Chemo-Enzymatic Strategies to Functionalize RNA with Biophysical Probes Using Template Dependent and Independent Polymerases

A thesis

submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

by

Jerrin Thomas George

20122026

**INDIAN INSTITUTE OF SCIENCE EDUCATION
AND RESEARCH PUNE**

2019

Luke: Master, moving stones around is one thing, but this is... totally different!
Yoda: No! No different! Only different in your mind. You must unlearn what you have learned.
Luke: All right, I'll give it a try.
Yoda: No! Try not. Do... or do not. There is no try.
[Luke tries to use the Force to levitate his X-Wing out of the bog, but fails in his attempt.]
Luke: I can't. It's too big.
Yoda: Size matters not. Look at me. Judge me by my size, do you? Hmm? Hmm. And well you should not. For my ally is the Force, and a powerful ally it is. Life creates it, makes it grow. Its energy surrounds us and binds us. Luminous beings are we, not this crude matter. You must feel the Force around you; here, between you, me, the tree, the rock, everywhere, yes. Even between the land and the ship.

-Star Wars: Episode V-The Empire Strikes Back, 1980

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pashan, Pune–411 008

Dr. Seergazhi G. Srivatsan
Associate Professor, Chemistry
Wellcome Trust DBT-India Alliance Senior Fellow

CERTIFICATE

Certified that the work incorporated in the thesis entitled *"Chemo-Enzymatic Strategies to Functionalize RNA with Biophysical Probes Using Template Dependent and Independent Polymerases"* submitted by Mr. Jerrin Thomas George was carried out by the candidate under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.

Date: August 13, 2019   Dr. Seergazhi G. Srivatsan

Place: Pune

DECLARATION

I declare that this written submission represents my ideas in my own words and where others ideas have included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: August 13, 2019

Place: Pune

Jerrin Thomas George

Reg. No: 20122026

Acknowledgements

Contents

Chapter 3: Posttranscriptional Chemo-Enzymatic Tailoring of RNA Employing a Terminal Uridylyl Transferase, SpCID1

Chapter 4: Utility of SpCID1 for High-Density Labeling of CRISPR Guide RNA, Developing FRET Probes and Site-Specific Internal Labeling

Chapter 5: A Responsive Fluorescent Nucleotide Enables the Probing of SpCID1-Mediated Terminal Uridylation

Abbreviations

ε	molar extinction coefficient	CRISPR	clustered regularly interspaced short palindromic repeats
μL	microliter		
μg	microgram		
μM	micro molar	CTP	cytidine monophosphate
2-EA	2-ethynyladenosine	Cu	copper
4-sU	4-thiouridine	CuAAC	copper(I)-catalyzed alkyne-azide cycloaddition
7-EAA	7-ethynyl-8-aza-7-deazaadenosine		
		CuSO4	copper sulphate
A	adenine	Cy3	cyanine 3
AAC	azide-alkyne cycloaddition	Cy5	cyanine 5
ACN	acetonitrile	DAPI	4', 6-diamidino-2-phenyl-indole
AdoMet	s-adenosyl-L-methionine		
agm^{2}C	2-agmatinylcytidine	DBCO	dibenzoazacyclooctyne
AMP	adenosine monophosphate	dCas9	nuclease deactivated or dead CRISPR associated protein 9
AMU	5-azidomethyluridine		
AMUTP	5-azidomethyl uridine triphosphate	DIBO	dibenzocyclooctyne
		DNA	deoxyribonucleic acid
APUTP	5-azidopropyl uridine triphosphate	DMAP	4-dimethylaminopyridine
		DMEM	Dulbecco's modified eagle medium
ATUTP	5-azidoTeg-uridine triphosphate		
		DMF	dimethylformamide
ATP	adenosine triphosphate	DMSO	N, N-dimethyl sulfoxide
BrU	5-bromouridine	DMEM	Dulbecco's Modified Eagle Medium
BSA	bovine serum albumin		
C	cytosine	DMSO	N, N-dimethyl sulfoxide
Cas9	CRISPR associated protein 9	DMF	Dimethylformamide
CIAP	calf intestinal alkaline phosphatase	DOTAP	*N*-[1-(2,3-Dioleoyloxy) propyl]-*N,N,N*-trimethyl-ammonium methylsulfate

ds	double stranded	IEDDA	inverse-electron demand Diels-Alder
Ds	7-(2-thienyl)imidazo[4,5-b]pyridine	HSV-1	herpes simplex virus
DTT	dithiothreitol	IMAC	immobilized metal affinity chromatography
EDC	1-ethyl-3-(3-dimethylamino-propyl)carbodiimide		
		in vitro	outside living organism
EDTA	ethylenediaminetetraacetic acid	*in vivo*	inside living organism
		K_d	apparent dissociation constant
EdU	ethynyl deoxyuridine		
em	emission	LB	lysogeny broth
EMSA	electrophoretic mobility shift assay	LPC	lipopolysaccharide
		LIF	leukamia inhibiting factor
ESI-MS	electrospray ionization mass spectrometry	MALDI-TOF	matrix assisted laser desorption ionisation-time of flight
EU	ethynyluridine		
FAM	Ffuorescein amidite	*max*	maximum
FBS	fetal bovine serum	MeOH	methanol
FISH	fluorescence in situ hybridization	$(MeO)_3PO$	trimethylphosphate
		mESC	mouse embryonic stem cell
FRET	Förster resonance energy transfer	mg	milligram
		MHz	megahertz
G	guanosine	miRNA	microRNA
GFP	green fluorescent protein	mM	millimolar
GlaTgs2	Giardia lamblia trimethylguanosine synthase	mRNA	messenger RNA
		MST	microscale thermophoresis
GTP	guanosine triphosphate	ng	nanogram
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	nm	nanometer
		nmol	nanomolar
		NMR	nuclear magnetic resonance
HPLC	high performance liquid chromatography	NTPs	nucleoside triphosphates
		N^6PA	N^6-propargyladenosine

OD$_{600}$	optical density at 600 nm	SpCID1	*Schizosaccharomyces pombe*
ON	oligonucleotide		caffeine induced death
Pa	pyrrole-2- carbaldehyde		suppressor 1
PAGE	polyacrylamide gel	ss	single stranded
	electrophoresis	TEAA	triethylammonium acetate
PAM	protospacer adjacent motif	TEAB	Triethylammonium
PAP	poly(A) polymerase		bicarbonate
PAR-CLIP	photoactivatable	TEG	triethylene glycol
	ribonucleoside-enhanced	THF	tetrahydrofuron
	crosslinking and immuno-	THPTA	tris-(3-hydroxypropyl
	precipitation		triazolylmethyl)amine
PBS	phosphate buffer saline	TLC	thin layer chromatography
PCR	polymerase chain reaction	tRNA	transfer RNA
Pd	palladium	Tris	Tris(hydroxymethyl) amino
pmol	picomolar		methane
POCl$_3$	phosphorus oxychloride	THPTA	tris-(3-hydroxypropyl
ppm	parts per million		triazolylmethyl)amine
PUP	poly(U) polymerase	TiaS	tRNA^{Ile2}-agm^{2}C synthetase
qPCR	real-time or quantitative	Trm1	tRNA methyltransferase
	polymerase chain reaction	TUTase	terminal uridylyl transferase
R$_f$	retention factor	U	uridine
RNA	ribonucleic acid	UMP	uridine monophosphate
RNP	ribonucleoprotein	UTP	uridine triphosphate
SDS	sodium dodecyl sulfate	UV	ultraviolet
SELEX	systematic evolution of	YES	yeast extract with
	ligands by exponential		supplements
	enrichment		
sgRNA	single guide RNA		
siRNA	small interfering RNA		
SPAAC	strain-promoted alkyne-azide		
	cycloaddition		

Synopsis

Chemo-Enzymatic Strategies to Functionalize RNA with Biophysical Probes Using Template Dependent and Independent Polymerases

Background and Aim: Ribonucleic acid (RNA) plays a multi-faceted role in cellular physiology wherein it can perform a wide variety of functions from catalysis to storage of genetic data. This is accomplished by forming complex secondary and tertiary structures which enable it to perform its associated function.[1-2] Unraveling these structures is crucial for gaining fundamental insight into various disease states and developing diagnostic tools. Various biophysical techniques like fluorescence, NMR, EPR, and X-ray crystallography are used to investigate the structure of RNA.[3-6] However, RNA has no intrinsic biophysical label which makes it imperative to label RNA with extrinsic functional tags.[7] Methods to label RNA can broadly be classified as covalent and non-covalent.[8] While non-covalent methods like antibodies, aptamers, and RNA-binding proteins are employed for RNA labeling,[9-11] a major drawback of these approaches are its limited cell permeability and selectivity. Also, some of these methods demand a specific RNA sequence/structure for binding and therefore, not a general approach.[9, 12-13]

Covalent approach is a more general technique to label RNA with functional tags and biophysical probes. In this strategy, RNA can either be synthesized chemically by solid-phase synthesis or enzymatically by RNA polymerases. One major drawback of solid-phase synthesis is the restriction in length of RNA synthesized due to decrease in coupling efficiency with each iteration of nucleotide addition. Alternatively, the promiscuity of certain RNA polymerases to accept functional substrates has been employed in labeling RNA with a variety of biophysical probes.[14] However, both these methods fail for bulky and sensitive modifications. This can be circumvented by using bioorthogonal click chemistry wherein the RNA is synthesized with a reactive handle and further functionalized postsynthetically with functional tags of interest.[11-15] The thesis focuses on the development of new chemo-enzymatic strategies for labeling RNA with biophysical probes using a template-dependent approach such as transcription and template-independent approach employing a terminal transferase enzyme.

The thesis is divided into Chapters 1-5.

Chapter 1: Labeling Technologies for Investigating RNA Structure and Function

With the need for introducing biophysical probes into RNA for deciphering the structure and function of RNA, various technologies have been developed to functionalize RNA.[8] In Chapter 1, we provide a brief overview of various covalent and non-covalent labeling technologies, followed by an in-depth discussion on different bioorthogonal chemistries for covalent functionalization of RNA. Further, we discuss various techniques to label RNA employing transcription, transferases and ribozymes, and their utility for labeling RNA *in vitro* and in cells.

Chapter 2: Vinyluridine as a Versatile Chemoselective Handle for the Posttranscriptional Chemical Functionalization of RNA

Vinyl tags have been used for post-synthetic labeling of protein and DNA.[40,49] However, the utility of such a tag as a versatile handle in developing RNA-labeling protocols has not been well-explored. In Chapter 2, we have synthesized 5-vinyl-modified uridine triphosphate and used it as modified nucleotide analog for postsynthetic functionalization of RNA using palladium-mediated oxidative Heck reaction and inverse electron-demand Diels-Alder reaction (IEDDA)(Figure i).[15] 5-vinyluridine triphosphate (VUTP) was incorporated into RNA via transcription reaction (with various template DNAs) which resulted in excellent incorporation efficiency of 5-vinyluridine when compared with natural UTP. Further, oxidative Heck reaction was performed on the vinyl-labeled RNA with various heterocycle conjugated boronic acids derivatives, which led to the formation of coupled products. Upon comparing the fluorescence of the coupled product with reactants, vinyl-labeled RNA and boronic acid derivative, we observed up to 170-fold enhancement in fluorescence.

One of the disadvantages of Pd-based coupling is the use of multiple reagents, including catalyst for conjugation. In order to explore the versatility of the vinyl-tagged RNA to undergo a reagentless coupling reaction, we sought to perform IEDDA reaction. For this, we synthesized tetrazine analogs having probes for fluorescence and pull-down. We performed IEDDA reaction on RNA using these tetrazine derivatives, which on reaction with vinyl functionality yielded coupled products. Also, we labeled a longer RNA having multiple vinyluridine which when reacted with Cy5-conjugated tetrazine, yielded fluorescent RNA having multiple Cy5 tags.

Altogether in Chapter 2, we report for the first time, oxidative Heck reaction on RNA by employing 5-vinyluridine. We demonstrate the utility of oxidative-Heck reaction for fluorogenic

coupling, which is highly useful for background-free staining of RNA in fluorescence cell imaging experiments. Subsequently, the versatility of this tag for introducing biophysical probes via IEDDA reaction, a reagentless bioorthogonal coupling reaction, was established.
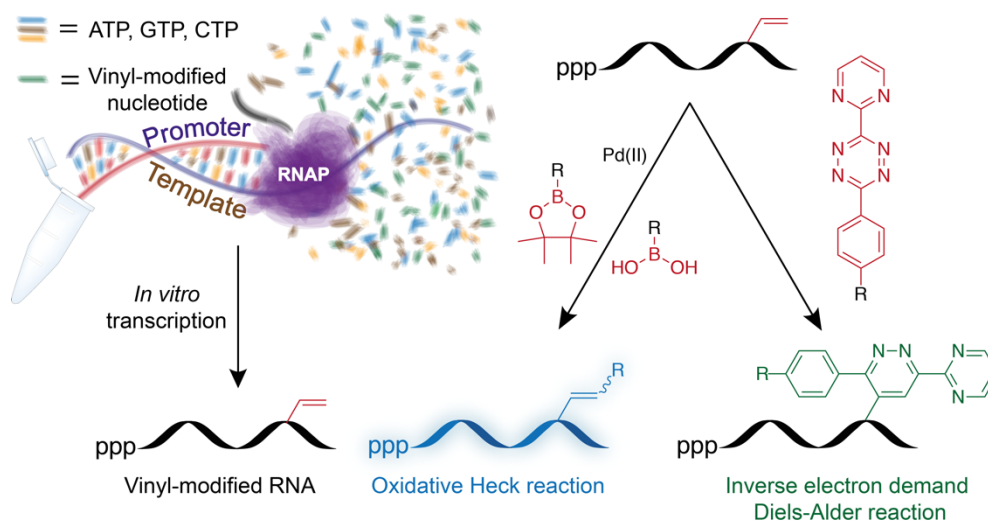
Figure i. Incorporation of vinyl-modified triphosphate into RNA transcripts by *in vitro* transcription reaction, followed by oxidative Heck reaction with boronic acid derivatives and IEDDA reaction with tetrazine derivatives, enabled the synthesis of RNA conjugated to various reporters and tags.

Chapter 3: Posttranscriptional Chemo-Enzymatic Tailoring of RNA Employing a Terminal Uridylyl Transferase, SpCID1

Although transcription reaction is one among the prominently used covalent chemo-enzymatic RNA labeling approaches, one major limitation is its inability to perform site-specific labeling of RNA. In Chapter 3, we develop a transferase-assisted chemo-enzymatic labeling technique using the promiscuity of terminal uridylyl transferase, SpCID1 for site-specific labeling of RNA at its 3' end. Based on crystal structure,[17] we hypothesized that modifications at 5-position of uridine might be tolerated by SpCID1 for terminal uridylation. The hypothesis was explored by cloning SpCID1 gene from fission yeast, *Schizosaccharomyces pombe,* and protein was expressed and purified. Using multiple azide, alkyne modified triphosphate and the vinyluridine triphosphate (Chapter 2), we compared its degree of incorporation with natural UTP wherein we found excellent incorporation of these modified triphosphate analogs into the 3' end of RNA (Figure ii).

Further, we tuned the stoichiometry of SpCID1 enzyme and modified triphosphate to control the addition of a single azide/alkyne UMP at the 3' end of RNA (Figure ii). We isolated the single azide-labeled RNAs with excellent yields and performed copper-catalyzed azide-alkyne

cycloaddition (CuAAC) reaction and strain-promoted azide-alkyne cycloaddition reaction (SPAAC) on the modified RNA to introduce a single fluorescent tag at the 3' end of RNA. The terminal uridylation approach presented in this chapter has tremendous utility for introducing single or multiple functional tags on biologically relevant RNAs, former for biophysical structure-analysis and later for attaching smart probes for understanding its cellular localization.
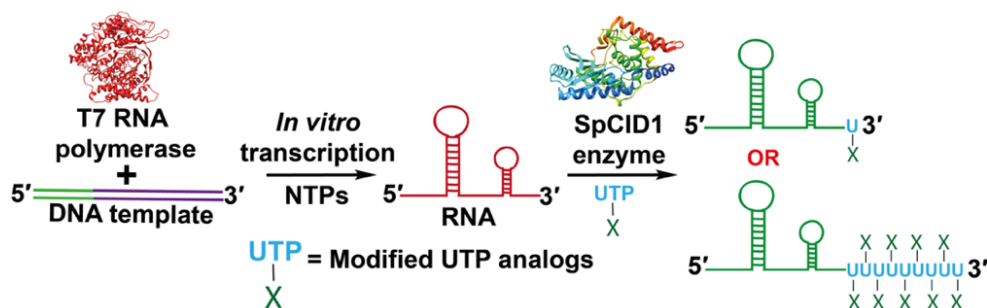
Figure ii. Single or multiple incorporations of modified UTP analogs at the 3' end of RNA using terminal uridylyl transferase, SpCID1. Structure of T7 RNA polymerase and SpCID1 adopted from PDB: 4RNP and 4FH3 using UCSF Chimera software.[16-17]

Chapter 4: Utility of SpCID1 for High-Density Labeling of CRISPR Guide RNA, Developing FRET Probes and Site-Specific Internal Labeling

A chemo-enzymatic strategy for site-specific labeling of RNA using multiple and single azide tags has been developed in Chapter 3. By, exploiting the ability of SpCID1 to introduce multiple site-specific azide tags, we label the functionally important, single guide RNA (sgRNA) of CRISPR-Cas9 system,[18-19] which allowed site-specific localization of azide tags to a gene locus for its further conjugation with functional probes. For this, we adopted a design strategy wherein multiple 5-azidomethyl uridine (AMU)-labeled guide RNA (azide-labeled sgRNA) synthesized by terminal uridylation could be functionalized using post-hybridization or pre-hybridization click chemistry depending on whether the bioorthogonal reaction was performed after or before ternary complex formation with dsDNA and dCas9 protein. By employing RNA secondary structure prediction, a single-stranded adenylate overhang was added at the 3' end of CRISPR sgRNA which allowed its efficient terminal uridylation with azide tags. The azide-labeled sgRNAs were subsequently click-functionalized using Cy3/biotin tags. Further, incubation of modified sgRNAs with dCas9 and target dsDNA, resulted in good binding of azide-labeled sgRNA (due to its minimum steric hindrance) and poor binding for click-functionalized sgRNA suggesting post-hybridization click chemistry as the method of choice. The azide-labeled sgRNA and dCas9 fused to eGFP were

localized to telomeric regions in mouse embryonic stem cells (mESCs) employing CRISPR-FISH.[20] This resulted in distinct nuclear puncta corresponding to azide-labeled telomere loci (Figure iii). Further, chromatin from mESCs was isolated, and the azide-labeled telomeric chromatin was reacted with biotin conjugated to alkyne/strained-alkyne employing post-hybridization click chemistry (Figure iii). Upon pull-down with streptavidin and qPCR quantification, we observed a ~12-fold and ~6-fold enrichment of telomere, respectively while using CuAAC and SPAAC reactions.
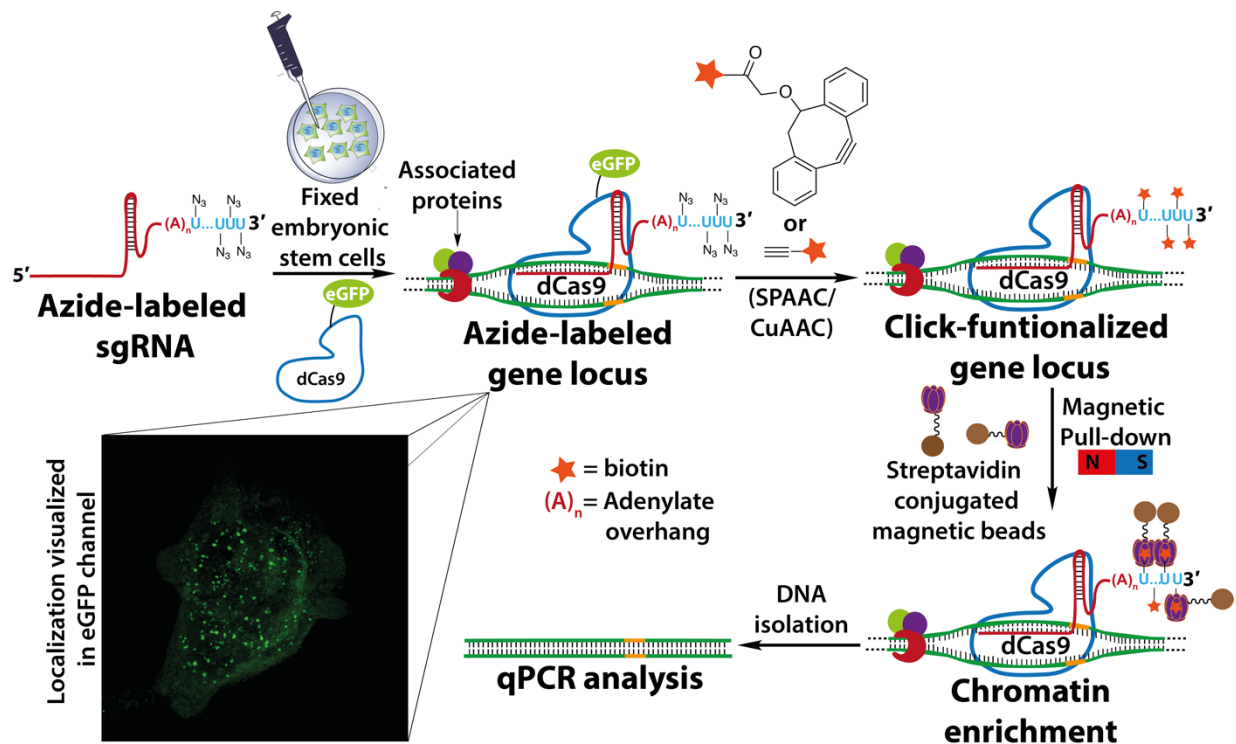
Figure iii. Azide-labeled sgRNA is localized in telomeric gene locus in fixed mouse embryonic stem cells followed by post-hybridization click chemistry on the chromatin using biotin tags. Further, the gene is enriched using streptavidin-conjugated magnetic beads.

Parallelly, utilizing the ability of SpCID1 to add a single azide UMP at the 3' end of RNA, we constructed FRET pairs on a potentially structured RNA. As a proof of principle, we further verified that FRET indeed occurs on the potentially structured RNA labeled using this chemo-enzymatic technique. We also demonstrated that the 3' azide-labeled RNA is compatible for splint-mediated ligation to obtain a longer RNA with a site-specific internal azide tag.

Altogether, in Chapter 4 we report the ability of SpCID1 enzyme to introduce multiple azide tags on sgRNA, which is exploited for the first time in bioorthogonal labeling of the

CRISPR-Cas9 system. Overall, this technology that we have devised can evolve as a universal approach for targeted labeling of gene loci using a minimally perturbing bioorthogonal handle as a 'trojan horse' on CRISPR-dCas9 system for site-directed delivery of functional warheads and tags while preserving its biological activity. Also, the ability of SpCID1 to introduce a single azide tag for FRET labeling and internal azide labeling can be further exploited for labeling biologically relevant RNAs for single molecule FRET experiments and for inserting site-specific internal biophysical probes on longer RNA.

Chapter 5: A Responsive Fluorescent Nucleotide Enables the Probing of SpCID1-Mediated Terminal Uridylation

Responsive fluorescent nucleotides have been previously employed for detecting variation in the local environment upon conformational switching of nucleic acids *in vitro*, in cells and, in cell-like confined environments.[21-23] SpCID1 has a tyrosine residue (Tyr212) in the catalytic cavity which is involved in stabilizing the last nucleobase of RNA and also UTP with the enzyme.[17,24] In this chapter, we perform direct incorporation of microenvironment sensing probes on 3' end of RNA, which is utilized for detecting RNA-SpCID1 interaction using the aromatic Tyr212 residue. For this, two modified nucleotides, 5-selenophene uridine triphosphate (SeUTP) and 5-benzofuran uridine triphosphate (BFUTP), reported previously from our laboratory, was used for terminal uridylation reactions (Figure iv).[25-26] We observed good incorporation of SeUTP at the 3' end of RNA, which saturated at +1 incorporation (Figure v). Surprisingly in the case of BFUTP, we observed multiple incorporation which did not get saturated at +4 incorporation. This was unexpected due to the fact that SeUTP and minimally perturbing triphosphate, AMUTP (used in Chapter 3) incorporated lesser modification at the 3' end as compared to relatively bulky BFUTP at similar reaction conditions. We hypothesized that this could be resulting from an aromatic stabilization of BFU and SpCID1 enzyme (possibly with Tyr 212). In order to investigate this aspect, we have synthesized a single BFU-labeled RNA and titrated it with SpCID1 enzyme, which resulted in concentration-dependent quenching in fluorescence (Figure vi). By also performing electrophoretic mobility shift assay and competitive binding assay, we have concluded that the quenching results from π-stacking interaction between Tyr 212 and the BFU probe on RNA.
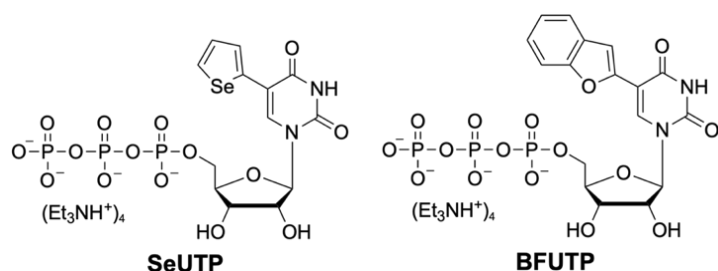
Figure iv. Microenvironment sensing UTP analogs used for terminal uridylation.[25-26]
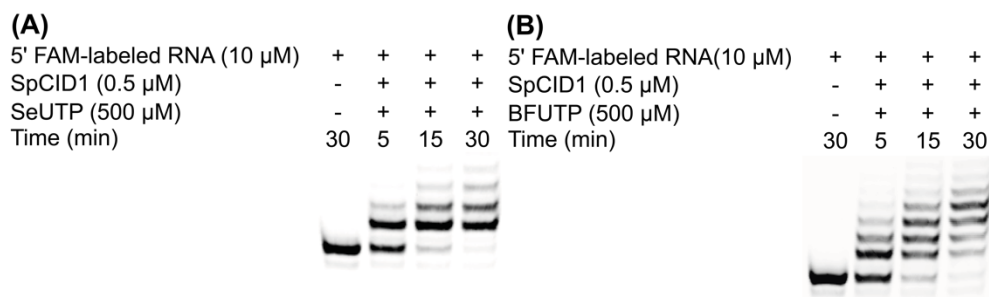
Figure v. Incorporation of (**A**) SeUTP and (**B**) BFUTP into 5′ FAM-labeled RNA using SpCID1.
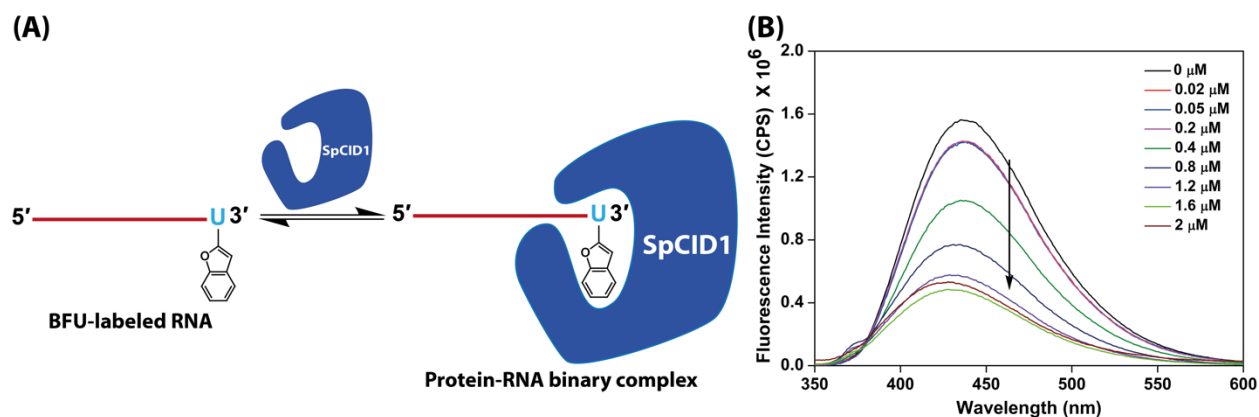
Figure vi. (**A**) Titration of BFU-labeled RNA with SpCID1. (**B**) Concentration-dependent quenching in the emission of BFU-labeled RNA with increasing SpCID1 enzyme.

Upon titrating BFUTP with increasing concentration of SpCID1, we observe a similar concentration-dependent quenching in fluorescence. Again, using competitive binding with natural UTP, we concluded that the quenching occurs due to stacking interaction between Tyr212 and BFU. Further, we calculated the dissociation constant of both BFU-labeled RNA and triphosphate, which experimentally validates the proposed model wherein UTP binding is the first step of enzyme cycle. Therefore direct attachment of microenvironment sensing molecules on UTP and on RNA using SpCID1 enables structural deconvolution of substrate-protein interactions in terminal uridylation.

References

1. Mortimer, S. A.; Kidwell, M. A.; Doudna, J. A., Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet* **2014,** *15* (7), 469–479.
2. Hermann, T.; Patel, D. J., RNA bulges as architectural and recognition motifs. *Structure* **2000,** *8* (3), R47–R54.
3. Sinkeldam, R. W.; Greco, N. J.; Tor, Y., Fluorescent analogs of biomolecular building blocks: Design, properties, and applications. *Chem. Rev.* **2010,** *110* (5), 2579–2619.
4. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G., Fluorescent nucleoside analogs: Probes for investigating nucleic acid structure and function. *Isr. J. Chem.* **2013,** *53* (6-7), 366–378.
5. Carlomagno, T., Present and future of NMR for RNA–protein complexes: A perspective of integrated structural biology. *Journal of Magnetic Resonance* **2014,** *241*, 126–136.
6. Ke, A.; Doudna, J. A., Crystallization of RNA and RNA–protein complexes. *Methods* **2004,** *34* (3), 408-414.
7. Wachowius, F.; Höbartner, C., Chemical RNA modifications for studies of RNA structure and dynamics. *ChemBioChem* **2010,** *11* (4), 469–480.
8. George, J. T.; Srivatsan, S. G., Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. *Methods* **2017,** *120,* 28–38.
9. 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.
10. Wu, B.; Chen, J.; Singer, R. H., Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Scientific Reports* **2014,** *4*, 3615.
11. Ozawa, T.; Natori, Y.; Sato, M.; Umezawa, Y., Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nature Methods* **2007,** *4*, 413.
12. Paige, J. S.; Wu, K. Y.; Jaffrey, S. R., RNA mimics of green fluorescent protein. *Science* **2011,** *333* (6042), 642–646.
13. Anhäuser, L.; Hüwel, S.; Zobel, T.; Rentmeister, A., Multiple covalent fluorescence labeling of eukaryotic mRNA at the poly(A) tail enhances translation and can be performed in living cells. *Nucleic Acids Res.* **2019,** *47* (7), e42–e42.
14. Holstein, J. M.; Rentmeister, A., Current covalent modification methods for detecting RNA in fixed and living cells. *Methods* **2016,** *98*, 18–25.
15. George, J. T.; Srivatsan, S. G., Vinyluridine as a versatile chemoselective handle for the post-transcriptional chemical functionalization of RNA. *Bioconjug Chem* **2017,** *28* (5), 1529-1536.
16. Sousa, R.; Chung, Y. J.; Rose, J. P.; Wang, B.-C., Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **1993,** *364* (6438), 593–599.
17. Lunde, B. M.; Magler, I.; Meinhart, A., Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res.* **2012,** *40* (19), 9815–9824.
18. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012,** *337* (6096), 816–821.
19. Doudna, J. A.; Charpentier, E., The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014,** *346* (6213).
20. Deng, W.; Shi, X.; Tjian, R.; Lionnet, T.; Singer, R. H., CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. USA* **2015,** *112* (38), 11870–11875.

21. Nuthanakanti, A.; Ahmed, I.; Khatik, S. Y.; Kayarat, S.; Srivatsan, S. G., Probing G-quadruplex topologies and recognition concurrently in real time and 3D using a dual-app nucleoside probe. *Nucleic Acids Res.* **2019**.

22. Manna, S.; Srivatsan, S. G., Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Advances* **2018,** *8* (45), 25673–25694.

23. Manna, S.; Sarkar, D.; Srivatsan, S. G., A dual-app nucleoside probe provides structural insights into the human telomeric overhang in live cells. *J. Am. Chem. Soc.* **2018,** *140* (39), 12622–12633.

24. Yates, L. A.; Fleurdépine, S.; Rissland, O. S.; De Colibus, L.; Harlos, K.; Norbury, C. J.; Gilbert, R. J. C., Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nature Structural &Amp; Molecular Biology* **2012,** *19*, 782.

25. Tanpure, A. A.; Srivatsan, S. G., A microenvironment-sensitive fluorescent pyrimidine ribonucleoside analogue: Synthesis, enzymatic incorporation, and fluorescence detection of a DNA abasic Site. *Chemistry – A European Journal* **2011,** *17* (45), 12820–12827.

26. Pawar, M. G.; Srivatsan, S. G., Synthesis, photophysical characterization, and enzymatic incorporation of a microenvironment-sensitive fluorescent uridine analog. *Org. Lett.* **2011,** *13* (5), 1114–1117.

List of publications

Based on the work presented in this thesis

1. **George, J. T**.; Srivatsan, S. G., Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. *Methods* **2017,** *120*, 28−38.
2. **George, J. T.**; Srivatsan, S. G., Vinyluridine as a Versatile Chemoselective Handle for the Post-transcriptional Chemical Functionalization of RNA. *Bioconjug Chem* **2017,** *28* (5), 1529−1536.
3. **George, J. T.**; Azhar, M.; Aich, M.; Sinha, D.; Ambi, U. B.; Maiti, S.; Chakraborty, D.; Srivatsan, S. G., TUTase mediated site-directed access to clickable chromatin employing CRISPR-dCas9. *bioRxiv* **2019**, 846980. (manuscript submitted)
4. **George, J. T.**; Ambi, U.; Srivatsan, S. G., Structural deconvolution of SpCID1 substrate binding using a responsive fluorescent nucleoside analog. (*manuscript under preparation*).

Other publication during MS-PhD

1. Sabale, P. M.; **George, J. T.**; Srivatsan, S. G., A base-modified PNA-graphene oxide platform as a turn-on fluorescence sensor for the detection of human telomeric repeats. *Nanoscale* **2014,** *6* (18), 10460−10469.

Chapter 1

Labeling Technologies for Investigating RNA Structure and Function[*]

1.1 Introduction

RNA plays a multi-protagonist role in cellular physiology, which includes storage and transfer of genetic information, catalytic processing of various RNA molecules and regulation of gene expression and epigenetic processes. RNA molecules perform these functions by utilizing its ability to fold into complex secondary and tertiary structures.[1-2] Compelling biochemical and biophysical investigations reveal that RNA, when executing its function, utilizes its inherent conformational dynamics to recognize and bind to nucleic acids, proteins and metabolites.[3-5] Further, the functioning of RNA is tightly regulated by its synthesis, maturation, transport, localization and degradation. Like any music note, the functional syncing between these events orchestrate the biological symphony. In order to understand RNA biogenesis and its functional intricacies, and draw cogent structure-function relationship, several biophysical techniques including fluorescence, NMR and X-ray crystallography have been developed.[6-9] Nucleosides that make up nucleic acids are practically non-fluorescent and do not contain intrinsic labels that can be efficiently utilized in the above-mentioned techniques. Therefore, the majority of these interrogations greatly rely on labeling methods that provide access to RNA labeled with appropriate reporter molecules.[10] These methods can be broadly categorized into covalent and non-covalent approaches (Figure 1.1). In this chapter, we mainly discuss the various covalent approaches to label RNA *in vitro* and in live cells with a greater emphasis on chemoenzymatic methods. Non-covalent approaches to label and visualize RNA in cellular environment will be discussed briefly, and the readers may refer to excellent reviews that have showcased the applications of non-covalent RNA labeling approaches.[11-13]
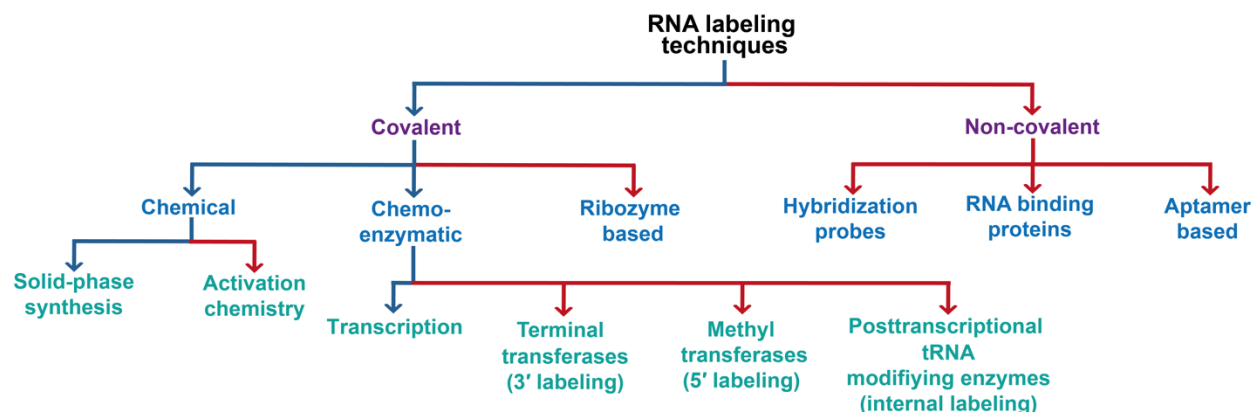
Figure 1.1. Flowchart elaborating different covalent and non-covalent approaches to label RNA. Synthetic and post-synthetic RNA tagging strategies represented with blue and red arrow lines.

1.2 Brief overview of methods to label and visualize RNA

1.2.1 Covalent methods to label RNA

Conventionally, solid-phase oligonucleotide (ON) synthesis chemistry and enzymatic reactions are employed in labeling RNA covalently with variety of probes.[14] In case of solid-phase ON synthesis, the modified nucleoside phosphoramidite substrate is site-specifically incorporated into an RNA ON by using iterations of deprotection, coupling, capping, and oxidation cycles. Although widely used, synthesis of functionalized phosphoramidite substrates could be laborious and, in several instances, the modified substrates do not survive the chemical manipulations used in the solid-phase ON synthesis protocols. Further, due to progressive reduction in coupling efficiency, synthesis of longer labeled RNA sequences (> 50 mer) result in poor yields. Alternatively, promiscuity of certain RNA polymerases and RNA processing enzymes to accept functionalized substrates has been employed in labeling RNA with variety of biophysical probes.[15] Bacteriophage RNA polymerases have been utilized in incorporating labels into RNA transcripts *in vitro* by using modified triphosphate substrates. Few modified nucleoside and nucleotide analogs have also been incorporated into cellular RNA transcripts by mammalian endogenous RNA polymerases. While this methodology is mild and can be used to incorporate sensitive modifications into longer RNA sequences, introducing bulky modifications and effecting site-specific labeling is very difficult. This issue has been partly negotiated by using unnatural base-pairs that are orthogonal to A-T and G-C pairing scheme and RNA ligating enzymes, which are discussed later in this chapter.

Postsynthetic chemical RNA labeling strategies employing activation chemistry using acid-amine or thiol-malemide Michael addition reaction have been used in several applications.

However, non-selectivity of these reaction and poor kinetics has impeded its use over time. Utility of these chemistries in the bioconjugation of ONs are reviewed in detail elsewhere.[16]

Recently, postsynthetic covalent strategies based on chemoselective reactions that have been successfully employed in labeling proteins, glycans, lipids and DNA ONs, have also been optimized to label RNA *in vitro* and in live cells.[15,17-19] In this methodology, a small reactive functionality is introduced into RNA by any one of the following methods–(i) chemically by solid-phase ON synthesis (ii) chemo-enzymatically using transcription reaction or posttranscriptionally utilizing RNA labeling enzymes. Following this step, the bioconjugation of RNA is achieved by performing a chemoselective reaction with a cognate reactive partner containing a desired reporter. The commonly used chemoselective reactions, which are also termed as bioorthogonal reactions in these cases due to their compatibility in biological systems, include Staudinger ligation, copper-catalyzed azide-alkyne cycloaddition (CuAAC), strain-promoted azide-alkyne cycloaddition (SPAAC) and inverse electron-demand Diels-Alder (IEDDA) reactions (Figure 1.2).
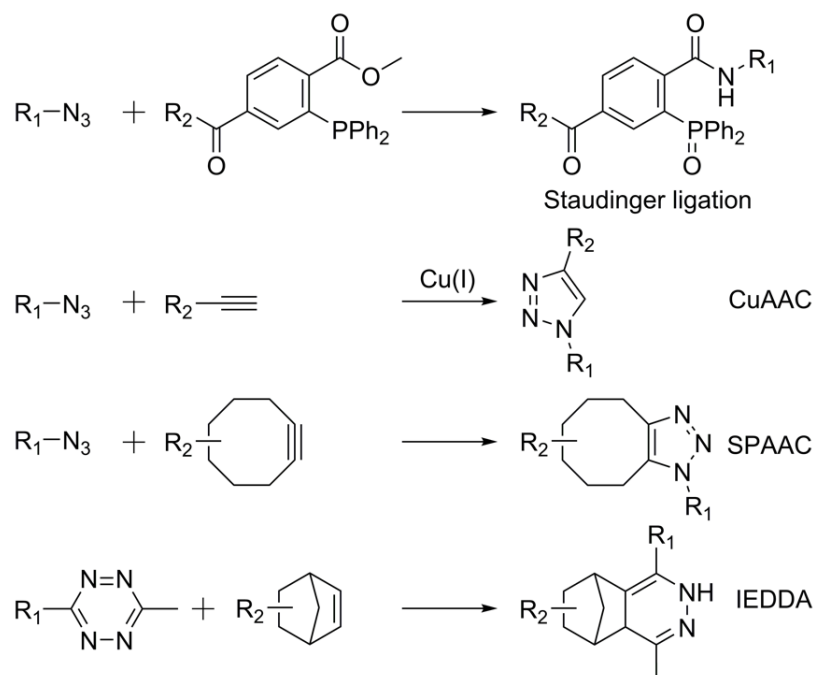
Figure 1.2. Types of bioorthogonal reaction commonly used in the labeling of RNA. Notably, the azide group serves as a common reactive species in Staudinger, CuAAC and SPAAC reactions.

1.2.2 Non-covalent methods to label RNA

Non-covalent approaches to label RNA are performed postsynthetically wherein *in situ* hybridization, RNA-binding proteins and aptamers are typically developed for the detection and

visualization of specific RNA sequences by fluorescence techniques.[11-12] RNA diagnostics based on hybridization method relies on fluorescently modified antisense oligomers (e.g. ONs, PNA and LNA) and molecular beacons (hairpin-shaped ONs containing a FRET pair), which upon binding to a target RNA motif endows fluorescence to the complex and facilitates the analysis of the target RNA by various fluorescence techniques including microscopy.[20-21] Again, these hybridization probes are either synthesized by chemical or enzymatic methods.

Genetically encoded systems have been developed to study RNAs in living cells. In one of the first examples, MS2 coat protein fused to GFP was used by Singer's group to locate and study the dynamics of RNA of interest, which was tagged to a series of MS2 RNA hairpin motifs.[22] This strategy was recently utilized in studying the trafficking of β-actin mRNAs in neuronal cells.[23] Singer and coworkers have improvised this technique and developed a background-free RNA imaging method by coexpressing two fusion proteins, which bind to the target RNA containing MS2 and PP7 binding sites.[24] A method utilizing sequence-specific RNA binding pumilio homolog domain was used by Umezawa and coworkwers to study the localization and dynamics of a mitochondrial RNA in single living cells.[25] Two pumilio homolog domain of human PUMILIO1 each having N- and C- terminal fragments of eGFP were brought together upon binding to the target RNA sequence to reconstitute eGFP, which facilitated the visualization of the RNA by fluorescence techniques.

Aptamers that bind to fluorophores and induce fluorescence emission have been used in imaging expressed RNAs. Jaffrey and coworkers have reported an aptamer-based approach wherein an RNA aptamer called spinach, obtained by directed evolution, is found to selectively bind to small molecule, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), which resembles GFP fluorophore.[26] Significant enhancement in emission of DFHBI on binding to spinach allowed its use in cellular application as a fluorogenic marker against spinach-tagged RNA. Improvements as well as *in vitro* selection against other fluorophores have led to the development of spinach2[27], broccoli[28], sulforhodamine[29], malachite green[30] and mango aptamers[31], which are brighter and exhibit emission at different wavelengths. However, it is important to note that insertion of large and or multiple RNA recognition elements into the target RNA may alter its folding and functioning.

1.3 Bioorthogonal chemistry as a tool for labeling nucleic acids

Pioneering work of Saxon and Bertozzi lead to the development of Staudinger ligation reaction, which represents the first example of a completely bioorthogonal reaction.[32] This reagent-free reaction exploits the facile reaction between an azide and a phosphine substrate to produce an aza-ylide intermediate, which is conveniently trapped by an internal acyl group to form a stable amide bond. This bioconjugation strategy became a valuable method for metabolic engineering of cell surface glycans, which was subsequently adapted to label proteins and nucleic acids.[33-36] In one of the examples, Weisbrod and Marx have successfully incorporated an azide-modified nucleotide into ONs by PCR and labeled them with a biotin tag by Staudinger ligation reaction.[37] Although this bioconjugation method is promising, poor stability of phosphine substrates in living systems as well as slow reaction kinetics have hampered its wide application.

Later, azide-alkyne cycloaddition (AAC) reaction in the presence of copper, developed independently by Sharpless and Meldal, superseded Staudinger ligation as a bioorthogonal tool to label glycans, proteins, lipids and more recently, nucleic acids.[38] The cytotoxicity of copper, which was a major concern in live cell applications, was largely overcome by using water-soluble Cu(I) stabilizing ligands such as tris(3-hydroxypropyltriazolylmethyl)amine (THPTA)[39-40] and tris-(benzyltriazolylmethyl)amine (TBTA)[41]. Bertozzi and coworkers, using an old physical organic concept, showed cycloaddition reaction between a strained alkyne (cyclooctyne derivative) and an azide substrate, which proceeded significantly faster than the linear alkynes.[42] This version of the copper-free cycloaddition reaction, termed as SPAAC reaction, has developed to an extent that it can be conveniently used to label glycans of zebrafish embryos[43], proteins[44], lipids[45-46] and nucleic acids[47]. Brown and coworkers have extended the utility of this reaction for internal DNA labeling by incorporating phosphoramidite monomer containing a strained cyclooctyne. The strained alkyne label on the ON efficiently reacted with complementary azide probes in the absence of any added reagents.[48]

Analogous to azide-alkyne cycloaddition reactions, inverse-electron demand Diels-Alder (IEDDA) reaction between a strained alkene and an electron deficient tetrazine has recently gained popularity as a nucleic acid bioconjugation strategy.[49] The key advantages of IEDDA reaction are, it is reagent free, significantly fast, specific and biocompatible. Typically, in this strategy, strained dienophiles like norbornene, trans-cyclooctene and cyclopropene derivatives are incorporated into ONs and subsequent functionalization is achieved by reacting with tetrazine probes.[50-55]

Although these bioorthogonal strategies have been successfully implemented in the analysis of glycans, proteins and DNA in both cell free and in vivo systems, their utility in the investigation of RNA, particularly in cellular environment, has been a major challenge. This is largely due to the low inherent hydrolytic stability of RNA and shortcomings in methods to incorporate chemoselective reactive handles into RNA *in vitro* and in cells. Furthermore, the reaction conditions that worked really well for other biopolymers had to be extensively optimized in case of RNA.[19] Nevertheless, chemists and chemical biologists interested in RNA biology have found ways to utilize this powerful technology in understanding RNA structure and function. Particularly, nucleosides, nucleotides and enzyme co-substrates, decorated with chemoselective reactive functionalities, which are compatible to posttranscriptional chemical modification, have been very valuable not only in visualizing RNA but also in studying RNA synthesis, turnover and processing in living cells. In the following sections, we provide a detailed overview of the application of different transcriptional or posttranscriptional chemo-enzymatic RNA labeling strategies based on bioorthogonal chemistry.

1.4 Chemo-enzymatic methods to label RNA employing bioorthogonal chemistry

Inaccessibility of appropriate phosphoramidites, instability of certain phosphoramidites in solid-phase ON synthesis conditions and incompatibility of modified nucleotides in enzymatic incorporation methods led to the development of postsynthetic ON modification strategies. Recent developments in RNA labeling technologies based on bioorthogonal chemistry have provided new opportunities to study RNA in cell-free and cellular settings.[15,56] For convenience, we have classified these chemo-enzymatic RNA labeling methods as transcriptional incorporation and transferase-assisted incorporation (Figure 1.3).

(A) Transcriptional incorporation
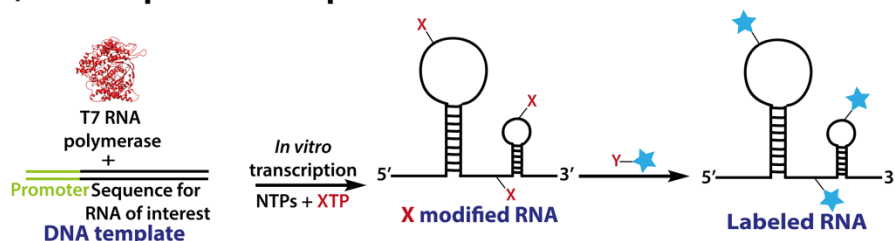
(B) Transferase-assisted incorporation
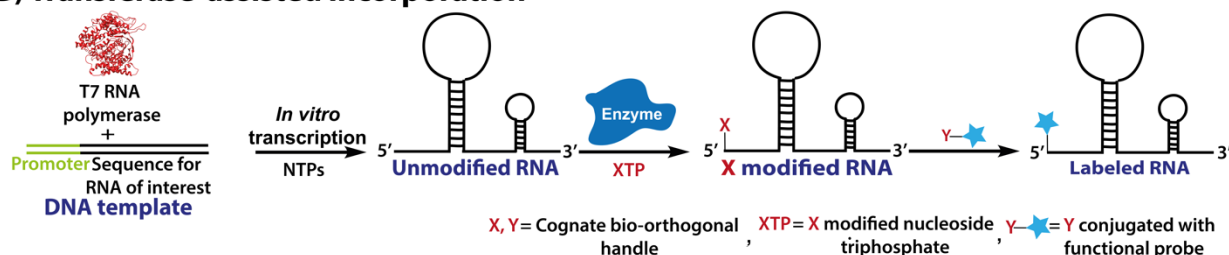
Figure 1.3. (**A**) Transcriptional incorporation and (**B**) Transferase-assisted incorporation of bioorthogonal labels for chemo-enzymatic labeling of RNA. Structure of T7 RNA polymerase adopted from PDB: 4RNP.[57]

1.4.1 Transcriptional incorporation

Azide-alkyne cycloaddition reaction

Postsynthetic bioconjugation of ONs by azide-alkyne cycloaddition (AAC) reaction largely preferred the installation of alkyne functionality into ONs because stable alkyne-modified substrates suitable for solid-phase as well as enzymatic incorporation could be readily prepared. Postsynthetic modification using CuAAC reaction was first established in DNA.[38,58] Seela and Sirivolu developed a series of 7-deazapurines and 5-substituted 2′-deoxyuridines bearing diyne groups with terminal triple bonds. These nucleosides were incorporated into DNA ONs by phosphoramidite chemistry and further functionalized by click reaction.[59-60] Carell and coworkers prepared the triphosphate of 5-ethynyl (EdUTP) and 5-diyne-substituted 2′-deoxyuridine analogs and successfully incorporated them into DNA by PCR.[61] These DNAs carrying high density of alkyne groups were suitable for click functionalization. While DNAs bearing flexible diyne group gave quantitative labeling, click reactions with DNAs containing more rigid alkyne (EdU) resulted in partial labeling. This methodology was further extended to label ONs with multiple tags including affinity and fluorescent probes.[62-63] Brown and coworkers click-stitched a DNA strand containing alkyne at the 5′ end and an azide at the 3′ end, which was incorporated by using NHS chemistry, to produce an intramolecular circular DNA.[64] The circular DNA was used as the template to synthesize covalently closed DNA catenane. Later, Salic and Nathan metabolically

incorporated EdU and an arabinofuranosyl ethynyluracil derivative into DNA, and used CuAAC reaction to detect the DNA synthesis in cells.[65-66]

Despite such successes with DNA, click chemistry approach to label RNA while keeping the RNA intact was a challenge. Click labeling of RNA was first demonstrated by Jao and Salic, wherein cellular RNA transcripts were visualized by click-staining metabolically incorporated 5-ethynyluridine (EU) with fluorescent azides (Figure 1.4).[67] However, studies to confirm if the EU-labeled RNA transcripts are functional and the intactness of clicked-labeled RNAs under CuAAC reaction conditions were not reported. Das and coworker systematically evaluated the conditions required for rapid RNA labeling with minimum degradation. For this purpose, RNA ONs containing alkyne and azide moieties were synthesized by chemical and *in vitro* transcription reactions in the presence of alkyne-modified phosphoramidites and an azide-modified guanosine initiator nucleotide, respectively. Their results indicated that the degradation of RNA could be significantly minimized by performing the reaction at a higher concentration of RNA (100 μM or more) in the presence of a Cu(I) stabilizing ligand (THPTA) or simply small amounts of acetonitrile as a cosolvent.[68-69]
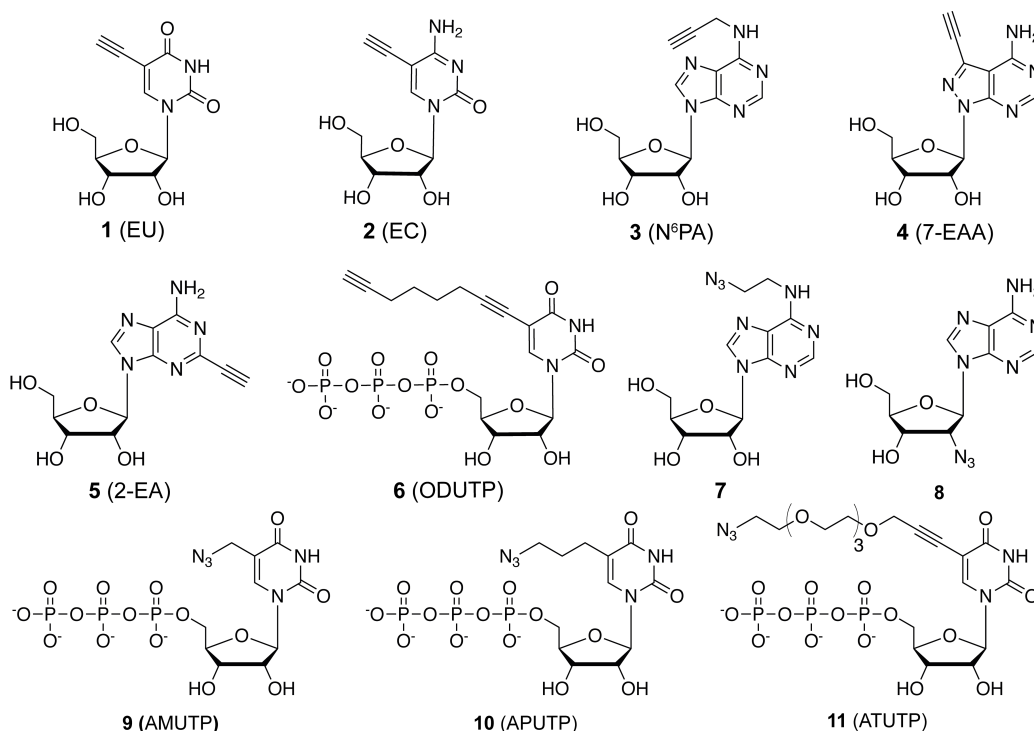
Figure 1.4. Chemical structure of alkyne- and azide-modified nucleoside and nucleotide analogs employed in posttranscriptional chemical functionalization of RNA using azide-alkyne cycloaddition reactions.

El-Sagheer and Brown employed intrastrand click ligation approach to produce a functional hammerhead ribozyme containing a triazole backbone at the catalytic site.[40] Two RNA strands containing an alkyne and azide group at 3′ and 5′ ends, respectively, were splint-ligated by using CuAAC reaction in the presence of a ligand, THPTA. The ribozyme prepared by this method cleaved the substrate at the predicted scissile phosphodiester bond similar to that of the native ribozyme. Carell and coworkers prepared different siRNA-receptor conjugates by click reacting alkyne-modified RNA with folate, cholesterol and arachidonoyl ethanol amide (anandamide) containing azide group. The anandamide-tagged siRNA exhibited very high transfection efficiency with very low toxicity, and enabled the gene silencing in neuronal and immune cells.[70]

More recently, Beal's group synthesized 7-ethynyl-8-aza-7-deazaadenosine (7-EAA) and the corresponding triphosphate (7-EAATP).[71] The ribonucleoside was supplemented to HeLa cells in culture, and metabolically incorporation of 7-EAA into cellular RNA transcripts was confirmed by Northern blotting the products obtained from a CuAAC reaction between isolated RNA and biotin-azide. Their results suggested that 7-EAA could be used in profiling polyadenylation in RNA. Further, the triphosphate of 7-EAA was found to serve as a good substrate for T7 RNA polymerase and *Escherichia coli* poly (A) polymerase.

The versatility of azide group to participate in a wide range of bioorthogonal reactions has been extensively utilized in labeling and visualizing glycans and proteins by CuAAC, SPAAC and Staudinger reactions. However, its wide application in labeling nucleic acids has been hampered because of the lack of stable azide substrates that can be metabolically introduced into cellular DNA and RNA. Moreover, incorporation of azide groups into ONs by using conventional phosphoramidite chemistry is not easy as azides undergo Staudinger type reaction with phosphoramidite substrates.[72] Apart from NHS ester chemistry and enzymatic incorporation by using initiator guanosine nucleotides, recently, few methods have been developed to end-label ONs with azide groups by using azide-modified solid supports and azide displacement reaction on the solid-support.[73-76] To address these problems and develop a modular method to label RNA with azide functionality in cell-free and cellular environments, we decided to use transcription reaction as a mode to install azide group into RNA. For this purpose, we developed a toolbox composed of 5-azidomethyl-, 5-azidopropyl-, and 5-azidoTeg-modified uridine triphosphates (Figure 1.4).[77-78] The triphosphates (**9-11**) were designed in such a fashion that the azide group was connected to the uracil ring via a short (methyl), medium (propyl) and long (tetraethylene glycol) linker.

Irrespective of the linker length, all the three triphosphates were efficiently incorporated into RNA transcripts by T7 RNA polymerase catalyzed *in vitro* transcription reaction. The azide-labeled RNA transcripts were conveniently functionalized posttranscriptionally with a variety of biophysical probes by CuAAC, SPAAC and Staudinger ligation reactions.[79] The labels ranged from fluorescence to affinity to amino acid. The RNA degradation was negligible and all the reactions afforded sufficient amounts of click-labeled RNA product for biophysical analysis. Further, high density functionalizing of RNA of 400-1100 bases length containing nearly 100-260 azide labels could be achieved by *in vitro* transcription reaction using triphosphates **9-11**. The azide-labeled RNA transcripts were effectively reverse transcribed to produce respective cDNA, which was amplified by PCR. Subsequent cloning and sequencing revealed that azide-modified UTPs are efficiently incorporated into RNA transcripts, which are copied by reverse transcriptase with high fidelity.[79] The ability to preserve the sequence information in a cycle of steps namely, transcription, reverse transcription and PCR indicates that the azide-modified UTPs could be highly suitable in expanding the chemical space of RNA libraries used in aptamers and ribozyme selection protocols.[80] Similar property of EdU has been exploited by Mayer and coworkers to establish click chemistry-based selection protocol to identify an aptamer that recognizes case cycle 3 GFP with high specificity.[81]

Hirao's group pioneered the development of one of the few unnatural base pairing schemes orthogonal to A-T and G-C pair. The hydrophobic base pairing interaction between 7-(2-thienyl)imidazo[4,5-b]pyridine (Ds) and pyrrole-2- carbaldehyde (Pa) was smartly utilized in site-specific enzymatic incorporation of nucleotide analogs into DNA and RNA.[82-84] The utility of this base pair was further extended by his group to introduce alkyne and azide functionalities site-specifically into RNA transcripts by *in vitr*o transcription reactions.[85-86] For this purpose, alkyne and azide-modified Pa ribonucleotides (**12-14**) were synthesized and efficiently incorporated opposite to Ds present in the coding region of the DNA template (Figure 1.5). The site-specifically labeled RNA transcripts were then functionalized by click reaction with probe of interest.

Recently, Jäschke and coworkers synthesized an octadynyl modified dinucleotide analog (OdUpG) as a transcriptional priming agent to incorporate alkyne group at the 5' end of the transcripts.[87] Bacteriophage RNA polymerases facilitated near-quantitative priming and bioconjugation with different azide labels using CuAAC reaction. An added advantage of these

transcripts is that they bear a ligation friendly 5' end, which could be transformed into a site-specific internal alkyne label by enzymatic ligation reaction.
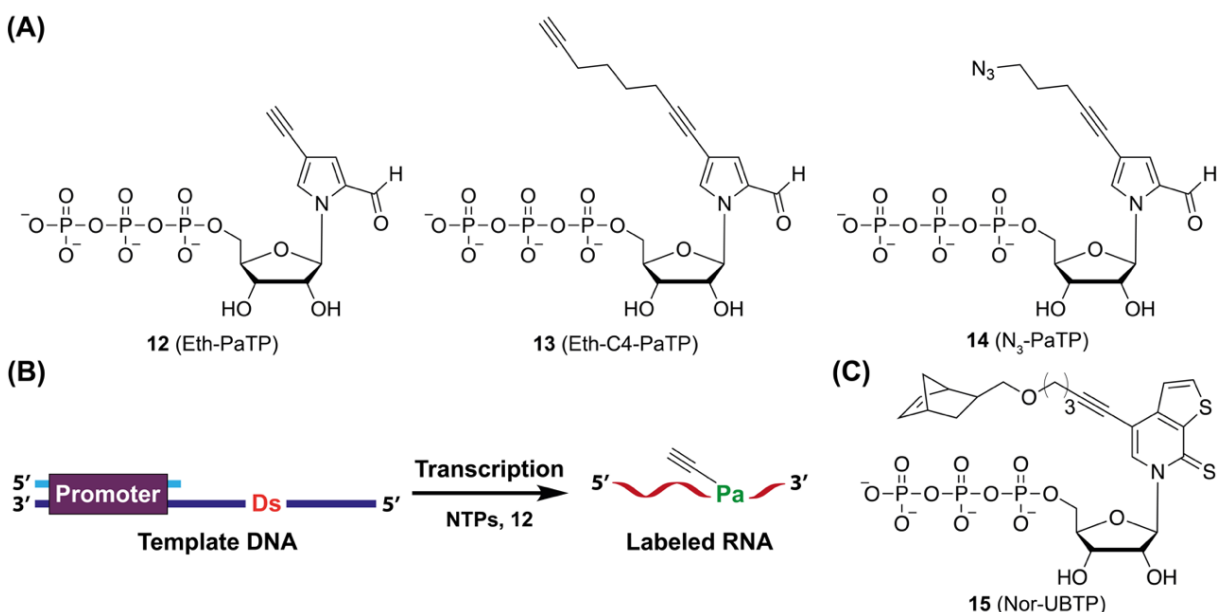
Figure 1.5. Posttranscriptional chemical modification using unnatural base pairing schemes. (**A**) Alkyne- and azide-modified unnatural nucleoside triphosphate, Eth-PaTP, Eth-C4-PaTP and N$_3$-PaTP developed by Hirao and coworkers.[85-86] (**B**) Transcriptional incorporation of modified Pa triphosphates *in vitro*, which is further, functionalized using Alexa azide/strained alkynes. (**C**) Norbornene functionalized unnatural nucleoside triphosphate, Nor-UBTP developed by Kath-Schorr and coworkers.[88] This nucleotide was effectively incorporated into RNA transcripts opposite to its cognate base pair (dNaM) in the coding region of DNA template. Further posttranscriptional modification by IEDDA enabled the labeling of RNA with fluorescent dyes.

Inverse electron-demand Diels-Alder reaction

RNA labeling by using IEDDA reaction was first demonstrated by Jäschke and co-workers. Norbornene-modified phosphoramidite and guanosine initiator nucleotides were synthesized and incorporated by solid-phase and *in vitro* transcription reaction, respectively.[89] The RNA ONs containing the strained dienophile were successfully modified by using IEDDA reactions in the presence of biotin- and dansyl-modified tetrazines. Further, Jäschke and coworkers utilized this posttranscriptional conjugation approach to prepare RNA-peptide conjugates. Guanosine and adenosine initiator nucleotides containing reactive dienophiles like norbornene and bicyclononyne were synthesized and efficiently 5' end labeled by transcription reaction.[90] The dienophile-labeled RNAs were subsequently conjugated to a tetrazine-derivatized peptide.

Kath-Schorr and coworkers demonstrated that a reaction between a norbornene-labeled RNA duplex and cognate fluorescent tetrazines in cells can be detected by flow cytometry and fluorescence microscopy.[91] This development points out that such an approach could be potentially useful in the detection of specific RNA sequences in cellular settings. Following this report, her group developed a method to site-specifically introduce IEDDA reaction-compatible alkene moiety into RNA by *in vitro* transcription reaction.[88] They took advantage of the hydrophobic base paring system (UB and NaM) developed by Romesberg's group.[92] Norbornene-modified UBTP **15** was synthesized and was incorporated into the transcribing RNA opposite to dNaM present in the coding region of the DNA template. The resulting site-specifically labeled RNA was tagged to fluorescent reporters by reacting with appropriate tetrazines.[88] In a similar approach, methyl cyclopropene, a dienophile that reacts faster and yields fewer isomers than norbornenes, was incorporated at predefined positions in RNA and posttranscriptionally conjugated to fluorescent tetrazines.[93] Royzen and coworkers developed a general fluorescent RNA labeling method, wherein a cytidine triphosphate analog containing a trans-cyclooctene group was introduced into miR-122 by *in vitro* transcription reaction.[94] The trans-cyclooctene-labeled miRNA was fluorescently labeled by reacting with fluorescein-modified tetrazine. While these strategies based on IEDDA reactions are useful, their utility in labeling endogenous RNA has not been realized yet. The existing dienophiles are definitely reactive and specific, but are bulky and hence, may not be suitable for metabolic incorporation by endogenous polymerases.

1.4.2 Transferase-assisted incorporation

Unlike incorporation of reactive groups into RNA during transcription reaction, clickable functionalities can be incorporated into RNA by using the promiscuity of certain RNA processing enzymes like transferases.[15] For example, an azide- or alkyne-modified co-substrate is used to transfer the reactive group to the RNA transcript of interest *in vitro* using an appropriate enzyme and further post-synthetic labeling is performed by using click reaction (Figures. 1.3B, 1.6 and 1.7).

Jäschke and coworkers evaluated the ability of nucleotidyl transferases namely, yeast and *E. coli* poly(A) polymerase (PAP) to incorporate a set of dNTPs/NTPs containing azide group at different positions of the nucleoside (C-8, C-2' and C-3').[95] PAP facilitated the incorporation of both base-modified and sugar-modified nucleotides into 3' end of RNA with moderate to good

efficiency (Figure 1.6). Further, these labeled RNAs were tagged with fluorophores. via CuAAC and SPAAC reactions. Splint ligation followed by click reaction also enabled site-specific internal labeling of RNA.

