The Synthesis of 2'-Modified RNAs for Probing Enzymes Implicated in Splicing and Gene Editing

Leonora Abdullahu

Department of Chemistry

McGill University

Montreal, QC, Canada

August 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Dedicated to my parents, Selvete and Vehbi Abdullahu.

For making the greatest sacrifice when leaving our home in Kosovo for an unknown future in Canada. You have gifted me with a world of opportunity and privilege. I owe every success to you. I am eternally grateful for your love, support, and encouragement.

Copyright Statements

Some of the material in this thesis has been adapted from published and/or submitted papers and is under copyright:

Sections 2.1, 2.3, 2.5, and 2.6 of **Chapter 2** have been reproduced from: "Munir, A.; Abdullahu, L.; Damha, M. J.; Shuman, S., Two-step mechanism and step-arrest mutants of Runella slithyformis NAD^+-dependent tRNA 2′-phosphotransferase Tpt1. *RNA* **2018,** *24* (9), 1144-1157. © CSHL Press for RNA Society, 2018" and "Banerjee, A.; Munir, A.; Abdullahu, L.; Damha, M. J.; Goldgur, Y.; Shuman, S., Structure of tRNA splicing enzyme Tpt1 illuminates the mechanism of RNA 2'-PO_4 recognition and ADP-ribosylation. *Nat Commun* **2019,** *10* (1), 218. © Nature Research, 2019"

Sections 3.1, 3.2, 3.3, 3.4.2, 3.5, 3.6, 3.7, 3.8.1 – 3.8.4 of **Chapter 3** have been adapted from: "Dantuluri, S.; Abdullahu, L.; Munir, A.; Katolik, A.; Damha, M. J.; Shuman, S., Substrate analogs that trap the 2'-phospho-ADP-ribosylated RNA intermediate of the Tpt1 (tRNA 2'-phosphotransferase) reaction pathway. *RNA* **2020,** *26* (4), 373-381. © CSHL Press for RNA Society, 2020"

Sections 4.1, 4.3, 4.4, 4.5, and 4.6 of **Chapter 4** have been adapted from "Ageely, E.A.; Chilamkurthy, R.; Jana, S.; Abdullahu, L.; O'Reilly, D.; Damha, M.J.; Gagnon, K.T., Gene Editing with CRISPR-Cas12a Guides Possessing Ribose-Modified 5' Pseudoknots. *Nat. Commun.* **2021**, *accepted,* MS# 20-43072. © Nature Research, 2021"

Appendix I is reproduced from "Munir, A.; Abdullahu, L.; Banerjee, A.; Damha, M. J.; Shuman, S., NAD^+-dependent RNA terminal 2' and 3' phosphomonoesterase activity of a subset of Tpt1 enzymes. *RNA* **2019,** *25* (7), 783-792. © CSHL Press for RNA Society, 2019"

Preface

The intellectual guidance and funding for the projects presented in this thesis were provided by Prof. Masad J. Damha.

I have been very fortunate to collaborate with world-class researchers throughout my PhD studies at McGill. I am grateful to all of these colleagues for their research efforts, knowledge, and advice during this time. These collaborations have been incredibly fruitful, and my projects could not have been realized without the significant contributions of the researchers listed herein.

In **Chapter 2**, Dr. Annum Munir carried out ^{32}P radiolabelling of our RNA substrates and performed the subsequent enzymatic assays with Tpt1, under the supervision of Prof. Stewart Shuman at the Memorial Sloan Kettering Cancer Center. Dr. Ankan Banerjee successfully produced co-crystals of Tpt1 in a product mimetic state, leading to useful mechanistic insights into Tpt1 activity.

In **Chapter 3**, Swathi Dantuluri carried out ^{32}P radiolabelling of our RNA substrates, and performed the subsequent enzymatic assays with Tpt1 under the supervision of Prof. Stewart Shuman at the Memorial Sloan Kettering Cancer Center.

In **Chapter 4**, Dr. Eman Ageely and Ramadevi Chilamkurthy, under the supervision of Prof. Keith Gagnon at Southern Illinois University, performed the splint ligations of our crRNA 5'-handles to universal guides as well as Cas12a enzyme expression and purification, while these experiments were carried out by Adrian Pater for the Cas9 enzyme. The cleavage assays reported in this chapter (and **Chapter 5**) were carried out in their lab. Dr. Daniel O'Reilly and Dr. Sunit Jana were the original members of the Damha lab who began this collaboration, and much of the initial work including the synthesis of the crRNA in Figure 4.8 and part of Figure 4.9 was done by them.

In **Chapter 5,** Seth Eddington, under the supervision of Prof. Keith Gagnon at Southern Illinois University, performed the splint ligations of our crRNA 5'-handles to universal guides as well as the cleavage assays for all of the RNAs mentioned.

Abstract

The exponential developments made in the field of mRNA vaccines, CRISPR technology and antisense therapies have solidified nucleic acid therapeutics' presence at the forefront of modern medicine. With an ability to engage targets that are otherwise undruggable by conventional therapeutics, nucleic acid-based drugs have opened new avenues for treating intractable diseases. Synthetic oligonucleotides are now recognized as the third major drug discovery platform, in addition to small molecules and antibodies. Additionally, they provide valuable probes to study key structure-function relations and offer mechanistic insights at the molecular level of important enzymatic processes. In this thesis, we highlight our efforts in synthesizing novel RNA molecules as well as their use for gaining mechanistic insights of tRNA 2'-phosphotransferase (Tpt1) and CRISPR-Cas12a systems.

Tpt1 is an essential agent of fungal tRNA splicing that removes the 2'-PO_4 group generated at the splice junction during RNA repair. While Tpt1 activity is essential for fungal growth, it is inessential in bacterial and mammalian taxa. As such, Tpt1 is seen as an attractive antifungal target. In light of this, we have implemented an optimized automated synthetic route for the synthesis of a series of chemically modified RNA substrates that mimic the splice junction of tRNA. The success of these analogues as Tpt1 substrates have uncovered useful insights into the kinetic mechanism and substrate specificity of this enzyme.

CRISPR-Cas12a systems have emerged as a leading technology for creation of model systems, therapeutic gene editing, and diagnostics. We have explored the functional impacts of chemical modification to an RNA pseudoknot structure essential to the gene editing activity of CRISPR-Cas12a. We found that significant modifications of this structure are tolerated *in vitro*, however, *in vivo* activity requires retention of several native ribose 2'-OH contacts. Our findings suggest that pseudoknot structure and chemistry may have some regulatory influence over Cas12a activity, and that chemical modification of the 5' pseudoknot will facilitate therapeutic development.

The rules for successful chemical modification of 2'-PO_4 RNAs and CRISPR RNAs (crRNAs) described in this thesis should accelerate the development of novel therapeutics and may be valuable for CRISPR-Cas gene editing and diagnostic applications.

Résumé

Les récentes avancées dans le domaine des thérapies à base d'oligonucléotides, tels que les vaccins en ARN messager, les outils CRISPR et les thérapies antisens, ont mis en lumière l'intérêt des acides nucléiques comme potentiels outils thérapeutiques puissants dans la médecine moderne. En effet, ces technologies permettent de traiter des maladies jusque-là très compliquées à cibler avec les thérapies conventionnelles. Les oligonucléotides de synthèses sont même désormais reconnus comme le troisième principal outil de l'arsenal thérapeutique, avec les molécules de synthèses et les anticorps. Par ailleurs, les acides nucléiques permettent également d'étudier des procédés biochimiques, tels que l'élucidation des relation structures-fonctions ou la compréhension des mécanismes moléculaires des enzymes. Ainsi, dans cette thèse, nous présentons la synthèse de nouvelles molécules d'ARN utilisées pour mieux comprendre les mécanismes de Tpt1 (tRNA 2-phosphotransferase) et du système CRISPR-Cas12a.

Tpt1 est un agent essential pour l'épissage des ARN de transfert (ARNt) des fonges. L'enzyme fonctionne en enlevant le groupe 2'-PO_4 au site d'épissage pendant la réparation de l'ARN. Tpt1 est essentielle pour la croissance fongique, alors qu'elle n'a pas d'utilité dans le métabolisme des bactéries et des mammifères. C'est pourquoi TpT1 est une cible intéressante pour des thérapies antifongiques. Nous avons donc optimisé une route de synthèse automatisée pour fabriquer des substrats en ARN, contenant des modifications chimiques, qui imitent la jonction d'épissage de l'ARNt. Ces analogues de substrats de Tpt1 ont permis de révéler de nouvelles informations sur la cinétique et la spécificité de cette enzyme.

Le système CRISPR-Cas12a est unes des technologies les plus à la pointe pour la création de nouveaux modèles géniques, pour l'édition de gêne ou pour les diagnostiques. Nous avons étudié l'impact des modifications chimiques sur une structure de pseudonœud d'ARN, qui est essentielle à l'activité d'édition génomique de CRISPR-Cas12a. Nous avons observé que les versions largement modifiées chimiquement de la structure sont bien tolérées *in vitro*. Cependant, l'activité *in vivo* nécessite de conserver certains motifs 2'-OH (motifs natifs). Notre recherche montre que la structure pseudonoeud, ainsi que la

composition chimique, peuvent toutes les deux avoir des effets sur l'activité de Cas12a, et que les modifications sur les positions 5' du pseudonoeud peuvent faciliter le développement thérapeutique.

Ainsi, les règles établies dans cette thèse pour les modifications chimiques d'ARN 2'-PO_4 et d'ARN CRISPR (crARNs) peuvent servir à accélérer le développement de nouvelles thérapies, ainsi que le développement d'outils d'édition géniques, ou même de diagnostiques.

Acknowledgements

This thesis is a culmination of ideas and experiences that would not have been possible without the encouragement of my mentors, colleagues, friends, and family, to whom I would like to express my deepest appreciation.

I would like to thank Professor Masad J. Damha for his unrivaled guidance and support throughout my PhD studies, and all of the opportunities he has provided me with. I have learned so much through our discussions which have shaped the way I think as a scientist. I am in awe of his wisdom and the impact he has on the Nucleic Acid Chemistry community. Dr. Damha has a real skill for identifying student's interests and talents, and then setting them up with the opportunities needed to develop them further. I cannot think of a better quality for a mentor. In Dr. Damha, I found a true mentor with a genuine concern for his students' success and well-being. Most importantly, I have learned just how far being an honest, kind, and compassionate person can get you. I am so grateful to have had the opportunity to conduct research through his lab.

To the Damha Lab – my home for the last five years. I could not have made a better choice for a lab. I have been so fortunate over the years to have a place that I would genuinely look forward to going into every day, and that was because of the people I had the pleasure of working with. To Danielle – who taught me *everything* I know about the gene machines and oligo synthesis and purification. I remember watching you troubleshoot with the machines and marveling at your determination and ingenuity. You were so welcoming to me from the very beginning. (P.S. the "Welcome Leonora" on my fumehood is still there, to this day, and no one is ever allowed to erase it!) To Hala, my lab mom and "deerest" friend. Thank you for helping me adjust to my new life in Montreal and letting me vent about being homesick every time I went longer than a few weeks without seeing my family. You have remained a loving friend throughout these past years and I can always count on you for sage advice and support.

Dan, James, and Roberto, – there was a time when we were the only four students in the lab and you three became like brothers to me. Dan, it is so inspiring to see the passion that you have for the field and I appreciate you being a constant source of information and suggestions whenever I have any issues with anything – even now from UMass, you're always just a quick text away! James, you're the only person that has been in the lab throughout the entire duration of my PhD… and I can't believe you've put up with all of my annoying antics this whole time. What a champ. Thank you for always giving me a new level of *cool* to aspire to. Roberto, I could write a separate thesis on just our friendship over the last five years. You have been a true companion on this journey, and I am lucky to know you. We have learned so much together – troubleshooting experiments and fixing machines with you was always such a fun time. You have been such a great friend beyond the lab as well. I believe very few PhD students are fortunate enough to have someone they can truly depend on, in all ways. I will cherish our time together.

To everyone else in the Damha lab, past and present - and namely Ada, Adam, Halle, Luisa, Elise, and Maryam (my first interaction with the Damha lab, who did a fabulous job selling the lab to me) - it has been such a pleasure working with you all. Also, to the Damha lab alumni who I have met at conferences and other events, thank you for always treating me so well. It is an honour to be among you!

To my extended lab family - Kaleena, Ashkan, Morten, and Josh – The last couple of years with you guys in the office have been some of the best! I have loved our intellectual discussions (ahem… gossip sessions), and our many, many coffee breaks. I also thank you - Morten, Kaleena, and Ashkan - for all of the wonderful suggestions you have made regarding my projects and experiments. The impact that your advice has had on this thesis is incredible.

Over the last few years I have had the pleasure of working with some wonderful Post Docs in the Damha Lab. Dr. Sunit Jana, thank you for all of your advice - especially in the beginning of my PhD - and your abundance of uplifting stories in my times of need. Dr. Miguel Garavís, the best post doc any lab could have, and an amazing friend! What a

pleasure it was having you in the lab. You were always willing to help with any issues I had, and did not hesitate to teach me new techniques, even when you were busy with your own things. You were always “down” to go out and have a good time outside of the lab as well. The Core 4 isn’t the same without you.

I would like to thank the McGill Chemistry Department and their wonderful staff who have made my experience in the department so pleasant. Particularly, I would like to thank Linda Del Paggio for her genuine interest and enthusiasm and our fun chats in the office, and Chantal Marotte for all her help throughout the years. I am also grateful to my committee members, Prof. Nicolas Moitessier and Prof. Hanadi Sleiman who have offered so much feedback and advice throughout my studies. Thank you to Dr. Alex Wahba and Nadim Saadeh for all of their Mass Spec efforts and their expertise. I would like to thank the incredible collaborators I have had the pleasure of working with – Prof. Stewart Shuman and Prof. Keith Gagnon and the students in their labs who have directly contributed to the success of the projects in this thesis. This thesis would not have been possible without your expertise and dedication.

Beyond the lab, I have been very fortunate to have an incredible support system who have made this experience so much more enjoyable for me. To my loving friends in Hamilton, Jessica, Joannie, Xochitl, Caman, and Rose, thank you for always being such supportive friends and keeping this friendship alive from a distance for so many years. To Allie – you have made this last year so much fun. Thank you for being such an honest and thoughtful friend. I cannot wait to continue to celebrate so many successes with you. To Ali, your sweetness, patience, and encouragement throughout this writing process have been a true blessing and I am so grateful for you.

À Daniel et Louise, pour m'avoir ouvert leur cœur et leur maison au cours des cinq dernières années. Je ne peux pas imaginer faire mes études de doctorat dans un environnement plus affectueux et de soutien. Je suis éternellement redevable à vous.

Finally, I would like to thank my family who have shown me nothing but love and support throughout my entire life and have pushed me to always reach for my highest potential. You have cheered for me during good times, and consoled me through hard times, and through it all, you have been so patient. To my brothers, Pajtim and Shpetim, thank you for being my best friends in life. I know I can count on you two to not let the "Dr" title go to my head. To my niece Adriana – thank you for always reminding me how cool it is to be a scientist. It melts my heart to hear when you say you want to become a scientist like Halla! To my parents, Selvete and Vehbi, every achievement I have made in my life is because of your incredible strength and guidance. It has not been easy being away from my family for so long, but if I have learned anything from my parents, it's that every sacrifice is worth it in the end.

Thank you.

Thank you to Dr. Damha, Dr. Daniel O'Reilly, and Roberto El-Khoury for their time and commitment in editing this thesis. I am also very grateful to Dr. Aurélie Lacroix for translating the Abstract of this thesis into French.

Table of Contents

List of Figures

Supplementary Figures

List of Tables

List of Abbreviations

2'-F-RNA (or 2'-F)	2'-fluororibonucleic acid
2'-FANA	2'-fluoroarabinonucleic acid
2'-O-MOE	2'-O-methoxyethyl
2'-OMe	2'-O-methyl
2'-P-ADPR	2'-phospho-ADP-ribosylated RNA
2'F-ANA	2'-deoxy-2'-fluoroarabinonucleic acid
A	alanine
Ac	acetyl
ACN (or MeCN/CH_3CN)	acetonitrile
AcOH	acetic acid
ADP	adenosine diphosphate
AGO2	Argonaute 2 protein
Ala	alanine
ALE	acetal levulinyl ester
AMP	adenosine monophosphate
ANA	arabinonucleic acid
Ar	argon
ara	arabinose
Arg	arginine
ASO (or AON)	antisense oligonucleotide
ATP	adenosine triphosphate
B	base (nucleobase)
BNA	bridged nucleic acid
bp	base pair
br	broad
Bz	benzoyl
Cas	CRISPR-associated
CD	circular dichroism
CE	cyanoethyl

CPD (or CPDase)	cyclic phosphodiesterase
CPG	controlled pore glass
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
d	doublet
dA	2′-deoxyadenosine
Dbr1	lariat debranching enzyme
dC	2′-deoxycytidine
DCA	dichloroacetic acid
DCM (or CH2Cl2)	dichloromethane
dd	doublet of doublets
ddd	doublet of doublet of doublets
DDT	dithiothreitol
dG	2′-deoxyguanosine
DIPEA	diisopropylethylamine
DMAP	dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DMTr	dimethoxytrityl
DNA	deoxyribonucleic acid
dq	doublet of quartets
DSB	double stranded break
dsDNA	double stranded DNA
dsRNA	double stranded RNA
dt	doublet of triplets
dT (or T)	2′-deoxythymidine
EDTA	Ethylenediaminetetraacetic acid
ESI	electrospray ionization
EtOAc	ethyl acetate
ETT	5-ethylthiotetrazole
FDA	Food and Drug Administration

Fmoc	fluorenylmethoxycarbonyl
FOA	fluoroorotic acid
gRNA	guide RNA
GTP	guanosine triphosphate
H	histidine
His	histidine
HPLC	high pressure liquid chromatography
HRMS	high resolution mass spectrometry
iBu	isobutaryl
IE (or IEX)	ion exchange
KO	knock-out
LCMS	liquid chromatography - mass spectrometry
Lev (or Lv)	levulinyl
LNA	locked nucleic acid
m	multiplet
min	minute
miRNA	micro RNA
MOPS	morpholinyl sulfonic acid (buffer)
mRNA	messenger RNA
MS	mass spectrometry
NAD	nicotinamide adenosine diphosphate
NHEJ	non-homologous end joining
NMN	nicotinamide mononucleotide
NMP	N-methylpyrrilidone
NMR	nuclear magnetic resonance
nt	nucleotide
nts	nucleotides
OD	optical density
ON	oligonucleotide
ONA	oxepane nucleic acid
ORF	open reading frame

Ox	oxepane
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
Ph	phenyl
PN	phosphoramidate
Pnkp (or Pnk)	polynucleotide kinase-phosphatase
PNPase	polynucleotide phosphorylase
PO (or P)	phosphate
pre-mRNA	precursor messenger RNA
PS	phosphorothioate
py	pyridine
q	quartet
R	arginine
rA	riboadenosine
rC	ribocytidine
rG	riboguanosine
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
Rnl	RNA ligase
RNP	ribonucleoprotein
RP	reverse phase
rRNA	ribosomal RNA
rU or U	uridine
s	singlet
s (or sec)	second
SDS	sodium dodecyl sulphate
SEC	size-exclusion chromatography
SEM	standard error of the mean

sgRNA	single guide RNA
siRNA	short interfering RNA
ssDNA	single stranded DNA
t	triplet
TBDMS	tert-butyldimethylsilyl
TBE	Tris-borate EDTA (buffer)
TCA	trichloroacetic acid
td	triplet of doublets
TEA (or Et3N)	triethylamine
TEAA	triethyl-ammonium acetate
Tf	triflate (triluoromethanesulfonyl)
THF	tetrahydrofuran
TLC	thin layer chromatography
Tm	melting temperature
Tpt1	tRNA 2'-phosphotransferase
tracrRNA	trans-activating RNA
TREAT-HF	triethylamine trihydrogen fluoride
Trl	tRNA ligase
tRNA	transfer RNA
Trp	tryptophan
UNA	unlocked nucleic acid
UnyLinker	universal linker (CPG)
UV	ultraviolet
WC	Watson-Crick (base pairing)
WT	wild type
YPD	yeast, peptone, dextose (liquid medium)
β	beta

Chapter 1

Introduction

1.1 History of Nucleic Acids

"The double helix is indeed a remarkable molecule. Modern man is perhaps 50,000 years old, civilization has existed for scarcely 10,000 years, and the United States for only just over 200 years; but DNA and RNA have been around for at least several billion years. All that time the double helix has been there, and active, and yet we are the First creatures on Earth to become aware of its existence."

