid structure and functionality. 8-methoxypsoralen (8-MOP) and 4'-aminomethyltrioxsalen (AMT) were designed in consideration of introducing electron-donating groups to improve quantum yield[8]. Especially for AMT, the modification of three methyl groups and aminomethyl group not only improves the photoreactivity of the compound, but also makes it more hydrophilic and permeable to cell membranes, which is critical for application in vivo (**Scheme 3-2**).
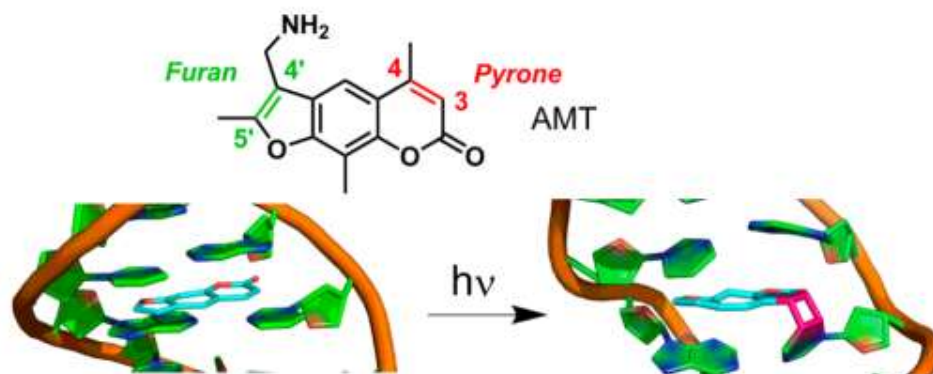
**Scheme 3-2 AMT crosslinking model. Green and red double bonds indicate the photo reactive bonds**. Similar to psoralen, AMT firstly intercalates into duplex of double stranded nucleic acid, and then undergoes cycloaddition between furan and pyrimidine or pyrone and pyrimidine upon UV irradiation.

Here, we developed an AMT based crosslinker for the purpose of capturing long range nucleic acid interactions (**Scheme 3-3**). Although AMT is widely used in study of RNA-RNA interactions, limitations are still obvious. Normally, non-coding RNA regulates transcription by hybridizing with target RNA. Traditional RNA crosslinkers like psoralen capture the interacting mRNA/non-coding RNA pair by ligating both of their uridines at the same time, which require staggered uridines existing at the hybridization region. The application of these small molecule crosslinker was limited by the weak capability of ligating long range uridines. Moreover, RNA-RNA interactions are mostly mediated by protein factors, which may prevent crosslinkers getting close to the hybridization region.

We here synthesized an AMT dimer which was linked by a length tunable linker. We assumed that two AMT could respectively react with uridines of two interacting RNAs，even if these uridines are not closely located. Moreover, the crosslinked uridines are not limited in hybridization region since long linker assembled AMT could reach uridines beyond the region. In addition, for the bulk protein mediated RNA-RNA interaction, steric hindrance could be overcome by the length tunable linker. We designed two oligomeric nucleic acids as a mimic of the duplex part of RNA to test if the AMT dimer could work as expected.
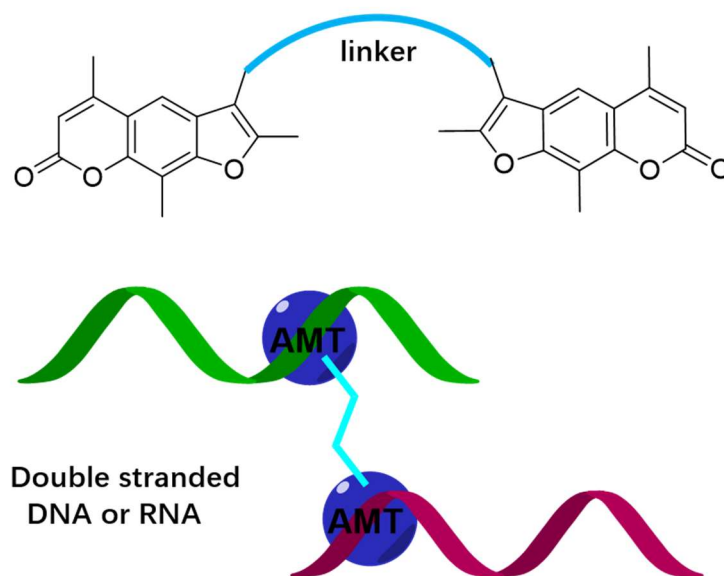
**Scheme 3-3 Model of AMT dimer crosslinks two double stranded DNA or RNA.** The light blue line indicates the length tunable linker, blue and purple belts denote two interacting double-stranded DNA or RNA and blue ball indicates AMT.

Here, we also describe an RNA and protein crosslinker, which is aimed to capture long range RNA-protein interactions (**Scheme 3-4**). As mentioned in chapter 1, traditional methods for studying of RNA-protein interactions mainly fall into three categories. One is formaldehyde based crosslinking and immunoprecipitation, which is limited by small size of formaldehyde. Formaldehyde can only crosslink lysine of target protein and amino group of RNA, and the interaction could be only captured when lysine is contained in the interaction sites. The second method is UV light induced crosslinking and immunoprecipitation, which requires uridines and photoactive amino acids on RNA and protein interaction sites. The third one is photoactivatable nucleoside analogues mediated crosslinking and immunoprecipitation, which also needs the interaction sites to contain uridine and photoactivatable amino acids.

The common limitation of the three traditional methods is the demand for special amino acid on RNA-protein interaction sites. To overcome this shortcoming, we synthesized an RNA protein crosslinker comprised of AMT and NHS ester conjugated by a size tunable linker. By using this reagent for the study of RNA-protein interactions, NHS side will crosslink with amino groups of protein and AMT side will crosslink to uridines of RNA. We expected that the introduction of length tunable linker could

overcome the limitations mentioned above. Because the flexibility of the linker enables AMT to react with nucleosides either within or around RNA and protein interaction sites, similarly, lysine around the interaction sites can also be captured by NHS.
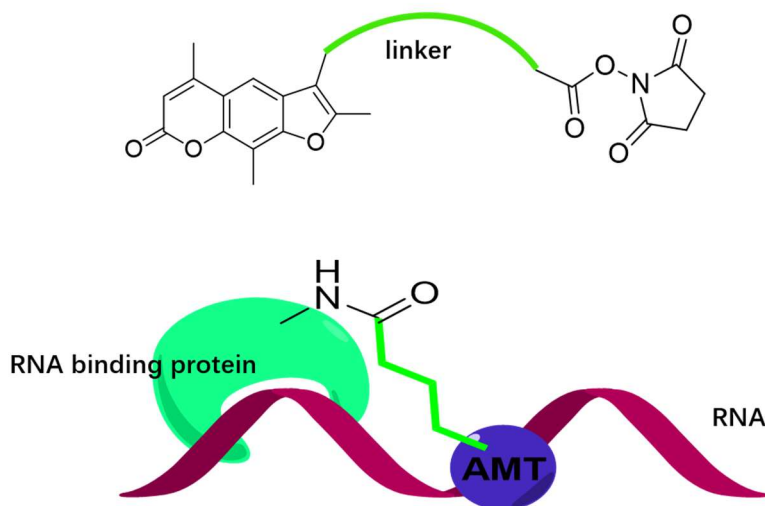


**Scheme 3-4 Model of AMT-NHS crosslinking RNA and protein**. Green line indicates the size tunable linker, light green spheroid indicates RNA binding protein, blue ball indicates AMT and purple belt denotes protein binding RNA. NHS ester forms amide bond with amino group on protein, and AMT reacts with uridine upon irradiation with 365 nm UV light. Both reaction sites are permitted to be beyond the RNA and protein interaction site.

### 3.2 Result and discussion

**Synthesis of AMT-NHS and AMT dime**r

We started synthesizing AMT-NHS from trimethyloxalen. Therefore, we modified the substrate with a chloromethyl group by nucleophilic addition of trioxane and electrophilic addition of hydrogen chloride. Then we introduced 3-mercaptopropionic acid as the linker of AMT and NHS. This linker was chosen to consist of three carbon atoms but it could be designed to be longer or shorter according to specific needs. Moreover, a more hydrophilic chain like PEG (polyethylene glycol) could be used as linker structure. Afterwards, we introduced NHS with N-hydroxysuccinimide with the help of dicyclohexylcarbodiimide (DCC) to obtain the target compound (AMT-NHS) (**Scheme 3-5**).
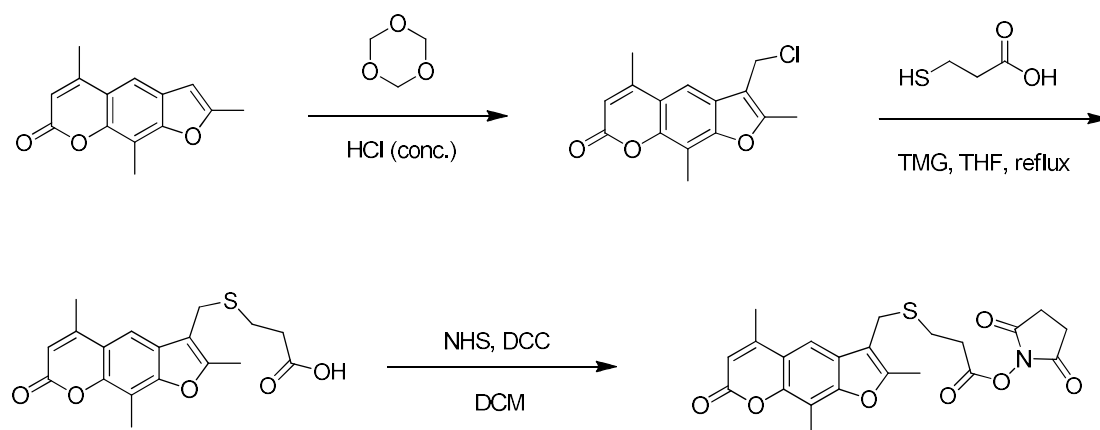
**Scheme 3-5 Synthesis of AMT-NHS**

We synthesized AMT dimer also by using trimethyloxalen as the starting reagent. Nucleophilic attack of N-hydroxymethylphthalimide was catalyzed by trifluoromethanesulfonic acid (TfOH) and trifluoroacetic acid (TFA). The phthalyl unit was removed by hydrazine hydrate to afford AMT. At last, AMT dimer was obtained by coupling of AMT and AMT-NHS (**Scheme 3-6**).
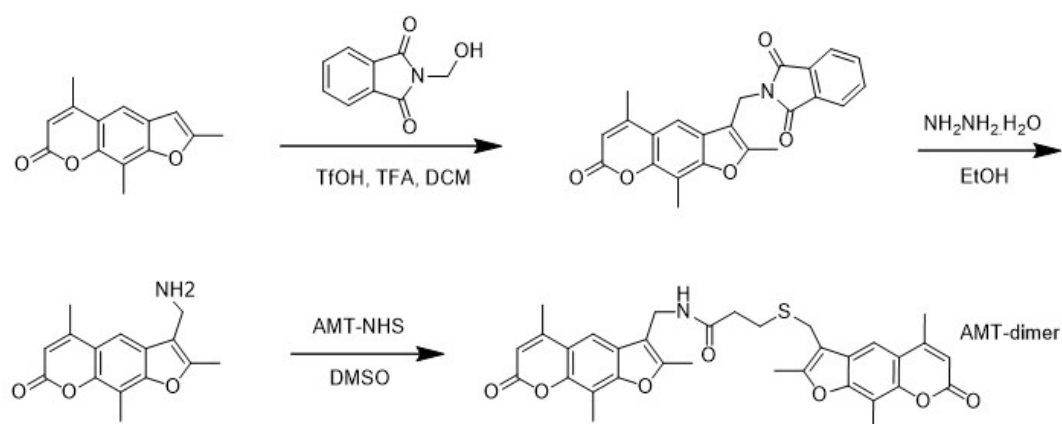


**Scheme 3-6 Synthesis of AMT dimer**

**In vitro verification of inter-strand crosslinking with AMT dimer**

We designed two oligonucleotides folded as hairpin with the intention of simulating the stem loop structure of RNA. We also inserted several 'AT' spacers in the sequence to form staggered thymidine, which is favorable for the crosslinking of AMT (**Figure 3-1**). We then incubated hairpin **1** and hairpin **2** with different concentration of AMT dimer in HEPES buffer (7.4) and irradiated with 365 nm UV light for 15 min. The crosslinked products were identified by native gel electrophoresis. Crosslinked

products were clearly observed from AMT dimer treated samples when comparing to samples free of AMT dimer. As expected, higher concentration of AMT dimer treated samples generated more crosslinked products. To eliminate the suspicion that nucleic acids might be crosslinked only by UV light irradiation, we set up a negative control group containing hairpin **1** and hairpin **2** that were irradiated with 365 nm UV light without addition of AMT dimer. We also tested if AMT dimer was able to crosslink hairpins without irradiation of UV light. It turned out that light induction and AMT dimer need to be present simultaneously to crosslink RNAs. In order to get more precise information, we analyzed the crosslinking products also on denaturing urea containing PAGE. We found that the products were comprised of hairpin dimer, trimer, tetramer and even more complex conjugates. (**Figure 3-2**).
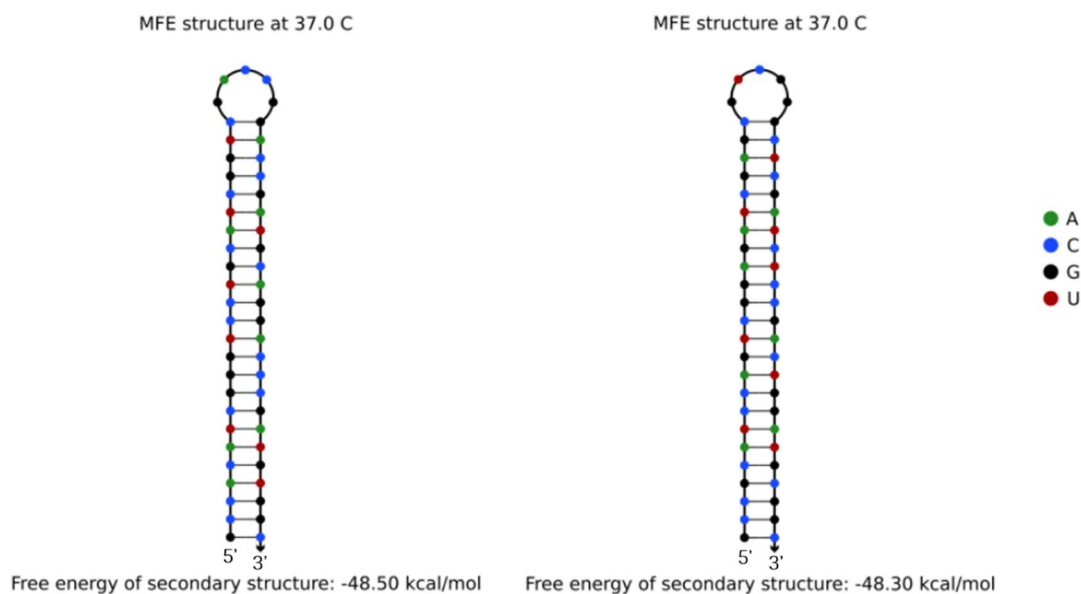


**Figure 3-1 Two hairpins were designed as mimics of RNA duplex, the stem loop structure was predicted by NUPACK.**
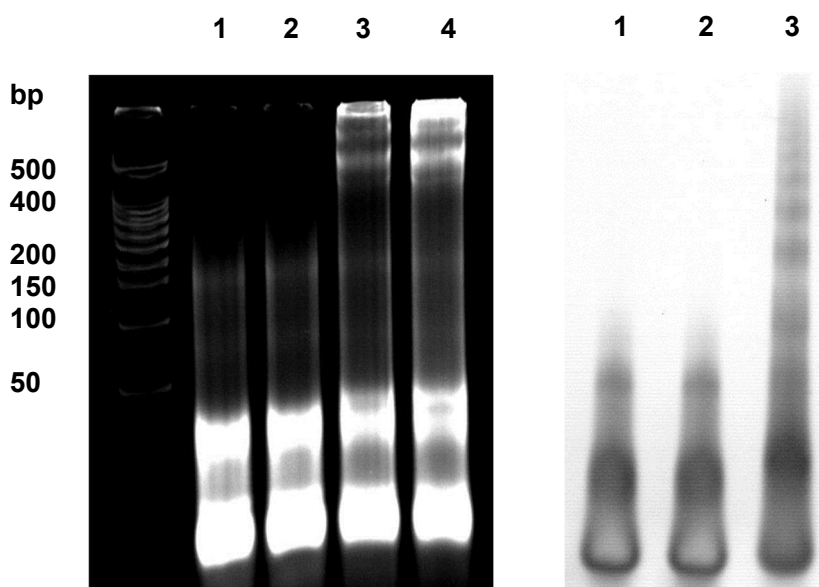
**Figure 3-2 AMT dimer crosslinks with hairpins in vitro. Left picture: 1**, 40 μM hairpin 1 and 40 μM hairpin 2 in HEPES buffer (pH 7.4); **2**, 40 μM hairpin 1 and 40 μM hairpin 2 in HEPES buffer (pH 7.4), irradiation at 365 nm for 15 min; **3**, 40 μM hairpin 1, 40 μM hairpin 2 and 120 μM AMT dimer in HEPES buffer (pH 7.4), irradiation at 365 nm for 15 min; **4**, 40 μM hairpin 1, 40 μM hairpin 2 and 400 μM AMT dimer in HEPES buffer (pH 7.4), irradiation at 365 nm for 15 min. Samples were analyzed in 2% agarose gel in TBE buffer. **Right picture: 1**, 50 μM hairpin 1 and 50 μM hairpin 2 in HEPES buffer (pH 7.4); **2**, 50 μM hairpin 1, 50 μM hairpin 2 and 150 μM AMT dimer in HEPES buffer (pH 7.4); **3**, 50 μM hairpin 1, 50 μM hairpin 2 and 500 μM AMT dimer in HEPES buffer (pH 7.4), irradiation at 365 nm for 15 min. Samples were analyzed on 6% Urea PAGE.

**RNA and protein crosslinking in vitro**

We also tested if AMT-NHS could be used as a protein-RNA crosslinker. First, we decided to prove if AMT-NHS could be efficiently ligated on protein. We chose an RNA binding protein, tiaS (tRNAIle agmatidine Synthetase), as the target protein for the crosslinking experiment. In consideration that AMT is fluorescent when activated by UV light, a bright band should be observed on PAGE (Poly Acrylamide Gel Electrophoresis) under UV light irradiation. We expressed tiaS in E. coli and in vitro transcribed tRNAIle for in vitro crosslinking. We incubated tiaS and AMT-NHS in PBS buffer (pH 7.4) and then analyzed the product on PAGE. As expected, bright bands were observed on the gel when exposed to UV light (**Figure 3-3, left**). In order to test

the crosslinking efficiency, we used higher concentration of AMT-NHS to ligate to tiaS and a brighter band was observed, which meant more lysines of tiaS reacted with AMT-NHS. Then we stained the gel with coomassie brilliant blue and compared with a sample not treated with AMT-NHS tiaS, which further confirmed that the observed bands were from AMT-NHS decorated tiaS (**Figure 3-3, right**).
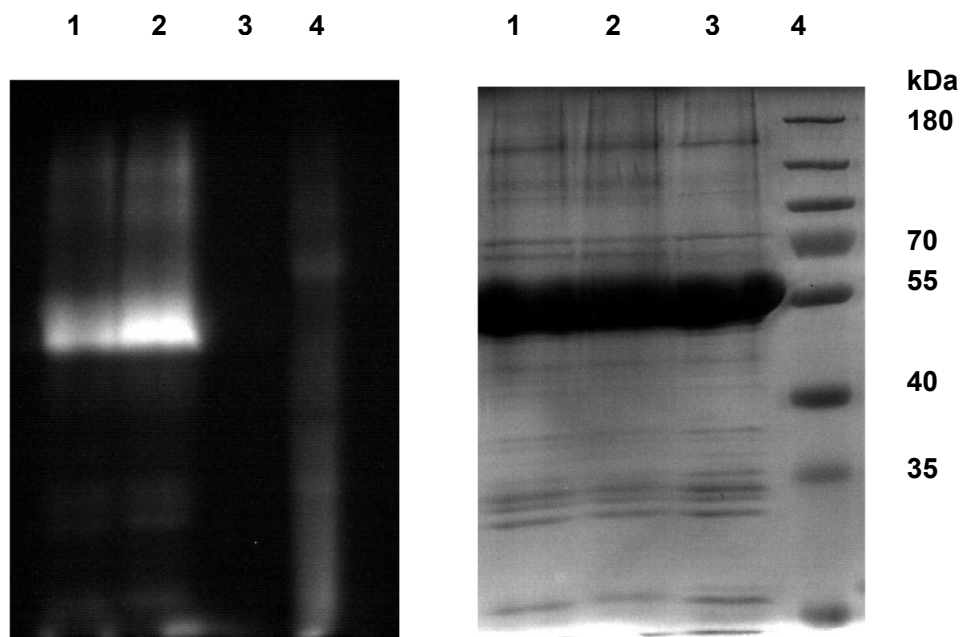


**Figure 3-3 AMT-NHS ligating on tiaS in vitro.** Left picture, **1**, 30 μM tiaS and 300 μM AMT-NHS in PBS (pH 7.4) buffer; **2**, 30 μM tiaS and 1.5 mM AMT-NHS in PBS (pH 7.4) buffer; **3**, 30 μM tiaS; **4**, Protein marker. The gel was illuminated with 300 nm UV light without staining, AMT-NHS treated tiaS is fluorescent. The higher the concentration of AMT-NHS the brighter of the protein (AMT/tiaS = 50/1). Right picture, the same gel was stained by coomassie brilliant blue. The size of tiaS did not change apparently with or without modification of AMT-NHS.