Recently, Pomerantz and coworkers have introduced functional tags and azide tag at the 3' of RNA using human polymerase theta (pol θ). Also, addition of functional tags can be controlled by using 2',3'-dideoxyuridine nucleotides.[96] However, non-templated addition with modified nucelotides results in visibly compromised yields indicated by the presence of large amounts of unlabeled RNA (reactant) observed in the reported gel images.
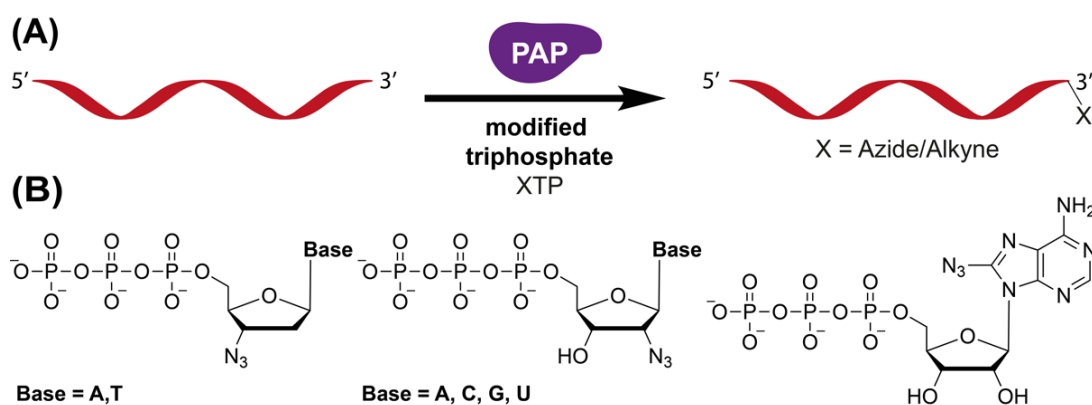
Figure 1.6. (**A**) Terminal 3′ labeling of RNA using PAP in the presence of modified nucleotides. (**B**) Azide-modified nucleotide analogs used by Jäschke and coworkers.[95]

Methyl transferases (MTases) are class of enzymes, which use S-adenosyl-L-methionine (AdoMet) cofactor to transfer a methyl group to the nucleobase of DNA and RNA.[97] Lately, the substrate tolerance of these enzymes has been used in tagging nucleic acids[98-99] and proteins[100] with functional markers. Helm and coworkers used tRNA methyltransferase (Trm1) and a modified cofactor, AdoEnYn, bearing a terminal alkyne to label tRNA[Phe]. The enzyme efficiently transferred the alkyne group at the N2 position of guanosine 26 of tRNA[Phe], which was subsequently click-functionalized with Alexa fluorophore for single molecule fluorescence experiments.[101] A combination of ribonucleoprotein RNA 2'-O-methyltransferase (box C/D RNP) and a synthetic AdoMet analog containing alkyne functionality (SeAdoYn) was used in site-specifically transferring a prop-2-ynyl group to both the wild-type and programmed sites in RNA transcripts.[102] The target RNA was further labeled with fluorescent tags using CuAAC reaction. Another MTase known to methylate miRNA in plants is HEN1 2'-O-methyl transferase. HEN1 transfers a methyl group to the 2' OH of 3' terminus of miRNA double strand. Alkyne and amine

containing AdoMets were used in transferring the respective functionality at the 3' end of double stranded miRNA and siRNA, which were then postsynthetically labeled using click and NHS ester chemistry.[103]

Rentmeister and coworkers used *Giardia lamblia* trimethylguanosine synthase (GlaTgs2) for introducing bioorthogonal reactive labels at the 5' end of RNA. This MTase is known to introduce a methyl group at the N2 position of the dinucleotide RNA cap, m^{7}GpppA using Ado-Met. Since the wild type GlaTgs2 showed a reduced preference for modified cofactors, an engineered variant (GlaTgs2-var1) was developed, which showed increased transfer efficiency. The enzyme variant, in the presence of synthetic AdoMet substrates, was used in covalently tagging alkyne and alkene groups onto RNA cap (Figure 1.7A). The mRNAs containing reactive handles were then successfully click-functionalized by CuAAC and thiol-ene reactions.[104] In a similar strategy, azide- and vinylbenzene-modified AdoMet analogs were used to transfer azide and vinyl-benzene residues to the RNA cap for further functionalization by SPAAC, IEDDA and photoclick reactions.[105-106] More recently, functionalized mRNAs obtained by this methodology using Ecm1 methyl transferase were aptly utilized in tuning the translation efficiency of specific mRNA sequences in living cells[107].

Suzuki and coworkers identified a new enzyme, tRNA^{Ile2}-agm^{2}C synthetase (TiaS), that catalyzes the formation of 2-agmatinylcytidine (agm^{2}C) at the wobble position of archael tRNA^{Ile2} in the presence of agmatine and ATP.[108] Based on the RNA sequence specificity and plasticity of TiaS, Wang and coworkers demonstrated the possibility of internal site- and sequence-specific transfer of azide and alkyne groups to a desired RNA *in vitro* and in mammalian cells (Figure 1.7B). Alkyne and azide modified agmatine analogs were synthesized and transferred specifically to C34 residue of the tRNA^{Ile2} using Tias.[109] The tRNAs containing reactive agmatines were then fluorescently labeled by using CuAAC and SPAAC reactions. While this method allows transfer of click groups at an internal position of RNA, the method is not applicable to other RNA sequences and therefore is not a general approach.
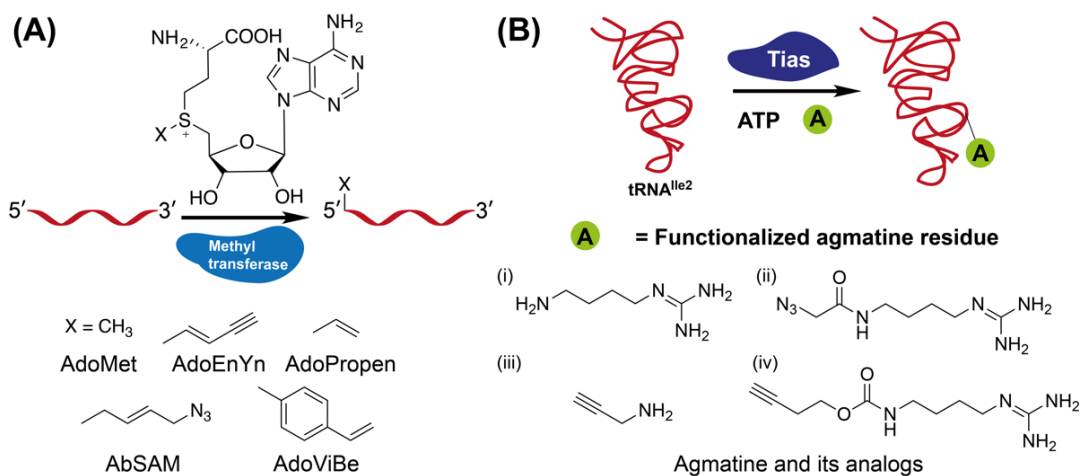
Figure 1.7. (**A**) Posttranscriptional labeling of 5′ RNA cap, exploiting the promiscuity of transferase enzymes to accept modified AdoMet cofactors, shown by Rentmeister group.[104,107] (**B**) Tias-mediated labeling of tRNA^{Ile2} using azide- and alkyne-containing agmatine analogs in the presence of ATP.[109]

1.5 Ribozyme-assisted approaches to label RNA

Self-alkylating ribozymes are also useful in labeling RNA with fluorescence and affinity tags. In one of the early reports, Szostak and Wilson used systematic evolution to isolate a ribozyme that selectively alkylated the N7 position of adenosine with a biotin tag.[110] Recently, Hemstra and coworkers reported a similar approach to introduce fluorescent marker on to RNA (Figure 1.8A). Here, the ribozyme was enriched via immunoprecipitation SELEX (IP SELEX) and the selected RNA was found to catalyze its reaction with fluorescein iodoacetamide to give the labeled RNA.[111] Liu's group reported a self-alkylating RNA that relied on probes derivatized with electrophilic groups to get selectively incorporated into a 42-nt catalytic RNA. Using the ribozyme, fluorescent and biotin probes with epoxy modification were incorporated into RNA posttranscriptionally.[112]

A different kind of ribozyme-based posttranscriptional labeling method was developed by Muller and coworkers. A twin ribozyme, HP-TW5 that could aid the insertion of a short RNA substrate, RS-TW5-X by strand exchange was engineered. Amine functionality was introduced on the substrate RNA (RS-TW5-X) using amine modified deoxythymidine residue (Figure 1.8B). Activation chemistry in the presence of fluorophore isothiocynate or succinamide ester was used to functionalize the RNA substrate. The pre- functionalized RNA substrate was then inserted into the ribozyme (HP-TW5) by strand exchange reaction.[113] Höbartner and coworker reported the use of deoxyribozymes to posttranscriptionally label RNA.[114] The DNAzymes in the presence of Tb^{3+} cofactor catalyzed the attachment of various modified guanosine monophosphates at the 2'-OH of internal adenosines of *in vitro* transcribed RNA. Guanosine triphosphate decorated with spin,

fluorescent, azide and biotin labels were successfully installed on to RNA using these DNAzymes (Figure 1.8C). Although this method is highly useful, ribozymes are restricted to a certain sequence of RNA and is therefore not a general approach.
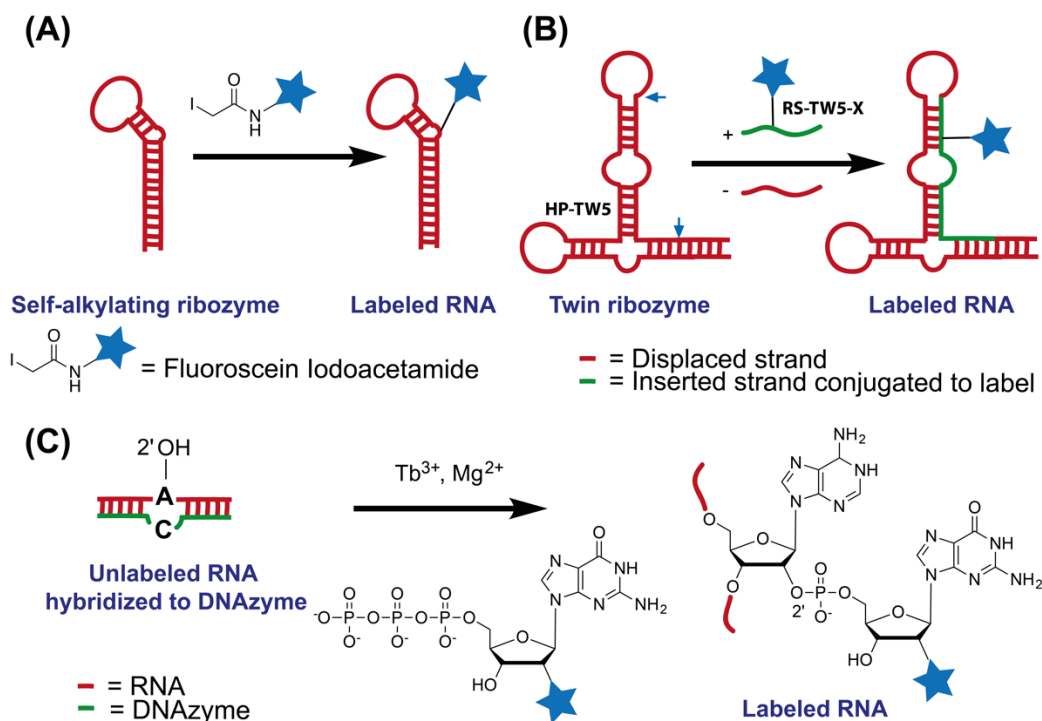
Figure 1.8. Ribozyme-assisted labeling of RNA. (**A**) Posttranscriptional labeling of RNA by self-alkylating ribozyme employing fluorescein iodoacetamide.[111] (**B**) Twin ribozyme (HP- TW5) mediated insertion of short-labeled RNA fragment (RS-TW-X) by strand exchange.[113] (**C**) DNAzymes-assisted labeling of 2′-OH of internal adenosine residues using labeled GTP in the presence of Tb^{3+}.[114]

1.6. Posttranscriptional labeling of RNA in cells

Labeling and imaging RNA in cells is more challenging than labeling RNA with probes in cell-free systems. In particular, microscopy methods to visualize RNA synthesis, transport and localization require labeling of RNA with probes that have high quantum yield and least fluorescence background. Apart from these characteristics, there should be a high local density of the fluorophore for obtaining a good signal to noise ratio. Some of the labeling protocols mentioned in the previous section have been used to detect exogenous and endogenous RNAs.

Wang and coworkers established an RNA imaging method using Tias enzyme, which specifically transfers a clickable group onto tRNA[Ile2] expressed in mammalian cells in the presence of a modified agmatine substrate.[109] A fusion RNA consisting of 5S ribosomal RNA and tRNA[Ile2]

was transcribed in U2OS cells expressing Tias enzyme. Cells grown in culture were supplemented with 2-propynylamine-modified agmatine and further were fixed and permeabilized. Click-staining the cells using sulfo-Cy5-azide enabled the selective detection of tRNA[Ile2]-5S fusion RNA.

Rentmeister and coworkers employed their 5'-cap RNA labeling method to visualize exogenous RNA in cells by click reaction. 5'-azide-labeled mRNAs obtained by transfer reaction using methyl transferase, Ecm1 and its co-substrate azide-modified AdoMet were transfected into HeLa cells. The live cells were treated with a strained cyclooctyne containing a fluorescent tag to facilitate SPAAC reaction, following which, the clicked RNA product in cells was visualized using a confocal microscope.[107] Recently, her group introduced multiple azide-modified adenylate residues onto the 3' end of eGFP and luciferase mRNA using yeast poly(A) polymerase which increases the translational efficiency of labeled mRNA. Further, the SPAAC functionalized mRNA has been used for simultaneous visualization of both mRNA and translated protein.[115]

Metabolic labeling of endogenous RNA and its posttranscriptional functionalization

Several labeling methods have been developed to investigate the steady-state level of specific RNA sequences and whole-cell RNA population in order to understand the interplay of RNA synthesis, processing and degradation in gene regulation and protein expression. Metabolic labeling of RNA transcripts and fluorescence in situ hybridization (FISH, discussed elsewhere) are the most widely used techniques for labeling and visualizing cellular RNAs.[11,21] In the metabolic labeling approach, a nucleoside analog, converted into its 5' nucleoside triphosphate by salvage pathway, is incorporated into RNA transcripts by endogenous enzymes. Alternatively, transfected nucleoside triphosphate analog can also be directly incorporated into transcripts by RNA polymerases. Following this step, the labeled RNA transcripts are either visualized inside the cells or isolated and analyzed by using antibodies against the nucleoside analog or by using bioorthogonal chemistry.

BrU and BrUTP are very popular metabolic labels used for studying the stability and steady-state levels of RNA.[116-118] Antibody against BrU can selectively label nascent RNA populations with fluorescent or biotin tags for RNA visualization or pull down. In a recent report, Ljungman and coworkers utilized BrU labeling combined with sequencing (Bru-Seq) and pulse-chase experiments for developing a methodology for genome-level investigation of RNA levels

and its stability.[119] While utilities of such halogenated nucleosides are invaluable, limited permeability of the nucleoside analogs and antibodies restrict their use to certain cell types and tissue samples.

Similar to BrU, 4-thiouridine (4sU) is also used in metabolic labeling of endogenous RNA for photo-crosslinking and for studying nascent RNA expression levels.[120] Rabani et al. studied the temporal changes in RNA level in mouse dendritic cells in response to lipopolysaccharide (LPS). Cells in culture were supplemented with 4sU and LPS. Total RNA was isolated, and newly transcribed RNAs labeled with 4sU were selectively and covalently labeled with biotin tag in reducing chemical environment.[121] The biotin-tagged RNA molecules were separated from whole-cell RNA by using streptavidin magnetic beads and sequenced. Advanced RNA quantification and computation modeling revealed that the dynamics of RNA production and degradation are tightly interconnected. Thiol-specific biotinylation on RNA can be performed by using pyridyldithiol-activated biotinylation reagents (biotin-HPDP) that selectively form a reversible disulphide bond between thiol-labeled RNA and biotin. Dölken and coworkers used this chemistry to selective pull-down 4sU labeled cellular RNAs, which were then subjected to quantitative RT-PCR and next generation sequencing to profile the RNA synthesis and decay.[122] 4sU labeling method was further used in understanding the herpes simplex virus (HSV-1) induced host shut-off mechanism. RNA profiling data revealed that HSV-1 disrupts the host-cell transcription termination process, but not its own genes.[123] 4sU is also used in photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) experiments.[124] Herein, the cellular RNA is labeled with 4sU and is UV irradiated to facilitate photo-crosslinking between nascent RNA and RNA binding protein. Despite its utility, the drawbacks of 4sU-based labeling method are inhibition of rRNA synthesis, nucleolar stress and cross reaction with other thiol-containing molecules in cells.[125]

Posttranscriptional labeling of cellular RNA transcripts decorated with bioorthogonal reactive handles is emerging as an alternative method to image and profile endogenous RNA (Figure 1.9). Jao and Salic provided the first example of the use of click chemistry in monitoring RNA synthesis and turnover in cells.[67] Cells in culture were supplemented with a clickable nucleoside analog (EU), which was specifically incorporated into nascent RNA transcripts by nucleoside salvage pathway. Further, click-staining the labeled RNA transcripts with fluorescent azide under CuAAC reaction conditions enabled the visualization of newly transcribing RNA in fixed cells. This method was extended to image transcription levels in organs of a mouse injected

with EU. Images were obtained by staining the ex vivo sections of organs using click reaction. In a similar approach, Zhang and coworkers described the posttranscriptional labeling of cellular RNA transcripts using 5-ethynylcytidine (EC).[126] While EC was efficiently incorporated into cellular RNA by endogenous RNA polymerases, 8-ethynyladenosine was not found to be a good substrate for metabolic labeling.
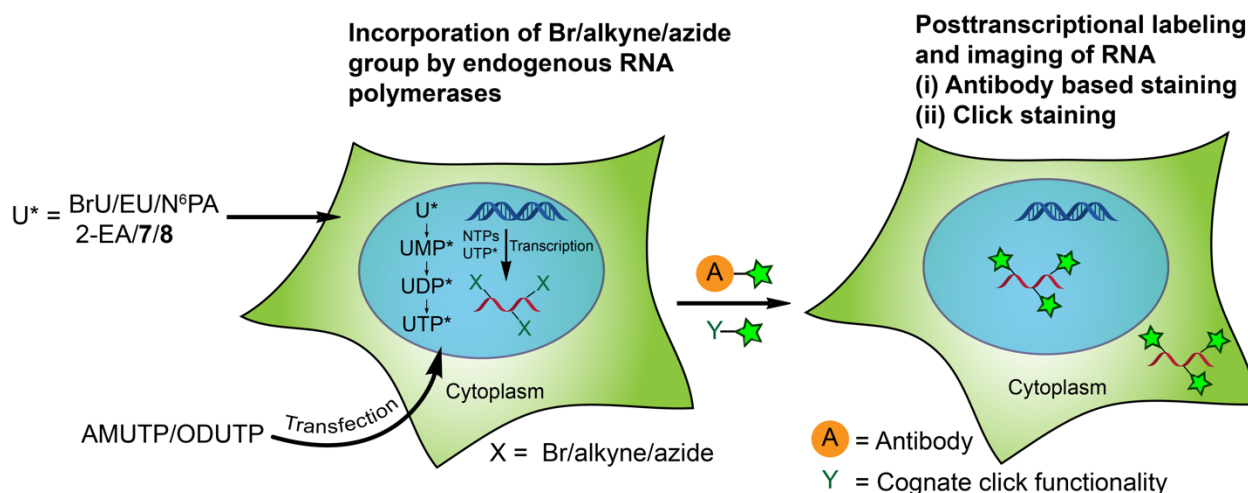
Figure 1.9. Metabolic labeling of cellular RNA: Modified nucleosides (U*) can be incorporated into RNA transcripts by ribonucleoside salvage pathway or modified nucleotides (UTP*) can be directly transfected to facilitate its incorporation by endogenous RNA polymerases. Further posttranscriptional modification has enabled the visualization and profiling of endogenous RNA transcripts.

Conrad and workers showed that N^6-propargyladenosine (N^6PA) is efficiently incorporated into RNA transcripts in mammalian cells by RNA polymerases and poly(A) polymerase. Labeling of N^6PA was used in fluorescence imaging and affinity enrichment of RNA transcripts using CuAAC reaction with appropriate azide-tags. Further, this method was utilized in studying the dynamics of poly(A) tailing of mRNA in cells by performing global pulse-chase analysis of poly(A) tail lengths.[127] In a similar approach, Jaffrey and coworkers labeled endogenous cellular RNA poly(A) tails with an alkyne modified adenosine analog, 2-ethynyladenosine (2-EA). This alkyne label was used as a poly(A) trap to pull down transcripts containing 2-EA by reacting with a biotin-conjugated azide. Next-generation sequencing of the pulled RNA transcripts enabled the stimuli-induced transcriptome-wide profiling of polyadenylation process.[128]

Interestingly, alkynes show distinctly strong Raman signal between 2100 and 2250 cm^{-1}, which falls in the Raman-silent region of the cells.[129-130] Bioconjugation of proteins, lipids and nucleic acids with this unique bioorthogonal vibrational frequency label has enabled the visualization of

these biopolymers in cells by using Raman spectroscopy.[131-132] Encouraged by the duality of the alkyne group, Srivatsan and coworkers developed 5-(1,7-octadinyl)uridine triphosphate (ODUTP) **6**, which contains a terminal alkyne suitable for click chemistry and an internal alkyne suitable for RNA spectroscopy (Figure 1.4).[133] ODUTP **6** was efficiently incorporate into RNA transcripts of various lengths by *in vitro* transcription reactions. The alkyne-labeled RNA transcripts were readily functionalized by using CuAAC reactions in the presence of various azide probes without affecting the internal alkyne. Interestingly, upon transfection, ODUTP was found to be efficiently and specifically incorporated into cellular RNA transcripts by endogenous RNA polymerases. Subsequent click staining with fluorescent azides enabled the visualization of newly transcribing RNA in mammalian cells by confocal microscopy. We expect that ODUTP containing a diyne label could potentially enable concurrent visualization of RNA in cells by click reaction and Raman spectroscopy.

The use of cytotoxic copper catalyst in CuAAC reaction to functionalize alkyne-labeled RNA transcripts hampered its implementation in live cell imaging experiments. In order to use click reaction for live cell imaging, a suitable azide based metabolic probe need to be exploited. Neef and Luedtke have recently shown that an azide-modified 2'-deoxyuridine analog can be incorporated into DNA, and by click reaction, DNA replication can be monitored by fluorescence microscopy.[134] In the meantime, we were interested in developing tools to incorporate azide functionality into cellular RNA transcripts.[77-78] We demonstrated the first example of azide labeling of cellular RNA transcripts by using 5-azidomethyl UTP (AMUTP). In our study, we found out that the nucleoside analog (AMU) was not a good substrate for metabolic labeling, and hence, transfected the cells in culture with its corresponding triphosphate (AMUTP) using a transfecting agent (DOTAP). AMUTP was found to be effectively incorporated into RNA transcripts by endogenous RNA polymerases, which enabled the imaging of newly transcribing RNA in fixed and live cells by using CuAAC and SPAAC reactions, respectively.[79]

More recently, Spitale and coworkers metabolically labeled and visualized cellular RNA using azide-modified purine analogs (**7** and **8**). Like N^{6}PA developed by Hang and Conrad, analogs **7** and **8** were found to be efficiently incorporated into RNA transcripts in cells by RNA polymerases. Live cell imaging and pull down of RNA transcripts were successfully demonstrated by using this method.[135]

1.7 Motivation and outline of the thesis

Understanding RNA structure and function in cell-free and cellular environments has greatly relied on ways to label RNA covalently. Direct incorporation of biophysical reporters into RNA, either by solid-phase or transcription has proved to be very useful over the years. However, in several instances the incompatibility of substrates to both the methods has severely restricted their applications, particularly in the context of cellular labeling. The introduction of bioorthogonal chemistry has provided newer opportunities to label RNA. Chemical biologists have wisely extended this chemistry that has been previously implemented for the bioconjugation of protein and glycan to develop practical methods to label and visualize RNA in cells and even in organisms. Recent use of reagent-free IEDDA reactions has provided new avenues of labeling exogenous RNA posttranscriptionally.[15,56] Most of these IEDDA-based approaches use relatively bulky and activated dienophiles, which may not suitable for cellular labeling and may hamper biological activity of RNA. The potential of other chemoselective reactions including palladium-mediated cross-coupling reactions have been recently used in the bioconjugation of DNA, however not been much explored in the context of RNA.[136-137] In Chapter 2, we describe the design and incorporation of a minimally perturbing vinyl-handle on RNA compatible for both IEDDA and palladium-catalyzed oxidative Heck reactions. By employing this palladium-mediated reaction, we design highly fluorogenic probes on RNA.

Transcription reaction has its drawback when site-specific labeling of RNA is needed. Site-specific chemo-enzymatic labeling methods make use of transferase-assisted approaches wherein selective 5', internal or 3' RNA labeling is possible. Both 5' and internal RNA labeling has its disadvantage especially when there is a demand for multiple labels on RNA. 3' RNA labeling employing transferase enzymes is a particularly interesting labeling approach with huge possibilities. In Chapter 3, we devise a novel chemo-enzymatic strategy to label RNA at its 3' end with a yeast terminal uridylyl transferase, SpCID1. To achieve this, we clone and express the enzyme and then control the enzyme to add a single or multiple azide labels at terminus of RNA, compatible for bioorthogonal click reactions.

In Chapter 4, we utilize the ability of SpCID1 for introducing multiple azide tags to construct high-density azide labels on CRISPR guide RNA without affecting the activity of CRISPR-Cas9 system. This novel technology is adopted for site-directed localization of azide tags onto gene locus of interest. Further, we confirm that the azide-tagged gene is compatible for

bioorthogonal click functionalization, thereby paving way for an innovative approach for localized delivery of any warhead onto the gene of interest using the combined power of CRISPR and bioorthogonal click chemistry. Subsequently, the ability of SpCID1 to introduce a single azide tag is used for constructing FRET pairs on RNA, and also for site-specific internal labeling of RNA which opens new possibilities of labeling native RNA for single molecule FRET studies.[138]

Finally, in Chapter 5, we investigate the ability of SpCID1 to introduce microenvironment sensing probes on RNA. We use this strategy for investigating possible aromatic interaction between Tyr212 in SpCID1 with terminal nucleobase of RNA and nucleotide. Further this is utilized for experimentally validating the proposed model wherein UTP binding is the first step of the catalytic cycle prior to RNA binding. Altogether, the thesis accomplishes the design of novel methods to chemoenzymatically label RNA for various applications, ranging from fluorogenic labeling to site-specific azide tagging of gene loci, constructing FRET probes and gaining fundamental insight into RNA-protein structural interactions.

1.8 References

1. Mortimer, S. A.; Kidwell, M. A.; Doudna, J. A., Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **2014,** *15* (7), 469–479.
2. Hermann, T.; Patel, D. J., RNA bulges as architectural and recognition motifs. *Structure* **2000,** *8* (3), R47–R54.
3. Shajani, Z.; Deka, P.; Varani, G., Decoding RNA motional codes. *Trends Biochem. Sci.* **2006,** *31* (8), 421–424.
4. Hall, K. B., RNA in motion. *Curr. Opin. Chem. Biol.* **2008,** *12* (6), 612-618.
5. Al-Hashimi, H. M.; Walter, N. G., RNA dynamics: It is about time. *Curr. Opin. Struct. Biol.* **2008,** *18* (3), 321–329.
6. Sinkeldam, R. W.; Greco, N. J.; Tor, Y., Fluorescent analogs of biomolecular building blocks: Design, properties, and applications. *Chem. Rev.* **2010,** *110* (5), 2579–2619.
7. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G., Fluorescent nucleoside analogs: Probes for investigating nucleic acid structure and function. *Isr. J. Chem.* **2013,** *53* (6–7), 366–378.
8. Carlomagno, T., Present and future of NMR for RNA–protein complexes: A perspective of integrated structural biology. *J Magn. Reson.* **2014,** *241*, 126–136.
9. Ke, A.; Doudna, J. A., Crystallization of RNA and RNA–protein complexes. *Methods* **2004,** *34* (3), 408–414.
10. Wachowius, F.; Höbartner, C., Chemical RNA modifications for studies of RNA structure and dynamics. *ChemBioChem* **2010,** *11* (4), 469-480.
11. Buxbaum, A. R.; Haimovich, G.; Singer, R. H., In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **2015,** *16* (2), 95–109.
12. Bao, G.; Rhee, W. J.; Tsourkas, A., Fluorescent probes for live-cell RNA detection. *Annu. Rev. Biomed. Eng.* **2009,** *11* (1), 25–47.

13. Dean, K. M.; Palmer, A. E., Advances in fluorescence labeling strategies for dynamic cellular imaging. *Nat. Chem. Biol.* **2014,** *10* (7), 512–523.
14. Blackburn, G. M.; Gait, M. J.; David, L.; Williams, D. M., *Nuclec Acids in Chemistry and Biology, third ed. .* Royal Society of Chemistry publishing: 2006.
15. Holstein, J. M.; Rentmeister, A., Current covalent modification methods for detecting RNA in fixed and living cells. *Methods* **2016,** *98*, 18–25.
16. Defrancq, E.; Singh, Y.; Spinelli, N., Chemical strategies for oligonucleotide-conjugates synthesis. *Curr. Org. Chem.* **2008,** *12* (4), 263–290.
17. Jewett, J. C.; Bertozzi, C. R., Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **2010,** *39* (4), 1272–1279.
18. Weisbrod, S. H.; Marx, A., Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem Commun.* **2008,** (44), 5675–5685.
19. Paredes, E.; Evans, M.; Das, S. R., RNA labeling, conjugation and ligation. *Methods* **2011,** *54* (2), 251–259.
20. Raj, A.; van den Bogaard, P.; Rifkin, S. A.; van Oudenaarden, A.; Tyagi, S., Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **2008,** *5* (10), 877–879.
21. Tyagi, S., Imaging intracellular RNA distribution and dynamics in living cells. *Nat. Methods* **2009,** *6* (5), 331–338.
22. 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), 437–445.
23. Buxbaum, A. R.; Wu, B.; Singer, R. H., Single beta-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* **2014,** *343* (6169), 419–422.
24. Wu, B.; Chen, J.; Singer, R. H., Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci. Rep.* **2014,** *4*, 3615.
25. Ozawa, T.; Natori, Y.; Sato, M.; Umezawa, Y., Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat. Methods* **2007,** *4*, 413.
26. Paige, J. S.; Wu, K. Y.; Jaffrey, S. R., RNA mimics of green fluorescent protein. *Science* **2011,** *333* (6042), 642–646.
27. 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), 1219–1224.
28. 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), 16299–16308.
29. Holeman, L. A.; Robinson, S. L.; Szostak, J. W.; Wilson, C., Isolation and characterization of fluorophore-binding RNA aptamers. *Fold Des.* **1998,** *3* (6), 423–431.
30. Grate, D.; Wilson, C., Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. USA* **1999,** *96* (11), 6131–6136.
31. Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J., RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.* **2014,** *9* (10), 2412–2420.
32. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* **2000,** *287* (5460), 2007–2010.
33. Prescher, J. A.; Bertozzi, C. R., Chemistry in living systems. *Nat. Chem. Biol.* **2005,** *1* (1), 13–21.

34. van Berkel, S. S.; van Eldijk, M. B.; van Hest, J. C. M., Staudinger ligation as a method for bioconjugation. *Angew. Chem. Int. Ed.* **2011,** *50* (38), 8806–8827.

35. Köhn, M.; Breinbauer, R., The Staudinger ligation—A gift to chemical biology. *Angew. Chem. Int. Ed.* **2004,** *43* (24), 3106–3116.

36. Kalia, J.; Raines, R. T., Advances in bioconjugation. *Curr. Org. Chem.* **2010,** *14* (2), 138–147.

37. Weisbrod, S. H.; Marx, A., A nucleoside triphosphate for site-specific labelling of DNA by the Staudinger ligation. *Chem. Commun.* **2007,** (18), 1828–1830.

38. El-Sagheer, A. H.; Brown, T., Click chemistry with DNA. *Chem. Soc. Rev.* **2010,** *39* (4), 1388–1405.

39. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G., Analysis and optimization of copper-catalyzed azide–alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed.* **2009,** *48* (52), 9879–9883.

40. El-Sagheer, A. H.; Brown, T., New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (35), 15329–15334.

41. Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V., Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **2004,** *6* (17), 2853–2855.

42. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted [3 + 2] azide−alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004,** *126* (46), 15046–15047.

43. Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R., In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **2008,** *320* (5876), 664–667.

44. Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A., Genetically encoded copper-free click chemistry. *Angew. Chem. Int. Ed.* **2011,** *50* (17), 3878–3881.

45. Neef, A. B.; Schultz, C., Selective fluorescence labeling of lipids in living cells. *Angew. Chem. Int. Ed.* **2009,** *48* (8), 1498–1500.

46. Jao, C. Y.; Roth, M.; Welti, R.; Salic, A., Biosynthetic labeling and two-color imaging of phospholipids in cells. *ChemBioChem* **2015,** *16* (3), 472–476.

47. Merkel, M.; Peewasan, K.; Arndt, S.; Ploschik, D.; Wagenknecht, H.-A., Copper-free postsynthetic labeling of nucleic acids by means of bioorthogonal reactions. *ChemBioChem* **2015,** *16* (11), 1541–1553.

48. Ren, X.; Gerowska, M.; El-Sagheer, A. H.; Brown, T., Enzymatic incorporation and fluorescent labelling of cyclooctyne-modified deoxyuridine triphosphates in DNA. *Biorg. Med. Chem.* **2014,** *22* (16), 4384–4390.

49. Wu, H.; Devaraj, N. K., Inverse electron-demand Diels–Alder bioorthogonal reactions. *Top. Curr. Chem.* **2015,** *374* (1), 3.

50. Schoch, J.; Wiessler, M.; Jäschke, A., Post-synthetic modification of DNA by inverse-electron-demand Diels−Alder reaction. *J. Am. Chem. Soc.* **2010,** *132* (26), 8846–8847.

51. Šečkutė, J.; Yang, J.; Devaraj, N. K., Rapid oligonucleotide-templated fluorogenic tetrazine ligations. *Nucleic Acids Res.* **2013**.

52. Rieder, U.; Luedtke, N. W., Alkene–tetrazine ligation for imaging cellular DNA. *Angew. Chem. Int. Ed.* **2014,** *53* (35), 9168–9172.

53. Bu; Batroff, E.; Niederwieser, A.; Abdel-Rahman, O. S.; Winter, R. F.; Wittmann, V.; Marx, A., Efficient labelling of enzymatically synthesized vinyl-modified DNA by an inverse-electron-demand Diels-Alder reaction. *Chem. Commun.* **2014,** *50* (74), 10827–10829.

54. Ren, X.; El-Sagheer, A. H.; Brown, T., Azide and trans-cyclooctene dUTPs: incorporation into DNA probes and fluorescent click-labelling. *Analyst* **2015,** *140* (8), 2671–2678.
55. Wang, K.; Wang, D.; Ji, K.; Chen, W.; Zheng, Y.; Dai, C.; Wang, B., Post-synthesis DNA modifications using a trans-cyclooctene click handle. *Org. Biomol. Chem.* **2015,** *13* (3), 909–915.
56. Kath-Schorr, S., Cycloadditions for studying nucleic acids. *Top. Curr. Chem.* **2016,** *374* (1), 4.
57. Sousa, R.; Chung, Y. J.; Rose, J. P.; Wang, B.-C., Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **1993,** *364* (6438), 593–599.
58. Gramlich, P. M. E.; Wirges, C. T.; Manetto, A.; Carell, T., Postsynthetic DNA modification through the copper-catalyzed azide–alkyne cycloaddition reaction. *Angew. Chem. Int. Ed.* **2008,** *47* (44), 8350–8358.
59. Seela, F.; Sirivolu, V. R., DNA containing side chains with terminal triple bonds: Base-pair stability and functionalization of alkynylated pyrimidines and 7-deazapurines. *Chem. Biodivers.* **2006,** *3* (5), 509–514.
60. Seela, F.; Sirivolu, V. R., Nucleosides and oligonucleotides with diynyl side chains: Base pairing and functionalization of 2′-deoxyuridine derivatives by the copper(I)-catalyzed alkyne–azide 'click' cycloaddition. *Helv. Chim. Acta* **2007,** *90* (3), 535–552.
61. Burley, G. A.; Gierlich, J.; Mofid, M. R.; Nir, H.; Tal, S.; Eichen, Y.; Carell, T., Directed DNA metallization. *J. Am. Chem. Soc.* **2006,** *128* (5), 1398–1399.
62. Gramlich, P. M. E.; Warncke, S.; Gierlich, J.; Carell, T., Click–click–click: Single to triple modification of DNA. *Angew. Chem. Int. Ed.* **2008,** *47* (18), 3442–3444.
63. Sirivolu, V. R.; Chittepu, P.; Seela, F., DNA with branched internal side chains: Synthesis of 5-tripropargylamine-dU and conjugation by an azide–alkyne double click Reaction. *ChemBioChem* **2008,** *9* (14), 2305–2316.
64. Kumar, R.; El-Sagheer, A.; Tumpane, J.; Lincoln, P.; Wilhelmsson, L. M.; Brown, T., Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *J. Am. Chem. Soc.* **2007,** *129* (21), 6859–6864.
65. Salic, A.; Mitchison, T. J., A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl. Acad. Sci. USA* **2008,** *105* (7), 2415–2420.
66. Neef, A. B.; Luedtke, N. W., Dynamic metabolic labeling of DNA in vivo with arabinosyl nucleosides. *Proc Natl Acad Sci U S A* **2011,** *108* (51), 20404–29409.
67. Jao, C. Y.; Salic, A., Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. USA* **2008,** *105* (41), 15779–15784.
68. Paredes, E.; Das, S. R., Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* **2011,** *12* (1), 125–131.
69. Paredes, E.; Das, S. R., Optimization of acetonitrile co-solvent and copper stoichiometry for pseudo-ligandless click chemistry with nucleic acids. *Bioorg. Med. Chem. Lett.* **2012,** *22* (16), 5313–5316.
70. Willibald, J.; Harder, J.; Sparrer, K.; Conzelmann, K.-K.; Carell, T., Click-modified anandamide siRNA enables delivery and gene silencing in neuronal and immune cells. *J. Am. Chem. Soc.* **2012,** *134* (30), 12330–12333.
71. Zheng, Y.; Beal, P. A., Synthesis and evaluation of an alkyne-modified ATP analog for enzymatic incorporation into RNA. *Bioorg. Med. Chem. Lett.* **2016,** *26* (7), 1799–1802.

72. Wada, T.; Mochizuki, A.; Higashiya, S.; Tsuruoka, H.; Kawahara, S.-i.; Ishikawa, M.; Sekine, M., Synthesis and properties of 2-azidodeoxyadenosine and its incorporation into oligodeoxynucleotides. *Tetrahedron Lett.* **2001,** *42* (52), 9215–9219.

73. Kiviniemi, A.; Virta, P.; Lönnberg, H., Utilization of intrachain 4′-C-azidomethylthymidine for preparation of oligodeoxyribonucleotide conjugates by click chemistry in solution and on a solid support. *Bioconjugate Chem.* **2008,** *19* (8), 1726–1734.

74. Pourceau, G.; Meyer, A.; Vasseur, J.-J.; Morvan, F., Azide solid support for 3′-conjugation of oligonucleotides and their circularization by click chemistry. *J. Org. Chem.* **2009,** *74* (17), 6837–6842.

75. Santner, T.; Hartl, M.; Bister, K.; Micura, R., Efficient access to 3′-terminal azide-modified RNA for inverse click-labeling Patterns. *Bioconjugate Chem.* **2014,** *25* (1), 188–195.

76. Qiu, J.; El-Sagheer, A. H.; Brown, T., Solid phase click ligation for the synthesis of very long oligonucleotides. *Chem. Commun.* **2013,** *49* (62), 6959–6961.

77. Rao, H.; Sawant, A. A.; Tanpure, A. A.; Srivatsan, S. G., Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chem. Commun.* **2012,** *48* (4), 498–500.

78. Rao, H.; Tanpure, A. A.; Sawant, A. A.; Srivatsan, S. G., Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nat. Protocols* **2012,** *7* (6), 1097–1112.

79. Sawant, A. A.; Tanpure, A. A.; Mukherjee, P. P.; Athavale, S.; Kelkar, A.; Galande, S.; Srivatsan, S. G., A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **2016,** *44* (2), e16.

80. Vaish, N. K.; Fraley, A. W.; Szostak, J. W.; McLaughlin, L. W., Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for in vitro selection of nucleic acids. *Nucleic Acids Res.* **2000,** *28* (17), 3316–3322.

81. Tolle, F.; Brändle, G. M.; Matzner, D.; Mayer, G., A versatile approach towards nucleobase-modified aptamers. *Angew. Chem. Int. Ed.* **2015,** *54* (37), 10971–10974.

82. Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S., An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat. Methods* **2006,** *3* (9), 729–735.

83. Kawai, R.; Kimoto, M.; Ikeda, S.; Mitsui, T.; Endo, M.; Yokoyama, S.; Hirao, I., Site-specific fluorescent labeling of RNA molecules by specific transcription using unnatural base pairs. *J. Am. Chem. Soc.* **2005,** *127* (49), 17286–17295.

84. Morohashi, N.; Kimoto, M.; Sato, A.; Kawai, R.; Hirao, I., Site-specific incorporation of functional components into RNA by an unnatural base pair transcription system. *Molecules* **2012,** *17* (3), 2855.

85. Ishizuka, T.; Kimoto, M.; Sato, A.; Hirao, I., Site-specific functionalization of RNA molecules by an unnatural base pair transcription system via click chemistry. *Chem. Commun.* **2012,** *48* (88), 10835–10837.

86. Someya, T.; Ando, A.; Kimoto, M.; Hirao, I., Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry. *Nucleic Acids Res.* **2015,** *43* (14), 6665–6676.

87. Samanta, A.; Krause, A.; Jaschke, A., A modified dinucleotide for site-specific RNA-labelling by transcription priming and click chemistry. *Chem. Commun.* **2014,** *50* (11), 1313–1316.

88. Domnick, C.; Eggert, F.; Kath-Schorr, S., Site-specific enzymatic introduction of a norbornene modified unnatural base into RNA and application in post-transcriptional labeling. *Chem. Commun.* **2015,** *51* (39), 8253–8256.

89. Schoch, J.; Ameta, S.; Jaschke, A., Inverse electron-demand Diels-Alder reactions for the selective and efficient labeling of RNA. *Chem. Commun.* **2011,** *47* (46), 12536–12537.

90. Ameta, S.; Becker, J.; Jäschke, A., RNA–peptide conjugate synthesis by inverse-electron demand Diels–Alder reaction. *Org. Biomol. Chem.* **2014,** *12* (26), 4701–4707.

91. Pyka, A. M.; Domnick, C.; Braun, F.; Kath-Schorr, S., Diels–Alder cycloadditions on synthetic RNA in mammalian cells. *Bioconjugate Chem.* **2014,** *25* (8), 1438–1443.

92. Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E., Optimization of an unnatural base pair toward natural-like replication. *J. Am. Chem. Soc.* **2009,** *131* (9), 3246–3252.

93. Eggert, F.; Kath-Schorr, S., A cyclopropene-modified nucleotide for site-specific RNA labeling using genetic alphabet expansion transcription. *Chem. Commun.* **2016,** *52* (45), 7284–7287.

94. Asare-Okai, P. N.; Agustin, E.; Fabris, D.; Royzen, M., Site-specific fluorescence labelling of RNA using bio-orthogonal reaction of trans-cyclooctene and tetrazine. *Chem. Commun.* **2014,** *50* (58), 7844–7847.

95. Winz, M.-L.; Samanta, A.; Benzinger, D.; Jäschke, A., Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res.* **2012,** *40* (10), e78.

96. Thomas, C.; Rusanov, T.; Hoang, T.; Augustin, T.; Kent, T.; Gaspar, I.; Pomerantz, R. T., One-step enzymatic modification of RNA 3′ termini using polymerase θ. *Nucleic Acids Res.* **2019,** *47* (7), 3272–3283.

97. Deen, J.; Vranken, C.; Leen, V.; Neely, R. K.; Janssen, K. P. F.; Hofkens, J., Methyltransferase directed labeling of biomolecules and its applications. *Angew. Chem. Int. Ed.* **2016**, *56* (19), 5182–5200.

98. Lukinavičius, G.; Lapienė, V.; Staševskij, Z.; Dalhoff, C.; Weinhold, E.; Klimašauskas, S., Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). *J. Am. Chem. Soc.* **2007,** *129* (10), 2758–2759.

99. Dalhoff, C.; Lukinavicius, G.; Klimasauskas, S.; Weinhold, E., Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol.* **2006,** *2* (1), 31–32.

100. Peters, W.; Willnow, S.; Duisken, M.; Kleine, H.; Macherey, T.; Duncan, K. E.; Litchfield, D. W.; Lüscher, B.; Weinhold, E., Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click Labeling. *Angew. Chem. Int. Ed.* **2010,** *49* (30), 5170–5173.

101. Motorin, Y.; Burhenne, J.; Teimer, R.; Koynov, K.; Willnow, S.; Weinhold, E.; Helm, M., Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic Acids Res.* **2011,** *39* (5), 1943–1952.

102. Tomkuvienė, M.; Clouet-d'Orval, B.; Černiauskas, I.; Weinhold, E.; Klimašauskas, S., Programmable sequence-specific click-labeling of RNA using archaeal box C/D RNP methyltransferases. *Nucleic Acids Res.* **2012,** *40* (14), 6765–6773.

103. Plotnikova, A.; Osipenko, A.; Masevičius, V.; Vilkaitis, G.; Klimašauskas, S., Selective Covalent labeling of miRNA and siRNA duplexes using HEN1 methyltransferase. *J. Am. Chem. Soc.* **2014,** *136* (39), 13550–13553.

104. Schulz, D.; Holstein, J. M.; Rentmeister, A., A chemo-enzymatic approach for site-specific modification of the RNA cap. *Angew. Chem. Int. Ed.* **2013,** *52* (30), 7874–7878.

105. Holstein, J. M.; Schulz, D.; Rentmeister, A., Bioorthogonal site-specific labeling of the 5'-cap structure in eukaryotic mRNAs. *Chem. Commun.* **2014,** *50* (34), 4478–4481.

106. Holstein, J. M.; Stummer, D.; Rentmeister, A., Enzymatic modification of 5'-capped RNA with a 4-vinylbenzyl group provides a platform for photoclick and inverse electron-demand Diels-Alder reaction. *Chem. Sci.* **2015,** *6* (2), 1362–1369.

107. Holstein, J. M.; Anhäuser, L.; Rentmeister, A., Modifying the 5'-cap for click reactions of eukaryotic mRNA and to tune translation efficiency in living cells. *Angew. Chem. Int. Ed.* **2016,** *55* (36), 10899–10903.

108. Ikeuchi, Y.; Kimura, S.; Numata, T.; Nakamura, D.; Yokogawa, T.; Ogata, T.; Wada, T.; Suzuki, T.; Suzuki, T., Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. *Nat. Chem. Biol.* **2010,** *6* (4), 277–282.

109. Li, F.; Dong, J.; Hu, X.; Gong, W.; Li, J.; Shen, J.; Tian, H.; Wang, J., A covalent approach for site-specific RNA labeling in mammalian cells. *Angew. Chem. Int. Ed.* **2015,** *54* (15), 4597–4602.

110. Wilson, C.; Szostak, J. W., In vitro evolution of a self-alkylatlng ribozyme. *Nature* **1995,** *374* (6525), 777–782.

111. Sharma, A. K.; Plant, J. J.; Rangel, A. E.; Meek, K. N.; Anamisis, A. J.; Hollien, J.; Heemstra, J. M., Fluorescent RNA labeling using self-alkylating ribozymes. *ACS Chem. Biol.* **2014,** *9* (8), 1680–1684.

112. McDonald, R. I.; Guilinger, J. P.; Mukherji, S.; Curtis, E. A.; Lee, W. I.; Liu, D. R., Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nat. Chem. Biol.* **2014,** *10* (12), 1049–1054.

113. Vauléon, S.; Ivanov, S. A.; Gwiazda, S.; Müller, S., Site-specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme. *ChemBioChem* **2005,** *6* (12), 2158–2162.

114. Büttner, L.; Javadi-Zarnaghi, F.; Höbartner, C., Site-specific labeling of RNA at internal ribose hydroxyl groups: Terbium-assisted deoxyribozymes at work. *J. Am. Chem. Soc.* **2014,** *136* (22), 8131–8137.

115. Anhäuser, L.; Hüwel, S.; Zobel, T.; Rentmeister, A., Multiple covalent fluorescence labeling of eukaryotic mRNA at the poly(A) tail enhances translation and can be performed in living cells. *Nucleic Acids Res.* **2019,** *47* (7), e42–e42.

116. Haukenes, G.; Szilvay, A. M.; Brokstad, K. A.; Kanestrøm, A.; Kalland, K. H., Labeling of RNA transcripts of eukaryotic cells in culture with BrUTP using a liposome transfection reagent (DOTAP). *Biotechniques* **1997,** *22* (2), 308–312.

117. Cmarko, D.; Verschure, P. J.; Martin, T. E.; Dahmus, M. E.; Krause, S.; Fu, X.-D.; van Driel, R.; Fakan, S., Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol. Biol. Cell* **1999,** *10* (1), 211–223.

118. Paulsen, M. T.; Veloso, A.; Prasad, J.; Bedi, K.; Ljungman, E. A.; Tsan, Y.-C.; Chang, C.-W.; Tarrier, B.; Washburn, J. G.; Lyons, R.; Robinson, D. R.; Kumar-Sinha, C.; Wilson, T. E.; Ljungman, M., Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (6), 2240–2245.

119. Paulsen, M. T.; Veloso, A.; Prasad, J.; Bedi, K.; Ljungman, E. A.; Magnuson, B.; Wilson, T. E.; Ljungman, M., Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods (San Diego, Calif.)* **2014,** *67* (1), 45–54.

120. Tani, H.; Akimitsu, N., Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling. *RNA Biol.* **2012,** *9* (10), 1233–1238.

121. Rabani, M.; Levin, J. Z.; Fan, L.; Adiconis, X.; Raychowdhury, R.; Garber, M.; Gnirke, A.; Nusbaum, C.; Hacohen, N.; Friedman, N.; Amit, I.; Regev, A., Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotech.* **2011,** *29* (5), 436–442.

122. Radle, B.; Rutkowski, A. J.; Ruzsics, Z.; Friedel, C. C.; Koszinowski, U. H.; Dolken, L., Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *J. Vis. Exp.* **2013,** (78), e50195.

123. Rutkowski, A. J.; Erhard, F.; L'Hernault, A.; Bonfert, T.; Schilhabel, M.; Crump, C.; Rosenstiel, P.; Efstathiou, S.; Zimmer, R.; Friedel, C. C.; Dölken, L., Widespread disruption of host transcription termination in HSV-1 infection. *Nature Commun.* **2015,** *6*, 7126.

124. Hafner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano Jr, M.; 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), 129–141.

125. Burger, K.; Mühl, B.; Kellner, M.; Rohrmoser, M.; Gruber-Eber, A.; Windhager, L.; Friedel, C. C.; Dölken, L.; Eick, D., 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol.* **2013,** *10* (10), 1623–1630.

126. Qu, D.; Zhou, L.; Wang, W.; Wang, Z.; Wang, G.; Chi, W.; Zhang, B., 5-Ethynylcytidine as a new agent for detecting RNA synthesis in live cells by "click" chemistry. *Anal. Biochem.* **2013,** *434* (1), 128–135.

127. Grammel, M.; Hang, H.; Conrad, N. K., Chemical reporters for monitoring RNA synthesis and poly(A) tail dynamics. *ChemBioChem* **2012,** *13* (8), 1112-1115.

128. Curanovic, D.; Cohen, M.; Singh, I.; Slagle, C. E.; Leslie, C. S.; Jaffrey, S. R., Global profiling of stimulus-induced polyadenylation in cells using a poly(A) trap. *Nat. Chem. Biol.* **2013,** *9* (11), 671–673.

129. Yamakoshi, H.; Dodo, K.; Palonpon, A.; Ando, J.; Fujita, K.; Kawata, S.; Sodeoka, M., Alkyne-tag raman imaging for visualization of mobile small molecules in live cells. *J. Am. Chem. Soc.* **2012,** *134* (51), 20681–20689.

130. Hong, S.; Chen, T.; Zhu, Y.; Li, A.; Huang, Y.; Chen, X., Live-cell stimulated raman scattering imaging of alkyne-tagged biomolecules. *Angew. Chem. Int. Ed.* **2014,** *53* (23), 5827–5831.

131. Wei, L.; Hu, F.; Shen, Y.; Chen, Z.; Yu, Y.; Lin, C.-C.; Wang, M. C.; Min, W., Live-cell imaging of alkyne-tagged small biomolecules by stimulated Raman scattering. *Nat. Methods* **2014,** *11* (4), 410–412.

132. Chen, Z.; Paley, D. W.; Wei, L.; Weisman, A. L.; Friesner, R. A.; Nuckolls, C.; Min, W., Multicolor live-cell chemical imaging by isotopically edited alkyne vibrational palette. *J. Am. Chem. Soc.* **2014,** *136* (22), 8027–8033.

133. Sawant, A. A.; Mukherjee, P. P.; Jangid, R. K.; Galande, S.; Srivatsan, S. G., A clickable UTP analog for the posttranscriptional chemical labeling and imaging of RNA. *Org. Biomol. Chem.* **2016,** *14* (24), 5832–5842.

134. Neef, A. B.; Luedtke, N. W., An azide-modified nucleoside for metabolic labeling of DNA. *ChemBioChem* **2014,** *15* (6), 789–793.

135. Nainar, S.; Beasley, S.; Fazio, M.; Kubota, M.; Dai, N.; Correa, I. R., Jr.; Spitale, R. C., Metabolic incorporation of azide functionality into cellular RNA. *Chembiochem* **2016,** *17* (22), 2149–2152.

136. Omumi, A.; Beach, D. G.; Baker, M.; Gabryelski, W.; Manderville, R. A., Postsynthetic guanine arylation of DNA by Suzuki−Miyaura cross-coupling. *J. Am. Chem. Soc.* **2011,** *133* (1), 42–50.

137. Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G., DNA modification under mild conditions by Suzuki–Miyaura cross-coupling for the generation of functional probes. *Angew. Chem. Int. Ed.* **2013,** *52* (40), 10553–10558.

138. Rinaldi, A. J.; Suddala, K. C.; Walter, N. G., Native purification and labeling of RNA for single molecule fluorescence studies. *Methods Mol. Biol.* **2015,** *1240*, 63–95.

Chapter 2

Vinyluridine as a Versatile Chemoselective Handle for the Posttranscriptional Chemical Functionalization of RNA[*]

2.1 Introduction

Functional modification of RNA *in vitro* and in cells have become necessary for understanding the regulation, structure, and function of RNA.[1-3] While solid-phase oligonucleotide (ON) synthesis and enzymatic labeling methods using bacteriophage RNA polymerases are widely used in the generation of labeled RNA suitable for cell-free and in-cell analysis, analogous approaches to label endogenous RNA are less prevalent.[4] Immunostaining of metabolically incorporated nucleoside analogs (e.g., BrU) using antibodies, structure-specific antibodies (e.g., G-quadruplexes), aptamers (e.g., spinach), and dyes have been employed to visualize cellular RNA structure and synthesis.[5-10] However, limited permeability and selectivity of the antibodies and compromised function of aptamer-tagged RNA have been the downsides of these methods. Alternatively, bio-orthogonal chemical reactions provide an easy route to label RNA *in vitro* and cellular milieu for a variety of applications.[11-15] In this approach, a reactive group is introduced into an RNA ON by either chemical or chemo-enzymatic (RNA polymerases, transferases) means, and further bioconjugation to label RNA is achieved by performing a chemoselective reaction with a cognate reactive partner containing a desired biophysical reporter. Several chemoselective reactions including azide-alkyne cycloaddition (AAC), Staudinger ligation, and inverse electron demand Diels-Alder (IEDDA) reactions have been effectively utilized in labeling protein, glycan, lipid, and DNA.[16-18] However, establishing RNA-labeling techniques under the conditions used in these reactions has always been a challenge due to the inherent instability of the RNA.[15]

Among the various bioorthogonal reactions, AAC reaction has been extensively used in labeling RNA postsynthetically.[19-27] Typically, in this reaction, a copper(I) stabilizing ligand is used to alleviate the toxic effects of copper ions, which is known to produce species harmful to nucleic acids in the redox environment.[14,28,29] In another strategy, the use of toxic copper has been

---

circumvented by employing strain-promoted AAC (SPAAC) reaction, wherein strained alkynes (e.g., cyclooctynes) are used as one of the reactive counterparts.[30-34]

However, SPAAC reaction efficiency in labeling nucleic acids in cells has been limited by the poor permeability and reactivity of bulky cyclooctyne probes.[33,34] Other concerns that have limited the application of AAC reactions are the involvement of alkyne substrates in side reactions such as homocoupling and thiol-yne addition and low stability of azide substrates in the chemical-labeling conditions.[33,35,36]

Analogous to azide and alkyne groups, the ability of alkene group to participate in a wide range of chemoselective reactions has been elegantly utilized in devising several bioconjugation strategies. Bioconjugation methods based on ene-thiol and ene-tetrazole reactions have been put to good use in cross-linking and functionalizing biomolecules, especially proteins.[37-39] However, the use of UV radiation and nonspecific reaction with cellular thiols has hampered their applications in in-cell analysis. Recently, protein labeling has been achieved by using oxidative Heck reaction between protein-bound alkenes and boronic acid reporters.[40-42] This reaction is particularly interesting because it can be effectively performed in aqueous buffer under near physiological conditions (~pH 7, room temperature and oxygen atmosphere) as compared to other analogous Pd-assisted biomolecular labeling reactions.[43,44] Very recently, base-modified 2′-deoxynucleosides, containing a minimally perturbing vinyl group as a dienophile, have been successfully incorporated *in vitro* into DNA ONs and metabolically into replicating DNA in cells.[45,46] Although the reaction rates are significantly lower as compared to strained dienophiles, vinyl-labeled DNA did enable the labeling as well as the visualization of DNA in cell-free and cellular environments by using IEDDA reactions.

Coupling reactions have been employed to attach heterocyclic aryl moieties via an alkene to natural nucleobases. Such conjugations have produced microenvironment-sensitive nucleoside probes, which have been used in studying the conformation of the G-quadruplex structure.[47] Similarly, fluorogenic reporters have been developed to specifically detect epigenetic and epitranscriptomic nucleobase modifications (e.g., 5-formylcytosine and 5-formyluracil) by using a chemoselective reaction between the base and trimethylindole derivatives.[48] Taking advantage of this knowledge, we rationalized that enzymatic incorporation of a minimally perturbing vinyl-modified ribonucleotide into RNA would facilitate the generation of microenvironment-sensitive fluoro- genic RNA reporters by post-transcriptional oxidative Heck reaction with appropriate

boronic acid derivatives (Figure 2.1). Also, we envisioned that vinyl-labeled RNA transcript would open up possibilities for investigating its potential to functionalize RNA by reagentless IEDDA reaction. Despite the successes with protein and DNA,[40,49] the utility of vinyl group as a versatile chemoselective handle in developing RNA-labeling protocols has not been well-explored.[50] In this context, here we describe the synthesis of 5-vinyl-modified UTP (VUTP) analog and its effective incorporation into RNA ONs by *in vitro* transcription reactions. Furthermore, we illustrate the post-transcriptional chemical functionalization of vinyl-labeled RNA transcripts with a variety of biophysical reporters by using oxidative Heck and IEDDA reactions. Notably, this is the first report on the bioconjugation of an RNA ON by employing Pd-mediated oxidative Heck reactions.
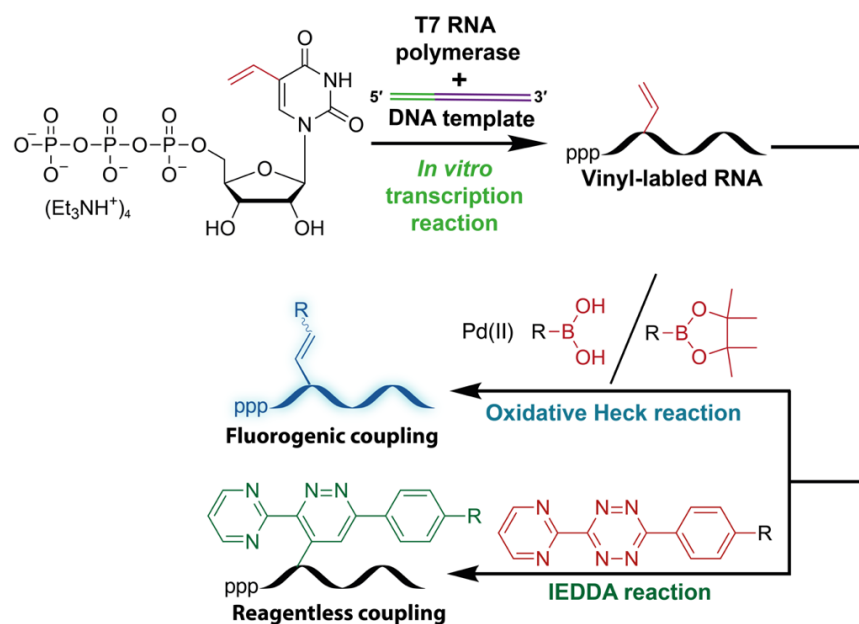
Figure 2.1. Diagram illustrating the post-transcriptional chemical labeling of RNA by using oxidative Heck and IEDDA reactions. Incorporation of VUTP into RNA transcripts, followed by oxidative Heck reaction with boronic acid derivatives and IEDDA reaction with tetrazine derivatives, enable the synthesis of RNA conjugated to various reporters and tags.

2.2 Results and Discussion

2.2.1 Synthesis and enzymatic incorporation of 5-vinyl uridine triphosphate

VUTP **3** was synthesized in simple steps starting from 5-iodouridine 1 (Scheme 2.1). 5-Vinyluridine (VU) **2** was synthesized by Stille cross-coupling reaction using vinyltributylstannane and a palladium catalyst, tris-(dibenzylideneacetone)dipalladium(0).[51] VU was further phosphorylated in the presence of POCl$_3$ and bis-tributylammonium pyrophosphate to afford VUTP.

Scheme 2.1. Synthesis of VUTP **3**. (**a**) Pd$_2$(dba)$_3$, P(fur)$_3$, vinyltributylstannane, DMF, 60 °C, 3 h, 53 %. (**b**) (i) POCl$_3$, proton sponge, trimethylphosphate, ~4 °C, 30 min, (ii) *bis*-tributylammonium pyrophosphate, tributyl amine, ~4 °C, 30 min, 21 %.

The ability of VUTP to serve as a UTP analog for RNA polymerase was studied by performing *in vitro* transcription reactions in the presence of a small series of DNA templates (Figure 2.2). The templates contained one or two dA residues in the coding region to guide the insertion of monophosphate of VUTP into RNA during the polymerization reaction. The 5′-end of each template ended with a dT residue to introduce a radiolabeled α-^{32}P adenosine at the 3′-end of the full-length transcript. Transcription was performed using GTP, CTP, UTP/3, and α-^{32}P ATP in the presence of T7 RNA polymerase. The transcription products were then electrophoresed on an analytical polyacrylamide gel and phosphorimaged.

Figure 2.2. Incorporation of the vinyl group into RNA ONs by in vitro transcription reaction using DNA templates **S1**−**S4** and VUTP **3**. "r" preceding the sequence in parentheses represents RNA ON transcripts.

The incorporation of VUTP **3** in the presence of template **S1** proceeded with very good efficiency at the +7 position, yielding a 10 mer full-length RNA transcript **4** (Figure 2.3, lane 2). Slower migration of the transcription product **4** in lane 2 hinted at the incorporation of **3** into the transcript (compare lanes 1 and 2). While a control reaction in the absence of UTP and **3** did not yield full-length RNA transcript, at equimolar concentrations, both natural UTP and **3** got incorporated into RNA (Figure 2.3 lanes 3 and 4, respectively). Templates **S2**−**S4** incorporated

modified UTP near the promoter region and at more than one site with good to excellent efficiency (Figure 2.3 lanes 6, 8, and 10).
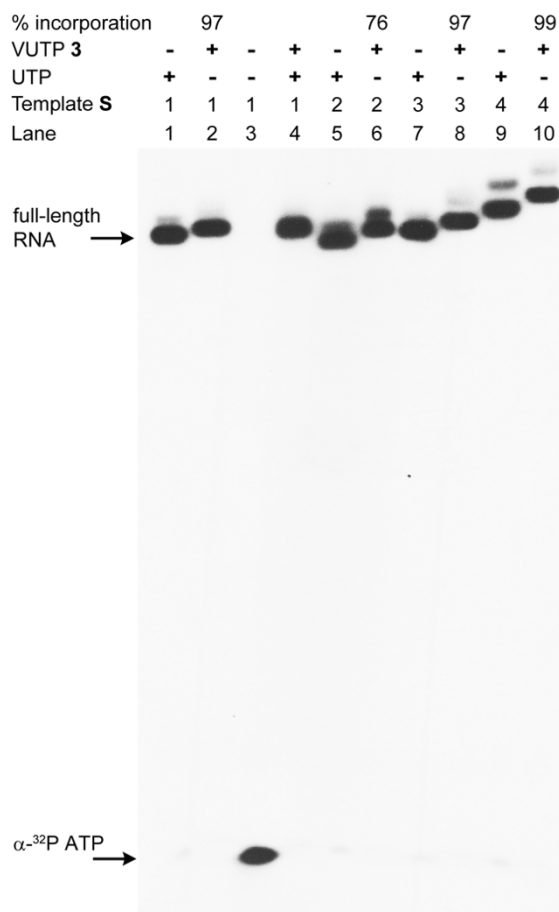
Figure 2.3. Phosphor image of transcripts obtained by transcription of templates **S1−S4** in the presence of UTP and **3**. "The percent of incorporation of **3** is reported relative to a control transcription with UTP.

Large-scale synthesis and characterisation

Transcript **4** was scaled up by performing large-scale reactions with **S1** in the presence of cold NTPs. Typically, a reaction in the presence of 75 pmol of the template gave nearly 15 nmol of the modified transcript after gel electrophoretic purification, which was further used for post-transcriptional chemical labeling experiments. The presence of VU in transcript **4** was checked by mass analysis (Figure 2.4A). Further, we characterized the transcript using enzymatic digestion, wherein we incubated it with RNase A, calf intestinal alkaline phosphatase, snake venom phosphodiesterase I, RNase T1 for 12 h at 37 °C. Products of digestion were analysed on HPLC against a standard set of natural nucleosides (rA, rU, rG, rC) and **2** (Figure 2.4B) and peaks

corresponding to each natural nucleoside and **2** was analysed by HRMS analysis which confirmed the presence of VU in the transcript (Table 2.1).
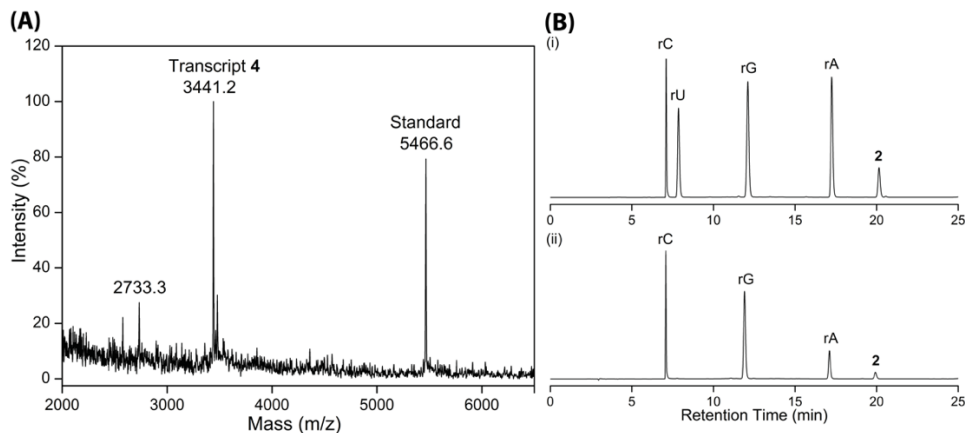
Figure 2.4. (**A**) MALDI-TOF mass spectrum of modified transcript **4** from large-scale transcription reaction with template **S1**. Spectrum is calibrated with respect to the +1 and +2 ion of an internal 18-mer DNA oligonucleotide standard (m/z for +1 and +2 ion are 5466.6 and 2733.3, respectively). (**B**) HPLC profile of ribonucleoside products obtained after enzymatic digestion of RNA ON **4** (260 nm). (i) Natural ribonucleosides (rA, rU, rG, rC) and vinyl modified ribonucleoside **2** mix. (ii) Transcript **4** digest.

Table 2.1. HRMS data for HPLC fractions of transcript **4** digest

HPLC fractions of the digest	Calcd.	Found
rC	$C_9H_{12}N_3O_5$: 242.0777 [M-H]$^-$	242.0797
rG	$C_{10}H_{12}N_5O_5$: 282.0838 [M-H]$^-$	282.0850
rA	$C_{10}H_{12}N_5O_4$: 266.0889 [M-H]$^-$	266.0898
2	$C_{11}H_{13}N_2O_6$: 269.0774 [M-H]$^-$	269.0784

2.2.2 Oxidative Heck reaction on vinyl-labeled RNA transcript

Electronically unbiased terminal alkene (e.g., allylic) substrates have been reported to undergo oxidative Heck reaction with boronic acids in the presence of water-soluble Pd-EDTA complex.[41] However, the feasibility of such a reaction on RNA has not been explored. In this regard, we sought to investigate the possibility of performing oxidative Heck reaction on vinyl-modified RNA with suitable boronic acid and ester substrates. For this purpose we chose a range of heterobicycles, namely benzothiophene-2- (**5**), benzofuran-2- (**6**), indole-5- (**7**), dibenzothiophene-4-boronic acid, (**8**) and a previously reported pinacol boronic ester, benzothiophene-2-vinyl boronic ester (**9**),[47]

which, upon conjugation to a uracil ring, could impart fluorescence to otherwise nonemissive nucleobase (Figure 2.5). Fluorescent nucleosides have been developed by attaching heterocycles to pyrimidine and purine bases, respectively, and such analogs incorporated into ONs have been used as probes in various nucleic acid studies.[2,52-57]
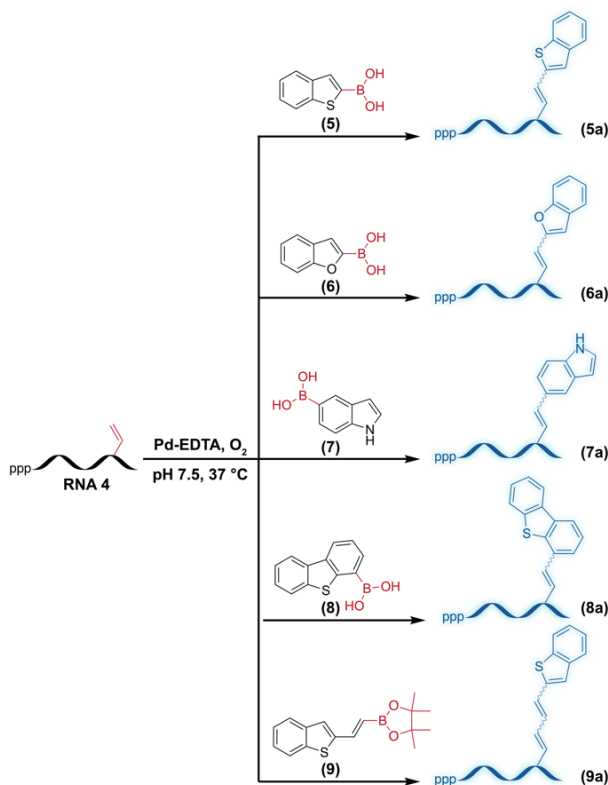
Figure 2.5. Oxidative Heck reaction between vinyl-modified RNA ON **4** and boronic acid substrates **5-8** or boronic ester **9** in the presence of Pd-EDTA complex.