- Francis Crick (1916–2004)

In 1868, Johann Friedrich Miescher, a Swiss biochemist at the University of Tübingen, undertook a research project with such unforeseeable significance, which all researchers aspire to do. Miescher set out to study the biochemical constituents of cells – specifically, leukocytes – and through these studies, discovered the presence of an unknown substance which would soon be coined "nuclein"[1]. Upon further characterization of this substance, Miescher noted throughout the isolation procedure that it was not a lipid, and that it could not be a protein either, as it was resistant to protease digestion[2]. Analyses of its elementary composition revealed that, unlike proteins, it contained large amounts of phosphorous and notably lacked sulphur. Miescher quickly realized that he had a novel molecule in his hands, and having isolated it from the nuclei of cells, he named this molecule nuclein. He noted in subsequent work that this compound is a characteristic component of all nuclei and that it may be linked to the function of this organelle. He even hinted at the idea that this compound may play a role in hereditary transmission through something akin to a code[3].

Albrecht Kossel was able to separate nuclein from its proteinaceous component to isolate a non-protein phosphorus-containing component and in 1889, Richard Altmann, a former student of Miescher, called this substance "nucleic acid" due to its acid precipitation properties. Kossel was able to digest nucleic acids down into smaller components, and through elemental analysis identified five nucleobases: thymine, adenine, cytosine, guanine, and uracil[2]. Subsequent studies of the digested fragments led Phoebus Levene to characterize the presence of a ribose sugar and specify the arrangement between the base, the sugar, and the phosphate. Levene is also credited with differentiating between ribonucleic acids (RNA) and deoxyribonucleic acids (DNA) at the sugar level[4].

It was in the mid-twentieth century that some of the most fundamental DNA discoveries were made. In 1944, Oswald T. Avery made the breakthrough suggestion that DNA, not proteins as previously widely believed, was the carrier of genetic information[5]. In 1949, Erwin Chargaff discovered that the base composition of DNA varies between species, but that within each species the bases are always present in fixed ratios of purines to pyrimidines: the same number of adenine as thymine bases and the same number of cytosine as guanine bases[6, 7]. This revelation expedited the discovery of the structure of the DNA double helix. Rosalind Franklin and Maurice Wilkins determined the X-ray diffraction pattern of DNA and through various analyses, determined that the phosphate groups must be exposed to water (on the outside of the double helix), while the nucleobases hidden away from water, on the inside of the helix[8, 9]. Piecing together all of the information up to this point, James Watson and Francis Crick proposed several potential models before establishing the famous double helical structure of DNA in 1953[10]. Chargaff's base-pairing rule helped them determine that the nucleobase sequence of one strand automatically dictates the sequence of the second, complementary strand, and thus one strand can be used as a template for copying genetic material[11]. For discovering the molecular structure of nucleic acids and determining their role in transferring genetic information; Watson, Crick and Wilkins were awarded the Nobel Prize in physiology and medicine in 1962. This historic discovery sparked immense interest in the field and in particular the synthesis, function, and potential applications of nucleic acids in translational research.

1.2 Nucleic Acid Structure

1.2.1 Structural Components of Nucleic Acids

Nucleic acids are vital molecules that are essential for carrying the genetic material of life in all living cells. It was the discovery of DNA as the carrier of the hereditary information that focused immediate attention on elucidation of its structure and biological function. Nucleic acids are long biopolymers composed of a linear array of monomeric units termed nucleotides. A nucleotide is a phosphorylated nucleoside composed of three components: a pentose sugar, a nitrogenous heterocyclic nucleobase and a phosphate group (**Figure 1.1a**). These nucleotide building blocks are elegantly assembled by polymerase enzymes within the cells to yield two major forms of genetic material: Deoxyribonucleic

Acid (DNA) and Ribonucleic Acid (RNA). RNA contains ribose sugar while the DNA sugar lacks the OH group at C-2 and is therefore termed 2-deoxyribose. The nucleobases are attached to the anomeric carbon (C1') of the sugar moiety *via* a β-glycosidic linkage to the N-1 nitrogen of pyrimidines and the N-9 nitrogen of purines. There are two types of nucleobases: monocyclic pyrimidines (cytosine (C), thymine (T) and uracil (U)) and bicyclic purines (adenine (A) and guanine (G)) (**Figure 1.1b**). DNA and RNA share the same purines and pyrimidines except that RNA contains uracil while DNA replaces uracil with thymine. An array of nucleotides is referred to as an oligonucleotide (ON), wherein nucleotides are linked consecutively *via* phosphodiester linkages between the 3'-hydroxyl group of one nucleotide and the 5'-hydroxyl group of the next nucleotide (**Figure 1.1c**). By convention, ON sequences are written from the 5'-end to the 3'-end.

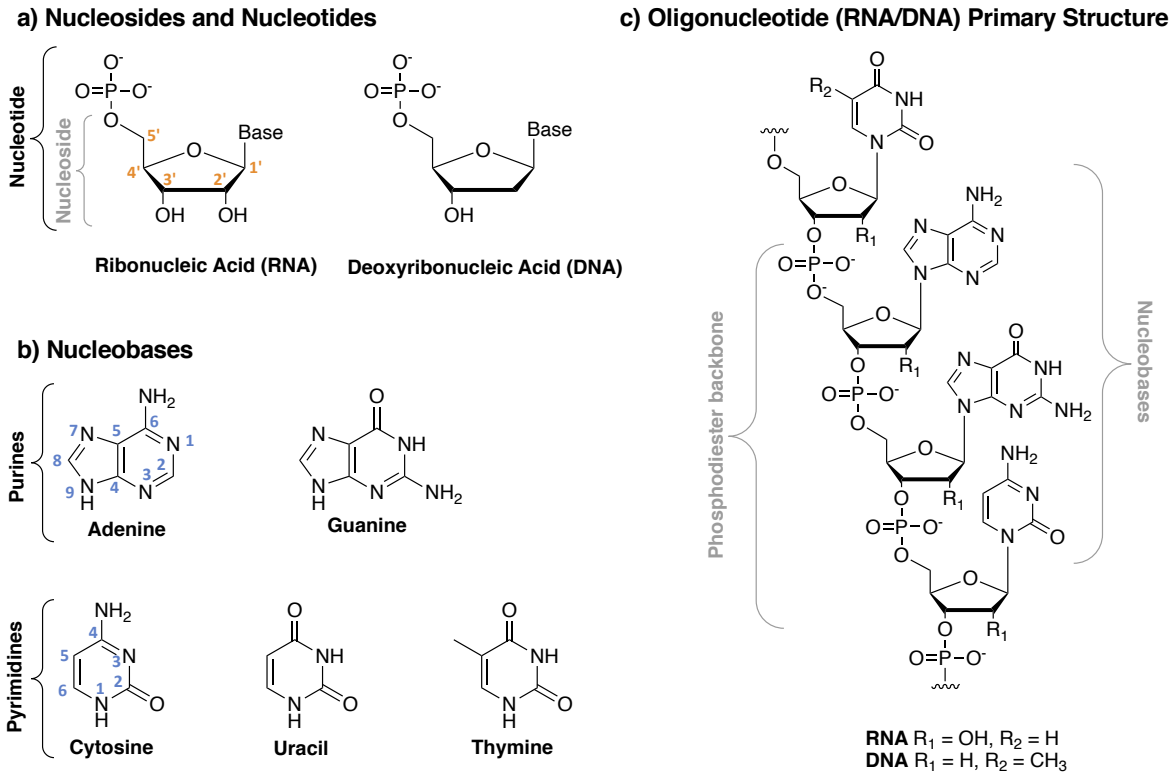


Figure 1.1. *Structural components of nucleic acids. a) The sugars ribose (RNA) and deoxyribose (DNA) and their numbering conventions. Nucleotides are phosphate esters of nucleosides. b) The nitrogen heterocycles used in RNA (A = adenine, G = guanine, C = cytosine, U = uracil) and DNA (A, G, C, T = thymine) and their atom numbering conventions.*

1.2.2 Conformation of nucleosides

Furanose sugar rings are naturally nonplanar with some atoms twisted out of plane to adopt the most stable conformation with minimized ring strain and non-bonded strain between their substituents. This twist, or nonplanarity, is called puckering. The sugar pucker is identified based on the orientation of carbon-2' and carbon-3' with respect to the plane of C1'- O4' - C4' (**Figure 1.2**). The term *endo* implies that ring atoms are above the plane, on the same side as the C5' and the nucleobase, while the term *exo* implies that they are below the plane, on opposite sides (**Figure 1.2**). By convention, the conformation of nucleoside sugar rings is described using a phase angle value (P) which is calculated from the exocyclic and endocyclic angles of a furanose ring. These values, shown in the "pseudorotational wheel" in **Figure 1.2**, determine whether the conformation falls within one of four minima: North (P = 0˚, C3'-*endo*), East (P = 90˚, O4'-*endo*), South (P = 180˚, C2'-*endo*) and West (P = 270˚, O4'-*exo*)[12]. Natural nucleosides in solution rapidly equilibrate between these conformations and with RNA sugars generally preferring the North (C3'-endo, C2'-exo) conformation, while DNA sugars tend to adopt a South (C2'-endo, C3'-exo) conformation[13]. For most substituted furanose sugars a combination of factors such as steric, anomeric and gauche effects participate in shifting the conformation equilibrium towards one of the puckers, significantly populating a limited segment of the pseudorotational wheel[14]. Thus, modified sugars favouring a north conformation are termed "RNA-like", while those favouring the south conformation are considered "DNA-like." These differences in conformation determine the overall shape of the helix.

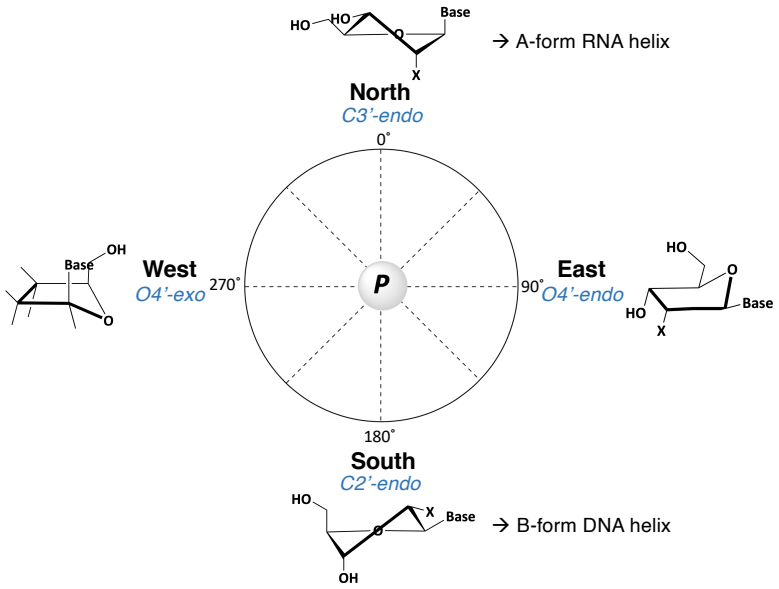


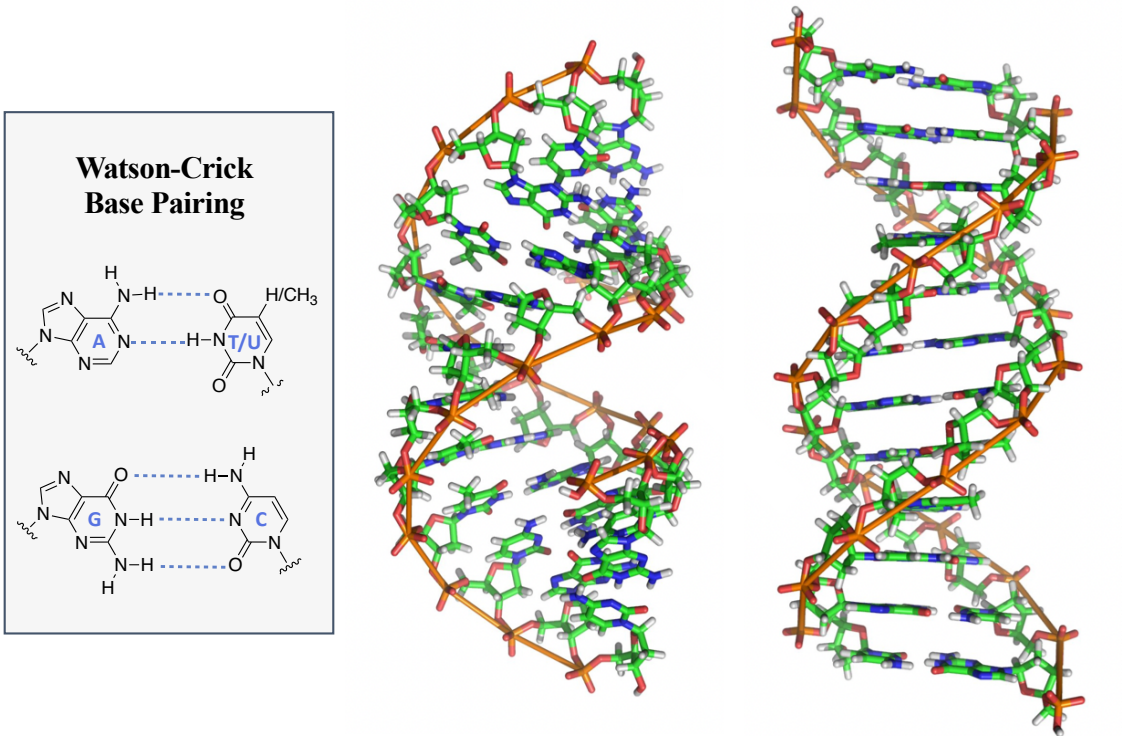
Figure 1.2. *Pseudoroational wheel and sugar pucker notations of the furanose ring, depicting the four conformational minima of nucleosides.*

1.2.3 The double helix

The DNA double helix is composed of two right-handed oligonucleotide chains, coiled around a central axis, with the two strands running antiparallel to one another[10]. The structure ultimately proved several theories, including Franklin's suggestion that the hydrophobic bases preferring to be on the inside of the structure, away from water, while the negatively charged phosphate backbone remains on the exterior. Additionally, a base pairing scheme with the nucleobases paired to each other *via* intermolecular non-covalent hydrogen bonding was proposed: Watson-Crick base pairing (**Figure 1.3**). This type of pairing confirmed Chargaff's observations that the ratios between G:C and A:T are always the same. Adenine forms two hydrogen bonds with thymine (or uracil) while guanine forms three hydrogen bonds with cytosine[10]. The helix is further stabilized through aromatic π-π base stacking of adjacent nucleobases[15-19]. Electrostatic repulsion between the negatively charged phosphates of the neighboring strands, however, destabilize the duplex[20]. Thus, the final structure of a nucleic acid duplex is the result of the balance among all these covalent and noncovalent interactions.

In the famous B-form double helix of DNA, each individual nucleotide has a *South* (*C2'-endo*) conformation and this leads to specific properties for the overall structure (**Figure 1.3**)[21, 22]. In contrast, RNA nucleotides, which have a 2′OH group, exhibit a *North*

(*C3'-endo*) conformation, which steers the helix conformation to an A-form (**Figure 1.3**)[23, 24]. The main reason for this is to avoid steric clash of the 2′OH in the major groove. Interestingly, DNA can also adopt an A-form helix and this is partly due to the inherent flexibility in the nucleotide conformation[25, 26].



Helical Parameters	A-form helix	B-form helix
Helix Sense	Right Handed	Right Handed
Sugar Pucker	C2′-*endo* (South)	C3′-*endo* (North)
Nucleotides per Turn	11	10
Twist per BP	32.7°	36°
Rise per BP	2.56 Å	3.3-3.4 Å
Major Groove Width	2.7 Å	11.7 Å
Major Groove Depth	13.5 Å	8.8 Å
Minor Groove Width	11.0 Å	5.7 Å
Minor Groove Depth	2.8 Å	7.5 Å

Figure 1.3. *Helical conformations and structural parameters of A- and B-form helices*[27].

1.3 The Central Dogma – Biological Role of Nucleic Acids

In molecular biology, the central dogma explains the flow of genetic information, from DNA to RNA to protein[28]. Originally coined by Watson and Crick in 1956, the central dogma suggests that the information present in DNA is essential to make up all proteins. Noting that DNA is restricted to cellular nuclei, where protein synthesis cannot occur, it was speculated that RNA serves as a transient messenger template for protein synthesis, originally copied from a DNA template.

1.3.1 The mechanism of gene expression

DNA is a very large molecule found in the nucleus of eukaryotic cells, which undergo self-replication during cell division[29]. Since DNA, the genetic material, is confined within the nucleus, RNA is used to shuttle genetic information from the nucleus to the cytoplasm *via* a process called transcription. In this process, DNA is converted into single-stranded RNA transcripts known as precursor mRNA (pre-mRNA)[30]. Normally during transcription, the 5'-end of the pre-mRNA is 'capped' by adding a 7-methyl-guanosine and subsequent methylation of one or two sugar residues of the adjacent nuc1eotides, leading to stabilization of the pre-mRNA[31-33]. Most pre-mRNA are poly-adenylated at the 3'-end, which stabilizes the RNA and assists in transportation of the mature mRNA into the cytoplasm[34]. The nascent pre-mRNA undergoes splicing (and other post-transcriptional modifications) during which the non-coding regions (introns) of RNA are removed, or spliced out, to join the coding regions (exons) into the appropriate mature messenger RNA (mRNA)[35]. mRNA is transported to ribosomes in the cytoplasm where it subsequently serves as a template for functional protein synthesis in a process known as translation[36]. Ribosomes are composed of proteins and catalytic RNAs (ribosomal RNA, rRNA), which use highly ordered transfer RNA (tRNA) to synthesize functional proteins from amino acids[37, 38]. tRNAs contain triplet codons which complement mRNA during translation, specifying one of the 20 natural amino acids, and sequentially attaching it to the growing peptide chain[38].

1.3.2 Faulty gene expression leads to genetic diseases

The process of gene expression and the biosynthesis of proteins is tightly controlled and regulated in cells. Although gene expression regulation occurs at all stages of transcription and translation, alterations or mutations can still occur. These mutations can

lead to changes in the expression or function of a protein, and ultimately give rise to genetic diseases. Traditionally, researchers have tackled these diseases at the protein level – usually designing and synthesizing small molecule drugs that bind to the protein to correct its function. More recently, however, nucleic acids have risen as an alternative to traditional small molecule pharmaceuticals, allowing us to intercept genetic diseases before RNAs are translated into proteins (**Figure 1.4.** Different types of oligonucleotide drugs and where they target relative to the Central Dogma. Traditional medicines only target proteins, while ON drugs can target genetic defects at all levels. Hence, oligonucleotides can be designed to hybridize and target a mRNA or a duplex DNA of interest in order to tackle diseases that were previously considered "un-druggable" with small molecule pharmaceuticals. These approaches are further discussed in Section 1.5 **Synthetic Oligonucleotides for Regulating Gene Expression**.