After verifying that AMT-NHS was able to react with tiaS efficiently, we next attempted to use this compound to crosslink tiaS and tRNAIle. We assumed that modification of NHS would not reduce the crosslinking efficiency of AMT. We incubated tiaS, tRNAIle and AMT-NHS in PBS (pH 7.4) buffer for 5 minutes and then irradiated the reaction mixture with 365 nm UV light for 15 minutes. The resulting crosslinking products were analyzed by SDS-PAGE. In order to compare with the traditional 254 nm UV light induced crosslinking method, we set up a sample wherein tiaS and tRNAIle were irradiated with 254 nm UV light for 15 minutes (**lane 2**), the crosslinking products bands mainly concentrate around the high molecule region (> 180

kDa). As for AMT crosslinking, the crosslinked product could be clearly observed and the size distribution of the products is slightly different with increasing the concentration of AMT-NHS (**lane 3-7**), which implies that the diversity of crosslinking products in vitro experiment may more depends on the concentration of RNA and protein, but not crosslinker. In order to investigate crosslinking efficiency , we observed that tiaS in **lane 2** was more than those in **lane 3-7,** and the crosslinked products were just the opposite. We conclude that the crosslinking efficiency of AMT-NHS is lower than traditional method in vitro experiment, and the products distribution is apparently different. (**Figure 3-4**). In consideration of the real environment of RNA and proteins existing in vivo is totally different from these in vitro experiments, such as the concentration of crosslinking components and the relative location of RNA and protein, we cannot draw any conclusion that the crosslinking efficiency would be different from the traditional method in vivo. However, the different product distribution tells that the AMT-NHS may capture some protein binding RNAs that traditional method miss.
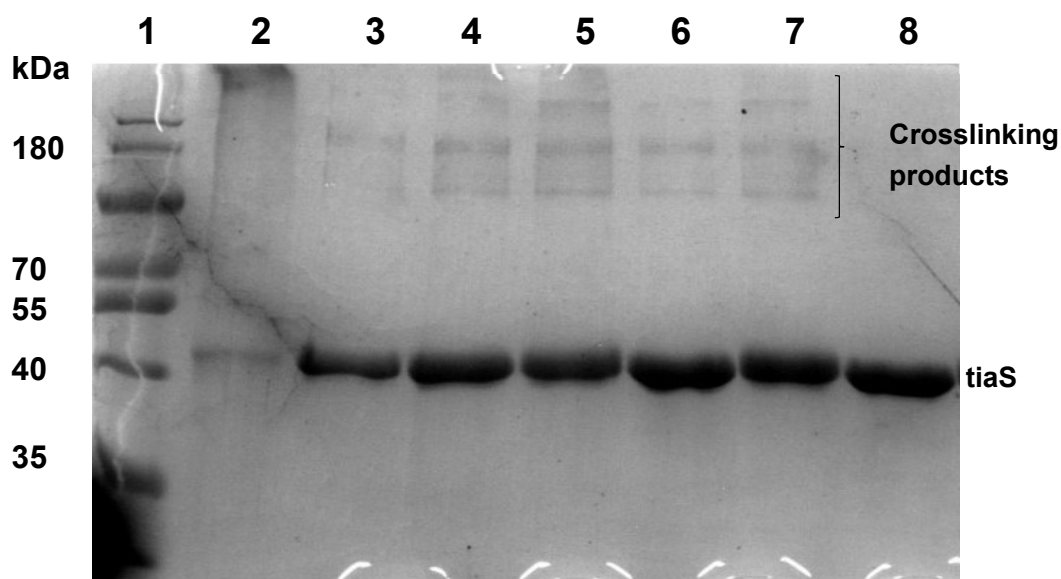


**Figure 3-4 tiaS and tRNAIle crosslinking in vitro. 1**, Protein marker; **2**, 10 μM tiaS and 10 μM tRNAIle in PBS (pH 7.4) buffer, 254 nm UV light for 15 min; **3**, 10 μM tiaS 10 μM tRNAIle and 50 μM AMT-NHS in PBS (pH 7.4) buffer, 365 nm UV light for 15 min; **4**, 10 μM tiaS，10 μM tRNAIle and 200 μM AMT-NHS in PBS (pH 7.4) buffer, 365 nm UV light for 15 min; **5**, 10 μM tiaS，10 μM tRNAIle and 500 μM AMT-NHS in PBS (pH 7.4) buffer, 365 nm UV light for 15 min; **6**, 10 μM tiaS，10 μM tRNAIle and 1 mM AMT-NHS in PBS (pH 7.4) buffer, 365 nm UV light for 15 min; **7**, 10 μM tiaS，10 μM tRNAIle and 2 mM AMT-NHS in PBS

(pH 7.4) buffer, 365 nm UV light for 15 min; **8**, 10 μM tiaS in PBS (pH 7.4) buffer. Crosslinking products were analyzed on 12 % SDS-PAGE, stained by coomassie brilliant blue.

## 3.3 Summary and conclusion

In this chapter, we synthesized a length tunable linker containing two AMT units to study RNA-RNA interactions. We also designed two hairpin shaped oligomeric nucleotides as mimic of duplex part of RNA. We finally confirmed that the AMT dimer is able to crosslink RNAs by irradiation of 365 nm UV light. Comparing to traditional RNA crosslinking methods, AMT dimer enables the exploration of long range RNA-RNA interactions, which may provide a broader scope of RNA interaction research.

Moreover, we synthesized a NHS ester modified with AMT to study RNA-protein interactions. We firstly verified that the novel reagent was able to effectively react with tiaS by utilizing the fluorescent property of AMT, and then we performed tiaS and tRNAIle crosslinking in vitro by illuminating with 365 UV light. The crosslinking products were analyzed on SDS-PAGE and finally we confirmed in vitro that AMT-NHS would be a novel crosslinker for studying RNA-protein interactions. Benefitting from the fact that the size of the linker is adjustable, this crosslinker should capture more comprehensive information in studying RNA-protein interactions in living cells.

## 3.4 Experimental section

### Materials and methods

The primers were synthesized from Sangon Biotech (Shanghai) Co., Ltd. Molecular Biology enzymes were purchased from New England BioLabs Inc. or Thermo Fisher Scientific. NTPs were purchased from Sangon Biotech (Shanghai) Co., Ltd. RNase inhibitor and TRIzol reagent were purchased from Thermo Fisher Scientific. Other chemicals were purchased from Sigma-Aldrich without further purification. NMR spectra were recorded on Varian Unity Inova (500 MHz). The agarose gel or the SDS-PAGE gel images were obtained from LX-BIO-2800 (KCBF). The UV crosslinking experiments were carried out in CL-1000 UV crosslinker (UVP), which has five 254 nm lamps (8 W) or five 365 nm lamps (8 W).

### Synthesis of 3-(chloromethyl)-2,5,9-trimethyl-7H-furo[3,2-g]chromen-7-one 1

A mixture of 1,3,5-trioxane (217 mg, 2.4 mmol) and concentrated HCl (20 ml) was vigorously stirring when trioxalen (500 mg, 2.2 mmol) was added portion wise. The reaction was stirred overnight and was then diluted with 100 ml water. The mixture was extracted with dichloromethane, washed three times with water and dried with sodium sulfate. The organic solvent was concentrated to obtain **1** as white solid (420 mg, 70%) $^1$H NMR (CDCl$_3$, 500 MHz) $\delta$ 7.60 (s, 1H), 6.26 (t, J = 9.2 Hz, 1H), 4.75 (s, 2H), 2.62 – 2.54 (m, 3H), 2.54 – 2.45 (m, 6H).
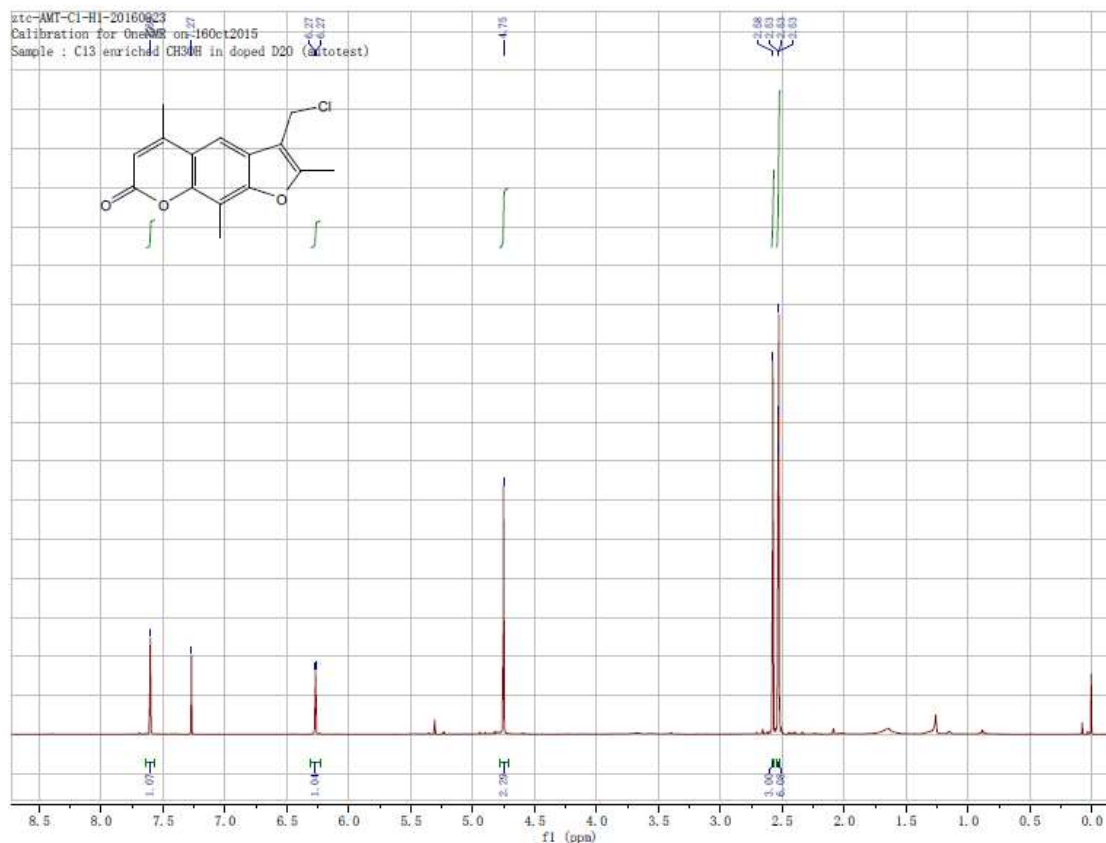


**Synthesis of 3-(((2,5,9-trimethyl-7-oxo-7H-furo[3,2-g]chromen-3-yl)methyl)thio) propanoic acid 2**

To a solution of compound **1** (100 mg, 0.36 mmol) and 3-mercaptopropanoic acid (38 μl, 0.44 mmol) in THF (4 ml), dimethylbiguanide (100 μl, 0.79 mmol) was added at room temperature. The reaction was heated to reflux for 24 h and then the solvent was evaporated to dryness. The mixture was dissolved in dichloromethane and washed by water and brine. Afterwards, the solvent was dried by sodium sulfate and evaporated to obtain a crude product. The crude product was further purified by silica gel column chromatography and eluted with dichloromethane to afford **2** as white solid (75 mg, 60%).

$^1$H NMR (d-DMSO, 500 MHz) δ 7.77 (s, 1H), 6.31 (s, 1H), 3.92 (s, 2H), 2.60 (t, J = 6.0 Hz, 2H), 2.53 (t, J = 7.0 Hz, 2H), 2.46 (m, 6H), 2.43 (s, 3H).
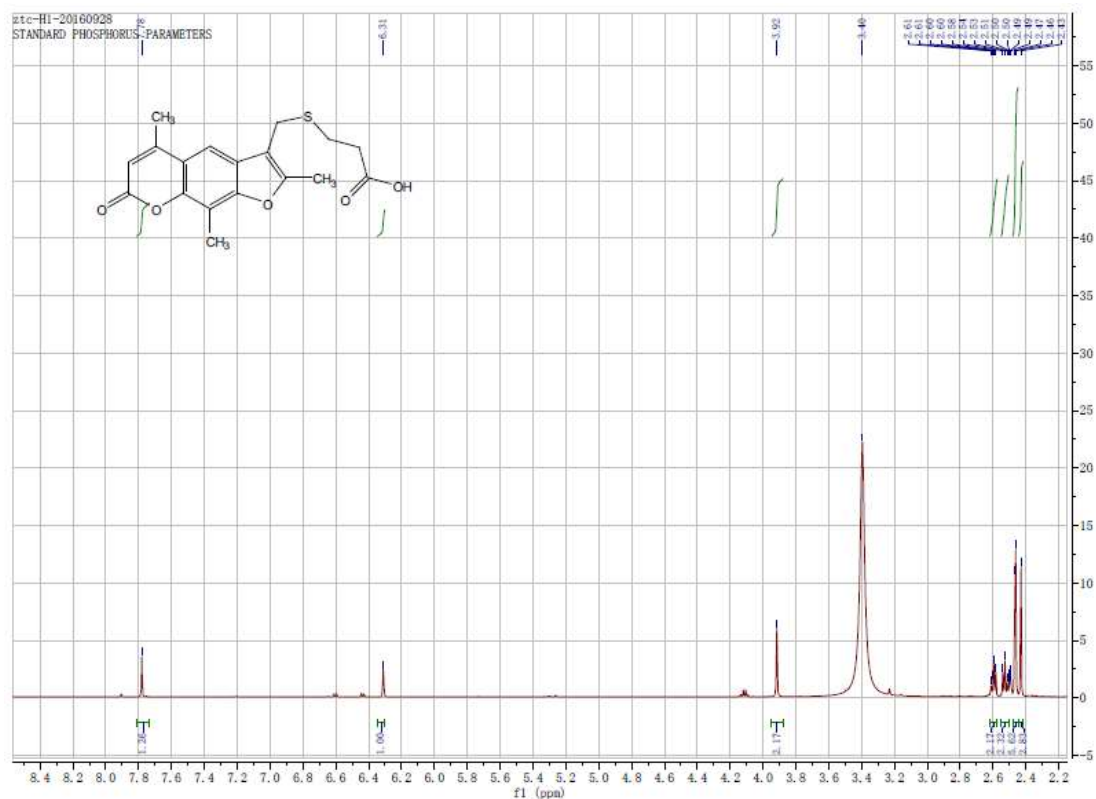


## Synthesis of 2,5-dioxopyrrolidin-1-yl 3-(((2,5,9-trimethyl-7-oxo-7H-furo[3,2-g]chromen-3-yl)methyl)thio)propanoate 3

Compound **2** (20 mg, 0.055 mmol) was dissolved in dichloromethane (3 ml) followed by addition of N-hydroxysuccinimide (7 mg, 0.061 mmol) and dicyclohexylcarbodiimide (12.5 mg, 0.061 mmol ) at room temperature. After stirring for 2 h, the mixture was extracted with dichloromethane and organic phase was washed with water and brine. Solvent was evaporated under vacuum to give the crude product, which was purified by silica gel column chromatography and eluted with ethyl acetate : hexane (4 : 1) to afford **3** as white solid (15 mg, 60%).

$^1$H NMR (CDCl$_3$, 500 MHz) δ 7.64 (s, 1H), 6.26 (s, 1H), 3.89 (s, 2H), 2.90 (t, J = 7.0 Hz, 2H), 2.85 (s, 4H), 2.82 (t, J = 6.5 Hz, 2H), 2.59 (s, 3H), 2.53 (s, 3H), 2.50 (s, 3H).
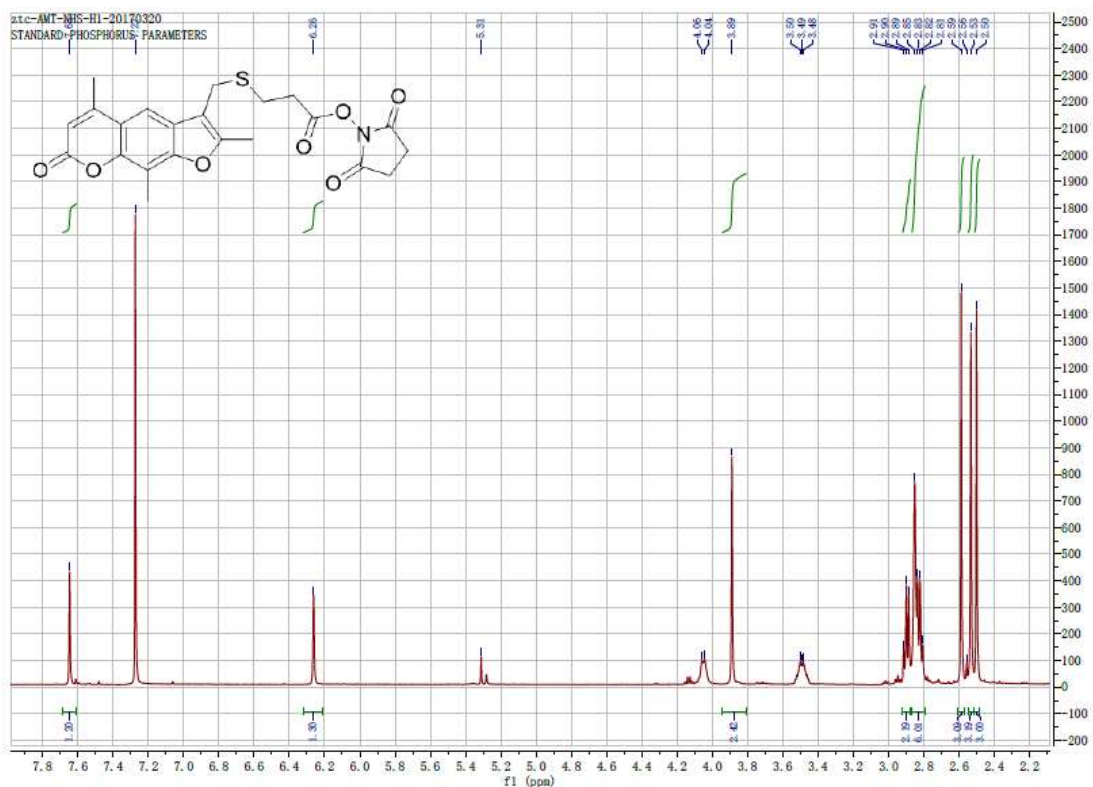
**Synthesis of 2-((2,5,9-trimethyl-7-oxo-7H-furo[3,2-g]chromen-3-yl)methyl) isoindoline-1,3-dione 4**

Trioxsalen (500 mg, 2.1 mmol) was dissolved in dichloromethane (15 ml) followed by addition of N-hydroxymethylphthalimide (800 mg, 4.2 mmol). Afterwards, a mixture of trifluoromethanesulfonic acid (230 μl, 2.5 mmol) and trifluoroacetic acid (5 ml) was added dropwise to the reaction. After stirring at room temperature for 24 h, saturated $NaHCO_3$ solution was added until no bubbles were generated anymore. The mixture was extracted with dichloromethane, washed with brine and dried by sodium sulfate. Then solvent was evaporated under vacuum to give **4** as white solid (510 mg, 60%).

$^1$H NMR (CDCl$_3$, 500 MHz) δ 7.95 (s, 1H), 7.84 (dd, J = 5.4, 3.1 Hz, 2H), 7.71 (dd, J = 5.4, 3.0 Hz, 2H), 6.24 (s, 1H), 4.91 (s, 2H), 2.71 (s, 3H), 2.57 (s, 3H), 2.55 (s, 3H).
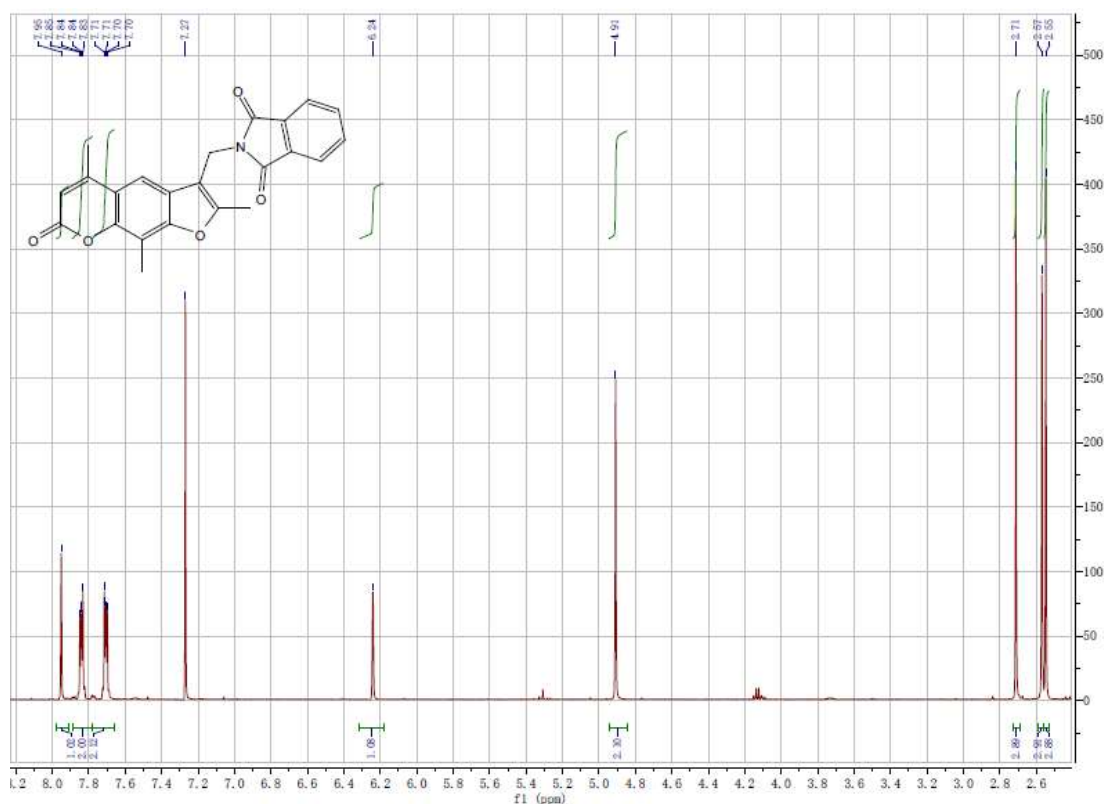


**Synthesis of 3-(aminomethyl)-2,5,9-trimethyl-7H-furo[3,2-g]chromen-7-one 5**

Compound **5** (510 mg, 1.3 mmol) was dissolved in ethanol (20 ml) followed by adding hydrazine hydrate (4 ml) at room temperature. Then the reaction mixture was heated to reflux for 2 h. Solvent was evaporated and 200 ml NaOH (0.1 M) was added. Then the solution was extracted with dichloromethane and the organic phase was washed and concentrated. The crude product was further purified by silica gel column chromatography and eluted with ethyl acetate : hexane (4 : 1) to afford **5** as white solid (270 mg, 80%).