For performing oxidative Heck reaction on RNA, transcript **4** was incubated with boronic acid **5** in the presence of Pd-EDTA complex at 37 °C. Aliquots of reaction mixture at various time intervals were resolved by PAGE under denaturing conditions. Almost complete conversion was observed at 18 h, and rewardingly, we observed an intense fluorescent band corresponding to the coupled product along with a minor band (Figure 2.6). Reactions with other substrates (**6-8**) also proceeded well (Figure 2.7). Further, products from large-scale reactions between transcript **4** and boronic acid **5-8** or boronic ester **9** were analyzed by reverse-phase high-performance liquid chromatography (RP-HPLC) at 260 (Figure 2.8) and 338 nm (conjugation of heterocycles to pyrimidine bases results in a strong absorption band near 338 nm).[2,58] Mass analysis of HPLC fractions corresponding to both 260 and 338 nm absorption bands revealed the formation of the

oxidative Heck products **5a-9a** (Table 2.2). Reactions performed at a 5 nmol scale of the RNA ON

4 provided 0.5 to 1.1 nmol of the coupled products (Table 2.3).
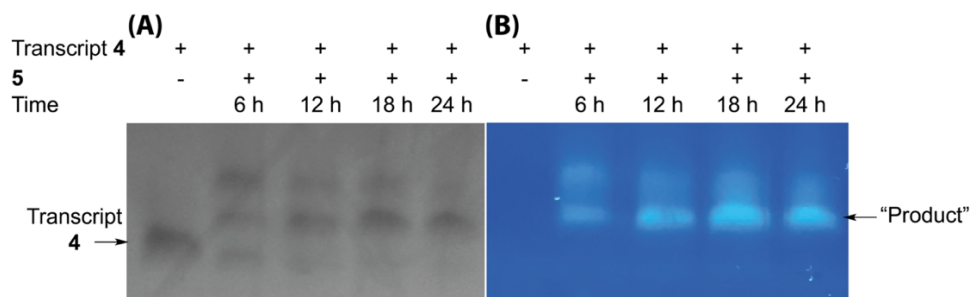
Figure 2.6. Oxidative Heck reaction between vinyl-modified RNA ON **4** and boronic acid substrate **5**. (**A**) UV-shadow image of the gel at 254 nm. (**B**) UV-transilluminator image of the gel at 364 nm.
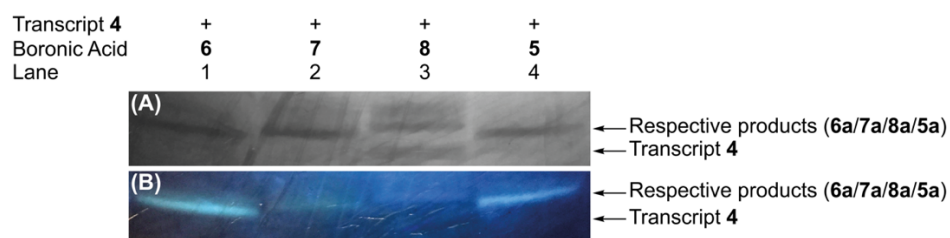
Figure 2.7. Oxidative Heck reaction between vinyl-modified RNA ON **4** and boronic acid substrate **5-8**. Products observed **6a** (lane 1), **7a** (lane 2), **8a** (lane 3) and **5a** (lane 4) was visualized on gel by (**A**) UV-shadow image at 254 nm and by (**B**) UV-transilluminator image at 364 nm.
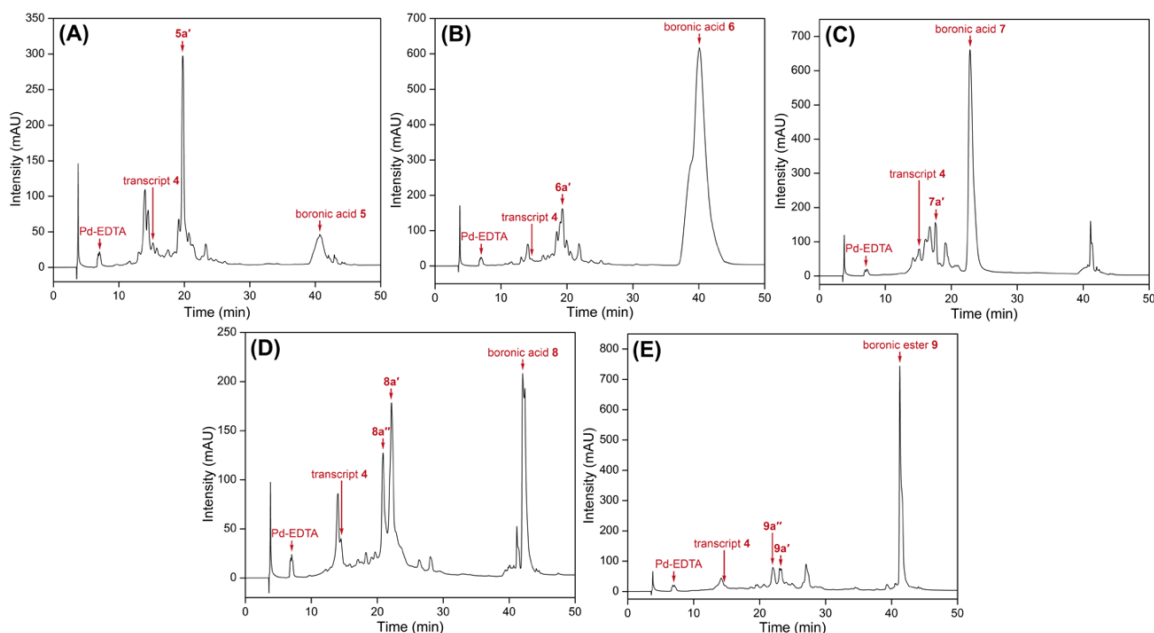
Figure 2.8. HPLC chromatogram of the reaction mixture of oxidative Heck reaction between transcript **4** and boronic acids **5-8** or boronic ester **9**. Reaction with **5** (**A**), **6** (**B**), **7** (**C**), **8** (**D**) and **9** (**E**). HPLC fractions designated as **5a′**, **6a′**, **7a′**, **8a′**, **8a″**, **9′** and **9a″** are the Heck-coupled RNA products obtained for substrates **5-9**, monitored at 260 nm wavelength.

Table 2.2. MALDI-TOF mass of oxidative Heck and IEDDA products[a]

Product	Calcd.	Found
5a′	3573.1	3572.5
6a′	3557.1	3557.3
7a′	3556.1	3556.4
8a′	3623.2	3624.2
8a′′	3623.2	3624.3
9a′	3599.2	3599.1
9a′′	3599.2	3599.1
10a	3647.2	3647.6
11a	3959.5	3959.7
12a	4358.0	4358.6

[a]Mass spectra in appendix-I

Table 2.3. Yields of oxidative Heck and IEDDA reactions between RNA transcript **4** and substrates **5-12**

Substrate[a]	Product	Isolated Yield (nmoles)	Isolated Yield (%)
5	5a′	1.0	20
6	6a′	0.6	12
7	7a′	0.5	10
8	8a′ + 8a′′	1.1	22
9	9a′ + 9a′′	0.6	12
10	10a	7.7	51
11	11a	4.7	31
12	12a	6.4	43

[a]Reactions with substrates **5-9** where performed on 5 nmol of RNA **4** (oxidative Heck reaction) and substrates **10-12** where performed on 15 nmol of RNA **4** (IEDDA reaction).

2.2.3 Regioisomers of oxidative Heck reaction

An oxidative Heck reaction between an electronically unbiased terminal alkene (e.g., allylic system) and arylboronic acid substrates typically gives linear and branched cross-coupled products.[40] However, reactions with electronically biased aryl vinyl systems (e.g., styrene) give predominantly linear trans-coupled product.[59,60] Formation of branched regioisomer product has also been reported using Pd in the presence of certain bulky ligands (e. g., 2,9-dimethyl-1,10-phenanthroline).[60] Although it is not a major concern in most bioconjugation strategies, we sought to determine the formation of different isomers in the oxidative Heck reaction of vinyl-labeled RNA transcript. First, we carried out a reaction between free vinyl-modified uridine **2** and boronic acid **5** under similar conditions used for RNA ligation. Interestingly, at the ribonucleoside level, a very low conversion was observed even after prolonged incubation time. Varying the Pd-ligand complex and oxidant did not give a better yield (data not shown). However, a fluorescent product was isolated in analytical quantities, and further characterization by NMR revealed the formation of a linear trans-isomer product **2a** (Figure 2.9A). Further, the coupled product gave a unique longer wavelength absorption band at ~338 nm as compared to native nucleoside and nucleic acid, which absorbs at 260 nm (Figure 2.9B). The coupled product **2a** also exhibited good fluorescence upon exciting at 338 nm (Figure 2.9C). Next, the major HPLC fraction corresponding to the RNA products (designated as **5a′**) of a reaction between vinyl-labeled RNA ON **4** and boronic acid **5** was subjected to enzymatic digestion in the presence of calf intestinal alkaline phosphatase, RNase A, RNase T1 and snake venom phosphodiesterase. Enzymatic digestion of **5a′** gave individual ribonucleosides as well as Heck-coupled ribonucleosides, whose identity was determined by comparing the HPLC chromatogram of the digested sample and oxidative Heck reaction product **2a** obtained from a reaction between uridine **2** and boronic acid **5** (Figure 2.10). The chromatogram of the digest revealed the presence of native ribonucleosides (rC, rG, and rA) and two peaks corresponding to coupled products absorbing at 338 nm. One of the peaks clearly matched with the trans-isomer product **2a**, whereas the chemical structure of the peak X could not be determined. However, both **2a** and **X** gave the same mass, indicating that a mixture of isomers (**X** may be the cis isomer) are formed in these reactions (Table 2.4). Formation of a mixture of isomers, which is an inherent "limitation" of this reaction, has also been observed in protein ligation experiments.[40,41] Although, at the present, this reaction may not be viable for introducing a specific type of

conformation-sensitive probe into RNA, development in chelate-control strategies could vastly improve the selective of the reaction.[60]
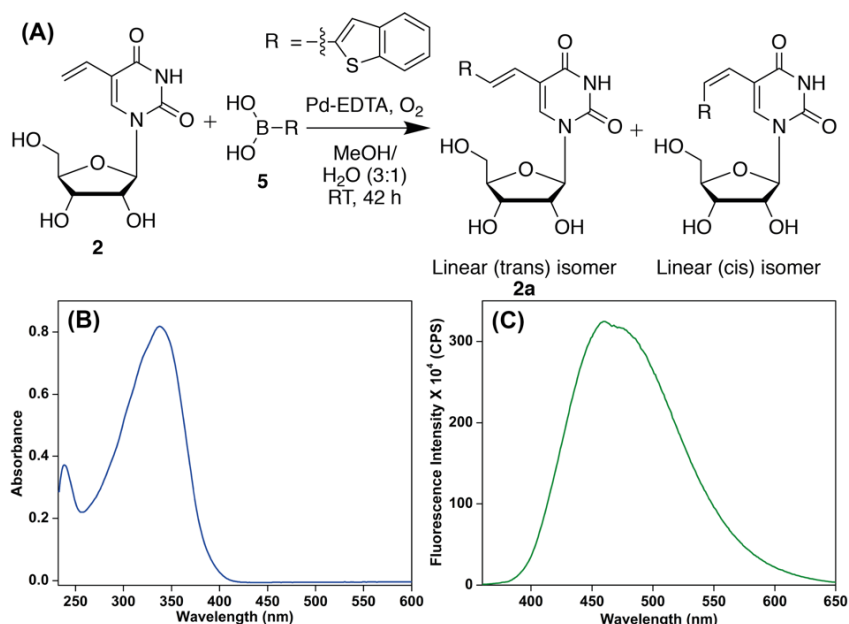
Figure 2.9. (**A**) Possible products of oxidative Heck reaction between modified nucleoside **2** and boronic acid **5**.[40] Linear (trans) isomer **2a** was isolated and characterized, while cis product was not isolable. (**B**) Absorption spectrum of product **2a** showing an intense band at 338 nm. (**C**) Fluorescence spectrum of **2a**. Sample was excited at 338 nm with excitation and emission slit width of 5 and 5 nm, respectively.
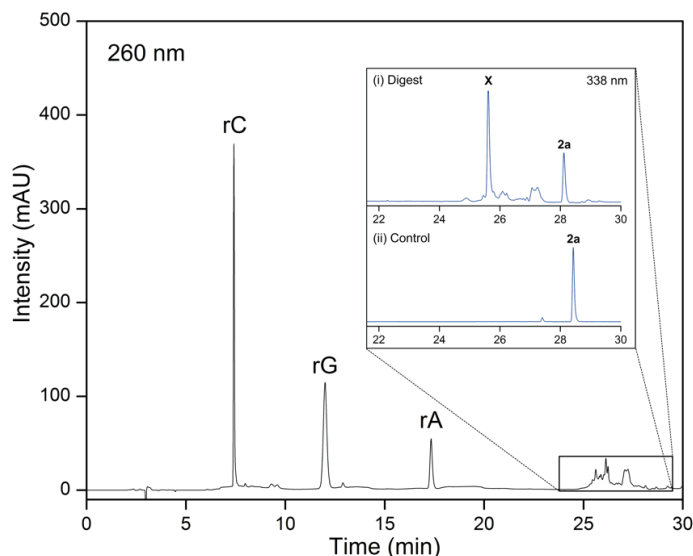
Figure 2.10. HPLC profile of ribonucleoside products obtained after enzymatic digestion of RNA **5a** (HPLC fraction **5a′**) at 260 nm. Inset (i) Chromatogram of enzymatic digest of **5a′** at 338 nm (zoom) (ii) Chromatogram of control ribonucleoside **2a** at 338 nm (zoom). Coupled products showed intense absorption at 338 nm as compared to 260 nm where natural ribonucleosides (rC, rG and rA) show maximum absorbance. **X**: possibly linear alkene (cis-isomer) or branched alkene; **2a**: linear alkene (trans-isomer).[40] Integrity of the peaks corresponding to **X** and **2a** isolated, where confirmed by HRMS analysis (Table 2.4).

Table 2.4. HRMS data for HPLC fractions of transcript **5a′** digest

HPLC fractions of the digest	Calcd.	Found
rC	$C_9H_{12}N_3O_5$: 242.0777 [M-H]$^-$	242.0783
rG	$C_{10}H_{12}N_5O_5$: 282.0838 [M-H]$^-$	282.0834
rA	$C_{10}H_{12}N_5O_4$: 266.0889 [M-H]$^-$	266.0895
2a	$C_{19}H_{17}N_2O_6S$: 401.0807 [M-H]$^-$	401.0766
X	$C_{19}H_{17}N_2O_6S$: 401.0807 [M-H]$^-$	401.0732

2.2.4 Fluorogenic coupling using oxidative Heck reaction

We had observed previously that the reaction mixture turned highly fluorescent after reaction as compared to before reaction when the vial was shined with UV-transilluminator at a wavelength of 364 nm (Figure 2.11). Therefore the fluorogenic nature of the RNA products were investigated by steady-state fluorescence spectroscopy. Benzothiophene-coupled RNA product **5a** (HPLC fraction corresponding to **5a′**) showed a ~40-fold enhancement in fluorescence intensity as compared to the boronic acid substrate **5** (Figures 2.11 and 2.12A). Interestingly, benzothiophene-alkene-coupled RNA product **9a** (HPLC fraction corresponding to **9a′**) showed a remarkable ~170-fold enhancement in fluorescence intensity as compared to substrate **9**. Furthermore, the emission profile of **9a** showed a significant bathochromic shift compared to RNA **5a** as a direct result of increased conjugation between the uracil and benzothiophene rings (Figures 2.5 and 2.12E). Reactions with benzofuran (**6**), indole (**7**), and dibenzothiophene (**8**) boronic acid substrates also produced fluorescent RNA products (Figure 2.12B-D). While benzofuran- and dibenzothio-phene-conjugated RNA products were reasonably fluorescent, indole-conjugated RNA was found to be weakly fluorescent. Taken together, this approach of functionalizing vinyl-labeled RNA transcripts by fluorogenic Heck-type coupling reaction with appropriate boronic acid/boronic ester substrates could provide direct access to RNA functionalized with fluorescent reporters.[48,61,62]
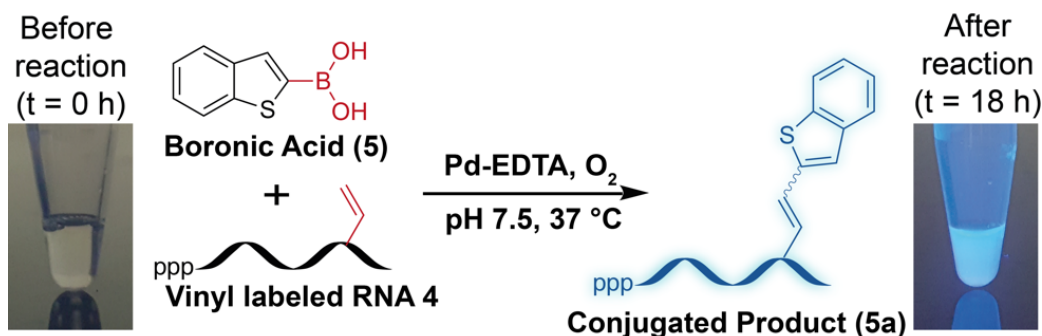
Figure 2.11. Oxidative Heck coupling between transcript **4** and boronic acid **5** as visualized in a vial by exciting using a UV-transilluminator at 364 nm before and after reaction.
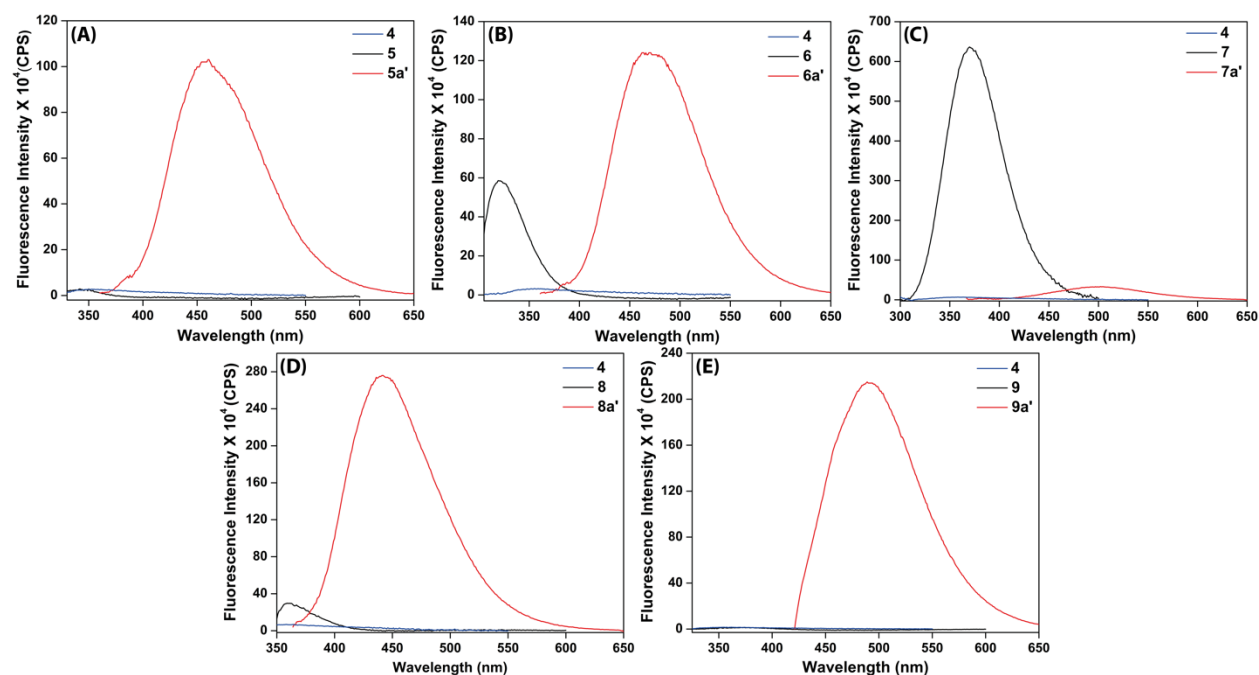
Figure 2.12. Fluorescence spectra (0.2 µM) of RNA transcript **4**, boronic acid/ester (black line) and RNA product (red line). RNA transcript **4** was excited at 289 nm. (**A**) **5**, **5a′** (excited at 310 nm and 340 nm with excitation and emission slit width of 8 nm and 9 nm, respectively), (**B**) **6**, **6a′** (excited at 291 nm and 340 nm with excitation and emission slit width of 9 nm and 10 nm, respectively), (**C**) **7**, **7a′** (excited at 273 nm and 340 nm with excitation and emission slit width of 12 nm and 14 nm, respectively) (**D**) **8**, **8a′** (excited at 328 nm and 340 nm with excitation and emission slit width of 12 nm and 14 nm, respectively), (**E**) **5**, **5a′** (excited at 310 nm and 372 nm with excitation and emission slit width of 7 nm and 8 nm, respectively).

2.2.5 IEDDA reaction on vinyl-labeled RNA transcript

Analogous to AAC chemistry, IEDDA reaction between an electron-rich dienophile and electron-deficient tetrazine is gaining particular attention as a useful biomacromolecular labeling strategy because this reaction is reagent-free, reasonably fast, highly selective, and biocompatible.[63-65] Bioconjugation strategies based on IEDDA reaction commonly use ONs labeled with reactive

dienophiles such as norbornene, trans-cyclooctene, and cyclopropene groups.[66-73] While the reaction rates of strained dienophiles with tetrazine derivatives are generally fast, such bulky dienophile substituents on nucleoside are particularly not suitable for labeling cellular nucleic acids as they may not serve as good substrates for endogenous polymerases. Despite low reactivity as compared to that of strained dienophiles, vinylated 2′-deoxynucleosides have been used as chemical reporters to label and visualize DNA in cells by IEDDA reaction.[45] Encouraged by these results, we decided to study the reactivity of vinyl label of transcript **4** toward diene counterpart in IEDDA reaction. For this purpose, we chose two tetrazine cores, which had reportedly shown a good reactivity against an electron-rich strained alkene.[66-73] The reactivity of transcript **4** was tested by using a commercially available tetrazine **10** and biotinylated (**11**) and Cy5-conjugated (**12**) tetrazines, which were synthesized by following an analogous literature procedure (Figures 2.13, 2.14 and Scheme 2.2).[64]
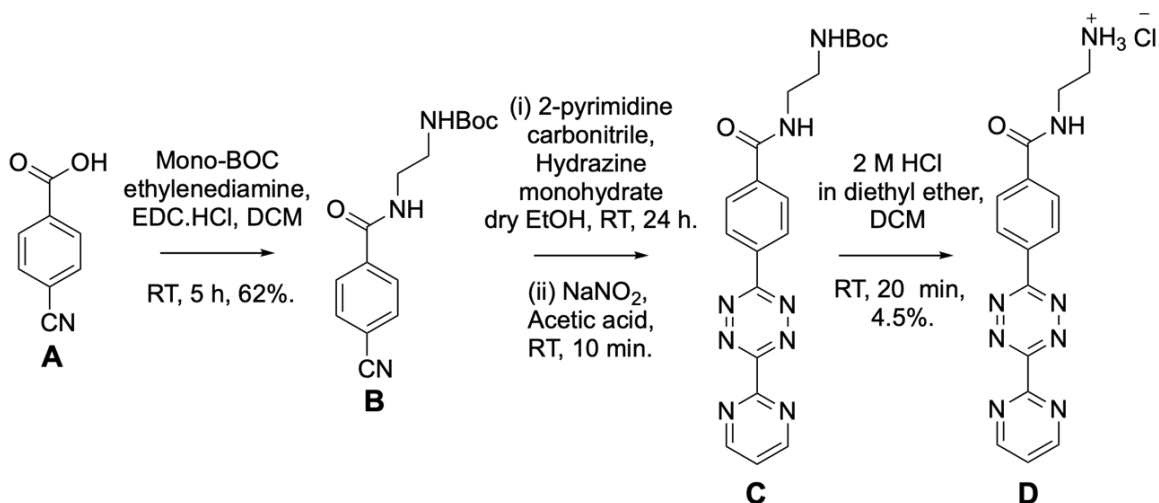
Scheme 2.2. Synthesis of N-(2-aminoethyl)-4-(6-(pyrimidin-2-yl)-1,2,4,5-tetrazin-3-yl)benzamide **D** (a) mono-BOC ethylene diamine, EDC.HCl, DMAP, CH$_2$Cl$_2$, RT, 5 h (b) (i) Hydrazine monohydrate, dry ethanol, 2-pyrimidine carbonitrile, 90℃, 12 h (ii) NaNO$_2$, acetic acid, RT, 10 min (c) 2 N HCl in diethyl ether, RT, 45 min.
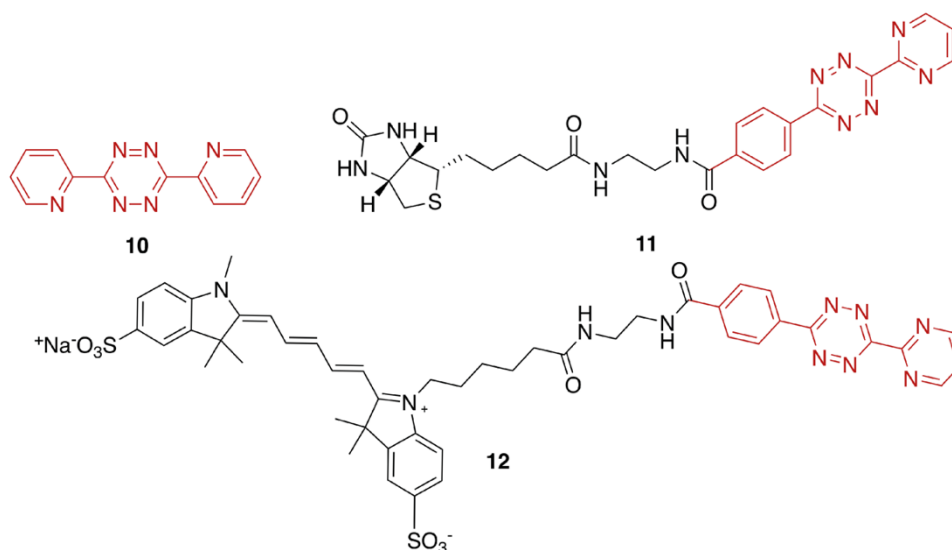
Figure 2.13. (**A**) Tetrazine substrates **10-12** used in this study. Tetrazine **11** and **12** were made by coupling tetrazine core **D** with biotin *N*-hydroxysuccinamide and sulfo-Cy5 *N*-hydroxysuccinamide ester (Scheme 2.2). Product **11** and **12** was visualized in HPLC (Figure 2.14).
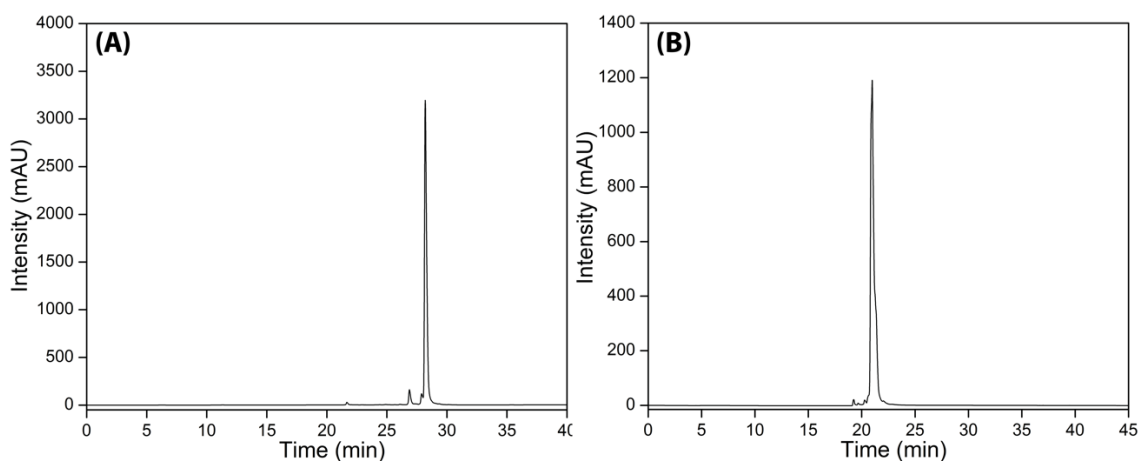
Figure 2.14. HPLC chromatogram for synthesized tetrazine substrates (**A**) **11** and (**B**) **12**.

Initially, the IEDDA reaction was performed using the free nucleoside **2** and a tetrazine substrate **10**. The reaction proceeded well, and almost complete conversion was observed after 21 h (Scheme 2.3). The reaction resulted in a mixture of ligated products (**2b** and **2c**), which were isolated and characterized. **2a** was found to be oxidized pyridazine product, whereas **2c** corresponded to the dihydropyridazine product, which underwent slow oxidation to form **2b**. Formation of a mixture of products and slow oxidation event has been documented earlier in DNA conjugation experiments.[45,74]
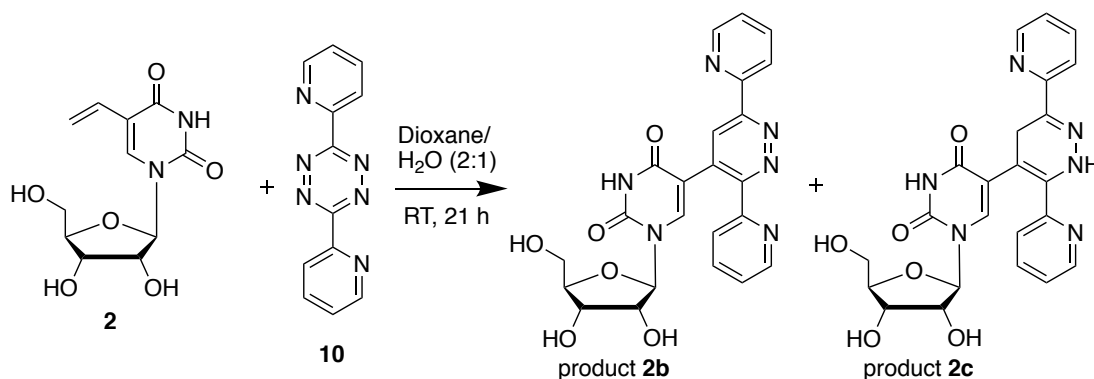
Scheme 2.3. IEDDA reaction between vinyl-modified nucleoside **2** and tetrazine **10**.

Next, IEDDA reactions were performed between RNA ON **4** and tetrazine substrates in Tris-HCl buffer (pH 7.5) at 37 °C, and aliquots of reaction mixture at various time intervals were analyzed by analytical PAGE under denaturing conditions. A reaction with tetrazine **10** was found to be almost complete in 12 h, whereas with substrates **11** and **12**, the reaction was partially complete even after 15 h (Figure 2.15). Reactions at elevated temperatures (45 and 55 °C) did not result in noticeable improvement in reaction efficiency (data not shown). This is in agreement with reports of less reactivity of the tetrazine core of **11** and **12** as compared to a tetrazine core of **10**.[65]
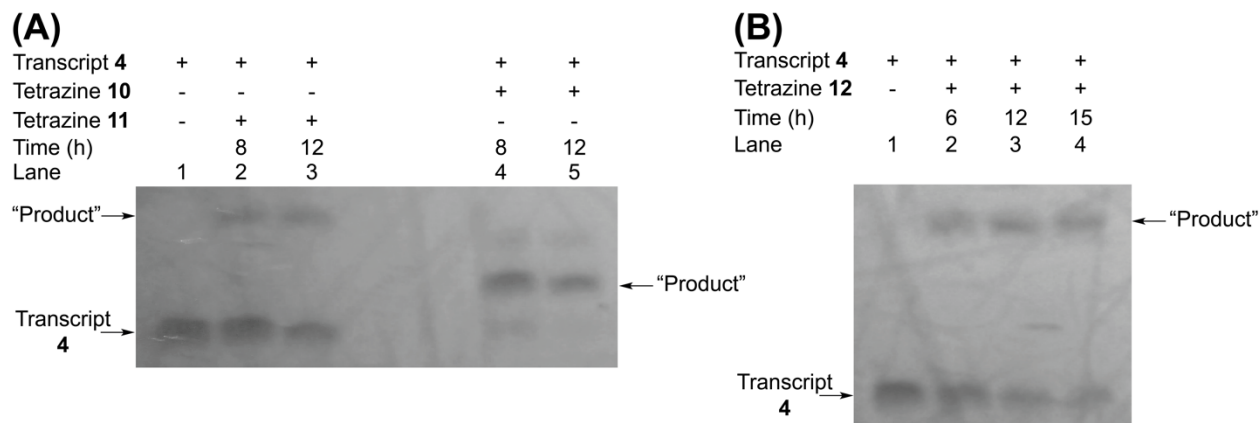
Figure 2.15. UV shadow image of PAGE resolved IEDDA reaction between transcript **4** and tetrazine **10-12**. (**A**) Reaction with tetrazine **10** and **11**. (**B**) Reaction with tetrazine **12**.

Large-scale reactions were then performed to isolate and characterize the RNA products. Reaction with substrates **10-12** produced major and minor (could not be isolated) bands (Figure 2.16). Mass analysis of major bands confirmed the formation of pyridazine products **10a-12a** (Table 2.2), similar to previous reports.[72] UV-vis profile of purified products was distinguishingly different from the profile of transcript **4**, and further fluorescence analysis also confirmed the

bioconjugation by IEDDA reaction (Figure 2.17). While HPLC analysis of the major band obtained from a reaction with symmetrical tetrazine **10** gave a single peak corresponding to the pyridazine **10a**, reactions with asymmetrical tetrazines **11** and **12** afforded mixture of pyridazine isomers (Figure 2.18). The formation of mixture of isomeric products is consistent with literature reports.[45,75]
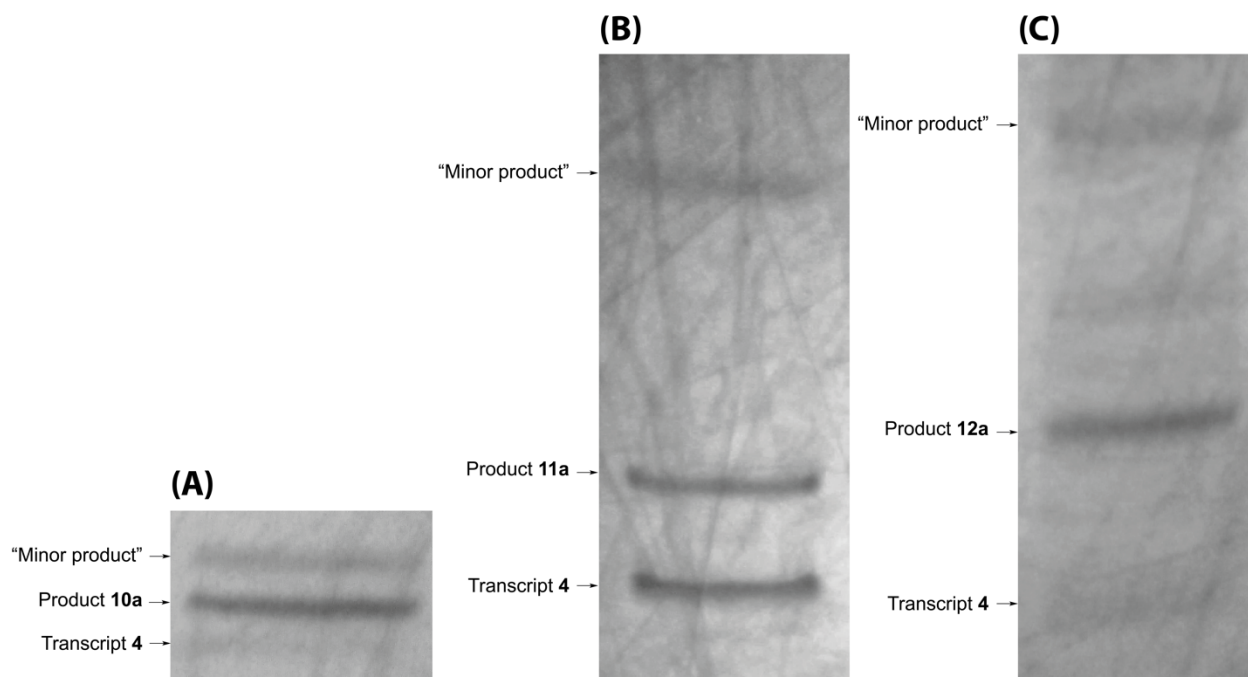
Figure 2.16. UV shadow image for large-scale IEDDA reaction between transcript **4** and (**A**) tetrazine **10** for 12 h, (**B**) tetrazine **11** for 12 h, and (**C**) tetrazine **12** for 15 h. Major bands corresponding to **10a**, **11a** and **12a** were isolated and characterized by mass analysis (Table 2.2). See Table 2.3 for the yields.
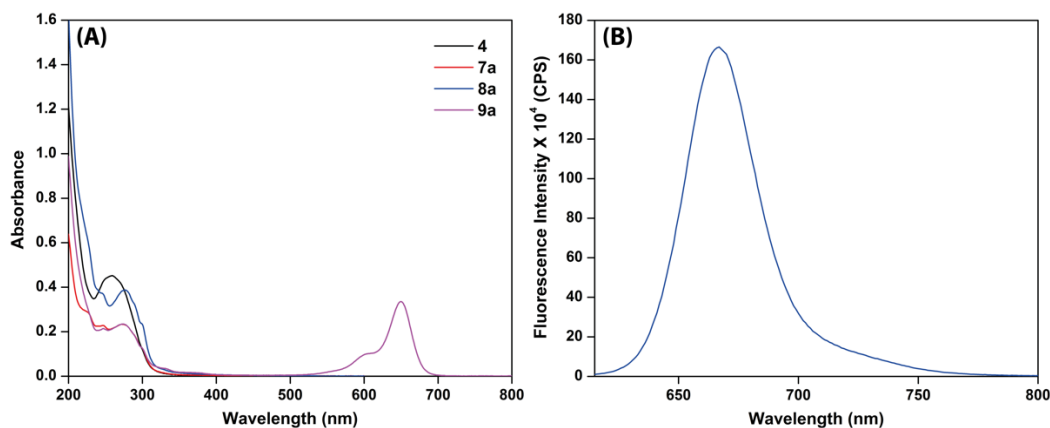
Figure 2.17. (**A**) Absorption profile of transcript **4**, RNA products **10a**, **11a** and **12a**. (**B**) Fluorescence spectra of RNA conjugated to sulfo-Cy5 tetrazine **12a.** Samples were excited at 600 nm and excitation and emission slit widths were maintained at 4 nm and 6 nm, respectively.
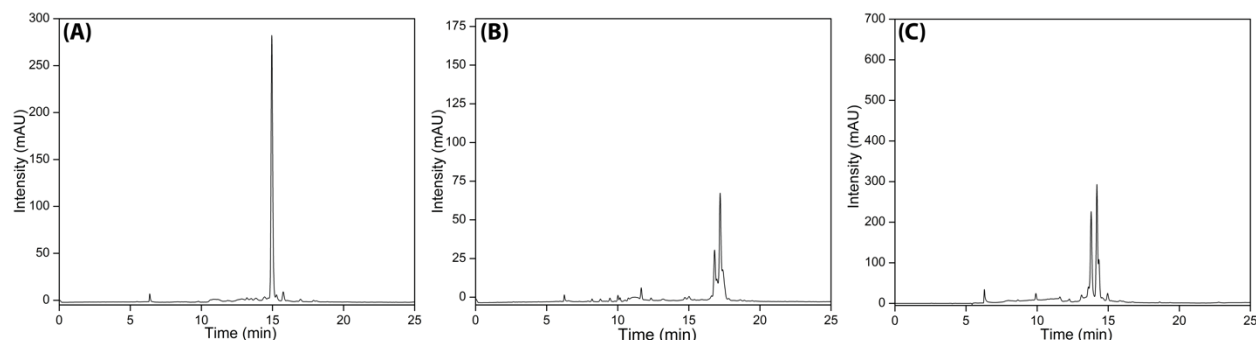
Figure 2.18. HPLC trace of IEDDA reaction products (**A**) **10a**, (**B**) **11a** and (**C**) **12a**. Mobile phase A: 50 mM triethylammonium acetate buffer (TEAA, pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0-10% B in 10 min and 10-100% B in 20 min. Reactions with asymmetrical tetrazines **11** and **12** afforded mixture of pyridazine isomers of **11a** and **12a**, respectively.

2.2.6 High-density vinyl labeling of longer RNA transcript and IEDDA reaction

High-density labeling of VUTP into a longer RNA transcript was studied by using a DNA template that would generate a 59 mer RNA containing nine modifications. T7 RNA polymerase effectively incorporated the vinyl analog into the 59 mer RNA **13** with efficiency comparable to that of natural UTP (Figure 2.19A). Transcript **13** was subsequently reacted with sulfo-Cy5 labeled tetrazine **12**, and the reaction product was resolved on a 2% agarose gel. The image captured using UV-transilluminator (364 nm) clearly revealed the fluorescence labeling of the longer RNA transcript (Figure 2.19B).
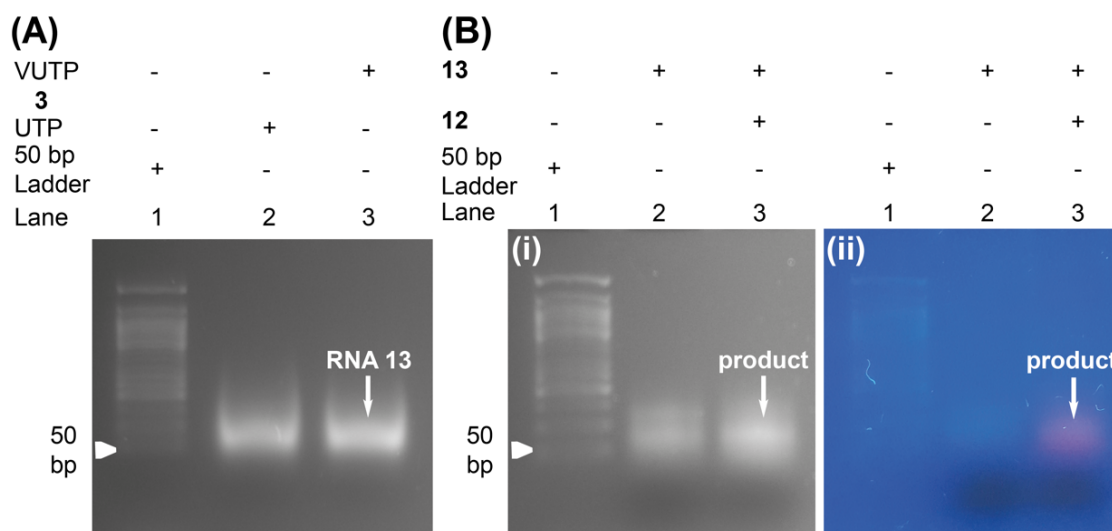
Figure 2.19. (**A**) Agarose gel picture of 59 mer control (lane 2) and VU-modified RNA transcript **13** (lane 3). Lane 1: 50 bp DNA ladder. (**B**) Agarose gel picture of IEDDA reaction between RNA **13** and Cy5-tetrazine **12**. UV-transilluminator image at (i) 254 nm and (ii) 364 nm. Lane 1: 50 bp DNA ladder.

2.3 Conclusions

We have successfully incorporated the vinyl functionality into RNA by *in vitro* transcription reaction using VUTP **3**. The results demonstrate that VU incorporated into RNA transcripts can serve as a useful handle for chemoselective functionalization of RNA in a modular fashion by oxidative Heck and IEDDA reaction. In particular, the generation of fluorogenic RNA by oxidative Heck reaction with boronic acid and ester is advantageous. It suggests that a screening reaction with appropriate boronic acid and ester substrates could provide direct access to RNA emitting at different wavelengths. Even though conjugation by utilizing oxidative Heck reaction resulted in a mixture of isomers, this study represents a promising initial step on further research on oxidative Heck reaction for the design of made-to-order biologically compatible regioselective and stereoselective ligands for palladium catalysis.[60,76] Such studies would pave the way for selective incorporation of desired biophysical probes (for example, microenvironment-sensing probes onto RNA). Furthermore, metabolic labeling of VU followed by post-transcriptional functionalization by these methods could potentially enable the imaging of RNA in cells. Taken together, the studies presented here are expected to complement other bioorthogonal RNA labeling strategies by providing alternative access to RNA labeled with biophysical probes and tags.

2.4 Experimental Section

2.4.1 Materials

2-pyrimidine carbonitrile, 4-cyanobenzoic acid, biotin, 3,6-di-2-pyridyl-1,2,4,5-tetrazine (tetrazine **10**), P(furyl)$_3$, *N,N*-diisopropylethylamine, Pd$_2$dba$_3$, hydrazine monohydrate were obtained from Sigma Aldrich. Tributyl(vinyl)tin was obtained from Alfa Aesar. POCl$_3$ was purchased from Acros Organics and was distilled before use. Benzothiophene-2- (**5**), benzofuran-2- (**6**), indole-5- (**7**) and dibenzothiophene-4- (**8**) boronic acids were purchased from Alfa Aesar. Benzothiophene-2-vinyl boronic ester (**9**) was kindly provided by Manisha Walunj.[47] Iodouridine was synthesized from uridine following well-known procedures. Activation of biotin was done using *N*-hydroxy succinamide. Sulfo-Cy5 *N*-hydroxy succinamide ester was purchased from Lumiprobe. DNA oligonucleotides (ONs) were purchased from Integrated DNA Technologie, Inc., and were purified by denaturing gel electrophoresis and were desalted using Sep-Pak Classic C18 cartridges (Waters Corporation). NTPs, T7 RNA polymerase, ribonuclease inhibitor (RiboLock), RNase A, RNase T1 were obtained from Thermo Fischer Scientific. Calf intestine

alkaline phosphatase (CIP) and snake venom phosphodiestrase I was purchased from Invitrogen and Sigma Aldrich. Reagents for buffer solutions were obtained from Sigma Aldrich (Bio Ultra Grade). Autoclaved water used in oxidative Heck reaction was purged with oxygen for 15 min before use.

2.4.2 Instrumentation

Mass measurements were recorded using Applied Biosystems 4800 Plus MALDI-TOF/TOF analyzer and HRMS Water Synapt G2 high definition mass spectrometers. Absorption experiments were carried out using UV-2600 Shimadzu spectrophotometer. NMR was recorded in Bruker Avance III HD Ascend 400 MHz and Bruker Advance III HD Ascend 600 MHz and processed in Mnova NMR software from Mestrelab Research. Fluorescence spectra were recorded in micro fluorescence cuvette (Hellma, path length 1.0 cm) using Horiba Jobin Yvon Fluorolog-3. Reversed-phase (RP) flash chromatographic (C18 Redi*SepRf* column) purifications were performed using a Teledyne ISCO, Combi Flash Rf. HPLC analysis was done on Agilent Technologies 1260 Infinity HPLC.

2.4.3 Synthesis of VUTP 3

5-Vinyluridine (2):[51] 5-Iodouridine **1** (0.30 g, 0.81 mmol, 1 equiv), Pd_2dba_3, (0.04 g, 0.04 mmol, 0.05 equiv), P(furyl)$_3$ (0.023 g, 0.10 mmol, 0.12 equiv) was dissolved in degassed anhydrous DMF (6 ml) and was purged with N_2 for 30 min. Tributyl(vinyl)tin (0.321 g, 1.01 mmol, 1.25 equiv) was added drop wise and the reaction was kept at 60 °C for 3 h. Once the reaction was complete, the reaction mixture was filtered through a celite pad and was subsequently washed with MeOH. Solvent was evaporated under vacuum and the residue was purified by silica gel column chromatography in $MeOH/CHCl_3$ to afford the product as a white solid (116 mg, 53%). TLC (MeOH:CHCl$_3$ = 25:75) R_f = 0.5; ^{1}H-NMR (400 MHz, d_6-DMSO): δ (ppm) 11.44 (s, 1H), 8.21(s, 1H), 6.36 (dd, J = 17.6, 11.6 Hz, 1H), 5.91 (dd, J = 17.8, 1.8 Hz, 1H), 5.78 (d, J = 4.8 Hz, 1H), 5.43 (d, J = 4.8 Hz, 1H), 5.25 (t, J = 4.4 Hz, 1H), 5.14–5.08 (m, 2H), 4.07–4.00 (m, 2H), 3.89–3.85 (m, 1H), 3.71–3.57 (m, 2H); ^{13}C NMR (100 MHz, d_6-DMSO): δ (ppm) 162.2, 149.90, 138.0, 128.7, 114.0, 110.8, 88.2, 84.7, 73.9, 69.5, 60.4; HRMS: (m/z): calculated for $C_{11}H_{13}N_2O_6$ [M-H]$^-$ = 269.0774, found = 269.0770; ε_{260} = 3540 M^{-1}cm^{-1}.

5-Vinyluridine-5'-triphosphate (VUTP) (3): Freshly distilled POCl$_3$ (52 μL, 0.56 mmol, 2.5 equiv) was added to 5-vinyluridine (0.060 g, 0.22 mmol, 1 equiv), proton sponge (0.057 g, 0.27 mmol, 1.2 equiv), trimethylphosphate (1 mL) in ice-cold conditions. The solution was stirred for 30 min at ~4 °C. Bis-tributylammonium pyrophosphate[77] (0.5 M in DMF, 1.1 mL, 2.5 equiv) and tributylamine (0.42 mL, 1.78 mmol, 8 equiv) was rapidly added to the solution in ice-cold conditions. After 30 min, the reaction was quenched with 1 M triethylammonium bicarbonate buffer (TEAB, pH 7.5, 15 mL), followed by washing with ethyl acetate (20 mL). Once the aqueous layer was evaporated, the residue was purified using a DEAE sephadex-A25 anion exchange column (10 mM–1 M TEAB buffer, pH 7.5) and was then further purified by RP flash column chromatography (C18 Redi*SepR$_f$*, 0–40% acetonitrile in 50 mM triethylammonium acetate buffer, pH 7.2, 40 min, flow rate: 7 mL/min). Desired fraction was lyophilized to afford the triphosphate product as triethylammonium salt (42.5 mg, 21%). ^{1}H-NMR (600 MHz, D$_2$O): δ (ppm) 7.98 (s, 1H), 6.52 (dd, J = 17.6, 11.5 Hz, 1H), 6.02–5.97 (m, 2H), 5.32 (d, J = 11.4 Hz, 1H), 4.49–4.39 (m, J = 5.9 Hz, 2H), 4.31–4.21 (m, 3H); ^{13}C NMR (151 MHz, D$_2$O): δ (ppm) 164.5, 151.2, 138.0, 127.6, 116.7, 113.1, 88.1, 83.6, 73.7, 69.7, 65.1; ^{31}P NMR (243 MHz, D$_2$O): δ (ppm) -9.76 (br, P$_γ$), -11.55 (br, P$_α$), -22.98 (br, P$_β$); HRMS: (m/z): calculated for C$_{11}$H$_{16}$N$_2$O$_{15}$P$_3$ [M-H]$^-$ = 508.9786, found = 508.9771.

2.4.4 *In vitro* transcription reaction using VUTP 3

Transcription reactions with α-^{32}P ATP

Annealing of duplex containing the T7 RNA polymerase consensus promoter DNA sequence and template **S1**–**S4** (5 μM) was done in TE buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) at 90 °C. After attaining room temperature, the solution was kept in an ice bath for 20 min followed by storing at -40 °C. Transcription was performed at 37 °C in 40 mM Tris-HCl buffer (pH 7.8) using 250 nM annealed promoter-template duplexes, 10 mM NaCl, 10 mM MgCl$_2$, 10 mM of dithiothreitol (DTT), 2 mM spermidine, 1 U/μL RNase inhibitor (Riboblock), 1 mM GTP, CTP, UTP and or modified UTP **3**, 20 μM ATP, 5 μCi α-^{32}P ATP and 3 U/μL T7 RNA in a 20 μL reaction volume. After 3.5 h, the reaction was quenched using 20 μL of loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8). The sample was heated at 75 °C for 3 min and then cooled in an ice bath. The samples (4 μL) were loaded onto sequencing 18% denaturing polyacrylamide gel and were electrophoresed. The bands corresponding to the

radioactive products were imaged using an X-ray sheet. The relative transcription efficiency was quantified using GeneTools software from Syngene. The incorporation efficiency of VUTP **3** into full-length RNA transcripts is reported with respect to incorporation of natural UTP in respective control reactions. All reactions were performed in duplicate and the errors in yields were < 2% (Figure 2.2 and 2.3).

Large-scale transcription reaction using template S1 and VUTP 3

Transcription was performed in 250 μL reaction volume using 2 mM ATP, GTP, CTP and the modified VUTP **3**, 20 mM $MgCl_2$, 0.4 U/μL RNase inhibitor (Riboblock), 300 nM annealed template and 800 units T7 RNA polymerase. The reaction solution was incubated at 37 °C for 12 h. The solution was made one-third its volume by speed vac followed by addition of loading buffer (50 μL) and was loaded onto a preparative 20% denaturing polyacrylamide gel and electrophoresed. After resolving, the gel was UV shadowed and appropriate band was excised. The transcript was extracted from gel using 0.3 M sodium acetate followed by desalting using a Sep-Pak C18 cartridge. Approximately 15 nmol of transcript **4** was obtained under these conditions when quantified using UV spectrophotometer ($\varepsilon_{260} = 84740$ $M^{-1}cm^{-1}$).

MALDI TOF mass analysis of transcripts

Sample for mass analysis was prepared using 3 μL of ~200 μM transcript, 1.5 μL of 100 μM DNA standard, 1 μL of 100 mM ammonium acetate buffer (pH 9) and 4 μL of matrix (saturated 3-hydropiccolinic acid). The solution was desalted using ion-exchange resin (Dowex 50W-X8, 100-200 mesh, ammonium form) and was characterized by MALDI-TOF mass spectrometry. The spectrum was then calibrated using internal DNA standard (Figure 2.4A, Table 2.2 and Appendix-I).

Enzymatic digestion of RNA transcript 4

Transcript **4** (4 nmol) was incubated with calf intestinal alkaline phosphatase (10 μL, 1 U/μL), RNase A (0.25 μg), snake venom phosphodiesterase I (0.01 U), 50 mM Tris-HCl buffer (pH 8.5, 40 mM $MgCl_2$, 0.1 mM EDTA) in a total volume of 100 μL for 12 h at 37 °C. Finally, the sample was treated with RNase T1 (0.2 U/μL) and kept for 4 h at 37 °C. The ribonucleoside mixture obtained was analyzed using RP-HPLC (Pheneomix-Luma C18 column, 250 x 4.6 mm, 5 micron)

in comparison with standard nucleoside (rC, rG, rA, **2**) at 260 nm. Mobile phase A: 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0−10% B in 20 min and 10−100% B in 15 min (Figure 2.4B and Table 2.1).

2.4.5 Oxidative Heck reaction between ribonucleoside 2 and boronic acid 5

Pd-EDTA complex was prepared by following a previously reported procedure.[41] To a solution of EDTA (0.138 g, 0.37 mmol, 5 equiv) in distilled water (1 mL), Pd(OAc)$_2$ (0.083 g 0.37 mmol 5 equiv) was added. The solution was purged with oxygen and further heated at 65 °C for 1 h under an atmosphere of oxygen. To the bright yellow solution of Pd-EDTA complex, a pre-mixed solution containing **2** (0.02 g, 0.07 mmol, 1equiv) and boronic acid **5** (0.066 g, 0.37 mmol, 5 equiv) in MeOH (3 mL) were added. The reaction was monitored for 42 h at room temperature. The reaction mixture was filtered through a celite pad and was washed with MeOH. Solvent was evaporated under vacuum and the residue was initially purified by RP flash column chromatography (C18 Redi*SepR$_f$*, 25–100% MeOH in Millipore water, 30 min). Further, the major product **2a** was purified by RP-HPLC using Pheneomix-Luma semi-prep C18 column to obtain an analytical quantity of product, Mobile phase A: Millipore water, mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 10-100 % B in 30 min. Reverse-Phase TLC (MeOH:H$_2$O = 70:30) R_f = 0.47; ^{1}H-NMR (600 MHz, d_6-DMSO): δ (ppm) 11.60 (s, 1H), 8.37 (s, 1H), 7.90 (d, J = 7.8 Hz, 1H), 7.77–7.73 (m, 2H), 7.41 (s, 1H), 7.37–7.32 (m, 2H), 6.71 (d, J = 16.2 Hz, 1H), 5.81 (d J = 4.2 Hz, 1H), 5.46 (d, J = 5.4 Hz, 1H), 5.35 (t, J = 4.8 Hz, 1H), 5.10 (d, J = 5.4 Hz, 1H), 4.13–4.11 (m, 1H), 4.07–4.05 (m, 1H), 3.90–3.89 (m, 1H), 3.79–3.76 (m, 1H), 3.66–3.63 (m, 1H); ^{13}C NMR (151 MHz, d_6-DMSO): δ (ppm) 162.0, 149.6, 143.0, 140.0, 139.0, 137.8, 124.8, 124.7, 123.4, 123.3, 123.0, 122.3, 121.7, 110.0, 88.4, 84.6 73.8, 69.2, 60.3; HRMS: (m/z): calculated for C$_{19}$H$_{17}$N$_2$O$_6$S [M-H]$^-$ = 401.0807, found = 401.0807.

Oxidative Heck reaction between transcript 4 and boronic acid 5-8/ester 9

Preparation of Pd-EDTA complex: To a solution of EDTA (2 mL, 8.0 mM, pH 7.5) in autoclaved water, Pd(OAc)$_2$ (3.6 mg) was added. The solution was purged with oxygen and heated at 65 °C for 1 h under an atmosphere of oxygen. The bright yellow solution of Pd-EDTA (8.0 mM) formed was filtered through a 0.45-micron syringe filter.

Analytical-scale reaction: Transcript **4** (7.55 μL, 1.33 mM, 1 equiv) was added to a solution

containing autoclaved water (69.95 μL) and DMSO (15 μL). Pd-EDTA complex (2.5 μL, 8 mM stock in autoclaved water, 2 equiv) was added followed by addition of boronic acid **5** (5 μL, 100 mM, 50 equiv). Total reaction volume was 100 μL containing 20% DMSO. The reaction was incubated at 37 °C. Aliquots (25 μL) were removed at 6 h, 12 h, 18 h and 24 h time points and were frozen at -40 °C. Control reaction sample (25 μL) in the absence of boronic acid was prepared as above. Further, loading buffer (10 μL) was added to each aliquot (25 μL) and samples were resolved by analytical PAGE (20%) under denaturing conditions. The gel was visualized using UV-shadowing (short wavelength UV, 254 nm) and on a UV-transilluminator (long wavelength UV, 365 nm) (Figure 2.6).

Large-scale oxidative Heck reaction: The scale up reaction was performed on 5 nmol of transcript **4.** To transcript **4** (4.92 μL, 1.02 mM, 1 equiv) in autoclaved water (33.83 μL) and DMSO (7.5 μL) was added Pd-EDTA complex (1.25 μL, 8 mM stock in H_2O, 2 equiv) followed by addition of boronic acid **5-8**/boronic ester **9 (**2.5 μL, 100 mM, 50 equiv). The final volume of solution was 50 μL maintaining 20% DMSO content. The solutions were incubated at 37 °C for 18 h. After incubation, the solutions were filtered through a 0.45 μm centrifuge tube filter and further washed with water (30 μL) to obtain an overall 12.5% DMSO content prior to RP-HPLC purification. The major products **5a-9a** were purified by RP-HPLC (Pheneomix-Luma C18 column, 250 x 4.6 mm, 5 micron). Mobile phase A: 50 mM triethylammonium acetate buffer (TEAA, pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0−30 % B in 35 min, 35−100% in 5 mins and 100% for 10 mins. Further, samples were lyophilized three times to remove remaining traces of TEAA. Further, the mass of product was confirmed by MALDI-TOF mass analysis (Figure 2.8, Table 2.2 and Appendix-I)

Enzymatic digestion of RNA transcript 5a (HPLC fraction 5a′)

Transcript **5a** (0.5 nmol) was incubated with calf intestinal alkaline phosphatase (10 μL, 1 U/μL), RNase A (0.25 μg), snake venom phosphodiesterase I (0.01 U), 50 mM Tris-HCl buffer (pH 8.5, 40 mM $MgCl_2$, 0.1 mM EDTA) in a total volume of 100 μL for 12 h at 37 °C. Finally, the sample was treated with RNase T1 (0.2 U/μL) and kept for 4 h at 37 °C. The ribonucleoside mixture obtained was analyzed using RP-HPLC (Pheneomix-Luma C18 column, 250 x 4.6 mm, 5 micron) in comparison with standard nucleoside (rC, rG, rA, **2a**) at 260 nm and 338 nm. Coupled products showed intense absorption at 338 nm as compared to 260 nm where the natural ribonucleosides

(rC, rG and rA) show maximum absorbance. Mobile phase A: 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0−10% B in 20 min and 10−100% B in 10 min (Figure 2.10 and Table 2.4).

2.4.6 Synthesis of biotin and Cy5 tagged tetrazine substrates 11 and 12

Tert-butyl(2-(4-cyanobenzamido)ethyl)carbamate (B): 4-cyanobenzoic acid (**A)** (2.85 g, 19.36 mmol, 1 equiv) was dissolved in dry CH_2Cl_2 (170 mL). Mono-BOC ethylenediamine (4.64 g, 28.89 mmol, 1.5 equiv), DMAP (1.18 g 9.68 mmol 0.5 equiv) and EDC.HCl (5.56 g, 28.89 mmol, 1.5 equiv) were added slowly under ice-cold conditions. After the addition, the reaction was continued at room temperature for 5 h. The reaction mixture was then washed with 5% citric acid followed by saturated $NaHCO_3$. The organic layer was extracted and dried using anhydrous Na_2SO_4. The crude product was purified by silica gel column chromatography using $MeOH/CH_2Cl_2$ as the mobile phase to obtain the product as a white solid (3.47 g, 62%). TLC ($MeOH:CH_2Cl_2$ = 10:90) R_f = 0.5; ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm); 7.93 (d, J = 8.4 Hz, 2H), 7.72–7.70 (m, 3H), 5.12 (br, 1H), 3.56–3.53 (m, 2H), 3.42–3.40 (m, 2H), 1.41 (s, 9H); ^{13}C NMR (101 MHz, $CDCl_3$): δ (ppm) 165.9, 158.2, 138.2, 132.4, 127.9, 118.2, 115.1, 80.5, 43.0, 39.8, 28.4. MALDI-TOF MS: (m/z): calculated for $C_{15}H_{19}N_3O_3K$ $[M+K]^+$ = 328.43, found = 328.10.

Synthesis of *N*-(2-aminoethyl)-4-(6-(pyrimidin-2-yl)-1,2,4,5-tetrazin-3-yl) benzamide (D)[40]: Tert-butyl(2-(4-cyanobenzamido)ethyl)carbamate (**B**) (3.47 g, 11.83 mmol, 1 equiv) and 2-pyrimidine carbonitrile (1.26 g, 11.93 mmol, 1 equiv) was dried in vacuum for 30 min and was dissolved in dry degassed EtOH (14 mL). Hydrazine monohydrate (3 mL) was added slowly, followed by heating at 90 °C for 24 h. After evaporating the solvent, the crude product was redissolved in acetic acid (375 mL), and $NaNO_2$ (1.84 g) was added proportion-wise. After 10 min, saturated $NaHCO_3$ was added slowly until the solution was neutral. The compound was extracted several times with ethyl acetate, and the organic layer was evaporated to dryness. The crude compound was purified by column chromatography ($MeOH/CHCl_3$) to obtain the product (**C**) as a pink solid (220 mg, 4.53 %). TLC ($MeOH:CH_2Cl_2$ = 5:95) R_f = 0.4.

The pink solid (**C**) (220 mg) was dissolved in dry CH_2Cl_2 (5 mL) followed by addition of 2 M HCl in dioxane (5 mL) to deprotect the BOC group. After stirring for 20 min at room temperature (RT), the compound was extracted by washing with distilled water. Aqueous layer

was evaporated and the crude product was purified by RP-HPLC to afford the compound (**D**) as a pink solid (25.5 mg). RP-HPLC was performed using Pheneomix-Luma semi-prep C18 column, Mobile phase A: Millipore water containing 0.1% TFA, mobile phase B: ACN containing 0.1% TFA. Flow rate: 1 mL/min. Gradient: 0–20 % B in 5 min, 20–85 % B in 40 min and 85–100 % B in 10 min. ^{1}H-NMR (400 MHz, D$_2$O): δ (ppm) 9.06 (d, J = 5.2 Hz, 2H), 8.51 (d, J = 8.0 Hz, 2H), 7.92 (d, J = 8.0 Hz, 2H), 7.80 (t, J = 5.0 Hz, 1H), 3.70 (t, J = 6 Hz, 2H), 3.27 (t, J = 6 Hz, 2H); ^{13}C NMR (101 MHz, D$_2$O): δ (ppm) 169.9, 164.1, 161.8, 158.7, 157.4, 137.1, 133.8, 128.8, 128.3, 123.9, 39.3, 37.5. HRMS: (m/z): calculated for C$_{15}$H$_{15}$N$_8$O [M+H]$^+$ = 323.1363, found = 323.1368.

Synthesis of biotin conjugated tetrazine (11): *N*-(2-aminoethyl)-4-(6-(pyrimidin-2-yl)-1,2,4,5-tetrazin-3-yl)benzamide (**D**) (7 mg, 19.51 µmol, 1 equiv) was added to biotin *N*-hydroxysuccinamide (6.6 mg, 19.51 µmol, 1 equiv), triethylamine (4.8 µL, 35.1 µmol, 1.8 equiv) in DMF (2 mL). After 12 h, the reaction mixture was evaporated and purified by semi-preparative RP-HPLC. Mobile phase A: Millipore water, mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0–100 % B in 40 min. See Figure 2.14A for HPLC spectrum of tetrazine **11.** HRMS: (m/z): calculated for C$_{25}$H$_{29}$N$_{10}$O$_3$S [M+H]$^+$ = 549.2145, found = 549.2145; C$_{25}$H$_{28}$N$_{10}$O$_3$SNa [M+Na]$^+$ = 571.1964, found = 571.1960.

Synthesis of sulfo-Cy5 conjugated tetrazine (12): *N*-(2-aminoethyl)-4-(6-(pyrimidin-2-yl)-1,2,4,5-tetrazin-3-yl) benzamide (**D**) (3.5 mg, 9.84 µmol, 1.5 equiv) was added to sulfo-Cy5 *N*-hydroxysuccinamide ester (5 mg, 6.65 µmol, 1 equiv) and triethyl amine (1.64 µL, 11.8 µmol, 1.8 equiv) in DMF (0.5 mL). After 12 h, the reaction mixture was evaporated and purified by semi-preparative RP-HPLC. Mobile phase A: Millipore water, mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0–80% B in 40 min and 80–100% B in 5 min. See Figure 2.14B for HPLC spectrum of tetrazine **12.** HRMS: (m/z): calculated for C$_{47}$H$_{51}$N$_{10}$O$_8$S$_2$ [M+H]$^+$ = 947.3327, found = 947.3337.