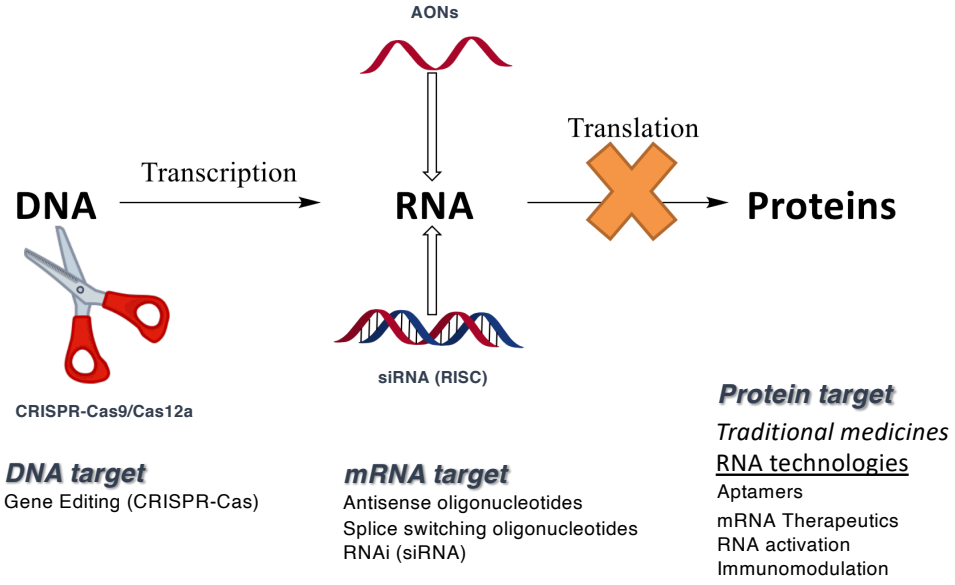


Figure 1.4. *Different types of oligonucleotide drugs and where they target relative to the Central Dogma. Traditional medicines only target proteins, while ON drugs can target genetic defects at all levels. Image courtesy of Dr. Daniel O'Reilly.*

1.4 The Chemical Synthesis of Nucleic Acids

1.4.1 Chemical synthesis of nucleosides and nucleotides

Following the great discovery of the structure and function of DNA, researchers quickly noticed the opportunity to chemically synthesize such molecules. It was in the early 1950's that A. M. Michelson and Sir Alexander Todd first synthesized 3' and 5'-phosphate analogues of each of the deoxynucleosides[39]. Shortly after, in 1955, they reported the first

synthesis of a dinucleotide by coupling two nucleotides together. This discovery initiated immense development towards the chemical synthesis of oligonucleotides. In recognition of his contributions, Todd received the Nobel Prize for Chemistry in 1957 for his work on nucleotides and nucleotide coenzymes. In the years that followed, several new developments were made in oligonucleotide synthesis, culminating eventually to the automated synthesis of gene fragments.

1.4.2 Solid phase synthesis of oligonucleotides

Perhaps the greatest technical advancement in the field of nucleic acid research was the development of an automated solid phase synthesis of oligonucleotides. While the first solid-supported synthesis of an oligonucleotide was achieved in 1965 by Bob Letsinger at Northwestern University, one of the first *automated* oligonucleotide synthesizer machines (colloquially referred to as a "Gene Machine") was built at McGill University in 1981, by Dr. Kelvin Ogilvie and his associates[40, 41]. This development reduced the time required to manufacture short DNA fragments from several weeks to just a few hours. The field witnessed tremendous growth in the synthesis of both DNA and RNA in the years that followed.

Several methods of solution-phase oligonucleotide synthesis have been devised over the years. However, the phosphoramidite method, pioneered by Caruthers and co-workers in the 1980s, is the gold standard method for DNA synthesis that has been used in the industry for almost 40 years. It provides fast reaction rates and high yield of products, making it possible to synthesize strands 20 to 150 nt in length[42]. Solid-phase synthesis is carried out on the solid support held between filters, in columns that enable all reagents and solvents to pass through freely. A typical solid phase synthesis cycle, along with the reagents and conditions commonly used in a standard automated oligonucleotide synthesis cycle are illustrated in **Figure 1.5**. The structures of the DNA and RNA 3'-phosphoramidites are also shown (**Figure 1.5** - bottom). The 2'-hydroxyl of these RNA monomers are protected with a *tert*-butyldimethylsilyl (TBDMS) group; the 5'-hydroxyl with the acid-labile dimethoxytrityl (DMTr), and the exocyclic amines of the nucleobases (A, G, and C) are protected in the form of amides[43-45]. T and U nucleobases do not require protection.

Oligonucleotide synthesis begins with a derivatized solid support (usually controlled pore glass (CPG) or polystyrene) and proceeds in the 3' to 5' direction, which is opposite to the 5' to 3' direction of DNA biosynthesis. The first step in the cycle is removal of the 5'-DMTr protecting group using an organic acid (trichloroacetic acid (TCA), or dichloroacetic acid (DCA)) in dichloromethane to free up the 5'-hydroxyl group. Next, the incoming phosphoramidite is first activated with a weak acid (e.g., 5-ethylthiotetrazole (ETT), pK_a 4.3) and coupled to the free 5'-hydroxyl of the nucleoside (or growing oligonucleotide) to form a phosphite triester linkage. Then a capping step is incorporated to ensure that any unreacted 5'-hydroxyl is "capped" in the form of an acetyl ester by reaction with acetic anhydride/pyridine/*N*-methylimidazole so that it cannot undergo further reaction in subsequent chain elongation steps. Next, the newly formed phosphite triester (P-III) linkage is oxidized with aqueous iodine/pyridine in tetrahydrofuran (THF) to afford the more stable phosphotriester (P-V) linkage. This cycle is repeated until the oligomer of desired length and base sequence is synthesized. Using this method, one complete cycle takes approximately 10-15 minutes depending on the phosphoramidite being coupled (Table 1.1). Once complete, the oligonucleotide is cleaved from the solid support and deprotected of any nucleobase and backbone protecting groups with concentrated aqueous ammonia ("NH_4OH"). For oligonucleotides containing RNA amidites, the TBDMS group is cleaved only after all other protecting groups have been cleaved. This is achieved using a mild fluorinating reagent such as triethylamine trihydrogen fluoride (NH_3/HF or TREAT-HF). The synthesized oligonucleotide can then undergo purification by either HPLC or gel electrophoresis followed by characterization (ESI-MS).

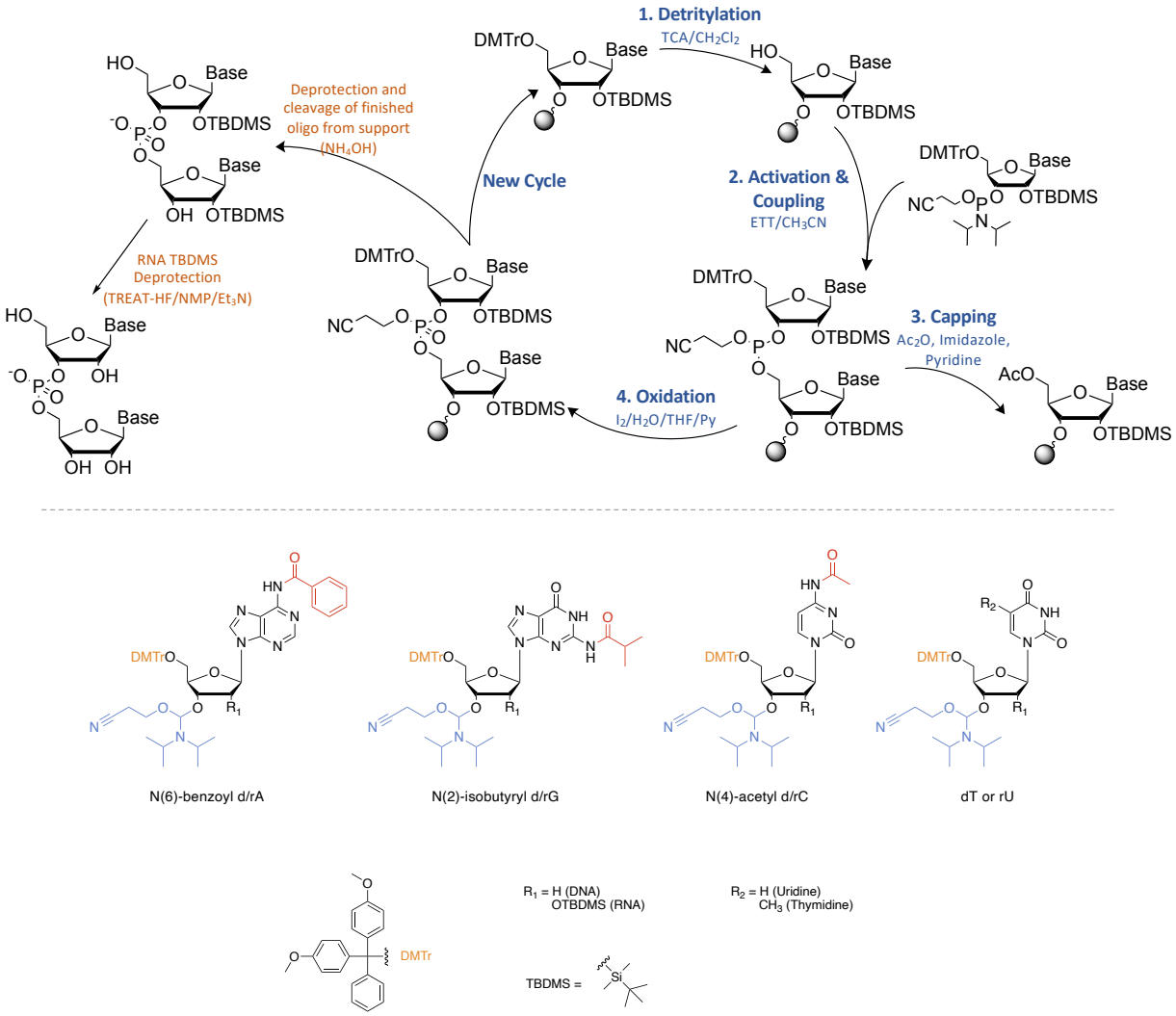


Figure 1.5. *(Top) Standard solid phase synthesis cycle for the synthesis of RNA. For DNA, the OTBDMS groups do not exist, and thus the final TBDMS deprotection is omitted. (Bottom) Phosphoramidite monomers used in the automated solid-phase synthesis of DNA and/or RNA oligonucleotides.*

Table 1.1 Typical timings used for the assembly of oligonucleotides via the phosphoramidite synthesis cycle.

Operation	Reagent/Solvent	Time
Detritylate	3 % TCA in DCM	90 s
Wash	Acetonitrile	30 s
Flush	Argon	10 s
Couple	0.15 M (RNA) or 0.1 M (DNA) phosphoramidite and 0.5 M ETT in MeCN	150 – 900 s (depending on amidite)
Wash	Acetonitrile	30 s
Flush	Argon	10 s
Cap	Ac_2O/py/THF and *N*-methylimidazole in MeCN	30 s
Wash	Acetonitrile	30 s
Flush	Argon	10 s
Oxidize	0.015 M I_2 in water/pyr/THF	45 s
Wash	Acetonitrile	30 s
Flush	Argon	10 s

1.4.2.1 The Solid Support

Different linkers can be used to connect the first phosphoramidite to the solid support. The linker must be stable to all reagents used in the solid-phase synthesis cycle, but cleavable under specific conditions at the end of the synthesis. The most commonly used linker is the succinyl linker[44]. Throughout this thesis, the commercially available UnyLinker solid support is used (**Figure 1.6**), except for syntheses described in Chapter 2, where we also used a special 3'-phosphate ON CPG which leaves a phosphate group at the oligonucleotide 3' terminus once it is released into solution.

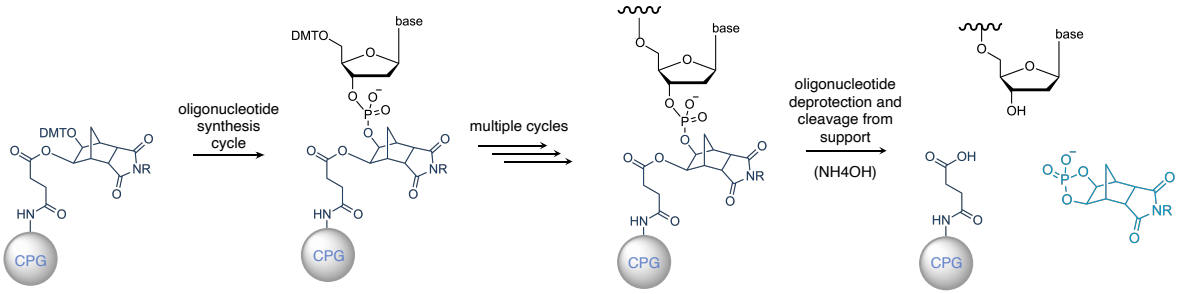


Figure 1.6. *Synthesis of the first base of an oligonucleotide on UnyLinker support (R = Ph), followed by cleavage of the oligonucleotide from the support.*

1.4.2.2 RNA 2'-hydroxyl protecting group

While DNA is generally stable under basic conditions, the 2'-hydroxyl group of RNA causes RNA to disintegrate when exposed to hydroxide. This is because, under basic conditions, the hydroxyl group can attack the vicinal 3'-phosphate linkage, resulting in the formation of a 2',3'-cyclic phosphate intermediate and release of a 5'-OH-RNA fragment[46]. Under acidic aqueous conditions RNA undergo isomerization to give 2'-5'-linkages. It is imperative, then, that for the chemical synthesis of oligonucleotides, the 2'-hydroxyl groups of RNA be protected throughout the synthesis and that such protecting group be stable through assembly of the RNA oligonucleotide chain. Silicon-based (silyl) moieties are typically used, as these are orthogonal to other acid- or base-labile protecting groups used throughout the synthesis[43]. The most popular group is the *t*-butyldimethylsilyl (TBDMS) protecting group, introduced by Ogilvie and co-workers[47]. While several silyl groups were surveyed, ultimately the 2'-*O*-TBDMS proved optimal and applicable for phosphoramidite coupling strategies[48].

The beta-cyanoethyl phosphate protecting group must be removed prior to TBDMS deprotection. This is to ensure the phosphotriester is transformed into a more stable phosphodiester, making it less susceptible to vicinal hydroxyl-initiated cleavage (**Figure 1.7**). Yet, it is important to note, that even the phosphodiester can undergo either chain cleavage (i.e. by nucleases or under basic conditions) or isomerization giving rise to 2'-5'-linkages (under acidic conditions). Thus, great care must be taken to ensure the integrity of RNA oligonucleotides.

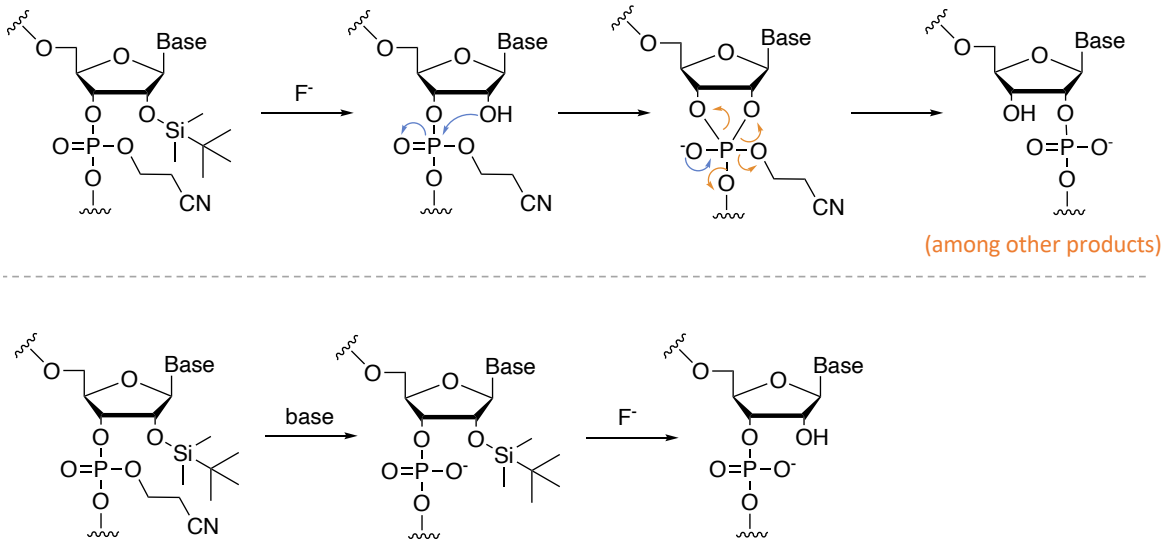


Figure 1.7. *Possible order of TBDMS deprotections. Top: Deprotection of silyl group prior to cyanoethyl cleavage triggers product disintegration. Bottom: Deprotection of silyl group after cyanoethyl cleavage stabilizes RNA.*

1.5 Synthetic Oligonucleotides for Regulating Gene Expression

1.5.1 Antisense-mediated gene targeting

In 1978, Zamecnik and Stephenson were the first to demonstrate the potential of oligonucleotides as chemotherapeutic agents by targeting mRNA to inhibit gene expression[49]. They reported the first synthetic DNA oligonucleotide delivered to cells, capable of inhibiting viral replication and oncogenic transformation caused by Rous sarcoma virus[49]. Their discovery birthed the new field of "antisense oligonucleotides" (ASOs). The term "antisense oligonucleotides" (ASOs) is used to describe oligonucleotides that bind to the "sense" mRNA in an anti-parallel, sequence specific manner via Watson-Crick base pairing. Today, ASOs are at the forefront of modern medicine, with several oligonucleotide therapeutics having been granted FDA approval for several different diseases[50]. They are generally 18-22 nt long and target a specific region of the mRNA inhibiting its translation through two distinct mechanisms: by sterically blocking the ribosome or splicing factors (steric block approach) or by triggering RNase H-mediated cleavage of the ASO:RNA heteroduplex (**Figure 1.8**) [49, 51]. Steric block ASOs are designed to bind the mRNA with high affinity, thus preventing ribosome assembly and leading to translational arrest without cleaving the mRNA[52], whereas ASOs that trigger

RNase H-mediated RNA cleavage are designed to form hybrids with RNA that resemble the native DNA:RNA heteroduplexes (RNA:RNA duplexes are not substrates for RNase H)[53].

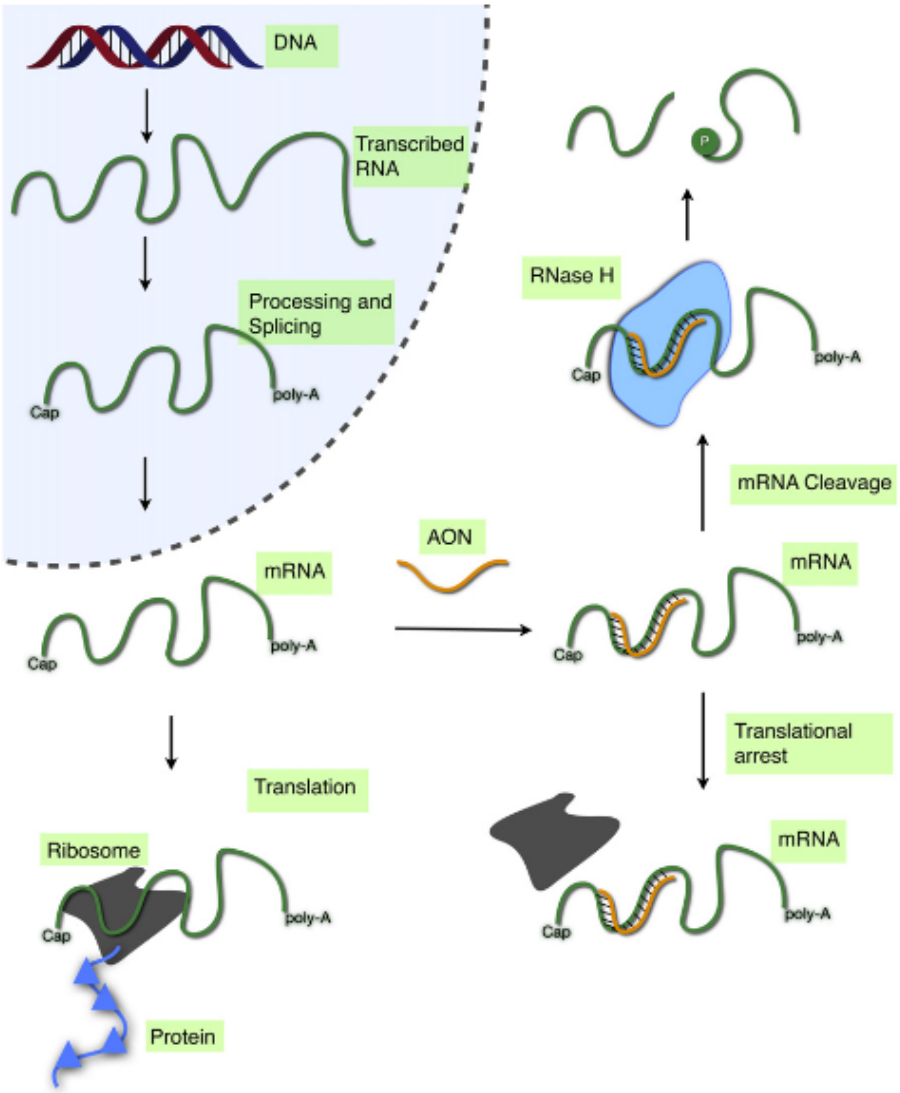


Figure 1.8. *Overview of the antisense approach for modulating gene expression via two routes: RNase H-mediated cleavage (top right), and steric block translational arrest (bottom right)[51].*

A different type of steric block mechanism includes ASOs that can modulate the splicing of mRNA. These ASOs are used in the modulation of alternative splicing in order to selectively *exclude* or *retain* a specific exon (exon skipping and exon inclusion, respectively)[54]. In these cases, the oligonucleotide 'masks' a splicing signal such that it becomes invisible to the spliceosome, leading to alterations in splicing decisions[55-57]. Typically, such splice correction approaches have been used to restore the translational

reading frame in order to rescue production of a therapeutic protein[58, 59]. However, the same technology can also be used for splice corruption, whereby an exon is skipped in order to disrupt the translation of the target gene[60].