$^1$H NMR (CDCl$_3$, 500 MHz) δ 7.60 (s, 1H), 6.24 (s, 1H), 3.96 (s, 2H), 2.57 (s, 3H), 2.51 (s, 3H), 2.48 (s, 3H).
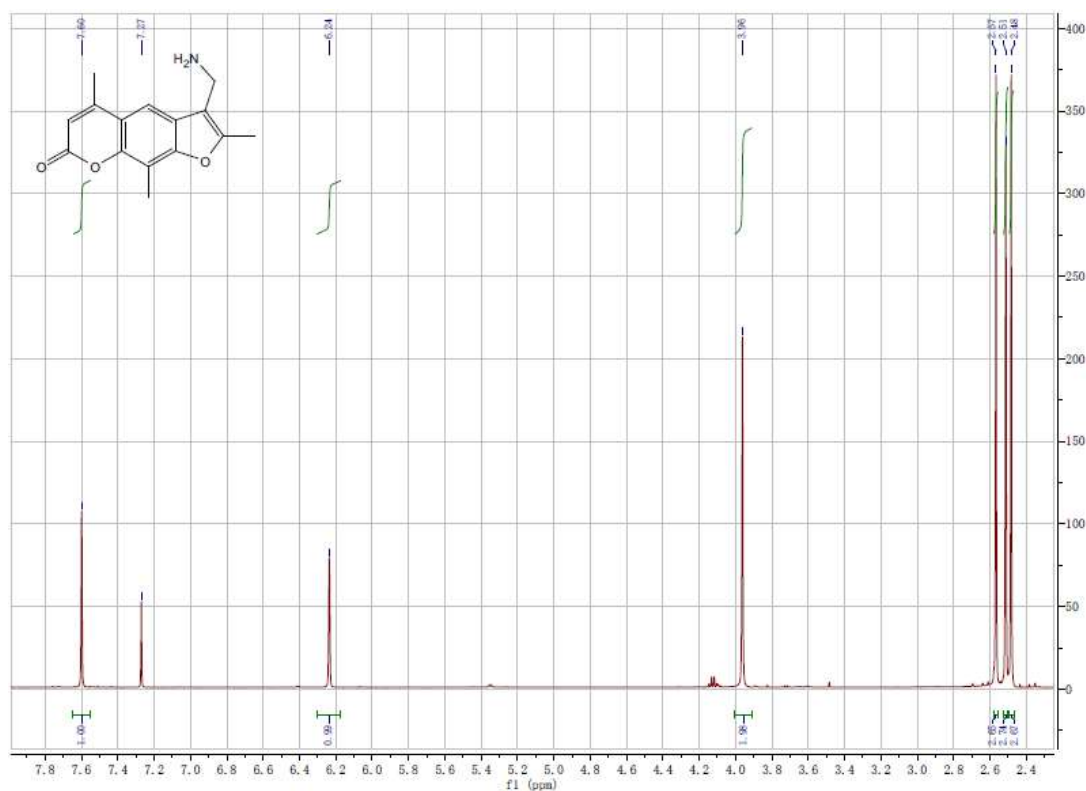
**Synthesis of N-(2,5,9-trimethyl-7-oxo-7H-furo[3,2-g]chromen-3-yl)-3- (((2,5,9-trimethyl-7-oxo-7H-furo[3,2-g]chromen-3-yl)methyl)thio)propanamide**

Compound **5** (9 mg, 0.034 mmol) was dissolved in DMSO (0.5 ml) followed by addition of compound **3** (15 mg, 0.034 mmol) in DMSO (0.5 ml). After stirring at room temperature for 1 h, the product was identified by LC-MS.

LC/MS (LC: gradient 10-90% MeOH [0.1% $HCO_2H$] over 15.0 min, 1.2 ml/min flow rate, retention time, 9.55 min; MS (ESI$^+$) (m/z): found 572.09 [M+H]$^+$; calculated, 572.17, [M+H]$^+$.

**TiaS expression and purification**

Transformation of DNA plasmid into competent BL21（DE3）*E. coli* cells

The competent cells were taken from -80℃ freezer and placed on ice for 5 min to thaw. Then 1 μL plasmid DNA was added into the thawed competent cells. The mixture was vortexed and incubate on ice for 0.5 h. Afterwards, the cells were heat shocked at 42℃ for 1 min and then placed on ice. The heat-shocked cells were then transferred into LB broth and incubated (37 ℃, 200 rpm) for 1 h. Then 100 μL cell culture was plated on LB ampicillin treated agar plates, which were then incubate at 37℃ overnight.

**Induction of protein expression**

The colony was picked from the plate and inoculate to 100 ml LB broth (100 ug/ml ampicillin), which was then incubate on shaker (37 ℃, 200 rpm) for overnight. 10 ml of the cell culture was inoculated to 1 L LB broth (100 ug/ml ampicillin) and incubated on shaker (37 ℃, 200 rpm) until the OD 600 nm reaches 0.6 (about 4 h). IPTG (isopropylthiogalactoside) was added to the cell culture (final concentration, 0.2 mM) and incubate on shaker (16 ℃, 160 rpm) for 16 h to induce protein expression.

**Cell lysis**

The cell culture was centrifuged at 4000 rpm for 15 minutes to collect the cells. The cell pellets was resuspended with lysis buffer (50 mM Tris-HCl (pH=8.0), 500 mM NaCl, 10 mM imidazole, 5% (v/v) Glycerol) and disrupted by ultrasonication (work 5 s, pause 5 s) for 40 minutes. The lysate was then centrifuged at 13000 rpm, 4 ℃ for 30 minutes and the supernatant was collect.

**Protein purification**

The supernatant was added 2 ml Ni-NTA agarose and incubated at 4 ℃ for 30 minutes. Afterwards, the mixture was purified on column and eluted with 50 ml wash buffer (50 mM Tris-HCl (pH=8.0), 500 mM NaCl, 30 mM imidazole, 5% (v/v) Glycerol) and 15 ml elution buffer (50 mM Tris-HCl (pH=8.0), 500 mM NaCl, 350 mM imidazole, 5% (v/v) Glycerol). The purified Tias was then analyzed on 12% (w/v) SDS-PAGE (**Figure 2-6**).The purified protein was concentrated with centrifugal filter (millipore) and further purified with AKTA pure chromatography system. The concentrated protein was then changed in to low salty buffer (50 mM Tris-HCl, pH=8.0, 50 mM NaCl)._The pump A and pump B system of AKTA was rinsed with water, low salty buffer and high salty buffer (50 mM Tris-HCl, pH=8.0, 1 M NaCl). The proteins were loaded from sample loop to HiTrap Heparin column which was pre-equilibrated by low salty buffer. The column was eluted with 200 ml elution buffer from 0% B to 100% B (flow rate 2 ml/min, 100 minutes). The target fractions were collected according to UV absorption at 280 nm (**Figure 2-5a**). The fractions were characterized on 12% (w/v) SDS-PAGE (**Figure 2-6**) and concentrated with centrifugal filter tube. The proteins were further purified with size exclusion column (Superdex 200, GL10/300) and eluted with SEC buffer (100 mM Tris-HCl (pH=8.0), 400 mM NaCl, 20 mM KCl, 5% (v/v) Glycerol), the fractions were collected according to UV absorption (**Figure 3-5b**). Eluent was then analyzed on 12% (w/v) SDS-PAGE (**Figure 3-6**) and concentrated for use.

**Figure 3-5a**
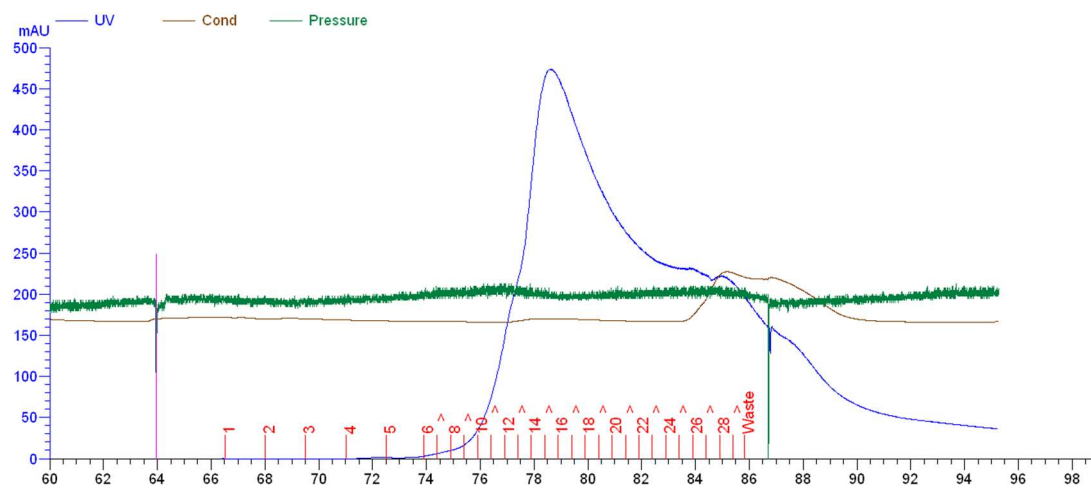


**Figure 3-5b**

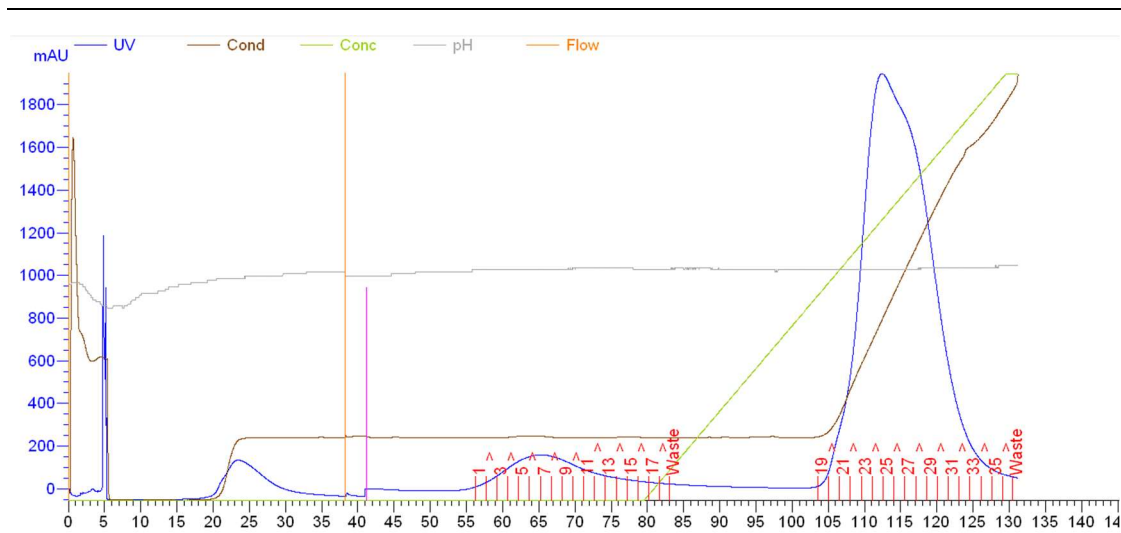**Figure 2-5 Purificaiton of TiaS on AKTA chromatography. 5a** Purification with HiTrap Heparin column; **5b** Purification with size exclusion column. Blue curve denotes UV absorption at 280 nm. Brown curve denotes conductivity of eluent. Red numbers denote number of fraction of eluent. Y axis denotes the UV absorption of eluent at 280 nm, X axis denotes the number of fractions according to collection time.
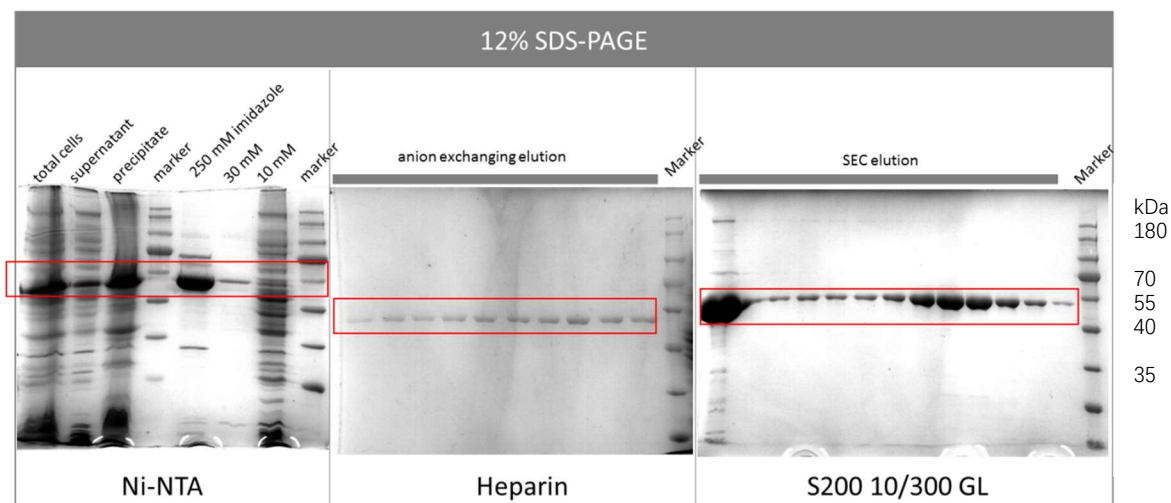
**Figure 3-6 Analyze TiaS on 12% (w/v) SDS-PAGE after purified with Ni-NTA column, Heparin column and S200 10/300 GL column.**

**In vitro transcription and purification of Af tRNA<sup>Ile</sup>**

Af tRNA<sup>Ile2</sup> template amplification

The DNA template was obtained by PCR amplification of the template sequence encoded on a plasmid.

Af tRNA<sup>Ile2</sup> template sequence

TAATACGACTCACTATAGGGCCCGTAGCTTAGCCAGGTCAGAGCGCCCGGC
TCATAACCGGGCGGTCGAGGGTTCGAATCCCTCCGGGCCCACCA

Primers

AftRNA<sup>Ile2</sup> F: TAATACGACTCACTATAGGGCCCGTAGC

Af tRNA<sup>Ile2</sup> R: TGGTGGGCCCGGAGGGATTC

**PCR components**

| Reagent | Quantity (500 µL) | final concentration |
|---|---|---|
| pET22b-Af tRNAIle2 plasmid | 1 µL | - |
| Af tRNA$^{Ile2}$ F (10 µM) | 10 µL | 0.2 µM |
| Af tRNA$^{Ile2}$ R (10 µM) | 10 µL | 0.2 µM |
| dNTP mix (10 mM) | 10 µL | 0.2 µM |
| 5X Pfu buffer | 100 µL | 1X |
| Pfu DNA polymerase (2.5 U/µL) | 5 µL | 25 U/mL |
| PCR H$_2$O | 364 µL | - |
| | 500 µL | |

**PCR program**

| Reaction | Temperature (℃) | Time | |
|---|---|---|---|
| initial denaturation | 95 | 3 min | |
| denaturation | 95 | 30 s | |
| annealing | 50 | 30 s | 35 cycles |
| extension | 72 | 4 min | |
| final extension | 72 | 10 min | |
| - | 12 | 15 min | |

**Purification of the DNA template**

The reaction solution was combined and added with equal volume of saturated phenol : chloroform : isoamylol (25 : 24 :1). The mixture was vortexed vigorously and centrifuged it at 12000 rpm for 15 minutes to form a two-phase solution. The upper aqueous phase was transfered to a new tube and added with equal volume of chloroform : isoamylol (24 : 1), the mixture was vortexed vigorously and centrifuged it at 12000 rpm for 15 minutes to form a two-phase solution. The upper phase was transferred to a new tube and added with 10% volume of 3M NaOAc (pH = 5.2, RNase free). Afterwards, the mixture was added with equal volume of isopropanol and stored

at -20 ℃ overnight. The mixture was then centrifuged at 12000 rpm for 15 minutes to form white precipitate at the bottom of tube. The supernatant was removed and the precipitate was washed with 75% ethanol, after centrifugation andd discard of liquid phase, the precipitate was dried in the air. Finally, the purified DNA template was analyzed on 2% (w/v) agarose gel (**Figure 3-7a**).

**RNA transcription**

| Component | Quantity (500 μL) | Final Concentration |
|---|---|---|
| 500 mM HEPES (pH=7.5) | 80 μL | 80 mM |
| $MgCl_2$ (200 mM) | 50 μL | 20 mM |
| spermidine (100 mM) | 5 μL | 1 mM |
| DDT (200 mM) | 25 μL | 10 mM |
| NTPs (20 mM) | 100 μL | 4 mM |
| DNA template | 2 μL | 2 μg/mL |
| T7 RNA polymerase | 50 μL | - |
| RNase inhibitor | 0.5 μL | 40 U/ml |
| Nuclease-free water | 188 μL | - |
| | 500 μL | |

The reaction was incubated at 37 ℃ for 5 h and the  solution was analyzed on 2% (w/v) agarose gel. In order to avoid contamination caused by template DNA, the template DNA was digested as following reaction:

| Component | Quantity | Final concentration |
|---|---|---|
| Af tRNA$^{Ile2}$ transcription solution | 1000 μL | - |
| RQ1 RNase-free DNase (1U/μL) | 10 μL | 10 U/ml |
| RQ1 DNase buffer (10X) | 100 μL | 1X |
| | 1110 μL | |

The reaction mixture was incubated at　37　℃　for 0.5 h and purified with TRIzol by manufacturer provided protocol. The purified RNA was analyzed on 2% (w/z) agarose gel (**Figure 2-7b**).
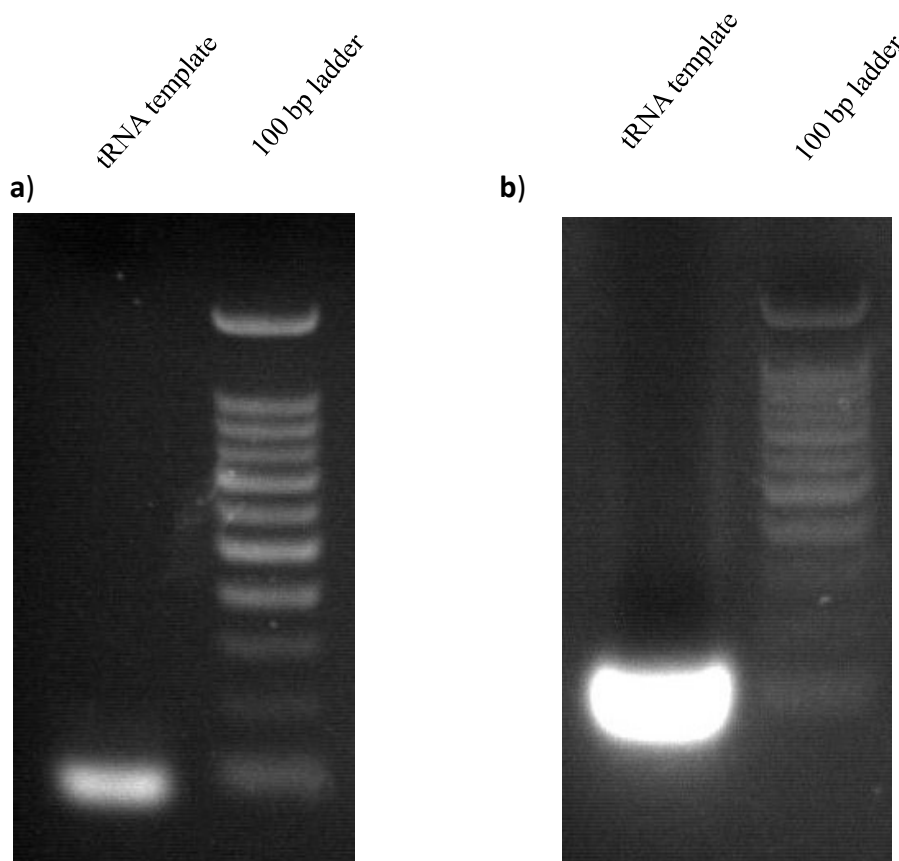


**Figure 3-7 Af tRNAIle2 template amplification (a) and in vitro transcription (b). Product was analyzed by agarose gel electrophoresis.**

## References

1.    Hemne, P. S.;   Kunghatkar, R. G.;   Dhoble, S. J.;   Moharil, S. V.; Singh, V., Phosphor for phototherapy: Review on psoriasis. *Luminescence : the journal of biological and chemical luminescence* **2017,** *32* (3), 260-270.

2.    Aw, J. G.;   Shen, Y.;   Wilm, A.;   Sun, M.;   Lim, X. N.;   Boon, K. L.;   Tapsin, S.;   Chan, Y. S.;   Tan, C. P.;   Sim, A. Y.;   Zhang, T.;   Susanto, T. T.;   Fu, Z.;   Nagarajan, N.; Wan, Y., In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol Cell* **2016,** *62* (4), 603-17.

3.    Calvet, J. P.; Pederson, T., Heterogeneous nuclear RNA double-stranded regions probed in living HeLa cells by crosslinking with the psoralen derivative aminomethyltrioxsalen. *Proc Natl Acad Sci U S A* **1979,** *76* (2), 755-9.

4.    Cimino, G. D.;   Gamper, H. B.;   Isaacs, S. T.; Hearst, J. E., Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annual review of biochemistry* **1985,** *54*, 1151-93.

5.    Welsh, J.; Cantor, C. R. J. J. o. M. B., Studies on the arrangement of DNA inside viruses using a breakable bis-psoralen crosslinker ☆. **1987,** *198* (1), 63-71.

6.    Diekmann, J.;   Gontcharov, J.;   Frobel, S.;   Torres Ziegenbein, C.;   Zinth, W.; Gilch, P., The Photoaddition of a Psoralen to DNA Proceeds via the Triplet State. *J Am Chem Soc* **2019,** *141* (34), 13643-13653.

7.    Frã¶Bel, S.;   Reiffers, A.;   Torres, Z. C.; Gilch, P. J. J. o. P. C. L., DNA Intercalated Psoralen Undergoes Efficient Photoinduced Electron Transfer. **2015,** *6* (7), 1260-1264.

8.    Heindel, N. D.;   Choudhuri, M.;   Ressner, J.; Foster, N. J. J. o. H. C., Aminomethyl psoralens. Electrophilic substitution of hydroxymethylphthalimide on linear furocoumarins. **2010,** *22* (1), 73-76.

# Chapter 4

# Verification of AMT based CLIP method in mammalian cells by study of an alternative splicing suppressor

## 4.1 Introduction

Alternative splicing is an RNA editorial method in the process of post-transcriptional modification[1-3], which consists of removing introns and ligating of the flanking exon fragments. Genes in mammalian cells contain several exons and introns, to increase the diversity of proteins translated from one gene, different mature mRNAs are formed by different combinations of exons in different situations (**Scheme 4-1**)[4].
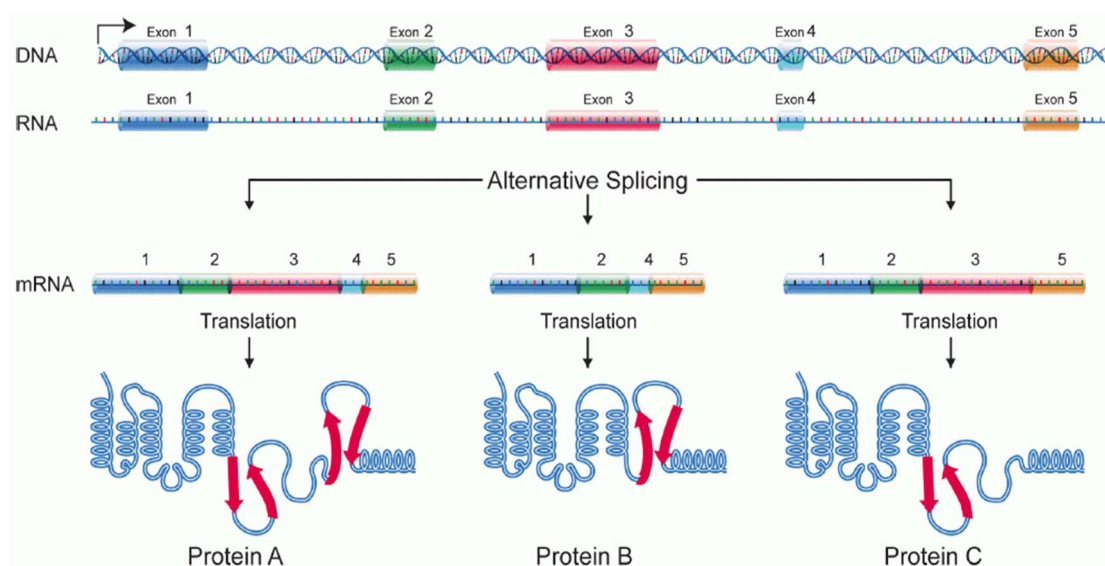


**Scheme 4-1 Diagram of alternative splicing.** Pre-mRNA transcribed from DNA consists of several introns and exons. During the processing of pre-mRNA, introns are cut out and exons are assembled to form a mature mRNA. Alternative splicing happens in the course of exon assembling. Some exons are skipped and the rest are retained under certain circumstances. Different combinations of particular exons under different conditions result in several different proteins coded by one DNA.