2.4.7 Inverse electron demand Diels-Alder reaction
IEDDA reaction between vinyl-modified nucleoside 2 and tetrazine

To a solution of vinyl uridine **2** (0.020 g, 0.07 mmol, 1 equiv) in 2:1 dioxane: H$_2$O (3 mL) was added tetrazine **10** (0.021 g, 0.09 mmol, 1.2 equiv).[45] The solution was stirred for 21 h and was

monitored by reverse phase-TLC. Solvent was evaporated under vacuum and the residue was purified by RP flash column chromatography (C18 Redi*SepRf*, 0–100% acetonitrile in Millipore water, 50 min). Product **2b** (12 mg, 34%) was isolated as a yellow solid. The product **2c** was characterized as unoxidised dihydropyridazine product by HRMS analysis. NMR could not be characterized for **2c** due to its slow oxidation to product **2b** as previously reported.[74] Reverse-Phase TLC (MeOH:H$_2$O = 70:30) R_f (**2b**) = 0.7, R_f (**2c**)= 0.6; For **2b**: ^{1}H-NMR (400 MHz, *d$_6$*-DMSO): δ (ppm) 11.46 (br, 1H), 8.83– 8.81(m, 1H), 8.65 (d, J = 8 Hz, 1H), 8.55–8.53 (m, 1H), 8.52 (s, 1H), 8.23 (s, 1H), 8.14 (d, J = 8 Hz, 1H), 8.09 (td, J = 7.8 Hz, 2 Hz, 1H), 7.99 (td, J = 7.8 Hz, 2 Hz, 1H), 7.63–7.59 (m, 1H), 7.45–7.42 (m, 1H), 5.82 (d, J = 4.8 Hz, 1H), 5.44 (d, J = 5.6 Hz, 1H), 5.11 (d, J = 5.2 Hz, 1H), 5.01 (t, J = 4.8 Hz, 1H), 4.03–3.99 (m, 1H), 3.96–3.92 (m, 1H), 3.85–3.81 (m, 1H), 3.59–3.54 (m, 1H), 3.50–3.45 (m, 1H) ^{13}C NMR (100 MHz, *d$_6$*-DMSO): δ (ppm) 160.8, 158.1, 157.1, 155.7, 152.6, 150.2, 149.9, 148.3, 139.3, 137.8, 137.0, 132.7, 126.1, 125.3, 123.9, 123.8, 121.2, 112.2, 88.2, 84.8, 73.8, 69.5, 60.4; HRMS: (m/z): calculated for **2b** C$_{23}$H$_{19}$N$_6$O$_6$ [M-H]$^-$ = 475.1366, found = 475.1377. HRMS: (m/z): calculated for **2c** C$_{23}$H$_{21}$N$_6$O$_6$ [M-H]$^-$ = 477.1523, found = 477.1532. The spectrum also contained a peak corresponding to the oxidized product **2a**.

IEDDA reaction between transcript 4 and tetrazine substrates 10-12

With tetrazine 10: A 50 μL reaction cocktail containing 20% DMSO was prepared. Transcript **4** (10.42 μL, 0.480 mM, 1 equiv) was added to a solution containing Tris-HCl buffer (10 μL, 100 mM, pH 7.5), NaCl (5 μL, 200 mM), DMSO (5 μL) and water (14.58 μL). Tetrazine **10** (5 μL, 20 mM stock in DMSO, 20 equiv) was added and incubated at 37 °C and aliquots (25 μL) were removed at 8 h and 12 h and were frozen at -40 °C. Control reaction was made by adding **4** (5.21 μL, 0.480 μM) to solution containing Tris-HCl buffer (5 μL, 100 mM, pH 7.5), NaCl (2.5 μL, 200 mM), autoclaved water (9.79 μL), DMSO (2.5 μL). Further, loading buffer (10 μL) was added to each aliquot and samples were resolved by analytical denaturing polyacrylamide gel (20%) under denaturing conditions. The gel was visualized using UV-shadowing method (short wavelength UV, 254 nm) (Figure 2.15).

Large-scale reaction with tetrazine 10: The scale up reaction was performed on 15 nmol of transcript **4**. Transcript **4** (21.1 μL, 0.713 mM, 1 equiv) was added to Tris-HCl buffer (30 μL, 100 mM, pH 7.5), NaCl (15 μL, 200 mM), autoclaved water (53.9 μL). Tetrazine **10** (30 μL, 10 mM

stock in DMSO, 20 equiv) was added to make a final volume of 150 µL with 20% DMSO content. The reaction solution was incubated at 37 °C for 12 h. The solution was concentrated in speed vac and was purified by preparative PAGE (20% gel) under denaturing condition. The product band (**10a**) was cut, eluted and desalted using Sep-Pak Classic C18 column. Further, the mass of major product **10a** was confirmed by MALDI-TOF mass analysis and HPLC trace was taken for the same (Figure 2.16-18, Table 2.2 and Appendix-I).

With tetrazine-biotin 11: A 200 µL reaction cocktails containing 20% DMSO content was prepared. Transcript **4** (8.63 µL, 0.58 mM, 1 equiv) was added to a solution containing Tris-HCl buffer (40 µL, 100 mM, pH 7.5), NaCl (20 µL, 200 mM) and autoclaved water (91.4 µL). Tetrazine-biotin **11** (40 µL, 2.5 mM stock in DMSO, 20 equiv) was added and incubated at 37 °C. Aliquots (100 µL) were removed at 8 h and 12 h time points and were frozen at -40 °C. Control reaction was made by adding transcript **4** (4.32 µL, 0.58 mM) to solution containing Tris-HCl buffer (20 µL, 100 mM, pH 7.5), NaCl (10 µL, 200 mM), autoclaved water (46.1 µL), DMSO (20 µL). Further, loading buffer (50 µL) was added to each aliquot (100 µL) and samples were resolved by analytical PAGE (20%) under denaturing conditions. The gel was visualized by UV-shadowing as above. A partial conversion of **4** occurred in 12 h (Figure 2.15).

Large-scale reaction with tetrazine-biotin 11: The scale up reaction was performed on 15 nmol of transcript **4**. Transcript **4** (24.6 µL, 0.609 mM, 1 equiv) was added to Tris-HCl buffer (120 µL, 100 mM), NaCl (60 µL, 200 mM), autoclaved water (275.4 µL). Tetrazine-biotin **11** (120 µL, 2.5 mM stock in DMSO, 20 equiv) was added to make a final volume of 600 µL with 20% DMSO content. The solution was incubated at 37 °C for 12 h. After incubation, the solution was concentrated in speed vac and major product **11a** was purified by denaturing polyacrylamide gel electrophoreses as above. Further, the mass of major product was confirmed by MALDI and HPLC trace was taken for the same (Figure 2.16-18, Table 2.2 and Appendix-I).

With sulfo-Cy5 tetrazine 12: Transcript **4** (13.25 µL, 0.566 mM, 1 equiv) was added to a solution containing Tris-HCl buffer (15 µL, 100 mM, pH 7.5), NaCl (7.5 µL, 200 mM) and autoclaved water (24.25 µL). Sulfo-Cy5 tetrazine **12** (15 µL, 10 mM stock in DMSO, 20 equiv) was added and reaction was incubated at 37 °C. Total volume of the reaction solution was 75 µL containing 20% DMSO. Aliquots (25 µL) were removed at 6 h, 12 h and 15 h time points and were frozen at

-40 °C. Control reaction was made by adding transcript **4** (4.42 μL, 0.566 mM) to solution containing Tris-HCl buffer (5 μL, 100 mM, pH 7.5), NaCl (2.5 μL, 200 mM), autoclaved water (8.18 μL), DMSO (5 μl). Further, loading buffer (10 μL) was added to each aliquot (25 μL) and samples were resolved by analytical polyacrylamide gel (20%) under denaturing conditions. The gel was visualized by UV-shadowing as above (Figure 2.15).

Large-scale reaction with sulfo-Cy5 tetrazine 12: The scale up reaction was performed on 15 nmol of transcript **4**. Transcript **4** (21.61 μL, 0.694 mM, 1 equiv) was added to Tris buffer (30 μL, 100 mM, pH 7.5), NaCl (15 μL, 200 mM), autoclaved water (53.39 μL). Sulfo-Cy5 tetrazine **12** (30 μL, 10 mM stock in DMSO, 20 equiv) was added to make a final volume of 150 μL having 20% DMSO. The reaction solution was incubated at 37 °C for 15 h. After incubation, the solution was concentrated in speed vac and major product **12a** was purified by denaturing polyacrylamide gel electrophoreses as above. Further, the mass of product was confirmed by MALDI-TOF mass analysis and HPLC trace was taken for the same (Figure 2.16-18, Table 2.2 and Appendix-I).

2.4.8 High-density incorporation of vinyl tags into longer RNA transcript and functionalization using IEDDA reaction

***In vitro* transcription of longer RNA with VUTP 3**

T7 RNA polymerase consensus promoter DNA ON was annealed to template **S5** (5 μM, see below for the sequence) in TE buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) at 90 °C. After cooling to RT, the solution was kept in an ice bath for 20 min followed by storing at -40 °C. Transcription was performed in 250 μL reaction volume using 2 mM ATP, GTP, CTP and followed by addition of modified VUTP **3** or natural UTP (control), 20 mM MgCl$_2$, 0.4 U/μL RNase inhibitor (Riboblock), 300 nM annealed template and 800 units T7 RNA polymerase. The reaction solution was incubated at 37 °C for 12 h. RNA **13** was purified by acid-phenol chloroform extraction (pH 4.5) and aqueous layer was washed in chloroform. Subsequently, the RNA was precipitated in isopropanol and washed with 75% ethanol and redissolved in water. The transcripts were visualized on a 2% agarose gel. RNA was stained using SYBR® safe gel staining dye. See below for sequence of template **S5** and RNA transcript **13** (Figure 2.19A).

DNA template S5
3' ATTATGCTGAGTGATATCCCTCTTCCCATAGGCCTAGCTTCAATCATCCG CCTCACTCTTCTCCACTGCCATGGTC 5'

Modified RNA transcript 13

5' GGGAGAAGGG**2**A**2**CCGGA**2**CGAAG**22**AG**2**AGGCGGAG**2**GAGAAGAGG**2**GA
CGG**2**ACCAG 3'

IEDDA reaction between 59 mer transcript 13 and sulfo-Cy5 tetrazine 12

Transcript **13** (35.5 μL, 0.28 mM, 1 equiv) was added to Tris-HCl buffer (6 μL, 500 mM, pH 7.5), NaCl (1.2 μL, 2.5 M), autoclaved water (77.3 μL), DMSO (20 μL). Sulfo-Cy5 tetrazine **12** (10.1 μL, 20 mM stock in DMSO, 20 equiv) was added to make a final volume of 150 μL containing 20% DMSO. The solution was incubated at 37 °C for 12 h. The reaction mixture was precipitated in isopropanol and washed in 75% ethanol and re-dissolved in water. The samples were resolved on a 2% agarose gel containing SYBR safe dye, and visualized using UV-transilluminator (254 nm and 364 nm) (Figure 2.19B).

2.5 References

1.   Wachowius, F.; Höbartner, C., Chemical RNA Modifications for studies of RNA structure and dynamics. *ChemBioChem* **2010,** *11* (4), 469−480.
2.   Sinkeldam, R. W.; Greco, N. J.; Tor, Y., Fluorescent analogs of biomolecular building blocks: Design, properties, and applications. *Chem. Rev.* **2010,** *110* (5), 2579−2619.
3.   Holstein, J. M.; Rentmeister, A., Current covalent modification methods for detecting RNA in fixed and living cells. *Methods* **2016,** *98*, 18−25.
4.   Baker, M., RNA imaging in situ. *Nat. Methods* **2012,** *9* (8), 787−790.
5.   Cmarko, D.; Verschure, P. J.; Martin, T. E.; Dahmus, M. E.; Krause, S.; Fu, X.-D.; van Driel, R.; Fakan, S., Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol. Biol. Cell* **1999,** *10* (1), 211−223.
6.   Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S., Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **2014,** *6* (1), 75−80.
7.   Wlodkowic, D.; Skommer, J.; Darzynkiewicz, Z., SYTO probes in the cytometry of tumor cell death. *Cytometry A* **2008,** *73A* (6), 496−507.
8.   Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M., Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (33), 14579−14584.
9.   Paige, J. S.; Wu, K. Y.; Jaffrey, S. R., RNA mimics of green fluorescent protein. *Science* **2011,** *333* (6042), 642−646.
10.   Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J. M. Y.; Monchaud, D., Visualization of RNA-quadruplexes in live cells. *J. Am. Chem. Soc.* **2015,** *137* (26), 8521−8525.
11.   Gramlich, P. M. E.; Wirges, C. T.; Manetto, A.; Carell, T., Postsynthetic DNA modification through the copper-catalyzed azide–alkyne cycloaddition reaction. *Angew. Chem. Int. Ed.* **2008,** *47* (44), 8350−8358.
12.   Weisbrod, S. H.; Marx, A., Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem. Commun.* **2008,** (44), 5675−5685.

13. Jao, C. Y.; Salic, A., Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (41), 15779−15784.
14. El-Sagheer, A. H.; Brown, T., New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (35), 15329−15334.
15. George, J. T.; Srivatsan, S. G., Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. *Methods* **2017,** *120*, 28−38.
16. El-Sagheer, A. H.; Brown, T., Click nucleic acid ligation: Applications in biology and nanotechnology. *Acc. Chem. Res.* **2012,** *45* (8), 1258−1267.
17. van Berkel, S. S.; van Eldijk, M. B.; van Hest, J. C. M., Staudinger ligation as a method for bioconjugation. *Angew. Chem. Int. Ed.* **2011,** *50* (38), 8806−8827.
18. Wu, H.; Devaraj, N. K., Inverse electron-demand Diels–Alder bioorthogonal reactions. *Top. Curr. Chem.* **2015,** *374* (1), 3.
19. Jayaprakash, K. N.; Peng, C. G.; Butler, D.; Varghese, J. P.; Maier, M. A.; Rajeev, K. G.; Manoharan, M., Non-nucleoside building blocks for copper-assisted and copper-free click chemistry for the efficient synthesis of RNA conjugates. *Org. Lett.* **2010,** *12* (23), 5410−5413.
20. Ishizuka, T.; Kimoto, M.; Sato, A.; Hirao, I., Site-specific functionalization of RNA molecules by an unnatural base pair transcription system via click chemistry. *Chem. Commun.* **2012,** *48* (88), 10835−10837.
21. Rao, H.; Tanpure, A. A.; Sawant, A. A.; Srivatsan, S. G., Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nat. Protoc.* **2012,** *7* (6), 1097−1112.
22. Curanovic, D.; Cohen, M.; Singh, I.; Slagle, C. E.; Leslie, C. S.; Jaffrey, S. R., Global profiling of stimulus-induced polyadenylation in cells using a poly(A) trap. *Nat. Chem. Biol.* **2013,** *9* (11), 671−673.
23. Holstein, J. M.; Schulz, D.; Rentmeister, A., Bioorthogonal site-specific labeling of the 5′-cap structure in eukaryotic mRNAs. *Chem. Commun.* **2014,** *50* (34), 4478−4481.
24. Phelps, K. J.; Ibarra-Soza, J. M.; Tran, K.; Fisher, A. J.; Beal, P. A., Click modification of RNA at adenosine: Structure and reactivity of 7-Ethynyl- and 7-Triazolyl-8-aza-7-deazaadenosine in RNA. *ACS Chem. Biol.* **2014,** *9* (8), 1780−1787.
25. Schmid, K.; Adobes-Vidal, M.; Helm, M., Alkyne-functionalized coumarin compound for analytic and preparative 4-thiouridine labeling. *Bioconjug Chem.* **2017**.
26. Cahova, H.; Winz, M.-L.; Hofer, K.; Nubel, G.; Jaschke, A., NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **2015,** *519* (7543), 374−377.
27. Sawant, A. A.; Mukherjee, P. P.; Jangid, R. K.; Galande, S.; Srivatsan, S. G., A clickable UTP analog for the posttranscriptional chemical labeling and imaging of RNA. *Org. Biomol. Chem.* **2016,** *14* (24), 5832−5842.
28. Paredes, E.; Das, S. R., Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* **2011,** *12* (1), 125−131.
29. Kislukhin, A. A.; Hong, V. P.; Breitenkamp, K. E.; Finn, M. G., Relative performance of alkynes in copper-catalyzed azide–alkyne cycloaddition. *Bioconjugate Chem.* **2013,** *24* (4), 684−689.
30. van Delft, P.; Meeuwenoord, N. J.; Hoogendoorn, S.; Dinkelaar, J.; Overkleeft, H. S.; van der Marel, G. A.; Filippov, D. V., Synthesis of oligoribonucleic acid conjugates using a cyclooctyne phosphoramidite. *Org. Lett.* **2010,** *12* (23), 5486−5489.

31. Someya, T.; Ando, A.; Kimoto, M.; Hirao, I., Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry. *Nucleic Acids Res.* **2015,** *43* (14), 6665−6676.

32. Santner, T.; Hartl, M.; Bister, K.; Micura, R., Efficient access to 3′-terminal azide-modified RNA for inverse click-labeling patterns. *Bioconjugate Chem.* **2014,** *25* (1), 188−195.

33. Neef, A. B.; Luedtke, N. W., An azide-modified nucleoside for metabolic labeling of DNA. *ChemBioChem* **2014,** *15* (6), 789−793.

34. Sawant, A. A.; Tanpure, A. A.; Mukherjee, P. P.; Athavale, S.; Kelkar, A.; Galande, S.; Srivatsan, S. G., A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **2016,** *44* (2), e16.

35. Ourailidou, M. E.; Zwinderman, M. R. H.; Dekker, F. J., Bioorthogonal metabolic labelling with acyl-CoA reporters: targeting protein acylation. *MedChemComm* **2016,** *7* (3), 399−408.

36. van Geel, R.; Pruijn, G. J. M.; van Delft, F. L.; Boelens, W. C., Preventing thiol-yne addition improves the specificity of strain-promoted azide–alkyne cycloaddition. *Bioconjugate Chem.* **2012,** *23* (3), 392−398.

37. Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q., Genetically encoded cyclopropene directs rapid, photoclick chemistry-mediated protein labeling in mammalian Cells. *Angew. Chem. Int. Ed.* **2012,** *51* (42), 10600−10604.

38. Li, Q.-F.; Yang, Y.; Maleckis, A.; Otting, G.; Su, X.-C., Thiol-ene reaction: a versatile tool in site-specific labelling of proteins with chemically inert tags for paramagnetic NMR. *Chem. Commun.* **2012,** *48* (21), 2704−2706.

39. Li, Y.; Yang, M.; Huang, Y.; Song, X.; Liu, L.; Chen, P. R., Genetically encoded alkenyl-pyrrolysine analogues for thiol-ene reaction mediated site-specific protein labeling. *Chem. Sci.* **2012,** *3* (9), 2766−2770.

40. Ourailidou, M. E.; van der Meer, J.-Y.; Baas, B.-J.; Jeronimus-Stratingh, M.; Gottumukkala, A. L.; Poelarends, G. J.; Minnaard, A. J.; Dekker, F. J., Aqueous oxidative Heck reaction as a protein-labeling strategy. *ChemBioChem* **2014,** *15* (2), 209−212.

41. Ourailidou, M. E.; Dockerty, P.; Witte, M.; Poelarends, G. J.; Dekker, F. J., Metabolic alkene labeling and in vitro detection of histone acylation via the aqueous oxidative Heck reaction. *Org. Biomol. Chem.* **2015,** *13* (12), 3648−3653.

42. Zhang, Y.; Pan, Y.; Liu, W.; Zhou, Y. J.; Wang, K.; Wang, L.; Sohail, M.; Ye, M.; Zou, H.; Zhao, Z. K., In vivo protein allylation to capture protein methylation candidates. *Chem. Commun.* **2016,** *52* (40), 6689−6692.

43. Omumi, A.; Beach, D. G.; Baker, M.; Gabryelski, W.; Manderville, R. A., Postsynthetic guanine arylation of DNA by Suzuki−Miyaura cross-coupling. *J. Am. Chem. Soc.* **2011,** *133* (1), 42−50.

44. Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G., DNA modification under mild conditions by Suzuki–Miyaura cross-coupling for the generation of functional probes. *Angew. Chem. Int. Ed.* **2013,** *52* (40), 10553−10558.

45. Rieder, U.; Luedtke, N. W., Alkene–Tetrazine ligation for imaging cellular DNA. *Angew. Chem. Int. Ed.* **2014,** *53* (35), 9168−9172.

46. Bu; Batroff, E.; Niederwieser, A.; Abdel-Rahman, O. S.; Winter, R. F.; Wittmann, V.; Marx, A., Efficient labelling of enzymatically synthesized vinyl-modified DNA by an inverse-electron-demand Diels-Alder reaction. *Chem. Commun.* **2014,** *50* (74), 10827−10829.

47. Blanchard, D. J. M.; Fadock, K. L.; Sproviero, M.; Deore, P. S.; Cservenyi, T. Z.; Manderville, R. A.; Sharma, P.; Wetmore, S. D., Photophysical properties of push-pull 8-

aryl-deoxyguanosine probes within duplex and G-quadruplex structures. *Journal of Materials Chemistry C* **2016,** *4* (14), 2915−2924.

48. Samanta, B.; Seikowski, J.; Höbartner, C., Fluorogenic labeling of 5-formylpyrimidine nucleotides in DNA and RNA. *Angew. Chem. Int. Ed.* **2016,** *55* (5), 1912−1916.

49. Dadová, J.; Orság, P.; Pohl, R.; Brázdová, M.; Fojta, M.; Hocek, M., Vinylsulfonamide and acrylamide modification of DNA for cross-linking with proteins. *Angew. Chem. Int. Ed.* **2013,** *52* (40), 10515−10518.

50. Holstein, J. M.; Stummer, D.; Rentmeister, A., Enzymatic modification of 5′-capped RNA with a 4-vinylbenzyl group provides a platform for photoclick and inverse electron-demand Diels-Alder reaction. *Chem. Sci.* **2015,** *6* (2), 1362−1369.

51. Wicke, L.; Engels, J. W., Postsynthetic on column RNA labeling via stille coupling. *Bioconjugate Chem.* **2012,** *23* (3), 627−642.

52. Srivatsan, S. G.; Tor, Y., Fluorescent pyrimidine ribonucleotide: Synthesis, enzymatic incorporation, and utilization. *J. Am. Chem. Soc.* **2007,** *129* (7), 2044−2053.

53. Pawar, M. G.; Nuthanakanti, A.; Srivatsan, S. G., Heavy atom containing fluorescent ribonucleoside analog probe for the fluorescence detection of RNA-ligand binding. *Bioconjugate Chem.* **2013,** *24* (8), 1367−1377.

54. Tanpure, A. A.; Srivatsan, S. G., Conformation-sensitive nucleoside analogues as topology-specific fluorescence turn-on probes for DNA and RNA G-quadruplexes. *Nucleic Acids Res.* **2015,** *43* (22), e149.

55. Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A., Positional impact of fluorescently modified G-tetrads within polymorphic human telomeric G-quadruplex structures. *ACS Chem. Biol.* **2015,** *10* (5), 1311−1318.

56. Suzuki, A.; Yanaba, T.; Saito, I.; Saito, Y., Molecular design of an environmentally sensitive fluorescent nucleoside, 3-deaza-2′-deoxyadenosine derivative: Distinguishing thymine by probing the DNA minor groove. *ChemBioChem* **2014,** *15* (11), 1638−1644.

57. Kanamori, T.; Ohzeki, H.; Masaki, Y.; Ohkubo, A.; Takahashi, M.; Tsuda, K.; Ito, T.; Shirouzu, M.; Kuwasako, K.; Muto, Y.; Sekine, M.; Seio, K., Controlling the fluorescence of benzofuran-modified uracil residues in oligonucleotides by triple-helix formation. *ChemBioChem* **2015,** *16* (1), 167−176.

58. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G., Fluorescent nucleoside analogs: Probes for investigating nucleic acid structure and function. *Isr. J. Chem.* **2013,** *53* (6-7), 366−378.

59. Shaikh, T. M.; Hong, F.-E., Palladium(II)-catalyzed Heck reaction of aryl halides and arylboronic acids with olefins under mild conditions. *Beilstein J. Org. Chem.* **2013,** *9*, 1578−1588.

60. Zheng, C.; Stahl, S. S., Regioselective aerobic oxidative Heck reactions with electronically unbiased alkenes: efficient access to α-alkyl vinylarenes. *Chem. Commun.* **2015,** *51* (64), 12771−12774.

61. Furukawa, K.; Abe, H.; Hibino, K.; Sako, Y.; Tsuneda, S.; Ito, Y., Reduction-triggered fluorescent amplification probe for the detection of endogenous RNAs in living human cells. *Bioconjugate Chem.* **2009,** *20* (5), 1026−1036.

62. He, Z.; Chen, Y.; Wang, Y.; Wang, J.; Mo, J.; Fu, B.; Wang, Z.; Du, Y.; Zhou, X., A rapidly photo-activatable light-up fluorescent nucleoside and its application in DNA base variation sensing. *Chem. Commun.* **2016,** *52* (55), 8545−8548.

63. Devaraj, N. K.; Hilderbrand, S.; Upadhyay, R.; Mazitschek, R.; Weissleder, R., Bioorthogonal turn-on probes for imaging small molecules inside living Cells. *Angew. Chem. Int. Ed.* **2010,** *49* (16), 2869−2872.

64. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W., Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* **2012,** *4* (4), 298−304.

65. Knall, A.-C.; Slugovc, C., Inverse electron demand Diels-Alder (iEDDA)-initiated conjugation: a (high) potential click chemistry scheme. *Chem. Soc. Rev.* **2013,** *42* (12), 5131−5142.

66. Schoch, J.; Wiessler, M.; Jäschke, A., Post-synthetic modification of DNA by inverse-electron-demand Diels−Alder reaction. *J. Am. Chem. Soc.* **2010,** *132* (26), 8846−8847.

67. Šečkutė, J.; Yang, J.; Devaraj, N. K., Rapid oligonucleotide-templated fluorogenic tetrazine ligations. *Nucleic Acids Res.* **2013,** *41* (15), e148−e148.

68. Asare-Okai, P. N.; Agustin, E.; Fabris, D.; Royzen, M., Site-specific fluorescence labelling of RNA using bio-orthogonal reaction of trans-cyclooctene and tetrazine. *Chem. Commun.* **2014,** *50* (58), 7844−7847.

69. Ameta, S.; Becker, J.; Jaschke, A., RNA-peptide conjugate synthesis by inverse-electron demand Diels-Alder reaction. *Org. Biomol. Chem.* **2014,** *12* (26), 4701−4707.

70. Pyka, A. M.; Domnick, C.; Braun, F.; Kath-Schorr, S., Diels–Alder cycloadditions on synthetic RNA in mammalian cells. *Bioconjugate Chem.* **2014,** *25* (8), 1438−1443.

71. Domnick, C.; Eggert, F.; Kath-Schorr, S., Site-specific enzymatic introduction of a norbornene modified unnatural base into RNA and application in post-transcriptional labeling. *Chem. Commun.* **2015,** *51* (39), 8253−8256.

72. Eggert, F.; Kath-Schorr, S., A cyclopropene-modified nucleotide for site-specific RNA labeling using genetic alphabet expansion transcription. *Chem. Commun.* **2016,** *52* (45), 7284−7287.

73. Lorenz, D. A.; Garner, A. L., A click chemistry-based microRNA maturation assay optimized for high-throughput screening. *Chem Commun (Camb)* **2016,** *52* (53), 8267−70.

74. Kore, A. R.; Yang, B.; Srinivasan, B., Synthesis of 5-[3,6-di(pyridin-2-yl)pyridazine-4-yl]-2'-deoxyuridine-5'-O-triphosphate-a potential probe for fluorescence detection and imaging DNA. *Tetrahedron Lett.* **2015,** *56* (6), 808−811.

75. Niederwieser, A.; Späte, A.-K.; Nguyen, L. D.; Jüngst, C.; Reutter, W.; Wittmann, V., Two-color glycan labeling of live Cells by a combination of Diels–Alder and click chemistry. *Angew. Chem. Int. Ed.* **2013,** *52* (15), 4265−4268.

76. Zhang, L.; Dong, C.; Ding, C.; Chen, J.; Tang, W.; Li, H.; Xu, L.; Xiao, J., Palladium-catalyzed regioselective and stereoselective oxidative Heck arylation of allylamines with arylboronic Acids. *Adv. Synth. Catal.* **2013,** *355* (8), 1570−1578.

77. Moffatt, J. G., A general synthesis of nucelosides-5′ triphosphates. *Can. J. Chem.* **1964**, *42*(3), 599–604.

2.6 Appendix-I: Characterization data of synthesized compounds

¹H-NMR of 5-vinyl uridine **2** in *d₆*-DMSO.
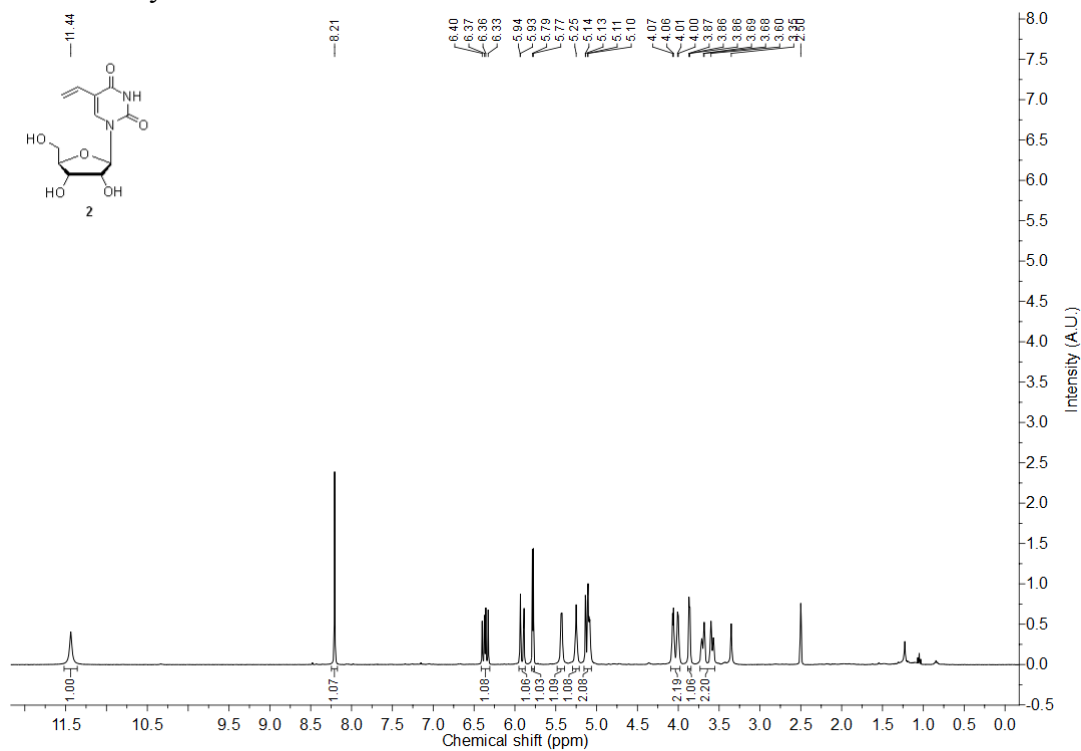
¹³C-NMR of 5-vinyl uridine **2** in *d₆*-DMSO.
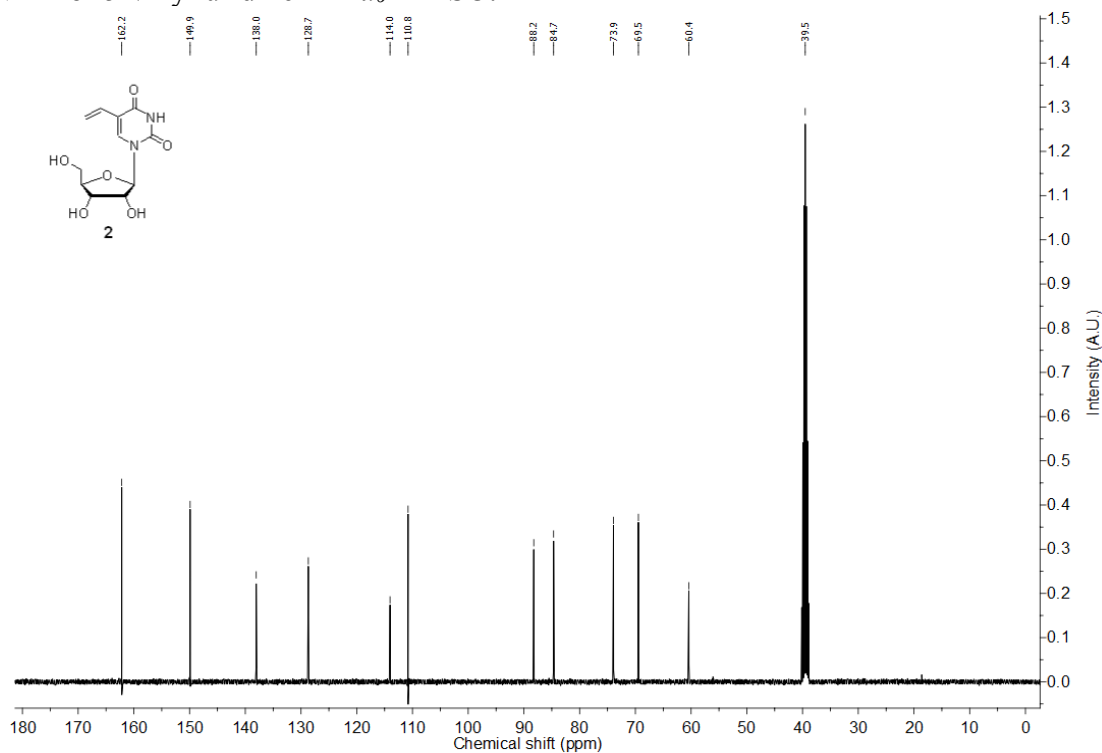
¹H-NMR of VUTP **3** in D₂O. Trace amounts of triethylammonium acetate buffer is present.
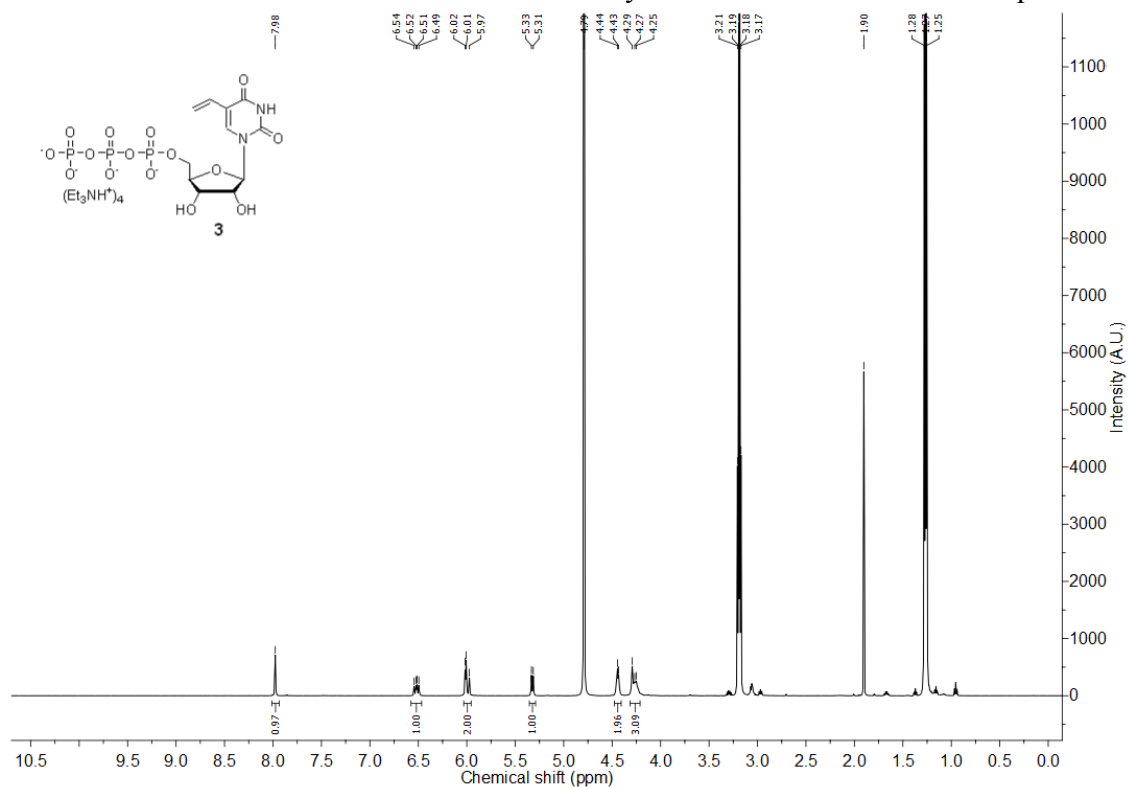
¹³C-NMR of VUTP **3** in D₂O. Trace amounts of triethylammonium acetate buffer is present.
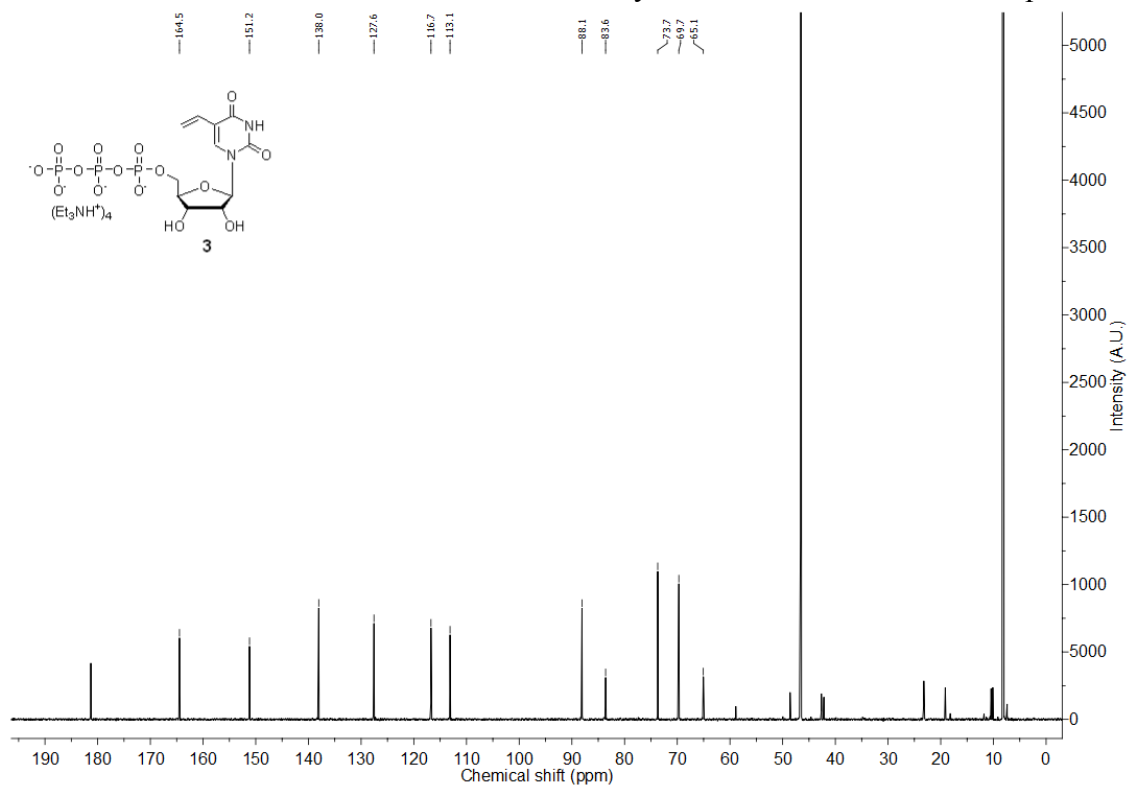
^{31}P-NMR of VUTP **3** in D_2O.
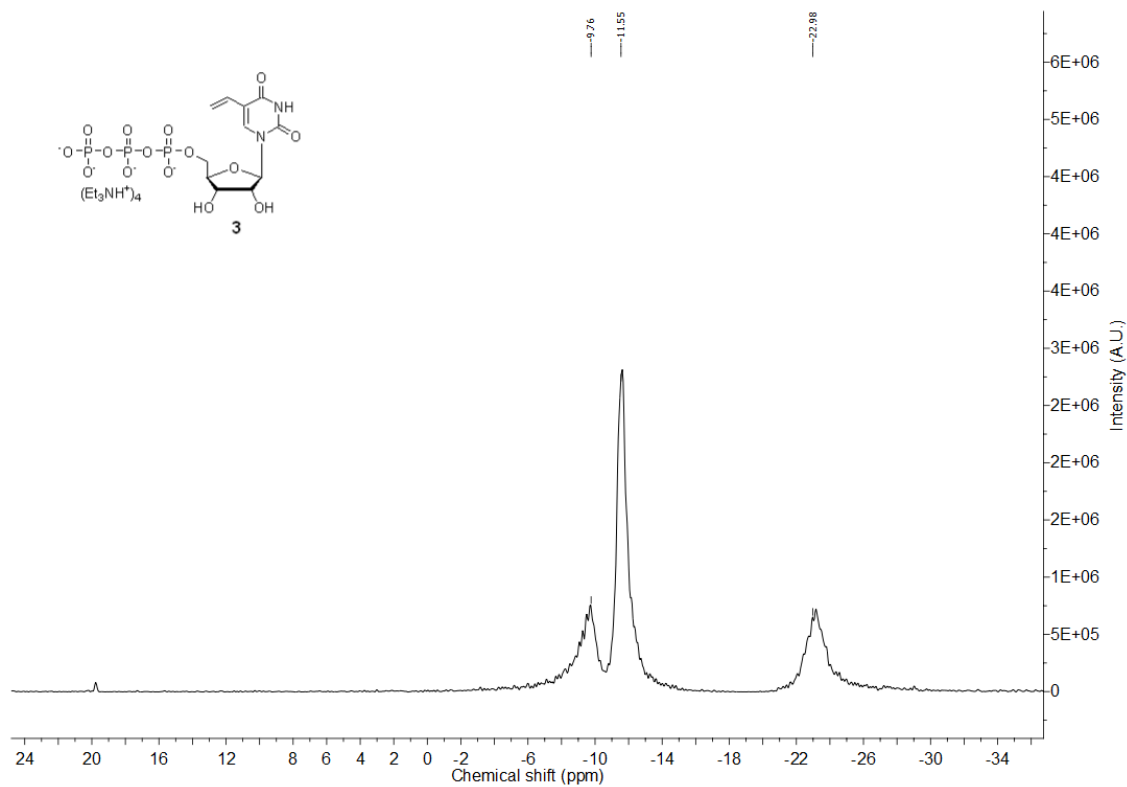
^{1}H-NMR of compound **2a** in d_6-DMSO.
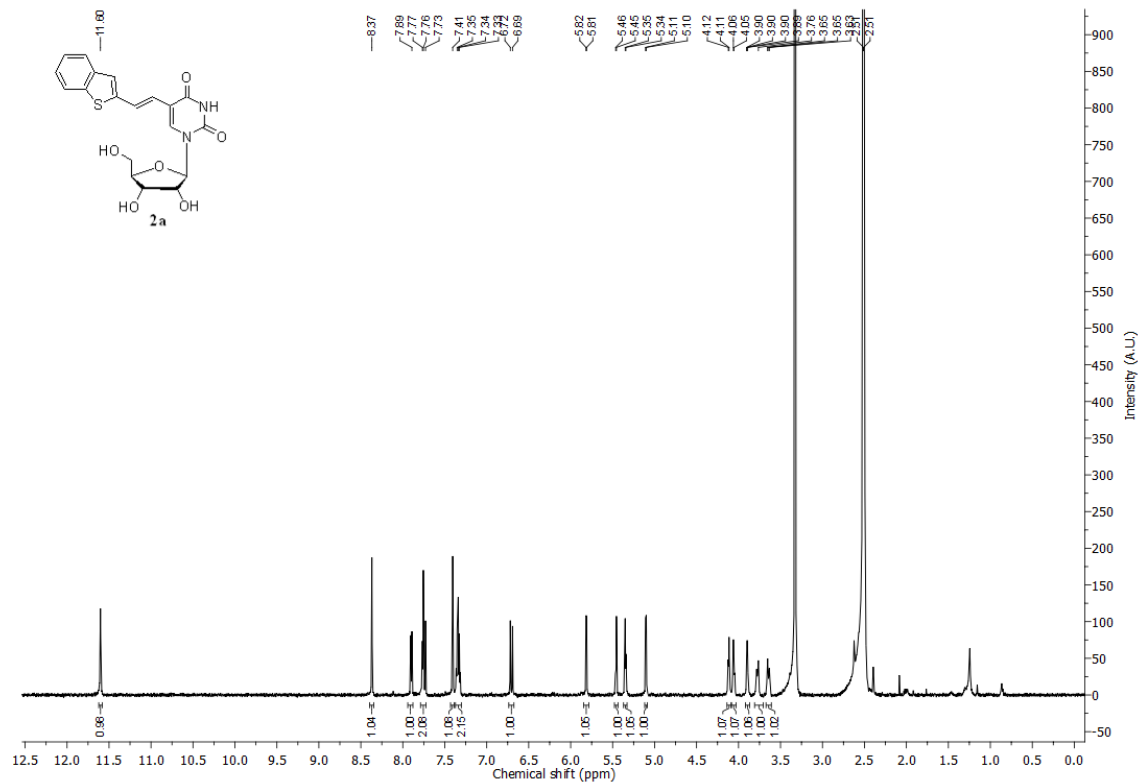
^{13}C-NMR of compound **2a** in d_6-DMSO.
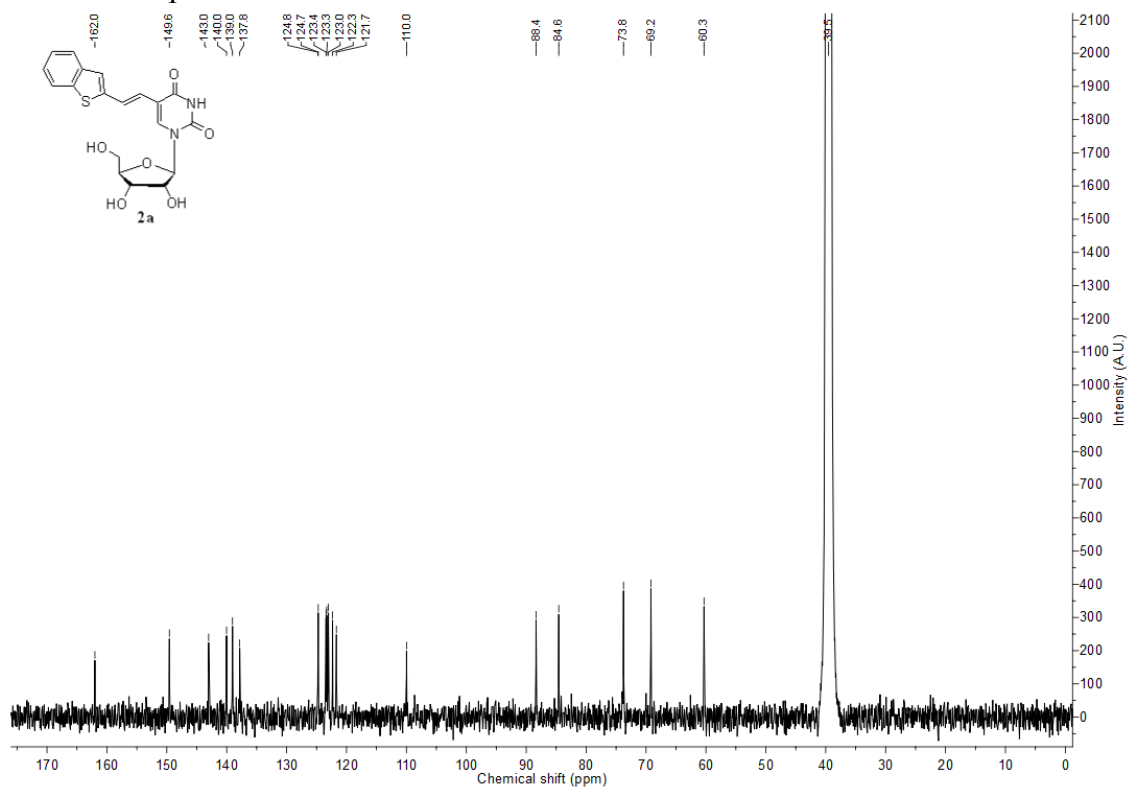
^{1}H-NMR of compound **B** in CDCl$_3$.
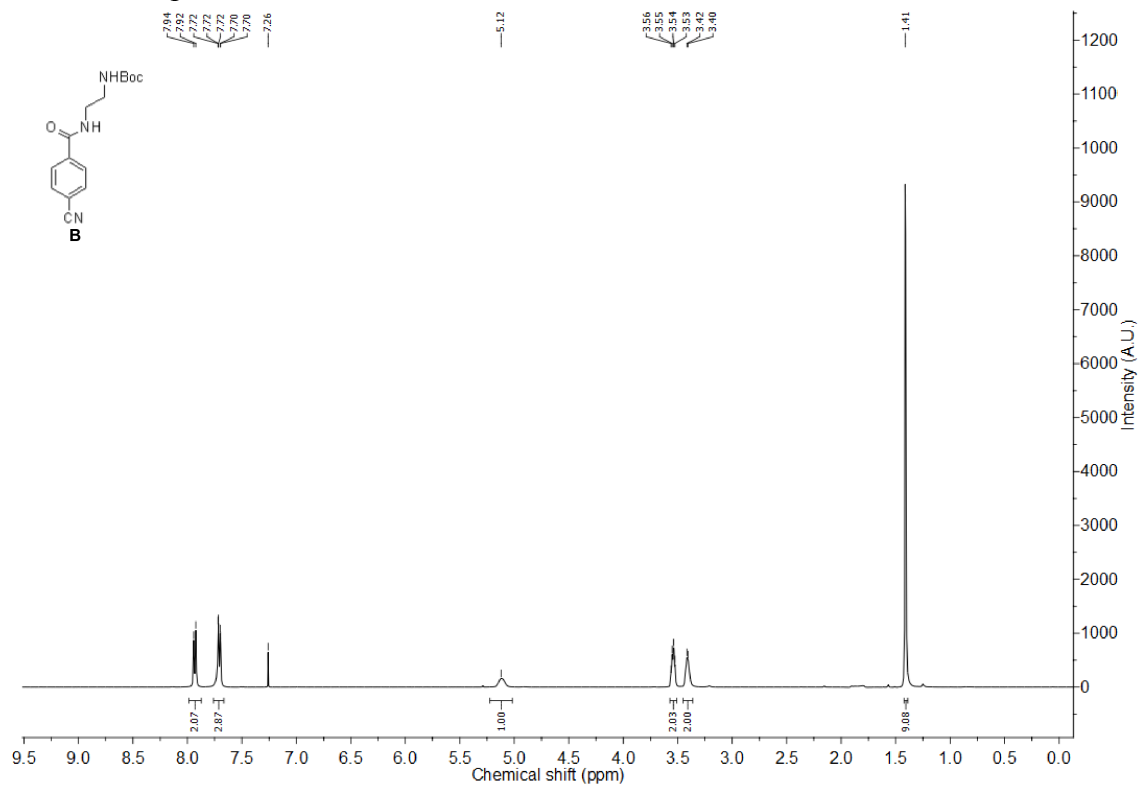
^{13}C-NMR of compound **B** in CDCl$_3$.
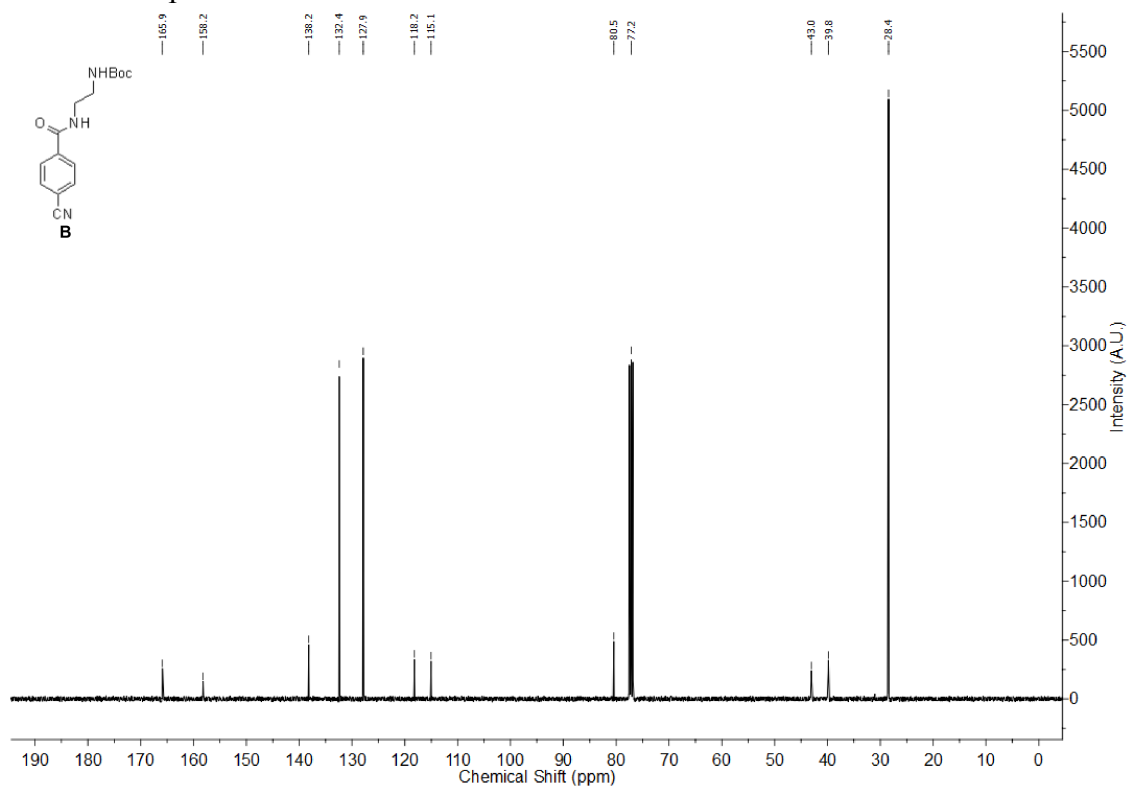
^{1}H-NMR of tetrazine **D** in D$_2$O.
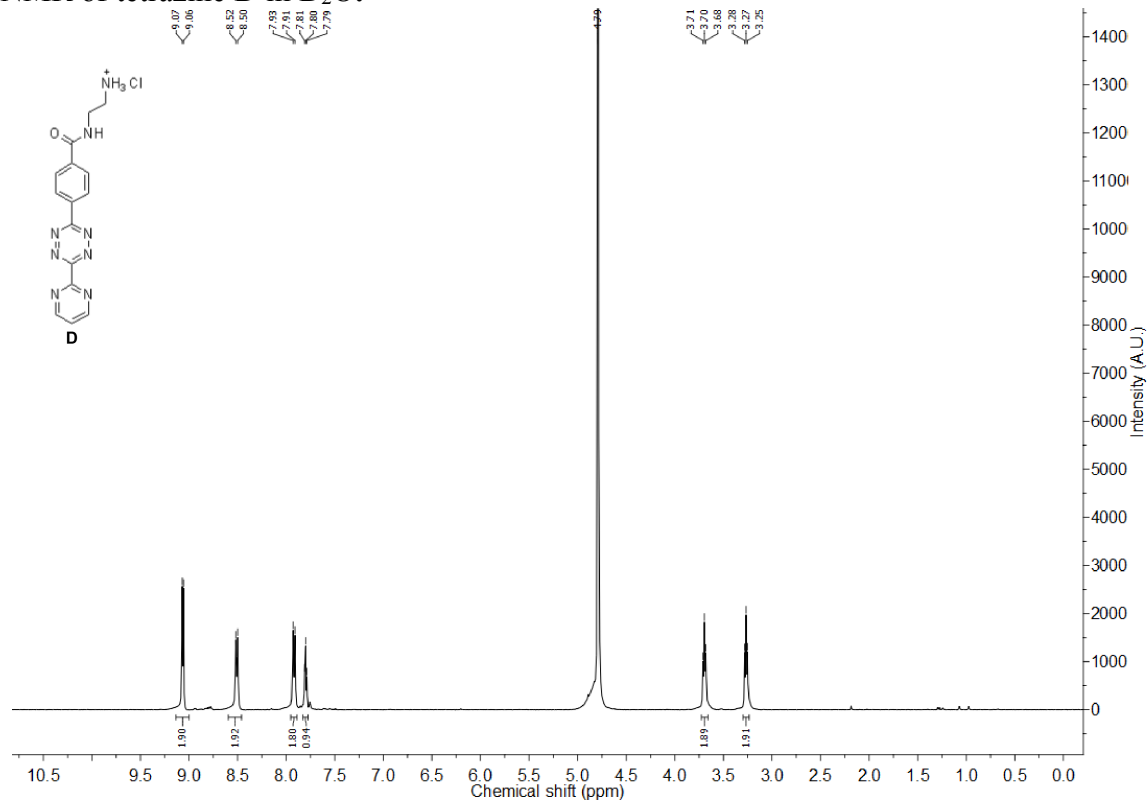
^{13}C-NMR of tetrazine **D** in D$_2$O.
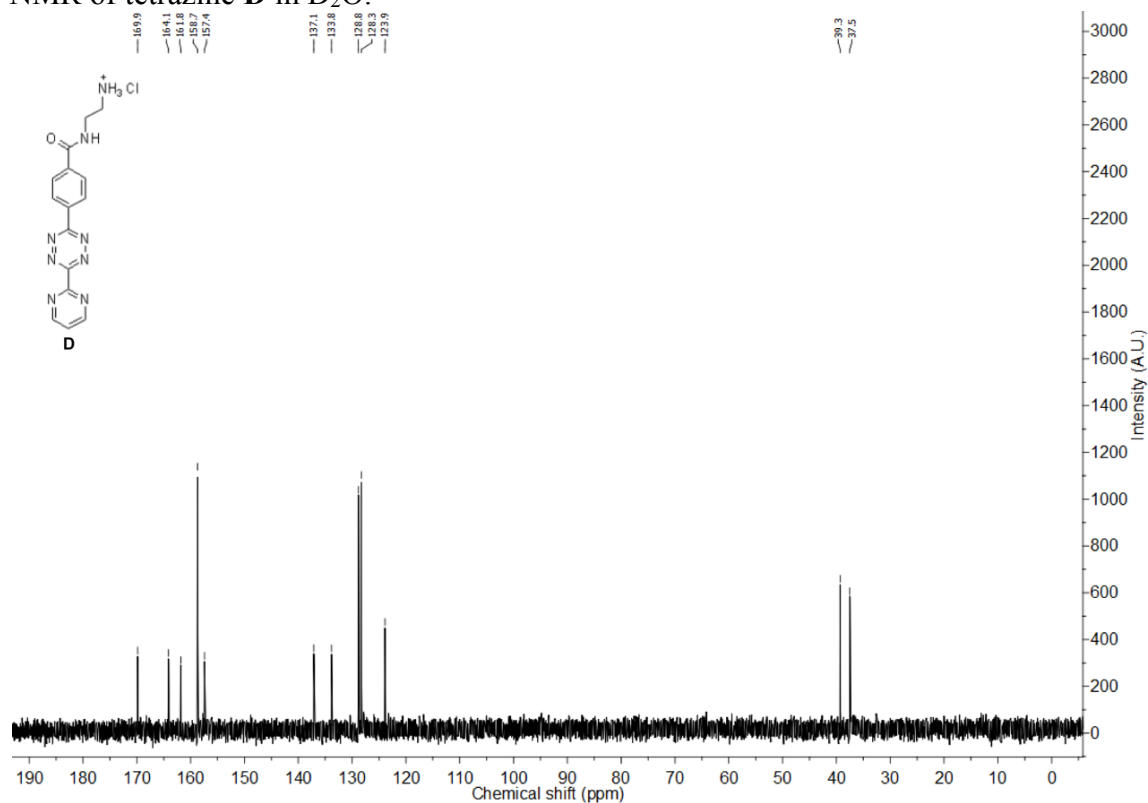
^{1}H-NMR of compound **2b** in *d$_6$*-DMSO. Trace amount of MeOH is present.
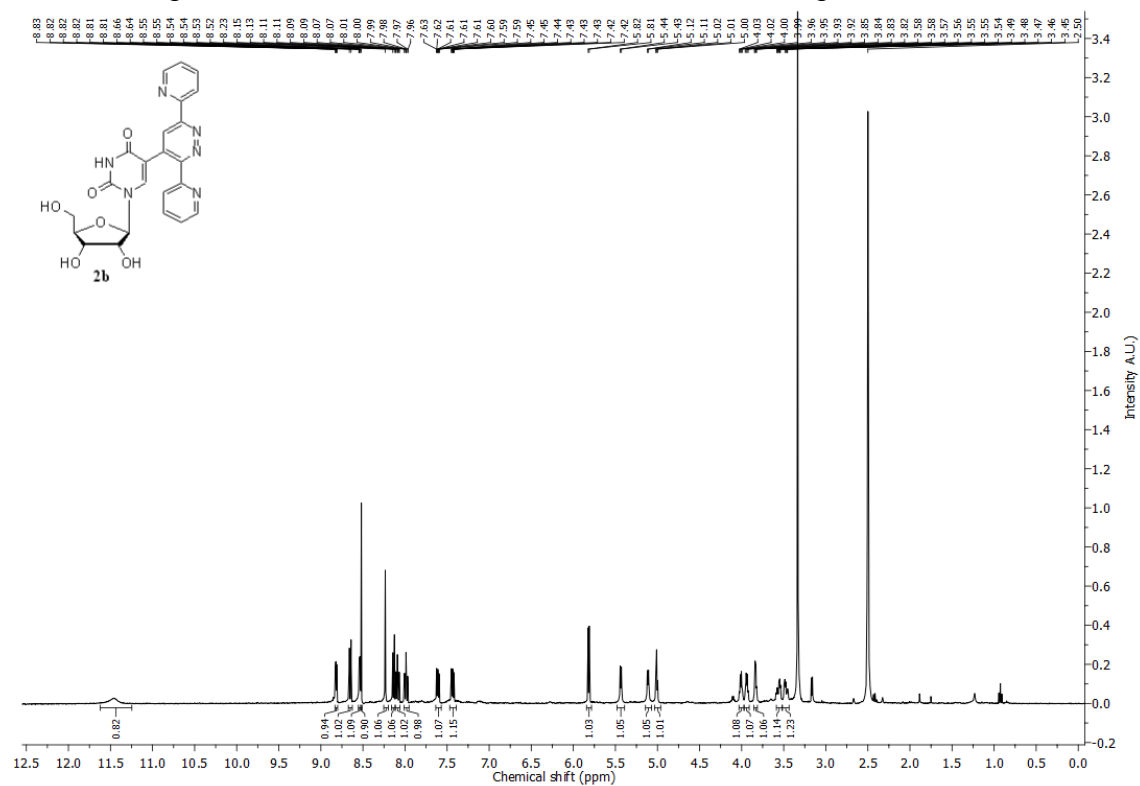
¹³C-NMR of compound **2b** in *d₆*-DMSO. Trace amount of MeOH is present.
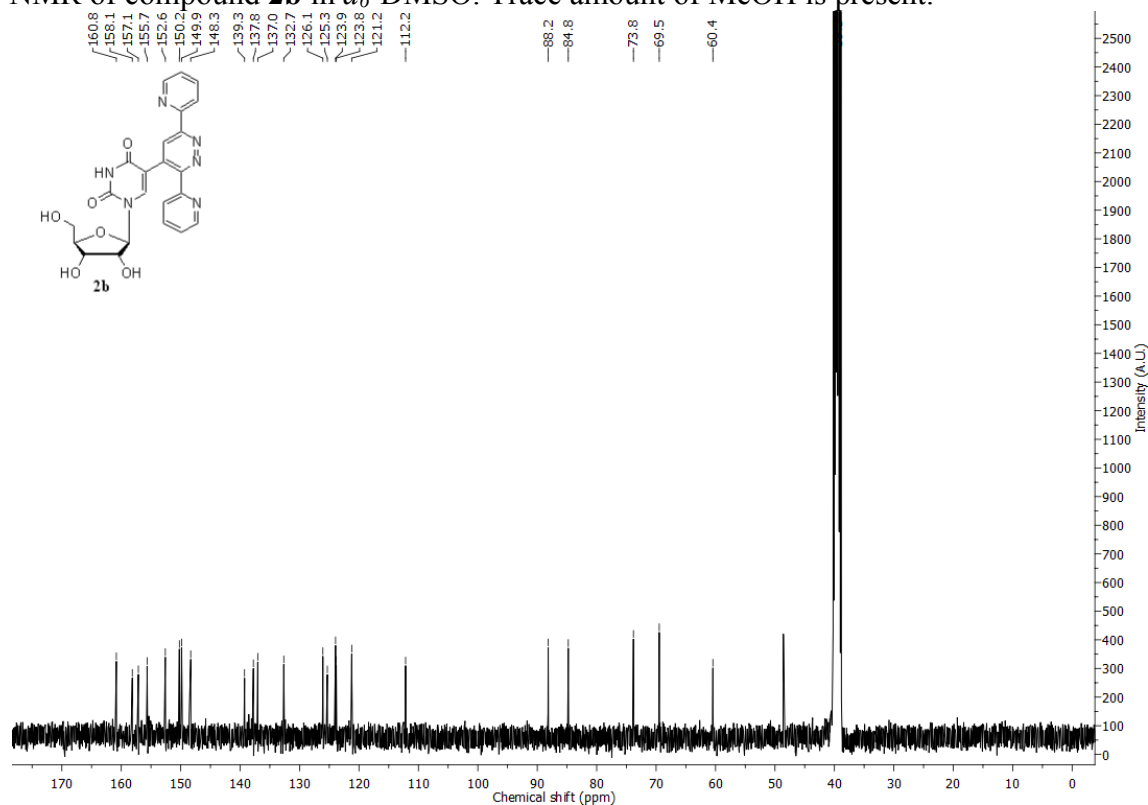
MALDI-TOF mass spectrum of oxidative Heck coupled RNA products corresponding to HPLC fractions[a]
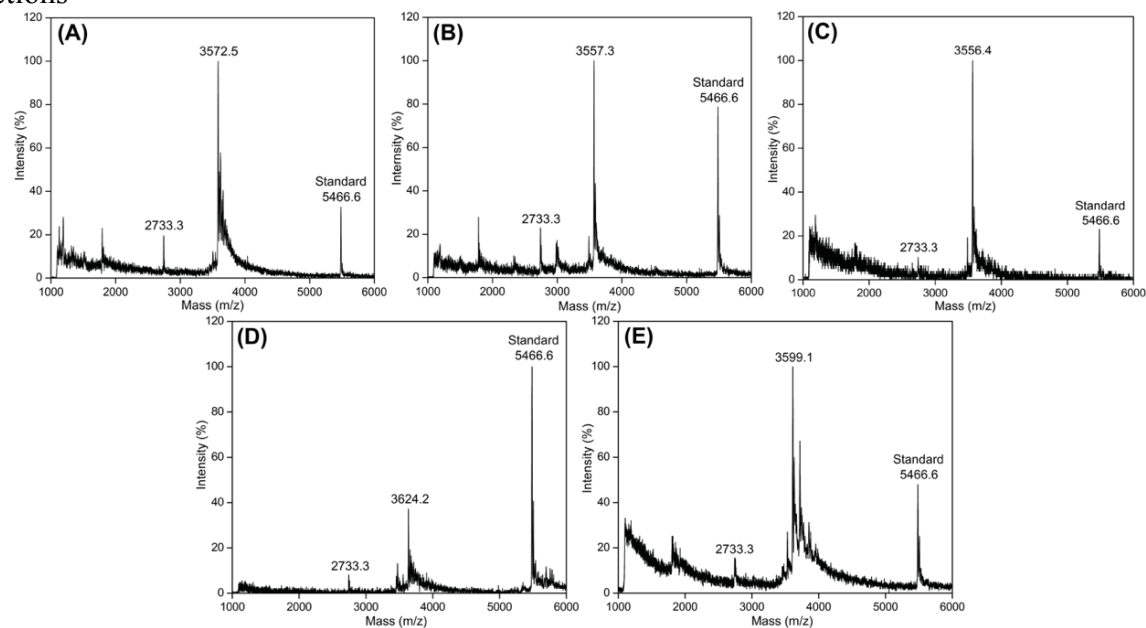
[a](A) **5a′**, (B) **6a′**, (C) **7a′**, (D) **8a′** and (E) **9a′** formed from large-scale reaction between transcript **4** and boronic acids **5-8**/boronic ester **9**. Spectrum is calibrated with respect to the +1 and +2 ion of an internal 18-mer DNA oligonucleotide standard (m/z for +1 and +2 ion are 5466.6 and 2733.3 respectively) (Table 2.2).

MALDI-TOF mass spectrum of IEDDA coupled RNA products[b]
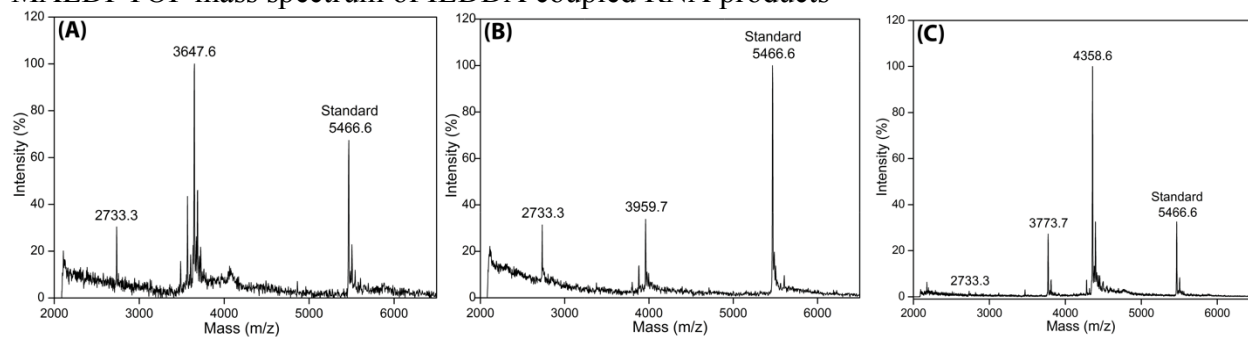
[b](**A**) **10a**, (**B**) **11a** and (**C**) **12a** formed from large-scale reaction between transcript **4** and tetrazines **10**, **11** and **12**, respectively. All spectra were calibrated with respect to the +1 and +2 ion of an internal 18-mer DNA ON standard (m/z for +1 and +2 ion are 5466.6 and 2733.3 respectively) (Table 2.2).