1.5.2 RNA interference (RNAi)-mediated gene silencing

RNAi is a regulatory mechanism that occurs in plants, animals, humans that uses short double stranded (ds) RNA molecules (microRNAs or miRNAs) to achieve gene silencing *via* endonucleolytic cleavage of RNA. RNAi is also important in the defense against viruses, particularly in lower organisms. The process was first discovered by Andrew Fire and Craig Mello in 1998, when they observed that a dsRNA was capable of influencing the expression of a particular gene in the nematode worm *Caenorhabditis elegans*[61]. Shortly after, it was shown that synthetic RNA duplexes of 21-nt long could exogenously trigger this pathway in cultured mammalian cells[62]. Eight years after these seminal works, Fire and Mello were awarded the Nobel Prize in Physiology or Medicine "for their discovery of RNA interference - gene silencing by double-stranded RNA."

Short interfering RNAs (siRNAs) are the 19-bp RNA duplexes with 3'-dinucleotide overhangs at both ends[62]. The “guide” or antisense strand of the duplex is complementary to the target mRNA and its role is to eventually induce cleavage of the target, whereas the other strand is called the “passenger” strand. siRNAs act to guide the Argonaute 2 protein (AGO2), as part of the RNAi silencing complex (RISC), to complementary target transcripts (**Figure 1.9** RISC-mediated gene silencing via siRNA. Small interfering RNAs (siRNAs) enter the RNA-induced silencing complex (RISC), which consists of Argonaute 2 protein (AGO2), DICER1 and TARBP2, and the passenger strand is discarded. The guide strand directs the RISC to complementary target genes that are cleaved by the slicer activity of AGO251.)[63]. Complete complementarity between the siRNA and the target transcript results in AGO2-mediated cleavage of the target mRNA (between nucleotides 10 and 11 of the guide strand), leading to gene silencing[64, 65]. On the other hand, if the guide strand is partially complementary to the mRNA, the mRNA is inhibited via translational repression, degradation or cleavage, as seen for microRNAs (miRNAs)[66]. Once the process is complete, the RISC complex releases the cleaved mRNA and the complex is free to bind to and cleave another mRNA target, making this process entirely catalytic. siRNA

therapeutics are very attractive over the ASO techniques mentioned above, as only a very small amount of siRNA is required to reach its target and achieve potent gene silencing.

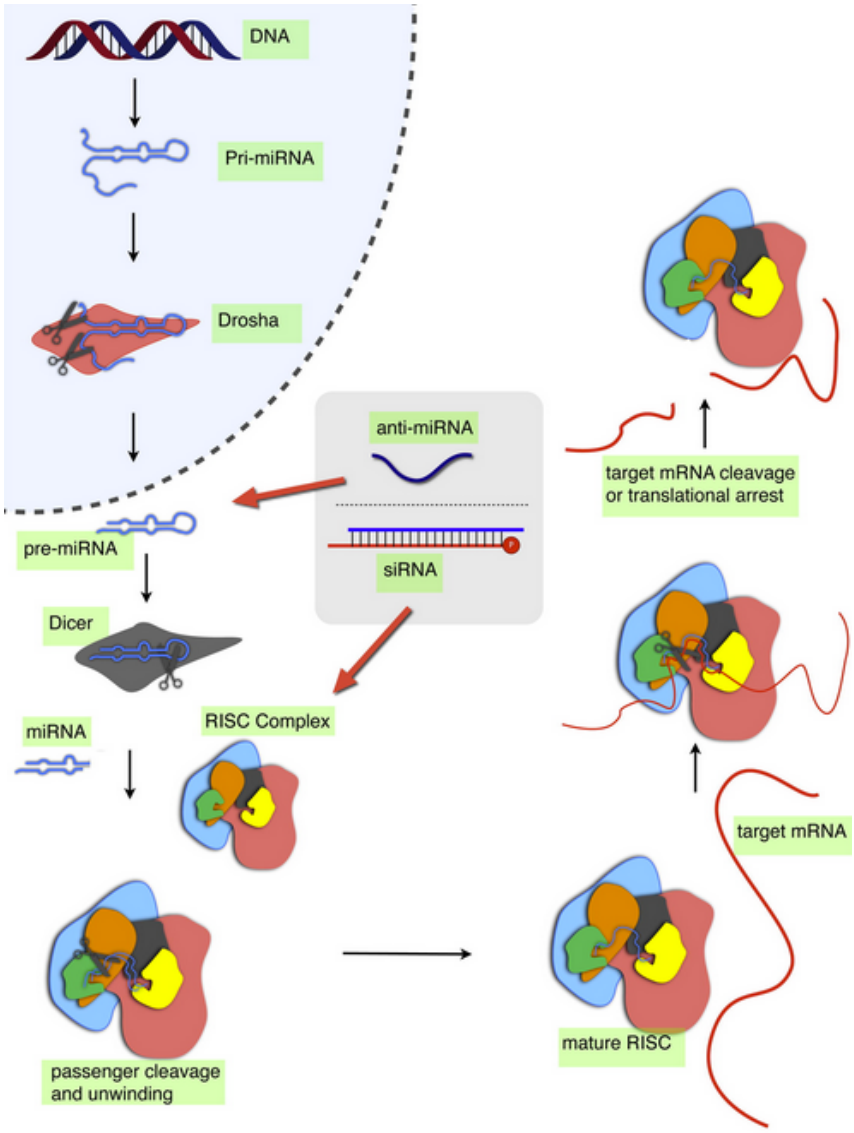


Figure 1.9 *RISC-mediated gene silencing via siRNA. Small interfering RNAs (siRNAs) enter the RNA-induced silencing complex (RISC), which consists of Argonaute 2 protein (AGO2), DICER1 and TARBP2, and the passenger strand is discarded. The guide strand directs the RISC to complementary target genes that are cleaved by the slicer activity of AGO2*[51].

1.5.3 Challenges facing oligonucleotide therapeutics

Statistically, a 17-mer oligonucleotide base sequence occurs just once in the base sequence of the human genome and antisense oligonucleotides would bind specifically to them due to base pairing specificity, Also, following translation, an mRNA can lead to production of many proteins (e.g., via alternative splicing pathways) and the traditional

drugs that bind to these proteins would be required in an equally large quantity. Relative to the proteins, the quantity of mRNA present is quite low and therefore a much smaller quantity of an equally efficacious drug would be required. However, that is not to say that oligonucleotide drugs do not come with their own set of drawbacks. The major drawbacks impeding the rapid development of oligonucleotide drugs is their inherent non-drug-like properties. Naturally occurring nucleic acids are very large molecules and highly negatively charged because of their ionized phosphate backbone. While such compounds can function well *in vitro*, their *in vivo* applications are highly limited. Furthermore, unmodified (native) oligonucleotide drugs exhibit poor cellular uptake and poor extracellular and intracellular stability due to nuclease degradation activity[54,67]. Additionally, oligonucleotides have the potential to induce off-target effects due to complementary binding between the oligonucleotide and unintended RNA with a sequence similar to the target RNA. Finally, immunostimulation and aptameric protein interactions pose further challenges[68]. Over the past few decades, nucleic acid chemists and biologists have worked intensely to develop strategies that overcome many of these obstacles. Recent progress has been rapid, broad, and exciting, with several oligonucleotide-based drugs already on the market and hundreds of candidates in early to late-stage clinical trials. In 2019 alone, over 150 companies were actively working on 500+ therapeutic projects, and 85% of the top 20 pharmaceutical companies had oligonucleotide therapeutic deals. A comprehensive review of the latest advances in this area, covering the approaches currently being pursued using nucleic acid therapeutics are documented in an excellent book by Agrawal and Gait[69].

1.6 Chemical Modifications of Nucleic Acids

The key to overcoming the hurdles encountered in oligonucleotide therapeutic development mentioned above has been the development of chemical modifications of nucleic acids. Nucleic acid chemists possess now a “toolbox” containing a diverse number of chemical modifications that upon incorporation into oligonucleotides allows them to control and adjust properties producing oligonucleotides with enhanced cellular uptake, binding affinity towards RNA targets, robust stability towards nuclease degradation while reducing their immunostimulatory capacity. A major advantage of informational drugs like

oligonucleotides (ONs) over traditional small molecule drugs is that the pharmacophore (molecular features that determine target specificity i.e. the sequence) of ONs can be optimized independently of the dianophore (molecular features that determine tissue distribution and metabolism) (**Figure 1.10**)[70]. When a dianophore for a particular tissue or cell is defined, it can be applied to a range of rationally designed pharmacophores.

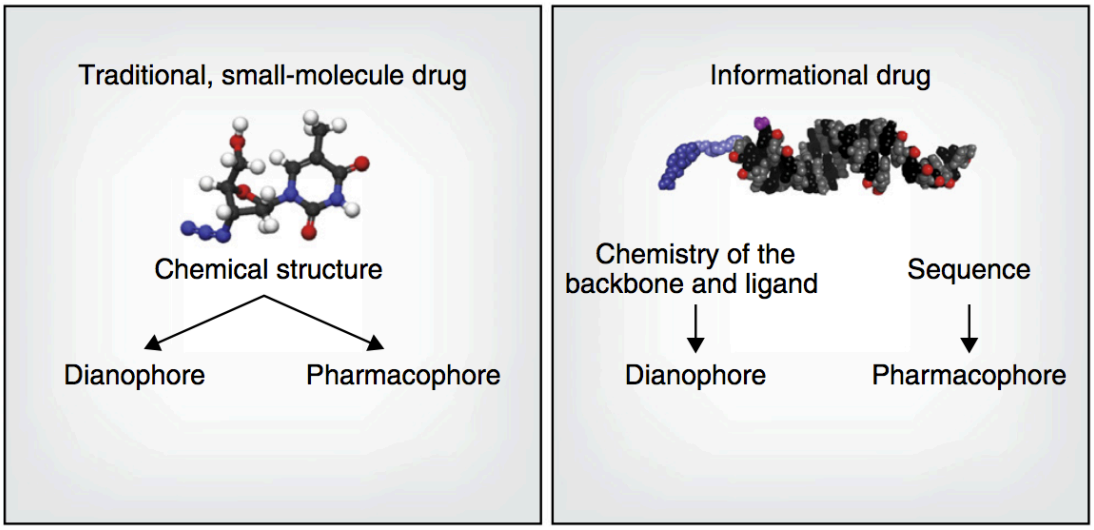


Figure 1.10. *Advantage of informational drugs over traditional, small-molecule drugs*[70].

Three major sites amenable for chemical modification of nucleic acids include (i) backbone modifications, (ii) sugar modifications, and (iii) nucleobase modifications.

1.6.1 Internucleotide linkage (backbone) modifications

The phosphodiester linkage in natural DNA and RNA is anionic in nature (pK_a ~2) and thus makes cellular uptake very difficult[51]. Even once in cells, these linkages are readily hydrolyzed by endo- and exonucleases. So, the modification of the phosphate moiety gives the opportunity to improve both nuclease stability as well as cellular uptake of oligonucleotides. Many internucleotide linkage modifications have been explored over the years, but one of the first modification to be incorporated into therapeutic oligonucleotides, the phosphorothioate (PS) modification, remains the most employed linkage in ASOs[71, 72,73] (and in many experiments reported in this thesis). This modification is created by replacing one nonbridging oxygen atom in the phosphate group with a sulfur atom. Besides increasing nuclease resistance, the PS linkage promotes cellular uptake,

increases ASO binding to plasma proteins and enhances tissue delivery, while retaining sufficient RNase H activation of target mRNA cleavage[74]. The sulfur atom creates a chiral centre at the phosphorus atom with two diastereoisomers: R_p and S_p[74]. All FDA approved ASO drugs to date utilize the PS modification and administer it as a mixture of the stereoisomers. However, recent efforts have been devoted to the synthesis of stereochemically pure PS-ASOs. It has been shown that PS linkages with R_p stereochemistry are more susceptible to nuclease degradation (vs S_p), however, they bind to target RNA with higher affinity[74]. As a result, the use of epimeric mixtures of PS is appropriate as it is a tradeoff between binding affinity and nuclease resistance. Some disadvantages of PS modifications worth mentioning include the decrease in thermal stability (0.5 ˚C per linkage) and increase in off-target effects due to non-specific protein binding. In fact, PS-ASO:protein interactions can affect distribution and tissue delivery, cellular uptake, intracellular trafficking, potency and toxicity. The toxicity issue can be addressed by altering the PS-ASO structure by additional chemical modifications to improve therapeutic index, and even target binding affinity. Some of these chemical modifications are described below[75].

1.6.2 Sugar modifications

The furanose sugar of nucleic acids has been extensively modified through the addition or substitution of substituents, restriction of the sugar conformation, as well as the expansion of the ring itself. Modifications at the sugar can tune the puckering of the furanose ring (described in Section 1.2.2) which consequently impacts the duplex conformation, binding affinity for target strands, as well as their interactions with proteins (i.e., serum stability and protein binding). Some notable examples of sugar modifications are shown in **Figure 1.11**, with the relevant ones pertaining to this thesis highlighted in grey.

Nucleotides are connected to each other through the 3'- and 5'- hydroxyls of their sugar rings, leaving the 2'-OH free and an obvious choice for modifications. Substitution of the 2'-OH of RNA with a fluorine atom (2'-F-RNA) was among the first studied RNA modifications[76, 77]. Fluorine is small (slightly smaller than Oxygen) and highly

electronegative and was found to help lock the nucleotide into a North conformation, maintaining and reinforcing the RNA-like pucker, making it an exceptional RNA mimic. This modification increases the binding affinity of ONs for target RNA due to stronger hydrogen bonding originating from inductive effects of the electronegative fluorine atom. Fluorine substitutions at the 2'-position of arabinonucleic acids (2'-F-ANA) have also been explored by our research group[78,79,80]. As the epimer of 2'-F-RNA, this modification adopts a South/East conformation, and is an excellent DNA mimic[81]. It is especially useful in antisense applications where it can act as a DNA mimic upon binding to a target RNA strand in order to recruit RNase H. 2'-F-ANA improves nuclease stability and is also well tolerated in the passenger strand of siRNA, when combined with other 2'- modifications[82-84].

Modifications of RNA with 2'-*O*-alkyl substituents have been extensively explored and found to generally present improved nuclease resistance and target binding affinity. 2'-*O*-methyl (2'-OMe) and 2'-*O*-methoxyethyl (2'-O-MOE) are RNA mimics which produce oligonucleotides with enhanced binding affinity and nuclease resistance[85, 86]. Additionally, 2'-OMe modifications are also known to decrease immunostimulatory effects while leaving potency unaffected. Although these RNA-like modifications do not illicit RNase H degradations, they are still widely used in ASOs in a "gapmer" strategy when combined with other DNA-like modifications. Additionally, their A-form helicity renders them well-suited for siRNAs as well.

Based on these modifications, researchers noted that the more confined or restricted the sugar ring is, the tighter it binds to target strands. This led to the development of "locked" nucleic acid (LNA). As the name suggests, the modification is locked in the North conformation through a 2'-*O*-methylene-4'-C bridge (**Figure 1.11**)[87, 88]. As a conformationally restricted, bicyclic nucleic acid, LNA has been observed to increase binding affinity to target RNA by an astounding 5-10 ˚C per insert, in addition to improved nuclease resistance and reduced immunostimulatory effects[87]. This is due to a few features: the locked sugar moiety renders this modification preorganized to match the A-form structure of target RNA; and there are additional hydrophobic interactions which improve base stacking. On the other hand, unlocked nucleic acids (UNA), lacking the bond between

C2' and C3', also exist and have been employed in applications where more flexibility is preferred (e.g. when targeting complex secondary structures)[89, 90, 91]. The high level of conformational flexibility accommodates binding of RNase H to the ASO and can increase the rate of mRNA cleavage[91]. They are also particularly useful in siRNA duplexes where the guide strand must dissociate from the passenger strand once it has been loaded onto RISC. Additionally, UNAs improve nuclease stability and lower toxicity[92-95].

2',5'-Linked RNA can be considered both a backbone modification, as well as a modification of the sugar, as the 2' and 3' substituents of the ring are being modified. Positive attributes of this modification such as improved nuclease stability as well as the simplicity and affordability of the modification made this an ideal modification to explore[96]. Beyond the furanose ring, expanded ring modifications have also been explored. Oxepane nucleic acids (ONA) were first described by our group as a means to introduce more flexibility and functionality to the sugar[97]. ONA:RNA hybrids are substrates for RNase H, although not as efficient as the native DNA:RNA hybrids. This modification continues to be explored by our group[98]. The sugar modifications mentioned above have been used in a variety of settings for probing substrate-enzyme interactions in Chapter 3, 4, and 5 of this thesis.

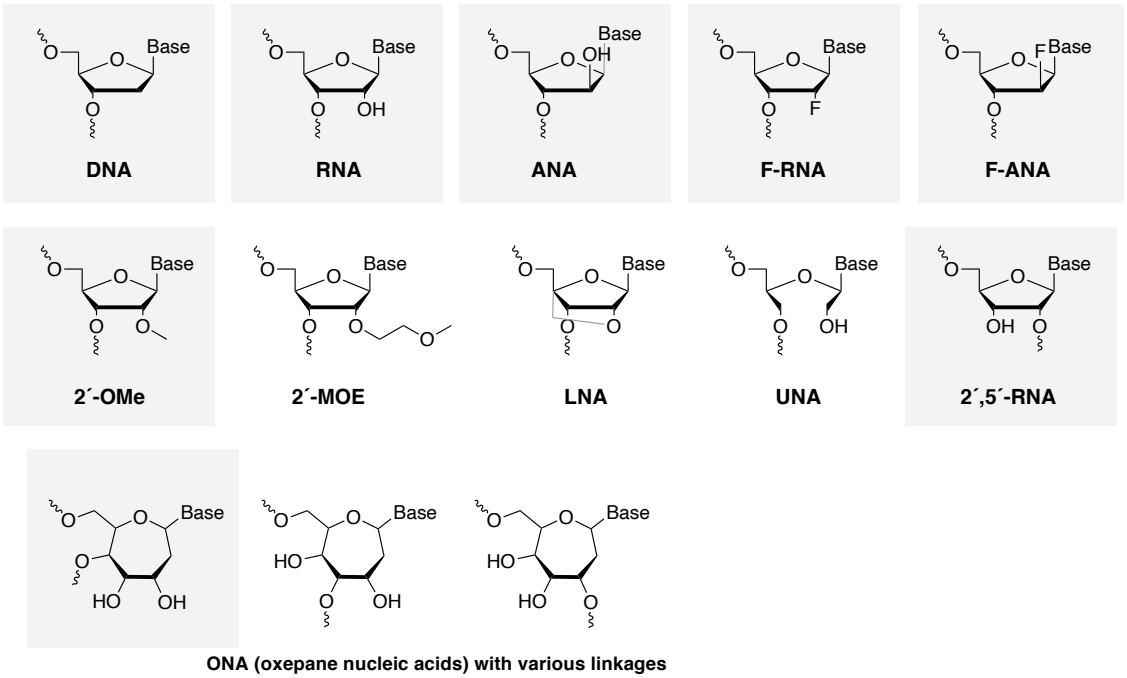


Figure 1.11. *An array of selected sugar modifications used in nucleoside analogues, with the ones studied in this thesis highlighted in grey.*

1.6.3 Nucleobase modifications

Modifications at the nucleobase have also been explored over the years, although not as heavily as the sugar and backbone modifications mentioned above. Such modifications allow for the modulation of the thermal stability of resulting oligonucleotides, the sugar conformation, as well as the strength of Watson-Crick base pairing interactions[51]. Some of the most notable ribo-specific nucleobase modifications include 5-methylcytosine and 5-methyluracil (thymine). These are modifications known to decrease immunostimulatory response. Although nucleobase modifications hold tremendous value in ON therapeutics, they have not been explored in this thesis and thus will not be elaborated upon any further.