Alternative splicing is carried out by the spliceosome, a molecular machine that is assembled by snRNAs (U1, U2, U4, U5 and U6) and associated protein factors[5, 6]. snRNAs determine the splice site of pre-mRNA with the assistance of protein factors (activator or suppressor). Generally, splicing activator U2AF binds with the polypyrimidine tract of an intron and recruits U2 to bind with the branch site of an intron. The intron is then cut out and the flanking exons are ligated. However, when a splicing suppressor PTB (poly-pyrimidine tract binding protein) binds to the polypyrimidine tract, U2AF will be blocked and the introns will not be excised (**Scheme 4-2**).
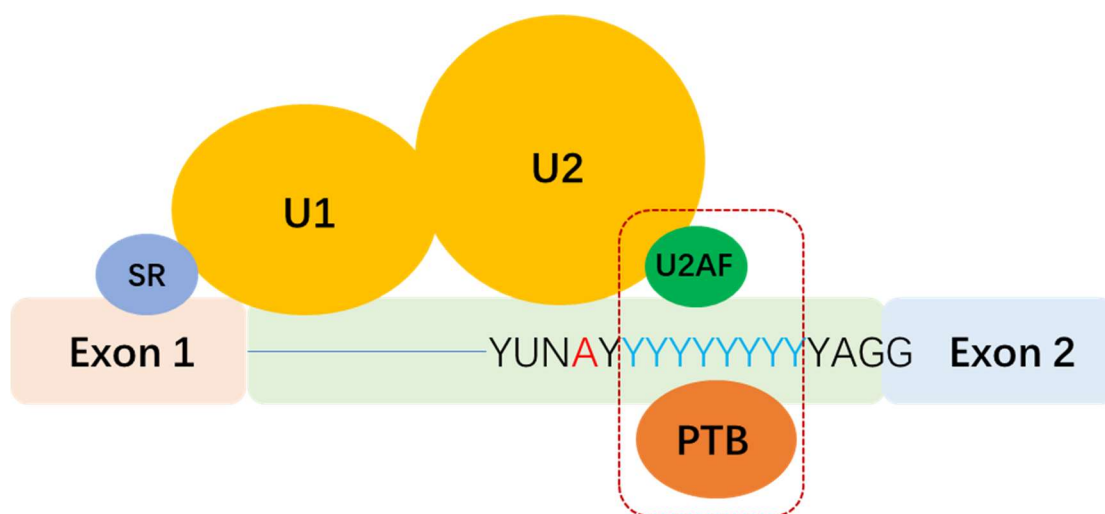
**Scheme 4-2 Key intermediate of alternative splicing.** This key intermediate is called spliceosome, a complex determines the end of an intron to be cut out and the end of an exon to be retained in the mature mRNA. U2AF as a splicing activator can bind with polypyrimidine tract (YYYYYYY) and recruit U2 to bind with branch point (A) of branch site (YUNAY), which retains Exon 2 in the mature mRNA. PTB, as an alternative splicing suppressor, can also bind to polypyrimidine tract, which blocks the interaction of U2AF and excludes Exon 2 in the mature mRNA.

In chapter **3** we proved that AMT-NHS is an available reagent for crosslinking of protein and RNA in vitro. Sine PTB is a well characterized alternative splicing suppressor[7, 8], here we determined to study PTB binding RNAs with our crosslinker and compare the results with the reported data obtained by traditional methods to verify if AMT-NHS based crosslinking strategy is suited for study of protein and RNA interactions in living cells [9].

**4.2 Results and discussion**

**AMT-NHS crosslinking efficiency test**

Before using AMT-NHS to study PTB binding RNAs, we firstly determined to test if it could efficiently crosslink PTB and RNA in living cells. We constructed a plasmid harbouring the gene of Myc tagged PTB and transfected it into mammalian cells. When Myc-PTB was expressed, the cells were treated with AMT-NHS and cultured at 37 ℃ to let the amino groups of PTB react with NHS ester. Then PTB and RNA was

crosslinked by irradiating with 365 nm UV light. After crosslinking, cells were lysed and the crosslinked PTB binding RNAs were collected by immunoprecipitation with anti-Myc beads, which is then separated on PAGE and finally analyzed by immunoblot (**Figure 4-1**). To verify the crosslinking products, we prepared samples by using traditional 254 nm UV light crosslinking method. The results obtained from immunoblot show that PTB and corresponding RNAs were efficiently crosslinked by AMT-NHS upon UV illumination. We also found that comparing with 254 nm UV light induced crosslinking method, AMT-NHS was more efficient. In addition, we set a control experiment that cells were treated with AMT-NHS without irradiating with UV light. Though it is expected that no crosslinking product should be formed, we observed slight bands from immunoblot result. This phenomenon might be caused by daylight lamp induced crosslinking.
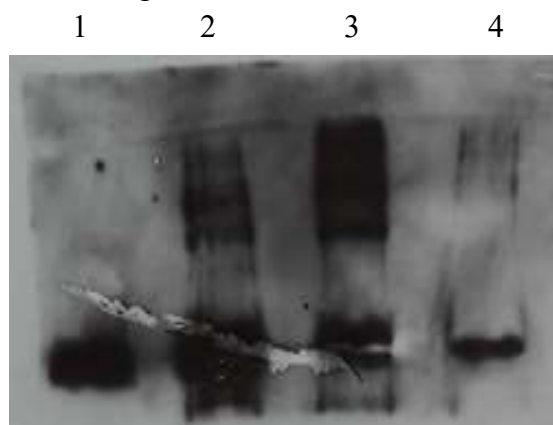


**Figure 4-1 Immunoblot analysis of RNA binding PTB in HEK 293 T.** The bottom band indicates PTB and the upper band denotes the RNA binding PTB. Crosslinking products were immuno-precipitated and size separated with gel electrophoresis, which was then transferred to PVDF membrane. The membrane was treated with primary antibody(anti-myc) and secondary and then illuminated with ECL. **lane 1**, Control group; **lane 2**, Cells were exposed to 254 nm UV light for 15 minutes; **lane 3**, Cells were cultured with AMT-NHS and then exposed to 365 nm UV light for 15 minutes; **lane 4**, Cells were cultured with AMT-NHS.
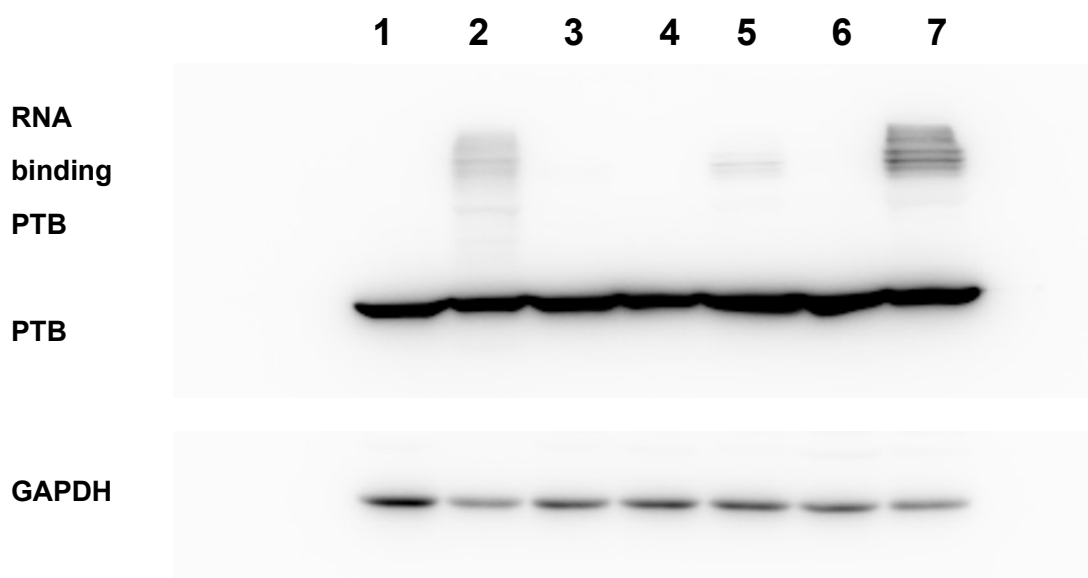
We also compared the crosslinking efficiency with traditional methods, CLIP and PAR-CLIP, which were performed as follows (**Figure 4-2**). We found that comparing to the most widely used CLIP method, AMT-NHS is apparently more efficient.

## CLIP method[10]

Myc-PTB expressed HEK 293T cells were irradiated with 254 nm UV light for 15 minutes and then lysed. The lysate was then treated with anti-Myc beads to pull down RNAs binding PTB. After that, crosslinking products were separated by polyacrylamide gel electrophoresis. Then the RNAs binding PTB were transferred from PAGE to PVDF membrane, which was then treated with specific antibody (anti-Myc, mouse) and secondary antibody (rabbit) in the following procedure. Finally, the crosslinking products were detected on the membrane by staining the secondary antibody with enhanced chemiluminescent substrate in the dark room.

## PAR-CLIP method[11]

Myc-PTB expressed HEK 293T cells were pre-treated with 4-sU (4-thiouridine) to replace the normal uridines on RNA, then irradiated with 365 nm UV light to induce RNA and protein crosslinking. Cell lysate was then dealt by using the same procedure as described in the CLIP method.



**Figure 4-2 Crosslinking efficiency comparison of three crosslinking methods.**

Crosslinking products were immuno-precipitated and size separated with gel electrophoresis, which was then transferred to PVDF membrane. The membrane was treated with primary antibody(anti-Myc) and secondary and then illuminated with ECL. **Lane1**, Control group; **Lane 2**, 254 nm UV light for 15 minutes; **lane 3**, 365 nm UV light for 15 minutes; **lane 4**, cells were treated with 4-sU 10 hours; **lane 5**, cells were treated with 4-sU 10 hours before

crosslinking, then exposed to 365 nm UV light for 15 minutes; **lane 6**, cells were treated with AMT-NHS at 37 ℃ for 15 minutes; **lane 7**, cells were treated with AMT-NHS at 37 ℃ for 15 minutes and then 365 nm UV light for 15 minutes.

**PTB binding RNA immunoprecipitation and sequencing**

After verifying the efficiency of AMT-NHS based RNA and protein crosslinking and products immunoprecipitation, we started further study of PTB binding RNAs by construction of cDNA library and deep sequencing of target RNAs. (**Scheme 4-3**) The crosslinked cells were lysed and treated with RNase to partially digest protein binding RNAs, which were then pulled down with anti-Myc magnetic beads. The RNA fragment bound PTB was then treated with proteinase K to afford RNA fragment with protein residues on crosslinking sites. The residues were further removed by irradiation with 254 nm UV light to reverse the AMT crosslinking. The resulting pure RNA fragment was then ligated with a 3' adaptor and reverse transcribed to generate cDNA, which is then circularized and PCR amplified to generate a cDNA library. The cDNA library was finally sequenced and analyzed by comparing with reported data.
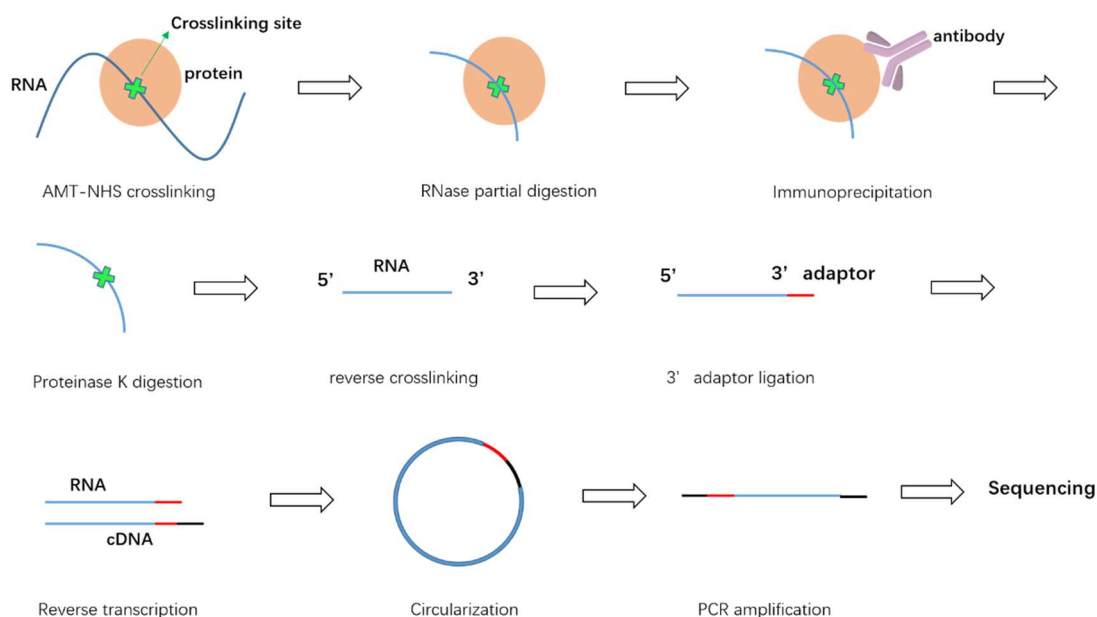


**Scheme 4-3 AMT-NHS based CLIP and sequencing.** In the first figure, the pink ball, the blue line and the green **X** denote RNA binding protein, protein binding RNA and crosslinking site, respectively. In the third figure, the purple **Y** shaped object is antibody (anti-Myc). In the sixth figure, the red bar indicates an introduced 3' adaptor, which is used for primer binding

during RNA reverse transcription. In the next figure, the black bar represents an RNA segment containing a restriction enzyme cutting site, which is cleaved to linearize the circularized DNA.

**Sequencing data analysis**

To test the suitability and stability of the AMT-NHS method, we prepared two parallel samples (**A** and **B**) for deep sequencing, which were obtained by the same protocol. (**Table 4-1**). In sample A, the total reads maps on 18335 genes, 30% of reads maps on rRNA and 48.77% of reads maps on coding RNA, which include 17.8 % reads mapping on exons. For sample B, the total reads maps on 10101 genes, 28.1% of reads maps on rRNA and 36.26% of reads maps on coding RNA, which include 12.98% reads mapping on exons. We also classified the data in terms of RPKM (Reads Per Kilobase per Million mapped reads), which is calculated by the formula as follows:

RPKM = total exon reads / mapped reads (million) * exon length (KB)

We found the RPKM distribution (RPKM > 1, 5, 10) of the two samples was very close.

**Table 4-1 A comparison of the reads mapping distribution between two parallel samples.**

| | Mapping | | | Gene number | | | |
|---|---|---|---|---|---|---|---|
| Sample | rRNA | Exons | Gene | Total | rpkm>1 | rpkm>5 | rpkm>10 |
| A | 30% | 17.80% | 48.77% | 18335 | 2698 | 1048 | 718 |
| B | 28.16% | 12.98% | 36.26% | 10101 | 2937 | 852 | 586 |

The intersection of the two sets of data was then calculated and we found that 88% of mapped transcripts in sample A are also mapped in sample B, which include 80% of reads that RPKM > 1. The high similarity of the data between sample **A** and sample **B** verified that AMT-NHS method was stable enough for study of RNA and protein interaction in living cell (**Figure 4-3**).
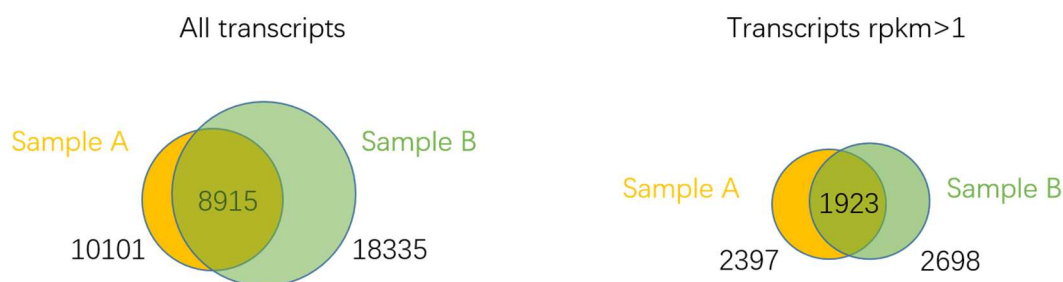
**Figure 4-3 A comparison of sequence data between sample A and sample B**

**Left side:** In sample A, reads maps on 10101 transcripts. In sample B, reads maps on 18335 transcripts. 8915 transcripts are mapped both by the reads of sample A and sample B. **Right side:** In sample A, reads maps on 2397 transcripts (rpkm > 1). In sample B, reads maps on 2698 transcripts (rpkm > 1) which include 1923 transcripts also mapped in sample A.

Afterward, we set out to test if the new method had some advantages when comparing with traditional methods. We assumed that the length flexible crosslinker could capture some protein binding RNAs that traditional CLIP could not. Hence, we compared the sequencing data with reported data by analyzing their intersection. We found that 36.6% of transcripts in sample A overlapped with reported data, and for sample **B**, 29.7% of mapped transcripts overlapped with reported data (**Figure 4-4**). We concluded that a large portion of protein binding RNAs captured by the new method are missed by traditional CLIP method. However, we can't draw a conclusion that the new method is able to obtain more information, because 55% transcripts mapped by traditional method are different from sample A, also, 75% mapped transcript by traditional method are different from sample B.

**Figure 4-4 Comparison of the sequenced data with reported data.** It was reported that
12195 PTB binding RNAs were captured by traditional method, which contains 6708 RNAs
that were also captured in sample A, and 3000 RNAs that were also captured in sample B.

At last, we listed the top 10 mapped transcripts and checked if these RNAs were
corresponding to alternative splicing (**Figure 4-5**);
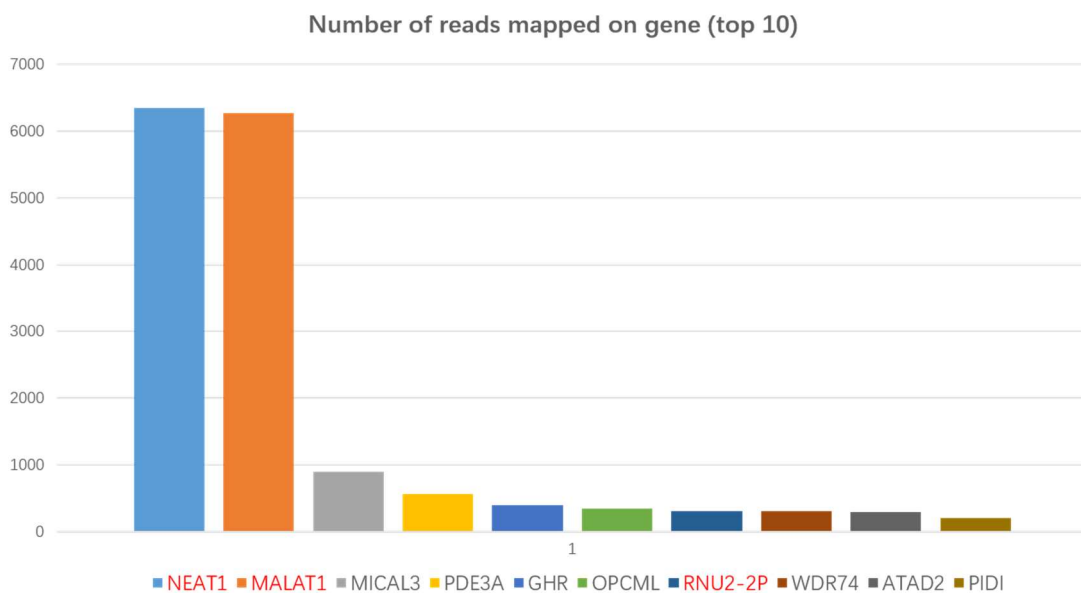


**Figure 4-5 Number of reads (top 10) mapped on gene**
The gene name marked in red is non-coding RNAs, and the rest of genes are protein coding
RNA. **MICAL3** (microtubule-associated monoxygenase, calponin and LIM domain containing
3), is processed by tissue specific alternative splicing, which has 5 isoforms. **PDE3A** (cGMP-
inhibited 3',5'-cyclic phosphodiesterase A) has more than 3 alternative splicing isoforms. GHR
(growth hormone receptor) has 4 alternative splicing isoforms. **OPCML** (opioid-binding
protein) has 4 alternative splicing isoforms. **WDR74** (WD repeat-containing protein 74) has 2
alternative splicing isoforms. **ATAD2** (ATPase family AAA domain-containing protein 2) has
2 alternative splicing isoforms. **PID1** (PTB-containing, cubilin and LRP1-interacting protein)
has 4 alternative isoforms.

**NEAT1 (nuclear enriched abundant transcript)**

NEAT1 is a nuclear long noncoding RNA, which is reported as an important cis acting element for regulation of alternative splicing[12, 13]. It is reported that NEAT1 is critical for the formation of paraspeckle, which is an important complex for post transcriptional regulation. NEAT1 is highly expressed in paraspeckle and nucleate it, so it is possible that PTB may be present in paraspeckle and regulate alternative splicing by effecting nucleation of NEAT1.

**MALAT1 (metastasis associated lung adenocarcinoma transcript 1)**

MALAT1 is a long noncoding RNA which is reported to localize the SR proteins in nuclear speckle by regulating their phosphorylation[14]. SR proteins containing long repeats of serine and arginine (SR represents serine and arginine) are involved in RNA splicing. SR proteins play important role in forming of spliceosome, they can assist U2 and U2AF to bind with cis acting element of pre-mRNA. It is possible that PTB may regulate alternative splicing by effecting phosphorylation of SR proteins, or it is possible that MALAT1 can localize PTB by regulating its phosphorylation.

**snRNA U2**

As mentioned before, U2 was recruited and located on branch site by U2AF protein[15], PTB prevents this interaction through occupying the polypyrimidine tract of intron. Which means PTB competes with U2AF to bind with polypyrimidine tract. It is also reported that PTB usually suppresses weak exon retention, but not sufficient to suppress robust exon, which is caused by insufficient binding site on polypyrimidine tract or the binding affinity is weaker than U2AF. So, in some cases, PTB binds on the polypyrimidine tract but does not stop U2AF binding around. As a result, U2 will be recruited and crosslinked with nearby PTB.

.

Except the above mentioned long non-coding RNAs that are involved in alternative splicing, the rest of top 10 mapped RNAs are protein coding RNAs, which all have several isoforms resulting from alternative splicing.

## 4.3 Summary and conclusion

In summary, we explored a new method to study RNA-protein interactions, which is based on photo reactivity of AMT. We synthesized a crosslinker consisting of AMT for binding to RNA and NHS ester for reacting with amino groups of protein. Besides, the introduction of the flexible linker enables the crosslinker to capture long range protein and RNA interactions, which is difficult to be realized by traditional methods.