Chapter 3

Posttranscriptional Chemo-Enzymatic Tailoring of RNA Employing a Terminal Uridylyl Transferase, SpCID1

3.1 Introduction

RNA labeling technologies utilizing chemo-enzymatic methods, taking advantage of the promiscuity of enzymes to incorporate modified residues, have particularly simplified the tedious process of solid-phase based incorporation. Traditional approaches employing bacteriophage RNA polymerases (T7, SP6, T3) and modified nucleotide analogs indiscriminately label all the corresponding nucleobases on RNA on which the modification was made.[1-2] Such a whole body labeling approach would perturb the native structure of RNA making it unusable for further downstream biological manipulations. Therefore, approaches for site-specific labeling of RNA at the terminus or at an internal position which is not accessible by traditional enzymatic labeling techniques are highly desired.[3-9] Given this demand, developing an easy-to-use chemo-enzymatic 3′ terminal-labeling strategy would enable both site-specific internal and high-density terminal labeling of RNA. High density labeling of RNA at terminus will be highly desirable for attaching probes which can particularly be useful in techniques like single molecule RNA FISH[10-11] and RNA pull-down[12] whereas site-specific internal and terminal labeling with a single modified probe will be useful for constructing RNA labeled with biophysical probe of interest for structural analysis.[4,7,13-16] Development of a dual-labeling technique which combines both the strategies will lead to a highly beneficial two-in-one approach for labeling RNA for understanding its structure-function relationship as well as its cellular dynamics.

Terminal uridylyl transferase (TUT), or poly(U) polymerase (PUP) is a template independent DNA β-like family of eukaryotic polymerase, which has been involved in variety of cellular functions.[17-20] These enzymes function as a key player in epitranscriptomics wherein it adds homopolymer UMP tails at the 3′ end of RNA which signals RNA biogenesis, processing, degradation or stabilization.[18,21] In the family of TUTs, a cytoplasmic terminal uridylyl transferase from *Schizosaccharomyces pombe* known as caffeine-induced death suppressor 1 (SpCID1) has been shown as a key post-transcriptional RNA labeling enzyme known to regulate mRNA

populations.[22] This enzyme adds homopolymer UMP-tails to polyadenylated tails of mRNAs like act1 or urg1 in a template-independent fashion. Although the enzyme tolerates the addition of both UMP and AMP residues *in vitro*, the enzyme specifically functions as terminal uridylyl transferase *in vivo*. A previous preliminary assay form Jäschke and co-workers using SpCID1 and sugar-modified azide moieties had showed relatively poor to nil incorporation efficiency.[7] The crystal structure of the enzyme solved with UTP bound form reveals that this could be because the sugar moieties of UTP is used by the enzyme for recognizing the triphosphate (Figure 3.1).[23-24] Hence, modified residues incorporated at these locations will be less tolerated. However, a careful evaluation of the catalytically active site of the enzyme shows that the 5-position of uridine is positioned far from the amino acid residues. From these observations, we hypothesize that the enzyme might be promiscuous to 5-base modified uridine triphosphates. Also, we envision that careful tuning of reaction conditions and appropriate choice of modified nucleoside triphosphate could control the enzyme to incorporate single or multiple modifications at 3′ end of RNA, which then can be used for site-specific internal labeling or high-density terminal labeling of RNA.
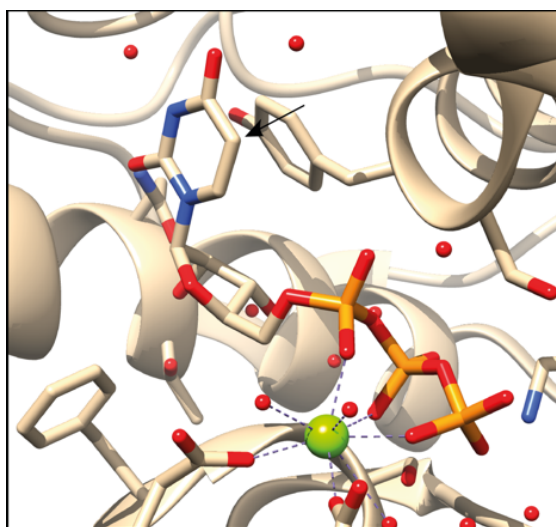
Figure 3.1. The crystal structure of SpCID1 bound to UTP (PDB 4FH5) generated with UCSF Chimera sofware.[23] Black arrow denotes the 5-postion of uridine triphosphate.

In order to achieve this, we came up with a design strategy wherein the enzyme, SpCID1 can be used to chemo-enzymatically functionalize an RNA of interest using click-modified nucleotide analogs, whereby giving access to single or multiple bioorthogonal tags on the 3′ end of RNA (Figure 3.2). The single modified RNA having the bioorthogonal tag can be used to synthesis a site-specific internal RNA by ligating the end-labeled RNA with another RNA in the

presence of a DNA splint and DNA/RNA ligases. The internal azide-labeled RNA could further be functionalized using copper-catalyzed or strain-promoted azide-alkyne cycloaddition (CuAAC or SPAAC) reactions.[25] The single or multiple click-labeled RNA can be also be conjugated employing click reaction to attach single or multiple functional probes/tags at the 3′ end of RNA.
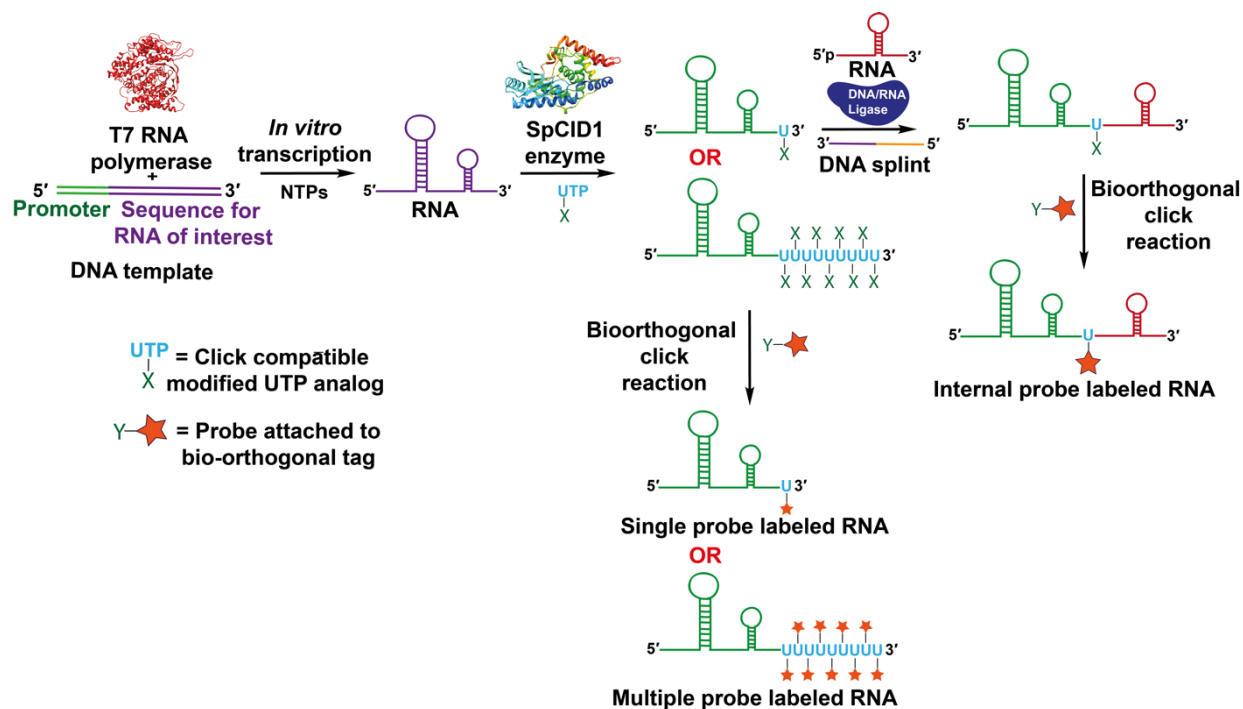
Figure 3.2. Design strategy for the incorporation modified nucleotides into RNA employing terminal uridylyl transferase, SpCID1. Structure of T7 RNA polymerase and SpCID1 adopted from PDB: 4RNP and 4FH3 using UCSF Chimera software.[23,26]

In this chapter, we clone SpCID1 gene from *S.pombe* fission yeast and express the recombinant protein in *E.coli*. We also synthesize 5-azido methyl uridine modified triphosphate which has previously been shown to metabolically label nascent RNA in cells.[25] Next, we investigate the ability of SpCID1 for incorporating single or multiple 5-base modified uridine residues at the 3′ end of RNA. Finally, we check the compatibility of the single azide-labeled RNA for functionalization employing SPAAC and CuAAC.

3.2 Results and Discussion

3.2.1 Cloning of SpCID1 gene into a bacterial expression vector

Synthesizing SpCID1 enzyme in-house requires the gene isolated from *S.pombe* and cloned into a bacterial expression vector. A truncated version of SpCID1 protein was chosen, encoding amino

acid residues 40-377 wherein the truncation does not compromise the enzyme activity, as evident from earlier literature reports.[27] We envisioned to insert the gene segment into pET28a vector using a restriction-free cloning approach (Figure 3.3).[28-29] In order to clone SpCID1 gene from *Schizosachromyces pombe*, the yeast was grown in YES media at 28 °C in 100 mL culture volume. The cells were pelleted down, and RNA was isolated using TRIzol reagent. Integrity of RNA was confirmed by electrophoresis in 1.5% agarose (Figure 3.4A). The RNA thus obtained was reverse transcribed using oligo dT primers to get the cDNA, corresponding to whole mRNA population. The cDNA was used for PCR amplifying SpCID1 gene using gene specific primers (Figures 3.4B, 3.6 and Table 3.1). A band corresponding to 1014 bp confirmed the amplification of SpCID1 gene of interest (Figures 3.3 and 3.4B) and further, the amplicon was isolated. In order to insert the SpCID1 gene amplicon into pET28a plasmid, overlapping regions corresponding to the empty plasmid were added at both 5′ and 3′ region of the gene by a second PCR using vector complementary primers, to get PCR amplicon 'megaprimer' which was confirmed by a band corresponding to 1065 bp on agarose gel (Figure 3.4C and Table 3.1). A gradient PCR (annealing temperatures 55, 58, 62 and 65 °C) was performed with 'megaprimer' and empty pET28a plasmid to insert SpCID1 gene into the vector. Also, as a negative control, PCR was performed without adding the 'megaprimer'. The empty pET28a plasmid in the reaction (usually methylated when isolated from bacteria) was degraded by Dpn1 restriction enzyme which specifically digests methylated recognition sites thereby leaving only the amplified clone in the PCR mixture. The PCR mixtures (for each annealing temperatures and control) was then individually transformed into chemically competent DH5α *E.coli* cells and then platted on LB agar with kanamycin (50 µg/mL) for 16-18 h. Seven bacterial colonies were obtained, for PCR corresponding to annealing temperature 65 °C. Other annealing temperatures yielded no colony. Also, no colony was observed in negative control confirming complete digestion of empty pET28a by Dpn1. Colonies obtained were picked and clone was isolated and amplified. Identity of positive clones were confirmed by PCR with T7 forward and reverse sequencing primers (Table 3.1). Six of seven colonies showed a positive clone and sequencing results for clones **3** with T7 sequencing primers, confirmed correct insertion of SpCID1 gene to get pET28a_SpCID1 plasmid (Figures 3.5 and 3.6).
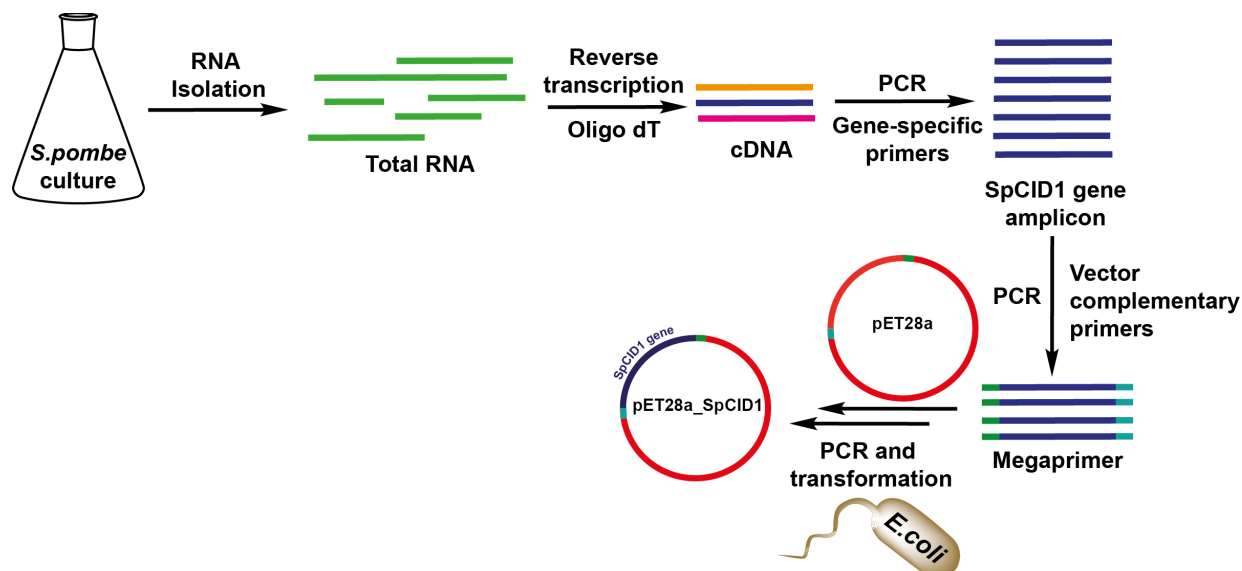
Figure 3.3. Work-flow for isolating and inserting SpCID1 gene into pET28a plasmid.
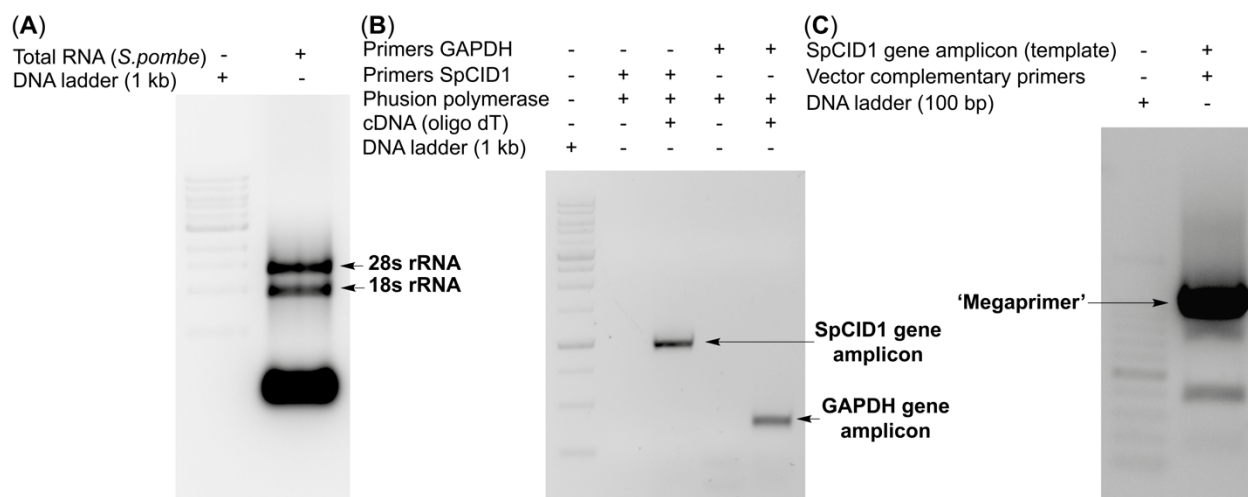
Figure 3.4. (**A**) Total RNA isolated from *S.pombe*. (**B**) Gene specific amplification of SpCID1 (1014 bp) and GAPDH (400 bp). (**C**) 'Megaprimer' amplification for SpCID1 (1065 bp). Refer, Figure 3.3 and Table 3.1.

Table 3.1. Sequence of RNA and primer DNAs used

DNA or RNA used	Sequence of oligonucleotide
5′ FAM-labeled RNA	5' (FAM) UUACCAUAGAAUCAUGUGCCAUACAUCA 3'
Gene specific forward primer, CID1FP	5' TCGCACAAGGAATTTACGAAGTTTTGC 3'
Gene specific reverse primer, CID1RP	5' GGCCTCCTCAAATAATGAATCATAAGGG 3'
GAPDH forward primer, GAPFP	5' TGGCAATTCCTAAGGTTGGTATTAACGGTT 3'
GAPDH reverse primer, GAPRP	5' CCGACAACGTACATGGGGGCGTC 3'
Vector complementary forward primer, CID1MF	5' GGTGCCGCGCGGCAGCCATATGTCGCACAAGGAATTTACGAAGTTTTGC 3'
Vector complementary reverse primer, CID1MR	5' CAGTGGTGGTGGTGGTGGTGCTCGAGCTAGGCCTCCTCAAATAATGAATCATAAGGG 3'
Sequencing primer T7 forward, T7FP	5' CGCGAAATTAATACGACTCACTATAGGG 3'
Sequencing primer T7 reverse, T7RP	5' GTTATGCTAGTTATTGCTCAGCGG 3'

Emerald PCR Mastermix	-	+	+	+	+	+	+	+	+	-
T7 (forward & reverse primer)	-	+	+	+	+	+	+	+	+	-
empty pET28a	-	-	-	-	-	-	-	-	+	-
Clone (DNA)	-	1	2	3	4	5	6	7	-	-
DNA ladder (100 bp)	-	-	-	-	-	-	-	-	-	+
DNA ladder (1 kb)	+	-	-	-	-	-	-	-	-	-
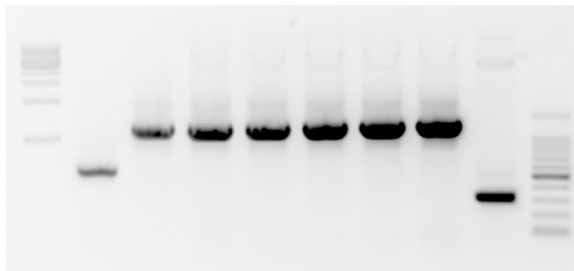
Figure 3.5. PCR for clones obtained from colonies 1-7 with T7 forward and reverse sequencing primer (Table 3.1). Positive clones 2-7 showed a band corresponding to 1274 bp.

5′...TCTCGATCCC<u>GCGAAATTAATACGACTCACTATAGGG</u>GAATTGTGAGCGG
ATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACC
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCT<u>GGTGCCGCGCG
GCAGCCATATG</u>**<u>TCGCACAAGGAATTTACGAAGTTTTGC</u>TATGAAGTGTAT
AATGAGATTAAAATTAGTGACAAAGAGTTTAAAGAAAAGAGAGCGGCAT
TAGATACACTTCGGCTATGCCTTAAACGAATATCCCCTGATGCTGAATTG
GTAGCCTTTGGAAGTTTGGAATCTGGTTTAGCACTTAAAAATTCGGATAT
GGATTTGTGCGTGCTTATGGATTCGCGCGTCCAAAGTGATACAATTGCG
CTCCAATTCTATGAAGAGCTTATAGCTGAAGGATTTGAAGGAAAATTTTT
ACAAAGGGCAAGAATTCCCATTATCAAATTAACATCTGATACGAAAAATG
GATTTGGGGCTTCGTTTCAATGTGATATTGGATTTAACAATCGTCTAGCT
ATTCATAATACGCTTTTACTTTCTTCATATACAAAATTAGATGCTCGCCTA
AAACCCATGGTCCTTCTTGTTAAGCATTGGGCCAAACGGAAGCAAATCA
ACTCTCCTTACTTTGGAACTCTTTCCAGTTATGGTTACGTCCTAATGGTT
CTTTACTATCTGATTCACGTTATCAAGCCTCCCGTCTTTCCTAATTTACTG
TTGTCACCTTTGAAACAAGAAAGATAGTTGATGGATTTGACGTTGGTT
TTGACGATAAACTGGAAGATATCCCTCCTTCCCAAAATTATAGCTCATTG
GGAAGTTTACTTCATGGCTTTTTTAGATTTTATGCTTATAAGTTCGAGCC
ACGGGAAAAGGTAGTAACTTTTCGTAGACCAGACGGTTACCTCACAAAG
CAAGAGAAAGGATGGACTTCAGCTACTGAACACACTGGATCGGCTGAT
CAAATTATAAAAGACAGGTATATTCTTGCGATTGAAGATCCTTTCGAGAT
TTCACATAATGTGGGTAGGACAGTTAGCAGTTCTGGATTGTATCGGATTC
GAGGGGAATTTATGGCCGCTTCAAGGTTGCTCAATTCTCGCTCATATCC
TAT<u>CCCTTATGATTCATTATTTGAGGAGGCC</u><u>TAGCTCGAGCACCACCACCA
CCACCACTG</u>AGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC
TGCTGCCA<u>CCGCTGAGCAATAACTAGCATAAC</u>CCCTTGGGGCCTCTAAAC...3′**

Figure 3.6. Representative sequence of SpCID1 gene in pET28a_SpCID1 plasmid (bold). Sequence corresponding to gene specific primers used for getting gene amplicon (blue). Sequence corresponding to vector complementary primer used for generating 'megaprimer'(underlined red). Sequence corresponding to sequencing primers T7 forward and reverse (underlined black) (See Table 3.1).

```
                   1         11        21        31        41        51        61        71        81        91
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) ATGTCGCACAAGGAATTTACGAAGTTTTGCTATGAAGTGTATAATGAGATTAAAATTAGTGACAAAGAGTTTAAAGAAAAGAGAGCGGCATTAGATACAC
Seq_forward            ATGTCGCACAAGGAATTTACGAAGTTTTGCTATGAAGTGTATAATGAGATTAAAATTAGTGACAAAGAGTTTAAAGAAAAGAGAGCGGCATTAGATACAC
Seq_reverse_complement ATGTCGCACAAGGAATTTACGAAGTTTTGCTATGAAGTGTATAATGAGATTAAAATTAGTGACAAAGAGTTTAAAGAAAAGAGAGCGGCATTAGATACAC

                   101       111       121       131       141       151       161       171       181       191
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) TTCGGCTATGCCTTAAACGAATATCCCCTGATGCTGAATTGGTAGCCTTTGGAAGTTTGGAATCTGGTTTAGCACTTAAAAATTCGGATATGGATTTGTG
Seq_forward            TTCGGCTATGCCTTAAACGAATATCCCCTGATGCTGAATTGGTAGCCTTTGGAAGTTTGGAATCTGGTTTAGCACTTAAAAATTCGGATATGGATTTGTG
Seq_reverse_complement TTCGGCTATGCCTTAAACGAATATCCCCTGATGCTGAATTGGTAGCCTTTGGAAGTTTGGAATCTGGTTTAGCACTTAAAAATTCGGATATGGATTTGTG

                   201       211       221       231       241       251       261       271       281       291
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) CGTGCTTATGGATTCGCGCGTCCAAAGTGATACAATTGCGCTCCAATTCTATGAAGAGCTTATAGCTGAAGGATTTGAAGGAAAATTTTTACAAAGGGCA
Seq_forward            CGTGCTTATGGATTCGCGCGTCCAAAGTGATACAATTGCGCTCCAATTCTATGAAGAGCTTATAGCTGAAGGATTTGAAGGAAAATTTTTACAAAGGGCA
Seq_reverse_complement CGTGCTTATGGATTCGCGCGTCCAAAGTGATACAATTGCGCTCCAATTCTATGAAGAGCTTATAGCTGAAGGATTTGAAGGAAAATTTTTACAAAGGGCA

                   301       311       321       331       341       351       361       371       381       391
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) AGAATTCCCATTATCAAATTAACATCTGATACGAAAAATGGATTTGGGGCTTCGTTTCAATGTGATATTGGATTTAACAATCGTCTAGCTATTCATAATA
Seq_forward            AGAATTCCCATTATCAAATTAACATCTGATACGAAAAATGGATTTGGGGCTTCGTTTCAATGTGATATTGGATTTAACAATCGTCTAGCTATTCATAATA
Seq_reverse_complement AGAATTCCCATTATCAAATTAACATCTGATACGAAAAATGGATTTGGGGCTTCGTTTCAATGTGATATTGGATTTAACAATCGTCTAGCTATTCATAATA

                   401       411       421       431       441       451       461       471       481       491
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) CGCTTTTACTTTCTTCATATACAAAATTAGATGCTCGCCTAAAACCCATGGTCCTTCTTGTTAAGCATTGGGCCAAACGGAAGCAAATCAACTCTCCTTA
Seq_forward            CGCTTTTACTTTCTTCATATACAAAATTAGATGCTCGCCTAAAACCCATGGTCCTTCTTGTTAAGCATTGGGCCAAACGGAAGCAAATCAACTCTCCTTA
Seq_reverse_complement CGCTTTTACTTTCTTCATATACAAAATTAGATGCTCGCCTAAAACCCATGGTCCTTCTTGTTAAGCATTGGGCCAAACGGAAGCAAATCAACTCTCCTTA

                   501       511       521       531       541       551       561       571       581       591
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) CTTTGGAACTCTTTCCAGTTATGGTTACGTCCTAATGGTTCTTTACTATCTGATTCACGTTATCAAGCCTCCCGTCTTTCCTAATTTACTGTTGTCACCT
Seq_forward            CTTTGGAACTCTTTCCAGTTATGGTTACGTCCTAATGGTTCTTTACTATCTGATTCACGTTATCAAGCCTCCCGTCTTTCCTAATTTACTGTTGTCACCT
Seq_reverse_complement CTTTGGAACTCTTTCCAGTTATGGTTACGTCCTAATGGTTCTTTACTATCTGATTCACGTTATCAAGCCTCCCGTCTTTCCTAATTTACTGTTGTCACCT

                   601       611       621       631       641       651       661       671       681       691
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) TTGAAACAAGAAAAGATAGTTGATGGATTTGACGTTGGTTTTGACGATAAACTGGAAGATATCCCTCCTTCCCAAAATTATAGCTCATTGGGAAGTTTAC
Seq_forward            TTGAAACAAGAAAAGATAGTTGATGGATTTGACGTTGGTTTTGACGATAAACTGGAAGATATCCCTCCTTCCCAAAATTATAGCTCATTGGGAAGTTTAC
Seq_reverse_complement TTGAAACAAGAAAAGATAGTTGATGGATTTGACGTTGGTTTTGACGATAAACTGGAAGATATCCCTCCTTCCCAAAATTATAGCTCATTGGGAAGTTTAC

                   701       711       721       731       741       751       761       771       781       791
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) TTCATGGCTTTTTTAGATTTTATGCTTATAAGTTCGAGCCACGGGAAAAGGTAGTAACTTTTCGTAGACCAGACGGTTACCTCACAAAGCAAGAGAAAGG
Seq_forward            TTCATGGCTTTTTTAGATTTTATGCTTATAAGTTCGAGCCACGGGAAAAGGTAGTAACTTTTCGTAGACCAGACGGTTACCTCACAAAGCAAGAGAAAGG
Seq_reverse_complement TTCATGGCTTTTTTAGATTTTATGCTTATAAGTTCGAGCCACGGGAAAAGGTAGTAACTTTTCGTAGACCAGACGGTTACCTCACAAAGCAAGAGAAAGG

                   801       811       821       831       841       851       861       871       881       891
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) ATGGACTTCAGCTACTGAACACACTGGATCGGCTGATCAAATTATAAAAGACAGGTATATTCTTGCGATTGAAGATCCTTTCGAGATTTCACATAATGTG
Seq_forward            ATGGACTTCAGCTACTGAACACACTGGATCGGCTGATCAAATTATAAAAGACAGGTATATTCTTGCGATTGAAGATCCTTTCGAGATTTCACATAATGTG
Seq_reverse_complement ATGGACTTCAGCTACTGAACACACTGGATCGGCTGATCAAATTATAAAAGACAGGTATATTCTTGCGATTGAAGATCCTTTCGAGATTTCACATAATGTG

                   901       911       921       931       941       951       961       971       981       991
Conservation       **********************************************************************************************
Desired(pET28a_SpCID1) GGTAGGACAGTTAGCAGTTCTGGATTGTATCGGATTCGAGGGGAATTTATGGCCGCTTCAAGGTTGCTCAATTCTCGCTCATATCCTATCCCTTATGATT
Seq_forward            GGTAGGACAGTTAGCAGTTCTGGATTGTATCGGATTCGAGGGGAATTTATGGCCGCTTCAAGGTTGCTCAATTCTCGCTCATATCCTATCCCTTATGATT
Seq_reverse_complement GGTAGGACAGTTAGCAGTTCTGGATTGTATCGGATTCGAGGGGAATTTATGGCCGCTTCAAGGTTGCTCAATTCTCGCTCATATCCTATCCCTTATGATT

                   1001      1011
Conservation       *******************
Desired(pET28a_SpCID1) CATTATTTGAGGAGGCCTAG
Seq_forward            CATTATTTGAGGAGGCCTAG
Seq_reverse_complement CATTATTTGAGGAGGCCTAG
```

Figure 3.7. Alignment of sequencing results of SpCID1 gene with the desired (pET28_SpCID1) plasmid with forward and reverse complement of sequencing results (obtained with T7 forward and reverse primers, Table 3.1). Sequence conservation denoted by asterix. Alignment of sequencing results performed using ClustalW and represented by UCSF Chimera software.

3.2.2 Expression of SpCID1 enzyme

The positive pET28a_SpCID1 plasmid was transformed in chemically competent BL21(DE3) *E.coli* cells and plated in LB agar having kanamycin and chloramphenicol. A colony was picked and inoculated in primary culture, of which 1% was inoculated to secondary culture. The secondary culture was incubated with varying concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.2-1 mM), either at 18, 28 or 37 °C. The cells were lysed and resolved by 12.5% SDS-PAGE to optimise the best IPTG and temperature for protein induction (Figure 3.8). At higher temperatures, SpCID1 protein went into inclusion bodies. Therefore, high expression of protein observed at these temperatures resulted in low protein content in supernatant after centrifugation.

From these results, the protein was observed to best express as a soluble form between 0.2-1 mM IPTG at 18 °C.

Optimisation of lysis and immobilized metal affinity chromatography (IMAC) purification was performed with Ni-IDA with increasing amount of imidazole concentration. Protein was lysed at 10 mM imidazole and was loaded onto Ni-IDA column. Subsequently the column was washed with 40-100 mM imidazole and protein eluted in 150-250 mM imidazole in increasing step gradients. For further purification, concentration of imidazole in wash buffer was fixed at 75 mM and elution buffer at 250 mM which yielded protein of good purity. The eluted protein was loaded and resolved by 12.5% SDS-PAGE (Figure 3.9A). The band corresponding to SpCID1 protein was also confirmed by western blot using Anti-6X His tag antibody (Figure 3.9B).

In order to express and purify SpCID1 enzyme in large-scale, pET28a_SpCID1 was transformed into chemically competent Rosetta™(DE3) *E.coli* cells and plated in LB agar having kanamycin and chloramphenicol. After 16 h, a colony was picked and inoculated in primary culture with both the antibiotics and the culture was grown to optical density at 600 nm (OD_{600}) of ~1.4. Further 2% of primary culture was inoculated in secondary culture and grown till OD_{600} of ~0.6 and cells were induced with 0.5 mM IPTG and incubated at 18 °C for 16 h. The bacterial cells were pelleted down and lysed by sonication. Lysate was centrifuged and the supernatant was used for IMAC using Ni-IDA beads. The protein was eluted at 250 mM imidazole, concentrated and buffer-exchanged to size-exclusion buffer using molecular weight cut-off columns. The protein was further purified by size-exclusion chromatography using fast protein liquid chromatography (FPLC), concentrated and dialyzed to make 50% glycerol stock (Figure 3.10). Concentration of SpCID1 enzyme obtained was quantified by Bradford assay using standard stocks of BSA. The expressed enzyme was used for the azide-modified nucleotide analogs. Also, a commercial acquired SpCID1 enzyme was used for the incorporation of alkyne- and vinyl-modified nucleotides on RNA. Concentration of commercial SpCID1 is reported in units (U) as calculated by the manufacturer.

pET28a_SpCID1 plasmid — — — — — + + + + + + + + + + | + + + + + + + + + + + + + + —
Histone plasmid (+ve control) + + + + — — — — — — — — — — — | — — — — — — — — — — — — — — —
Total lysate + — + — + — + — + — + — + — — | + — + — + — + — + — + — + — —
Supernatant after centrifugation — + — + — + — + — + — + — + + | — + — + — + — + — + — + — + —
IPTG (mM) — — 0.5 0.5 — — 0.2 0.2 0.5 0.5 1 1 — — | 0.2 0.2 0.5 0.5 1 1 — — 0.2 0.2 0.5 05 1 1 —
Temperature 37 37 37 37 — 18 18 18 18 18 18 18 18 28 28 | 28 28 28 28 28 28 37 37 37 37 37 37 37 37 —
Protein ladder (10-250 kDa) — — — — — + — — — — — — — — — | — — — — — — — — — — — — — — + (kDa)

Figure 3.8. Protein expression at different IPTG concentrations and temperatures. The total lysate after sonication and supernatant after centrifuging was loaded and resolved by 12.5% SDS-PAGE. Histone protein was used as a positive control.

(A)
Supernatant (after centrifugation) — — — + — — — — —
Cell pellet (after centrifugation) — — + — — — — — —
Uninduced lysate — + — — — — — — —
Wash buffer — — — — — — — — +
Elution buffer (fraction) — — — — 1 2 4 5 —
Protein Ladder (10-250 kDa) + — — — — — — — —

SpCID1 (41 kDa)

(B)
Supernatant (after centrifugation) — — + — —
Cell pellet (after centrifugation) — + — — —
Ni-IDA bead (before eluting) — — — + —
Eluted protein — — — — +
DNA ladder (12-225 kDa) + — — — —

SpCID1 (41 kDa)

Figure 3.9. (**A**) The protein was purified by Ni-IDA chromatography. Protein SpCID1 was eluted in 150-250 mM imidazole. (**B**) Western blot obtained after immunostaining for 6X His-tag using Anti His-tag mouse IgG2B monoclonal primary antibody and Goat Anti-Mouse IgG secondary antibody conjugated to horseradish peroxidase.

(A) Molecular Mass(kDa) 440 158 66 44 17
SpCID1 fraction (41 kDa)

(B) Fraction — — 1 2 3 4 5 6 7 8
Protein ladder + — — — — — — — — —

SpCID1 (41 kDa)

Figure 3.10. (**A**) Size exclusion chromatogram of IMAC purified SpCID1 enzyme as compared with protein standards. (**B**) Fraction corresponding to SpCID1 was loaded and resolved by 12.5% SDS-PAGE and was further stained with Coomassie blue.

3.2.3 Click-compatible modified UTP analogs

Click-modified nucleotide analogs for performing terminal uridylation reaction

To perform terminal uridylation employing 5-base modified UTPs, we envisioned to use a series of click-compatible uridine triphosphate analogs previously reported from our group (Figure 3.11). Azide modified triphosphate, AMUTP, APUTP and ATUTP with increasing linker length having a methyl, propyl and tetraethylene glycol linker between the azide group and nucleobase was used for terminal uridylation[25,30-31]. An alkyne-labeled triphosphate, EUTP and alkyne with increasing linker length namely, ODUTP and 5-vinyl uridine triphosphate, VUTP synthesized in Section 2.2.1 were also used.[32-33]
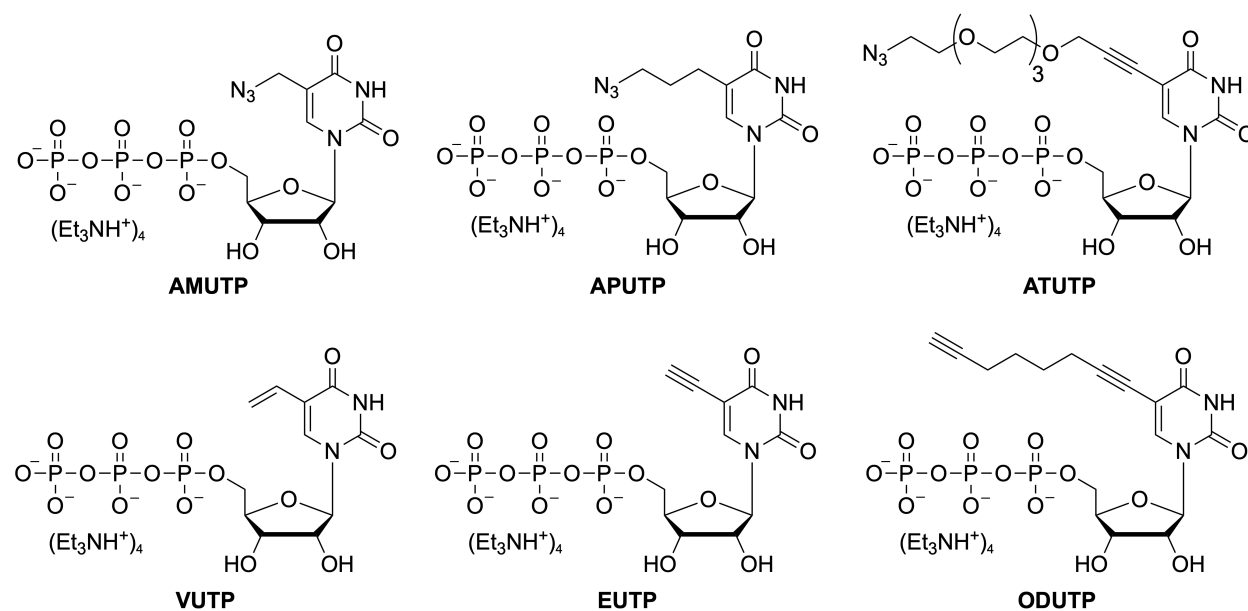

Figure 3.11. Click compatible modified UTP analogs used for terminal uridylation.[25,30-33]

Synthesis of 5-Azidomethyl uridine triphosphate

5-Azidomethyl uridine triphosphate (**AMUTP**) was synthesized by previously reported procedures (Scheme 3.1).[25,34] Uridine nucleoside was acetyl protected at 2′ and 3′ hydroxyl group and further this protected nucleoside (**a**) was reacted with paraformaldehyde in presence KOH to obtain acetyl-protected 5-hydroxymethyl uridine (**b**). The precursor nucleoside (**b**) was reacted with trimethylsilyl chloride and subsequently with sodium azide in a single-pot reaction to obtain acetyl-protected azidomethyl uridine (**c**). The acetyl group was deprotected by acetic acid to obtain 5-azidomethyl uridine nucleoside (**d**, **AMU**). The azide-modified nucleoside was

triphosphorylated using $POCl_3$ and bis-tributylammonium pyrophosphate to get 5-azidomethyl uridine triphosphate (**e**, **AMUTP**).
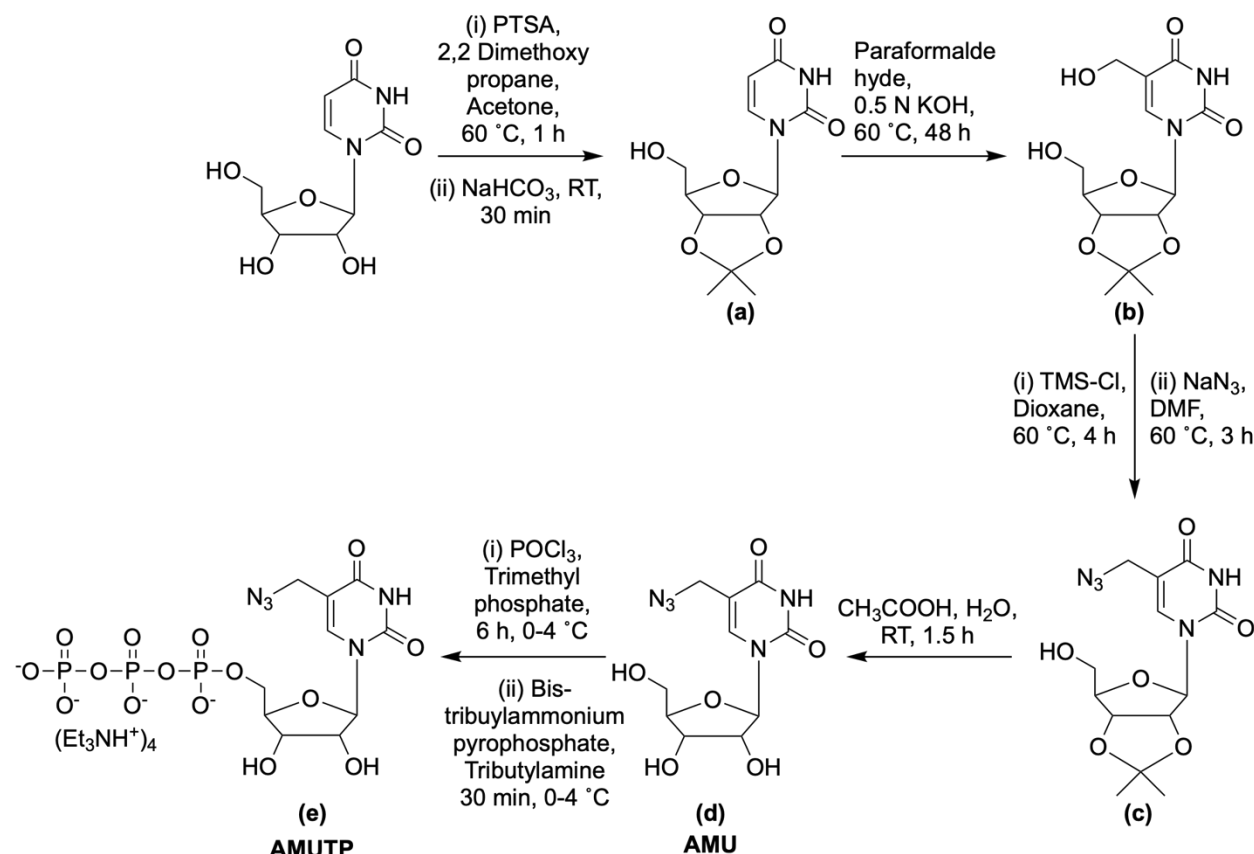
Scheme 3.1. Synthesis of AMUTP.[25,34]

3.2.4 Enzymatic incorporation of modified nucleotides

To investigate the ability of SpCID1 enzyme to incorporate 5-position base-modified UTP analogs, a 5′-FAM labeled RNA oligonucleotide was used as read-out to visualize RNA on gel by fluorescence (Table 3.1). In-house expressed SpCID1 (10.25 pmol, 0.5 μM) or commercial SpCID1 (2U) was incubated with 500 μM of UTP or modified triphosphate with 5′-FAM labeled RNA (10 μM) in a final volume of 20 μL. Aliquots of reaction were taken at different time points and 12.5 pmol of RNA samples were loaded on 20% denaturing polyacrylamide gel. The gel was imaged using Typhoon gel scanner at FAM wavelength. A biphasic incorporation corresponding to processive and distributive enzyme addition was observed for natural UTP (Figures 3.12 and 3.13).[27,35] Distributive enzyme addition involves the dissociation of enzyme from substrate RNA after every UMP addition giving rise to RNA of shorter length.[36] Processive enzyme addition

results in longer RNA due to multiple reactions before enzyme dissociates from the RNA. This is thought to be a result of increase in binding affinity of enzyme (20 times) upon addition of 6-7 uridine residues.[35] Rewardingly all the nucleotide analogs served as good substrate for the enzyme. Terminal uridylation with all the modified triphosphate prominent displayed distributive enzyme incorporation (Figures 3.12 and 3.13). Minimally perturbing modified nucleotide analogs like AMUTP, EUTP and VUTP showed small amount of processive incorporation visualized upon changing the exposure of the gel.

SpCID1 (0.5 µM)	+	-	+	+	+	+	+	+	+	+	+	+	+	+
5' FAM labeled RNA (10 µM)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UTP (500 µM)	-	-	+	+	+	-	-	-	-	-	-	-	-	-
AMUTP (500 µM)	-	-	-	-	-	+	+	+	-	-	-	-	-	-
APUTP (500 µM)	-	-	-	-	-	-	-	-	+	+	+	-	-	-
ATUTP (500 µM)	-	-	-	-	-	-	-	-	-	-	-	+	+	+
Time (min)	30	30	5	15	30	5	15	30	5	15	30	5	15	30
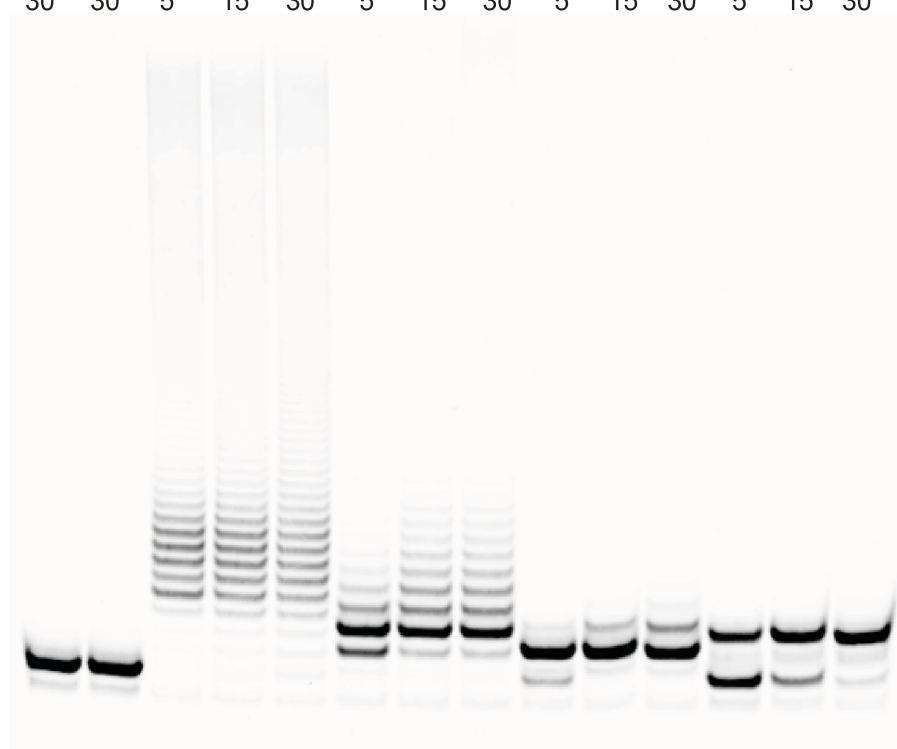
Figure 3.12. Incorporation of AMUTP, APUTP and ATUTP as compared to natural UTP.

Commercial SpCID1 (2 U)	+	-	+	+	+	+	+	+	+	+	+	+	+	+
5' FAM-labeled RNA (10 µM)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UTP (500 µM)	-	-	+	+	+	-	-	-	-	-	-	-	-	-
VUTP (500 µM)	-	-	-	-	-	+	+	+	-	-	-	-	-	-
EUTP (500 µM)	-	-	-	-	-	-	-	-	+	+	+	-	-	-
ODUTP (500 µM)	-	-	-	-	-	-	-	-	-	-	-	+	+	+
Time (min)	30	30	5	10	30	5	10	30	5	10	30	5	10	30
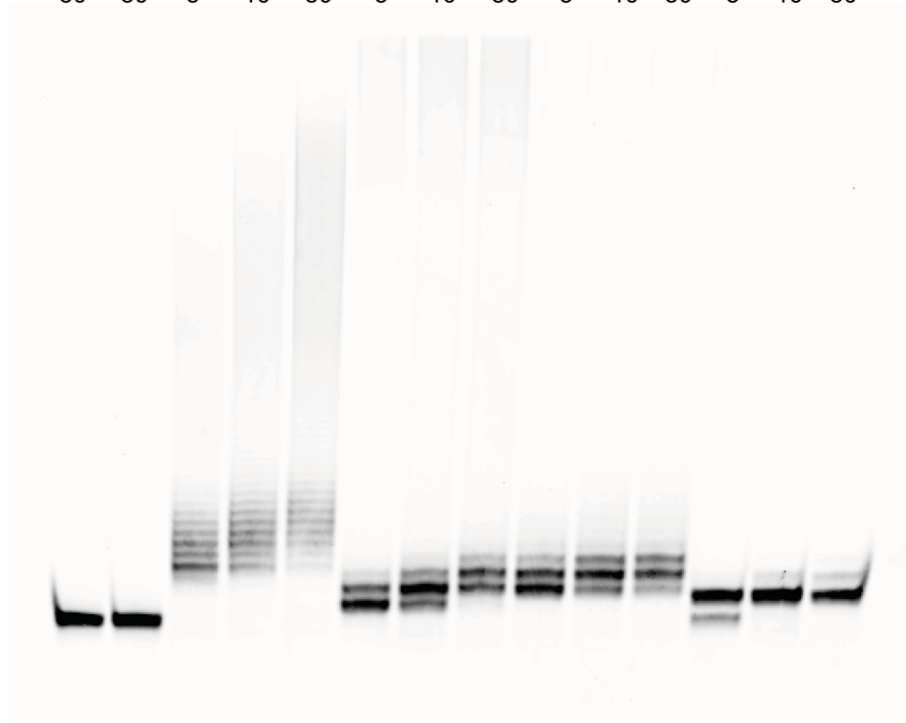
Figure 3.13. Incorporation of VUTP, EUTP and ODUTP as compared to natural UTP.

3.2.5 Controlling degree of incorporation of modified nucleotides

In order to understand the variation in incorporation efficiency upon increasing enzyme and modified UTP concentration, the number of modifications incorporated was monitored by varying enzyme concentration keeping triphosphate concentration constant and vice versa. While changing commercial SpCID1 concentration from 0.5 to 8 U, the percentage of glycerol was kept constant. This is because an increased amount of glycerol in reaction can affect the overall incorporation efficiency, possibly due to solvent friction. As speculated, we observed an increase in degree of incorporation with increase in the concentration of the enzyme (Figure 3.14). Higher enzyme concentration (8 U) resulted in majorly 3-4 modified UMP incorporation in 15 and 30 min. Therefore, by increasing the concentration of the enzyme, one can introduce multiple modified analogs at the 3′ end of RNA, which could be advantageous in high density labeling for signal amplification.

EUTP (0.5 mM)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5' FAM-labeled RNA (10 µM)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Commercial SpCID1 (U)	-	0.5	0.5	0.5	1	1	1	2	2	2	4	4	4	8	8	8
Time (min)	30	5	10	30	5	10	30	5	10	30	5	10	30	5	10	30
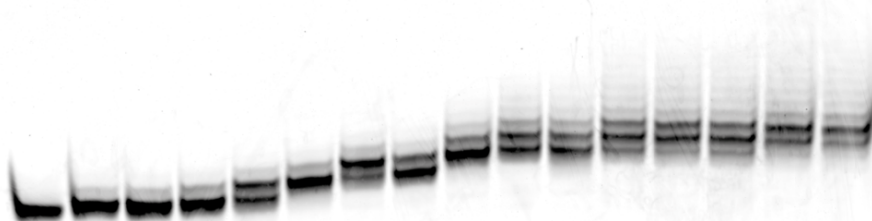
Figure 3.14. Effect of SpCID1 concentration on terminal uridylation reaction.

In order to the study changes in incorporation efficiency upon increasing the concentration of modified substrate, EUTP concentration was varied from 0.5 to 4 mM (50 to 400 equiv with respect to the concentration of RNA). A saturation in the degree of incorporation was observed till 2 mM of EUTP (Figure 3.15). However, upon increasing the concentration to 400 equiv (4 mM) of EUTP, a sudden decrease in the degree of incorporation corresponding to an inhibition of enzyme activity was observed. Limiting the modified triphosphate or the enzyme concentration should also decrease the degree of incorporation. Therefore, these assays hints that one could control the degree of 3′ incorporation of modified residues either by varying enzyme or triphosphate concentration and thereby effectively assisting in addition of single or multiple modified residues at the 3′ end of RNA.

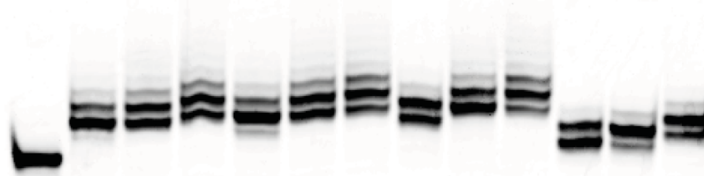	Commercial SpCID1 (U)	+	+	+	+	+	+	+	+	+	+	+	+	+
5' FAM-labeled RNA (10 µM)	+	+	+	+	+	+	+	+	+	+	+	+	+
EUTP (mM)	-	0.5	0.5	0.5	1	1	1	2	2	2	4	4	4
Time (min)	30	5	10	30	5	10	30	5	10	30	5	10	30

Figure 3.15. Effect of EUTP concentration on terminal uridylation reaction.

3.2.6 Single incorporation of modified nucleotides into 3′ end of RNA by varying the enzyme and triphosphate stoichiometry

A single incorporation of click compatible UTP analogs on 3′ end of RNA is particularly desirable for conjugating biophysical probes on RNA for structure-analysis (Figure 3.16). For achieving single modification of RNA using alkyne- and vinyl-modified uridine triphosphates, the terminal

uridylation reaction was performed with varying concentration of SpCID1 in a final volume of 100 μL. When the commercial SpCID1 was used, concentration was varied in the range of 2-5 U. A near saturation of single incorporation was achieved in 30 min for all alkyne- and vinyl-modified uridine triphosphates (Figure 3.17). Therefore, controlling enzyme concentration influences the overall catalytic cycles available on RNA substrate and thereby yielding mainly the single 3′ end modified RNA.
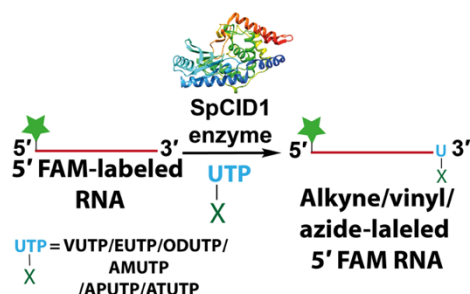

Figure 3.16. Incorporation of a single alkyne, vinyl or azide tag at the 3′ end of RNA. Structure of SpCID1 adopted from PDB:4FH3 using UCSF Chimera software.[23]
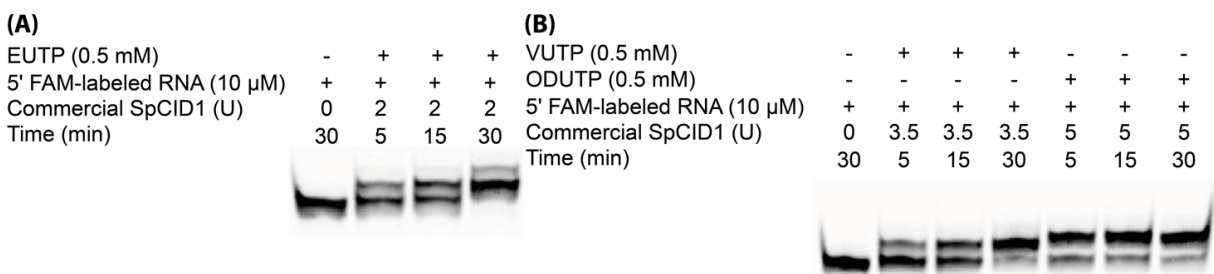

Figure 3.17. Saturation of single incorporation at end of 30 min for VUTP, EUTP and ODUTP.

To obtain a better controlled incorporation, we sought to vary both stoichiometry of enzyme and triphosphate. Single incorporation of azide-modified uridine analogs were optimized by varying the concentration of nucleotide and the concentration of the enzyme in a 20 μL reaction volume (Figure 3.18). When controlling the degree of incorporation, the nucleotide concentration used for reaction, was typically varied from 1-5 equivalence as compared to that of RNA thereby specifically limiting the incorporation to a single modification. The RNA (200 pmol, 10 μM) was incubated with 1 μL of SpCID1 (3.42 pmol or 10.25 pmol). A single incorporation of azide at the 3′ end of RNA was observed within 15 min for AMUTP and APUTP and 30 min for ATUTP. Further, large-scale synthesis of AMU-labeled RNA was carried out. The minimally perturbing azide handle on AMU is a versatile tag unlike alkyne and is compatible for SPAAC, CuAAC and

Staudinger ligation reaction. The azide incorporated RNAs were synthesized in 1 nmol scale in batches (5 x 200 pmol reactions) and was purified by RP-HPLC. The reaction volume was kept constant at 20 μL to maintain effective heat transfer to the reaction vile for consistency in incorporation efficiency. A typical reaction on 1 nmol scale of 5′ FAM-labeled RNA yielded 0.89, 0.95 and 0.76 nmol of RNA products labeled with AMU, APU and ATU. Mass of RNA products obtained were confirmed by direct infusion in ESI-MS (Table 3.2).
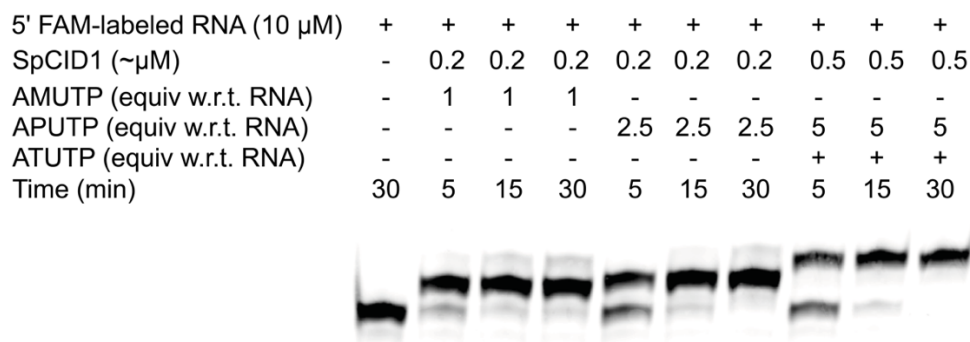
Figure 3.18. Single incorporation of azide tags at 3′ end of RNA using AMUTP, APUTP and ATUTP.

3.2.7 SPAAC reaction on 3' azide-labeled RNA

In order to check if the 3′ azide-labeled RNA is compatible for SPAAC reaction. AMU-labeled 5′ FAM RNA (0.3 nmol, 10 μM) was incubated with Cy3-DBCO in a 30 μL reaction volume at 37 °C for 2 h. RNA was precipitated and clicked product was loaded on 20% denaturing polyacrylamide gel and was visualized in FAM and Cy3 channels using gel scanner. The gel showed almost complete conversion of AMU-labeled 5′ FAM RNA. A typical 0.3 nmol reaction yielded 0.15 nmol of isolable product after precipitation step (Figure 3.19). Mass of the product was confirmed by ESI-MS (Table 3.2).

Table 3.2. Mass and yield of large-scale isolated AMU-, APU- and ATU-labeled 5′ FAM RNA and click conjugated products for AMU-labeled 5′ FAM RNA

Product RNA	Calcd. Mass	Found Mass	Isolated yield (nmole)	Isolated yield (%)
AMU-labeled 5′ FAM RNA	9750.0	9749.6	0.89[a]	89
APU-labeled 5′ FAM RNA	9778.1	9777.6	0.95[a]	95
ATU-labeled 5′ FAM RNA	9950.3	9949.8	0.76[a]	76
Cy3-FAM RNA	10733.2	10732.6	0.15[b]	51
Alexa 594-FAM RNA	10509.9	10509.1	0.16[b]	52

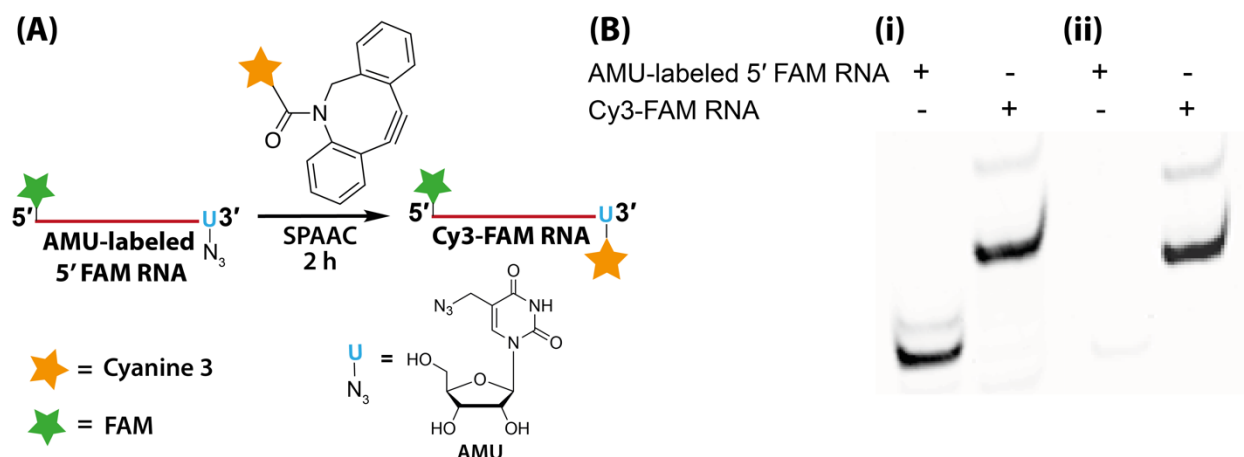Isolated yields for reactions performed on [a]1 and [b]0.3 nmol scale.

Figure 3.19. (**A**) Strain-promoted azide-alkyne cycloaddition reaction between AMU-labeled 5′ FAM RNA and Cy3-DBCO. (**B**) AMU-labeled 5′ FAM RNA and Cy3-FAM RNA imaged by Typhoon gel scanner in (i) FAM and (ii) Cy3 wavelength.

3.2.8 CuAAC reaction on 3′ azide-labeled RNA

In order to perform copper-catalyzed click reaction on azide RNA, $CuSO_4$ in presence of THPTA ligand was reduced to Cu(I) using sodium ascorbate. Further, Alexa 594 alkyne and AMU-labeled 5′ FAM RNA (0.3 nmol, 10 µM) was added to a final volume of 30 µL. The reaction was incubated at 37 °C for 2 h and further the RNA was precipitated to obtain the clicked product. The product was loaded on 20% denaturing polyacrylamide gel and was visualized in FAM and Alexa 594 wavelength using gel scanner (Figure 3.20). The gel showed almost complete conversion of AMU-labeled 5′ FAM RNA. A typical 0.3 nmol reaction yielded 0.16 nmol of isolable product. Mass of the product was confirmed by ESI-MS (Table 3.2).
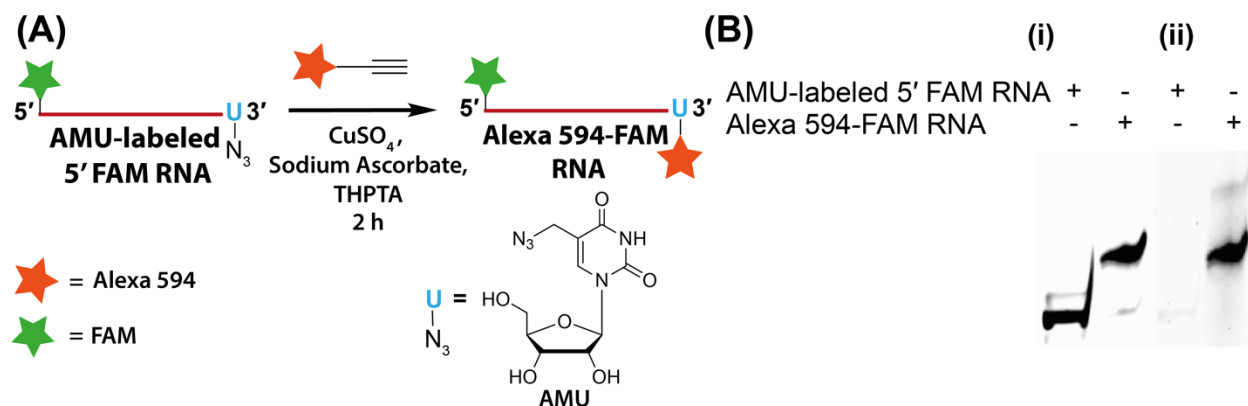

Figure 3.20. (**A**) Copper-catalyzed azide alkyne cycloaddition reaction between AMU-labeled 5′ FAM RNA and Alexa 594 alkyne. (**B**) AMU-labeled 5′ FAM RNA and Alexa 594-FAM RNA imaged by Typhoon gel scanner in (i) FAM and (ii) Alexa wavelength.

3.3 Conclusions

A novel chemo-enzymatic labeling strategy for incorporating modified nucleotide analogs on RNA has been developed by utilizing terminal uridylyl transferase, SpCID1. To achieve this, the SpCID1 gene was successfully cloned into a bacterial expression vector and the protein was expressed and purified. Analysis of the reported crystal structure of SpCID1 enzyme[23] indicated its potential for incorporating 5-nucleobase modified uridine triphosphate analogs on RNA. By taking this lead, we have experimentally demonstrated the ability of SpCID1 enzyme to incorporate a single or multiple azide-, alkyne- and vinyl-modified UMP(s) at the 3′ end of RNA with excellent efficiency, by varying stoichiometry of enzyme and triphosphate. The methodologies presented herein are indeed useful for the incorporation of clickable handles on RNA by SpCID1 which is compatible for copper-catalyzed and strain-promoted azide-alkyne cycloaddition reactions. Being the first report of labeling RNA with clickable handles by employing SpCID1, the methodology presented here opens door for exploring terminal uridylation as a new technology for signal amplification via high-density end-labeling and constructing RNA-labeled with suitable biophysical reporters.

3.4 Experimental Section

3.4.1 Materials

Schizosachromyces pombe was kindy gifted by Dr. Devyani Haldar (CDFD, Hyderabad). Luria Broth, agar, yeast extract was purchased from HiMedia. Commercial SpCID1, Phusion® High-Fidelity DNA Polymerase, Dpn1 was acquired from New England Biolabs. EmeraldAmp GT PCR master mixa and DNA ladders were purchased from Takara. POCl₃ was purchased from Acros Organics and was distilled before use. DNA oligonucleotides (ONs) were purchased from Integrated DNA Technologie, Inc. and Sigma Aldrich. NTPs, T7 RNA polymerase, ribonuclease inhibitor (RiboLock), Alexa 594 alkyne, Superscipt III, TRIzol were obtained from Thermo Fischer Scientific. PrimeSTAR GXL DNA Polymerase was purchased from TakaraBio. DH5α, BL21 (DE3) and Rosetta™(DE3) competent cells were kindly gifted by Dr. Thomas J. Pucadyil (IISER Pune). pET28a plasmid was acquired from Novagene. The modified triphosphates used, APUTP, ATUTP, EUTP and ODUTP was kindly provided by Dr. Arun Tanpure and Dr. Anupam Sawant. Cy3-DBCO was obtained from Click Chemistry Tools. Anti-6X His tag® antibody and HRP Goat Anti-Mouse (IgG) secondary antibody from abcam were kindly gifted by Dr. Siddhesh

S. Kamat. Reagents for buffer solutions were obtained from Sigma Aldrich (Bio Ultra Grade). Vivaspin 20 (10 kDa) molecular-weight cut off columns were acquired from GE Healthcare. NucleoSpin® Gel and PCR Clean-up kit was purchased from Macherey-Nagel. QIAprep Spin Miniprep Kit was purchased form Qiagen.

3.4.2 Instrumentation

Fluorescent RNA was resolved by polyacrylamide on OWL S4S sequencing gel electrophoresis instrument and was imaged using Typhoon TRIO+ Variable mode Imager. RNA was quantified by measuring absorbance in UV-2600 Shimadzu or NanoDrop™ 2000c spectrophotometer. Bacterial cultures were incubated in New Brunswick Innova 4230 Refrigerated Incubator Shaker and Thermo Scientific Forma Orbital Shaker 480. Mass analysis for AMUTP was performed using using Applied Biosystems 4800 Plus MALDI-TOF/TOF analyser. NMR of AMUTP was recorded in Bruker AVANCE III HD ASCEND 400 MHz spectrometer. Size exclusion chromatography was performed on Biorad NGC™ Chromatography system on a HiLoad 16/600 Superdex 200 preparative column. ESI-MS mass analysis of RNA was performed using Waters SYNAPT G2-Si Mass Spectrometry instrument in negative mode. HPLC analysis was done on Agilent Technologies 1260 Infinity HPLC. Reversed-phase (RP) flash chromatographic (C18 RediSep*Rf* column) purifications were performed using a Teledyne ISCO, Combi Flash Rf.