1.7 RNA Repair Enzymes

Cells and viruses exploit a variety of RNA repair pathways to maintain or manipulate RNA structure in response to purposeful RNA breakage inflicted by transesterifying endoribonucleases, e.g., during physiological RNA processing (tRNA splicing) and under conditions of cellular stress (virus infection, unfolded protein

response)[99]. RNA cleavage by transesterification yields 2',3'-cyclic-PO_4 and 5'-OH ends. RNA repair enzymes capable of sealing 2',3'-cyclic-PO_4/5'-OH breaks are present in diverse taxa in all phylogenetic domains of life.

There are two styles of repair pathways that differ with respect to whether the RNA ligation event results either in a "clean" 3'-5' phosphodiester at the repair junction or a junction in which there is a covalent modification of the 2'-OH adjacent to the newly formed 3'-5' phosphodiester (**Figure 1.12**). The clean junction pathways are exemplified by: (i) the two-component bacteriophage T4 RNA repair system in which an end-healing enzyme polynucleotide kinase-phosphatase (Pnkp) converts 2',3'-cyclic-PO_4 and 5'-OH ends to 2'-OH, 3'-OH and 5'-PO_4 ends that are then sealed by an ATP-dependent 3'-OH/5'-PO_4 RNA ligase (Rnl1)[100-102]; and (ii) the Escherichia coli GTP-dependent RNA ligase RtcB that joins 2',3'-cyclic-PO_4 and 5' -OH ends via RNA-3'-PO_4 and RNA-(3')pp(5')G intermediates[103, 104].

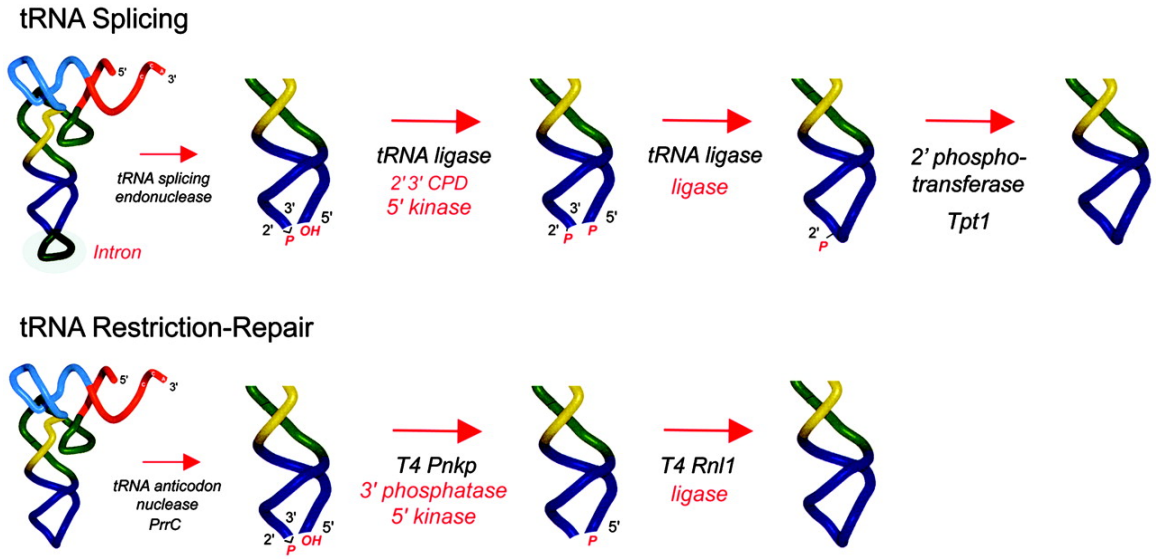


Figure 1.12. *Yeast tRNA-splicing and phage tRNA-restriction-repair pathways. In tRNA splicing, the pre-tRNA is cleaved at the exon–intron junctions in the anticodon loop by a tRNA-splicing endonuclease, which leaves a 2',3'-cyclic phosphate end on the proximal half-molecule and a 5'-OH on the distal half-molecule. The ends are then remodeled and sealed by tRNA ligase (Trl1), a multifunctional protein with 2',3'-CPD, 5' kinase, and ligase activities. The residual 2'-PO_4 at the splice junction is then removed by the* ***NAD^+-dependent 2'-phosphotransferase (Tpt1).*** *In tRNA-restriction repair, the mature tRNALys is cleaved in the anticodon loop by PrrC, which leaves 2',3'-cyclic phosphate and 5'-OH ends. The ends are then healed by T4 Pnkp, which removes the phosphate at the 3' end and phosphorylates the 5' terminus. T4 Rnl1 then seals the 3'-OH and 5'-PO_4 termini to form a standard 3'-5' phosphodiester linkage and thus restore tRNAL ys function in protein synthesis. Figure adapted with permission from reference 105.*

There are two types of 2'-OH junction modifications that can occur during, or be coupled to, enzymatic RNA ligation. In the case of the Hen1-Pnkp RNA repair system found in diverse bacteria, the Hen1 methyltransferase installs a ribose 2'-OCH_3 mark at the

3' terminal nucleoside of the RNA break, such that subsequent ATP-dependent 3'-OH/5′-PO_4 RNA ligation embeds the 2'-OCH_3 mark at the splice junction[106-109]. The advantage of this system is that the 2'-OCH_3 modification immunizes the repaired RNA against iterative cleavage by a transesterifying endoribonuclease.

The second type of junction modification pertains to the tRNA splicing system of fungi and plants, wherein a multidomain tRNA ligase catalyzes three reactions that heal and seal the 2',3'-cyclic-PO_4 and 5'-OH ends of the tRNA exons[110, 111]. In the healing phase of fungal tRNA splicing, the 5'-OH end is converted by a polynucleotide kinase to a 5'-PO_4, and the 2',3'-cyclic-PO_4 is hydrolyzed to a 3'-OH, 2'-PO_4 end by a cyclic phosphodiesterase (CPDase) (**Figure 1.13**). An ATP-dependent RNA ligase then joins these ends to form a spliced tRNA containing a 2'-PO_4, 3'-5' phosphodiester at the splice junction. The fungal/plant RNA ligase is unique among polynucleotide sealing enzymes in that ligation is strictly dependent on the 2'-PO_4 moiety of the 3'-OH strand[112]. Discrimination of 2′-PO_4 versus 2'-OH ends provides a quality control checkpoint to ensure that only the broken RNAs that have undergone healing by the CPDase can be sealed. A potential disadvantage of this requirement is that the 2'-PO_4 at the tRNA splice junction in the anticodon loop precludes proper tRNA function in codon recognition during translation[113]. Thus, it is necessary to remove the junction 2'-PO_4 as the final step in fungal tRNA splicing, a reaction performed by a dedicated tRNA 2′-phosphotransferase enzyme named Tpt1 that was discovered and characterized by the laboratory of Eric Phizicky (**Figure 1.13**)[114, 115].

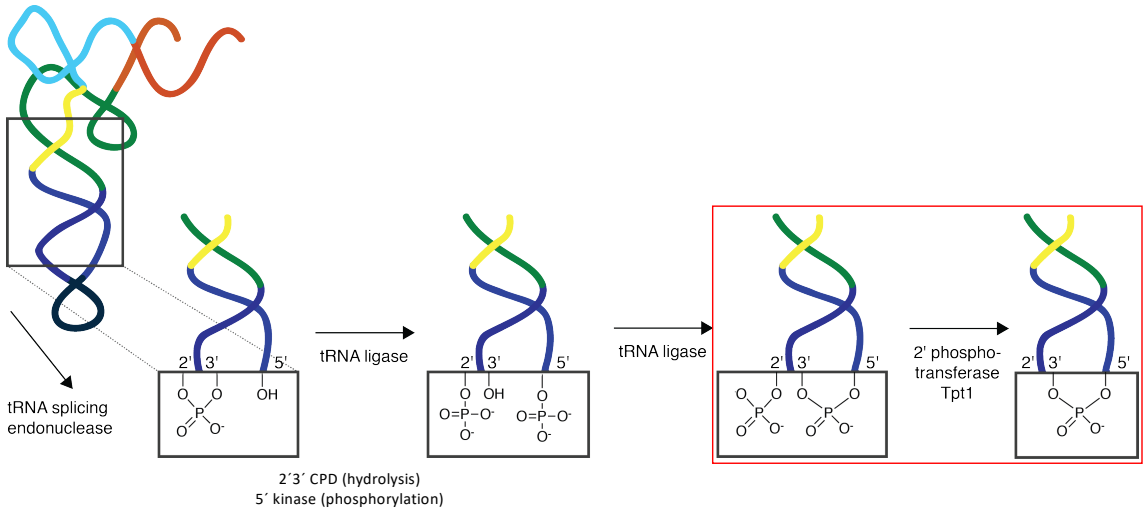


***Figure 1.13**. Mechanism of repair in tRNA splicing system of fungi and plants at the molecular level, with Tpt1 step indicated in red. Image adapted with permission from reference 105.*

Tpt1, which is essential for growth of the budding yeast *Saccharomyces cerevisiae*[116], catalyzes the transfer of the tRNA 2'-PO_4 to NAD^+ to form ADP-ribose 1″-2″ cyclic phosphate and nicotinamide[117]. The Tpt1 mechanism entails two chemical steps[118, 119]. First, NAD^+ reacts with the tRNA 2'-phosphate to expel nicotinamide and generate a 2' -phospho-ADP-ribosylated RNA intermediate (**Figure 1.14**). Then, transesterification of the ADP-ribose 2'-OH to the tRNA 2′-phosphate displaces the tRNA product and generates ADP-ribose 1″-2″ cyclic phosphate. This mechanism is expanded upon further in Chapters 2 and 3. Tpt1 exemplifies a family of structurally homologous proteins distributed widely in eukaryal, archaeal, and bacterial proteomes[120]. Because Tpt1 homologs are found in bacterial species (e.g., *E. coli*) that have no known intron-containing tRNAs and no known pathways to generate RNAs with internal 2'-PO_4 modifications, it is possible that members of the Tpt1 family from such taxa may dephosphorylate other (non-RNA) substrates in vivo via a shared mechanism of phosphoryl transfer to NAD^+ to form ADP-ribose 1″-2″ cyclic phosphate. However, it is certainly not the case that such Tpt1 homologs are specific for non-RNA substrates, insofar as the *E. coli* homolog (named KptA) can perform the same NAD^+-dependent RNA 2'-phosphoryl transfer reactions *in vitro* as yeast Tpt1[118, 121] and can complement an otherwise lethal *S. cerevisiae* tpt1Δ mutation in vivo[120, 122].

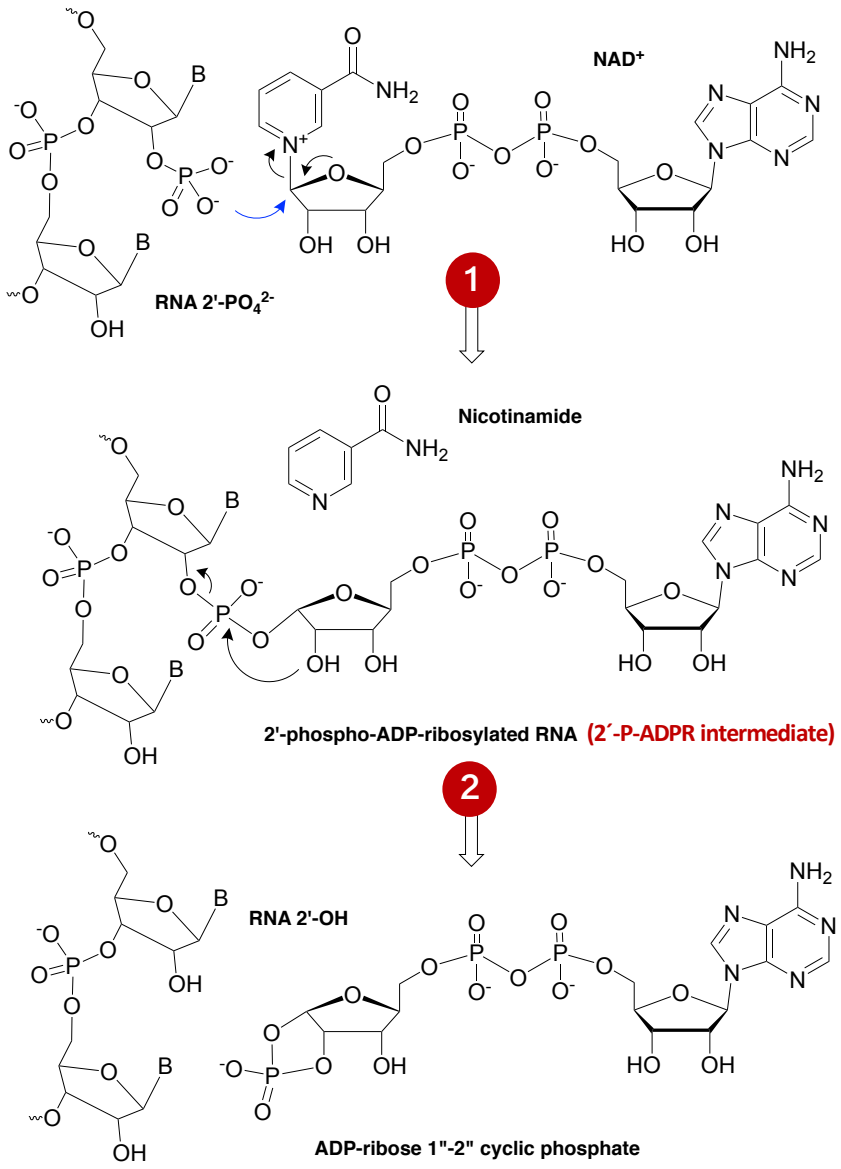


Figure 1.14. *Two step mechanism of action of Tpt1.* [117-119]

1.7.1 Tpt1 as an attractive antifungal agent

We view Tpt1 as a promising target for discovery of antifungals, based on the fact that the enzymatic mechanism of tRNA splicing in metazoan is entirely different from that of fungi and does not result in a junction 2'-PO_4[99, 104]. Whereas mammals do have Tpt1, it plays no essential role in mammalian physiology, insofar as a *tpt1*-KO mouse develops normally and has no defects in protein synthesis[123]. We hypothesize that inhibitors of Tpt1 should serve as highly selective antifungal agents. Chapters 1 and 2 of this thesis are

dedicated to the discovery of underlying mechanisms of Tpt1 enzymes as well as our efforts in designing and synthesizing oligonucleotide substrates for probing enzyme activity.

1.8 Gene Editing and CRISPR

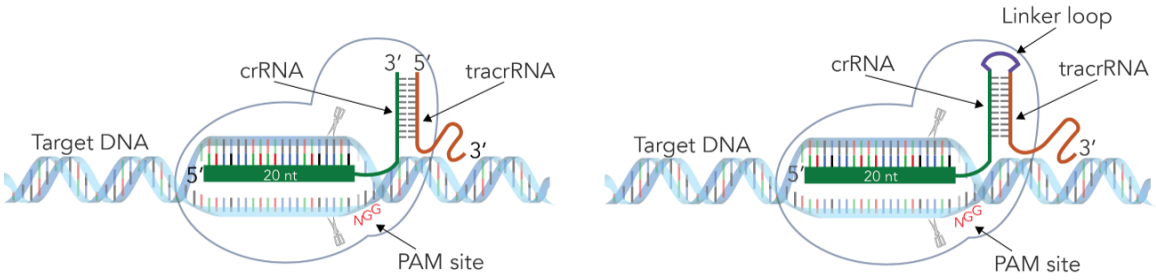
1.8.1 CRISPR technologies

CRISPR (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats) is a method of gene editing that uses proteins and specific guide RNAs to edit host genes. CRISPR was initially discovered in bacteria (*Streptococcus pyogenes*) and characterized by Francisco Mojica as an adaptive immunity response mechanism against viruses[124]. In 2012, Emmanuelle Charpentier and Jennifer Doudna then studied the system further and learned to repurpose it to serve a new role in genome editing[125]. Since then, CRISPR technology has witnessed tremendous growth, even earning Charpentier and Doudna the Nobel Prize in Chemistry in 2020. This growth is due to the ease of use and efficiency of the system, which have allowed it to be rapidly adopted by laboratories across the world for various applications.

1.8.2 CRISPR-Cas9

The CRISPR-associated protein (Cas9) is an endonuclease that uses two RNA molecules, trans-activating crRNA (tracrRNA, pronounced "tracer RNA") and CRISPR RNA (crRNA), which interact with each other to form a functional guide RNA (gRNA) (**Figure 1.15**). Cas9 uses the tracrRNA portion of the guide as a handle, while the crRNA spacer sequence directs the complex to a matching DNA sequence. For recognition and binding to target DNA, Cas9 requires the protospacer adjacent motif (PAM), as a short conserved sequence located just downstream of the non-complementary strand of the target dsDNA[126]. Recognition of the PAM (5′-NGG-3′) triggers dsDNA melting, enabling crRNA strand invasion and base pairing, and ultimately enabling Cas9 to introduce a site-specific double-strand break in the DNA (**Figure 1.15**A). The cleaved DNA is then repaired by non-homologous end joining (NHEJ) or homologous recombination, resulting in a modified gene sequence[127]. The dual tracrRNA:crRNA was later engineered as a single guide RNA (sgRNA) that retains two critical features: a sequence at the 5′ end that determines the DNA target site by base-pairing and a duplex RNA structure at the 3′ side

that binds to Cas9 (**Figure 1.15**B)[125, 128]. This finding created a simple two-component system in which changes in the guide sequence of the sgRNA program Cas9 to target any DNA sequence of interest.



A. 2-part crRNA:tracrRNA complex B. Single fusion sgRNA trigger

Figure 1.15. *Types of CRISPR-Cas9 guide RNA. (A) Wild-type S. pyogenes CRISPR systems use separate crRNA (targets DNA) and tracrRNA (binds Cas protein). (B) Single-guide RNA (sgRNA)is a result of joining the crRNA and tracrRNA into a fusion sgRNA by adding a short linker sequence between the two components. This figure is property of IDT Technologies and is used with permission.*

1.8.3 CRISPR-Cas12a *vs* Cas9

Until recently, the majority of CRISPR genome editing has been carried out using the Cas9 system. In recent years, research has begun to examine the potential of other systems for gene editing, such as CRISPR Cas12a. Cas12a has its own characteristic features which make it an ideal alternative for the Cas9 system. For example, Cas12a makes staggered cuts whereas Cas9 makes blunt cuts in the genome[129]. Cas9 makes a double stranded break (DSB) at the target site by cutting each strand in opposite directions, at the same position (**Figure 1.16**A). The nuclease active site of Cas12a cuts the target strand of the DNA in *cis* and the non-target strand of DNA in *trans* position (**Figure 1.16**B)[130]. This nuclease can only embed one DNA strand at a time, so the target and non-target DNA strands are presumably cleaved sequentially. This sequential cleavage of DNA elucidates the mechanism of staggered-end DNA break induced by Cas12a.