We selected PTB, a widely reported RNA binding protein as the target for testing the function of the new method. We firstly verified the crosslinking efficiency of the AMT-NHS by in living cell crosslinking, immunoprecipitation and immunoblot analysis. We confirmed that the crosslinking efficiency of AMT-NHS was higher than CLIP and PAR-CLIP methods. We then pulled down PTB binding RNAs and constructed the cDNA library for deep sequencing. By comparing the sequence data of parallel samples, we verified that the method was stable enough for further application. At last, we compared the sequencing data with reported data. We found that a certain part of PTB binding RNAs captured by the AMT-NHS were related to alternative splicing, which was corresponding to reported results. In the meantime, we also found that a great part of PTB binding RNAs were missed when using traditional methods. We finally concluded that the AMT-NHS based CLIP was a stable and efficient method to capture protein and RNA interactions in mammalian cells. We suggest that a combination of AMT-NHS based CLIP and traditional CLIP methods will capture more comprehensive RNA and protein interactions.

## 4.4 Experimental section

### Materials and methods

The primers were synthesized from Sangon Biotech (Shanghai) Co., Ltd. Molecular Biology enzymes were purchased from New England BioLabs Inc. or Thermo Fisher Scientific. NTPs were purchased from Sangon Biotech (Shanghai) Co., Ltd. RNase inhibitor and TRIzol reagent were purchased from Thermo Fisher Scientific. Other chemicals were purchased from Sigma-Aldrich without further purification. The concentration of RNAs was measured on NanoDrop 2000 spectrophotometer. The SDS-PAGE gel images were obtained from LX-BIO-2800 (KCBF). The UV crosslinking experiments were carried out in CL-1000 UV crosslinker (UVP), which

has five 365 nm lamps (8 W). The reverse crosslining experiments were carried out in CL-1000 UV crosslinker (UVP), which has five 254 nm lamps (8 W).

**Buffer components**

| buffer | components |
|---|---|
| NP-40 buffer | 150 mM NaCl, 1.0% NP-40, 50 mM Tris-HCl, pH 8.0, Protease inhibitors |
| loading buffer | 4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris-HCl |
| Running buffer | 25 mM Tris base, 190 mM glycine, 0.1% SDS |
| Transfer buffer | 25 mM Tris base, 190 mM glycine, 20% methanol |
| Blocking buffer | 5% skimmed milk in TBST buffer |
| TBS (10x) | for 1 L, 24.23 g Trizma HCl, 80.06 g NaCl Dissolve in 800 mL distilled water pH to 7.6 with HCl and top up to 1 L |
| TBST | for 1 L, 100 mL TBS 10x, 900 mL distilled water, 1 mL Tween 20 |
| High stringency buffer | 15 mM Tris–HCl, pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% TritonX-100 |
| High salt buffer | 15 mM Tris–HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% TritonX-100, 1% Na-deoxycholate, 1 M NaCl |
| Low salt buffer | 15 mM Tris–HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% TritonX-100, 1% Na-deoxycholate, 150 mM NaCl |

**Reaction solution preparation**

| Composition | Initial concentration | volume | Final concentration |
|---|---|---|---|
| AMT-NHS | 30 mM | 67 µl | 500 µM |
| Digitonin | 5 mg/ml | 80 µl | 100 µg/ml |
| DMSO | - | 333 µl | - |
| PBS buffer | 1x | 3520 µl | - |

4ml

AMT-NHS stock solution was diluted in 333 μl DMSO and then PBS buffer was added. The mixture was vigorously shaken for 2 minutes. At last, digitonin was added and the final solution was completely mixed with a pipet, and the solution was kept at 37 ℃ for crosslinking.

**Cell culture and DNA plasmid transfection**

HEK 293T cells were cultured in humidified atmosphere containing 5% carbon dioxide, at 37 ℃.The complete culture medium consists of DMEM (dulbecco's modified eagle medium), 10% FBS (fetal calf serum), 2mM L-glutamine and 100u/ml penicillin-streptomycin. Cells were adhesive cultured on 10 cm dishes and passaged every after 48h. Cells were passaged as the following steps: The original culture medium was removed, the cells were washed once with 5 ml PBS and the supernatant was discarded. The cells were incubated with 5 ml trypsin for 5 minutes and added with another 5 ml complete culture medium to transfer all the mixture to a 15 ml centrifuge tube which was then centrifuged at 1500 rpm for 5 minutes. The supernatant was discarded and the cells were resuspended in 10 ml complete culture medium for dividing into 5 dishes. Each dish was added with another 10 ml complete culture medium and cultured at 37 ℃ (5% $CO_2$) . When the cells reach to about 80% confluency, the complete culture medium was removed and gently rinsed with 5 ml PBS buffer. pcmv-PTB-cMyc plasmid was transfected with lipofectamine 2000 (Thermo Fisher Scientific). The transfection protocol was performed following the supplier's instruction. After 5 h of infection (37 ℃ 5% $CO_2$), the supernatant was removed and 10 ml complete culture medium was added to each dish, which was cultured at 37 ℃ (5% $CO_2$).

**PTB and RNA crosslinking by AMT-NHS**

After 72 h of infection, the complete culture medium was removed and the cells were washed with 5 ml PBS buffer. Then each dish was gently added with 4 ml reaction solution and incubated at 37 ℃, 5% $CO_2$ for 30 minutes. After removing the reaction mixture, each dish was gently added with 2 ml PBS buffer and placed on ice. The ice cooled dishes were transferred to an UV crosslinker and irradiated with 365 nm UV light for 15 minutes (0.15J/cm$^2$).

**PTB and RNA crosslinking by iCLIP method**

After 72 h of infection, the complete culture medium was discarded and the cells were washed with 5 ml PBS buffer. The buffer was removed and another 5 ml ice cooled PBS buffer was added to each dish which was then moved to ice and irradiated with 254 nm UV light for 15 minutes ($0.15 \text{ J/cm}^2$).

**PTB and RNA crosslinking by PAR-CLIP method**

10 h before crosslinking, 1 ml 4-sU solution (1M stock solution) was added to the culture medium, and the cells were cultured at 37 ℃ (5% $CO_2$). After 12 h, the culture medium was discarded and cells were washed with 5 ml PBS buffer. The buffer was discarded and another ice cooled PBS buffer was added to each dish which was moved on ice and irradiated with 365 nm UV light for 15 minutes ($0.15 \text{ J/cm}^2$).

The AMT-NHS, iCLIP method and PAR-CLIP method crosslinked cells were treated with the same protocol in the rest of procedures.

**Extraction of proteins from adherent cells**

The crosslinking solution was removed from the cell dishes and the cells were washed with ice-cold PBS. Cells were scraped from the dishes by a cell scraper and collected in 15 ml centrifuge tubes which was centrifuged at 1500 rpm at 4℃ for 5 min. After removing of the supernatant, each tube was added ice-cold lysis buffer (1 ml for one dish) and 1/500 RNase I (2 μl for one dish). The tubes were incubated at 37℃ for 3 minutes with shaking and then put back on ice. The lysate was centrifuged at 12000 rpm, 4℃ for 15 min and then the supernatant was collected in a 15 ml centrifuge tube which was placed on ice.

**Anti-Myc magnetic beads preparation**

50 μl anti-Myc magnetic beads was pipetted into a 1.5 ml centrifuge tube and placed on a magnetic rack. The tube was added with 1 ml miliQ water, vortexed for 1 minute and moved to the magnetic rack, and the water was removed. The tube was added with 1 ml lysis buffer, vortexed for 1 minute and moved to the magnetic rack, and the lysis buffer was removed. The magnetic beads was washed twice with lysis buffer and used for immuno-precipitation.

**Immuno-precipitation**

The lysis buffer was removed from the anti-Myc beads by placing the tube on magnetic stand. Then the tube was added with 1 ml cell lysate and rotated at 4℃ for overnight. Afterwards, the supernatant was discarded and the beads was washed sequentially with 1x high stringency buffer, 1x high salt buffer and 1x low salt buffer.

**Reverse crosslinking and RNA purification**

The magnetic beads in low salt buffer was transferred to 10 cm dishes and irradiated with 365 nm UV light for 15 minutes (0.15 J/cm$^2$). Then the buffer and beads was collected in a 1.5 ml centrifuge tube which was put on a magnetic rack for separation. The supernatant was moved to a new ice cooled 1.5 ml centrifuge tube for further purification. 0.5 ml of the supernatant was transferred to a new 1.5 ml tube and added with 250 μl TE buffer saturated phenol, 240 μl chloroform and 10 μl isopentanol (25:24:1). The tube was shaken vigorously for 30 s and then centrifuged at 12000 rpm for 2 minutes. Then the upper aqueous phase (about 500 μl) was transferred to a new 1.5 ml tube which was then added with 480 μl chloroform and 20 μl isopentanol (24:2). The tube was shaken vigorously for 30 s and centrifuged at 12000 rpm for 5 minutes. Then the upper aqueous phase (about 500 μl) was transferred to a new 1.5 ml tube which was then added with 50 μl sodium acetate (3M, pH=4.8) and 500 μl isopropanol. The solution was mixed vigorously and precipitated overnight at -20℃. Finally, 1 μl glycogen (Thermo Fisher Scientific) was added to the solution if no precipitate was observed, the solution was mixed thoroughly and incubated overnight at -20℃.

**Running the crosslinker product with SDS-PAGE**

10 μl of the former prepared beads was added with protein loading buffer and incubated in boiling water for 5 minutes. Afterwards, the beads was discarded and the supernatant was loaded into 12% SDS-PAGE gel, as well as appropriate amount of pre-stained protein marker, then the gel was run for 1.5h at 100 V.

**Transferring the protein to PVDF membrane**

The gel below 35 KD ladder was cut off according to the protein marker. The PVDF membrane was activated by soaking in methanol for 1 minute and in transfer buffer for 5 minutes. The filter papers and sponges were also soaking in transfer buffer for 5 minutes, afterwards, the gel and membrane were assembled as a sandwich which was arranged as an order of anode – sponges – filter papers – PVDF membrane – gel – filter

papers – sponges – cathode. The sandwich was placed into a transfer buffer filled transfer cassette and run for 75 minutes at 200 mA.

## Immunoblotting

The membrane was blocked by incubating in blocking buffer for 1 h at 4℃ under agitation. Afterwards, the membrane was incubated in 15 ml primary antibody (anti-Myc, mouse, diluted with blocking buffer) buffer at 4℃ for 12 h while agitating. Then the membrane was washed with 10 ml TBST for 5 minutes, and repeat the wash for 3 times. The membrane was incubated with 15 ml secondary antibody (anti-mouse, rabbit, diluted in blocking buffer) for 1.5 h at 4℃ while agitating and washed 3 times with TBST buffer for 5 minutes at 4℃ under agitation. Afterwards, the membrane was incubated in ECL (enhanced chemiluminescent) substrate working solution for 5 minutes and transferred to a plastic sheet. At last images were obtained from an imaging system.

## References

1. Black, D. L., Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **2003,** *72* (1), 291-336.
2. Matlin, A. J.; Clark, F.; Smith, C. W. J., Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology* **2005,** *6* (5), 386-398.
3. Pan, Q.; Shai, O.; Lee, L. J.; Frey, B. J.; Blencowe, B. J., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* **2008,** *40* (12), 1413-1415.
4. Wikipedia.org. S.v. "alternative splicing." Retrieved June 22 2022 from https://encyclopedia.thefreedictionary.com/alternative+splicing

5. Shi, Y., Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* **2017,** *18* (11), 655-670.
6. Sohail, M.; Xie, J., Diverse regulation of 3' splice site usage. *Cell Mol Life Sci* **2015,** *72* (24), 4771-93.
7. Gooding, C.; Roberts, G. C.; Smith, C. W., Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. *RNA (New York, N.Y.)* **1998,** *4* (1), 85-100.
8. Wagner, E. J.; Garcia-Blanco, M. A., Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol* **2001,** *21* (10), 3281-8.

9.    Xue, Y.;   Zhou, Y.;   Wu, T.;   Zhu, T.;   Ji, X.;   Kwon, Y. S.;   Zhang, C.;   Yeo, G.;   Black, D. L.;   Sun, H.;   Fu, X. D.; Zhang, Y., Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **2009,** *36* (6), 996-1006.

10.   Sugimoto, Y.;   König, J.;   Hussain, S.;   Zupan, B.;   Curk, T.;   Frye, M.; Ule, J., Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome biology* **2012,** *13* (8), R67.

11.   Hafner, M.;   Landthaler, M.;   Burger, L.;   Khorshid, M.;   Hausser, J.;   Berninger, P.;   Rothballer, A.;   Ascano, M., Jr.;   Jungkamp, A. C.;   Munschauer, M.;   Ulrich, A.;   Wardle, G. S.;   Dewell, S.;   Zavolan, M.; Tuschl, T., Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **2010,** *141* (1), 129-41.

12.   Cooper, D. R.;   Carter, G.;   Li, P.;   Patel, R.;   Watson, J. E.; Patel, N. A., Long Non-Coding RNA NEAT1 Associates with SRp40 to Temporally Regulate PPARγ2 Splicing during Adipogenesis in 3T3-L1 Cells. *Genes* **2014,** *5* (4), 1050-63.

13.   Wegener, M.; Müller-McNicoll, M., Nuclear retention of mRNAs – quality control, gene regulation and human disease. *Seminars in Cell & Developmental Biology* **2018,** *79*, 131-142.

14.   Tripathi, V.;   Ellis, J. D.;   Shen, Z.;   Song, D. Y.;   Pan, Q.;   Watt, A. T.;   Freier, S. M.;   Bennett, C. F.;   Sharma, A.;   Bubulya, P. A.;   Blencowe, B. J.;   Prasanth, S. G.; Prasanth, K. V., The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **2010,** *39* (6), 925-38.

15.   van der Feltz, C.; Hoskins, A. A., Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. *Critical reviews in biochemistry and molecular biology* **2019,** *54* (5), 443-465.

# Chapter 5

# Incorporation of a photo-crosslinkable amino acid into protein by expending the genetic code

## 5.1 Introduction

Proteins are one of the most essential elements in organism. They play various roles like structure material, catalyst, transporter which are essential for the cell living. These sophisticated machineries are actually constructed by 20 basic amino acids. Chemists have been trying to modify natural proteins with noncanonical amino acids to endow them with some new properties and functions. Though the chemical modification of this type of bio-macromolecule is still a challenge. A series of modification approaches have been explored.

The most widely used chemical method is solid phase peptide synthesis [1]. Small size peptides (less than 100 amino acids) could be synthesized by this method with natural or unnatural amino acids (UAA). This method could also be used for ligation of several peptide fragments, which eliminates size limitation to some extent[2-4]. However, ligation on specific sites on a peptide needs protecting group manipulation, which limits its application.

Proteins are synthesized on the ribosome, in which amino acids are transported to mRNA templates by tRNA. Amino acids are recognized and loaded on tRNA with catalysis of aminoacyl tRNA synthetases (aaRS). The amino acid loaded tRNA matches template mRNA by codon and anticodon base paring, and the transported amino acids are assembled according to the sequence of the mRNA.

Biochemists have developed an in vitro UAA incorporation method based on protein translation process in vivo [5-8]. In this method, unnatural amino acids are chemically modified on a nucleotide by aminoacylation, which is then enzymatically incorporated in a blank codon suppressor (tRNA$_{BL}$). The decorated tRNA$_{BL}$ is then subjected to an in vitro protein translation system to synthesize target protein. By combining with micro-injection, unnatural amino acid modified tRNA$_{BL}$ could be incorporated into cells, followed by protein or peptide translation in living cells. However, this method is still limited by using stoichiometric artificial tRNAs and the instability of such tRNAs in living cells.

In 2001, Perter G. Shultz's lab developed a novel method for incorporation of a wide range of unnatural amino acids (UAA) into the defined sites of protein in living cells by expanding the genetic code[9-11]. In this approach, a special tRNA and aaRS pair is selected from a pre-constructed tRNA and aaRS library, which is then expressed in cells and react with UAA to form UAA modified tRNA. UAA attached tRNA is then transported into ribosome to suppress a blank codon of mRNA (**Scheme 5-1**)[12]. To ensure the specificity of UAA incorporation, the tRNA/aaRS pair should be orthogonal, which means the introduced tRNA only recognizes UAA with catalysis of the introduced aaRS but not endogenous aaRSs. In the meantime, the selected aaRS only catalyze aminoacylation of the introduced tRNA but not endogenous tRNAs. In addition, the blank codon must be only suppressed by the introduced tRNA but not by other endogenous tRNAs. Lastly, the UAA must be stable and permeable to living cells.
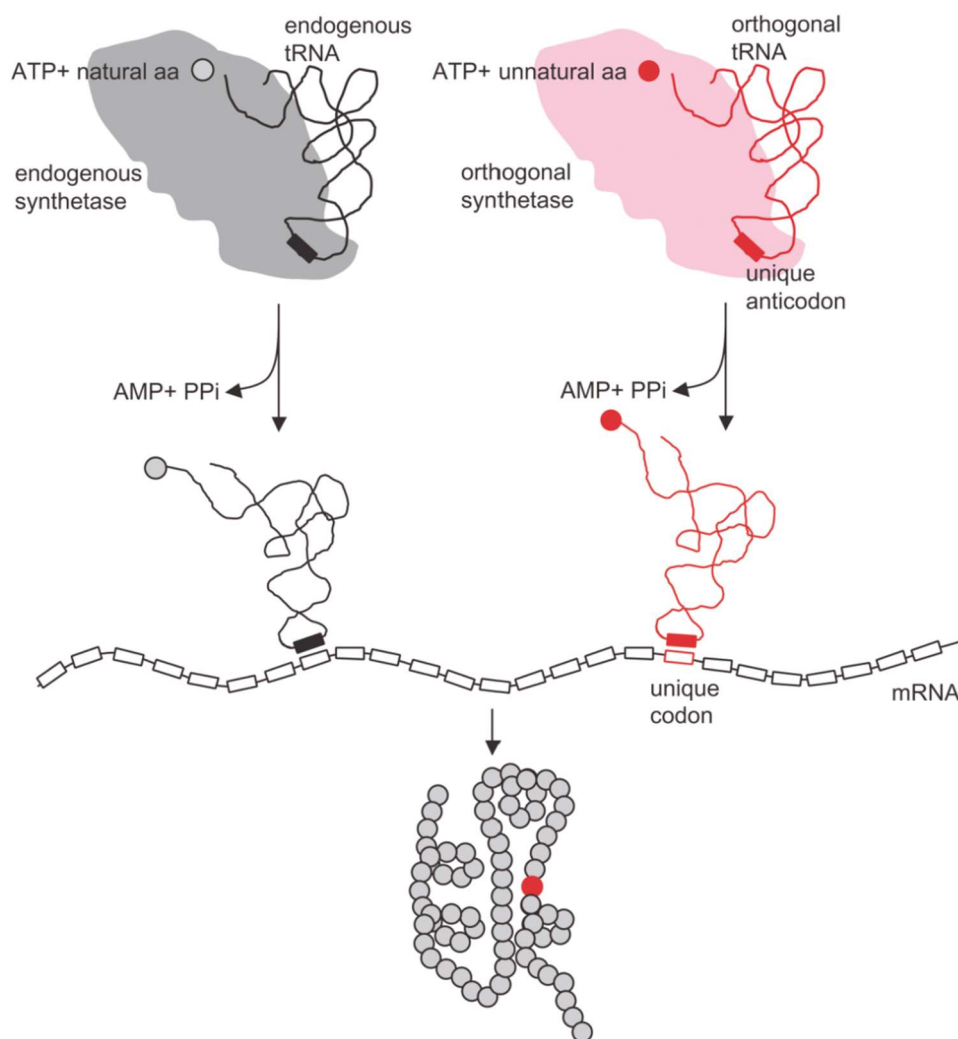


**Scheme 5-1 A general process of incorporating an unnatural amino acid into the specific site of a protein in vivo.**

In the beginning, Perter G. Shultz's group tried to obtain an orthogonal tRNA and aaRS system for UAA incorporation in Escherichia coli (*E. coli*) [13-16]. An mRNA harbouring amber stop codon (UAG) was chosen for recognizing the target tRNA, which was taken from an archaea Methanococcus jannaschii. The reason for choosing UAG as the blank codon is that UAG is the least used nonsense codon in *E. coli*. tRNA was chosen from archaea because a tRNA from different species is barely recognized by endogenous aaRSs of *E. coli*. Comparing to *E. coli* tRNA, Methanococcus jannaschii tRNA (*Mj*tRNA) has a different recognizing element with cognate aaRS (*Mj*TyrRS). The first *Mj*tRNA used in E. coli is tyrosyl-tRNA (*Mj*tRNA$_{CUA}^{Tyr}$), because *Mj*TyrRS has the smallest binding domain with anticodon loop of *Mj*tRNA$_{CUA}^{Tyr}$, which is favorable for mutation of the anticodon to CUA with least effect on aminoacylation efficiency (**Figure 5-1**).

*Mj*tRNA$_{CUA}^{Tyr}$ was later found to be recognized by endogenous aaRS of *E. coli*, thus a *Mj*tRNA$_{CUA}^{Tyr}$ mutant library was constructed by randomizing several nucleotides which did not significantly affect its basic function. Then a selection process was carried out to obtain an *Mj*tRNA$_{CUA}^{Tyr}$ mutant that only interacts with heterogenous *Mj*TyrRS.
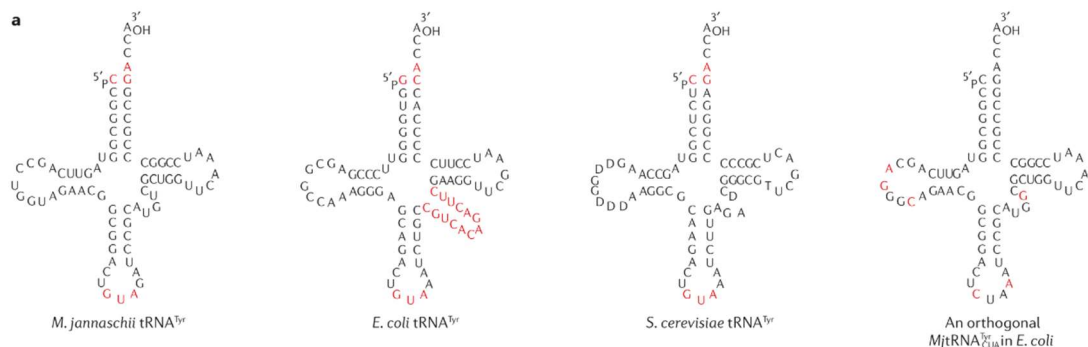


**Figure 5-1 An orthogonal *Mj*tRNA$_{CUA}^{Tyr}$ derived from *M. jannaschii* tRNA$^{Tyr}$.**

Briefly, the selection process started with a negative selection. A plasmid harbouring the gene of *Mj*tRNA$_{CUA}^{Tyr}$ library and toxic barnase (with UAG stop codons in several permissive sites) was transformed into *E. coli*. In this step, if *Mj*tRNA$_{CUA}^{Tyr}$ is recognized by endogenous aaRS, the cells expressing barnase will get killed. Cells survive only if the *Mj*tRNA$_{CUA}^{Tyr}$ has no function or can't be recognized by endogenous aaRS. *Mj*tRNA$_{CUA}^{Tyr}$ mutants were then extracted from surviving cells and transformed into *E. coli* for the positive selection. The *E. coli* cells used in positive selection contain the

gene of $Mj$TyrRS and $\beta$-lactamase, which have an UAG stop codon in a permissive site. In the presence of ampicillin, cells containing nonfunctional $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ mutants are not able to express $\beta$-lactamase and will be killed and only leaving clones containing $Mj$TyrRS aminoacylated $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$. As a result, the obtained $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ will only be aminoacylated by $Mj$TyrRS but not by $aa$RS of *E. coli*.