3.4.3 Cloning of SpCID1 gene, its expression and purification
Yeast culture and total RNA isolation

Schizosachromyces pombe was grown in a primary culture in YES media (15 mL) after which it was inoculated to a secondary culture in the same media (100 mL) and grown for 24 h at 28 °C until OD_{600} of the culture medium was roughly 1.4. The fission yeast cells were then pelleted down and small quantity of pelleted culture was transferred into a micro-centrifuge tube. To the tube, TRIzol (1 mL) was added and the pellet was homogenized. Chloroform (0.2 mL) was added to the homogenized solution, mixed properly and was centrifuged at 12,000 rpm at 4 °C. The upper aqueous layer containing the RNA was carefully transferred to a separate tube, followed by the addition of 0.5 mL isopropanol. The solution was incubated at -20 °C for 1 h followed by centrifugation at 12,000 rpm at 4 °C. The pellet was washed with 1 mL of 75% ethanol in nuclease free water and again centrifuged. The RNA pellet thus obtained was resuspended in fresh nuclease

free water. The integrity of RNA was confirmed by 1.5% agarose gel electrophoresis in 1X TAE buffer (Figures 3.3 and 3.4A).

cDNA synthesis and amplification of SpCID1 gene

The total RNA obtained (1.9 μg) was incubated with oligo $(dT)_{20}$ (5 μM) or gene specific reverse primers (0.05 μM) and dNTPs (1 mM) in nuclease free water in a final volume of 10 μL. The cocktail was heated at 65 °C for 5 min in a PCR machine and flashed cooled to 4 °C in 30 s. To the cocktail was added, $MgCl_2$ (5 mM), DTT (0.01 mM), RNAse OUT (2 U/μL), Superscript III (10 U/μL) and 2 μL of 10X RT buffer (200 mM Tris-HCl, 500 mM KCl, pH 8.4) in a final volume of 20 μL. The reverse transcription cocktail was heated to 25 °C for 5 min, 50 °C for 50 min. Further the enzyme was heat denatured at 85 °C for 5 min and then cooled to 4 °C. To the reaction cocktail, 1 μL of *E.coli* RNase H (2 U) was added and incubated for 20 min at 37 °C.

In order to amplify the SpCID1 gene from cDNA, gene specific primers CID1FP and CID1RP was used. GAPFP and GAPRP primers were used for amplifying GAPDH, which was used as a positive control. For performing a PCR for amplifying the gene, 2 μL of cDNA was incubated with forward and reverse primers (0.5 μM), dNTPs (0.2 mM), Phusion® High-Fidelity DNA Polymerase (0.1 U/μL) in 1X Phusion HF buffer in a final volume of 20 μL. PCR conditions: heat denaturation at 98 °C for 3 min, 35 cycles of (denaturing: 98 °C for 20 s, annealing: 58 °C for SpCID1 and 62.2 °C for GAPDH for 30 s, extension: 72 °C for 40 s), final extension at 72 °C for 5 min, 4 °C hold.

The products were visualized on 1.5% agarose gel run in 1X TAE buffer. Further, the PCR product corresponding to SpCID1 gene was isolated using NucleoSpin® Gel and PCR Clean-up kit to obtain 2.65 μg of gene specific amplicon (Figures 3.3, 3.4B and Table 3.1).

PCR amplification of 'Megaprimer'

In order to add overlapping regions corresponding to empty pET28a plasmid to the 3′ and 5′ of SpCID1 gene, vector complementary primers CID1MF and CID1MR (0.6 μM) were incubated with SpCID1 gene amplicon (100 ng), dNTPs (0.2 mM), PrimeSTAR GXL DNA polymerase (0.025 U/μL) in 1X PrimeSTAR GXL buffer in a final volume of 50 μL. PCR conditions: heat denaturation at 98 °C for 3 min, 35 cycles of (denaturing: 98 °C for 30 s, annealing: 55 °C for 30 s, extension: 68 °C for 1.33 min), final extension at 68 °C for 10 min, 4 °C hold. The resulting

amplicon was visualized in 1.5% agarose run in 1X TAE buffer. Further the product amplicon was isolated using Mega clear PCR clean-up kit to obtain 1.84 µg of 'megaprimer' (Figures 3.3, 3.4C and Table 3.1).

Restriction-free cloning using PCR

Gradient PCR was set for inserting the gene into pET28a vector. The 'megaprimer' (500 ng), pET28a template (100 ng), dNTPs (0.2 mM), and PrimeSTAR GXL DNA polymerase (0.05 U/µL) was incubated in 1X PrimeSTAR GXL buffer in a final volume of 25 µL. PCR conditions: heat denaturation at 98 °C for 2 min, 35 cycles of (denaturing: 98 °C for 10 s, annealing: 55, 58, 62 and 65 °C for 30 s, extension: 68 °C for 6.5 min), final extension at 68 °C for 10 min, 4 °C hold. Also, as a negative control, PCR was performed without adding the 'megaprimer'. Further, digestion of unamplified pET28a was performed with Dpn1 (20 U). Each of the gradient PCR reactions cocktail were transformed in chemically competent *E.Coli.* DH5α cells by providing heat shock at 42 °C for 1 min and was chilled immediately at ice for 2 min. LB (300 µL)was added to the transformed cells and was incubated at 37 °C for 1 h at 180 rpm shaking. The bacterial cells were plated overnight on an LB agar having kanamycin (50 µg/mL). After 12 h, bacterial colonies were picked individually and inoculated in separate 15 mL cultures. No colonies were observed for negative control without megaprimer. After reaching an optical density at 600 nm(OD_{600}) of ~1.4, cells corresponding to each colony were pelleted down and plasmids were isolated individually using QIAprep Spin Miniprep kit using the manufacturer's protocol. The isolated plasmids were resuspended in autoclaved water and was used for PCR for screening positive clones (Figure 3.3).

PCR of isolated plasmids corresponding to clones 1-7

The isolated clones corresponding to DNA isolated from each colony was used as a template for PCR using T7 forward and reverse sequencing primer (T7FP and T7RP) with Emerald Taq DNA polymerase. For each clone, primers (0.5 µM), EmeraldAmp GT PCR master mix (10 µL), clone DNA (100 ng) was added to a final volume of 20 µL in autoclaved water. PCR conditions: heat denaturation at 94 °C for 3 min, 35 cycles of (denaturing: 94 °C for 30 s, annealing 58 °C for 30 s, extension: 68 °C for 1.5 min), final extension at 68 °C for 10 min, 4 °C hold. Also, as a positive control, PCR was performed with empty pET28a (100 ng). Further clone **3** was sequenced and was

found to have a positive match with the designed pET28a_SpCID1 construct (Figures 3.5, 3.7 and Table 3.1).

Optimization of expression of SpCID1 enzyme

For investigating the appropriate IPTG concentration and temperature required for expression of SpCID1 enzyme, pET28a_SpCID1 plasmid was transformed in chemically competent BL21 (DE3) cells and streaked in LB agar plate having kanamycin (50 μg/mL). After 16 h, a colony was picked from the plate and incubated in primary culture with kanamycin until OD_{600} of ~1.4 was reached. From this, a 50% glycerol stock was made for further streaking. Bacteria was grown in LB having kanamycin for all downstream culturing.

One percent of primary culture was inoculated in a secondary culture (15 mL) and grown till OD_{600} of ~0.6 after which expression was induced with 0.2, 0.5 or 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). As an uninduced control, no IPTG was added to check leaky expression from plasmid. The culture flasks were incubated at varying temperatures of 18 °C for 16 h, 28 °C for 12.5 h or 37 °C for 4.5 h. Cells were pelleted down and washed in 1X phosphate buffer saline (PBS). The cells were then resuspended in 500 μL of 50 mM HEPES, 500 mM NaCl, 5 mM mercaptoethanol, 30 μM PMSF and sonicated on ice (1 s pulse on, 3 s pulse off at 60% amplitude for 1 min) in a 1.5 mL micro-centrifuge tube to obtain cell lysate. The total cell lysate was then centrifuged at 14,500 rpm for 5 min. Total lysate and supernatant after centrifugation (30 μL) was mixed with 12.5 μL of SDS denaturing protein loading buffer, boiled at 95 °C and loaded onto 12.5% SDS-PAGE gel and resolved. The bands corresponding to the protein was visualized by staining gel in Coomassie blue and by further de-staining. As a positive control, plasmid expressing histone protein known to get induced with 0.5 mM IPTG and 37 °C was transformed, cultures lysed, and was run on gel (Figure 3.8).

Optimization of lysis and immobilized metal affinity chromatography for protein isolation

In order to purify the protein, glycerol stock containing pET28a_SpCID1 plasmid transformed in Rosetta™(DE3) cells were streaked on LB agar plate with kanamycin (50 μg/mL) and chloramphenicol (25 μg/mL). A colony was picked and used for growing primary culture (7.5 mL LB with kanamycin and chloramphenicol at 180 rpm shaking. Further 2% of primary culture was inoculated in secondary culture (200 mL LB) having both the antibiotics and grown till OD_{600} of

~0.6. After attaining the required OD$_{600}$, the flask was brought to 18 °C, induced with 0.5 mM IPTG and incubated at 18 °C for 16 h with 180 rpm shaking. The cells were pelleted down in a 50 mL centrifuge tube and washed with PBS. All downstream procedures were performed at 4 °C. The cells were resuspended in 7 mL lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10 mM Imidazole, 30 μM PMSF) and were sonicated on ice (20 s pulse on, 30 s pulse off at 80% amplitude for 2 min) in a centrifuge tube. Further, the lysate was centrifuges at 14,500 rpm for 30 min and supernatant was mixed with Ni-IDA affinity beads and incubated for 90 min. After 90 min, the sample was passed through a polypropylene column and the Ni-IDA beads were washed with varying concentration of imidazole in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 40-100 mM imidazole, 30 μM PMSF) with increasing step gradients in a volume of 5 mL each. The protein was eluted in varying concentration of imidazole in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 150-250 mM imidazole, 30 μM PMSF). The elute and wash buffer fraction were boiled at 90 °C with SDS denaturing loading buffer. Further samples were loaded and run on 12.5% denaturing loading gel. Uninduced lysate, the cell pellet and supernatant after centrifugation was also loaded. The gel was visualized by Coomassie blue gel staining (Figure 3.9A).

Western blot analysis of SpCID1 protein using Anti-His antibody

Fraction corresponding to cell pellet after centrifugation, lysate after centrifugation, small quantity of Ni-IDA resin bound with protein and purified protein was boiled with denaturing SDS loading buffer and run on 12.5% SDS-PAGE. The gel was then sandwiched with PVDF membrane and stacked on the blotting apparatus. The protein in gel was transferred to PVDF membrane by electro blotting at 100 V for 2 h at 4 °C using pre-chilled transfer buffer. The membrane was removed and washed with 10 mL of PBST buffer (1X PBS with 0.1% Tween-20) for 5 min with gentle shaking on a gel rocker and was further blocked with 10 mL of 5% skimmed milk in PBST buffer (1X PBS with 0.1% Tween-20) for 1 h with gentle shaking. The membrane was washed again three times with 10 mL of 1X PBST for 5 min each and incubated with 2 ml of 5% skimmed milk in PBST having Anti-6X His tag® antibody (1:1000 dilution) in a laminated sheet at 4 °C for 12 h with overnight rocking. The primary antibody was removed and PVDF membrane was washed three times in 10 mL of 1X PBST for 5 min. Secondary antibody, HRP Goat Anti-Mouse (IgG) was diluted in 5% skimmed milk in 1X PBST (1:10,000) and was incubated with PVDF membrane for 3 h at room temperature. Further, the membrane was washed three times in 1X PBST for 5 min

and incubated with 50% solution of luminol in hydrogen peroxide and imaged using chemiluminescence in G:Box (Figure 3.9B).

Large-scale purification and size-exclusion chromatography

Two percent of primary culture having cells previously transformed with pET28a_SpCID1 was inoculated in secondary culture in multiple batches of 1 L culture volume having both kanamycin and chloramphenicol. Cells were induced with 0.5 mm IPTG at OD_{600} of ~0.6 and grown at 18 °C for 16 h with 180 rpm shaking. The culture was pelleted down and lysed on ice in 50 mL lysis buffer (20 s pulse on, 30 s pulse off at 80% amplitude for 10 min). The lysed cells were centrifuged at 14,500 rpm for 30 min. To the supernatant, 10 ml Ni-IDA resin was added and incubated for 90 min under slow rotation at 4 °C. The resin was poured and packed in a glass column for gravity assisted elution. Further, the column with resin was washed with 500 mL of wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 75 mM imidazole, 30 µM PMSF) and protein was eluted in elution buffer (50 mM Tris pH 7.5, 150 mM NaCl, 250 mM imidazole, 30 µM PMSF). The eluted protein was buffer exchanged to size exclusion buffer (20 mM HEPES pH 6.8, 150 mM NaCl, 5 mM 2-mercaptoethanol, 30 µM PMSF) at 4 °C and concentrated by molecular weight cut-off columns. The concentrated enzyme (10 mL) was injected in fast protein liquid chromatography (FPLC) into HiLoad 16/600 Superdex 200 preparative column and eluted in size exclusion buffer at a flow rate of 1 mL/min. Further fractions corresponding to SpCID1 enzyme was dialysed three times against storing buffer (20 mM HEPES pH 6.8, 150 mM NaCl, 5 mM 2-mercaptoethanol, 50% glycerol), aliquoted and stored at -40 °C. The protein concentration was calculated using Bradford assay against a standard solution of BSA which yielded protein of 10.25 µM and 107.4 µM stock concentrations (Figure 3.10).

3.4.4 Synthesis of 5-azidomethyluridine-5'-triphosphate (e, AMUTP)

5-azidomethyl uridine was prepare by previously reported procedures.[25,34] To 5-azidomethyl uridine (0.150 g, 0.50 mmol, 1 equiv), was added freshly distilled $POCl_3$ (117 µL, 1.25 mmol, 2.5 equiv), trimethylphosphate (1.5 mL) in ice-cold conditions. The solution was stirred for 20 h at ~4 °C. Tributylamine (1.2 mL, 5.01 mmol, 10 equiv) and bis-tributylammonium pyrophosphate (0.5 M in DMF, 5 mL, 5 equiv) and was added simultaneously to the solution in ice-cold conditions. After 30 min, the reaction was quenched with 1 M triethylammonium bicarbonate buffer (TEAB,

pH 7.5, 15 mL), followed by washing with ethyl acetate (20 mL). The aqueous layer was evaporated-off and the residue obtained was purified using a DEAE sephadex-A25 anion exchange column (10 mM–1 M TEAB buffer, pH 7.5). Further, it was purified by RP flash column chromatography (C18 Redi*SepRf*, 0–50% acetonitrile in 50 mM triethylammonium acetate buffer, pH 7.5, 50 min). Desired fraction was lyophilized to afford the triphosphate product as triethylammonium salt (43.6 mg, 9.2%) (Scheme 3.1). ^{1}H NMR (400 MHz, D$_2$O): δ (ppm) 8.04 (s, 1H), 5.93 (d, J = 4.16 Hz, 1H), 4.36 (br, 2H), 4.21–4.17 (m, 5H); ^{31}P NMR (162 MHz, D$_2$O): δ (ppm) -10.93 (br, P$_γ$), -11.88 (br, P$_α$), -22.47 (br, P$_β$); MALDI-TOF MS (m/z) negative mode: Calculated for C$_{10}$H$_{15}$N$_5$O$_{15}$P$_3^-$ [M-H]$^-$ = 538.17, found = 537.96 (Appendix-II).

3.4.5 Terminal uridylation of RNA using SpCID1 and modified UTPs

To label RNA at the 3′ end, 5′ FAM-labeled RNA (10 μM) was incubated with natural UTP or modified triphosphates (500 μM, 50 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and 1 μL of in-house expressed SpCID1 (10.25 pmol) or commercial SpCID1 (2 U) in a final volume of 20 μL. After 5, 10/15 min and 30 min, 5 μL aliquots of reaction mixture (50 pmol) was mixed with 15 μL of denaturing loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8) and heat-denatured at 75 °C for 3 min. Further, 5 μL of the inactivated reaction mixture (12.5 pmol) was loaded on to a 20% denaturing polyacrylamide gel and imaged using Typhoon gel scanner at FAM wavelength. The control reactions were maintained with same concentration of 5′ FAM-labeled RNA, Tris-HCl, NaCl, MgCl$_2$, and DTT without any nucleoside triphosphates and with or without SpCID1 enzyme (Figures 3.12, 3.13 and Table 3.1).

3.4.6 Controlling the degree of incorporation of modified triphosphate
Effect of SpCID1 concentration on 3′ RNA tailing

5′ FAM-labeled RNA (10 μM) was incubated with EUTP (500 μM, 50 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and commercial SpCID1 (0.5, 1, 2, 4 and 8 U) in a final volume of 20 μL, keeping overall glycerol percentage constant at 10%. After 5, 10 and 30 min, aliquots of 5 μL of reaction (50 pmol) was mixed with denaturing loading buffer (15 μL) and heat-denatured at 75 °C for 3 min. Further, 5 μL of the inactivated reaction mixture (12.5 pmol) was

loaded on to a 20% denaturing polyacrylamide gel and imaged using Typhoon scanner at FAM wavelength. The control reaction was maintained with same concentration of EUTP, 5′ FAM-labeled RNA, Tris-HCl, NaCl, $MgCl_2$ and DTT without the enzyme (Figure 3.14).

Effect on modified nucleoside triphosphate concentration on 3′ RNA tailing

5′ FAM-labeled RNA (10 μM) was incubated with EUTP (0.5-4mM, 50-400 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and commercial SpCID1 (2 U) in a final volume of 20 μL. After 5, 10 and 30 min, 5 μL aliquots of reaction (50 pmol) was mixed with denaturing loading buffer (15 μL) and heat-denatured at 75 °C for 3 min. Further, 5 μL of the inactivated reaction mixture (12.5 pmol) was loaded on 20% denaturing polyacrylamide gel and imaged using Typhoon scanner at FAM wavelength. The control reaction was maintained with same concentration of enzyme, 5′ FAM-labeled RNA, Tris-HCl, NaCl, $MgCl_2$ and DTT without the triphosphate (Figure 3.15).

3.4.7 Incorporation of single modified nucleotide on RNA

Optimization of single modification of 5′ FAM-labeled RNA with base-modified vinyl and alkyne UTPs

Single terminal uridylation of RNA at the 3′ end with alkyne and vinyl tag was optimized by incubating 5′ FAM-labeled RNA (10 μM) for 5, 10 and 30 min with VUTP, EUTP or ODUTP (500 μM, 50 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and commercial SpCID1 (2 U for EUTP, 3.5 U for VUTP or 5 U for ODUTP) in a final volume of 100 μL. After 5, 10 and 30 min, 5 μL aliquots of reaction (50 pmol) was mixed with denaturing loading buffer (15 μL) and heat-denatured at 75 °C for 3 min. Further, 5 μL of the inactivated reaction mixture (12.5 pmol) was loaded on to a 20% denaturing polyacrylamide gel and imaged using Typhoon scanner at FAM wavelength. (Figures 3.16 and 3.17).

Optimization of single modification of 5′ FAM-labeled RNA with base-modified azide UTPs

Single terminal uridylation of RNA at the 3′ end was optimized by incubating 5′ FAM-labeled RNA (10 μM) for 5, 15 and 30 min with AMUTP (10 μM), APUTP (25 μM) or ATUTP (50 μM)

in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/µL) and in-house expressed SpCID1 (3.42 pmol for AMUTP and APUTP, 10.25 pmol for ATUTP) in a final volume of 20 µL. After 5, 15 and 30 min, 5 µL aliquots of reaction (50 pmol) was mixed with denaturing loading buffer (15 µL) and heat-denatured at 75 °C for 3 min. Further, 5 µL of the inactivated reaction mixture (12.5 pmol) was loaded on to a 20% denaturing polyacrylamide gel and imaged using Typhoon gel scanner at FAM wavelength (Figures 3.16 and 3.18).

Scale-up of single azide-modified RNA

5′ FAM-labeled RNA (10 µM, 200 pmol) was incubated with AMUTP (10 µM), APUTP (25 µM) or ATUTP (50 µM) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/µL) and in-house expressed SpCID1 (3.42 pmol for AMUTP and APUTP, 10.25 pmol for ATUTP) in a final volume of 20 µL and was incubated for 15 min with AMUTP, APUTP or 30 min for ATUTP. Typically for a 1 nmol reaction, five small scale reactions of 200 pmol of RNA (20 µL) was incubated at respective time points, heat-inactivated at 75 °C for 3 min and was either purified by reverse-phase high-performance liquid chromatography (RP-HPLC) or by ethanol precipitation. When purified by RP-HPLC a tailing reaction using modified UTP analogs AMUTP, APUTP and ATUTP yielded 89, 95 and 76% of product RNA. Further, the mass of the RNA was confirmed by ESI mass analysis in negative mode by direct infusion of 100 pmol of RNA in 50% ACN (LCMS grade) in 10 mM triethylamine and 100 mM hexafluoro-2-propanol (Table 3.2).

3.4.8 Strain-promoted azide-alkyne cycloaddition reaction on azide RNA

For performing SPAAC reaction, AMU-labeled 5′ FAM RNA (10 µM, 300 pmol) was incubated with Cy3-DBCO (1 mM), in final volume of 30 µL in autoclaved water. The reaction was incubated at 37 °C for 2 h followed by addition of 30 µL of 5 M ammonium acetate and 300 µL ethanol. The reaction mixture was precipitated kept at -20 °C overnight followed by centrifugation at 15,000 rpm for 15 min. The pellet was washed with 500 µL pre-chilled 75% ethanol in autoclaved water to remove salts and unconjugated dye followed by centrifugation (15,000 rpm) for 15 min. The RNA pellet obtained was dried and nuclease free autoclaved water was added to

dissolve the RNA to obtain 0.15 pmol of product Cy3-FAM RNA. Further, the mass of the RNA was confirmed by ESI mass analysis (Figure 3.19 and Table 3.2).

### 3.4.9 Copper-catalyzed azide-alkyne cycloaddition reaction on azide RNA

For performing copper-catalyzed click reaction on azide-labeled RNA, $CuSO_4$ (0.2 mM) was reduced with sodium ascorbate (2 mM) in the presence of THPTA ligand (1 mM) and further Alexa 594 alkyne (1 mM) and AMU-labeled 5′ FAM RNA (10 μM, 300 pmol) was added to a final volume of 30 μL. The reaction was incubated at 37 °C for 2 h. RNA was precipitated by adding 30 μL of 5 M sodium acetate and 300 μL ethanol and kept at -20 °C overnight followed by centrifugation at 15,000 rpm for 15 min. The pellet thus obtained was washed with 500 μL pre-chilled 75% ethanol in autoclaved water and centrifuged at 15,000 rpm for 15 min. The RNA pellet obtained was dried and resuspended in nuclease-free autoclaved water to obtain 0.16 nmol of Alexa 594-FAM RNA. Further, the mass of the RNA was confirmed by ESI mass analysis (Figure 3.20 and Table 3.2).

3.5 References

1. George, J. T.; Srivatsan, S. G., Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. *Methods* **2017,** *120*, 28–38.
2. Holstein, J. M.; Rentmeister, A., Current covalent modification methods for detecting RNA in fixed and living cells. *Methods* **2016,** *98*, 18–25.
3. Motorin, Y.; Burhenne, J.; Teimer, R.; Koynov, K.; Willnow, S.; Weinhold, E.; Helm, M., Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic Acids Res.* **2011,** *39* (5), 1943–1952.
4. Holstein, J. M.; Schulz, D.; Rentmeister, A., Bioorthogonal site-specific labeling of the 5′-cap structure in eukaryotic mRNAs. *Chem. Commun.* **2014,** *50* (34), 4478–4481.
5. Muttach, F.; Rentmeister, A., One-pot modification of 5′-capped RNA based on methionine analogs. *Methods* **2016,** *107*, 3–9.
6. Plotnikova, A.; Osipenko, A.; Masevičius, V.; Vilkaitis, G.; Klimašauskas, S., Selective covalent labeling of miRNA and siRNA duplexes using HEN1 methyltransferase. *J. Am. Chem. Soc.* **2014,** *136* (39), 13550–13553.
7. Winz, M.-L.; Samanta, A.; Benzinger, D.; Jäschke, A., Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res.* **2012,** *40* (10), e78.
8. Tomkuvienė, M.; Clouet-d'Orval, B.; Černiauskas, I.; Weinhold, E.; Klimašauskas, S., Programmable sequence-specific click-labeling of RNA using archaeal box C/D RNP methyltransferases. *Nucleic Acids Res.* **2012,** *40* (14), 6765–6773.

9.   Alexander, S. C.; Busby, K. N.; Cole, C. M.; Zhou, C. Y.; Devaraj, N. K., Site-Specific Covalent Labeling of RNA by Enzymatic Transglycosylation. *J. Am. Chem. Soc.* **2015,** *137* (40), 12756–12759.

10.   Mondal, M.; Liao, R.; Nazaroff, C. D.; Samuel, A. D.; Guo, J., Highly multiplexed single-cell in situ RNA and DNA analysis with bioorthogonal cleavable fluorescent oligonucleotides. *Chem. Sci.* **2018,** *9* (11), 2909–2917.

11.   Pitchiaya, S.; Heinicke, L. A.; Custer, T. C.; Walter, N. G., Single molecule fluorescence approaches shed light on intracellular RNAs. *Chem. Rev.* **2014,** *114* (6), 3224–3265.

12.   Panda, A. C.; Martindale, J. L.; Gorospe, M., Affinity pulldown of biotinylated RNA for detection of protein-RNA complexes. *Bio Protoc.* **2016,** *6* (24).

13.   Paredes, E.; Das, S. R., Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem* **2011,** *12* (1), 125–131.

14.   Dojahn, C. M.; Hesse, M.; Arenz, C., A chemo-enzymatic approach to specifically click-modified RNA. *Chem. Commun.* **2013,** *49* (30), 3128–3130.

15.   Keyhani, S.; Goldau, T.; Blümler, A.; Heckel, A.; Schwalbe, H., Chemo-enzymatic synthesis of position-specifically modified RNA for biophysical studies including light control and NMR spectroscopy. *Angew. Chem. Int. Ed.* **2018,** *57* (37), 12017–12021.

16.   Rinaldi, A. J.; Suddala, K. C.; Walter, N. G., Native purification and labeling of RNA for single molecule fluorescence studies. *Methods Mol. Biol.* **2015,** *1240*, 63–95.

17.   Kwak, J. E.; Wickens, M., A family of poly(U) polymerases. *RNA* **2007,** *13* (6), 860–867.

18.   Wickens, M.; Kwak, J. E., A tail tale for U. *Science* **2008,** *319* (5868), 1344–1345.

19.   Kroupova, A.; Ivaşcu, A.; Jinek, M.; Reimão-Pinto, M. M.; Ameres, S. L., Structural basis for acceptor RNA substrate selectivity of the 3′ terminal uridylyl transferase Tailor. *Nucleic Acids Res.* **2018,** *47* (2), 1030–1042.

20.   Aphasizhev, R.; Sbicego, S.; Peris, M.; Jang, S.-H.; Aphasizheva, I.; Simpson, A. M.; Rivlin, A.; Simpson, L., Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase). *Cell* *108* (5), 637–648.

21.   Menezes, M. R.; Balzeau, J.; Hagan, J. P., 3′ RNA Uridylation in epitranscriptomics, gene regulation, and disease. *Front. Mol. Biosci.* **2018,** *5* (61).

22.   Rissland, O. S.; Norbury, C. J., The Cid1 poly(U) polymerase. *BBA-Gene Regul. Mech.* **2008,** *1779* (4), 286–294.

23.   Lunde, B. M.; Magler, I.; Meinhart, A., Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res.* **2012,** *40* (19), 9815–9824.

24.   Yates, L. A.; Fleurdépine, S.; Rissland, O. S.; De Colibus, L.; Harlos, K.; Norbury, C. J.; Gilbert, R. J. C., Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat. Struct. Mol. Biol.* **2012,** *19*, 782.

25.   Sawant, A. A.; Tanpure, A. A.; Mukherjee, P. P.; Athavale, S.; Kelkar, A.; Galande, S.; Srivatsan, S. G., A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **2016,** *44* (2), e16.

26.   Sousa, R.; Chung, Y. J.; Rose, J. P.; Wang, B.-C., Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **1993,** *364* (6438), 593–599.

27.   Munoz-Tello, P.; Gabus, C.; Thore, S., Functional Implications from the Cid1 Poly(U) Polymerase Crystal Structure. *Structure* **2012,** *20* (6), 977–986.

28.   van den Ent, F.; Löwe, J., RF cloning: A restriction-free method for inserting target genes into plasmids. *J. Biochem. Bioph. Methods* **2006,** *67* (1), 67–74.

29. Unger, T.; Jacobovitch, Y.; Dantes, A.; Bernheim, R.; Peleg, Y., Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.* **2010,** *172* (1), 34–44.

30. Rao, H.; Sawant, A. A.; Tanpure, A. A.; Srivatsan, S. G., Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chem. Commun.* **2012,** *48* (4), 498-–500.

31. Rao, H.; Tanpure, A. A.; Sawant, A. A.; Srivatsan, S. G., Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nat. Protoc.* **2012,** *7* (6), 1097–1112.

32. Sawant, A. A.; Mukherjee, P. P.; Jangid, R. K.; Galande, S.; Srivatsan, S. G., A clickable UTP analog for the posttranscriptional chemical labeling and imaging of RNA. *Org. Biomol. Chem.* **2016,** *14* (24), 5832–5842.

33. George, J. T.; Srivatsan, S. G., Vinyluridine as a Versatile Chemoselective Handle for the Post-transcriptional Chemical Functionalization of RNA. *Bioconjug Chem.* **2017,** *28* (5), 1529–1536.

34. Leszczynska, G.; Leonczak, P.; Wozniak, K.; Malkiewicz, A., Chemical synthesis of the 5-taurinomethyl(-2-thio)uridine modified anticodon arm of the human mitochondrial tRNALeu(UUR) and tRNALys. *RNA* **2014,** *20* (6), 938–947.

35. Munoz-Tello, P.; Gabus, C.; Thore, S., A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure. *Nucleic Acids Res.* **2014,** *42* (5), 3372–3380.

36. Chelico, L.; Pham, P.; Goodman, M. F., Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. *Philos. Trans. Royal Soc. B* **2009,** *364* (1517), 583–593.

3.6 Appendix-II: Characterization data of synthesized compounds

¹H NMR of AMUTP (400 Mhz, D₂O)
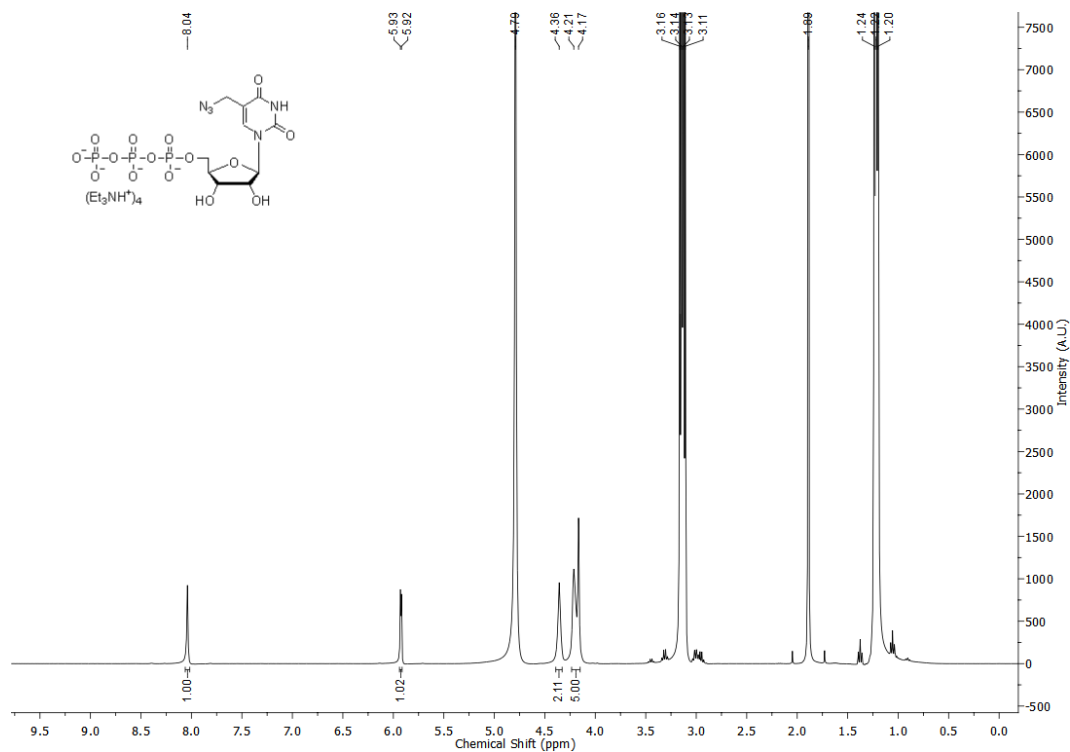
³¹P NMR of AMUTP (162 Mhz, D₂O)
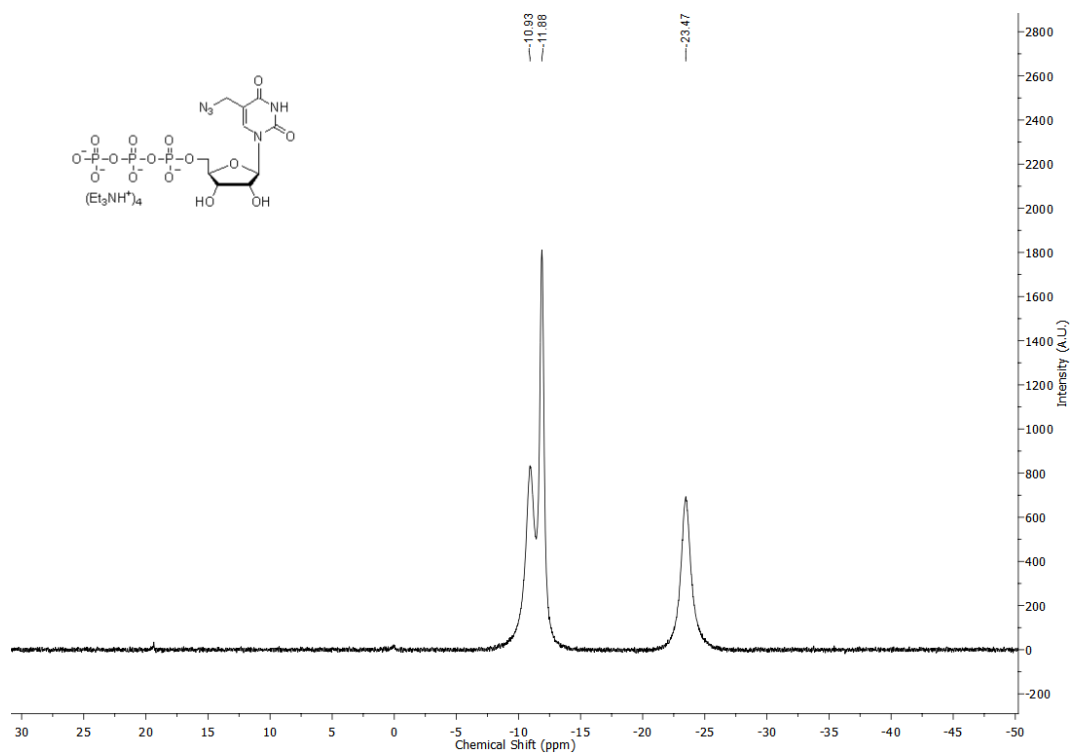
Chapter 4

Utility of SpCID1 for High-Density Labeling of CRISPR Guide RNA, Developing FRET Probes and Site-Specific Internal Labeling

4.1 Introduction

The introduction of functional tags specifically at the 3′ end of RNA of interest using SpCID1 terminal uridylating enzyme opens up tremendous possibilities for labeling RNA with single or multiple labels. Consecutive incorporation of modified nucleosides using conventional methods like solid-phase synthesis is particularly challenging, due to the decrease in efficiency upon incorporating multiple modified phosphoramidates. Also, the length of oligomers, synthesized by solid-phase chemistry is restricted to <50 bases.[1] Alternatively, one can synthesize labeled RNA enzymatically with high density labels by T7 RNA polymerase mediated transcription using modified NTPs.[2-4] However, this approach results in modification incorporated non-selectively throughout the RNA, which could possibly affect the native structure and folding of RNA and or perturb critical RNA-protein interactions thereby limiting its utility. Multiple tags incorporated into RNA using terminal transferases provides the ability to tailor high-density functional labels selectively at the 3′ end which can be designed to be elsewhere from sites of RNA-RNA or RNA-protein interaction and thereby preserving the functional integrity of RNA.[5] These high-density labels at the 3′ end has its precedence over conventional tags as it helps in signal amplification which can be used for ensemble or single molecule fluorescence[6] or pull-down[7-9] experiments used for a variety of applications from sensing RNAs of low copy number[10], to investigating RNA-protein interactions[7, 9, 11] and to detect viruses[12]. In chapter 3, we had optimized the enzyme, SpCID1, to incorporate single or multiple modified nucleotides at the 3′ end of RNA. The objective of this chapter is to explore various utility of this labeling approach. To check the ability of SpCID1 enzyme to introduce multiple clickable tags specifically at 3′ end with minimal perturbation of structure and function of RNA, we envisioned to introduce labels on a functional RNA, the CRISPR guide RNA. We conceived that employing clickable CRISPR guide RNA, we could localize useful chemical functionalities site-specifically to gene locus of interest.

Sections 4.2.2-4, 4.4.4-5 has been partially performed together with Azhar, Meghali and Dipanjali at Dr. Souvik Maiti and Dr. Debojyoti Chakraborty Lab, Institute of Genomics and Integrative Biology, New Delhi.

Further, we explore the possibility of constructing RNA labeled with FRET pairs using the selective incorporation of a single bioorthogonal tag at the 3′ end of RNA. This technique provides easy enzymatic access to biophysical probes, compatible for RNA ensemble or single molecule FRET experiments. Next, in order to prove that modification incorporated using SpCID1 can also utilized for site-specific internal labeling, we construct an internal label on a longer RNA using the 3′ terminal uridylated RNA having single bioorthogonal tag.

CRISPR gene editing tool and beyond

The bacterial adaptive defense system, namely CRISPR-Cas9 has been in the spotlight for its ability to be used as a simple and easy-to-use genome editing tool.[13-18] The type II CRISPR-Cas bacterial immune system from which the gene editing tool is derived, consist of series of repetitive sequences having interspaced non-repetitive sequences called spacers corresponding to acquired region from viral phages.[14, 16, 19] The region is transcribed into an RNA called CRISPR-RNA (crRNA) having sequences of this acquired spacer region. The defense system encompasses additional genetic elements for a non-coding RNA namely, trans-acting CRISPR RNA (tracr RNA) which forms a transcript complementary to the repetitive sequence and therefore binds to the crRNA. This enable both the RNAs to function together as the guide RNA. The defense system also involves operons having CRISPR-associated genes (Cas) which expresses various proteins that help in acquiring and integration of spacer elements from viral gene vectors. Also, the expression of a nuclease protein from the Cas gene which binds to crRNA and tracr RNA guides the ribonucleoprotein (RNP) complex for target cleavage. Doudna and coworkers, used the type II CRISPR system from *Streptococcus pyogenes* which uses a dsDNA nuclease protein namely, Cas9 for genome editing.[18, 20] The crRNA and tracr RNA for this CRISPR system was combined into one *in vitro* transcribable RNA namely, single guide RNA (sgRNA) with a protospacer sequence complementary to dsDNA for cleavage (Figure 4.1A). The sgRNA binds to Cas9 nuclease protein to form the ribonucleoprotein (RNP) complex which searches dsDNA for a protospacer adjacent motif (PAM) having an 'NGG' sequence, where 'N' can be any nucleotide. Upon binding to PAM, the RNP complex melts the dsDNA and checks for sequence complementarity between dsDNA and sgRNA. Upon target binding, the RNP complex using RuvC and HNH nuclease domains, cleaves the dsDNA.
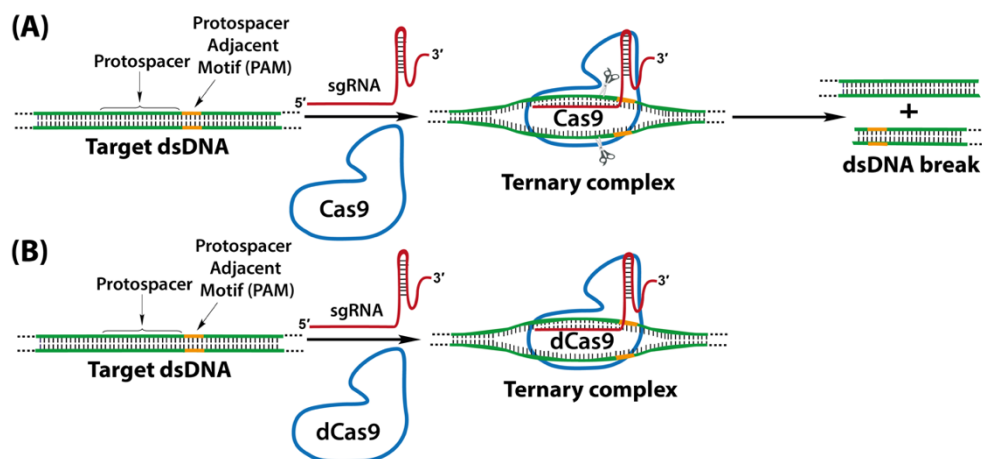
Figure 4.1. The CRISPR tool (**A**) Cas9 mediated cleavage of dsDNA (**B**) dCas9 binds efficiently to form the ternary complex which is used for localizing to specific chromatin regions.

The easy-of-use and the ability to target any genomic region by changing the guide RNA sequence sets this technology apart from other traditionally used technologies like ZFN and TALENS.[15] A double mutant of Cas9 protein namely, dCas9 (mutated at RuvC and HNH nuclease domains) can bind effectively to target dsDNA and therefore be used for localizing the ternary complex sequence-selectively to chromatin regions (Figure 4.1B). This mutant protein, dCas9 has received tremendous attention for its versatility for gene visualization, transcription activation, repression, or CRISPR based chromatin pull-down (Figure 4.2).[15, 17, 21-24] This have been achieved by attaching genetically fused proteins, epitope tags on dCas9 or aptamers on sgRNA.[15, 24-25] However there lacks a simplified, robust and versatile approach to conjugate small molecules probes and modulators on CRISPR system. This would require an sgRNA synthesized chemically by solid-phase synthesis which is restricted its length. For example, sgRNA has been synthesized with a fluorescent dye, generated by splint-ligating a chemically synthesized RNA oligonucleotide, was used in establishing a CRISPR-based FISH assay to visualize gene loci in fixed cells.[26] One of the most advanced bioconjugation technique in the past decade uses bioorthogonal tools employing copper or copper-free click chemistries for conjugating proteins, drugs, fluorophores, pull-down tags or any biologically functional warhead.[3, 27-30] Recently, chemically synthesized crRNAs containing chemo-selective reactive handles have been used in stitching tracrRNA to yield full-length sgRNA or donor DNA to facilitate homology directed repair.[31-32] However, incorporating bioorthogonal clickable tags on CRISPR-Cas9 active complex is tedious and hasn't been reported as far as our best knowledge. This is because selective covalent incorporation of bioorthogonal tags require genetically incorporated clickable tags on amino acid for protein tagging[33] or

chemically synthesized guide RNA[21, 32] having clickable tags which doesn't hamper its activity. Herein, we employed an innovative chemo-enzymatic strategy to introduce a bioorthogonal click tag at 3′ end of sgRNA using SpCID1 preserving the CRISPR activity. Particularly, this technology makes possible, site-directed localization of azide tags to genetic locus of interest in cells exploiting the utility of CRISPR. Next, by employing click chemistry (CuAAC or SPAAC) it is possible to perform targeted conjugation of any functional warhead or tag onto the azide-labeled gene locus. Such a technology opens up tremendous possibilities for CRISPR-Cas9 assisted functional tagging by providing site-directed click functionalization of genetic locus.
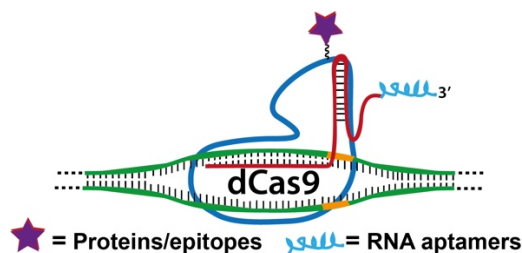
Figure 4.2. dCas9 protein can be covalently attached to a functional protein or non-covalently labeled with variety of functional probes using epitopes on dCas9 and aptamers on sgRNA.[15, 17]

RNA FRET pairs and site-specific internal labeling

FRET pairs on RNA have tremendously expanded our understanding of RNA structural changes upon RNA-RNA or RNA-protein interactions. Structure determination of RNA performed using ensemble or single molecule FRET studies have provided detailed information of their conformational states which in turn plays a decisive role in its function. FRET pairs on RNA has been used for solving structure of various ribozymes,[34-35] riboswitches[36] and ribonucleoprotein complexes.[37-38] Conventionally FRET probes are attached on RNA using NHS ester chemistry, periodate chemistry and RNA ligation.[39] However, these techniques are tedious and gives reduced isolable yields of product RNA. Chemo-enzymatic techniques can be used to add single functional tag on 3′ end of RNA with the assistance of 3′ sugar-protected nucleotide, dideoxy nucleotide or reversible terminators.[40-41] However, using these approaches for site-specific internal labeling using RNA ligation is not possible due to the absence of free 3′ hydroxyl. Also, employing reversible terminators add to the complexity of the technique wherein an additional reagent is required for deprotection.[42] Terminal uridylation of RNA employing SpCID1 furnishes the ability to add a single modified UMP residue without 3′ hydroxyl protected nucleotide, dideoxy nucleotide or reversible terminators. This is achieved by controlling the stoichiometry of enzyme

and modified triphosphate analog. In this chapter we focus on the utility of this single addition for constructing RNA FRET pairs and further we also investigate whether 5-azidomethyl uridine (AMU) modified RNA can be deployed for site-specific internal labeling of RNA.

4.2 Results and Discussion

4.2.1 Utility of SpCID1 for high-density labeling of CRISPR sgRNA

Design strategy for terminal uridylation of sgRNA

In order for functionalize the CRIPSR sgRNA with multiple clickable tags, we adopted a design strategy wherein the RNA is posttranscriptionally terminal uridylated using SpCID1 enzyme with AMUTP to yield high-density azide tags at its 3′ end (Figure 4.3). Further, we hypothesized that the azide-labeled sgRNA obtained can be used for (A) post-hybridization or (B) pre-hybridization click chemistry (Figure 4.3). For post-hybridization click chemistry, the azide modified sgRNA is incubated with dCas9 and target dsDNA to form ternary complex having multiple azide tags which can be further functionalized using SPAAC or CuAAC reaction to yield the labeled ternary complex having the functional tag. For pre-hybridization click approach, the azide-labeled sgRNA is first functionalized using SPAAC or CuAAC reaction to yield the probe/functional tag conjugated sgRNA which can be further be hybridized to dCas9 and target dsDNA to give the labeled ternary complex having the functional tag. Thus, the attachment of multiple functional tags on the labeled ternary complex will pave way for enabling site-directed labeling of gene loci with functional labels.
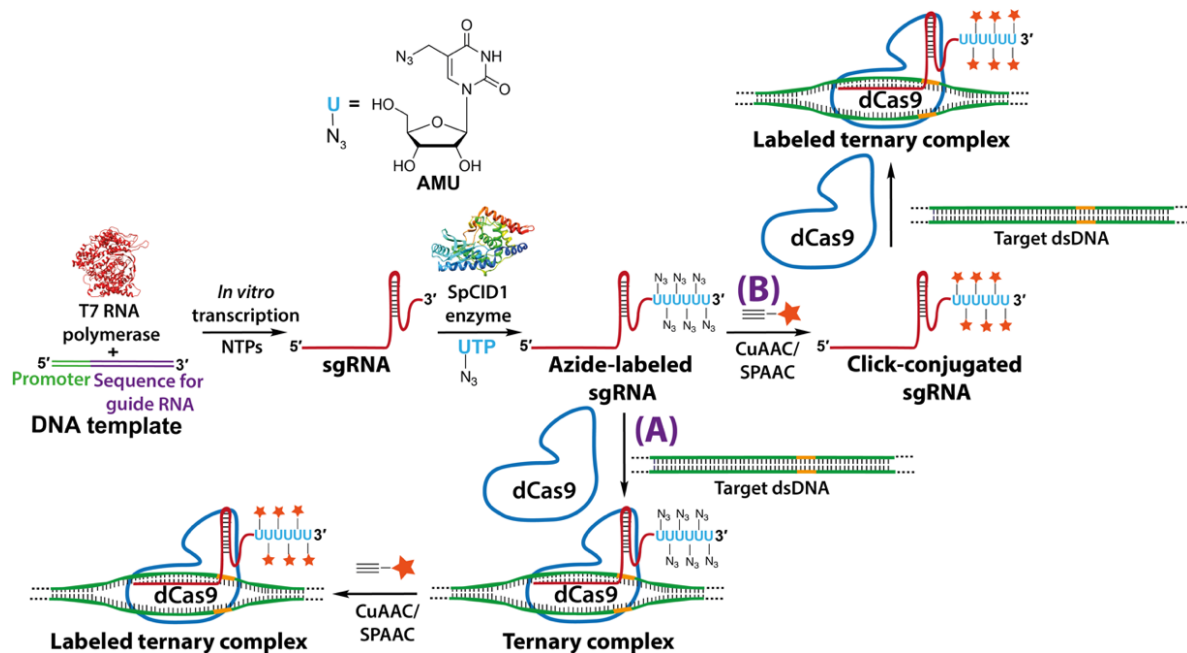
Figure 4.3. *In vitro* transcribed CRISPR sgRNA is terminal uridylated using SpCID1. The terminal uridylated RNA can further be functionalized using (**A**) post-hybridization click or (**B**) pre-hybridization click chemistry. Either strategy results in the formation of labeled ternary complex having multiple functional tags. (Structure of T7 RNA polymerase and SpCID1 adopted from PDB: 4RNP and 4FH3)[43-44]

Design of CRISPR sgRNA suitable for terminal uridylation reaction utilizing secondary structure prediction tools

The sgRNA for modifications were designed with a protospacer targeting either the telomere repeat region or eGFP gene.[23, 45] In order to label to CRISPR sgRNA with azide functionalities, we initially used a conventional sgRNA design[23] to synthesize sgRNA targeting the telomere region (*in vitro* transcribed from double stranded guide RNA template, CT) and incubated it with SpCID1 enzyme and AMUTP (Table 2). However, we observed no visible incorporation of AMU residues at the 3′ end when RNA was resolved by gel electrophoresis. This is because, the tracr region of CRISPR sgRNA being RNA pol III transcribed, ends with a strong hairpin loop structure at its 3′ end (Figure 4.4).[46] This can also be seen in the crystal structure solved for guide RNA bound to Cas9 nuclease protein.[19] *In vitro* experiments have revealed that the enzyme, SpCID1 requires a ~13 nucleotide stretch of single stranded RNA for efficient binding and reaction.[47] *In vivo*, the enzyme SpCID1 is known to bind to polyadenylated mRNA[48] and it is known that homopolymer A residues on RNA also exist in single stranded form.[49-50] In our experience of terminal uridylating various RNAs with SpCID1, we have observed that SpCID1 mediated terminal uridylation efficiency decreases when RNA forms strong secondary structures at its 3′

end. In order to facilitate the enzyme to bind and add modified UMPs at the end of CRISPR guide RNA, we decided to add a single stranded nucleotide overhangs at its 3′ end. From secondary structure prediction, extension of guide RNA with an additional 15 mer homopolymer of adenosine at its 3′ end, preserved the secondary structure of guide RNA and introduced a free single stranded RNA overhang (Figure 4.4). Homopolymer pyrimidine overhangs at the 3′ end has been reported to increase the stability of sgRNA in CRISPR systems and its efficiency.[51-52] Based on these considerations we sought to synthesize sgRNA having 15 mer adenylate overhang at its 3′ end and use this as unlabeled sgRNA control for all our experiments.
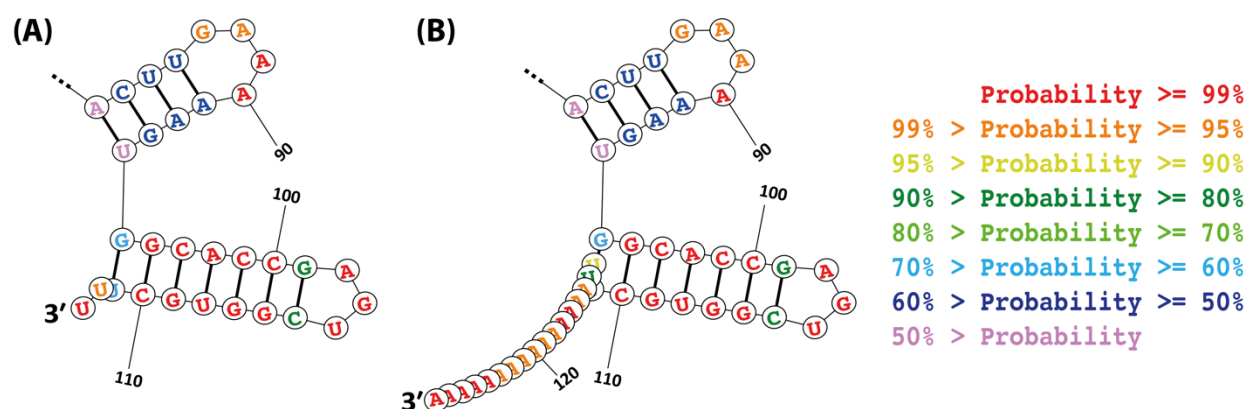
Figure 4.4. 3′ end of secondary structure predicted (**A**) conventional sgRNA design[23] (**B**) 15 AMP tailed sgRNA design (unlabeled sgRNA), color coded with probability of structure formation. Structure prediction of RNA was performed using online software, RNAstructure.[53]

Terminal uridylation of CRISPR guide RNA

The dsDNA template for sgRNA was synthesized by PCR and the amplicon obtained was used as template for *in vitro* transcription using natural NTPs and T7 RNA polymerase to obtain 2-3 nmoles of unlabeled sgRNA In order to perform terminal uridylation, unlabeled sgRNA **1** (10 pmoles) was incubated with the enzyme SpCID1 (1 µM) for 30 min at 37 °C with natural UTP or AMUTP (0.5 mM) in a final volume of 20 µL. Reaction was quenched and products were resolved by 8.5% denaturing PAGE. Unlabeled sgRNA **1** showed a slightly diffused band which can be attributed to slight variation in number adenylate residues added at the 3′ end as opposed to the designed 15 mer adenylate overhang. This is due to transcript slippage when T7 RNA polymerase transcribes a homopolymer A tail.[54] Nevertheless, the addition of a minimum ~10-15 residues at the 3′ end of unlabeled sgRNA can be observed on gel while comparing with conventionally used sgRNA lacking the 15 mer adenylate tail (data not shown). Terminal uridylation reaction,

rewardingly gave bands of lower mobility as observed on polyacrylamide gel stained with Stains-All (Figure 4.5C) or by 3.5% denaturing agarose gel stained with SYBR® Safe gel staining reagent. The reaction yielded products corresponding to distributive and processive incorporation for AMUTP.[55-56] The modified triphosphate, AMUTP incorporated with lesser efficiency as compared to natural UTP as observed from greater shift in product bands for natural UTP (Figure 4.5C). Further the azide-labeled sgRNAs were synthesized in large-scale by terminal uridylation of unlabeled sgRNA **1-2** (Table 1).
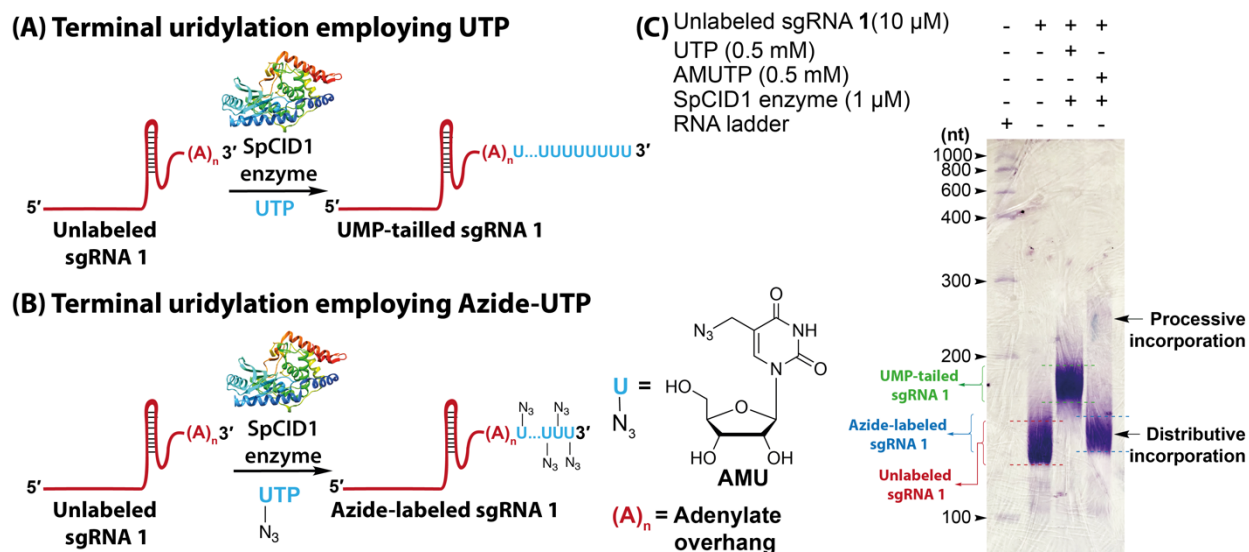
Figure 4.5. Terminal uridylation of unlabeled sgRNA **1** employing (**A**) natural UTP (**B**) 5-azidomethyl uridine triphosphate (AMUTP). (**C**) The bands corresponding to the uridylated product was resolved and visualized on 8.5% denaturing polyacrylamide gel using Stains-All reagent. The reaction resulted in formation of major band corresponding to distributive incorporation and minor amounts of processive incorporation for AMUTP. Structure of SpCID1 adopted from PDB: 4FH3.

Table 1. Mass and yield of large-scale isolated RNA products.

Product RNA	Calcd. Mass	Found Mass	Isolated yield (nmole)*	Isolated yield (%)
AMU-labeled 5' Cy5 RNA	6938.5	6938.0	0.85[a]	85
Cy5-Cy3 RNA	7921.7	7921.5	0.48[b]	80
Azide-labeled sgRNA 1	-	-	0.90[a]	90
Azide-labeled sgRNA 1'	-	-	1.86[c]	93
Azide-labeled sgRNA 2	-	-	1.91[c]	96
Cy3-conjugated sgRNA 1	-	-	0.27[d]	90
Cy3-conjugated sgRNA 2	-	-	1.65[e]	92
Biotin-conjugated sgRNA 1	-	-	0.23[d]	77

*Isolated yields for reactions performed on [a]1, [b]0.6, [c]2, [d]0.3, [e]1.8 nmol scale.

Labeling CRISPR sgRNA using stain-promoted azide-alkyne cycloaddition reaction with fluorescent and pull-down probes.

Bioconjugation of fluorescent/pull-down tag on azide-labeled sgRNA **1** for pre-hybridization click approach is achieved by incubating the modified RNA with Cy3-DBCO or biotin-DBCO for 2 h at 37 °C (Figure 4.6A). A typical 0.3 nmol reaction with Cy3-DBCO or biotin-DBCO yielded 0.27 and 0.23 nmol of Cy3 and biotin-conjugated sgRNA **1** (Table 1). SPAAC reaction on RNA was confirmed by resolving on 8.5% denaturing PAGE. Cy3-labeled product imaged in Typhoon gel scanner at Cy3 wavelength showed bands corresponding to distributive and processive incorporation of azide UMPs, labeled with Cy3 (Figure 4.6Bii). Further gel staining with Stains-All reagent showed product bands corresponding to lower migration as compared to the starting, processive and distributive AMU incorporated, azide-labeled sgRNA **1** (Figure 4.6Bi). Large scale SPAAC reaction was also performed on azide-labeled sgRNA **1-2** to get Cy3-conjugated sgRNAs **1-2** (Table 1).
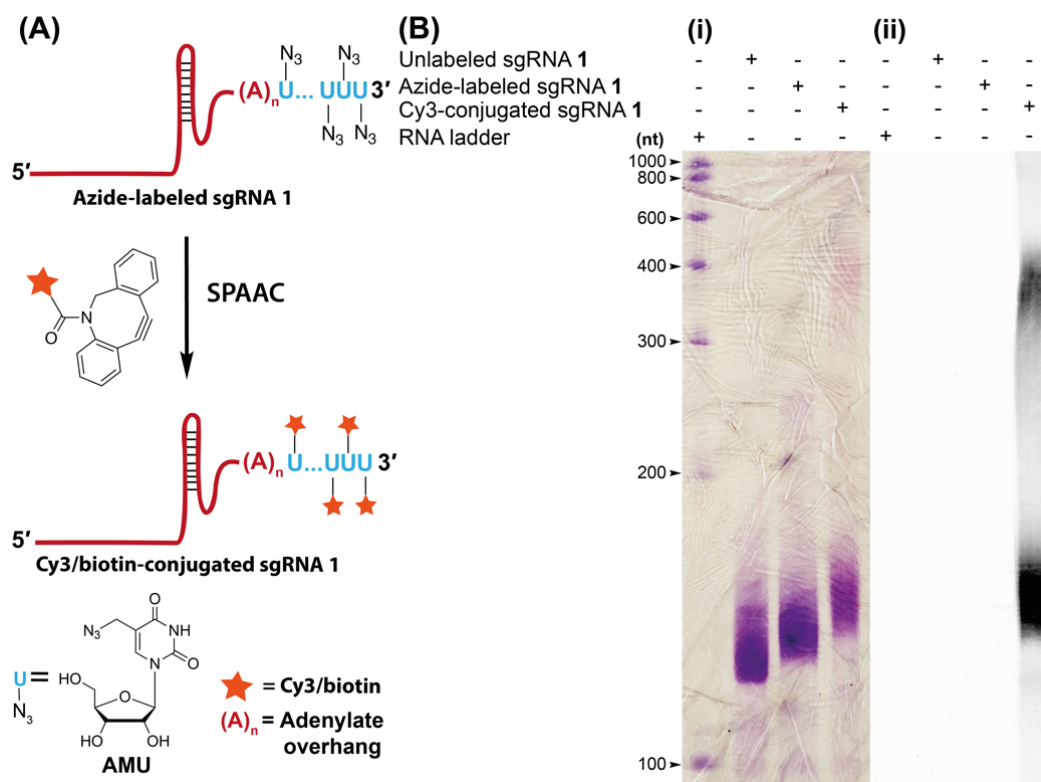
Figure 4.6. (A) SPAAC reaction on azide-labeled sgRNA **1** with Cy3/biotin-DBCO. **(B)** Product of SPAAC reaction with Cy3-DBCO was resolved and visualized by 8.5% denaturing PAGE using (i) Stains-All staining and (ii) Typhoon gel scanner at Cy3 wavelength.

4.2.2 *In vitro* cleavage assay with modified guide RNA

In order to explore whether the modified guide RNAs were functional, we sought to perform a cleavage reaction with the nuclease protein, Cas9 and a target dsDNA (216 bp) corresponding to a segment of eGFP gene, prepared by PCR (Table 2 and Figure 4.7A-C). eGFP was chosen as cleavage target since *in vitro* cleavage with telomere dsDNA (repetitive sequence) causes a frame shift of the protospacer resulting in multiple cleavage bands.

For performing *in vitro* cleavage using CRISPR-Cas9 genome editing tool, 1:1 equivalence of the unlabeled, azide-labeled or Cy3-conjugated sgRNA and Cas9 were incubated at room temperature for 10 min which enabled the formation of the ribonucleoprotein (RNP complex). Further, the target dsDNA was added to the RNP complex in binding buffer and incubated for 30 min at 37 °C. The reaction was quenched by adding formamide gel-loading buffer and was heated at 95 °C for 10 min. The cleavage products were visualized in 1.5% agarose gel. Bands when using unlabeled, azide-labeled or Cy3-conjugated sgRNA **2** showed a cleavage band (Figure 4.7). Both the cleavage products were visualized as a single band owing to its similar size. A decrease in efficiency of cleavage was observed with azide-labeled sgRNA **2** and Cy3-labeled sgRNA **2** potentially owing to an increase in steric hindrance. Nevertheless, the assay confirms that all the modified guide RNAs used were indeed functional.
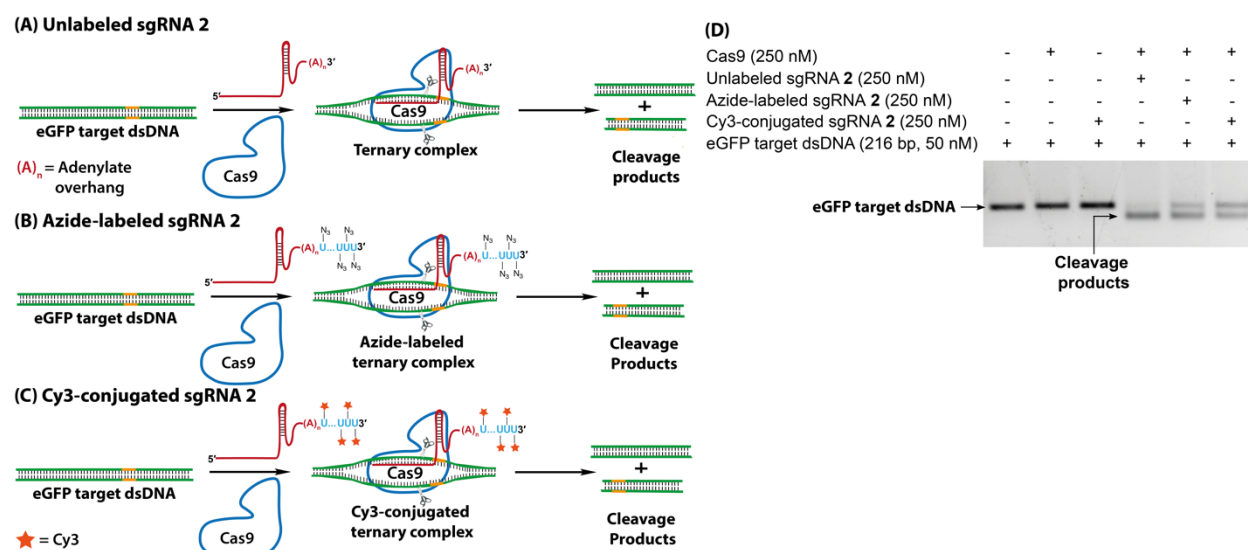
Figure 4.7. eGFP target dsDNA was incubated with Cas9 nuclease protein and (**A**) unlabeled sgRNA (**B**) azide-labeled or (**C**) Cy3-conjugated sgRNA **2**. (**D**) The cleavage products were run on 2.5% agarose gel stained with EtBr and was imaged.

Table 2. Sequence of RNA, primer and template DNA used.