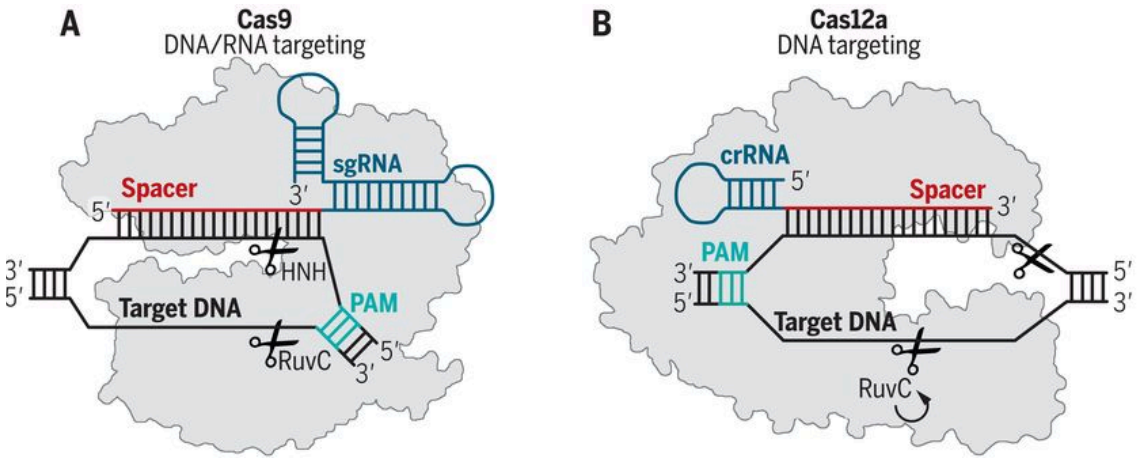


Figure 1.16. *(A) CRISPR-Cas9 with a sgRNA (blue) encoding a spacer/guide (red) bound to target dsDNA (black) upstream to a PAM (teal). CRISPR-Cas12a with crRNA (blue) encoding a spacer/guide (red) bound to a complememtary dsDNA (black) downstream to a PAM (teal).*

Cas12a functions using only a single, 39-42 nucleotide long crRNA. This greatly simplifies the chemical synthesis of Cas12a guide RNA and is one of the main reasons we looked to study this system. Another reason is the fact that Cas9 needs a downstream (3′) NGG protospacer-adjacent motif (PAM) to recognize and bind to DNA, while Cas12a requires an upstream (5′) TTN (**Figure 1.16**)[129]. Consequently, TT is slightly more common than GG in the human genome, which means Cas12a can appeal to a larger library of possible gene targets through recognition of a T-rich PAM. Cas12a has a lower tolerance for mismatches in the spacer/guide of the crRNA and thus is found to be more efficient and highly specific compared to Cas9, with very little off-target cleavage capacity.

In this thesis we will study *Acidominococcus species* Cas12a (*As*Cas12a). The structure of AsCas12a crRNA reveals 20 nt stem (5′ handle) sequence and a spacer (guide segment) sequence of 23 nt in length (**Figure 1.17**). The 5′ handle adopts a pseudoknot structure consisting of a stem and a loop region. The pseudoknot, starting from -1 to -20 bases, consists of five Watson-Crick base pairs, one noncanonical U–U base pair, one UCUU tetraloop, one reverse Hoogsteen A–U base pair and three 5′-end bases (**Figure 1.17**). The hydrogen bonds formed within stem and loop regions stabilizes the pseudoknot structure. The bases U (-1), U (-10), U (-16), and A (-18) are conserved across Cas12a homologs indicating formation of similar tetraloop pseudoknot is crucial for the efficiency of endonuclease activity of Cas12a.

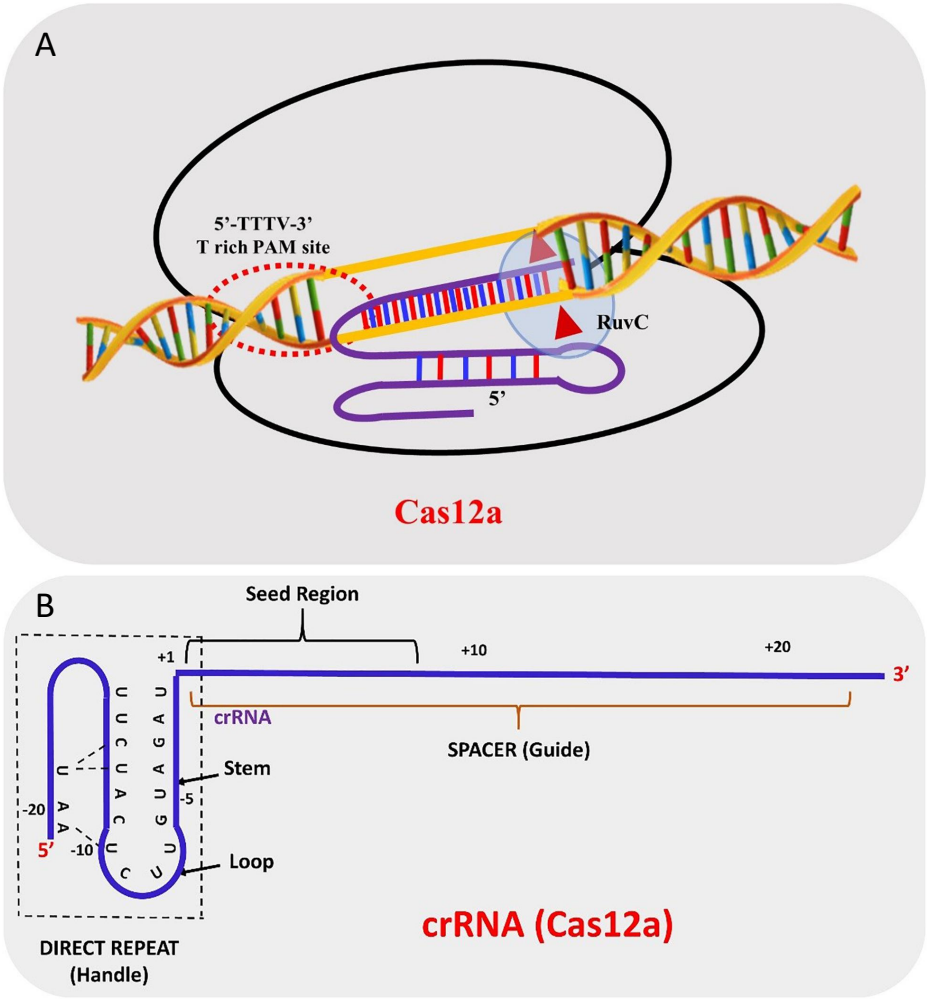


Figure 1.17. *(A) Schematic representation of Cas12a crRNA with the target strand DNA association. (B) Schematic representation of mature crRNA derived from the maturation of pre-crRNA. Figure adapted with permission from reference* [131].

1.8.4 Differential *cis* and *trans* cleavage activity of Cas12a

A unique property of Cas12a is *trans* activity, which is an unlocking of non-sequence-specific single-stranded DNA cleavage (ssDNase) activity with high catalytic turnover[132, 133]. *Trans* activity requires that the Cas12a RNA first bind to and cut a target DNA strand, which is referred to as *cis* activity (**Figure 1. 18**). *Trans* activity has been exploited to create molecular diagnostics that can detect very small amounts of DNA and amplify the detection signal through cleavage of ssDNA substrates that can be detected, such as by fluorescence, antibodies on test strips, and nanopores[132, 134, 135]. The digestion of the ssDNA reporter by *trans* activity is intrinsically tied to the *cis* cleavage of the specific target DNA because both mechanisms use the same catalytic residues[136, 137]. Mutation of

neighboring amino acids or REC (recognition lobe) linker and lid regions of Cas12a, however, could allow *trans* but not *cis* activity when Cas12a was bound to a complementary single-strand DNA target[136]. However, it appears that *cis* and *trans* activity cannot fully be uncoupled for recognition of double-stranded targets *via trans* activity without *cis* cleavage of the target strand[136, 137].

1.8.4.1 CRISPR-Cas12a as a diagnostic tool

The need for time- and cost-effective nucleic acid detection methods is ever increasing in human genotyping and pathogen detection. CRISPR-Cas12a applications for the diagnosis of invading nucleic acids have been reported. Li and colleagues reported that Cas12a performs collateral cleavage of non-target ssDNAs upon formation of the Cas12a/crRNA/target DNA ternary complex[133]. In addition to cutting dsDNA in a crRNA-dependent manner (*cis*), Cas12a can also exert its ability over ssDNA substrates (*trans*) (**Figure 1. 18**). This non-specific cleavage is only triggered once Cas12a has bound to and cleaved its complementary target and allows Cas12a to attack other invading nucleic acids in the vicinity by collateral damaging.

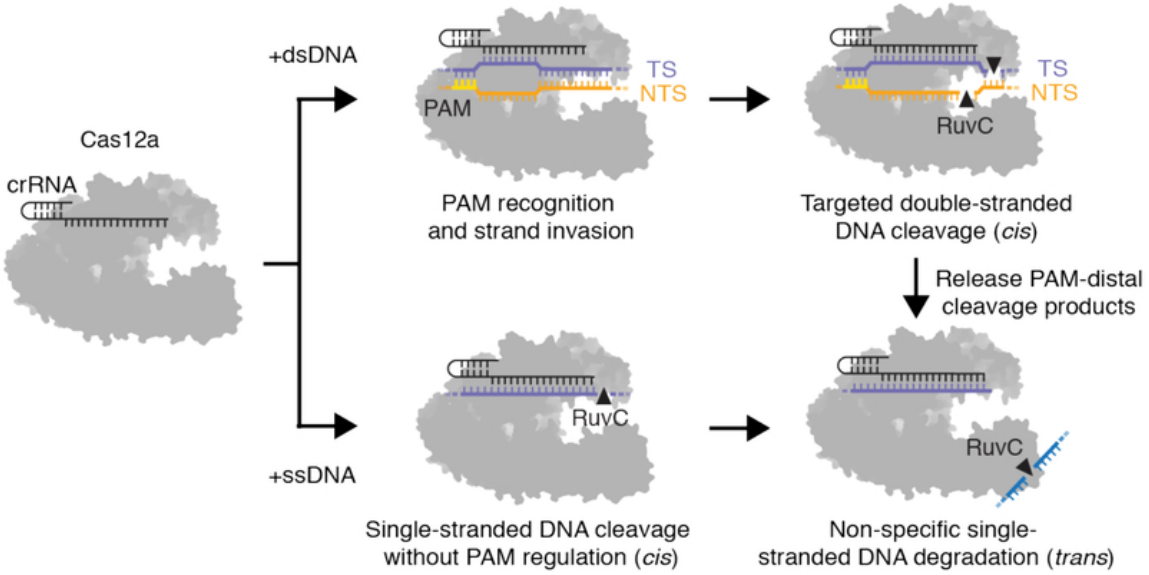


Figure 1. 18. *Model for PAM-dependent and PAM-independent activation of cis and trans-cleavage by Cas12a. The Cas12a-crRNA complex binds to a complementary dsDNA in a PAM-dependent manner (top) or ssDNA in a PAM-independent manner (bottom), which is sufficient to unleash indiscriminate ssDNase activity by RuvC nuclease. Cas12a can also release its PAM-distal cleavage products, which exposes the RuvC active site for multiple rounds of non-specific ssDNA degradation. Figure adapted with permission from reference [132].*

With promising applications in the creation of model systems, therapeutic gene editing, and diagnostics, the CRISPR-Cas12a system has leapt to the forefront of gene

editing therapeutics. These applications could benefit from modifications that can enhance guide RNA nuclease resistance or tune enzyme properties. For Cas12a, attempts to fully modify the crRNA guide have proven unsuccessful[138-140]. However, the 5' pseudoknot handle posed the greatest challenge as it appeared to tolerate very little modification to the ribose, suggesting that successful modification of this unique structure would be the major limitation to greater modification of Cas12a guides for therapeutic development. The 5' pseudoknot handle of the Cas12a system is the subject of our studies in Chapters 4 and 5 and is described in further detail therein.

1.9 Thesis Objectives

In this thesis, we investigate 2'-modified nucleic acids to probe two understudied enzymes, namely Tpt1 and Cas12a. Specifically, we focus on the design, synthesis, and activity of 2' modified oligonucleotides to gain mechanistic insights into the function of these enzymes. This endeavor allowed us to design and identify oligonucleotides substrates that can be further pursued for therapeutic drug development.

In Chapter 2, we introduce the Tpt1 enzyme, in the context of RNA repair enzymes. We probe the substrate specificity of this enzyme by exposing it to various synthetic RNA analogues that mimic the branch structure of tRNA intermediates. A key contribution was to develop a method to chemically synthesize various RNA oligonucleotides with an internal 2'-phosphate branch for biochemical studies of exemplary Tpt1 enzymes. Studies by the Shuman lab (Memorial Sloan Kettering Cancer Center) using our RNA molecules, including transient-state kinetic analysis of *Runella* Tpt1, and mutants thereof, established a two-step mechanism and identified active site amino acids that act specifically as catalysts of the second step of the Tpt1 pathway[141]. Our synthetic substrates also aided in further studies of the *Clostridium* Tpt1 enzyme that led to the determination of its crystal structure in a product mimetic state[142].

In Chapter 3, we extend this study beyond substrate mimics, and work towards the development of potentially therapeutic substrates for Tpt1, through the incorporation of modifications of the RNA residues flanking the 2'-phosphate, as well as modifications of the 2'-phosphate itself. Specifically, we present the synthesis of a series of RNA substrate analogs with modified sugars (2'-deoxyribose, 2'-fluororibose, 2'-OMe ribose, arabinose,

2'F-arabinose) and phosphorothioate (PS) linkages flanking the 2'-PO_4 junction[143]. In an attempt to trap RNA-2'-phospho-ADP-ribose intermediates, we present our efforts to synthesize modified RNAs with 2'-phosphonate and 2'-phosphoramidate groups.

Chapters 4 and 5 are dedicated to our efforts in understanding the CRISPR-Cas12a gene editing system. Chapter 4 builds upon an ongoing collaborative effort between our lab and that of Prof. Keith Gagnon (Southern Illinois U), in which we systematically incorporate modifications to crRNA to characterize structure-activity relationships within CRISPR-Cas12 and identify which modifications are compatible with editing activity in cells. As well, we probe the importance of putative 2'-hydroxyl contacts between the crRNA and Cas12 and Cas9 using a series of crRNAs that are modified at the predicted 2'-OH contact positions.

In Chapter 5, we introduce two isomeric forms of RNA, namely arabinonucleic acid (ANA) and 2',5'-RNA, into an already modified cRNAs to further probe chemical and structural requirements of *Acidominococcus species* Cas12a (*As*Cas12a). Furthermore, native gel electrophoresis, UV-melting experiments and circular dichroism studies (CD) were carried out to assess the secondary structure of the non-canonical pseudoknot structure that mediates conserved binding to Cas12a.

The rules for successful chemical modification of 2'-P-RNA and crRNAs described in this thesis should accelerate therapeutic development and may be valuable for CRISPR-Cas gene editing and diagnostic applications.

1.10 References

1. Dahm, R., Friedrich Miescher and the discovery of DNA. *Dev Biol* **2005,** *278* (2), 274-88.
2. Dahm, R., Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **2008,** *122* (6), 565-81.
3. Lamm, E.; Harman, O.; Veigl, S. J., Before Watson and Crick in 1953 Came Friedrich Miescher in 1869. *Genetics* **2020,** *215* (2), 291-296.
4. Levene, P. A.; London, E. J., On the Structure of Thymonucleic Acid. *Science* **1928,** *68* (1771), 572-3.
5. Avery, O. T.; Macleod, C. M.; McCarty, M., Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* **1944,** *79* (2), 137-58.
6. Chargaff, E., Structure and function of nucleic acids as cell constituents. *Fed Proc* **1951,** *10* (3), 654-9.
7. Chargaff, E.; Vischer, E.; et al., The composition of the desoxypentose nucleic acids of thymus and spleen. *J Biol Chem* **1949,** *177* (1), 405-16.
8. Franklin, R. E.; Gosling, R. G., Molecular configuration in sodium thymonucleate. *Nature* **1953,** *171* (4356), 740-1.
9. Franklin, R. E.; Gosling, R. G., Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **1953,** *172* (4369), 156-7.
10. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953,** *171* (4356), 737-8.
11. Watson, J. D.; Crick, F. H. C., Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953,** *171* (4361), 964-967.
12. Altona, C.; Sundaralingam, M., Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *Journal of the American Chemical Society* **1972,** *94* (23), 8205-8212.
13. Rich, A., The double helix: a tale of two puckers. *Nature Structural & Molecular Biology* **2003,** *10* (4), 247-249.
14. Guschlbauer, W.; Jankowski, K., Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Research* **1980,** *8* (6), 1421-1433.
15. Hunter, C. A., Sequence-dependent DNA structure. The role of base stacking interactions. *J Mol Biol* **1993,** *230* (3), 1025-54.
16. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X. F.; Sheils, C. J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T., Experimental Measurement of Aromatic Stacking

Affinities in the Context of Duplex DNA. *Journal of the American Chemical Society* **1996,** *118* (34), 8182-8183.

17. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T., Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J Am Chem Soc* **2000,** *122* (10), 2213-2222.
18. Kool, E. T., Hydrogen bonding, base stacking, and steric effects in dna replication. *Annu Rev Biophys Biomol Struct* **2001,** *30*, 1-22.
19. Sponer, J.; Leszczynski, J.; Hobza, P., Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **2001,** *61* (1), 3-31.
20. Manning, G. S., The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* **1978,** *11* (2), 179-246.
21. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci U S A* **1981,** *78* (4), 2179-83.
22. Ghosh, A.; Bansal, M., A glossary of DNA structures from A to Z. *Acta Crystallogr D Biol Crystallogr* **2003,** *59* (Pt 4), 620-6.
23. Wang, A. H.; Fujii, S.; van Boom, J. H.; Rich, A., Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc Natl Acad Sci U S A* **1982,** *79* (13), 3968-72.
24. Arnott, S.; Chandrasekaran, R.; Millane, R. P.; Park, H. S., RNA-RNA, DNA-DNA, and DNA-RNA Polymorphism. *Biophys J* **1986,** *49* (1), 3-5.
25. Saenger, W.; Hunter, W. N.; Kennard, O., DNA conformation is determined by economics in the hydration of phosphate groups. *Nature* **1986,** *324* (6095), 385-8.
26. Yanagi, K.; Privé, G. G.; Dickerson, R. E., Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J Mol Biol* **1991,** *217* (1), 201-14.
27. Chapter 2 DNA and RNA Structure. In *Nucleic Acids in Chemistry and Biology (3)*, The Royal Society of Chemistry: 2006; pp 13-76.
28. Crick, F., Central dogma of molecular biology. *Nature* **1970,** *227* (5258), 561-3.
29. Bessman, M. J.; Kornberg, A.; Lehman, I. R.; Simms, E. S., Enzymic synthesis of deoxyribonucleic acid. *Biochim Biophys Acta* **1956,** *21* (1), 197-8.
30. Weiss, S. B.; Nakamoto, T., Net Synthesis of Ribonucleic Acid with a Microbial Enzyme Requiring Deoxyribonucleic Acid and Four Ribonucleoside Triphosphates. *Journal of Biological Chemistry* **1961,** *236* (3), PC18-PC20.
31. Chow, L. T.; Gelinas, R. E.; Broker, T. R.; Roberts, R. J., An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **1977,** *12* (1), 1-8.
32. Jacob, F.; Monod, J., Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **1961,** *3*, 318-56.

33. Furuichi, Y.; Muthukrishnan, S.; Shatkin, A. J., 5'-Terminal m-7G(5')ppp(5')G-m-p in vivo: identification in reovirus genome RNA. *Proc Natl Acad Sci U S A* **1975,** *72* (2), 742-5.