After the selection of the specific $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ for $Mj$TyrRS, an orthogonal $Mj$TyrRS also needs to be selected to specially recognize UAA and decorate it on $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$. Before selection, an $Mj$TyrRS library was constructed by randomizing several amino acids of the $Mj$TyrRS in the UAA binding pocket. Normally, the mutation sites are decided according to the structure of UAA. After the construction of $Mj$TyrRS library, several rounds of positive selection and negative selection were performed to obtain an orthogonal $Mj$TyrRS. Lastly, the gene of the selected $Mj$TyrRS was sequenced and co-expressed with the gene of $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ in cells to incorporate the desired UAA into UAG mutated target protein.

In this chapter, we are aiming to incorporate a photo-crosslinking reactive unnatural amino acid into a protein by expanding the genetic code. Unlike the natural amino acids absorbing UVC (100 nm – 280 nm) or UVB (280 nm – 315 nm), we intend to introduce an UAA that absorbing UVA (315 nm – 400 nm) in a protein, which should improve the photo-crosslinking reactivity of the protein in terms of UVA (365 nm) triggered protein/RNA crosslinking.

As reviewed in chapter 1, protein and RNA interaction researches were mainly based on photo-crosslink methods. Uridines crosslink with amino acids when irradiated by 254 nm UV light (CLIP method). 4-thiouridine crosslinks with amino acids when illuminated with 365 nm UV light (PAR-CLIP method) (**Scheme 5-2**)[17].

Natural amino acids involved in photo-crosslink methods are aromatic compounds like tyrosine, phenyl alanine and tryptophan. However, the abundance of aromatic amino acids in a protein is not high enough to occupy every protein/RNA binding site. Thus some protein/RNA interactions may not be captured if there is no aromatic amino acid around the protein/RNA binding sites. There are two ways to introduce one or several photo-crosslink reactive amino acid around the predicted binding sites. One is to incorporate a natural photo active amino acid by point mutation, the other one is to incorporate a more photo reactive aromatic UAA by expanding genetic code.

Here, we synthesized Cl-naphthol alanine a photo reactive compound and performed three rounds of positive selection and two rounds of negative reaction to obtain several orthogonal aaRSs. At last, we verified the fidelity of the aaRS by incorporating Cl-naphthol alanine into myoglobin and measuring its molecular weight.
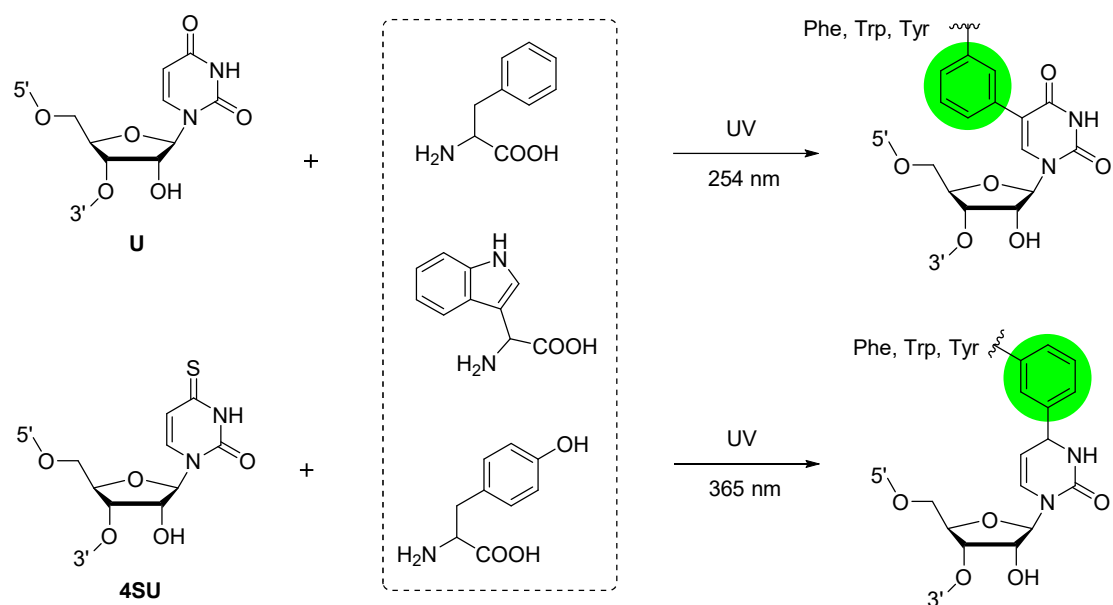


**Scheme 5-2 Aromatic amino acids based protein and RNA crosslinking method.**

## 5.2 Results and discussion

### Synthesis of Cl-naphthol alanine

We started synthesis of Cl-naphthol alanine from 6-methoxy-2-naphthoic acid, which was chlorinated by compound **2** in the presence of acetic acid. The resulting compound **3** was then esterified by methyl iodine to afford compound **4**. The ester was then reduced by lithium aluminum hydride to get alcohol **5**, which was chlorinated with thionyl dichloride to generate compound **6**. Afterwards, compound **7** was used to introduce an amino acid precursor to compound **6** to afford **8**. At last, we used 47% hydrobromic acid to break the methyl ether bond, deprotect acyl group of amine and hydrolyze ester to afford the target product **9** (**Scheme 5-3**). After obtaining Cl-naphthol alanine, we measured its absorption and fluorescence spectra (**Figure 5-2**). The absorption spectrum has two peaks at 330nm and 350nm. The emission spectrum was obtained by exciting the compound with 350 nm UV light.
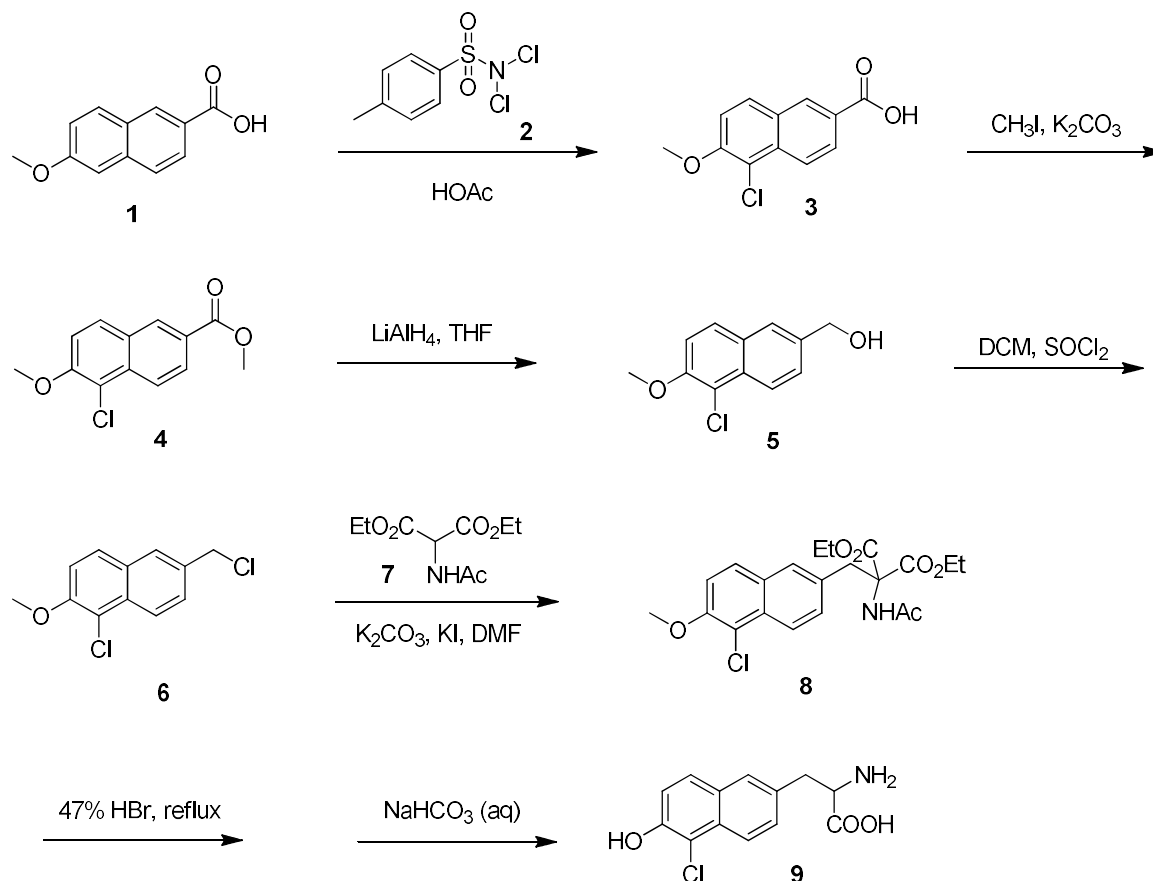
**Scheme 5-3 Synthesis of Cl-naphthol alanine**
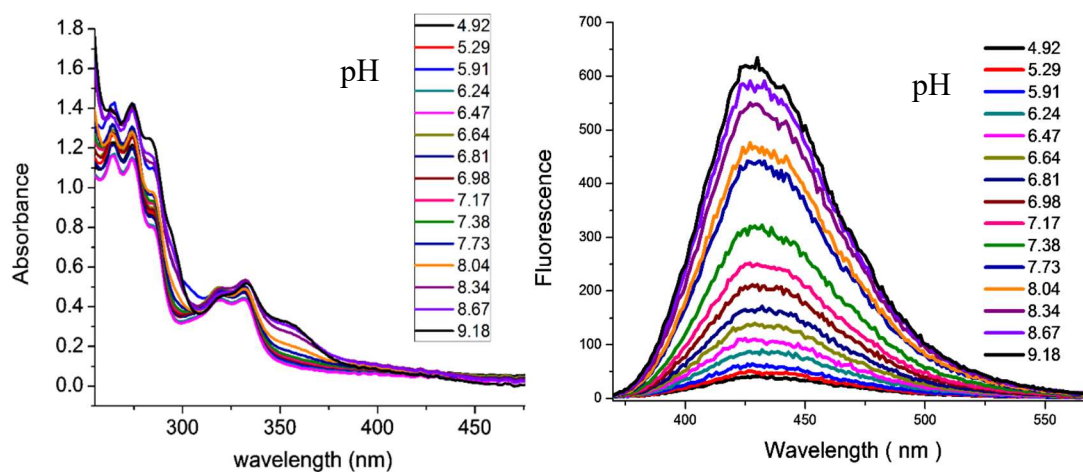


**Figure 5-2 Absorbance and fluorescence spectra of Cl-naphthol alanine.** The absorption and fluorescence spectra of Cl-naphthol alanine were measured at different pH.

## Selection of specific aaRS for Cl-naphthol alanine in *E. coli*

To obtain one or several specific *Mj*TyrRS for ligation of Cl-naphthol alanine and *Mj*tRNA$_{\text{CUA}}^{\text{Tyr}}$, several rounds of selection were performed to exclude the *Mj*TyrRSs that did not work in *E. coli* or catalyzed aminoacylation of natural amino acids. As a result, the surviving *Mj*TyrRS only catalyzes aminoacylation of Cl-naphthol alanine. The selection process consists of three rounds of positive selection and two rounds of negative selection (**Scheme 5-4**).

## Positive selection

The selection process started from positive selection, which is based on resistance to chloramphenicol. First of all, we constructed an *Mj*TyrRS mutant library (pBK-lib-jw1) by randomizing Tyr32, Leu65, Phe108, Gln109, Asp158, Leu162 and any one of the six residues (Ile63, Ala67, His70, Tyr114, Ile159, Val164), which was either mutated to Gly or kept unchanged. After that, we constructed a plasmid pREP(2)/YC that harboured the gene of *Mj*tRNA$_{\text{CUA}}^{\text{Tyr}}$, the GFP gene under control of T7 promoter, the T7 RNA polymerase gene with a TAG codon at a permissive site and the chloramphenicol acetyltransferase (CAT) gene with a TAG codon at a permissive site. We then transformed plasmid pBK-lib-jw1 into pREP(2)/YC contained *E. coli* cells, which were then cultured in the presence of chloramphenicol and Cl-naphthol alanine. As a result, if the *Mj*tRNA$_{\text{CUA}}^{\text{Tyr}}$ mutants recognize homogenous amino acids or Cl-naphthol alanine, it will express CAT to keep the cell alive in the presence of chloramphenicol. If the *Mj*tRNA$_{\text{CUA}}^{\text{Tyr}}$ mutants neither recognize natural amino acids nor Cl-naphthol alanine, the cells will not express CAT and be killed by chloramphenicol. Besides, green fluorescence was detected from the survived clones.

The *Mj*TyrRS mutants obtained from positive selection consist of two types of *Mj*TyrRS. Type I, *Mj*TyrRS aminoacylates *Mj*tRNA$_{\text{CUA}}^{\text{Tyr}}$ with Cl-naphthol alanine; Type II, *Mj*TyrRS aminoacylates *Mj*tRNA$_{\text{CUA}}^{\text{Tyr}}$ with natural amino acids of host cells. Thus a round of negative selection is needed to exclude the type II *Mj*TyrRS mutants in the following process.

## Negative selection

In order to exclude *Mj*TyrRS mutants that recognize homogenous amino acids, a toxic barnase dependent negative selection system was constructed. We constructed a

plasmid (pLWJ17B3) that harbouring the gene of $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ and barnase with three TAG codons at permissive sites.

Cells survived in positive selection were collected and the plasmid pBK-lib-jw1 was then extracted for the following procedure. Afterwards, we transformed plasmid pBK-lib-jw1 into pLWJ17B3 contained *E. coli* competent cells, which were then cultured in the absence of Cl-naphthol alanine. In this step, if the $Mj$TyrRS aminoacylate $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ with natural amino acids, the TAG codons of barnase gene will be suppressed and toxic barnase will be expressed to kill the cells. If the $Mj$TyrRS are not able to recognize natural amino acids, the barnase will not be expressed and the cells will survive. We extracted the plasmid pBK-lib-jw1 from the surviving cells and performed the second rounds of positive and negative selection. After the third round of positive selection, we finally obtained an orthogonal $Mj$TyrRS/ $Mj$tRNA$_{\text{CUA}}^{\text{Tyr}}$ pair.
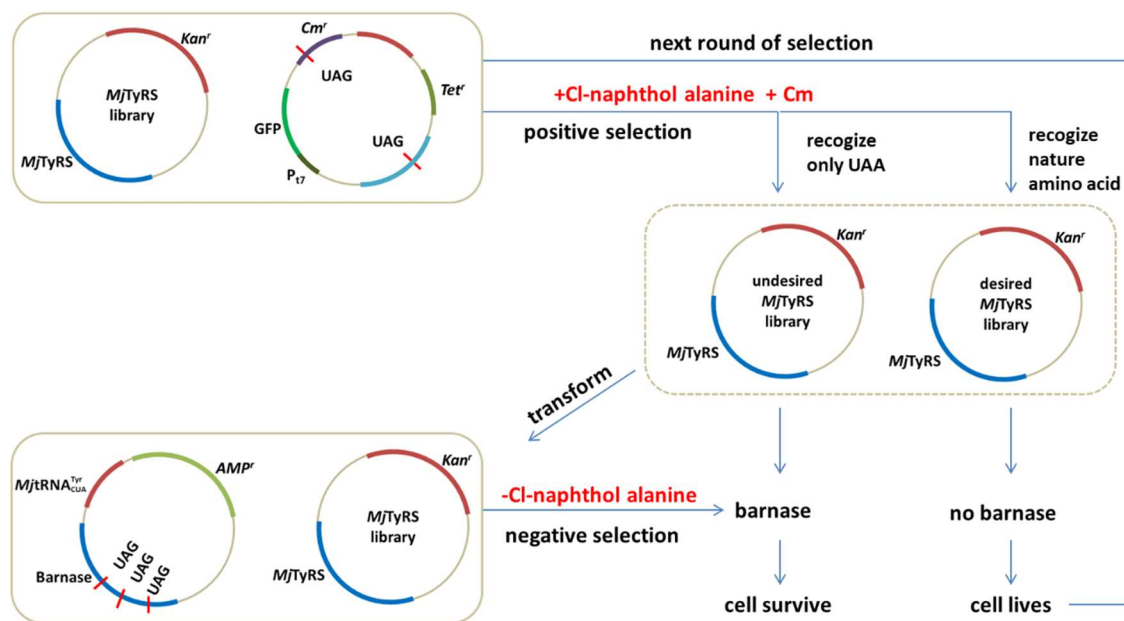


**Scheme 5-4 Combination of positive and negative selection process.**

In order to improve the fidelity of the $Mj$TyrRS, we performed the second round of positive-negative selection and the last round of positive selection. In the last positive selection, plasmid (pBK-lib-jw1) extracted from second round of negative selection was transformed into positive selection cells and cultured in the presence of Cl-naphthol alanine. The survived clones were selected and transferred to 96-well plate, which were filled with LB. The 96-well cultured cells were then replica-spotted on two sets of culture plates. One set of culture plate was supplemented with Cl-naphthol alanine, tetracycline, kanamycin and chloramphenicol at concentrations of 60, 80, 100 and 120

μg/mL. The other set of culture plate was supplemented with tetracycline, kanamycin and chloramphenicol at concentrations of 0, 20, 40 and 60 μg/mL, but no Cl-naphthol alanine.

After culturing the two sets of plate for 60 h, we found that some clones died in low concentration of chloramphenicol in the absence of Cl-naphthol alanine, but survived in high concentration of chloramphenicol in the presence of Cl-naphthol alanine (**Figure 5-7**), which meant these clones containing *Mj*TyrRS that only interacted with Cl-naphthol alanine but not with endogenous amino acids. We finally obtained clones harbouring the orthogonal *Mj*ClNARS (ClNARS stands for Cl-naphthol alanine aminoacyl tRNA synthetase) that can precisely incorporate Cl-naphthol alanine into the specific site of target protein. We picked the surviving clones and extracted the pBK-ClNARS, which were then sequenced to learn the mutation sites of *Mj*ClNARS (**Table 5-1**).
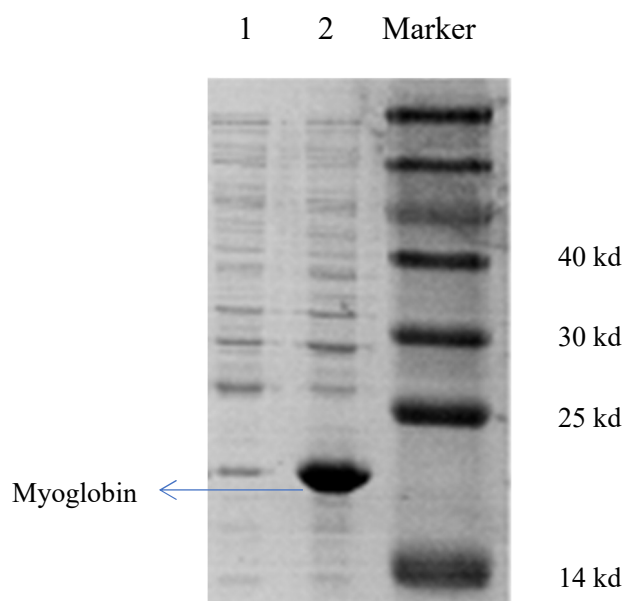


**Figure 5-3 Coomassie-Blue stained SDS-PAGE of Cl-naphthol alanine incorporated myoglobin (arrow) expressed in the presence (lane 2) and absence (lane 1) of UAA.**

To learn the fidelity of The *Mj*TyrRS/ *Mj*tRNA$_{CUA}^{Tyr}$ pair, we decided to incorporate Cl-naphthol alanine into sperm whale myoglobin. We firstly constructed a plasmid pBAD/JYAMB-Ser4TAG that harboured the gene of TAG mutated myoglobin and *Mj*tRNA$_{CUA}^{Tyr}$. We then co-transformed pBK-ClNARS with pBAD/JYAMB-ser4TAG into TOP10 *E. coli* competent cells. Lastly, we expressed myoglobin mutant, purified the protein and analyzed its molecular weight with SDS-PAGE (**Figure 5-3**) and MALDI-TOF MS (**Figure 5-4**).
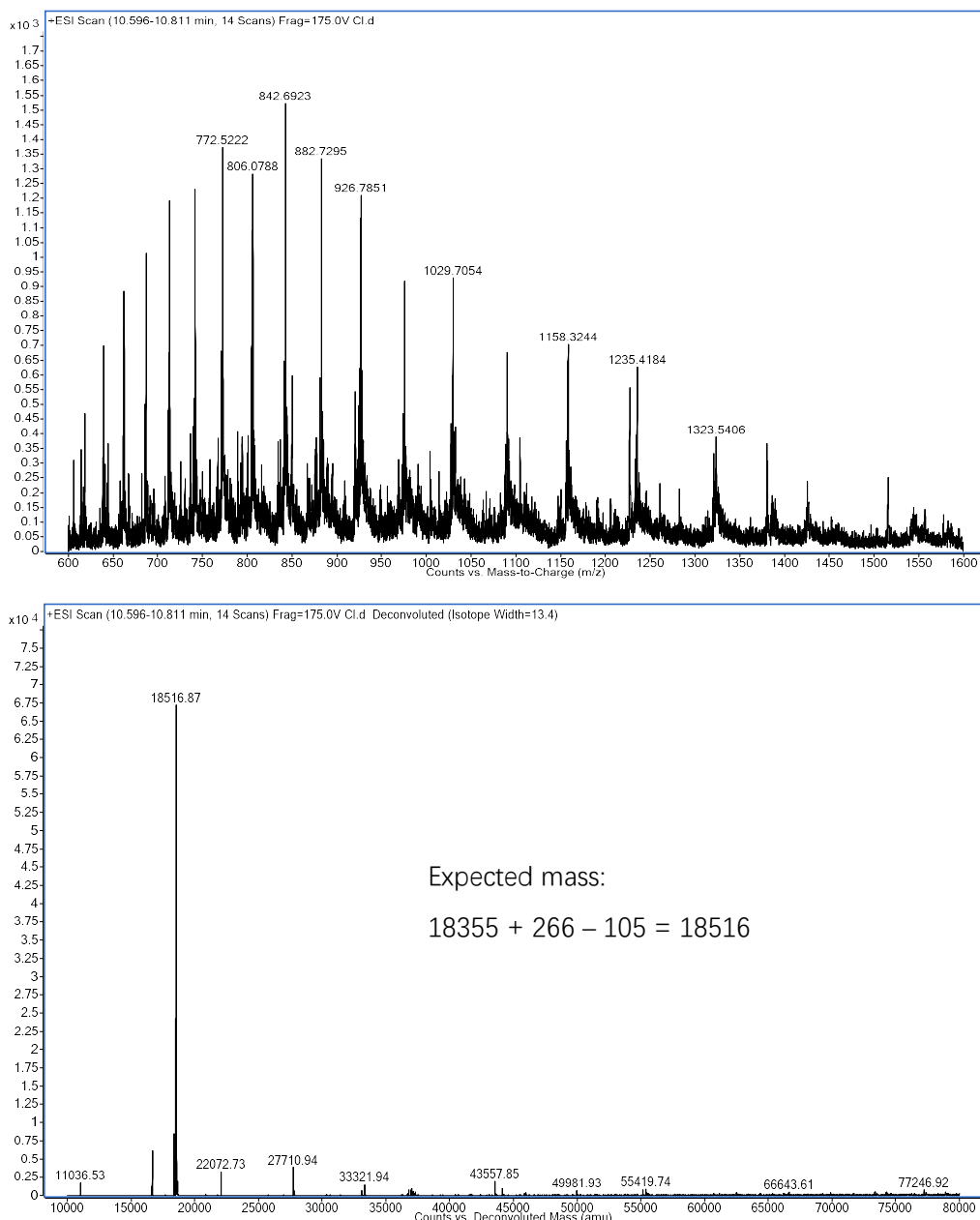


**Figure 5-4 ESI-MS spectra of Cl-naphthol alanine incorporated Myoglobin.** The top picture is the original spectra (m/z). The bottom picture is the deconvoluted spectra. Expected mass is 18516 Da, found at 18516 Da.