DNA or RNA used	Sequence of oligonucleotide
sgRNA forward primer, CFP1	5' GCGTAATACGACTCACTATAGGGTTAGG 3'
sgRNA forward primer, CFP1'	5' GCGTAATACGACTCACTATAGTTAGGGTTAG 3'
sgRNA forward primer, CFP2	5' GCGTAATACGACTCACTATAGGCGA 3'
sgRNA reverse primer, CRP	5' TTTTTTTTTTTTTTTAAAGCACCGACTCGGTGC 3'
sgRNA template, CT1	5' GCGTAATACGACTCACTATAGGGTTAGGGTTAGGGTTAGGGTTAGTTTAAGAGCTATGCTGGAAAC-AGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT 3'
sgRNA template, CT1'	5' GCGTAATACGACTCACTATAGTTAGGGTTAGGGTTAGGGTTGTTTAAGAGCTATGCTGGAAACAG-CATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT 3'
sgRNA template, CT2	5' GCGTAATACGACTCACTATAGGCGAGGGCGATGCCACCTAGTTTAAGAGCTATGCTGGAAACAG-CATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT 3'
eGFP Template forward primer, eGFPF	5' AGGGCGAGGAGCTGTTCA 3'
eGFP Template reverse primer, eGFPR	5' GGTAGCGGCTGAAGCACT 3'
qPCR telomere forward primer, TeloF	5' GGTTTTTGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT 3'
qPCR telomere reverse primer, TeloR	5' TCCCGACTATCCCTATCCCTATCCCTATCCCTATCC-CTA 3'
5' FAM-labeled target DNA strand 1, TS1	5' FAM TAATGAATTCCCCAATACCCTAACCCTAACCCTAACCCTAACCCGTTCATATAA 3'
Target DNA strand 2, TS2	5' TTATATGAACGGGTTAGGGTTAGGGTTAGGGTTAGGGTATTGGGGAATTCATTA 3'
5' Cy5-labeled RNA	5' Cy5 GCGUGUCGUGCAGCCUCCG 3'
RNA 2	5' GGGUGCUCAGUACGAGAGGAACCGCACCC 3'
Splint DNA	5' GGGTGCGGTTCCTCTCGTACTGAGCACCCATGATGTATGGCACATGATTCTATGGTAA 3'

4.2.3 Binding affinity of modified sgRNA to dCas9 and target dsDNA

Electrophoretic mobility shift assay for modified CRISPR guide RNAs

The formation of ternary complex while using dCas9 the double mutant of Cas9 and modified guide RNAs can confirm whether a post-hybridization or pre-hybridization click strategy is feasible in the context of localizing modified sgRNAs to chromatin regions. For investigating the binding affinity of modified sgRNA to dsDNA and dCas9, unlabeled, natural UMP-tailed, azide-labeled, Cy3-conjugated or biotin-conjugated sgRNA **1** were incubated with dCas9 for 10 min in room temperature for the formation of RNP complex. The target dsDNA (labeled with 6-FAM on one strand) having a short telomere region, was prepared by annealing, DNA strands, TS1 and TS2. Further, this dsDNA (labeled with FAM on one strand) was added to RNP complex in binding buffer and incubated for 30 min at 37 °C (Figure 4.8A-D). Bands corresponding to the ternary complex were resolved by native PAGE at 4 °C and imaged using a gel scanner in FAM and Cy3 wavelength. The unlabeled sgRNA **1** showed complete binding to form the ternary complex evident from the formation of band displaying decreased migration as compared to dsDNA (Figure 4.8E). In contrary, UMP-tailed sgRNA showed very poor binding wherein majority of fraction remained in unbound state visualized as unbound dsDNA band. The azide-labeled sgRNA **1** showed excellent binding of ~85-90% with very less amount of the unbound fraction. It is to be noted that since AMUTP is incorporated with less efficiency compared to natural UTP in terminal uridylation reaction, there is lesser addition of azide modified UMPs at the 3′ end of azide-labeled

sgRNA **1** as opposed to natural UMPs in UMP-tailed sgRNA **1** (Figure 4.5C). The large number of UMP residues present at the 3′ end of UMP-tailed sgRNA **1** could hinder its binding to dCas9. Therefore, we can presume that the lesser incorporation efficiency of AMUTP during terminal uridylation, has indirectly assisted in dCas9 binding to azide-labeled sgRNA **1**. Biotin and Cy3-conjugated sgRNA **1** showed poor binding to form ~20-25% of ternary complex formation while majorly existed as unbound dsDNA. This can be attributed to the bulkiness induced by click conjugated probes which could decrease the binding of dCas9 protein to guide RNA. The bulkiness of modified sgRNAs used can also be visualized by the shifts in migration of the ternary complexes wherein, ternary complex corresponding to unlabeled sgRNA, migrates much faster than complex of modified sgRNAs (Figure 4.8E, dotted line). 5-azidomethyl uridine (AMU) being a conservative modification minimally hinders ternary complex formation and thereby could allow efficient binding of azide-labeled RNP complex to target dsDNA.
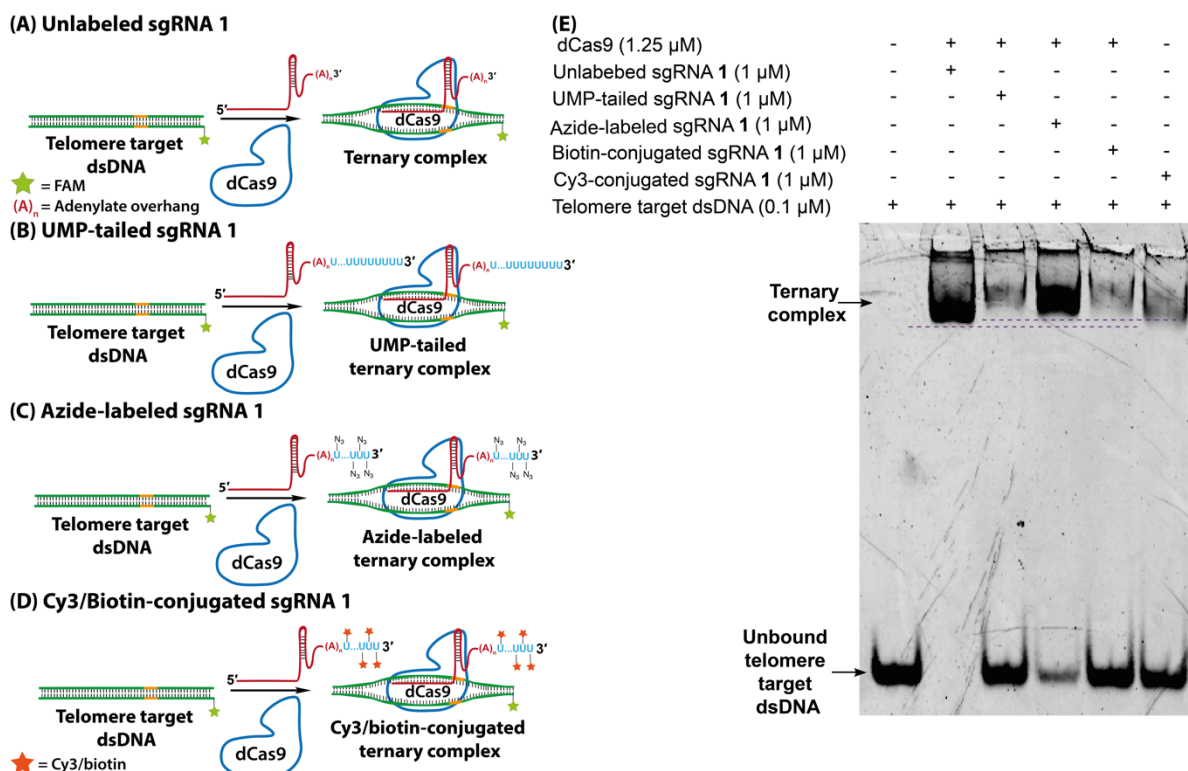
Figure 4.8. The (**A**) unlabeled, (**B**) UMP-tailed, (**C**) azide-labeled, (**D**) Cy3 or biotin-conjugated sgRNA **1** was incubated with dCas9 and target dsDNA to form the ternary complex. (**E**) The ternary complex was resolved in native 8% polyacrylamide gel and imaged in Typhoon gel scanner at FAM wavelength. A shift in migration of the ternary complex was observed with increase in bulkiness of modified sgRNA used (compare with violet line).

Quantitative measure of binding affinity of modified sgRNA to dCas9 employing microscale thermophoresis

In the CRISPR mechanism, it is known that sgRNA binds to Cas9/dCas9 and subsequently this binary complex then binds to the target dsDNA to form a ternary complex.[57] The major structural changes in protein complex occur during the formation of guide RNA-dCas9 RNP binary complex as compared to RNP-target dsDNA binding.[19, 58] Therefore, any inhibition of RNP complex formation would prevent binding to the target dsDNA. In order to investigate whether the changes in binding efficiency as visualized by EMSA is occurring as result of shift in equilibrium in modified sgRNA and dCas9 binary complex formation, we performed microscale thermophoresis (MST).[59-60] For this, we titrated dCas9 fused to eGFP namely, dCas9-eGFP (180 nM) with increasing concentration of unlabeled sgRNA **2**, azide-labeled sgRNA **2** or Cy3-conjugated sgRNA **2** (0.4 nM-12.5 μM). An apparent dissociation constant of 98.1 ± 22.1 nM was observed for the unlabeled sgRNA. The azide-labeled sgRNA and Cy3-conjugated sgRNA showed an apparent dissociation constant of 87.4 ± 19.8 nM and 799.5 ± 87.6 nM. Indeed, we observe that the binding of Cy3-conjugated sgRNA to dCas9 is inhibited as opposed to azide-labeled sgRNA (Figure 4.9). The MST and EMSA measurements together confirm that azide-labeled sgRNA binds effectively to form the ternary complex with almost good binding as unlabeled RNA. Importantly we can conclude from the assays that post-hybridization click chemistry is the method of choice for functionalizing azide-labeled sgRNAs as opposed to pre-hybridization click chemistry.
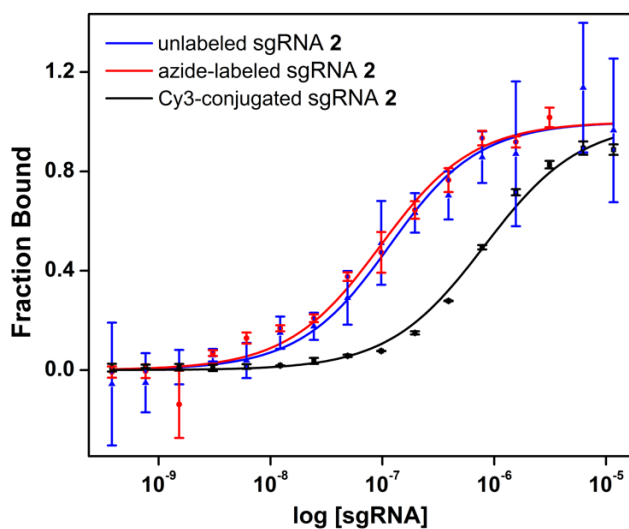
Figure 4.9. Binding of unlabeled sgRNA **2**, azide-labeled sgRNA **2** and Cy3-conjugated sgRNA **2** to dCas9-eGFP protein measured by microscale thermophoresis.

4.2.4 Site-specific cellular localization of azide-labeled sgRNA on gene locus and post hybridization functionalization using click chemistry.

CRISPR-FISH with azide-labeled guide RNA on mouse embryonic stem cells (mESC)

From EMSA and MST analysis we have concluded that azide-labeled sgRNAs can effectively bind to dCas9 and target dsDNA Herein we wanted to investigate if this can potentially be used for localizing azide tags to specific gene locus employing CRISPR system using a post-hybridization click approach. sgRNA and dCas9 tagged to a fluorophore protein/tag has been extensively used for imaging genomic loci.[23, 26] In order to check whether the azide-labeled sgRNA can indeed localize to telomeric regions as similar to unlabeled sgRNA, we performed CRISPR-FISH on mouse embryonic stem cells (mESC). Fixed, permeabilized mouse mES cells were incubated with RNP complex having either unlabeled sgRNA **1′**, azide-labeled sgRNA **1′** (targeting telomeric region) or azide-labeled sgRNA **2** (negative control) and dCas9-eGFP (200 nM, 1:1 equiv). Unlabeled and azide-labeled sgRNA **1′** has a 3-nucleotide omission in telomere protospacer region to match the sequence of guide sequence used previously for biotinylating and pull-down using CRISPR.[61] Nuclear puncta corresponding to telomere were visualized for both unlabeled sgRNA **1′** and also for azide-labeled sgRNA **1′** (Figure 4.10-11). The small decrease in binding of azide-labeled sgRNA **1′** over unlabeled sgRNA **1′** is evident from increase in background eGFP signal. Cells treated with unlabeled sgRNA **2** showed diffused background staining with no visible puncta whereas incubation with dCas9-eGFP alone resulted in aggregates mostly localized outside the nucleus. This confirms the cellular localization of azide-labeled sgRNA to telomeric regions.
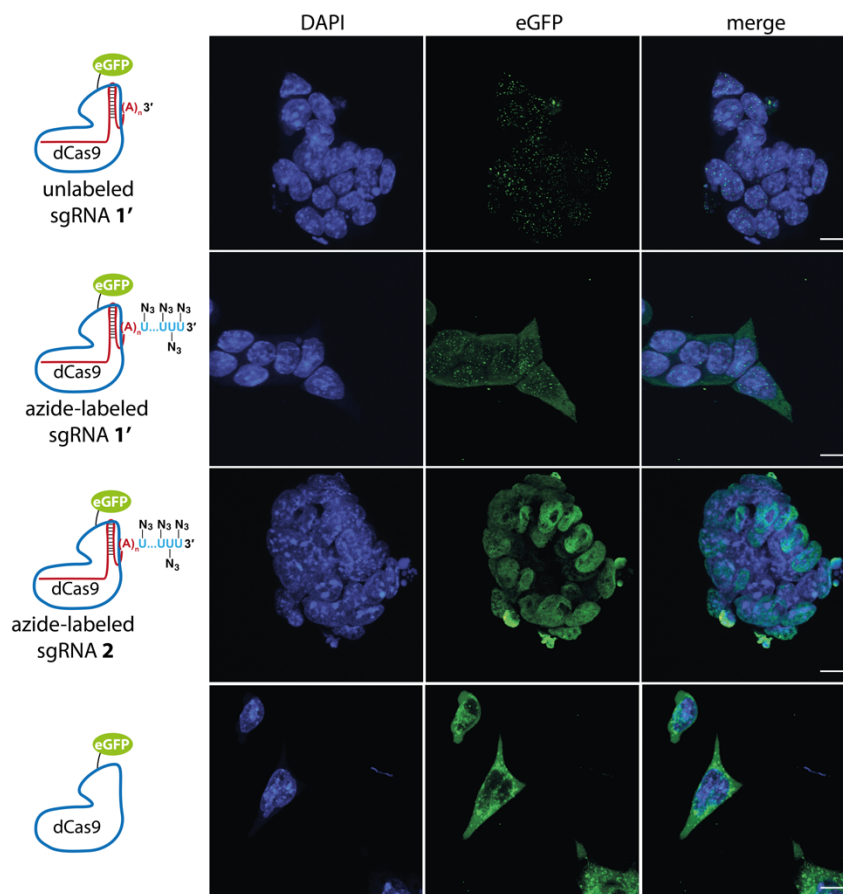
Figure 4.10. mESCs were treated with dCas9-eGFP alone or its RNP complex with unlabeled sgRNA **1′** (targeting telomere), azide-labeled sgRNA **1′** (targeting telomere) and azide-labeled sgRNA **2** (targeting eGFP: control). The cells were imaged in DAPI and eGFP channels and maximally projected Z-stacks are shown. First two rows: RNP complex of unlabeled sgRNA **1′** and azide-labeled sgRNA **1′** directs the localization to telomeric regions visualized as nuclear puncta. Third row: RNP complex of azide-labeled sgRNA **2** and dCas9-eGFP non-specifically distributes throughout the nucleus with no observable puncta. Last row: incubation with dCas9-eGFP alone resulted in aggregates, which were mostly localized outside the nucleus (scale bar, 10 μm).
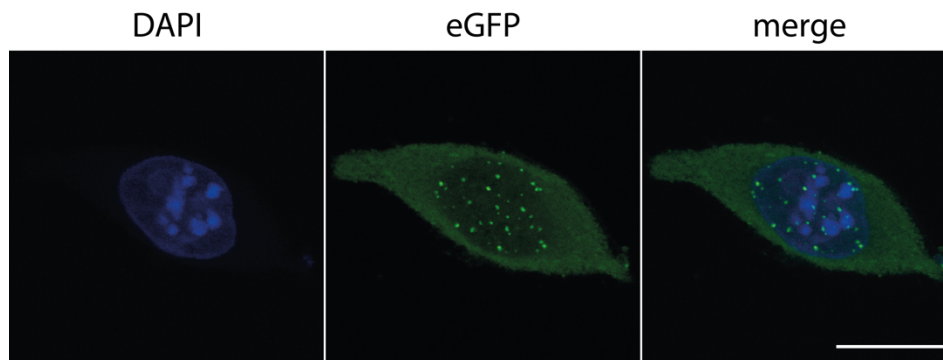
Figure 4.11. Maximally projected Z-stack image showing nuclear puncta localized to telomeres in a single embryonic stem cell obtained using RNP complex of azide-labeled sgRNA **1′** and dCas9-eGFP (scale bar, 10 μm).[23, 26]

Chromatin enrichment using post-hybridization click chemistry

Next, we wanted to investigate whether employing the post-hybridization click approach, we can perform targeted delivery of azide to gene loci for bioorthogonal click functionalization. For this as a proof-of-concept, we investigated the possibility of performing CuAAC or SPAAC functionalization of azide tags localized on gene loci with biotin tags for enrichment of chromatin regions using chromatin precipitation and streptavidin-biotin pull-down (Figure 4.12A). In order to perform this, we fixed, permeabilized mouse embryonic stem cells (mESC) and incubated it with azide-labeled sgRNA **1′** and dCas9-eGFP RNP complex (200 nM, 1:1 equiv) which would localize to telomeric regions. As a negative control, cells were treated with RNP complex having azide-labeled sgRNA **2** (targeting eGFP gene) and dCas9-eGFP (200 nM, 1:1 equiv) which should not enrich target telomeric regions. Using azide-labeled sgRNA **2** as a control, serves normalizing the background resulting from any non-specific enrichment from click reaction. The bound complex in telomere labeled cells can be confirmed by the presence of nuclear puncta as observed while performing CRISPR-FISH.[26] The cells were then washed and cross-linked using formaldehyde. After quenching the cross-linking reaction with glycine, the mESC cells were harvested. Cells were nuclear isolated and the cross-linked nuclear pellet was then fragmented by sonication until chromatin fragments of the range of 200-500 bp was observed when visualized by agarose gel. For samples wherein SPAAC reaction was performed, the lysate supernatant was treated additionally with iodoacetamide to reduce non-specific reaction. Click reaction was performed either using biotin-PEG-alkyne, $CuSO_4$, sodium ascorbate and THPTA or biotin sDIBO alkyne for 1 h at room temperature for both the samples (treated with azide-labeled sgRNA **1′** and **2**). sDIBO was used for SPAAC as DBCO/DIBAC based alkynes are known to show non-specific binding to cellular thiols upon lysate reaction.[62] The lysate was washed and buffer exchanged and lysate was then incubated with streptavidin coated magnetic beads. After overnight incubation, the beads were washed and further the captured chromatin region were eluted. qPCR was performed reaction in triplicate for samples treated with SPAAC or CuAAC reaction using telomeric qPCR primers used previously.[63] The fold enrichment for telomeric region was calculated for normalized pull down with azide-labeled sgRNA **1′** over normalized pull-down with azide-labeled sgRNA **2** (against eGFP). A higher enrichment was observed for telomere over eGFP for both CuAAC and SPAAC reaction respectively (Figure 4.12B). The telomeric region was enriched ~12 and ~6 fold for samples treated with CuAAC and SPAAC reaction respectively (Figure 4.12C). A better pull-

down was observed for CuAAC possibly owing to its fast kinetics and lower background over SPAAC.[62, 64] A twice higher telomere enrichment is observed for pull-down with CuAAC in comparison to the recent-reported CRISPR-peroxidase APEX2 enrichment system using APEX2 on MS2 coat protein non-covalently bound to sgRNA targeting telomere.[65] Also having azide tag opens the versatility of functionalizing azide-labeled loci with such functional proteins or small molecules probes having a suitable alkyne/strained alkyne tag. Overall, this confirms that post-hybridization click approach can indeed be used to functionalize targeted gene loci with highly functional probes which can be used for a variety of application.
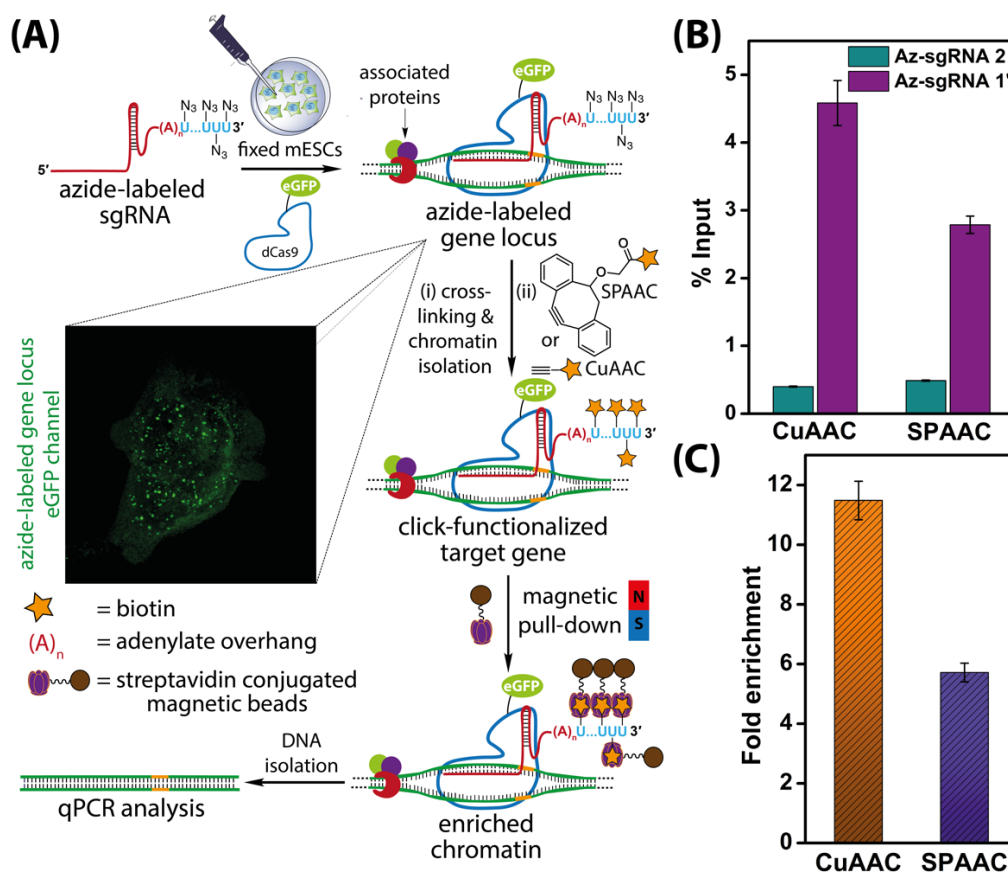
Figure 4.12. (**A**) Schematic diagram illustrating the steps involved in chromatin capture by *in situ* click reaction performed on the target-bound ternary complex using biotin-alkynes. While CuAAC reaction was performed using a biotin substrate containing a terminal alkyne, SPAAC reaction was performed using biotin-conjugated to a strained alkyne (sDIBO). Biotinylated chromatin was enriched using streptavidin beads and subjected to qPCR analysis. (**B**) qPCR analysis of enriched chromatin obtained by CuAAC and SPAAC reactions using telomere-targeting azide-labeled sgRNA **1′** and control non-targeting azide-labeled sgRNA **2**. The enrichment is expressed relative to respective inputs after click reaction step. (**C**) A plot showing the fold enrichment of telomere DNA using azide-labeled sgRNA **1′** normalized over azide-labeled sgRNA **2** for CuAAC and SPAAC reactions (see experimental section for details).

4.2.5 Utility of SpCID1 for developing FRET probes and site-specific internal labeling

Terminal uridylation of a potentially structured RNA with a single azide tag

In chapter 3, we had varied the enzyme and triphosphate stoichiometry to fine tune the addition of a modified UMP at the 3′ end of RNA. In order to explore the utility of the labeling technology for constructing FRET pairs employing terminal uridylation, we chose a potentially structured RNA using secondary structure prediction tools (Figure 4.13A).[53] The RNA structure was also designed to be not too strong so as to inhibit SpCID1 binding.
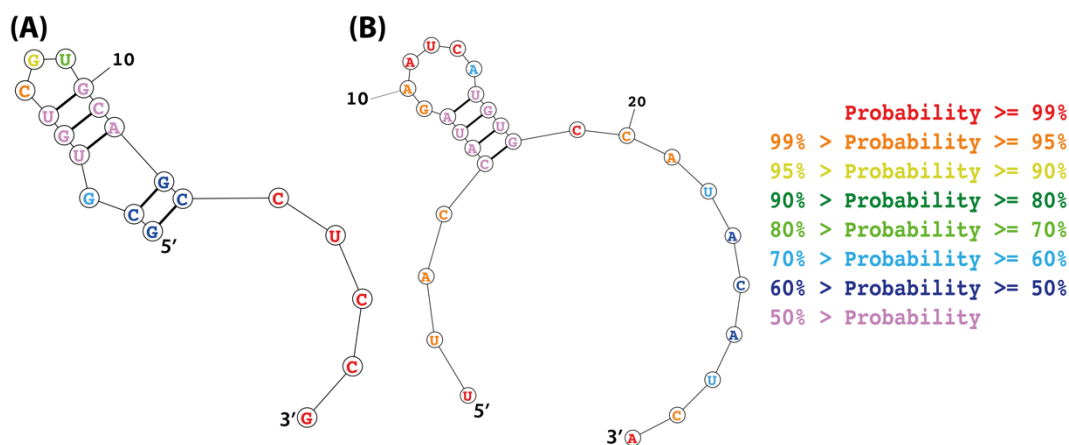
Figure 4.13. Secondary structure prediction of (**A**) 5′ Cy5-labeled RNA and (**B**) 5′ FAM-labeled RNA used in chapter 3. Based on prediction, an addition of a single UMP residue at the 3′ end of both RNAs doesn't change the overall secondary structure of RNA. RNA Structure prediction performed using online software, RNAstructure.[53]

To construct an RNA showing FRET process, we acquired an RNA pre-labeled at its 5′ end with a Cy5 tag (Table 2). Cy3-Cy5 FRET systems are one of the most popular FRET pairs used for single molecule studies on RNA.[66] The 5′ Cy5-labeled RNA (10 µM) was incubated with SpCID1 (1 µM) and AMUTP (0.5 mM) at 37 °C for 5, 15 and 30 min after which aliquots of reaction mixture were quenched and loaded on to 20% denaturing polyacrylamide gel. The reaction gave almost maximum yield of single incorporation with minor amounts of starting RNA and double azide-labeled RNA (Figure 4.14). A higher stoichiometry of SpCID1 enzyme and AMUTP was required for single incorporation of AMU at the 3′ end of 5′ Cy5-labeled RNA as compared to the 5′ FAM-labeled RNA used in chapter 3. This indirectly agrees with the observed structure prediction wherein 5′ Cy5-labeled RNA formed a stronger hairpin with lesser accessible 3′ single stranded overhang as compared to 5′ FAM-labeled RNA (Figure 4.13). In order to synthesize

AMU-labeled 5′ Cy5 RNA in large-scale, several 20 μL (200 pmol) reactions were performed for 15 min and product was isolated by HPLC chromatography. Typically, a 1 nmol reaction yielded 0.85 nmol of AMU-labeled 5′ Cy5 RNA (Table 1).
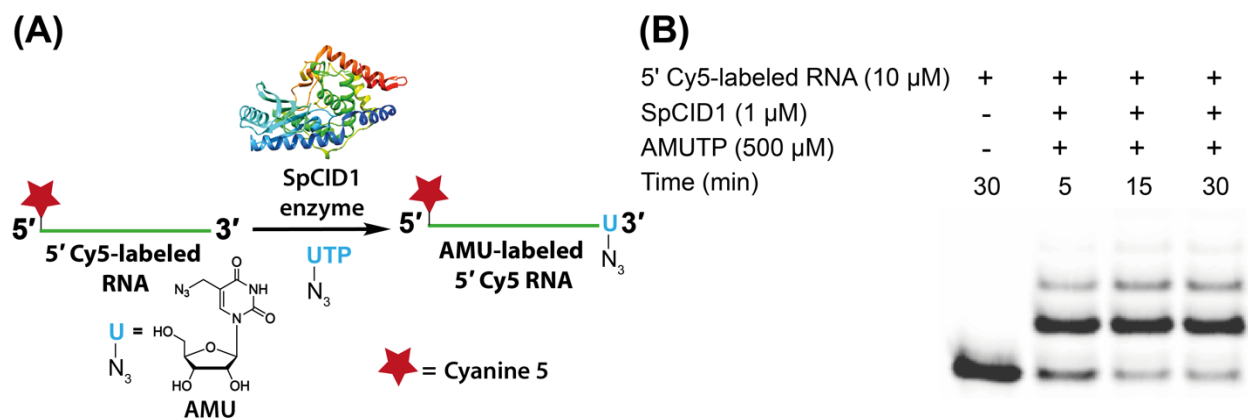

Figure 4.14. (**A**) The 5′ Cy5-labeled RNA was subjected to uridylation reaction using SpCID1 and AMUTP to get a single azide incorporated, AMU-labeled 5′ Cy5 RNA. (**B**) The reaction was monitored by gel to optimize maximum single incorporation of AMU.

Click labeling of terminal uridylated RNA

The 3′ azide-labeled RNA was subjected to SPAAC reaction using Cy3-DBCO (1 mM) for 2 h at 37 ˚C (Figure 4.15). The RNA was precipitated and further purified by gel electrophoresis. Band corresponding to Cy5-Cy3 dual-labeled RNA was gel extracted and eluted. A typical 0.6 nmol reaction yielded, 0.48 nmol of Cy5-Cy3 RNA (Table 1).
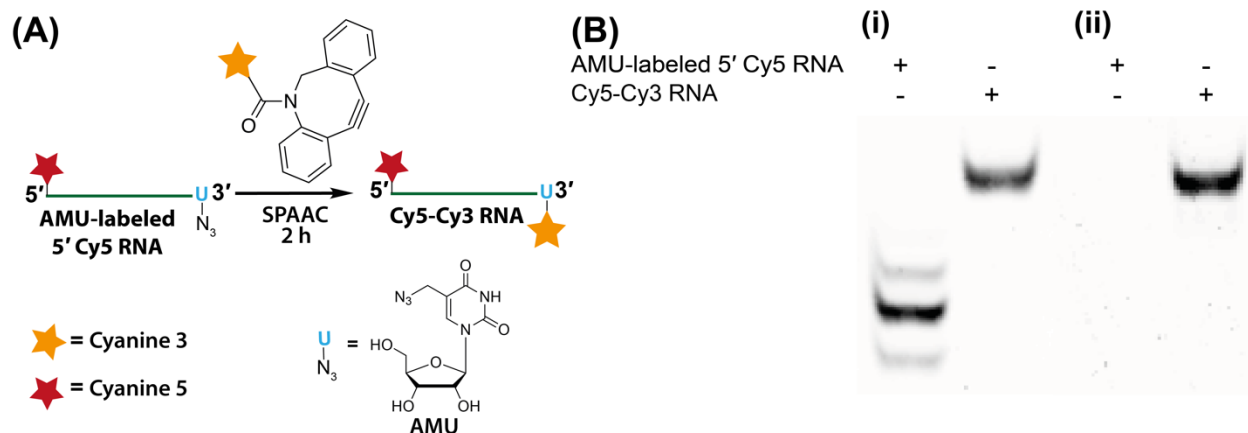
Figure 4.15. (**A**) AMU-labeled 5′ Cy5 RNA was click reacted with Cy3-DBCO to get dual labeled Cy5-Cy3 RNA. (**B**) Gel electrophoresis of purified samples corresponding to AMU-labeled 5′ Cy5 RNA (purified by HPLC) and Cy5-Cy3 RNA (purified by gel) visualized in typhoon gel scanner at (**i**) Cy5 and (**ii**) Cy3 wavelength.

Föster resonance energy transfer on dual Cy5-Cy3 labeled RNA

We have investigated Föster resonance energy transfer (FRET) in Cy5-Cy3 RNA by comparing its emission with 5′ Cy5-labeled RNA and Cy3-DBCO. These compounds were excited at 520 nm which corresponds to Cy3 (donor) absorption. The dual-labeled RNA showed strong emission corresponding to the Cy5 (acceptor), which was found to be higher in intensity than emission of 5′ Cy5-labeled RNA excited at 520 nm (Figure 4.16). To further confirm the FRET process in Cy5-Cy3 RNA, the dual labeled RNA was degraded using RNase A. A time-dependent decrease in Cy5 (acceptor) emission intensity with a concomitant increase in Cy3 (donor) emission intensity almost matching the emission intensity of Cy3-DBCO was observed when excited at 520 nm (Figure 4.16B). These fluorescence measurements confirm the presence of high FRET in RNA when labeled with Cy3 and Cy5 pairs, similar to that observed earlier.[67]
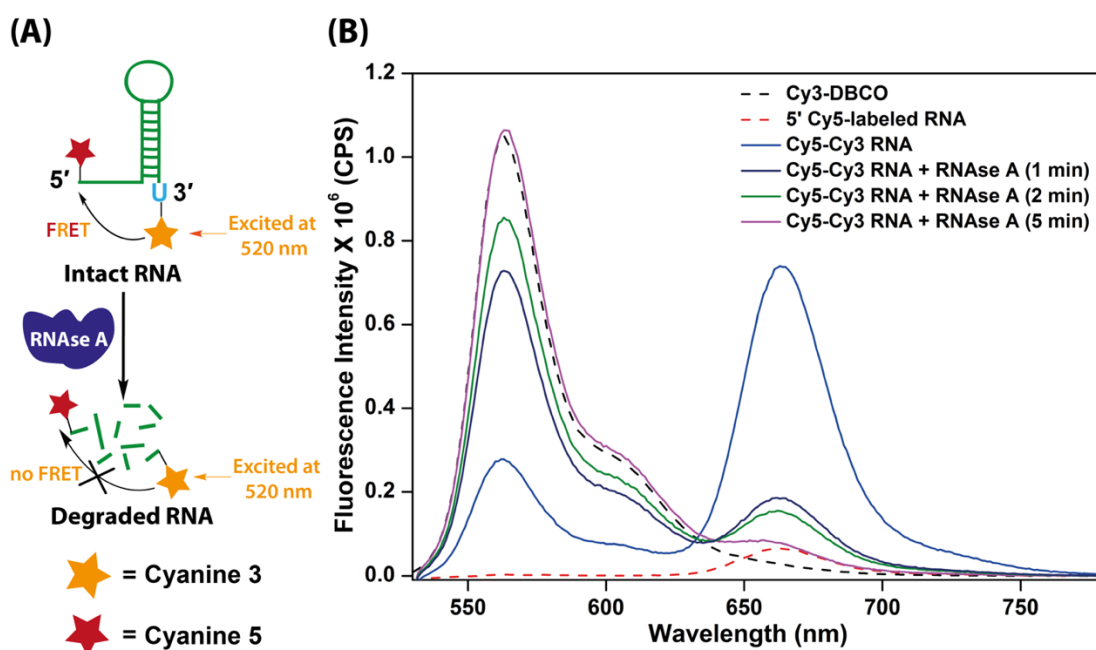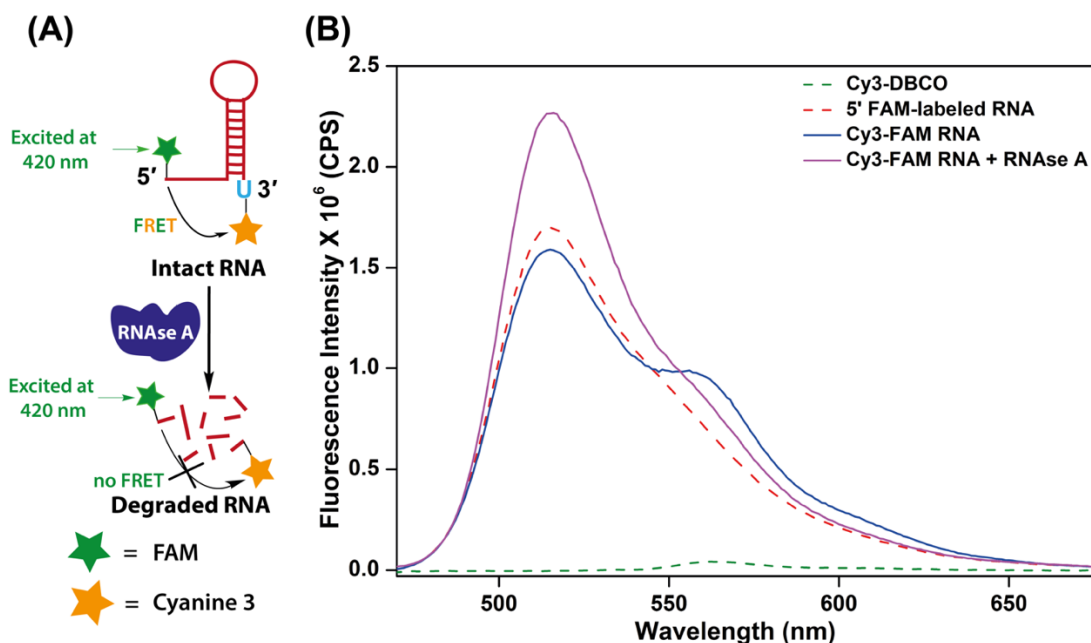

Figure 4.16. (**A**) Schematic showing Cy3 and Cy5 label at 5′ and 3′ end of Cy5-Cy3 RNA. To validate the presence of FRET, Cy5-Cy3 RNA was degraded by RNase A enzyme which cleaves RNA at pyrimidine nucleotides. (**B**) Cy3-DBCO, 5′ Cy5-labeled, Cy5-Cy3 RNA was excited at 520 nm and emission was recorded. Upon RNase A treatment, fluorescence was recorded at 1, 2 and 5 min, respectively. Excitation and emission slit width of 5 and 5 nm was maintained for all experiments.

Emission spectra of fluorescein (FAM) and absorption spectra of Cy3 similar to Cy3-Cy5 pair, shows good spectral overlap and has been used for FRET studies.[68] To investigate if the dual labeled FAM-Cy3 RNA synthesized in the previous chapter shows FRET, we performed a similar

experiment wherein we compared its fluorescence with 5′ FAM-labeled RNA and Cy3-DBCO. All compounds were excited at 450 nm wavelength corresponding to FAM (donor) absorption (Figure 17). A fluorescence band corresponding to Cy3 emission was observed for the FAM-Cy3 RNA. Upon degrading the RNA with RNase A enzyme, the emission intensity of Cy3 (acceptor) decreased accompanied with a small increase in FAM (donor) emission intensity when excited at 450 nm. This further validates the existence of FRET on the dual labeled RNA (Figure 4.17B).

Figure 4.17. (**A**) Schematic showing FAM and Cy3 label at 5′ and 3′ end of Cy3-FAM RNA. To validate the presence of FRET, Cy3-FAM RNA was degraded by RNase A enzyme which cleaves RNA at pyrimidine nucleotides. (**B**) Cy3-DBCO, 5′ FAM-labeled RNA, Cy3-FAM RNA was excited at 450 nm and emission was recorded. Upon RNase A treatment, fluorescence of degraded RNA was recorded. Excitation and emission slit width of 5 and 5 nm was maintained for all experiments.

4.2.6 Site-specific internal labeling of terminal uridylated azide-labeled RNA

Enzymatic ligation of RNA is typically accomplished using an RNA having free 3′ hydroxyl and a second RNA, monophosphorylated at its 5′ end. An additional ssDNA called the splint DNA, complementary to both RNA is used so as to bring both free hydroxyl and monophosphorylated 5′ end within close proximity (Figure 4.18A). Also, usage of a ssDNA splint reduces chances of self-ligation. Addition of 5-azidomethyl uridine monophosphate (AMUMP) at the 3′ end of RNA employing terminal uridylyl transferase gives access to RNA having free 3′ hydroxyl group which can be utilized for RNA ligation to synthesize a longer RNA having site-specific internal label. For performing RNA ligation, we monophosphorylated an RNA (RNA **2**) using T4 polynucleotide

kinase and the product obtained was purified by HPLC (Table 2). The monophosphorylated RNA **2** (5 equiv) was incubated with AMU-labeled 5′ FAM RNA (1 equiv) and complementary ssDNA splint in the presence of T4 DNA Ligase or T4 RNA Ligase 2. The reaction was incubated for 1 h at 37 °C. Further the ssDNA splint was degraded by DNase I treatment for 20 min at 37 °C and heat inactivated. The product of ligation was resolved on 16% denaturing PAGE and was imaged by gel scanner at FAM wavelength. Almost complete conversion of the starting material was observed when using both T4 DNA Ligase and T4 RNA Ligase 2. A new band of lower migration corresponding to the ligated product was visualized (Figure 4.18B).
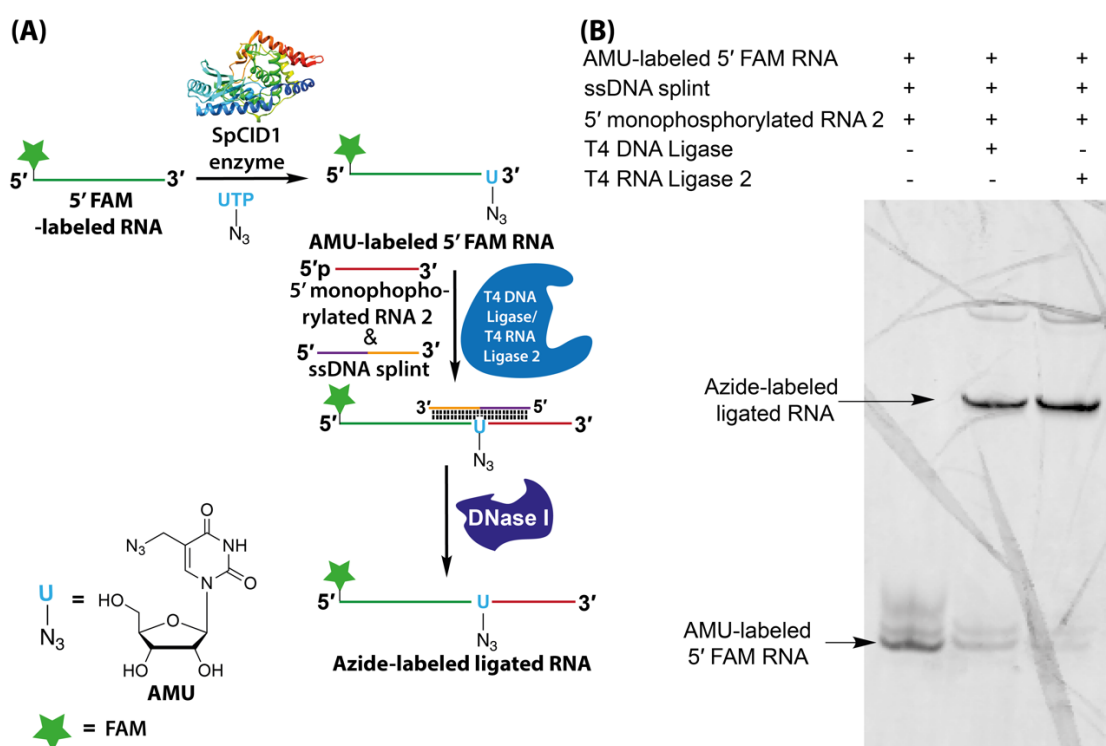
Figure 4.18. (**A**) RNA was terminal uridylated to get the AMU-labeled 5′ FAM RNA which is ligated with a 5′ monophosphorylated RNA **2** using a complementary ssDNA splint and T4 DNA Ligase/T4 RNA ligase 2. The ssDNA was degraded by DNase I treatment and the products were analyzed by PAGE. (**B**) The product was resolved on 16% PAGE and imaged in Typhoon scanner at FAM wavelength. Structure of SpCID1 adopted from PDB: 4FH3.

4.3 Conclusions

Terminal uridylation of RNA can be used for a broad range of application depending on the need for single or multiple modifications. Site-specific high-density azide labeling of a highly functional long RNA, the CRISPR guide RNA is possible using SpCID1 enzymes. Herein we report for the

first time, an enzyme has been used to introduce multiple bioorthogonal clickable tags on CRISPR sgRNA. Further, we find that CRISPR sgRNA having multiple azide tags minimally perturb the binding and formation of the ternary complex with dCas9 and target dsDNA and therefore post-hybridization click is the method of choice for incorporating functional tags on the ternary complex. Using this approach, we have developed a novel technique for localizing azide-labeled sgRNAs site-specifically to gene locus of interest in fixed cells. Subsequently the azide tagged gene loci were cross-linked and subjected to post-hybridization click chemistry to introduce multiple biotin tags. These biotin-tagged telomeric gene regions were enriched by streptavidin-based pull-down using magnetic beads. Therefore, the very fact that post-hybridization click chemistry works on CRISPR system open up new possibilities for functionalization any gene locus with bioorthogonal click tools for site-specific delivery of functional tag or warheads utilizing the combined power of CRISPR-Cas9 and bioorthogonal click chemistry.

Next, by utilizing the single 3′ azide-labeled RNA, we prove that terminal uridylation can be used to construct RNA FRET pairs which opens a new mean to label RNA for ensemble or single molecule FRET studies. In turn, such a strategy can be employed for chemo-enzymatic labeling of RNA in native state for single molecule FRET studies.[39] We also show that the 3′ azide modified RNA transcript are processed by ligases to produce long site-specific internal labeled RNA transcripts. An internal click labels on a RNA transcript can in principle be used for diverse application from designing internal biophysical probes to attaching any small molecule or macro-molecule having the cognate functionality on any U residue on RNA[29].

4.4 Experimental Section

4.4.1 Materials

DNA oligonucleotides (ONs) were purchased from Integrated DNA Technologie, Inc. and Sigma Aldrich. T7 RNA polymerase, NTPs, ribonuclease inhibitor RiboLock, SUPERase In™ RNase Inhibitor RiboRuler, Low Range RNA Ladder, MEGAclear™ Transcription Clean-Up kit, DMEM, FBS, Proteinase K Click-iT™ Biotin sDIBO Alkyne were obtained from Thermo Fischer Scientific. Cy3-DBCO, biotin-DBCO and Biotin-PEG4-Alkyne were obtained from Click Chemistry Tools. NucleoSpin® PCR Clean-up kit and QIAquick PCR purification kit was purchased from Macherey-Nagel. Qiagen. Sep-Pak Classic C18 cartridges for desalting oligonucleotides was obtained from Waters Corporation. EmeraldAmp GT PCR master mix and

DNA ladders were purchased from Takara. 5′ Cy5-labeled RNA was purchased from Dharmacon Inc. Stains-All, Iodoacetamide and reagents for buffer (Bio Ultra Grade) were acquired from Sigma Aldrich. EmeraldAMP GT PCR master mix, TB Green™ Premix Ex Taq™ II was obtained from TakaraBio. Taq DNA polymerase was purchased from Geneaid. cOmplete EDTA-free protease inhibitor cocktail was purchased from Roche. Standard treated quartz capillary for MST was acquired from nanoTEMPER technologies. All RNA experiments were performed in nuclease-free water.

4.4.2 Instruments

RNA was quantified by measuring absorbance at 260 nm in UV-2600 Shimadzu or NanoDrop™ 2000c spectrophotometer. ESI-MS mass analysis of RNA was performed using Waters SYNAPT G2-Si Mass Spectrometry instrument in negative mode. HPLC analysis was performed on an Agilent Technologies 1260 Infinity HPLC. Polymerase chain reaction was performed on Mastercycler® pro PCR machine from Eppendorf. Fluorescent RNA was resolved by polyacrylamide on OWL S4S sequencing gel electrophoresis instrument and was imaged using Typhoon TRIO+ Variable mode Imager. Fluorescence spectra were recorded in micro fluorescence cuvette (Hellma, path length 1.0 cm) on Horiba Fluoromax 4 spectrofluorometer. MST was measure Monolith NT.115 from nanoTEMPER technologies. Sonication for chromatin precipitation was performed in Bioruptor® Plus sonication device from Diagenode. qPCR was performed in LightCycler® 480 Instrument II from Roche.

4.4.3 Terminal functionalization of CRISPR guide RNA
PCR amplification of sgRNA template

In order to PCR amplify the template required for *in vitro* transcribing CRISPR sgRNAs, the sgRNA forward and reverse primers, CFP and CRP (1 μM) were incubated with sgRNA template, CT (44 nM) in EmeraldAMP GT PCR master mix in a final volume of 25 μL. PCR conditions: heat denaturation at 94 °C for 1 min, 35 cycles of (denaturing: 94 °C for 20 s, annealing: 64 °C for 20 s, extension: 68 °C for 30 s), final extension at 68 °C for 5 min, 4 °C hold. Several small-scale PCR were performed to isolate template required for large-scale transcription. The amplicon was purified using NucleoSpin® Gel and PCR Clean-up kit (Table 2).

In vitro transcription of CRISPR sgRNA

Transcription reaction of CRISPR sgRNA was performed using ATP, GTP, CTP and UTP (2 mM), RiboLock RNase inhibitor (0.4 U/μL), dsDNA template (generated by PCR, 300 nM) and T7 RNA polymerase (3.2 U/μL) in 40 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 20 mM DTT, 10 mM NaCl and 1 mM spermidine in a final reaction volume of 250 μL. The reaction was incubated at 37 °C for 12 h. The transcription reaction was centrifuged and unlabeled sgRNAs were isolated by MEGAclear™ Transcription Clean-Up kit using manufacture's protocol. The purity and integrity of isolated RNA were confirmed by 3.5% denaturing agarose gel run in 1X MOPS buffer and visualized using SYBR® Safe gel staining reagent. A typical 250 μL transcription reaction yielded 2-3 nmol of unlabeled sgRNA **1-2**.

Terminal uridylation of sgRNA using SpCID1

In vitro transcribed unlabeled sgRNA (10 μM, 200 pmol) was incubated with natural UTP or 5-azidomethyl uridine triphosphate, AMUTP (500 μM, 50 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and in-house expressed, SpCID1 (20.5 pmol) in a final volume of 20 μL. The reaction was incubated at 37 °C for 30 min and further heat-inactivated by incubating at 75 ˚C for 3 min. The reaction mixture (150 pmol) was resolved in 8.5% denaturing polyacrylamide gel and was visualized by Stains-All reagent. The reaction was also resolved by 3.5% denaturing agarose gel run in 1X MOPS buffer and visualized using SYBR® Safe gel staining reagent.

Typically for a 1 nmol reaction, five small-scale 20 μL (200 pmol) reactions were performed. After heat-inactivation, the reactions were pooled together and azide-labeled sgRNA was precipitated by adding 1 volume of 5 M ammonium acetate, 10 volume of ethanol and incubated for 2 h at -20 °C. The sample was centrifuged at 15,000 rpm for 15 min and RNA pellet obtained was washed with chilled 75% ethanol in nuclease free water and resuspended in nuclease free water to obtain ~90% yield of azide-labeled sgRNA **1**. Reaction with unlabeled sgRNA **1′** and **2** performed in 2 nmol scale, gave azide-labeled sgRNA products in 93% and 96% yield (Figure 4.5 and Table 1).

Functionalization of azide-labeled sgRNA using click chemistry

In order to conjugate azide-labeled sgRNA with various functional tags, the guide RNA (10 μM)

was incubated with Cy3-DBCO or biotin-DBCO (1 mM) in a final volume of 30 μL for 2 h at 37 °C. The product obtained were precipitated overnight using 1 volume of 5 M ammonium acetate and 10 volume of ethanol and centrifuged at 15000 rpm for 15 min. RNA pellet obtained, was further washed with 75% ethanol in water and resuspended in nuclease free water to obtain the product, Cy3 and biotin-conjugated sgRNA **1** in ~90 and ~77% yield. Reaction with 1.8 nmol of azide-labeled sgRNA **2** (in 180 μL volume) yielded product Cy3-conjugated sgRNA **2** in 92% yield. Unlabeled sgRNA **1**, azide-labeled sgRNA **1** and Cy3-conjugated sgRNA **1** (~150 pmol) were resolved by 8.5% denaturing polyacrylamide gel and imaged by staining with Stains-All reagent and in Typhoon gel scanner in Cy3 wavelength (Figure 4.6 and Table 1).

4.4.4 *In vitro* cleavage assay with modified guide RNAs

PCR for generating target dsDNA eGFP template

In order to PCR generate the 216 bp dsDNA template required for checking cleavage, eGFP forward and reverse primers, eGFPF and eGFPR (0.5 μM), dNTPs (0.2 mM) were incubated with linearized eGFP-N1 plasmid (100 ng) with Taq DNA polymerase (2.5 U) in 1X PCR buffer final volume of 50 μL. PCR conditions: heat denaturation at 95 ˚C for 5 min, 35 cycles of (denaturing: 95 ˚C for 30 s, annealing: 58 ˚C for 30 s, extension: 72 ˚C for 30 s), final extension at 72 ˚C for 5 min, 4 ˚C hold. Several small-scale PCR were performed to isolate template required for large-scale transcription. The amplicon was purified using PCR Clean-up kit (Table 2).

***In vitro* cleavage of eGFP target dsDNA using modified sgRNAs**

In vitro cleavage using CRISPR-Cas9 genome editing tool was performed in 1:1 equivalence wherein 1.5 μL of 2.5 μM, unlabeled, azide-labeled or Cy3-conjugated sgRNA **2** were incubated with 1.5 μL of 2.5 μM Cas9 at room temperature for 10 min. Target dsDNA generated by PCR was added to the RNP complex (final concentration 50 nM of dsDNA and 250 nM of RNP complex) in 20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT to make a final volume of 15 μL. The cleavage reaction was incubated for 30 min at 37 °C. Further proteinase K (40 μg) was added to the reaction and incubated at 55 °C for 30 min. Further proteinase K was deactivated by incubating at 70 °C for 10 min followed by the addition of RNase A (10 μg). The reaction was quenched by the addition of DNA gel loading buffer and was loaded onto 2% agarose and resolved imaged using EtBr (Figure 4.7 and Table 2).

4.4.5 Binding of azide-labeled sgRNA to dCas9 and target dsDNA and cellular functionalization of gene loci with azide tags

Electrophoretic mobility shift assay for modified CRISPR guide RNA and dCas9 complex

In order to check the binding of modified guide RNA to dsDNA and dCas9, unlabeled, UMP-tailed, azide-labeled, Cy3-conjugated or biotin-conjugated sgRNA **1** (2 μM) and dCas9 (2.5 μM) in the presence of RiboLock RNase inhibitor (4 U/μL) were incubated in a volume of 5 μL at room temperature for 10 min. Further, 5 μL solution of target dsDNA (0.2 μM) and DTT (20 μM) was added to the ribonucleoprotein complex in binding buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 1 mg/mL BSA, 0.1% Triton X-100, 5% glycerol) to make a final reaction volume of 10 μL. Final concentration of Cas9-sgRNA RNP is 1 μM and target dsDNA is 0.1 μM. The binding reaction was incubated for 30 min at 37 °C. The samples (0.5 pmol of 5′ FAM-labeled DNA) were resolved by 8% native polyacrylamide gel at 4 °C in 1X TBE buffer. Both gel and running buffer was supplemented with a final concentration of 10 mM $MgCl_2$. The gel was imaged in Typhoon gel scanner at FAM wavelength (Figure 4.8).

Microscale thermophoresis (MST) measurement for calculating binding between modified sgRNA and dCas9

To investigate the binding of modified sgRNA to dCas9, we incubated samples having dCas9-eGFP (180 nM) in 20 mM HEPES (pH 7.5), 150 mM KCl, 10 mM $MgCl_2$ 1 mM DTT and varying concentrations of unlabeled, azide-labeled and Cy3-conjugated sgRNA **2** (0.4 nM to 12.5 μM) in a final volume of 10 μL for 10 min at room temperature. The samples (4-5 μL) were loaded onto standard treated quartz capillary and excited at 60% LED power and 40% MST power at 25 °C using blue filter and MST was recorded (Figure 4.9).

CRISPR FISH using azide-labeled sgRNA

Mouse embryonic stem cells (R1/E) were seeded in coverslips and cultured in DMEM in presence of leukamia inhibiting factor (LIF) at 37 °C and 5% CO2 to maintain cells its pluripotent state. Post-seeding (24 h), the media was removed and cells were washed in 1X PBS. The cells were then fixed in pre-chilled 50% acetic acid in methanol and incubated at -20 °C for 20 min. The cells were washed 3 times with 1X PBS for 5 min with gentle shaking. Further, the cells were permeabilized by incubating in hybridization buffer (20 mM HEPES (pH 7.5), 100 mM KCl, 5

mM $MgCl_2$, 1 mM DTT, 2% BSA, 5% glycerol and 0.1% Triton-X) at 37 °C for 30 min. Either dCas9-eGFP (200 nM) alone or dCas9-eGFP (200 nM) with unlabeled sgRNA **1′**, azide-labeled sgRNA **1′** or azide-labeled sgRNA **2** (200 nM) and RiboLock RNase inhibitor (1.6 U/µL), were incubated for 20 min at room temperature for the formation of RNP complex in hybridization buffer in a final volume of 50 µL. The cells were incubated with RNP complex in a humid chamber for 4 h at 37 °C and was then washed 3 times with hybridization buffer for 5 min. The cover-slips were stained with 5 µg/mL DAPI and mounted on glass slides for imaging. The cells were visualized in Leica TCS SP8 in DAPI and eGFP channel. Images were deconvoluted on Leica LAS X software using blind deconvolution and Z-stack were projected (maximum intensity) using Fiji image analysis software (Figure 4.10-11).

Chromatin precipitation using azide-labeled sgRNA

Mouse embryonic stem cells (R1/E) were seeded in 10 cm dish and cultured in DMEM in presence of leukamia inhibiting factor (LIF) at 37 °C and 5% CO_2. At around >70% confluency cells were fixed and permeabilized as performed for CRISPR-FISH. dCas9-eGFP (200 nM), azide-labeled sgRNA **1′** or **2** and SUPERaseIn RNAse Inhibitor (0.1 U/µL), were incubated for 20 min at room temperature for the formation of RNP complex in hybridization buffer in a final volume of 1 mL. The cells were incubated with RNP complex in a humid chamber for 1.5 h at 37 °C and then washed with 1X PBS for 5 min. Cells were cross-linked with 1% formaldehyde in DMEM for 10 min at room temperature with gentle shaking. Further, glycine (125 mM) was added to quench the extra formaldehyde and incubated for 5 min. Cells were harvested and washed twice with 1X protease inhibitor cocktail (PIC) in 1X PBS. For nuclear isolation, cells were resuspended in 2 ml of PBS followed by the addition of 2 mL of nuclear isolation buffer (40 mM Tris HCl pH 7.5, 1.28 M sucrose, 20 mM $MgCl_2$ and 4% Triton X-100) in a final volume of 10 mL and incubated on ice for 20 min with frequent mixing. The nuclear isolate was centrifuged at 2500 g for 15 min at 4 °C. The nuclear pellet was resuspended in 1 mL RNA immunoprecipitation buffer (25 mM Tris HCl pH 7.5, 150 mM KCl, 5 mM EDTA pH 8.0, 0.5 mM DTT, 0.5% Igpal, 0.1 U/µL SUPERaseIn RNAse Inhibitor, 1X PIC) and sonicated for 4 cycles of 15 min (30s on and 30s off) at 4 °C in Bioruptor® Plus sonication device. After sonication the fragment size (200-500 bp) was confirmed by agarose gel electrophoresis. The lysate was centrifuged at 13,300 rpm for 10 min and supernatant was transferred to a fresh vial. For SPAAC reaction, lysates corresponding to azide-

labeled sgRNA **1′** and **2** were treated with 50 mM iodoacetamide in a final volume of 1 mL for 60 min at room temperature. Samples were buffer-exchanged (10 times volume) to 1X PBS in 10 kDa cut-off columns at 4 °C. Further samples for CuAAC reaction (corresponding to azide-labeled sgRNA **1′** and **2**) were treated with 100 μM Biotin PEG4 Alkyne, 1 mM, $CuSO_4$, 1 mM sodium ascorbate, 1 mM THPTA and SPAAC reaction (corresponding to azide-labeled sgRNA **1′** and **2**) were treated with 100 μM Biotin sDIBO Alkyne in a final volume of 500 μL having 0.1 U/μL SUPERaseIn RNAse Inhibitor. The reactions were incubated for 1 h at room temperature. The excess biotin was removed by buffer exchanging (20 times volume) to 1X PBS in 10 kDa size cut-off columns to a final volume of 1 mL. Meanwhile, 50 uL of Dynabeads MyOne Streptavidin C1 magnetic beads for each sample were washed thrice with 1X PBS. 5% volume inputs (50 μL) were taken out, stored at -20 °C and remaining samples were treated with 50 μL of Dynabeads MyOne Streptavidin C1 magnetic beads along with 0.1 U/μL SUPERaseIn RNAse Inhibitor and incubated overnight with gentle rotation at 4 °C. Next, pull-down samples were washed with 1 ml of each wash buffer (1X low salt buffer: 20 mM Tris HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS, 1X high salt buffer: 20 mM Tris HCl pH 8.0, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS, 1X LiCl buffer: 10 mM Tris HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 1% sodium deoxycholate, 1% Igpal, 2X TE buffer: 10 mM Tris HCl pH 8.0, 1 mM EDTA pH 8.0) for 5 min at 4 °C with gentle rotation. The pull-down and input samples were eluted from the beads with 500 μL of elution buffer (100 $NaHCO_3$, 1% SDS) by incubating at room temperature for 15 min with gentle rotation. In order to de-crosslink the chromatin samples, 200 mM NaCl was added to the sample and incubated at 65 ˚C overnight. The samples were treated with proteinase K (20 μg), 10 μL of 0.5 M EDTA pH 8.0 and 20 μL of 1 M Tris HCl pH 7.5 and incubated at 42 ˚C for 60 min. Further DNA from the samples were isolated using PCR-clean up kit using manufacture's protocol (Figure 4.12A).

Quantitative real-time PCR reaction for checking enrichment of telomeric region

In order to investigate the enrichment of telomere pull-down over eGFP, 2 μL of 1% of input or 2 μL pull-down samples corresponding to azide-labeled sgRNA **1′** and **2** were taken in 15 μL of reaction having 7.5 μL TB Green Premix Ex Taq II and 1 μL of TeloF and TeloR primer (0.5 μM each). The qPCR conditions: heat denaturation at 95 °C for 1 min, 40 cycles of (denaturing: 95 °C for 10 sec, annealing: 52 °C for 30 s, extension: 72 °C for 30 s), final extension at 72 °C for 5 min.

The C_t values for input were adjusted to 100% and then pull-down values were normalized with 100% input. Further, the enrichment for telomere region with azide-labeled sgRNA **1′** over azide-labeled sgRNA **2** (targeting eGFP) was calculated. Samples were performed in triplicate for qPCR (Figure 4.12B and Table 2).

$$Fold\ enrichment = 2^{-\left(\frac{Target\left(\frac{Pulldown_{Azide\ labeled\ sgRNA\ \mathbf{1′}}\ ^{C_t}}{Input_{Azide\ labeled\ sgRNA\ \mathbf{1′}}\ ^{C_t}}\right)}{Control\left(\frac{Pulldown_{Azide\ labeled\ sgRNA\ \mathbf{2}}\ ^{C_t}}{Input_{Azide\ labeled\ sgRNA\ \mathbf{2}}\ ^{C_t}}\right)}\right)}$$

4.4.6 Terminal uridylation on a potentially structure RNA for constructing FRET pair on RNA

Optimization of single modification of a structured RNA with base-modified azide analogs

Terminal uridylation of RNA was performed at the 3′ end by incubating 5′ Cy5-labeled RNA (10 µM,) in presence of AMUTP (500 µM, 50 equiv), Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/µL) and in-house expressed, SpCID1 (20.5 pmol) in a final volume of 20 µL. At 5, 15 and 30 min, 5 µL aliquots of the reaction was added to denaturing loading buffer (15 µL) and was heat denatured at 75 °C and 2 µL of heat inactivated reaction mixture (5 pmol) was loaded in 20% denaturing polyacrylamide gel and was electrophoresed (Figure 4.14 and Table 2).

Large scale synthesis of AMU-labeled 5′ Cy5 RNA

5′ Cy5-labeled RNA (10 µM, 200 pmol) was incubated with AMUTP (500 µM, 50 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/µL) and in-house expressed SpCID1 (20.5 pmol) in a final volume of 20 µL for 15 min.

Typically for a 1 nmol reaction, five small scale reactions of 200 pmol of RNA was incubated for 15 min, heat-inactivated at 75 °C for 3 min and was purified by semi-preparative reverse-phase high-performance liquid chromatography (RP-HPLC). RP-HPLC performed in Mobile phase A: 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0-30% B in 30 min and 30-100% B in 10 min. When purified by RP-HPLC a tailing reaction using AMUTP gave a yield of ~85% of AMU-labeled 5′ Cy5 RNA. Further, the mass of

the RNA was confirmed by ESI mass analysis in negative mode by direct infusion of 100 pmol of RNA in 50% ACN (LCMS grade) in 10 mM triethylamine and 100 mM hexafluoro-2-propanol (Table 1).

Strain-promoted azide-alkyne cycloaddition reaction on AMU-labeled 5′ Cy5 RNA

For performing a SPAAC reaction on azide modified ON, AMU-labeled 5′ Cy5 RNA (10 μM, 600 pmol) was incubated with Cy3-DBCO (1 mM), in autoclaved water in final volume of 60 μL. The reaction was incubated at 37 °C for 2 h followed by addition of 60 μL ammonium acetate (5 M) and 600 μL ethanol. The reaction mixture was precipitated at -20 °C overnight followed by centrifugation at 15,000 rpm for 15 min. RNA pellet was washed with 500 μL of chilled 75% ethanol in autoclaved water to remove salts and unconjugated dye followed by centrifugation (15,000 rpm) for 15 min. The pellet obtained was dried and nuclease free autoclaved water was added to dissolve the RNA. The RNA obtained was loaded onto 20% denaturing polyacrylamide gel and electrophoresed. The resolved band corresponding to the fluorescent product, Cy5-Cy3 RNA was excised. The RNA was eluted from gel using 0.5 M sodium acetate and desalting using a Sep-Pak C18 cartridge to obtain a yield of ~80%. Equal amount of HPLC purified AMU-labeled 5′ Cy5 RNA and Cy5-Cy3 RNA was loaded onto denaturing polyacrylamide and was visualized by Typhoon scanner at Cy3 and Cy5 wavelengths. Mass of Cy5-Cy3 RNA was confirmed by ESI-MS by direct infusion (Figure 4.15 and Table 1).

4.4.7 FRET on Cy5-Cy3 RNA and Cy3-FAM RNA

Steady-state fluorescence measurement of Cy5-Cy3 RNA

The 5′ Cy5-labeled RNA, Cy3-DBCO or Cy5-Cy3 RNA (40 nM) was taken in HEPES buffer (10 mM, pH 7.0) in a final volume of 200 μL in a quartz cuvette. All samples were excited at 520 nm and emission was recorded from 530-800 nm keeping both the excitation and emission slit width as 5 nm and 5 nm on a Horiba FluoroMax 4 spectrofluorometer. RNase A (0.1 μg) was added to Cy5-Cy3 RNA (40 nM) in a final volume of 200 μL in HEPES buffer (10 mM, pH 7.0) and fluorescence was recorded at 1, 2 and 5 min respectively. (Figure 4.16).

Steady-state fluorescence measurement of Cy3-FAM RNA

The 5′ FAM-labeled RNA, Cy3-DBCO or Cy3-FAM RNA (40 nM) was taken in HEPES buffer

(10 mM, pH 7.0) in a final volume of 200 μL in a quartz cuvette. All samples were excited at 450 nm and emission was recorded from 460-800 nm keeping both excitation and emission slit width as 5 nm and 5 nm on a Horiba Fluoromax 4 spectrofluorometer. RNase A (0.1 μg) was added to Cy3-FAM RNA (40 nM) in a final volume of 200 μL in HEPES buffer (10 mM, pH 7.0) and fluorescence was recorded (Chapter 3 and Figure 4.17).

4.4.8 Site-specific internal labeling of long RNA using RNA ligation

Monophosphorylation of RNA 2

RNA 2 (6 μM, 300 pmol) was incubated with ATP (1 mM), RiboLock RNase inhibitor (0.4 U/μL), T4 Polynucleotide Kinase reaction buffer (70 mM Tris-HCl pH 7.6, 10 mM MgCl$_2$, 5 mM DTT) and T4 Polynucleotide (10 U/μL) in a final volume of 50 μL at 37 °C for 30 min. The reaction was heat inactivated at 65°C. Multiple reactions were performed and 1200 pmol of 5′ monophosphorylated RNA 2 was purified by RP-HPLC. RP-HPLC performed in Mobile phase A: 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0-30% B in 30 min and 30-100% B in 10 min (Table 2).

Splint ligation of RNA for internal labeling

AMU-labeled 5′ FAM RNA (0.5 μM) was incubated with 5′ monophosphorylated RNA 2 (2.5 μM), splint DNA (2.5 μM), RiboLock RNase inhibitor (1 U/μL) in 1X T4 DNA Ligase buffer (50 mM Tris HCl pH 7.5, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP) at 90 °C for 1 min followed by slow cooling for 15 min for effective annealing of RNA and DNA duplex. T4 RNA Ligase 2 (10 U) or T4 DNA Ligase (1340 U) was added to make a final volume of 20 μL and the sample were incubated at 37 °C for 1 h. The enzyme was heat inactivated by heating at 80 °C for 10 min followed by cooling and addition of DNase I (10 U). The DNase I reaction was incubated at 37 °C for 20 min followed by 5 min heat inactivation at 80 °C. As control, reaction was performed without the addition of either of the ligating enzymes. The products (1 pmol) were loaded on 15% denaturing polyacrylamide gel and was electrophoresed. The gel was visualized in Typhoon gel scanner in FAM wavelength (Figure 4.18 and Table 2).