34. Philipson, L.; Wall, R.; Glickman, G.; Darnell, J. E., Addition of polyadenylate sequences to virus-specific RNA during adenovirus replication. *Proc Natl Acad Sci U S A* **1971,** *68* (11), 2806-9.

35. Brody, E.; Abelson, J., The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **1985,** *228* (4702), 963-7.

36. Palade, G. E., A small particulate component of the cytoplasm. *J Biophys Biochem Cytol* **1955,** *1* (1), 59-68.

37. Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A., The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **2000,** *289* (5481), 920-930.

38. Hoagland, M. B.; Stephenson, M. L.; Scott, J. F.; Hecht, L. I.; Zamecnik, P. C., A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* **1958,** *231* (1), 241-57.

39. Hayes, D. H.; Michelson, A. M.; Todd, A. R., Nucleotides. Part XXX. Mononucleotides derived from deoxyadenosine and deoxyguanosine. *Journal of the Chemical Society (Resumed)* **1955,** (0), 808-815.

40. Alvarado-Urbina, G.; Sathe, G. M.; Liu, W. C.; Gillen, M. F.; Duck, P. D.; Bender, R.; Ogilvie, K. K., Automated synthesis of gene fragments. *Science* **1981,** *214* (4518), 270-4.

41. Letsinger, R. L.; Ogilvie, K. K., Convenient method for stepwise synthesis of oligothymidylate derivatives in large-scale quantities. *Journal of the American Chemical Society* **1967,** *89* (18), 4801-4803.

42. Beaucage, S. L.; Caruthers, M. H., Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* **1981,** *22* (20), 1859-1862.

43. Ogilvie, K. K.; Schifman, A. L.; Penney, C. L., The synthesis of oligoribonucleotides. III. The use of silyl protecting groups in nucleoside and nucleotide chemistry. VIII. *Canadian Journal of Chemistry* **1979,** *57* (17), 2230-2238.

44. Damha, M. J.; Giannaris, P. A.; Zabarylo, S. V., An improved procedure for derivatization of controlled-pore glass beads for solid-phase oligonucleotide synthesis. *Nucleic Acids Res* **1990,** *18* (13), 3813-21.

45. Pon, R. T.; Damha, M. J.; Ogilvie, K. K., Modification of guanine bases by nucleoside phosphoramidite reagents during the solid phase synthesis of oligonucleotides. *Nucleic Acids Res* **1985,** *13* (18), 6447-65.

46. Watts, J. K.; Katolik, A.; Viladoms, J.; Damha, M. J., Studies on the hydrolytic stability of 2'-fluoroarabinonucleic acid (2'F-ANA). *Org Biomol Chem* **2009,** *7* (9), 1904-10.

47. Ogilvie, K. K., The tert-Butyldimethylsilyl Group as a Protecting Group in Deoxynucleosides. *Canadian Journal of Chemistry* **1973,** *51* (22), 3799-3807.

48. Ogilvie, K. K., Isomerization of tert-butyldimethylsilyl protecting groups in ribonucleosides. **1981**.

49. Stephenson, M. L.; Zamecnik, P. C., Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc Natl Acad Sci U S A* **1978,** *75* (1), 285-8.

50. Aartsma-Rus, A.; Corey, D. R., The 10th Oligonucleotide Therapy Approved: Golodirsen for Duchenne Muscular Dystrophy. *Nucleic Acid Ther* **2020,** *30* (2), 67-70.

51. Deleavey, G.; Damha, M. J., Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing. *Chemistry & biology* **2012,** *19*, 937-54.

52. Bennett, C. F.; Swayze, E. E., RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu Rev Pharmacol Toxicol* **2010,** *50*, 259-93.

53. Stein, H.; Hausen, P., Enzyme from Calf Thymus Degrading the RNA Moiety of DNA-RNA Hybrids: Effect on DNA-Dependent RNA Polymerase. *Science* **1969,** *166* (3903), 393-395.

54. Watts, J. K.; Corey, D. R., Silencing disease genes in the laboratory and the clinic. *J Pathol* **2012,** *226* (2), 365-79.

55. Roberts, T. C.; Langer, R.; Wood, M. J. A., Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery* **2020,** *19* (10), 673-694.

56. Dominski, Z.; Kole, R., Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc Natl Acad Sci U S A* **1993,** *90* (18), 8673-7.

57. Singh, R. N.; Singh, N. N., Mechanism of Splicing Regulation of Spinal Muscular Atrophy Genes. *Adv Neurobiol* **2018,** *20*, 31-61.

58. Aartsma-Rus, A.; Straub, V.; Hemmings, R.; Haas, M.; Schlosser-Weber, G.; Stoyanova-Beninska, V.; Mercuri, E.; Muntoni, F.; Sepodes, B.; Vroom, E.; Balabanov, P., Development of Exon Skipping Therapies for Duchenne Muscular Dystrophy: A Critical Review and a Perspective on the Outstanding Issues. *Nucleic Acid Ther* **2017,** *27* (5), 251-259.

59. Wan, L.; Dreyfuss, G., Splicing-Correcting Therapy for SMA. *Cell* **2017,** *170* (1), 5.

60. Ward, A. J.; Norrbom, M.; Chun, S.; Bennett, C. F.; Rigo, F., Nonsense-mediated decay as a terminating mechanism for antisense oligonucleotides. *Nucleic Acids Res* **2014,** *42* (9), 5871-9.

61. Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C., Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **1998,** *391* (6669), 806-11.

62. Elbashir, S. M.; Harborth, J.; Lendeckel, W.; Yalcin, A.; Weber, K.; Tuschl, T., Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **2001,** *411* (6836), 494-498.

63. Liu, J.; Carmell, M. A.; Rivas, F. V.; Marsden, C. G.; Thomson, J. M.; Song, J. J.; Hammond, S. M.; Joshua-Tor, L.; Hannon, G. J., Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **2004,** *305* (5689), 1437-41.

64. Roberts, T. C., The microRNA Machinery. *Adv Exp Med Biol* **2015,** *887*, 15-30.

65. Schürmann, N.; Trabuco, L. G.; Bender, C.; Russell, R. B.; Grimm, D., Molecular dissection of human Argonaute proteins by DNA shuffling. *Nat Struct Mol Biol* **2013,** *20* (7), 818-26.

66. Lam, J. K. W.; Chow, M. Y. T.; Zhang, Y.; Leung, S. W. S., siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy - Nucleic Acids* **2015,** *4*.

67. Geary, R. S., Antisense oligonucleotide pharmacokinetics and metabolism. *Expert Opin Drug Metab Toxicol* **2009,** *5* (4), 381-91.

68. Marques, J. T.; Williams, B. R., Activation of the mammalian immune system by siRNAs. *Nat Biotechnol* **2005,** *23* (11), 1399-405.

69. Agrawal, S.; Gait, M. J., History and Development of Nucleotide Analogues in Nucleic Acids Drugs. *Advances in Nucleic Acid Therapeutics, RSC,* **2019,** 1-21.

70. Khvorova, A.; Watts, J. K., The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol* **2017,** *35* (3), 238-248.

71. Eckstein, F., A dinucleoside phosphorothioate. *Tetrahedron Letters* **1967,** *8* (13), 1157-1160.

72. Eckstein, F., Nucleoside phosphorothioates. *Annu Rev Biochem* **1985,** *54*, 367-402.

73. Eckstein, F., Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Acid Ther* **2014,** *24* (6), 374-87.

74. Eckstein, F., Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev* **2000,** *10* (2), 117-21.

75. Crooke, S. T.; Vickers, T. A.; Liang, X.-h., Phosphorothioate modified oligonucleotide–protein interactions. *Nucleic Acids Research* **2020,** *48* (10), 5235-5253.

76. Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A., Applications of Fluorine in Medicinal Chemistry. *J Med Chem* **2015,** *58* (21), 8315-59.

77. Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V., Fluorine in medicinal chemistry. *Chemical Society Reviews* **2008,** *37* (2), 320-330.

78. Kois, P.; Tocik, Z.; Spassova, M.; Ren, W.-Y.; Rosenberg, I.; Soler, J. F.; Watanabe, K. A., Synthesis and Some Properties of Modified Oligonucleotides. II. Oligonucleotides Containing 2′-Deoxy-2′-fluoro-β-D-arabinofuranosyl Pyrimidine Nucleosides. *Nucleosides and Nucleotides* **1993,** *12* (10), 1093-1109.

79. El-Khoury, R.; Damha, M. J., 2′-Fluoro-arabinonucleic Acid (FANA): A Versatile Tool for Probing Biomolecular Interactions. *Accounts of Chemical Research* **2021,** *54* (9), 2287-2297.

80. Wilds, C. J.; Damha, M. J., 2'-Deoxy-2'-fluoro-beta-D-arabinonucleosides and oligonucleotides (2'F-ANA): synthesis and physicochemical studies. *Nucleic Acids Res* **2000,** *28* (18), 3625-35.

81. Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M., Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymines: implications of conformational preorganization for duplex stability. *Nucleic Acids Res* **1998,** *26* (10), 2473-80.

82. Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J., Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing. *Nucleic Acids Research* **2010,** *38* (13), 4547-4557.

83. Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J., Differential stability of 2'F-ANA*RNA and ANA*RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res* **2010,** *38* (7), 2498-511.

84. Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; González, C., The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res* **2012,** *40* (18), 9329-39.

85. Wan, W. B.; Seth, P. P., The Medicinal Chemistry of Therapeutic Oligonucleotides. *Journal of Medicinal Chemistry* **2016,** *59* (21), 9645-9667.

86. Egli, M.; Minasov, G.; Tereshko, V.; Pallan, P. S.; Teplova, M.; Inamati, G. B.; Lesnik, E. A.; Owens, S. R.; Ross, B. S.; Prakash, T. P.; Manoharan, M., Probing the influence of stereoelectronic effects on the biophysical properties of oligonucleotides: comprehensive analysis of the RNA affinity, nuclease resistance, and crystal structure of ten 2'-O-ribonucleic acid modifications. *Biochemistry* **2005,** *44* (25), 9045-57.

87. Sørensen, M. D.; Kvaernø, L.; Bryld, T.; Håkansson, A. E.; Verbeure, B.; Gaubert, G.; Herdewijn, P.; Wengel, J., alpha-L-ribo-configured locked nucleic acid (alpha-L-LNA): synthesis and properties. *J Am Chem Soc* **2002,** *124* (10), 2164-76.

88. Nielsen, K. M. E.; Petersen, M.; Håkansson, A. E.; Wengel, J.; Jacobsen, J. P., α-L-LNA (α-L-ribo Configured Locked Nucleic Acid) Recognition of DNA: An NMR Spectroscopic Study. *Chemistry – A European Journal* **2002,** *8* (13), 3001-3009.

89. Langkjaer, N.; Pasternak, A.; Wengel, J., UNA (unlocked nucleic acid): a flexible RNA mimic that allows engineering of nucleic acid duplex stability. *Bioorg Med Chem* **2009,** *17* (15), 5420-5.

90. Campbell, M. A.; Wengel, J., Locked vs. unlocked nucleic acids (LNAvs.UNA): contrasting structures work towards common therapeutic goals. *Chemical Society Reviews* **2011,** *40* (12), 5680-5689.

91. Mangos, M. M.; Min, K. L.; Viazovkina, E.; Galarneau, A.; Elzagheid, M. I.; Parniak, M. A.; Damha, M. J., Efficient RNase H-directed cleavage of RNA promoted by antisense DNA or 2'F-ANA constructs containing acyclic nucleotide inserts. *J Am Chem Soc* **2003,** *125* (3), 654-61.

92. Giannaris, P. A.; Damha, M. J., Oligoribonucleotides containing 2',5'-phosphodiester linkages exhibit binding selectivity for 3',5'-RNA over 3',5'-ssDNA. *Nucleic Acids Res* **1993,** *21* (20), 4742-9.

93. Fluiter, K.; Mook, O. R.; Vreijling, J.; Langkjaer, N.; Højland, T.; Wengel, J.; Baas, F., Filling the gap in LNA antisense oligo gapmers: the effects of unlocked nucleic acid (UNA) and 4'-C-hydroxymethyl-DNA modifications on RNase H recruitment and efficacy of an LNA gapmer. *Mol Biosyst* **2009,** *5* (8), 838-43.

94. Wasner, M.; Arion, D.; Borkow, G.; Noronha, A.; Uddin, A. H.; Parniak, M. A.; Damha, M. J., Physicochemical and biochemical properties of 2',5'-linked RNA and 2',5'-RNA:3',5'-RNA "hybrid" duplexes. *Biochemistry* **1998,** *37* (20), 7478-86.

95. Habibian, M.; Harikrishna, S.; Fakhoury, J.; Barton, M.; Ageely, E. A.; Cencic, R.; Fakih, H. H.; Katolik, A.; Takahashi, M.; Rossi, J.; Pelletier, J.; Gagnon, K. T.; Pradeepkumar, P. I.; Damha, M. J., Effect of 2'-5'/3'-5' phosphodiester linkage heterogeneity on RNA interference. *Nucleic Acids Res* **2020,** *48* (9), 4643-4657.

96. Torrence, P. F.; Maitra, R. K.; Lesiak, K.; Khamnei, S.; Zhou, A.; Silverman, R. H., Targeting RNA for degradation with a (2'-5')oligoadenylate-antisense chimera. *Proceedings of the National Academy of Sciences* **1993,** *90* (4), 1300-1304.

97. Sabatino, D.; Damha, M. J., Oxepane Nucleic Acids: Synthesis, Characterization, and Properties of Oligonucleotides Bearing a Seven-Membered Carbohydrate Ring. *Journal of the American Chemical Society* **2007,** *129* (26), 8259-8270.

98. Habibian, M.; Martínez-Montero, S.; Portella, G.; Chua, Z.; Bohle, D. S.; Orozco, M.; Damha, M. J., Seven-Membered Ring Nucleoside Analogues: Stereoselective Synthesis and Studies on Their Conformational Properties. *Organic Letters* **2015,** *17* (21), 5416-5419.

99. Popow, J.; Schleiffer, A.; Martinez, J., Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences* **2012,** *69* (16), 2657-2670.

100. Das, U.; Shuman, S., Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res* **2013,** *41* (1), 355-65.

101. Wang, L. K.; Lima, C. D.; Shuman, S., Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *Embo j* **2002,** *21* (14), 3873-80.

102. Wang, L. K.; Ho, C. K.; Pei, Y.; Shuman, S., Mutational analysis of bacteriophage T4 RNA ligase 1. Different functional groups are required for the nucleotidyl transfer and phosphodiester bond formation steps of the ligation reaction. *J Biol Chem* **2003,** *278* (32), 29454-62.

103. Tanaka, N.; Chakravarty, A. K.; Maughan, B.; Shuman, S., Novel mechanism of RNA repair by RtcB via sequential 2',3'-cyclic phosphodiesterase and 3'-Phosphate/5'-hydroxyl ligation reactions. *J Biol Chem* **2011,** *286* (50), 43134-43.

104. Chakravarty, A. K.; Shuman, S., The sequential 2',3'-cyclic phosphodiesterase and 3'-phosphate/5'-OH ligation steps of the RtcB RNA splicing pathway are GTP-dependent. *Nucleic Acids Res* **2012,** *40* (17), 8558-67.

105. Schwer, B.; Sawaya, R.; Ho, C. K.; Shuman, S., Portability and fidelity of RNA-repair systems. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (9), 2788-2793.

106. Chan, C. M.; Zhou, C.; Huang, R. H., Reconstituting bacterial RNA repair and modification in vitro. *Science* **2009,** *326* (5950), 247.

107. Mui Chan, C.; Zhou, C.; Brunzelle, J. S.; Huang, R. H., Structural and biochemical insights into 2'-O-methylation at the 3'-terminal nucleotide of RNA by Hen1. *Proc Natl Acad Sci U S A* **2009,** *106* (42), 17699-704.

108. Jain, R.; Shuman, S., Active site mapping and substrate specificity of bacterial Hen1, a manganese-dependent 3' terminal RNA ribose 2'O-methyltransferase. *Rna* **2011,** *17* (3), 429-38.

109. Jain, R.; Shuman, S., Bacterial Hen1 is a 3' terminal RNA ribose 2'-O-methyltransferase component of a bacterial RNA repair cassette. *Rna* **2010,** *16* (2), 316-23.

110. Greer, C. L.; Peebles, C. L.; Gegenheimer, P.; Abelson, J., Mechanism of action of a yeast RNA ligase in tRNA splicing. *Cell* **1983,** *32* (2), 537-546.

111. Sawaya, R.; Schwer, B.; Shuman, S., Genetic and biochemical analysis of the functional domains of yeast tRNA ligase. *J Biol Chem* **2003,** *278* (45), 43928-38.

112. Remus, B. S.; Shuman, S., A kinetic framework for tRNA ligase and enforcement of a 2'-phosphate requirement for ligation highlights the design logic of an RNA repair machine. *RNA* **2013,** *19 5*, 659-69.

113. Spinelli, S. L.; Consaul, S. A.; Phizicky, E. M., A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction. *Rna* **1997,** *3* (12), 1388-400.

114. McCraith, S. M.; Phizicky, E. M., An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. *Journal of Biological Chemistry* **1991,** *266* (18), 11986-11992.

115. McCraith, S. M.; Phizicky, E. M., A highly specific phosphatase from Saccharomyces cerevisiae implicated in tRNA splicing. *Mol Cell Biol* **1990,** *10* (3), 1049-1055.

116. Culver, G. M.; McCraith, S. M.; Consaul, S. A.; Stanford, D. R.; Phizicky, E. M., A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. *J Biol Chem* **1997,** *272* (20), 13203-10.

117. Culver, G. M.; McCraith, S. M.; Zillmann, M.; Kierzek, R.; Michaud, N.; LaReau, R. D.; Turner, D. H.; Phizicky, E. M., An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate. *Science* **1993,** *261* (5118), 206.

118. Spinelli, S. L.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Transient ADP-ribosylation of a 2′-Phosphate Implicated in Its Removal from Ligated tRNA during Splicing in Yeast. **1999,** *274* (5), 2637-2644.

119. Steiger, M. A.; Jackman, J. E.; Phizicky, E. M., Analysis of 2'-phosphotransferase (Tpt1p) from Saccharomyces cerevisiae: evidence for a conserved two-step reaction mechanism. *RNA* **2005,** *11* (1), 99-106.

120. Spinelli, S. L.; Malik, H. S.; Consaul, S. A.; Phizicky, E. M., A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proceedings of the National Academy of Sciences* **1998,** *95* (24), 14136-14141.

121. Steiger, M. A.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Substrate recognition by a yeast 2'-phosphotransferase involved in tRNA splicing and by its Escherichia coli homolog. *Biochemistry* **2001,** *40* (46), 14098-105.

122. Sawaya, R.; Schwer, B.; Shuman, S., Structure-function analysis of the yeast NAD+-dependent tRNA 2'-phosphotransferase Tpt1. *RNA (New York, N.Y.)* **2005,** *11* (1), 107-113.

123. Harding, H. P.; Lackey, J. G.; Hsu, H.-C.; Zhang, Y.; Deng, J.; Xu, R.-M.; Damha, M. J.; Ron, D., An intact unfolded protein response in Trpt1 knockout mice reveals phylogenic divergence in pathways for RNA ligation. *RNA (New York, N.Y.)* **2008,** *14* (2), 225-232.

124. Mojica, F. J. M.; Ferrer, C.; Juez, G.; Rodríguez-Valera, F., Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular Microbiology* **1995,** *17* (1), 85-93.

125. 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.

126. Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **2014,** *507* (7490), 62-7.

127. Platt, F. M.; d'Azzo, A.; Davidson, B. L.; Neufeld, E. F.; Tifft, C. J., Lysosomal storage diseases. *Nat Rev Dis Primers* **2018,** *4* (1), 27.

128. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M., RNA-guided human genome engineering via Cas9. *Science* **2013,** *339* (6121), 823-6.

129. Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.;

Koonin, E. V.; Zhang, F., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **2015,** *163* (3), 759-71.