## 5.3 Summary and conclusion

In this chapter, we synthesized an unnatural amino acid Cl-naphthol alanine. We prepared an aaRS mutant library and constructed a selection system for generation of a specific aaRS to aminoacylate a blank codon suppressor (tRNA$_{CUA}$) with Cl-naphthol alanine. In order to obtain a highly specific aaRS, we performed three rounds of positive selection and two rounds of negative selection. We finally obtained several clones harbouring orthogonal $Mj$TyrRS/ $Mj$tRNA$_{CUA}^{Tyr}$ pairs. We learnt the mutated sites of the $Mj$TyrRS by sequencing the extracted selection plasmids. At last, we verified the specificity of selected $Mj$TyrRS by incorporating Cl-naphthol alanine into TAG mutated Myoglobin.

In the future, the Cl-naphthol alanine could be incorporated into RNA binding proteins to study RNA and protein interactions through UVA based crosslinking method. Moreover, it could be decorated on any predicted interaction sites of target protein to highly improve information capture efficiency.

## 5.4 Experimental section

**Materials and methods**

The primers were synthesized from Sangon Biotech (Shanghai) Co., Ltd. Molecular Biology enzymes were purchased from New England BioLabs Inc. or Thermo Fisher Scientific. NTPs were purchased from Sangon Biotech (Shanghai) Co., Ltd. Other chemicals were purchased from Sigma-Aldrich without further purification. Fluorescence spectra were measured on Varioskan Flash spectral scanning multimode reader (Thermo Scientific) and absorbance spectra were measured on NanoDrop 2000 spectrophotometer. NMR spectra were recorded on Varian Unity Inova (500 MHz). The agarose gel or the SDS-PAGE gel images were obtained from LX-BIO-2800 (KCBF). Protein purification was performed at AKTA UPC 900 FPLC system (GE healthcare). Protein mass spectra were recorded on a thermo LTQ-orbitrap at the IBP, CAS (Beijing, China).

**Synthesis of Cl-naphthol alanine**

**Synthesis of 5-chloro-6-methoxy-2-naphthoic acid (3)**

Under argon, 1 g 6-methoxy-2-naphthoic acid (5 mmol) was dissolved in 35 ml acetic acid, then 0.6 g N,N-dichloro-4-methylbenesulfonamide (2.5 mmol) was added and the reaction was stirred at 40 ℃. After 1h, the reaction was cooled to room temperature, reaction solution was filtrated and the solid was washed by acetic acid to obtain compound **3** as white solid (1.07 g, 92%). The compound was directly used for the next step without any further purification.

NMR (400 MHz, d-DMSO): σ, 13.03 (s, 1H), 8.62 (s, 1H), 8.18 (d, J = 9.2 Hz, 1H), 8.14 (d, J = 8.8 Hz, 1H), 8.07 (dd, J = 8.8 Hz, 1.2 HZ, 1H), 7.65 (d, J = 9.2 Hz, 1H), 4.03 (s, 3H)

**Synthesis of methyl 5-chloro-6-methoxy-2-naphthoate (4)**

Under argon, 0.5 g 5-chloro-6-methoxy-2-naphthoic acid (2.1 mmol) was dissolved in 20 ml DMF, followed by adding 0.44 g potassium carbonate (3.2 mmol) and 0.33 g iodomethane (2.3 mmol). The reaction mixture was stirred at room temperature for overnight. The reaction was extracted with ethyl acetate, the organic phase was washed with 30 ml water for three times and with brine for three times. The organic phase was dried with sodium sulfate and vacuum evaporated to afford compound **4** (0.47 g, 90%), which was directly used for next step without any further purification.

NMR (400 MHz, d-DMSO): σ, 8.63 (d, J = 1.6 Hz, 1H), 8.19 (d, J = 9.2 Hz, 1H), 8.14 (d, J = 8.8 Hz, 1H), 8.05 (dd, J = 9.2 Hz, 1.6 HZ, 1H), 7.66 (d, J = 9.2 Hz, 1H), 4.04 (s, 3H), 3.91 (s, 1H)

**Synthesis of (5-chloro-6-methoxynaphthalen-2-yl)methanol (5)**

Under argon, 0.091 g LiAlH$_4$ (2.4 mmol) was dissolved in 10 ml dry THF and the solution was cooled to 0 ℃. Afterwards, 5 ml THF solution of methyl 5-chloro-6-methoxy-2-naphthoate (0.3 g, 1.3 mmol) was slowly added to the LiAlH$_4$ solution. The reaction was warmed to room temperature and stirred for 30 minutes. Then, 2 ml HCl (1M) was added and the reaction was extracted with ethyl acetate. The organic phase was washed with water and brine, then dried with sodium sulfate and vacuum evaporated to afford compound **5** (0.24 g, 92%), which was directly used for next step.

NMR (400 MHz, d-DMSO): σ, 8.04 (d, J = 8.4 Hz, 1H), 7.94 (d, J = 8.8 Hz, 1H), 7.85 (s, 1H), 7.59 (dd, J = 8.8 Hz, 1.6 HZ, 1H), 7.53 (d, J = 9.2 Hz, 1H), 5.35 (t, J = 5.6 Hz, 1H), 4.66 (d, J = 5.6 Hz, 2H), 3.99 (s, 3H)

**Synthesis of 1-chloro-6-(chloromethyl)-2-methoxynaphthalene (6)**

Under argon, 0.2 g (5-chloro-6-methoxynaphthalen-2-yl)methanol (0.9 mmol) was dissolved in 10 ml dry dichloromethane, and the solution was cooled to 0 ℃. In the meantime, 0.36 g thionyl chloride (3 mmol) was dissolved in 5 ml dichloromethane and the solution was slowly dropped to the former solution. The reaction was stirred at room temperature. After 1 h, the reaction mixture was vacuum evaporated to dryness and the generated product **6** was directly used for next step without any further purification （0.19 g, 90%）.

NMR (400 MHz, d-DMSO): σ, 8.09 (d, J = 8.8 Hz, 1H), 8.01 (s, 1H), 7.97 (d, J = 9.2 Hz, 1H), 7.67 (dd, J = 8.8 Hz, 1.2 HZ, 1H), 7.59 (d, J = 9.2 Hz, 1H), 4.93 (s, 2H), 4.00 (s, 3H)

**Synthesis of diethyl 2-acetamido-2-((5-chloro-6-methoxynaphthalen-2-yl)methyl) malonate (8)**

Under argon, 0.32 g 1-chloro-6-(chloromethyl)-2-methoxynaphthalene (1.3 mmol) was dissolved in 20 ml DMF, which was followed by adding 0.32 g diethyl 2-acetamidomalonate (1.4 mmol), 0.19 g potassium carbonate (1.4 mmol) and 0.015 g potassium iodide (0.1 mmol), the reaction was stirred at room temperature for overnight. The reaction was extracted with ethyl acetate, the organic phase was alternately washed with 30 ml water and 30 ml brine for three times. The organic phase was dried with sodium sulfate and vacuum evaporated to afford compound **8** (0.42 g, 75%), which was directly used for next step.

NMR (400 MHz, d-DMSO): σ, f8.07 (s, 1H), 7.99 (d, J = 8.8 Hz, 1H), 7.92 (d, J = 9.2 Hz, 1H), 7.54 (m, 2H), 7.24 (dd, J = 8.4 Hz, 1.2 Hz, 1H), 4.17 (qd, J = 7.2 Hz, 2.0 Hz, 4H), 3.98 (s, 3H), 3.59 (s, 2H), 1.98 (s, 3H), 1.19 (t, 6H)

**Synthesis of 2-amino-3-(5-chloro-6-hydroxynaphthalen-2-yl)propanoic acid (9)**

Under argon, 1 g diethyl 2-acetamido-2-((5-chloro-6-methoxynaphthalen-2-yl) malonate (2.4 mmol) was dissolved in 15 ml hydrobromic acid (47%), and the reaction mixture was heated to reflux for 20 h. Afterwards, the acid was removed by vacuum evaporation. The resulting solid was dissolved in 10 ml water and the solution pH was adjusted to 5-6 by adding 2M sodium carbonate. The solution was filtrated to afford compound **9** as white solid (0.6 g, 95%).

NMR (400 MHz, D₂O): σ, 7.97 (d, $J$ = 8.1 Hz, 1H), 7.60 (m, 2,1H), 739 (dd, $J$ = 8.1 Hz, 2.1 Hz, 1H), 7.07 (d, $J$ = 9.1 Hz, 1H), 3.56 (dd, $J$ = 7.6 Hz, 5.2 Hz, 1H), 3.11 (dd, $J$ = 13.6 Hz, 5,2 Hz, 1H), 2.92 (dd, $J$ = 13.6 Hz, 7.6 Hz, 1H). MS (ESI⁺) (m/z): found 266.0 [M+H]⁺; calculated, 266.1, [M+H]⁺.
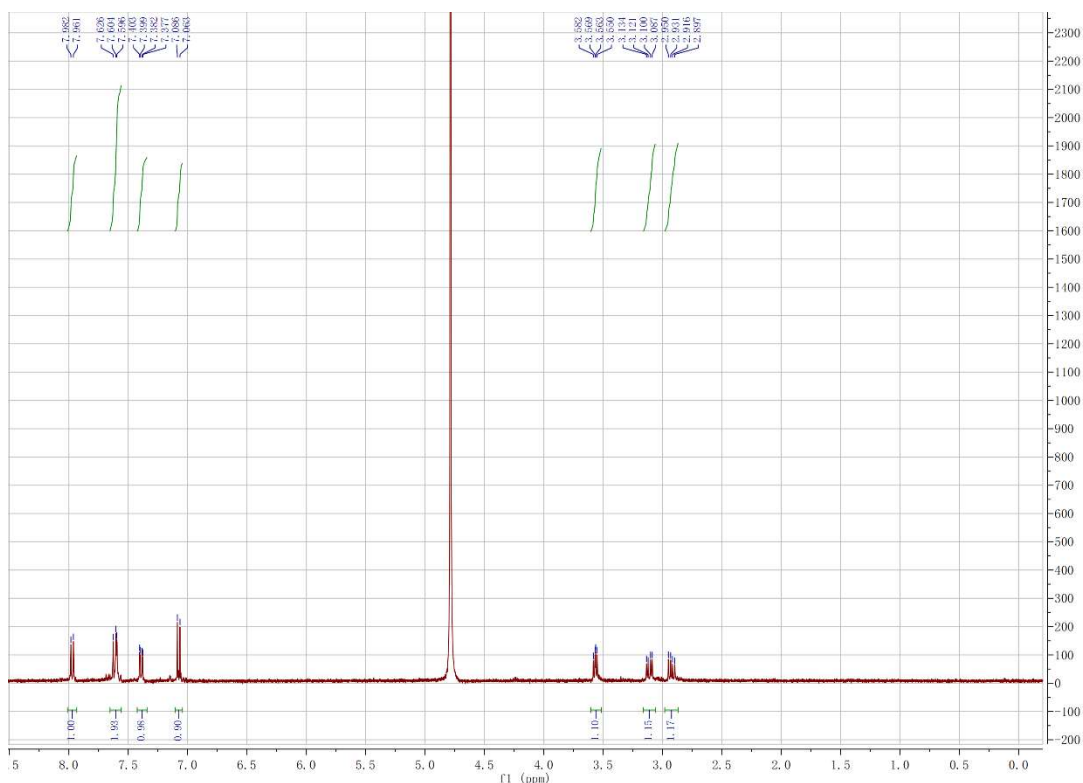


**Figure 5-5 Proton NMR spectrum of Cl-naphthol alanine.**

Plasmids used: Plasmid pBK-lib-jw1 encodes a library of M. Jannaschii tyrosyl tRNA synthetase (TyrRS) mutants randomized at residues Tyr32, Leu65, Phe108, Gln109, Asp158 and Leu162; and any one of the six residues (Ile63, Ala67, His70, Tyr114, Ile159, Val164) was either mutated to Gly or kept unchanged. Plasmid pREP(2)/YC encodes $Mj$tRNA$_{CUA}^{Tyr}$, the chloramphenicol acetyltransferase (CAT) gene with a TAG codon at residue 112, the GFP gene under control of the T7 promoter, and a Tetr marker; plasmid pLWJ17B3 encodes $Mj$tRNA$_{CUA}^{Tyr}$ under the control of the lpp promoter and rrnC terminator, the barnase gene (with three amber codons at residues 2, 44 and 65) under the control of the S4 ara promoter, and an Amp marker. Plasmid pBAD/JYAMB-4TAG encodes the mutant sperm whale myoglobin gene with an arabinose promoter and rrnB terminator, $Mj$tRNA$_{CUA}^{Tyr}$ with an lpp promoter and rrnC terminator, and a tetracycline resistance marker.

## Selection method of *Mj*TyrRS

### Electro competent cell preparation

In order to reach a high capacity library (> $10^9$ different clones) of the aaRS mutant, it is critical to prepare efficient electroporation competent cells for the transformation of the aaRS library. At the first day, we prepared 5 L autoclaved MiliQ water, 2 L autoclaved 10% glycerol, four sterilized 500 ml centrifuge bottles, more than two hundred 0.5 ml Eppendorf tubes and 900 ml sterilized LB without antibiotics. We chilled these materials except LB in 4°C cold room. We then streaked frozen stocked cells onto LB plate (no antibiotics) and cultured in 37°C for overnight. On the second day, we picked a single colony from the LB plate and inoculated 100 ml LB (no antibiotics), which was shaken at 37°C for overnight. On the third day, we inoculated the 100 ml cultured cells into 900 ml LB and shook at 37°C until OD600 reached 0.4-0.5. We chilled the cells on ice for 0.5 h in the cold room. After cooling the centrifuge rotor and bottles for 5 minutes, we poured 250 ml chilled cells into each centrifuge bottle, which was centrifuged at 3500 rpm for 15 minutes at 4°C. The supernatant was removed and the cells were resuspended in 1 L ice chilled water. Cells were spun at 3500 rpm at 4°C for 15 minutes and the wash procedure was repeated twice. After the last wash, we discarded the supernatant and resuspended the cells with 20 ml ice chilled 10% glycerol in 50 ml centrifuge tube, which was centrifuged at 3500 rpm at 4°C for 15 minutes. We then discarded the supernatant and resuspended the cells with 2 ml ice chilled 10% glycerol. We transferred every 100 μl cell portion to an ice chilled 0.5 ml Eppendorf tube, which were then frozen in liquid nitrogen. We lastly stored the frozen electro-competent cells at -70°C for use.

### Genetic selection of the specific aaRS mutant for Cl-naphthol alanine

We constructed the pBK-lib-jw1plasmid harboring $2\times10^9$ MjTyrRS mutants by using standard PCR method. DH10B *E.coli* cells containing positive plasmid pREP(2)/YC were transformed with pBK-lib-jw1, the transformants were then recovered in SOC until OD600 reach around 1 (about 1 hour). The cells were centrifuged and washed twice with LB media. After the last wash, cells were centrifuged and resuspended with proper volume of LB to make OD600 reach 3. Then, every 0.5 ml cells were plated on a LB-agar plate, which was supplemented with kanamycin (50 mg/ml), chloroamphenicol (60 mg/ml), tetracycline (15 mg/ml) and Cl-naphthol alanine (1 mM).

The plates were incubated at 37 °C for 60 hours. Surviving cells were collected by adding LB medium (with 50 mg/ml kanamycin) and scrapped from the plates. Cells were cultured in LB medium at 37 °C for 2 hours and lysed, followed by extraction and purification of the plasmids with agarose gel electrophoresis. Plasmid pBK-lib-jw1 was collected from 2-3 K band of the agarose gel (**Figure 5-6**). The collected pBK-lib-jw1 was extracted and purified as described in the DNA extraction kit (Promega: PureYield™ Plasmid Miniprep System) manual. The negative selection electro-competent cells harbouring pLWJ17B3 were thawed and transformed with 3 μl purified pBK-lib-jwl. The cell suspension was mixed thoroughly and transferred into pre-chilled electrotransfection cuvette. Cells were transfected at 4500 V, for 5 seconds with one pulse. The content of the cuvette was added to 700 μl LB medium and the cells were transferred into 1.5 ml Eppendorf tubes, which were incubated at 37 °C for 1 h. We then plated 1 μl, 10 μl or 100 μl cells on LB plate supplemented with ampicillin (50 mg/ml) and kanamycin (50 mg/ml) to test the transfection efficiency ($> 10^8$). The rest of the cells were plated on LB plates supplemented with 2% arabinose, ampicillin (50 mg/ml) and kanamycin (50 mg/ml), which were incubated at 37 °C for 8 h. The surviving cells were collected and the pBK mutants were extracted as described above. The pBK mutants were carried through another round of positive selection, negative selection and last round of positive selection.
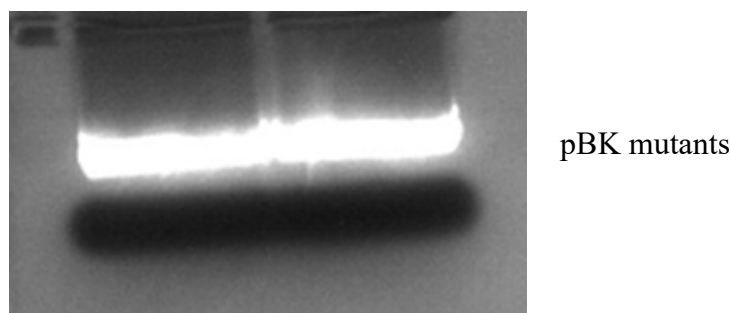


pBK mutants

**Figure 5-6 Agarose gel electrophoresis of pBK mutants which were extracted from clones that survived the positive selection.**

**Verification of the target *Mj*TyRS mutant *Mj*ClNARS (ClNA stands for Cl-naphthol alanine)**

96 surviving clones were picked after last positive selection and cultured in 96-well plates filled with 2 ml LB at 37 °C for 8 h. The cells were replica-spotted on two sets of LB-agar plates. One plate was supplemented with 1 mM Cl-naphthol alanine,

tetracycline (15 mg/ml), kanamycin (50 mg/ml) and chloramphenicol at concentrations of 60, 80, 100 and 120 mg/ml. The other plate was supplement with tetracycline (15 mg/ml), kanamycin (50 mg/ml) and chloramphenicol at concentrations of 0, 20, 40 and 60 mg/ml. After incubating the plates at 37 °C for 60 h, some clones were visible (survived) in the plate containing 80 mg/ml chloramphenicol and 1 mM Cl-naphthol alanine, but invisible (died) in the plate containing 40 mg/ml chloramphenicol but no Cl-naphthol alanine (**Figure 5-7a**). We picked these clones and inoculated them into a 96-well plate, which was replica-spotted on the same two sets of plates. After 48 h incubation at 37 °C, we picked the clones that survived in 80 mg/ml chloramphenicol in the presence of Cl-naphthol alanine, but died in 40 mg/ml chloramphenicol in the absence of Cl-naphthol alanine (**Figure 5-7b**). The surviving clones were separately cultured and plasmid mutants (*Mj*ClNARS) were extracted for the following DNA sequencing and protein expression.

**a**
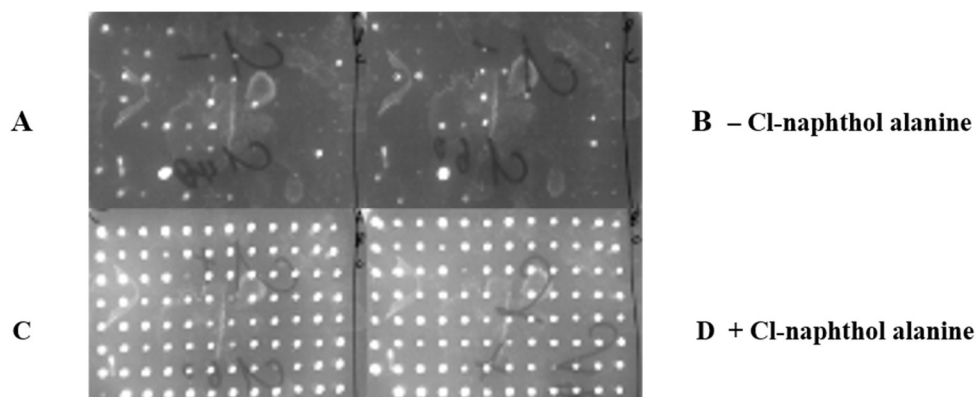


**b**
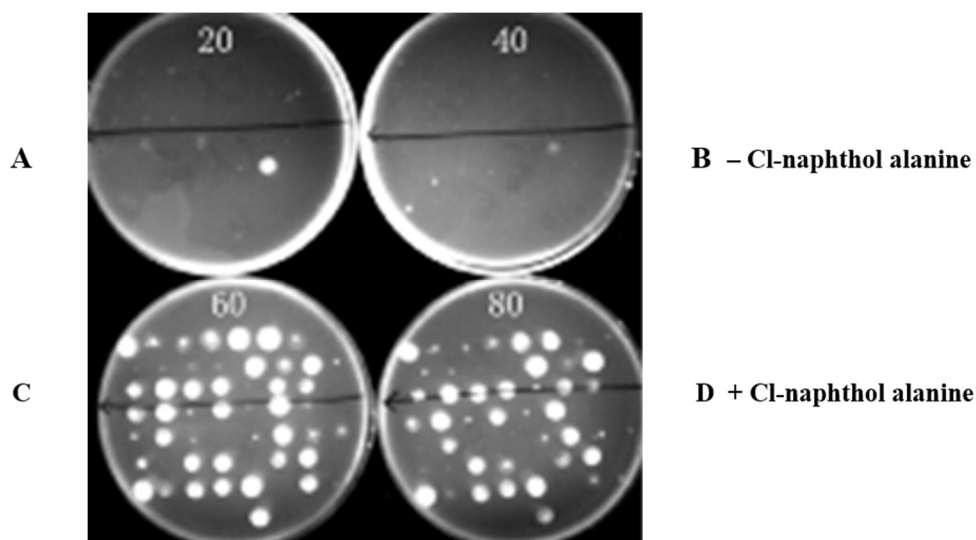
**Figure 5-7 Verification of the *Mj*ClNARS by culturing clones on plates that were supplement with different concentrations of chloramphenicol.**

**(a)** Plate **A**: 40 mg/ml Chloramphenicol, no UAA. Plate **B**: 60 mg/ml Chloramphenicol, no UAA. Plate **C**: 60 mg/ml Chloramphenicol, 1 mM Cl-naphthol alanine.   Plate **D**: 80 mg/ml Chloramphenicol, 1 mM Cl-naphthol alanine.