4.5 References

1. El-Sagheer, A. H.; Brown, T., New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (35), 15329–15334.
2. Sawant, A. A.; Tanpure, A. A.; Mukherjee, P. P.; Athavale, S.; Kelkar, A.; Galande, S.; Srivatsan, S. G., A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **2016,** *44* (2), e16.
3. George, J. T.; Srivatsan, S. G., Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. *Methods* **2017,** *120,* 28–38.
4. Holstein, J. M.; Rentmeister, A., Current covalent modification methods for detecting RNA in fixed and living cells. *Methods* **2016,** *98*, 18–25.
5. Anhäuser, L.; Hüwel, S.; Zobel, T.; Rentmeister, A., Multiple covalent fluorescence labeling of eukaryotic mRNA at the poly(A) tail enhances translation and can be performed in living cells. *Nucleic Acids Res.* **2019,** *47* (7), e42–e42.
6. Pitchiaya, S.; Heinicke, L. A.; Custer, T. C.; Walter, N. G., Single molecule fluorescence approaches shed light on intracellular RNAs. *Chem. Rev.* **2014,** *114* (6), 3224–3265.
7. Fareh, M.; Joo, C., Probing RNA–protein interactions with single-molecule pull-down Assays. In *Nanoscale Imaging: Methods and Protocols*, Lyubchenko, Y. L., Ed. Springer New York: New York, NY, 2018; pp 267–285.
8. Jain, A.; Liu, R.; Ramani, B.; Arauz, E.; Ishitsuka, Y.; Ragunathan, K.; Park, J.; Chen, J.; Xiang, Y. K.; Ha, T., Probing cellular protein complexes using single-molecule pull-down. *Nature* **2011,** *473*, 484.
9. Panda, A. C.; Martindale, J. L.; Gorospe, M., Affinity pulldown of biotinylated RNA for detection of protein-RNA complexes. *Bio Protoc.* **2016,** *6* (24), e2062.
10. George, L.; Indig, F. E.; Abdelmohsen, K.; Gorospe, M., Intracellular RNA-tracking methods. *Open Biol.* **2018,** *8* (10), 180104.
11. Serebrov, V.; Moore, M. J., Single molecule approaches in RNA-Protein interactions. In *RNA Processing: Disease and Genome-wide Probing*, Yeo, G. W., Ed. Springer International Publishing: Cham, 2016; pp 89–106.
12. Vyboh, K.; Ajamian, L.; Mouland, A. J., Detection of viral RNA by fluorescence in situ hybridization (FISH). *J. Vis. Exp.* **2012,** (63), e4002-e4002.
13. Guha, T. K.; Wai, A.; Hausner, G., Programmable genome editing tools and their regulation for efficient genome engineering. *Comput. Struct. Biotec.* **2017,** *15*, 146–160.
14. Sander, J. D.; Joung, J. K., CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **2014,** *32*, 347.
15. Wang, H.; La Russa, M.; Qi, L. S., CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **2016,** *85* (1), 227–264.
16. Thurtle-Schmidt, D. M.; Lo, T.-W., Molecular biology at the cutting edge: A review on CRISPR/CAS9 gene editing for undergraduates. *Biochem. Mol. Biol. Educ.* **2018,** *46* (2), 195–205.
17. Hsu, Patrick D.; Lander, Eric S.; Zhang, F., Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **2014,** *157* (6), 1262–1278.
18. Doudna, J. A.; Charpentier, E., The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014,** *346* (6213).
19. Jiang, F.; Doudna, J. A., CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **2017,** *46* (1), 505–529.

20. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012,** *337* (6096), 816–821.

21. Zalatan, Jesse G.; Lee, Michael E.; Almeida, R.; Gilbert, Luke A.; Whitehead, Evan H.; La Russa, M.; Tsai, Jordan C.; Weissman, Jonathan S.; Dueber, John E.; Qi, Lei S.; Lim, Wendell A., Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **2015,** *160* (1), 339–350.

22. Qin, P.; Parlak, M.; Kuscu, C.; Bandaria, J.; Mir, M.; Szlachta, K.; Singh, R.; Darzacq, X.; Yildiz, A.; Adli, M., Live cell imaging of low- and non-repetitive chromosome loci using CRISPR-Cas9. *Nature Commun.* **2017,** *8*, 14725.

23. Chen, B.; Gilbert, Luke A.; Cimini, Beth A.; Schnitzbauer, J.; Zhang, W.; Li, G.-W.; Park, J.; Blackburn, Elizabeth H.; Weissman, Jonathan S.; Qi, Lei S.; Huang, B., Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **2013,** *155* (7), 1479–1491.

24. Fujita, T.; Yuno, M.; Fujii, H., enChIP systems using different CRISPR orthologues and epitope tags. *BMC Res. Notes* **2018,** *11* (1), 154.

25. Anton, T.; Karg, E.; Bultmann, S., Applications of the CRISPR/Cas system beyond gene editing. *Biol. Methods Protoc.* **2018,** *3* (1).

26. Deng, W.; Shi, X.; Tjian, R.; Lionnet, T.; Singer, R. H., CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (38), 11870–11875.

27. Prescher, J. A.; Bertozzi, C. R., Chemistry in living systems. *Nat. Chem. Biol.* **2005,** *1* (1), 13–21.

28. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **2009,** *48* (38), 6974–6998.

29. Horisawa, K., Specific and quantitative labeling of biomolecules using click chemistry. *Front. Physiol.* **2014,** *5* (457).

30. Speers, A. E.; Cravatt, B. F., Profiling enzyme activities in vivo using click chemistry Methods. *Chem. Biol.* **2004,** *11* (4), 535–546.

31. Taemaitree, L.; Shivalingam, A.; El-Sagheer, A. H.; Brown, T., An artificial triazole backbone linkage provides a split-and-click strategy to bioactive chemically modified CRISPR sgRNA. *Nature Communications* **2019,** *10* (1), 1610.

32. Lee, K.; Mackley, V. A.; Rao, A.; Chong, A. T.; Dewitt, M. A.; Corn, J. E.; Murthy, N., Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife* **2017,** *6*, e25312.

33. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nature Commun.* **2014,** *5*, 4740.

34. Vušurović, N.; Altman, R. B.; Terry, D. S.; Micura, R.; Blanchard, S. C., Pseudoknot formation seeds the twister ribozyme cleavage reaction coordinate. *J. Am. Chem. Soc.* **2017,** *139* (24), 8186–8193.

35. Ditzler, M. A.; Alemán, E. A.; Rueda, D.; Walter, N. G., Focus on function: Single molecule RNA enzymology. *Biopolymers* **2007,** *87* (5-6), 302–316.

36. Manz, C.; Kobitski, A. Y.; Samanta, A.; Keller, B. G.; Jäschke, A.; Nienhaus, G. U., Single-molecule FRET reveals the energy landscape of the full-length SAM-I riboswitch. *Nat. Chem. Biol.* **2017,** *13*, 1172.

37. Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A., A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **2017,** *3* (8), eaao0027.

38. Parks, J. W.; Kappel, K.; Das, R.; Stone, M. D., Single-molecule FRET-Rosetta reveals RNA structural rearrangements during human telomerase catalysis. *RNA* **2017,** *23* (2), 175–188.

39. Rinaldi, A. J.; Suddala, K. C.; Walter, N. G., Native purification and labeling of RNA for single molecule fluorescence studies. *Methods Mol. Biol.* **2015,** *1240*, 63–95.

40. Thomas, C.; Rusanov, T.; Hoang, T.; Augustin, T.; Kent, T.; Gaspar, I.; Pomerantz, R. T., One-step enzymatic modification of RNA 3′ termini using polymerase θ. *Nucleic Acids Res.* **2019,** *47* (7), 3272–3283.

41. Paredes, E.; Das, S. R., Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **2011,** *12* (1), 125–131.

42. Winz, M.-L.; Samanta, A.; Benzinger, D.; Jäschke, A., Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res.* **2012,** *40* (10), e78.

43. Sousa, R.; Chung, Y. J.; Rose, J. P.; Wang, B.-C., Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **1993,** *364* (6438), 593–599.

44. Lunde, B. M.; Magler, I.; Meinhart, A., Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res.* **2012,** *40* (19), 9815–9824.

45. O'Sullivan, R. J.; Karlseder, J., Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **2010,** *11* (3), 171–81.

46. Ui-Tei, K.; Maruyama, S.; Nakano, Y., Enhancement of single guide RNA transcription for efficient CRISPR/Cas-based genomic engineering. *Genome* **2017,** *60* (6), 537-545.

47. Yates, L. A.; Fleurdépine, S.; Rissland, O. S.; De Colibus, L.; Harlos, K.; Norbury, C. J.; Gilbert, R. J. C., Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat. Struct. Mol. Biol.* **2012,** *19*, 782.

48. Rissland, O. S.; Norbury, C. J., The Cid1 poly(U) polymerase. *BBA- Gene Regul. Mech.* **2008,** *1779* (4), 286–294.

49. Saenger, W.; Riecke, J.; Suck, D., A structural model for the polyadenylic acid single helix. *J. Mol. Biol.* **1975,** *93* (4), 529–34.

50. Skinner, G. M.; van den Hout, M.; Broekmans, O.; Dekker, C.; Dekker, N. H., Distinguishing single- and double-stranded nucleic acid molecules using solid-state nanopores. *Nano Lett.* **2009,** *9* (8), 2953–2960.

51. Mu, W.; Zhang, Y.; Xue, X.; Liu, L.; Wei, X.; Wang, H., 5′ capped and 3′ polyA-tailed sgRNAs enhance the efficiency of CRISPR-Cas9 system. *Protein Cell* **2019,** *10* (3), 223–228.

52. Bin Moon, S.; Lee, J. M.; Kang, J. G.; Lee, N.-E.; Ha, D.-I.; Kim, D. Y.; Kim, S. H.; Yoo, K.; Kim, D.; Ko, J.-H.; Kim, Y.-S., Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3′-overhang. *Nature Commun.* **2018,** *9* (1), 3651.

53. Reuter, J. S.; Mathews, D. H., RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **2010,** *11* (1), 129.

54. Molodtsov, V.; Anikin, M.; McAllister, W. T., The presence of an RNA:DNA hybrid that is prone to slippage promotes termination by T7 RNA polymerase. *J. Mol. Biol.* **2014,** *426* (18), 3095–3107.

55. Chelico, L.; Pham, P.; Goodman, M. F., Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. *Philos. Trans. Royal Soc. B: Biol. Sciences* **2009,** *364* (1517), 583–593.

56. Munoz-Tello, P.; Gabus, C.; Thore, S., A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure. *Nucleic Acids Res.* **2014,** *42* (5), 3372-3380.

57. Zhu, X.; Clarke, R.; Puppala, A. K.; Chittori, S.; Merk, A.; Merrill, B. J.; Simonović, M.; Subramaniam, S., Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **2019**.

58. Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A., Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014,** *343* (6176), 1247997.

59. Jerabek-Willemsen, M.; André, T.; Wanner, R.; Roth, H. M.; Duhr, S.; Baaske, P.; Breitsprecher, D., MicroScale thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **2014,** *1077*, 101–113.

60. Seidel, S. A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods (San Diego, Calif.)* **2013,** *59* (3), 301–315.

61. Gao, X. D.; Tu, L.-C.; Mir, A.; Rodriguez, T.; Ding, Y.; Leszyk, J.; Dekker, J.; Shaffer, S. A.; Zhu, L. J.; Wolfe, S. A.; Sontheimer, E. J., C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9–APEX2. *Nature Methods* **2018,** *15* (6), 433–436.

62. van Geel, R.; Pruijn, G. J. M.; van Delft, F. L.; Boelens, W. C., Preventing thiol-yne addition improves the specificity of strain-promoted azide–alkyne Cycloaddition. *Bioconjugate Chem.* **2012,** *23* (3), 392–398.

63. Cawthon, R. M., Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **2002,** *30* (10), e47–e47.

64. Cserép, G. B.; Herner, A.; Kele, P., Bioorthogonal fluorescent labels: a review on combined forces. *Methods App. Fluores.* **2015,** *3* (4), 042001.

65. Qiu, W.; Xu, Z.; Zhang, M.; Zhang, D.; Fan, H.; Li, T.; Wang, Q.; Liu, P.; Zhu, Z.; Du, D.; Tan, M.; Wen, B.; Liu, Y., Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res.* **2019,** *47* (9), e52–e52.

66. Iqbal, A.; Arslan, S.; Okumus, B.; Wilson, T. J.; Giraud, G.; Norman, D. G.; Ha, T.; Lilley, D. M. J., Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (32), 11176-11181.

67. Hengesbach, M.; Kobitski, A.; Voigts-Hoffmann, F.; Frauer, C.; Nienhaus, G. U.; Helm, M., RNA intramolecular dynamics by single-molecule FRET. *Curr. Protoc. Nucleic Acid Chem.* **2008,** *34* (1), 11.12.1–11.12.22.

68. Olejko, L.; Cywiński, P. J.; Bald, I., An ion-controlled four-color fluorescent telomeric switch on DNA origami structures. *Nanoscale* **2016,** *8* (19), 10339–10347.

Chapter 5

A Responsive Fluorescent Nucleotide Enables the Probing of SpCID1-Mediated Terminal Uridylation

5.1 Introduction

Posttranscriptional modification of RNA in organisms plays a pivotal role in modulating, its stability, processing and degradation.[1] One such important class of modification is polyadenylation of pre-mRNA which is performed by template-independent enzymes namely, poly(A) polymerases.[2-3] An alternate category of template independent enzyme, which has been less explored is terminal uridylyl transferases (TUTase) or poly(U) polymerases (PUP). These enzymes belonging to β-like family of eukaryotic polymerase governs RNA stability and degradation by uridylation of 3′ end of coding and ncRNA.[4-7] Identification of 3′ uridylation of novel RNA within the epitranscriptome is still actively in progress and is employed using advanced biochemical and biophysical techniques which has aided in discovery of new obscure pathways of RNA regulation.[8-9] One of these enzymes namely, caffeine-induced death suppressor 1 (SpCID1) was identified in *Schizosaccharomyces pombe* during an overexpression screen designed to understand signaling cascade involved in S-M checkpoint.[10] Similar to SpCID1 in yeast, Ce3-5 in *C. elegans*, At1 in *A. thaliana* and mammalian TUTase like Hs2, Hs3 have been thought to have a conserved mechanism of action.[10-11] U6 TUT1, a human nuclear terminal transferase analog plays a crucial role in the stability of U6 snRNA, important for spliceosomal assembly.[12-13] Also, several terminal uridylating enzymes have been shown to play an important role in microRNA biogenesis.[8, 14-15] Despite a plethora of regulatory functional roles, comparably less biophysical characterizations have been performed on this important class of polymerases.[13,16-18] Especially, few studies in solution state have been performed on these enzymes, which have enabled the understanding of the working mechanism by comparing with crystal structure data.[8] Structures of SpCID1 has been solved in apo form and in the UTP bound form.[16,19] Structure of enzyme solved with dinucleotide, ApU shows its binding to the catalytically active site of the enzyme.[20] Yet, there hasn't been any crystal data of multi-nucleotide long RNA bound to SpCID1 enzyme or RNA-UTP-SpCID1 ternary complex. It has been speculated that SpCID1 binds to RNA using a basic patch of repetitive

sequences containing Arg and Lys which directs the 3′ end of RNA into the active site.[19] In all these crystal structures, it has been observed that UTP is stabilized by a series of amino acids and importantly by a tyrosine residue (Tyr212) which π stacks with the nucleobase (Figure 5.1A). The crystal structure of terminal uridylating enzyme, DmTailor solved recently with bound RNA reveals the π-stacking of the 3′ terminal nucleobase with this conserved tyrosine residue (Figure 5.1B).[21] In crystal of TbTUT4, it has been observed that the 3′ terminal nucleobase of RNA, UTP and tyrosine (Tyr189 in TbTUT4) forms a triple π stacking sandwich.[22] These studies suggest that this conserved tyrosine (Tyr212) is an important motif in terminal uridylyl transferase, SpCID1 for binding of RNA and triphosphate.[16] Constructing a smart probe on RNA or triphosphate for sensing microenvironment changes induced by such a stacking interactions could help understand its real-time binding.[23-24] Terminal uridylyl transferases has been recently studied for its role in tumorigenesis as a potential therapeutic target.[4,14] An assay, which senses real-time binding of the enzyme to its substrate could serve as a platform for high-throughput molecular screening of inhibitors for terminal uridylating enzymes.
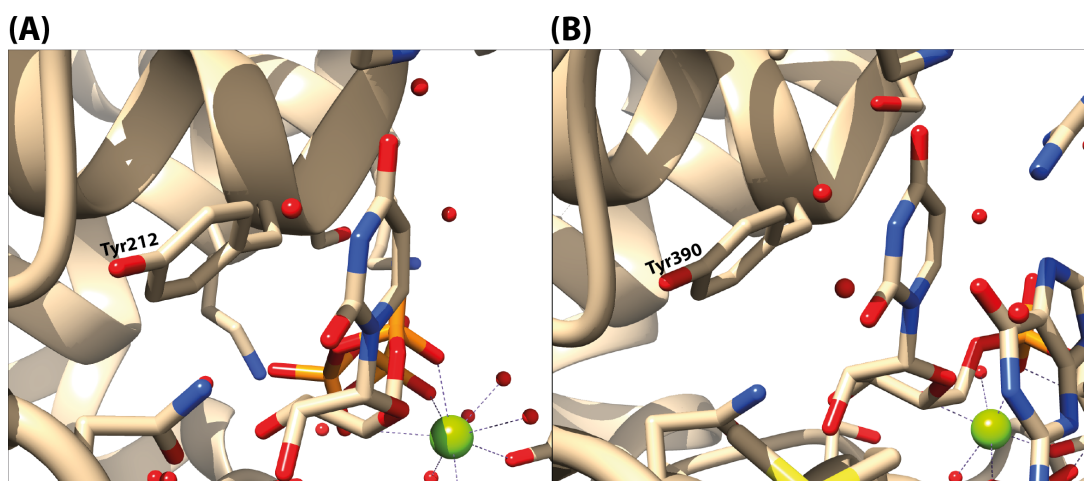
Figure 5.1. (**A**) Crystal structure of SpCID1 bound to UTP shows uracil stacking with Tyr212 (PDB: 4FH5).[16] (**B**) Crystal structure of DmTailor bound to RNA shows last nucleobase, uracil stacking with Tyr390 (PDB: 6I0V).[21] Structure generated using UCSF Chimera software.

Design of various responsive fluorescent nucleotide has been previously carried out in our group taking inspiration from the indole core of aromatic amino acid, tryptophan.[25] These microenvironment sensing nucleotide analogs have been shown to detect variations in local environment upon conformational switching in nucleic acids *in vitro*, in cells and in cell-like

confined environments.[25-27] We envisioned that uridylating RNA with these reporter molecules would enable devise a novel method to sense in real-time, the RNA processing enzymes.

5.2 Results and Discussion

5.2.1 Microenvironment sensing nucleotides for terminal uridylation

Two microenvironment responsive nucleotides, 5-selenophene uridine triphosphate (SeUTP) and 5-benzofuran uridine triphosphate (BFUTP) reported previously from our group were used for terminal uridylation reaction (Figure 5.2).[28-29] SeU has been used earlier for sensing ligand-induced fluorescent changes in ribosomal A-site and also has helped in structure determination of G-quadruplex by taking advantage of the anomalous X-ray scattering of Se.[26] BFU has been previously shown to sense depurination activity of RIPs toxin[30] and conformational dynamics in telomere G-quadruplexes[31]. An artificial nucleic acid, PNA labeled with these reporter molecules have been shown to exhibit quenching in fluorescence and blue-shift upon π stacking with graphene oxide.[32]
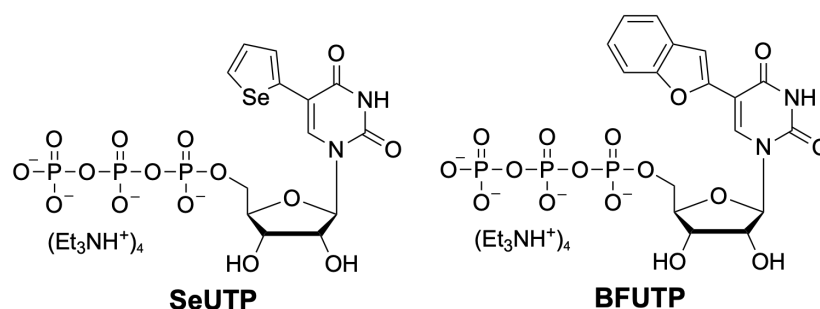
Figure 5.2. Microenvironment sensing UTP analogs used for terminal uridylation.[28-29]

5.2.2 Direct incorporation of microenvironment sensing nucleotides on RNA

To investigate whether direct 3′ enzymatic incorporation of microenvironment sensing probes tagged at the 5-position of uridine is possible, we incubated the terminal uridylyl transferase, SpCID1 (10.25 pmol) with 50 equivalence of SeUTP or BFUTP and 5′ FAM-labeled RNA (10 μM) wherein the 5′ end of RNA, is modified with a fluorescein (Table 5.1). Aliquots of reaction were heat-denatured at 5, 15 and 30 min and was resolved by 20% polyacrylamide gel. SeU got incorporated into 5′ FAM-labeled RNA with good efficiency and tends to saturate at +1 modification in 30 min (Figure 5.3A). Surprisingly, BFU exhibited multiple rounds of distributive incorporation with no visible saturation even after 30 min (Figure 5.3B).[20]

In Chapter 3, we have observed that a minimally perturbing modified triphosphate analog, 5-azidomethyl uridine triphosphate (AMUTP), prominently exhibited distributive enzyme addition which saturates at +2 incorporation, concomitantly showed small amounts of greater than +2 incorporation and trace amounts of products corresponding to processive incorporation (Figure 3.12). Therefore, the fact that a bulky modification like BFUTP incorporates ~4 modified residues in similar conditions is quite unexpected. Additionally, multiple rounds of BFUMP addition at the 3′ end, hints the fact that the RNA product of one BFU addition acts a good substrate for further enzymatic addition as compared to that of SeU. This might be a result of an additional molecular stabilization of BFUTP or the RNA product of BFU addition with the SpCID1 protein, possibly through an aromatic stabilization with catalytic core. It is highly probable that the aromatic BFU probe might be undergoing a π-π stacking interaction with Tyr212 and possibly Phe88 which has been observed to stabilize natural UTP in SpCID1 closed conformation in the catalytic cycle.[33] This hypothesis can be investigated using the fluorescence properties of BFU.

(A)

5' FAM labeled RNA (10 µM)	+	+	+	+
SpCID1 (0.5 µM)	-	+	+	+
SeUTP (500 µM)	-	+	+	+
Time (min)	30	5	15	30

(B)

5' FAM labeled RNA(10 µM)	+	+	+	+
SpCID1 (0.5 µM)	-	+	+	+
BFUTP (500 µM)	-	+	+	+
Time (min)	30	5	15	30
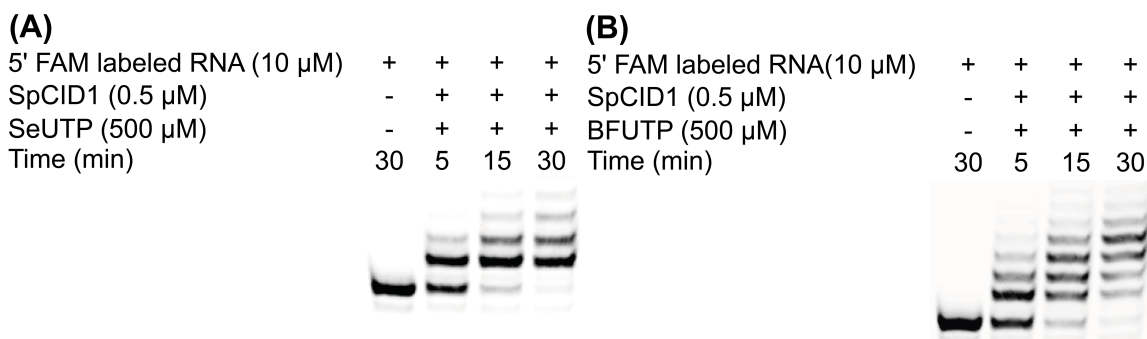
Figure 5.3. Incorporation of (**A**) SeUTP and (**B**) BFUTP into 5′ FAM-labeled RNA using SpCID1 (Table 5.1).

Table 5.1. Sequence of RNA

RNA used	Sequence of oligonucleotide
5′ FAM-labeled RNA	5′ (FAM) UUACCAUAGAAUCAUGUGCCAUACAUCA 3'
Unlabeled RNA	5' UUACCAUAGAAUCAUGUGCCAUACAUCA 3'

5.2.3 Single incorporation of a microenvironment sensing nucleotide at the 3′ end of RNA

In order to investigate the interaction of the microenvironment sensing probe with the catalytic core of SpCID1, a single BFU at the 3′ end of RNA is required. For achieving this, we incubated the 5′ FAM-labeled RNA with limited stoichiometry of the enzyme and the triphosphate (Table

5.1). Since the substrate exhibited highly efficient distributive incorporation, we could not maximize the addition to a saturation of single BFU-labeled product unlike in the case of click labeled triphosphates. This is because the RNA product of one BFU addition is immediately processed as substrate for next addition. For achieving a maximum single modification on RNA using BFUTP, the terminal uridylation reaction was optimized by keeping the concentration of SpCID1 constant and varying concentration of BFUTP in the range of 10-50 µM. Products of reaction was resolved by 20% denaturing polyacrylamide gel. Reaction corresponding to 20 µM of BFUTP at 30 min gave up to ~40% single labeled product band when visualized in Typhoon gel scanner at FAM wavelength (Figure 5.4A). Further, multiple 20 µL reactions were performed on 5′ FAM-labeled RNA or unlabeled RNA (lacking the 5′ FAM-label) with this optimized condition for large-scale isolation of products (Table 5.1). Products of reaction on unlabeled RNA were resolve by 20% denaturing polyacrylamide and visualized using Stains-All reagent (Figure 5.4B). The desired products, BFU-labeled 5′ FAM RNA or BFU-labeled RNA was isolated by reverse-phase HPLC (RP-HPLC) purification and mass of the products were confirmed by ESI-MS (Figure 5.4C and Table 5.2). A typical large-scale terminal uridylation reaction with 1 nmol of 5′ FAM-labeled RNA and 4 nmol reaction of unlabeled RNA gave 0.21 nmol of BFU-labeled 5′ FAM RNA and 1.68 nmol of BFU-labeled RNA, respectively (Table 5.2).
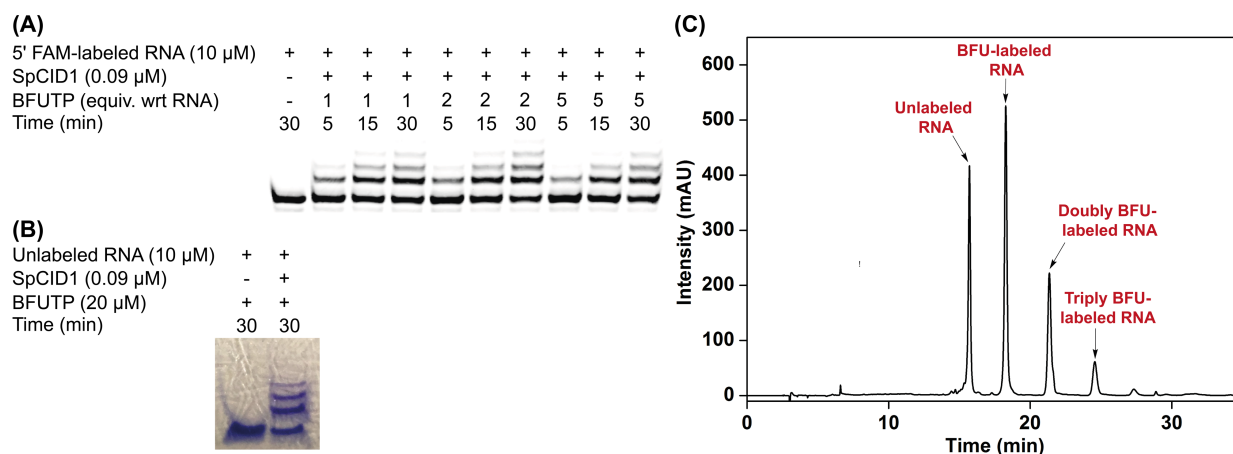
Figure 5.4. (A) Optimization of best condition for single incorporation of BFU at the 3′ end of 5′ FAM-labeled RNA resolved by 20% PAGE. **(B)** Products of large-scale terminal uridylation reaction optimized for yielding maximum single incorporation of BFU at the 3′ end of unlabeled RNA were resolved by 20% PAGE and visualized by Stains-All reagent. **(C)** The reaction on unlabeled RNA performed in large-scale was purified by RP-HPLC (monitored at an absorption of 260 nm).

Table 5.2. Mass and yield of large-scale isolated RNA products

Product RNA	Calcd. Mass	Found Mass	Isolated yield (nmole)*	Isolated yield (%)
BFU-labeled 5′ FAM RNA	9811.1	9810.5	0.21[a]	21
BFU-labeled RNA	9273.7	9273.3	1.68[b]	42

*Isolated yields for reactions performed on [a]1, [b]4 nmol scale.

5.2.4 Binding of SpCID1 enzyme to RNA modified with BFU at the 3′ end

Further, we explored whether BFU-labeled RNA can be used for sensing the binding of RNA to SpCID1 enzyme. For this, BFU-labeled RNA was incubated with varying concentration of SpCID1 for 30 min at room temperature and fluorescence was recorded by exciting the samples at 330 nm (corresponding to the absorption of the BFU probe, Figure 5.5A). Upon titrating BFU-labeled RNA with increasing SpCID1, a concentration-dependent quenching of fluorescence intensity accompanied by a small blue shift in emission maximum was observed (Figure 5.5B). It has been shown earlier that BFU exhibits a quenching and blue shift in fluorescence upon decrease in microenvironment polarity.[28] We have reported that PNA labeled with BFU showed observable quenching in fluorescence and a blue shift in emission maxima upon undergoing π-π stacking interaction with graphene oxide.[32] Therefore, these observations indicate that the quenching in fluorescence could possibly be due to BFU stacking with Ty212 and minor changes in micropolarity. Upon fitting the normalized fluorescence change to a Hill equation, we observed an apparent K_d of 524 ± 38 nM (Figure 5.5C).

(A)

(B)

(C)
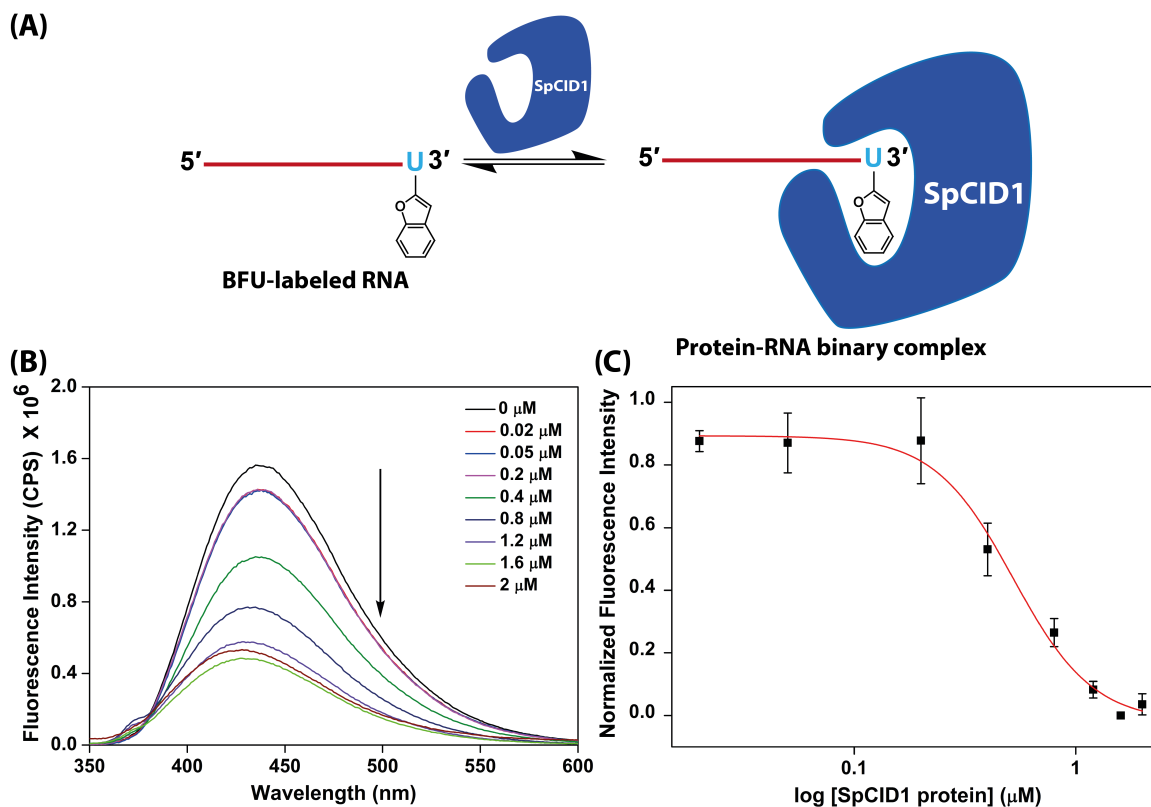
Figure 5.5. (A) Titration of BFU-labeled RNA with SpCID1. **(B)** BFU-labeled RNA (0.2 μM) and SpCID1 enzyme (0-2 μM) was excited at 330 nm and fluorescence emission was recorded keeping an excitation and emission slit width of 5 and 6 nm. **(C)** Curve fit for the above titration at emission maximum of BFU-labeled RNA (λ_{em} = 435 nm) using Hill equation.

To validate that SpCID1 enzyme in indeed binding to BFU-modified RNA, we titrated BFU-labeled 5′ FAM RNA with SpCID1 and monitored the binary complex formation using electrophoretic mobility shift assay (EMSA) (Figure 5.6A). Upon visualizing the RNA in gel, a concentration-dependent depletion of unbound RNA band accompanied by the formation of binary complex band of reduced migration was observed upon increasing SpCID1 enzyme (Figure 5.6B). Almost complete binding was observed at ~60 equiv of SpCID1 over RNA. Fitting the data to Hill equation resulted in an apparent K_d of ~5.4 μM (Figure 5.6C). Previously using this nonequilibrium technique,[34] K_d of RNA binding to SpCID1 has been measured as ~1 and ~21 μM for 15 nucleotide homopolymer (U)$_{15}$ and (A)$_{15}$ RNA.[20] Therefore, fluorescence provides a direct real-time read-out of binding of RNA and SpCID1 in solution-state at equilibrium.

(A)

(B)

(C)
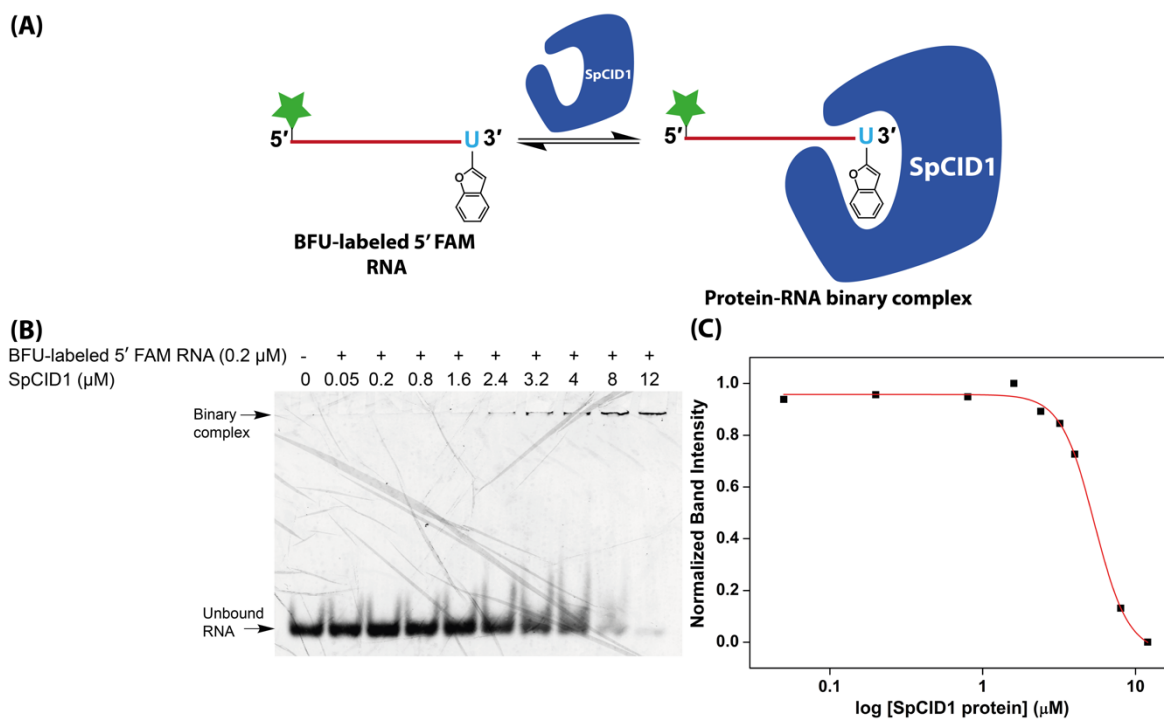
Figure 5.6. (A) Titration of BFU-labeled 5′ FAM RNA with SpCID1. **(B)** BFU-labeled 5′ FAM RNA and SpCID1 enzyme (0-12 μM) were incubated and resolved by 15% native polyacrylamide gel. The protein RNA binary complex remained at the gel well. **(C)** Curve fit for the titration measured using normalized band intensity employing Hill equation. Normalized band intensity calculated using ratio of unbound BFU-labeled 5′ FAM RNA in the presence of SpCID1 to unbound BFU-labeled 5′ FAM RNA sample without SpCID1 detected at FAM wavelength.

Competitive binding assay with unlabeled RNA

The enzyme requires the 3′ nucleotide of RNA to enter its active site for terminal uridylation reaction. EMSA proves the binding of RNA to SpCID1 enzyme. Next objective is to confirm that the quenching of fluorescence results from specific binding of BFU in RNA to the active site of the enzyme and not through a non-specific binding between BFU and enzyme. For confirming this, we performed a competitive assay wherein SpCID1 (1.6 μM) and BFU-labeled RNA (0.2 μM) were titrated with increasing concentration of unlabeled RNA (0-10 equiv with respect to BFU-labeled RNA) and variation of fluorescence was recorded (Figure 5.7A). A concentration-dependent recovery of fluorescence from the quenched state was observed upon increasing the concentration of unlabeled RNA. The results indicate that the unlabeled RNA competes with BFU-labeled RNA for binding onto the catalytic active site (Figure 5.7B). The recovery in fluorescence is due to the displacement of the modified RNA from the active site by the unlabeled RNA wherein the 3′ nucleobase π stacks with this conserved Tyr212.[20-21] It is to be noted that a 10-fold concentration of unlabeled RNA compared to BFU-labeled RNA did not completely recover the

fluorescence. This indicates that BFU-labeled RNA binds more effectively as compared to unlabeled RNA to the active site.
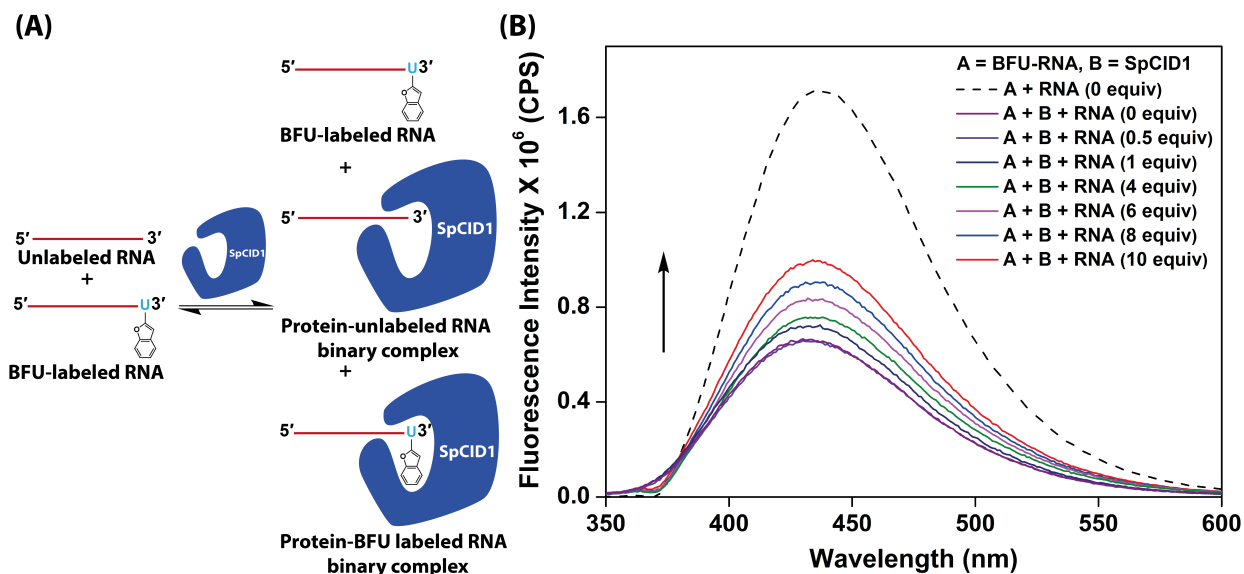
Figure 5.7. (**A**) Titration of BFU-labeled RNA and SpCID1 with unlabeled RNA. (**B**) BFU-labeled RNA (BFU-RNA, 0.2 μM) and SpCID1 enzyme (1.6 μM) were incubated with unlabeled RNA (1-10 equiv with respect to BFU-labeled RNA). Samples were excited at 330 nm and fluorescence was recorded keeping an excitation and emission slit width of 5 and 6 nm.

5.2.5 Binding of SpCID1 enzyme to modified triphosphate, BFUTP

Further we investigated whether BFU probe could sense the binding of nucleotide to SpCID1 protein. For this, BFUTP (0.2 μM) was incubated with increasing concentration of SpCID1 enzyme (0-2 μM) for 30 min at room temperature (Figure 5.8A). The samples were excited at 330 nm corresponding to the absorption of BFUTP and fluorescence was recorded. A concentration-dependent quenching of fluorescence and a small blue shift in emission maxima was observed upon increasing SpCID1 concentration (Figure 5.8B). Upon fitting the data to Hill equation, we observed an apparent K_d of 181 ± 35 nM, which is lesser than the RNA (Figure 5.8C). These results further indicate the possibility of a stacking interaction between Tyr212 and BFUTP.

(A)

(B)

(C)
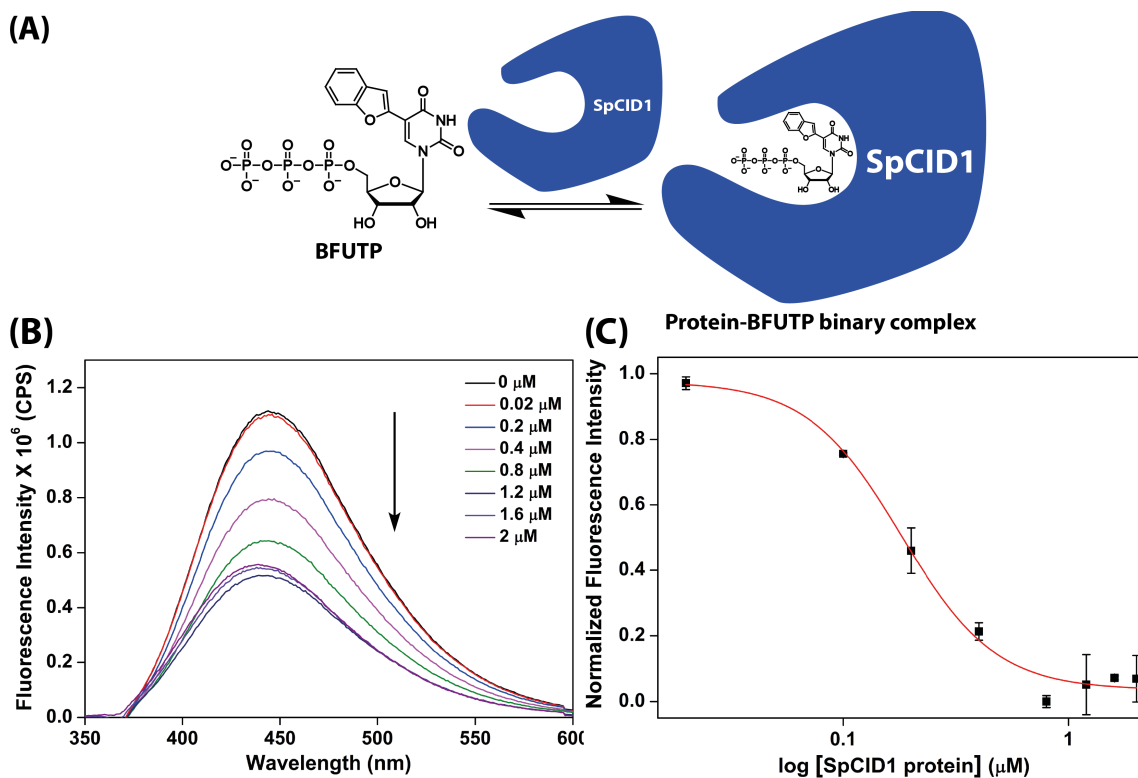
Figure 5.8. (A) Titration of BFUTP with SpCID1. **(B)** BFUTP (0.2 μM) and SpCID1 enzyme (0-2 μM) was excited at 330 nm and fluorescence emission was recorded keeping an excitation and emission slit width of 5 and 6 nm. **(C)** Curve fit for the above titration at at emission maximum of BFUTP (λ_{em} = 436 nm) with Hill equation.

Competitive binding assay with natural UTP

Similar to the experiment performed for BFU-labeled RNA and SpCID1 enzyme, a competitive assay with natural UTP and BFUTP were used to confirm whether the quenching is as a result of specific binding between BFUTP and SpCID1 (Figure 5.9A). Multiple samples having BFUTP (0.2 μM) and SpCID1 (0.8 μM) were incubated with increasing concentration of natural UTP (0-200 equiv with respect to BFUTP) for 30 min at room temperature. The samples were excited at 330 nm and fluorescence was recorded. Titration of UTP resulted in a recovery of fluorescence intensity from quenched state (Figure 5.9B). Since it has been proven previously that UTP π stacks with this conserved Tyr212,[16] the observed recovery in fluorescence pinpoints direct competition of UTP over BFUTP for stacking to Tyr212. We thus conclude that the interaction between BFUTP and SpCID1 is a specific interaction. It is to be noted that a 200-fold concentration of UTP over BFUTP was not sufficient enough to completely recover the fluorescence implying that BFUTP binds better to SpCID1 as compared to UTP.
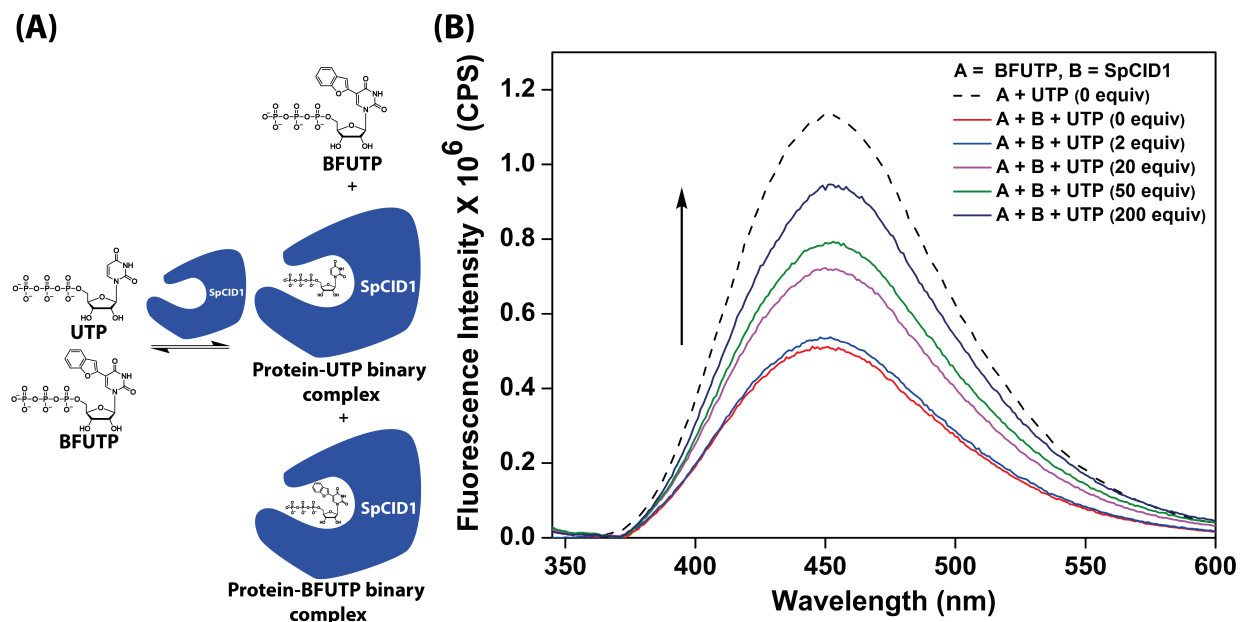
Figure 5.9. (**A**) Titration of BFUTP and SpCID1 with natural UTP. (**B**) BFU-labeled RNA (BFU-RNA, 0.2 μM) and SpCID1 enzyme (0.8 μM) was incubated with UTP (1-200 equiv with respect to BFU-labeled RNA) and samples were excited at 330 nm and fluorescence was recorded keeping an excitation and emission slit width of 5 and 6 nm.

5.3 Conclusions

Terminal uridylation of RNA has been shown to be responsible for regulating RNA population by signaling it for stability or degradation. Herein we introduced highly useful microenvironment sensing probes namely, SeU and BFU, directly at the 3′ end of RNA employing terminal uridylating enzyme, SpCID1. These probes are useful for the biophysical solution-state characterization of SpCID1 which has not been well explored. Terminal uridylation using BFUTP results in higher efficiency of distributive incorporation of BFU residues added at the 3′ end of RNA as compared to SeUTP or minimally perturbing triphosphate, AMUTP used in Chapter 3. We hypothesized that this multiple distributive addition could possibly be resulting from aromatic stabilization of BFUTP or the BFU at the 3′ end of BFU-labeled RNA with amino acid residues in the enzyme. This hypothesis was tested by constructing an RNA having single BFU residue at the 3′ end. Incubation of BFU-labeled RNA or BFUTP with SpCID1 resulted in concentration-dependent quenching of BFU probe upon binding to the catalytic active site. Using EMSA and competitive assay we have validated that the quenching is a result of a direct interaction of BFU label on RNA or BFUTP with the tyrosine amino acid. Both these observation falls in line with the reported π-π stabilization of bound UTP[16]/dinucleotide RNA[20] with Tyr212 in SpCID1 and

terminal nucleobase of RNA with the Tyr390 of DmTailor[21]. Further we calculated the apparent dissociation constant for BFU-labeled RNA and BFUTP in real-time in solution-state. The dissociation constant for BFUTP was found to be lesser than BFU-labeled RNA which corroborates with the existing model wherein it has been proposed that UTP binding is the first step of SpCID1 catalytic cycle prior to binding of RNA to form the ternary complex.[33] Altogether we devise a novel strategy using terminal uridylyl transferase, SpCID1 for introducing microenvironment sensing probes into RNA for biophysical characterization of itself. Certainly, such a technique can be further expanded as general strategy for sensing the binding of polymerases having an aromatic stacking interaction with incoming ON or triphosphate. Also, terminal RNA uridylation has been shown to induce tumorigenesis and hence they are potential therapeutic targets.[4,14] Therefore, this strategy of monitoring the real-time substrate binding helps in the development of assays for therapeutic screening of potential small-molecule inhibitors for uridylating enzymes.

5.4 Experimental Section

5.4.1 Materials

5′ FAM-labeled RNA and unlabeled RNA (Table 5.1) was purchased from Dharmacon Inc and deprotected according to the manufacturer's protocol. RiboLock RNase inhibitor, natural UTP was obtained from Thermo Fischer Scientific. BFUTP and SeUTP was kindly provided by Arun Tanpure and Ashok Nuthanakanti. Stains-All and reagents for buffer used in reaction (Bio Ultra Grade) were acquired from Sigma Aldrich. All terminal uridylation reactions in this chapter were performed with the in-house expressed SpCID1 enzyme synthesized in Chapter 3.

5.4.2 Instrumentation

RNA was quantified by measuring absorbance in UV-2600 Shimadzu or NanoDrop™ 2000c spectrophotometer. ESI-MS mass analysis of RNA was performed using Waters SYNAPT G2-Si Mass Spectrometry instrument in negative mode. HPLC analysis was performed on an Agilent Technologies 1260 Infinity HPLC. Fluorescent RNA was resolved by polyacrylamide on OWL S4S sequencing gel electrophoresis instrument and was imaged using Typhoon FLA 9500 Variable mode Imager. Fluorescence spectra were recorded in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on Horiba Fluoromax 4 spectrofluorometer.

5.4.3 3′ RNA labeling using SpCID1 and microenvironment responsive nucleotides

To label RNA at the 3′ end, 5′ FAM-labeled RNA (10 μM) was incubated with SeUTP or BFUTP (500 μM, 50 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and 1 μL of in-house expressed SpCID1 enzyme (10.25 pmol) in a final volume of 20 μL. After 5, 15 min and 30 min, 5 μL aliquots of reaction mixture (50 pmol) were mixed with 15 μL of denaturing loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8) and heat-denatured at 75 °C for 3 min. Further, 5 μL of the inactivated reaction mixture (12.5 pmol) was loaded on to a 20% denaturing polyacrylamide gel and imaged using Typhoon gel scanner at FAM wavelength. The control reactions were performed with same concentration of 5′ FAM-labeled RNA, Tris-HCl, NaCl, $MgCl_2$, and DTT without addition of SpCID1 enzyme and modified UTP (Figure 5.3 and Table 5.1).

5.4.4 Optimization of single modification on 5′ FAM-labeled RNA with BFUTP

Terminal uridylation of RNA with a single BFU at the 3′ end was optimized by incubating 5′ FAM-labeled RNA (10 μM) for 5, 15 and 30 min at varying concentrations of BFUTP (10-50 μM, 1-5 equiv) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and in-house expressed SpCID1 (1.71 pmol) in a final volume of 20 μL. 5 μL aliquots of reactions were heat-denatured at 5, 15 and 30 min and product RNA (12.5 pmol) was resolved by 20% denaturing polyacrylamide. The control reactions were performed with same concentration of RNA, Tris-HCl, NaCl, $MgCl_2$, and DTT without addition of SpCID1 enzyme and BFUTP (Figure 5.4A and Table 5.1).

Large-scale reaction for single BFU modification on 5′ FAM-labeled RNA or unlabeled RNA

Single incorporation of RNA at the 3′ end was performed by incubating 5′ FAM-labeled RNA or unlabeled RNA (10 μM) with BFUTP (20 μM) in the presence of Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (2 mM), RiboLock RNase inhibitor (1 U/μL) and in-house expressed, SpCID1 (1.71 pmol) in a final volume of 20 μL for 30 min. Reaction with unlabeled RNA (100 pmol) was analyzed by 20% denaturing polyacrylamide gel and stained with Stains-All gel staining reagent. The control reactions were performed with same concentration of unlabeled RNA, Tris-HCl, NaCl, $MgCl_2$, and DTT without addition of SpCID1 enzyme and

BFUTP. Further, several small-scale reactions are performed for large-scale isolation of product. The reaction products were purified by RP-HPLC. RP-HPLC performed in mobile phase A: 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: ACN. Flow rate: 1 mL/min. Gradient: 0-30% B in 30 min and 30-100% B in 10 min. Further, the mass of product RNA was confirmed by ESI mass analysis in negative mode by direct infusion of 100 pmol of RNA in 50% ACN (LCMS grade) in 10 mM triethylamine and 100 mM hexafluoro-2-propanol. HPLC purification gave 21% of BFU-labeled 5′ FAM RNA and 42% of BFU-labeled RNA (Figure 5.4B-C and Table 5.1, 5.2).

5.4.5 Steady-state fluorescence binding assay for BFU-labeled RNA/BFUTP, and SpCID1 enzyme.

A series of samples having either BFU-labeled RNA or BFUTP (0.2 μM) were incubated with increasing concentrations of SpCID1 enzyme (0-2 μM) in Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), $MgCl_2$ (10 mM), DTT (1 mM) and RiboLock RNase inhibitor (0.2 U/μL) in a final volume of 200 μL maintaining an overall ~0.9 % glycerol. The samples were incubated for 30 min at room temperature followed by steady-state fluorescence measurements. Steady-state fluorescence of individual samples were performed by exciting samples at 330 nm wavelength and fluorescence was recorded from 340 to 600 nm maintaining an excitation emission slit width of 5 and 6 nm respectively. Control samples without BFU-labeled RNA or BFUTP showed no fluorescence. The fluorescence quenching assay was performed in duplicate and standard deviation was calculated. Dose dependent quenching of fluorescence observed at 435 nm for BFU-labeled RNA and 436 nm for BFUTP was fit to Hill equation (Origin 8.5) by plotting normalized fluorescence intensity (F_N) versus log of protein concentration. A Hill coefficient (n) of 2.4 and 2.1 was observed for binding of BFU-labeled RNA and BFUTP to SpCID1 (Figure 5.5 and 5.8).

$$F_N = \frac{F_i - F_s}{F_0 - F_s}$$

F_i is the fluorescence emission intensity measured at each protein concentration. F_0 and F_S are the fluorescence emission intensity measured in the absence of protein SpCID1 and at saturation point where *n* is the Hill coefficient measuring the degree of cooperativity in binding.

$$F_N = F_0 + (Fs - F_0)\left(\frac{[\text{SpCID1}]^n}{[K_a]^n + [\text{SpCID1}]^n}\right)$$

5.4.6 Competition assay with unlabeled RNA and UTP

Steady-state fluorescence competition assay for BFU-labeled RNA and SpCID1 enzyme, titrated with unlabeled RNA.

Multiple samples having BFU-labeled RNA (0.2 μM) and SpCID1 enzyme (1.6 μM) were incubated with increasing concentration of unlabeled RNA (0-2 μM) in Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (1 mM) and Riboblock RNase Inhibitor (0.2 U/μL) in a final volume of 200 μL maintaining an overall ~0.8% glycerol. The samples were incubated for 30 min at room temperature followed by steady-state fluorescence measurements. Steady-state fluorescence of individual samples were performed by exciting samples at 330 nm wavelength and fluorescence was recorded from 340 to 600 nm maintaining an excitation emission slit width of 5 and 6 nm respectively. Control sample without BFU-labeled RNA showed no fluorescence (Figure 5.7).

Steady-state fluorescence competition assay for BFUTP and SpCID1 enzyme, titrated with natural UTP.

A series of samples having BFU-labeled RNA (0.2 μM) and SpCID1 enzyme (0.8 μM) was incubated with increasing concentration of UTP (0-40 μM) in Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (1 mM) and Riboblock RNase Inhibitor (0.2 U/μL) in a final volume of 200 μL maintaining an overall ~0.4 % glycerol. The samples were incubated for 30 min at room temperature followed by steady-state fluorescence measurements. Steady-state fluorescence of individual samples were performed by exciting samples at 330 nm wavelength and fluorescence was recorded from 340 to 600 nm maintaining an excitation emission slit width of 5 and 6 nm respectively. Control sample without BFUTP showed no fluorescence (Figure 5.9).

5.4.7 Electrophoretic mobility shift assay for BFU-labeled 5′ FAM RNA and SpCID1 enzyme.

A series of samples having BFU-labeled 5′ FAM RNA (0.2 μM) was titrated with SpCID1 enzyme (0-12 μM) in Tris-HCl buffer (10 mM, pH 7.9 at 25 °C), NaCl (50 mM), MgCl$_2$ (10 mM), DTT (1 mM) and Riboblock RNase Inhibitor (0.2 U/μL) in a final volume of 20 μL maintaining overall 5.6% glycerol. The samples (2 pmol of BFU-labeled RNA) were incubated for 30 min at room temperature followed by the addition of 20 μL of native loading buffer. The samples were loaded

into native 12% polyacrylamide gel maintaining a 10 mM $MgCl_2$ concentration. The gel was run in 1X TBE, PAGE running buffer supplemented with 10 mM $MgCl_2$. The gel was resolved at 4 °C and was imaged in Typhoon gel scanner at FAM wavelength. Since the bound fraction was stuck in the well, binding curve was generated by calculating relative band intensity corresponding to BFU-labeled 5′ FAM RNA on addition of SpCID1 compared to sample without SpCID1 in Fiji Image analysis software and was fit to Hill equation (Origin 8.5) using equation mentioned in section 5.4.5.

F_i is the fluorescence band intensity measured at each protein concentration. F_0 and F_S are the fluorescence band intensity measured in the absence of protein SpCID1 and at saturation point where n is the Hill coefficient measuring the degree of cooperativity in binding. A Hill coefficient (n) of 3.9 was observed for binding of BFU-labeled 5′ FAM RNA to SpCID1 (Figure 5.6).

5.5 References

1. Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C., Dynamic RNA modifications in gene expression regulation. *Cell* **2017,** *169* (7), 1187–1200.
2. Dreyfus, M.; Régnier, P., The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* **2002,** *111* (5), 611–613.
3. Laishram, R. S., Poly(A) polymerase (PAP) diversity in gene expression – star-PAP vs canonical PAP. *FEBS Lett.* **2014,** *588* (14), 2185–97.
4. Munoz-Tello, P.; Rajappa, L.; Coquille, S.; Thore, S., Polyuridylation in eukaryotes: A 3′-end modification regulating RNA life. *BioMed Res. Int.* **2015,** *2015*, 12.
5. Rissland, O. S.; Mikulasova, A.; Norbury, C. J., Efficient RNA polyuridylation by poncanonical poly(A) polymerases. *Mol. Cell. Biol.* **2007,** *27* (10), 3612–3624.
6. Zigáčková, D.; Vaňáčová, Š., The role of 3′ end uridylation in RNA metabolism and cellular physiology. *Philos. Trans. Royal Soc. B: Biol. Sciences* **2018,** *373* (1762), 20180171.
7. Menezes, M. R.; Balzeau, J.; Hagan, J. P., 3′ RNA uridylation in epitranscriptomics, gene regulation, and disease. *Front. Mol. Biosci.* **2018,** *5* (61).
8. Kim, B.; Ha, M.; Loeff, L.; Chang, H.; Simanshu, D. K.; Li, S.; Fareh, M.; Patel, D. J.; Joo, C.; Kim, V. N., TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *EMBO J.* **2015,** *34* (13), 1801–1815.
9. Chung, C. Z.; Jaramillo, J. E.; Ellis, M. J.; Bour, D. Y. N.; Seidl, L. E.; Jo, D. H. S.; Turk, M. A.; Mann, M. R.; Bi, Y.; Haniford, D. B.; Duennwald, M. L.; Heinemann, I. U., RNA surveillance by uridylation-dependent RNA decay in Schizosaccharomyces pombe. *Nucleic Acids Res.* **2019,** *47* (6), 3045–3057.
10. Rissland, O. S.; Norbury, C. J., The Cid1 poly(U) polymerase. *BBA-Gene Regul. Mech.* **2008,** *1779* (4), 286–294.
11. Kwak, J. E.; Wickens, M., A family of poly(U) polymerases. *RNA* **2007,** *13* (6), 860–867.
12. Trippe, R.; Guschina, E.; Hossbach, M.; Urlaub, H.; Lührmann, R.; Benecke, B.-J., Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* **2006,** *12* (8), 1494–1504.

13. Yamashita, S.; Takagi, Y.; Nagaike, T.; Tomita, K., Crystal structures of U6 snRNA-specific terminal uridylyltransferase. *Nature Commun.* **2017,** *8*, 15788.
14. Lehrbach, N. J.; Armisen, J.; Lightfoot, H. L.; Murfitt, K. J.; Bugaut, A.; Balasubramanian, S.; Miska, E. A., LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. *Nat. Struct. Mol. Biol.* **2009,** *16*, 1016.
15. Heo, I.; Ha, M.; Lim, J.; Yoon, M.-J.; Park, J.-E.; Kwon, S. C.; Chang, H.; Kim, V. N., Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **2012,** *151* (3), 521–532.
16. Lunde, B. M.; Magler, I.; Meinhart, A., Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res.* **2012,** *40* (19), 9815–9824.
17. Yashiro, Y.; Tomita, K., Function and eegulation of human terminal uridylyltransferases. *Front. Genet.* **2018,** *9* (538).
18. Yamashita, S.; Nagaike, T.; Tomita, K., Crystal structure of the Lin28-interacting module of human terminal uridylyltransferase that regulates let-7 expression. *Nature Commun.* **2019,** *10* (1), 1960.
19. Yates, L. A.; Fleurdépine, S.; Rissland, O. S.; De Colibus, L.; Harlos, K.; Norbury, C. J.; Gilbert, R. J. C., Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat. Struct. Mol. Biol.* **2012,** *19*, 782.
20. Munoz-Tello, P.; Gabus, C.; Thore, S., A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure. *Nucleic Acids Res.* **2014,** *42* (5), 3372–3380.
21. Kroupova, A.; Ivaşcu, A.; Jinek, M.; Reimão-Pinto, M. M.; Ameres, S. L., Structural basis for acceptor RNA substrate selectivity of the 3′ terminal uridylyl transferase Tailor. *Nucleic Acids Res.* **2018,** *47* (2), 1030–1042.
22. Stagno, J.; Aphasizheva, I.; Aphasizhev, R.; Luecke, H., Dual role of the RNA substrate in selectivity and catalysis by terminal uridylyl transferases. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (37), 14634.
23. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G., Fluorescent nucleoside analogs: probes for investigating nucleic acid structure and function. *Isr. J. Chem.* **2013,** *53* (6–7), 366–378.
24. Sinkeldam, R. W.; Greco, N. J.; Tor, Y., Fluorescent Analogs of Biomolecular Building Blocks: Design, Properties, and Applications. *Chem. Rev.* **2010,** *110* (5), 2579–2619.
25. Manna, S.; Srivatsan, S. G., Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv.* **2018,** *8* (45), 25673–25694.
26. Nuthanakanti, A.; Ahmed, I.; Khatik, S. Y.; Kayarat, S.; Srivatsan, S. G., Probing G-quadruplex topologies and recognition concurrently in real time and 3D using a dual-app nucleoside probe. *Nucleic Acids Res.* **2019**.
27. Manna, S.; Sarkar, D.; Srivatsan, S. G., A dual-app nucleoside probe provides structural insights into the human telomeric overhang in live cells. *J. Am. Chem. Soc.* **2018,** *140* (39), 12622–12633.
28. Tanpure, A. A.; Srivatsan, S. G., A microenvironment-sensitive fluorescent pyrimidine ribonucleoside analogue: synthesis, enzymatic incorporation, and fluorescence detection of a DNA Abasic Site. *Chem.: Eur. J.* **2011,** *17* (45), 12820–12827.
29. Pawar, M. G.; Srivatsan, S. G., Synthesis, Photophysical characterization, and enzymatic incorporation of a microenvironment-sensitive fluorescent uridine analog. *Org. Lett.* **2011,** *13* (5), 1114–1117.

30. Tanpure, A. A.; Srivatsan, S. G., Synthesis and photophysical characterisation of a fluorescent nucleoside analogue that signals the presence of an abasic site in RNA. *ChemBioChem* **2012,** *13* (16), 2392–2399.
31. Tanpure, A. A.; Srivatsan, S. G., Conformation-sensitive nucleoside analogues as topology-specific fluorescence turn-on probes for DNA and RNA G-quadruplexes. *Nucleic Acids Res.* **2015,** *43* (22), e149.
32. Sabale, P. M.; George, J. T.; Srivatsan, S. G., A base-modified PNA-graphene oxide platform as a turn-on fluorescence sensor for the detection of human telomeric repeats. *Nanoscale* **2014,** *6* (18), 10460–9.
33. Durrant, B. P.; Harlos, K.; Yates, L. A.; Gilbert, R. J. C.; Fleurdépine, S.; Norbury, C. J., Structural plasticity of Cid1 provides a basis for its distributive RNA terminal uridylyl transferase activity. *Nucleic Acids Res.* **2015,** *43* (5), 2968–2979.
34. Hellman, L. M.; Fried, M. G., Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2007,** *2* (8), 1849–61.

Final Conclusions and Outlook

In this thesis, we have developed novel chemo-enzymatic technologies to functionalize RNA with biophysical probes and functional tags using the promiscuity of template-dependent and independent polymerases. We demonstrated the incorporation of a vinyl tag into RNA oligonucleotides using *in vitro* transcription, which facilitated its modular posttranscriptional functionalization using palladium-mediated oxidative Heck reaction and reagentless inverse-electron demand Diels-Alder reaction. This method enabled the construction of fluorogenic RNA oligonucleotides, which could be potentially utilized in background-free detection of RNA in cells.

A major limitation of transcriptional reaction is non-selective incorporation of modified nucleotide all throughout the RNA. To tackle this, we developed a site-specific chemo-enzymatic approach to label RNA at its 3' end using SpCID1 and further demonstrated its tunability for adding single or multiple biorthogonal tags. We applied this novel strategy for tailoring multiple azide tags on CRISPR sgRNA for site-specific labeling of gene locus with azide tags. Further, bioorthogonal functionalization of the gene was achieved using post-hybridization click reaction, which in principle can be expanded for tagging any desired gene. The ability to control SpCID1 activity to incorporate a single modified nucleotide at 3' end was utilized in assembling FRET pair-labeled RNAs. This end-labeling approach followed by a ligation reaction was also successfully implemented in introducing site-specific internal clickable labels onto a longer RNA. Taking further, incorporation of a microenvironment sensing nucleotide analog made possible the structural deconvolution of terminal uridylation process by SpCID1.

The possibilities of SpCID1-mediated chemo-enzymatic site-specific 3' labeling strategy is tremendous. The technology we have devised on CRISPR can evolve as a universal approach for targeted labeling of gene loci wherein a minimally perturbing bioorthogonal handle can be exploited as a 'trojan horse' on CRISPR-dCas9 system for site-directed delivery of functional warheads and tags while preserving its biological activity. Apart from the applications of terminal uridylation we have discussed in the thesis, we envision that site-specific incorporation of multiple clickable labels can be utilized in (i) amplifying signal in single molecule RNA-FISH experiments, (ii) developing sensitive RNA diagnostic tools and (iii) for detection of low-expressing RNA species in cells. Also, the direct chemo-enzymatic incorporation of microenvironment sensing nucleosides and construction of FRET tools would enable investigation of RNA structure and

RNA-ligand/protein interactions. Altogether, the versatility of SpCID1 to append single or multiple modified nucleotide analogs can be applied as a two-in-one labeling technology for understanding RNA structure, function and its cellular dynamics without perturbing its biological activity.

Copyrights and Permissions