**(b)** Plate **A**: 20 mg/ml Chloramphenicol, no UAA. Plate **B**: 40 mg/ml Chloramphenicol, no UAA. Plate **C**: 60 mg/ml Chloramphenicol, 1 mM Cl-naphthol alanine.   Plate **D**: 80 mg/ml Chloramphenicol, 1 mM Cl-naphthol alanine.

| | Y32 | I63 | L65 | A67 | H70 | K91 | K101 | F108 | Q109 | Y114 | V149 | Y151 | D158 | I159 | L162 | V164 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cl10 | E | | P | | S | | | | | | | | G | | N | |
| Cl11 | E | | S | | | | | | | | | | G | | I | G |
| Cl2 | E | | T | S | | | R | | | | | | G | | V | G |
| Cl3 | E | | A | | N | | | | A | | | | G | S | F | |
| Cl12 | E | | P | S | N | | | | | | | | G | S | N | S |
| Cl19 | E | | A | | R | | | S | P | | | C | L | | G | |
| Cl21 | E | | T | S | N | | | G | H | G | | | G | | S | G |
| Cl22 | E | | N | | | | | V | G | G | | | G | | T | G |
| Cl23 | E | | A | | S | | | | | | | | G | S | H | |
| Cl27 | E | | | | N | | | | | | | | G | | N | |
| Cl29 | E | | T | | | | | | | | I | | G | | T | |
| Cl31 | E | | I | | | | | | | | | | G | | E | |
| Cl35 | D | | Y | | G | | | I | W | | | | G | | A | |
| Cl42 | E | | P | S | N | | | | | | | | G | G | H | |
| Cl45 | E | | I | | | | | | | | | | G | | E | |
| Cl46 | E | | | | S | | | | | | | | G | | N | |
| Cl50 | E | G | H | G | G | | | T | I | | | | G | Y | G | Y |
| Cl52 | E | | C | | | | | | A | | | | G | | I | G |
| Cl53 | Q | | G | | S | | | | H | | | | G | | D | G |
| Cl54 | E | | S | | R | | | V | H | | | | G | | C | S |
| Cl56 | E | | T | G | | | | | | | | | G | | C | S |

**Table 5-1 The mutation sites of of *Mj*ClNARS.**

## The expression of the myoglobin mutant

To express the mutant myoglobins, plasmid pBADJYAMB-Ser4TAGHis was co-transformed with pBK-ClNARS into TOP10 E. coli competent cells. Cells were amplified in LB media (5 mL) supplemented with kanamycin (50 μg/mL) and tetracycline (15 μg/mL). A starter culture (1 mL) was used to inoculate 100 mL of liquid LB supplemented with appropriate antibiotics and Cl-naphthol alanine (1 mM). Cells were then grown at 37°C to OD600 of 0.5, and protein expression was induced by the addition of 0.2% arabinose. After 12 h of growth at 37°C, cells were harvested by centrifugation. Myoglobin mutant was then purified by Ni-NTA affinity chromatography (Qiagen) and size exclusion chromatography to obtain the desired protein with the UAA.

## References

1.    Stewart, J. M., Solid Phase Peptide Synthesis. *Journal of Macromolecular Science Chemistry* **2006,** *10* (1), 259-288.

2.    Dawson, P. E. M., T.W. & Clark-Lewis, I & Kent, S.B.H. , Synthesis of proteins by native chemical ligation. *Science* **1994,** *266*, 776-779.

3.    Camarero, J. M., Alexander., Synthesis of Proteins by Native Chemical Ligation Using Fmoc-Based Chemistry. *Protein & Peptide Letters* **2005,** *12* (8), 723-728.

4.    Kochendoerfer, G. G.;   Chen, S. Y.;   Mao, F.;   Cressman, S.;   Traviglia, S.;   Shao, H.;   Hunter, C. L.;   Low, D. W.;   Cagle, E. N.;   Carnevali, M.;   Gueriguian, V.;   Keogh, P. J.;   Porter, H.;   Stratton, S. M.;   Wiedeke, M. C.;   Wilken, J.;   Tang, J.;   Levy, J. J.;   Miranda, L. P.;   Crnogorac, M. M.;   Kalbag, S.;   Botti, P.;   Schindler-Horvat, J.;   Savatski, L.;   Adamson, J. W.;   Kung, A.;   Kent, S. B.; Bradburne, J. A., Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* **2003,** *299* (5608), 884-7.

5.    Hecht, S. M., Alford, B. L., Kuroda, Y., & Kitano, S. , "Chemical aminoacylation" of tRNA's. *Journal of Biological Chemistry* **1978,** *253* (13), 4517-4520.

6.    Robertson, S. A.;   Noren, C. J.;   Anthony-Cahill, S. J.;   Griffith, M. C.; Schultz, P. G., The use of 5'-phospho-2 deoxyribocytidylylriboadenosine as a facile route to chemical aminoacylation of tRNA. *Nucleic Acids Res* **1989,** *17* (23), 9649-60.

7.    Eisenhauer, B. M.; Hecht, S. M., Site-specific incorporation of (aminooxy)acetic acid into proteins. *Biochemistry* **2002,** *41* (38), 11472-8.

8.    Murakami, H.;   Hohsaka, T.;   Ashizuka, Y.;   Hashimoto, K.; Sisido, M., Site-directed incorporation of fluorescent nonnatural amino acids into streptavidin for highly sensitive detection of biotin. *Biomacromolecules* **2000,** *1* (1), 118-25.

9.    Wang, L.;   Brock, A.;   Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.

10.   Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annual review of biochemistry* **2010,** *79* (1), 413-44.

11.   Xie, J.; Schultz, P. G., A chemical toolkit for proteins — an expanded genetic code. *Nat Rev Mol Cell Biol* **2006,** *7* (10), 775-782.

12.   Wang, L.; Schultz, P. G., Expanding the genetic code. *Chem Commun (Camb)* **2002,**   (1), 1-11.

13.   Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chem Biol* **2001,** *8* (9), 883-90.

14.   Wang, L.;   Magliery, T. J.;   Liu, D. R.; Schultz, P. G., A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *Journal of the American Chemical Society* **2000,** *122* (20), 5010-5011.

15.   Fechter, p.;   Rudinger-Thirion, J.;   Tukalo, M.; Giege, R., Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNA(Tyr) are conserved but expressed differently. *European journal of biochemistry* **2001,** *268* (3), 761-767.

16.   Steer, B. A.; Schimmel, P., Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *J Biol Chem* **1999,** *274* (50), 35601-6.

17. Ascano, M.; Hafner, M.; Cekan, P.; Gerstberger, S.; Tuschl, T., Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* **2012,** *3* (2), 159-77.

# Summary

# Summary

RNA as one of the most important bio-macromolecules of life has attracted biologists to develop a variety of tools to study how it regulates cellular function. Generally, these tools are divided into two groups: RNA imaging and RNA crosslinking-seq. RNA imaging denotes visualization of target RNA by labelling it with fluorophores, like FISH, MS2 system, or RNA aptamer systems. RNA crosslinking methods refers to investigating the function of target RNA, which is captured by crosslinking with the counterparts it interacts with. Usually, these methods are followed by deep sequencing of the corresponding cDNA library like HITS-CLIP, iCLIP, or PAR-CLIP. This thesis (Chapter **1**) firstly reviewed the widely applied aforementioned methods and analyzed the advantages and disadvantages by comparing them with each other.

In chapter **2**，inspired by a reported RNA imaging tool that was composed of an RNA aptamer (Broccoli) and a GFP mimic chromophore (DFHBI)，we designed and synthesized an chromophore (D-P) consisting of DFHBI and a ligand (paromomycin) that had very high binding affinity with its RNA aptamer (Apt-P). Meanwhile, we also prepared the corresponding RNA aptamer (Apt-DP) consisting of Broccoli and Apt-P. We assumed that the new chromophore D-P should have higher binding affinity with Apt-DP than DFHBI/Broccoli, which should be an ideal method to reduce background signal in RNA imaging in living cells. The in vitro experimental data shows that the binding affinity of DFHBI/Broccoli is dramatically improved by introducing paromomycin to the chromophore. In cell RNA imaging results also proves that Apt-DP has a much higher specificity to target RNA. We demonstrated in this chapter that improvement of chromophore/aptamer binding affinity could be realized by introducing a ligand/aptamer pair that has higher bind affinity.

In chapter **3** we developed two psoralen based length tunable crosslinkers for studying RNA-RNA interactions or RNA-protein interactions. The crosslinkers were designed to improve traditional crosslinkers' intrinsic shortage that specific nucleotides or amino acids were required at the interaction sites. We designed an RNA-RNA crosslinker that was composed of two psoralen analogues conjugated with a length tunable bridge (AMT-dimer). The strategy was extended to design a RNA-protein crosslinker that was composed of a psoralen analogue and a NHS group connected by a length tunable bridge (AMT-NHS). We assumed that the introduction of the bridge would enable the crosslinker to locate the crosslinking sites at a wider range but not only at the interaction

sites. We synthesized the crosslinkers and verified in vitro experiment that RNA-RNA interactions can be effectively harnessed by the AMT dimer, and RNA-protein interactions could be effectively tracked by AMT-NHS. Both conjugates pave the way for application of these crosslinkers in living cells.

After proving AMT-NHS is able to crosslink protein and RNA in vitro，in chapter **4**, we decided to test if this crosslinker could be applied to study RNA-protein interactions in living cells. In order to compare with the traditional cosslinking methods, a well-studied alternative splicing suppressor PTB was chosen as the target protein. PTB was expressed in living cells and cosslinked to protein binding RNAs by using AMT-NHS or traditional crosslinking methods. The crosslinking efficiency of those methods was compared through immunoblot analysis of immuno-precipitated crosslinking products. AMT-NHS proved to be an effective protein-RNA crosslinker. The crosslinked RNAs were further analyzed by comparing with reported sequence data that was obtained by traditional crosslinking method. We finally proved that AMT-NHS CLIP is a stable and efficient method for studying RNA-protein interactions, and most importantly it captures a large portion of interactions that traditional CLIP method may lose.

In the CLIP method, a large portion of protein binding RNAs are missed due to the lack of photoactivatable amino acids around the RNA and protein interaction sites. In order to overcome this limitation, in chapter **5**, we intended to incorporate a photoactivatable amino acid into a protein on a specific site. We designed and synthesized the amino acid and constructed a selection system for generation of a specific aaRS to aminoacylate a blank codon suppressor with the amino acid. After three rounds of positive selection and two rounds of negative selection, clones containing an ideal tRNA (blank codon) / aaRS pair was obtained and the unnatural amino acid was successfully incorporated into a protein at the chosen site. Site specific incorporation of a photoactivatable amino acid into an RNA binding protein paves the way to develop a novel and powerful CLIP method to capture more comprehensively RNA-protein interactions.

# Samenvatting

RNA als een van de belangrijkste componenten van het organisme trok biologen aan om een verscheidenheid aan hulpmiddelen te ontwikkelen om te bestuderen hoe het cellulaire activiteit reguleert. Over het algemeen zijn deze tools verdeeld in twee groepen: RNA-beeldvorming en RNA-crosslinking-seq. RNA-beeldvorming duidt op het visualiseren van doel-RNA door etikettering met fluorofoor, zoals FISH, MS2-systeem, RNA-aptameersysteem enzovoort. RNA-crosslinking-methode verwijst naar het onderzoeken van de functie van doel-RNA dat wordt vastgelegd door crosslinking met de tegenhangers waarmee het interageert, deze methoden worden normaal gesproken gevolgd door diepe sequencing van de overeenkomstige cDNA-bibliotheek zoals HITS-CLIP, iCLIP, PAR-CLIP enzovoort. Dit proefschrift (Hoofdstuk **1**) heeft eerst de veelgebruikte bovengenoemde methoden besproken en de voor- en nadelen geanalyseerd door ze met elkaar te vergelijken.

In hoofdstuk **2**，geïnspireerd door een gerapporteerde RNA-beeldvormingstool die was samengesteld uit een RNA-aptameer (Broccoli) en een GFP-nabootsende chromofoor (DFHBI), hebben we een chromofoor (DP) ontworpen en gesynthetiseerd bestaande uit DFHBI en een ligand (paromomycine) die zeer hoge bindingsaffiniteit met zijn RNA-aptameer (Apt-P). Ondertussen hebben we ook het overeenkomstige RNA-aptameer (Apt-DP) bereid, bestaande uit Broccoli en Apt-P. We gingen ervan uit dat de nieuwe chromofoor D-P een hogere bindingsaffiniteit met Apt-DP zou moeten hebben dan DFHBI/Broccoli, wat een ideale methode zou moeten zijn om het achtergrondsignaal in RNA-beeldvorming in levende cellen te verminderen. De in vitro experimentele gegevens laten zien dat de bindingsaffiniteit van DFHBI/Broccoli dramatisch wordt verbeterd door paromomycine in de chromofoor te introduceren. In cel-RNA-beeldvormingsresultaten bewijzen ook dat Apt-DP een veel hogere specificiteit heeft om RNA te targeten. We hebben in dit hoofdstuk aangetoond dat verbetering van de chromofoor/aptameer bindingsaffiniteit kan worden gerealiseerd door het introduceren van een ligand/aptameer paar dat een hogere bindingsaffiniteit heeft.

In hoofdstuk **3** hebben we twee op psoraleen gebaseerde lengte-afstembare crosslinkers ontwikkeld voor het bestuderen van RNA/RNA-interactie of RNA/eiwit-interactie. De crosslinkers zijn ontworpen om het intrinsieke tekort van traditionele crosslinkers te verbeteren, aangezien specifieke nucleotiden of aminozuren nodig waren op de interactieplaatsen. We ontwierpen een RNA- en RNA-crosslinker die was samengesteld

uit twee psoraleen-analogen geconjugeerd met een in lengte instelbare brug (AMT-dimeer). De strategie werd ook gebruikt om een RNA- en eiwit-crosslinker te ontwerpen die was samengesteld uit een psoraleen-analoog en een NHS-groep die was geligeerd met een afstembare brug op lengte (AMT-NHS). We gingen ervan uit dat de introductie van de brug de crosslinker in staat zou stellen om de crosslinking-sites op een groter bereik te lokaliseren, maar niet alleen op de interactiesites. We hebben de crosslinkers gesynthetiseerd en in vitro experiment geverifieerd dat RNA en RNA effectief kunnen worden gecrosslinkt door AMT-dimeer, en RNA en eiwit kunnen effectief worden gecrosslinkt door AMT-NHS, wat de weg vrijmaakt voor toepassing van de crosslinkers in levende cellen.

Nadat we in hoofdstuk **4** hadden bewezen dat AMT-NHS in staat is om eiwit en RNA in vitro te verknopen, hebben we besloten om te testen of deze verknoper kan worden toegepast om RNA en eiwitinteractie in levende cellen te bestuderen. Om te vergelijken met de traditionele cosslinking-methoden, werd een goed onderzochte alternatieve splitsingsonderdrukker PTB gekozen als het doeleiwit. PTB werd tot expressie gebracht in levende cellen en gekoppeld aan eiwitbindende RNA's met behulp van AMT-NHS of traditionele verknopingsmethoden. De verknopingsefficiëntie van die methoden werd vergeleken door middel van Western-blot-analyse van immunogeprecipiteerde verknopingsproducten. AMT-NHS bleek een effectieve eiwit- en RNA-crosslinker te zijn. De verknoopte RNA's werden verder geanalyseerd door ze te vergelijken met gerapporteerde sequentiegegevens die waren verkregen met de traditionele verknopingsmethode. We hebben eindelijk bewezen dat AMT-NHS CLIP een stabiele en efficiënte methode is voor het bestuderen van RNA- en eiwitinteractie, en het belangrijkste is dat het een groot deel van de interacties vastlegt die de traditionele CLIP-methode kan verliezen.

In de CLIP-methode wordt een groot deel van de eiwitbindende RNA's gemist vanwege het ontbreken van fotoactiveerbaar aminozuur rond de RNA- en eiwitinteractieplaatsen. Om de beperking te doorbreken, wilden we in hoofdstuk **5** een fotoactiveerbaar aminozuur inbouwen in een eiwit op een specifieke plaats. We hebben het aminozuur ontworpen en gesynthetiseerd en een selectiesysteem geconstrueerd voor het genereren van een specifiek aaRS om een blanco codon suppressor te aminoacyleren met het aminozuur. Na 3 ronden van positieve selectie en 2 ronden van negatieve selectie,

werden klonen met een ideaal tRNA (blanco codon) / aaRS-paar verkregen en werd het onnatuurlijke aminozuur met succes opgenomen in een eiwit op de gekozen plaats. Plaatsspecifieke opname van een fotoactiveerbaar aminozuur in een RNA-bindend eiwit maakt de weg vrij om een nieuwe en krachtige CLIP-methode te ontwikkelen om meer uitgebreide RNA- en eiwitinteractie vast te leggen.

# Acknowledgement

In the end of my thesis, it is the right moment to look back and acknowledge all the people who contributed to this thesis and supported me during my Ph.D. studies.

First and foremost, I would like to give my deepest gratitude to my supervisor Prof. Andreas Herrmann, thank you for offering me the opportunity to study in PCBE. I still remember the first time we discussed the project I was going to start, I was amazed by your interdisciplinary experience in chemistry and molecular biology, as an organic chemist, I used to satisfy with my professional capacity, but you made me know this was absolutely not enough. I really appreciated your guidance, inspiring ideas and discussions we had. I learned a lot of polymer chemistry and biology knowledge from you, and you were always so patient when you explaining so many obscure mechanisms to me.

I would like to thank Prof. Jiangyun Wang, thank you for offering the opportunity to finish my projects in IBP. I still remember the first time I went to your lab, you gave me a bunch of molecular biology books and ask me to transform from an organic chemist to a bio-chemist ASAP. Thank you for giving me so many valuable suggestions and pushing me to finish the transformation. I would also thank you for providing us the relaxing and comfortable working and resting environment where I can lift dumbbell and skip rope in the office.

I would also like to give my gratitude to all the members of reading committee: Prof. Romana Schirhagl, Prof. Gerard Roelfes and Prof. Andreas Marx. Thank you so much for reading and approving the manuscript, also thank you for your valuable comments which helped me to polish my manuscript.

I would give my sincere gratitude to Prof. A. J. Boersma. Arnold, thank you so much for revising the DFHBI manuscript and data analysis, I appreciated very much your valuable suggestions and I was really impressed on your rigorous scientific attitude.

Special thanks to Ursula, Thank you so much for your great help for preparation of my thesis defense. You were so patient and positive when I met difficulties when organizing reading committee and arranging defense time. The process must be very tough without your help, I wish you all the best in the future.

Many thanks to my colleagues in PCBE who helped me a lot both in my work and life. Lifei, I always admire the attitude you treat friends, family and research. It was lucky for me to be in the same office with you, I really enjoyed daily discussions about work and life with you, your flexible mindset inspired me a lot. Kai, you are the most passionate workmate I have ever met, I always tell myself a great scientist should be like this. I still clearly remember the first day I arrived in Groningen, you and Lifei helped me to carried my heavy luggage to my apartment with your bicycle, it will never be possible for me to take a 20 kg luggage case in one hand when riding bicycle. Thank you so much for your help in my work and life. Zhuojun, and Qing, Thanks a lot for the delicious dinner your prepared for celebrating by birthday when my wife was not there, thank you for your kindness. Jingyi, when I was planning to travel in EU, you were always the first one to recommend me many interesting places, thank you for your detailed advices and tips. Jing, I enjoyed the time when we finished lunch and hanging around Zernike institute，wish you and Jiaying have a bright future. Hongyan, thank you for taking care of my cat when I was on traveling, that was a big help for me. Chao, do you still remember the drunk night during the conference trip that you slept in bathtub? I enjoyed the experience very much. Gurudas and Avishake，I was so happy to talk with you guys after lunch or during coffee time, thank you for sharing so many funny stories and Indian cultures. Konstantin, you are always the first guy I would come to ask for help when I encountered synthesis problems, thank you for your patience and helpful suggestions, I also really enjoyed the chat and smoking time with you almost every afternoon. Agnieszka，Thank you for teaching me the protocol of selective modification of antibiotics through APGs. Jan Willem, you are always willing to help everyone, thank you for teaching me how to use instruments of our lab. I would also like to thank Wei, Lei, Jun, Pei, Eliza, Mark, Feng, Alina and Jennifer, thank you for being my colleague, it was a memorable experience to work with you.

Very special thanks to all my friends in Groningen, Jiquan, you are my first roommate and first friend I made in the Netherland, thank you for showing me all around Groningen when I first came to this new city. I also thank you for introducing me to so many friends who made me no longer lonely in this foreign land. Thank you so much for teaching me cook so many Chinese food, I wish you all the best in the future. Wenjun and Tao, thank you for taking care of me like family, I really find the feeling

of home with your company. You know that I am a northwestern Chinese, so you always invited me when you cooked noodles and dumplings, thank you for the hometown flavor. I wish you all find your own happiness and Nuannuan has a happy childhood. I would like to thank Guowei, Zhiyuan, Yuantao, Xujing, Qiuyan, Boqun, Qingkai, Yangshuo, Zhangbo and all my friends in Groningen, I really enjoyed every party and every football game together with you, you made my life colorful in the Netherlands.

Xuzhen, It was so lucky for me to accomplish the RNA imaging and RNA/protein crosslinking projects with you. Thank you for teaching me so many biological techniques and knowledges, I appreciated the time we were doing experiment, playing basketball and drinking beers together, it was a great memory in my life. Yangfan, Thank you for giving me so many suggestions and teaching me Western blot techniques. Yazhou, thank you for you great help both in and outside lab. Liuqi, thank you for being always so patient to explain many biological mechanisms to me. I would also like to thank many friends in IBP, Du Kang, Jiajun, Wangli, Zhangfeng, Fuying, Shuqin, Liuchuan and all colleagues in Prof. Wang's lab, wish you all the best.

最后，我想感谢我的家人，感谢父母把我带到这个世界，把我培养成人。感谢岳父岳母还有两个姐姐这些年的帮助和支持。我想感谢我的妻子胡洋，谢谢你这些年的照顾和陪伴，你为我们这个小家付出了太多，余生一定会是稳稳的幸福！最后的最后，我有很多话要对我的女儿说，可是又不知从哪说起，爸爸希望你能勇敢、善良，爸爸希望你有一份平和的心态，爸爸妈妈永远爱你！