130. Bayat, H.; Modarressi, M. H.; Rahimpour, A., The Conspicuity of CRISPR-Cpf1 System as a Significant Breakthrough in Genome Editing. *Curr Microbiol* **2018,** *75* (1), 107-115.
131. Bandyopadhyay, A.; Kancharla, N.; Javalkote, V. S.; Dasgupta, S.; Brutnell, T. P., CRISPR-Cas12a (Cpf1): A Versatile Tool in the Plant Genome Editing Tool Box for Agricultural Advancement. *Frontiers in Plant Science* **2020,** *11* (1589).
132. Chen, J. S.; Ma, E.; Harrington, L. B.; Da Costa, M.; Tian, X.; Palefsky, J. M.; Doudna, J. A., CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **2018,** *360* (6387), 436-439.
133. Li, S. Y.; Cheng, Q. X.; Liu, J. K.; Nie, X. Q.; Zhao, G. P.; Wang, J., CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res* **2018,** *28* (4), 491-493.
134. Broughton, J. P.; Deng, X.; Yu, G.; Fasching, C. L.; Servellita, V.; Singh, J.; Miao, X.; Streithorst, J. A.; Granados, A.; Sotomayor-Gonzalez, A.; Zorn, K.; Gopez, A.; Hsu, E.; Gu, W.; Miller, S.; Pan, C. Y.; Guevara, H.; Wadford, D. A.; Chen, J. S.; Chiu, C. Y., CRISPR-Cas12-based detection of SARS-CoV-2. *Nat Biotechnol* **2020,** *38* (7), 870-874.
135. Nouri, R.; Jiang, Y.; Lian, X. L.; Guan, W., Sequence-Specific Recognition of HIV-1 DNA with Solid-State CRISPR-Cas12a-Assisted Nanopores (SCAN). *ACS Sensors* **2020,** *5* (5), 1273-1280.
136. Stella, S.; Mesa, P.; Thomsen, J.; Paul, B.; Alcón, P.; Jensen, S. B.; Saligram, B.; Moses, M. E.; Hatzakis, N. S.; Montoya, G., Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **2018,** *175* (7), 1856-1871.e21.
137. Swarts, D. C.; Jinek, M., Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Mol Cell* **2019,** *73* (3), 589-600.e4.
138. McMahon, M. A.; Prakash, T. P.; Cleveland, D. W.; Bennett, C. F.; Rahdar, M., Chemically Modified Cpf1-CRISPR RNAs Mediate Efficient Genome Editing in Mammalian Cells. *Mol Ther* **2018,** *26* (5), 1228-1240.
139. Li, B.; Zhao, W.; Luo, X.; Zhang, X.; Li, C.; Zeng, C.; Dong, Y., Engineering CRISPR-Cpf1 crRNAs and mRNAs to maximize genome editing efficiency. *Nat Biomed Eng* **2017,** *1* (5).
140. Kim, H.; Lee, W. J.; Oh, Y.; Kang, S. H.; Hur, J. K.; Lee, H.; Song, W.; Lim, K. S.; Park, Y. H.; Song, B. S.; Jin, Y. B.; Jun, B. H.; Jung, C.; Lee, D. S.; Kim, S. U.; Lee, S. H., Enhancement of target specificity of CRISPR-Cas12a by using a chimeric DNA-RNA guide. *Nucleic acids research* **2020,** *48* (15), 8601-8616.
141. Munir, A.; Abdullahu, L.; Damha, M. J.; Shuman, S., Two-step mechanism and step-arrest mutants of Runella slithyformis NAD+-dependent tRNA 2′-phosphotransferase Tpt1. **2018,** *24* (9), 1144-1157.

142. Banerjee, A.; Munir, A.; Abdullahu, L.; Damha, M. J.; Goldgur, Y.; Shuman, S., Structure of tRNA splicing enzyme Tpt1 illuminates the mechanism of RNA 2'-PO4 recognition and ADP-ribosylation. *Nat Commun* **2019,** *10* (1), 218.

143. Dantuluri, S.; Abdullahu, L.; Munir, A.; Katolik, A.; Damha, M. J.; Shuman, S., Substrate analogs that trap the 2'-phospho-ADP-ribosylated RNA intermediate of the Tpt1 (tRNA 2'-phosphotransferase) reaction pathway. *Rna* **2020,** *26* (4), 373-381.

Chapter 2

RNA repair enzymes – understanding Tpt1 mechanism *via* synthetic substrates

2.1 Introduction

The enzyme Tpt1 is an essential agent of fungal and plant tRNA splicing that removes the 2′-PO_4 at the splice junction generated by fungal and plant tRNA ligases[1]. Tpt1 catalyzes a unique two-step reaction whereby the internal RNA 2′-PO_4 attacks NAD^+ to form an RNA- 2′-phospho-ADP-ribosyl intermediate that undergoes transesterification to yield 2′-OH RNA and ADP-ribose-1″,2″-cyclic phosphate products (**Figure 2.1**)[2-6].

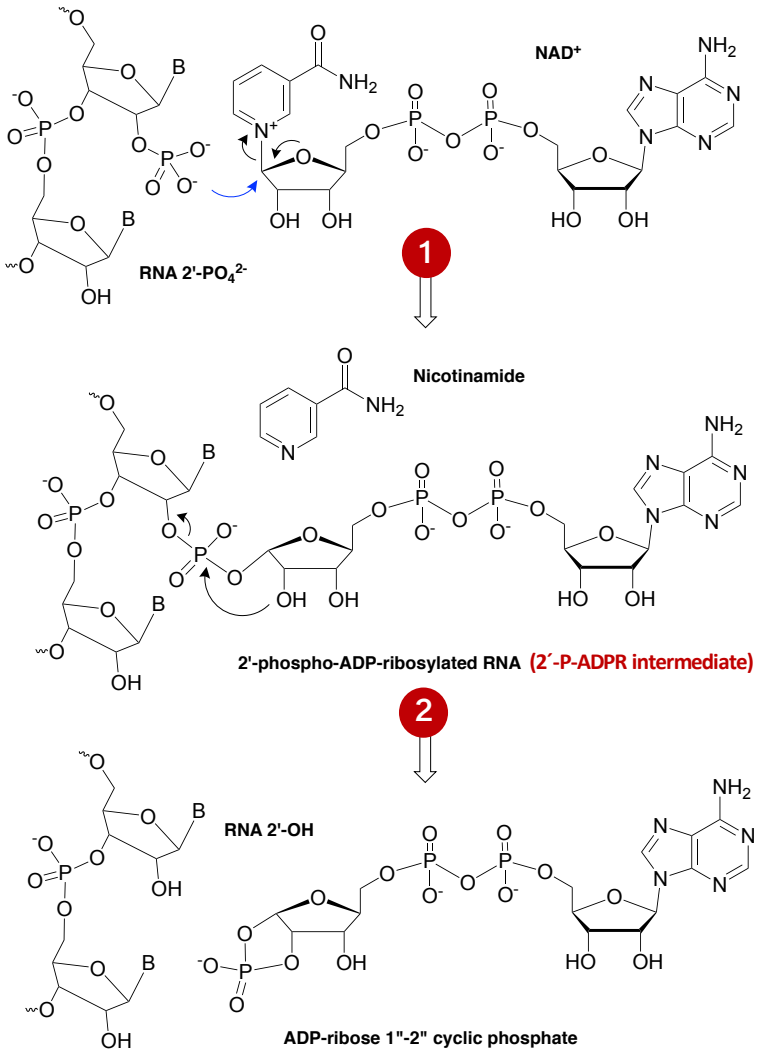


Figure 2.1. *Two-step mechanism of Tpt1.*[3]

Tpt1 homologs are distributed widely in archaeal, bacterial, and metazoan taxa[7]. Many of the bacterial species with Tpt1 homologs have no known intron-containing tRNAs and/or no known pathways to generate RNAs with internal 2′-PO_4 modifications. In

metazoa and archaea, the mechanism of tRNA exon ligation is entirely different from that of fungi and plants and does not result in a junction 2′-PO_4 [8]. Bacterial and human Tpt1 homologs have NAD^+-dependent RNA 2′-phosphotransferase activity in vitro and can genetically complement an otherwise lethal tpt1Δ deletion in *Saccharomyces cerevisiae*, signifying that they are capable of removing the tRNA splice junction 2′-PO_4 in vivo[4, 6, 7, 9]. It is clear by the criteria of genetic complementation of yeast tpt1Δ and direct biochemical assay of recombinant Tpt1 proteins that Tpt1 enzymes from widely divergent organisms all possess NAD^+-dependent RNA 2′-phosphotransferase activity[6, 7, 9, 10].

Whereas it is not obvious what reactions Tpt1 performs in taxa that lack a fungal-type RNA ligase, it is clear that Tpt1 is not essential for their viability and fitness under laboratory conditions, insofar as genetic ablation of *E. coli* Tpt1 (also known as KptA) and mouse Tpt1 has no phenotypic consequences[7, 11]. These findings underscore the high value of Tpt1 as a target for antifungal drug discovery, predicated on pharmacological inhibition of Tpt1′s essential function in fungal tRNA splicing and its contribution to the fungal unfolded protein response[1, 12-14].

Shuman and colleagues previously reported a purification and biochemical analysis of the HD-Pnk end-healing enzyme[15]. Preliminary study of the ligase-like protein indicated that it reacts with ATP to form a covalent enzyme–AMP intermediate characteristic of classic polynucleotide ligases but does not ligate any of the DNA and RNA substrates with 3′ -OH and 5′ -PO_4 termini that were tested. Because the natural substrates for repair by the Runella HD-Pnk and ligase enzymes are not known, we tentatively hypothesize that genetic clustering with a Tpt1 homolog might signal the participation of an RNA 2′-phosphotransferase in a novel bacterial repair pathway. To provide a foundation for this idea, the Shuman and Damha laboratories collaborated in analyzing *Runella slithyformis* Tpt1 enzyme biochemically and genetically. We show here that RslTpt1 is biologically active as an RNA 2′-phosphotransferase in vivo, as gauged by yeast tpt1Δ complementation. We also show that it is biochemically active as an NAD^+-dependent RNA 2′-phosphotransferase in vitro, by using a series of chemically synthesized RNAs with an internal 2′-PO_4 modification. We outline, in detail, the methods used to synthesize the RNA substrates employed in these studies. We define, by alanine scanning, four

conserved amino acids that are essential for RslTpt1 function, and we illuminate their distinct roles during the two chemical steps of the Tpt1 reaction.

2.2 Synthesis of Tpt1 Substrate Mimics for Gauging Tpt1 Activity

The principal stumbling blocks to early mechanistic studies of Tpt1 were the difficulty in preparing a defined 2′-PO_4 RNA substrate in sufficient quantity and the need for post-reaction processing steps for separation and quantification of substrates and products. Then, in the late 90s, the Phizicky laboratory deployed chemically synthesized 2′ -PO_4 RNAs[16] as substrates for SceTpt1 and *E. coli* KptA, which facilitated dissection of the two-step reaction pathway shared by SceTpt1 and EcoKptA[4, 5, 17]. Here, we have developed an optimized, fully automated, solid-phase synthesis of RNAs of varying length, containing an internal 2′-PO_4 using readily available phosphoramidites and reagents.

2.2.1 ALE protecting group paves the way for downstream phosphitylation

Commercially available RNA phosphoramidites are typically protected at the 2′ position with *tert*-butyldimethylsilyl (TBDMS) groups which can be removed by way of fluorinating reagents to yield 2′-OH groups of RNA. This desilylation takes place once the oligonucleotide has been cleaved from the solid support and all other protecting groups have been removed (ammonolysis: ethanol/concentrated aqueous ammonia (1:3 v/v)). To be able to incorporate a single phosphate moiety at a desired location along an RNA, we employed a phosphoramidite with a protecting group that is orthogonal to the rest of the RNA (TBDMS), such that it can be selectively removed from the oligonucleotide, on the solid support, to provide a single OH as a handle for downstream phosphitylation. To that end, we looked to a protecting group previously developed in our lab – the acetal levulinyl ester (ALE) protecting group for the 2′-hydroxyl of ribonucleosides[18]. The ALE protecting group modality was developed in our lab as an optimization of the existing 2′-*O*-levulinyl (Lev) protecting group (**Figure 2.2**). By displacing the acetyl group two atoms further away, it prevents 2′↔3′ isomerisation from the vicinal 3′ phosphate, making it a much more suitable phosphoramidite for oligo synthesis. Both of these groups, labile to hydrazinolysis, arose as promising modalities for the synthesis of long RNA chains that remain attached to a solid polymer support or to a glass or chip surface. They have been

especially significant towards the synthesis of RNA microarrays for many uses (such as in siRNA, CRISPR guide sequences, etc.), where fluoride-labile 2′-*O*-protecting groups (like TBDMS and other silyl groups) are unsuitable as they are incompatible with glass substrates used in RNA chip fabrication[19].

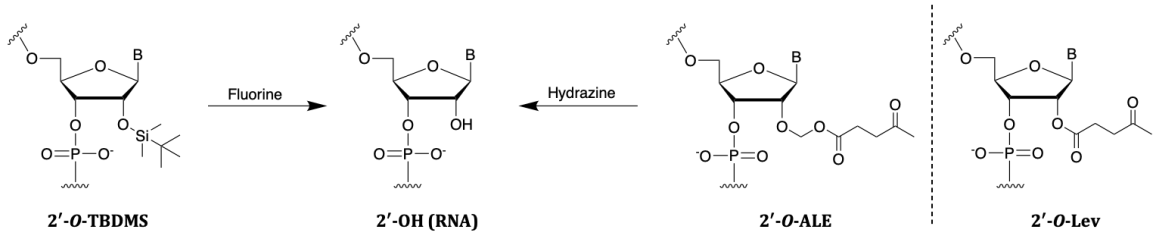


Figure 2.2. *Different RNA protecting groups and their deprotection conditions.*

The synthesis of 2′-*O*-ALE uridine phosphoramidite was first reported by Lackey *et al.* in 2008 (**Figure 2.3**)[18]. Briefly, uridine (1) was reacted with 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane in pyridine, then DMSO was used to initiate a Pummerer rearrangement to give compound **3**. The resulting sulfide then underwent a series of *in situ* reactions containing the key reagent, sodium levulinate, to afford the ALE group on compound 4. This compound then underwent standard procedures to give the desired phosphoramidite. This synthesis has since been reproduced with other nucleobases (A, G, and C) and these phosphoramidites are now commercially available through ChemGenes Corporation (MA, USA). For the purpose of our Tpt1 studies, we employed only the 2′-*O*-ALE adenosine phosphoramidite, as explained below.

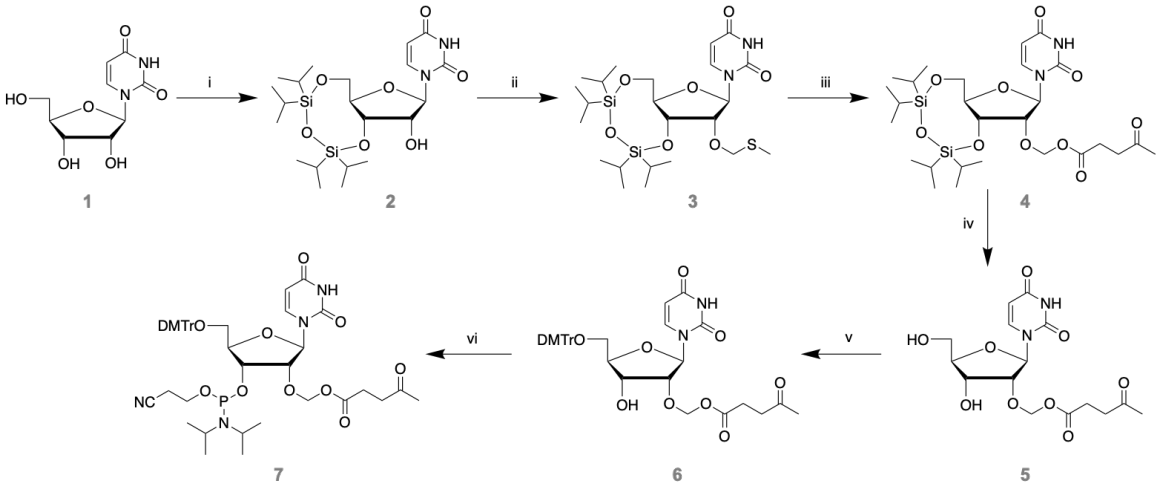


Figure 2.3. *Synthesis of 5'-DMTr-2'-ALE-3'-phosphoramidite-U monomer. i)TIPDSiCl$_2$, pyr. ii) DMSO, AcOH, Ac$_2$O iii) a) SO$_2$Cl$_2$, CH$_2$Cl$_2$ b) Sodium levulinate, 15-C-5, CH$_2$Cl$_2$ iv) NEt$_3$-HF, THF v) DMTrCl, pyr. vi) (i-Pr$_2$N)PClOCE, Hunig's base, THF. (Figure adapted from ref 18)*

2.2.2 Synthesizing RNAs with internal PO_4

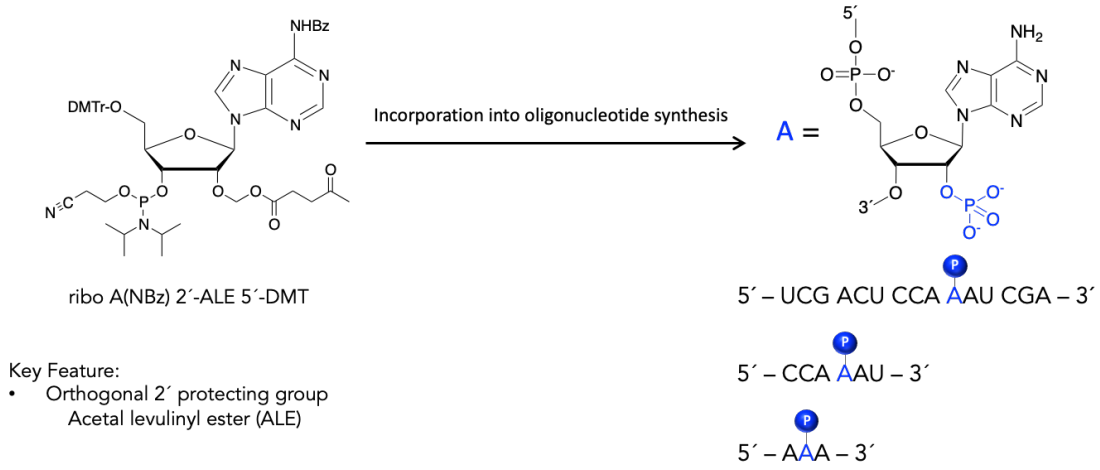


Figure 2.4. *Introducing a single phosphate moiety at a specific position along an oligonucleotide chain.*

With an orthogonal phosphoramidite in hand, we first synthesized a 15-mer RNA oligonucleotide containing an internal 2′ -PO_4 that corresponds to the splice junction of the anticodon stem-loop of yeast tRNATrp (5′-UCGACUCCAA$^{(2'\text{-}PO4)}$AUCGA-3′) from 3′ to 5′, *via* conventional solid phase synthesis using standard RNA phosphoramidites along with 5′-DMTr-2′-ALE-3′-OCE (**Figure 2.4**). Once the oligonucleotide was fully synthesized, and before the final 5′-DMTr removal, the cyanoethyl phosphate protecting groups were removed by passing a mixture of anhydrous triethylamine/acetonitrile solution (2:3 v/v) through the controlled-pore glass (CPG) solid support. This step ensures that all phosphate 3′-5′ linkages, including that vicinal to the 2′-ALE group, are in the diester form, thus preventing chain cleavage in the ensuing removal of the 2′-ALE group. To remove the ALE group, a freshly prepared solution of 0.5 M hydrazine hydrate in pyridine/AcOH (3:2 v/v) was then flowed through the solid support (**Figure 2.5**). The newly exposed internal 2′-OH group was coupled to bis(2-cyanoethyl)-N,N-diisopropylphosphoramidite, and then oxidized to the phosphate triester shown (**Figure 2.6**). Following this step, the DMTr group at the 5′ end of the oligonucleotide was removed under mildly acidic conditions, and the oligo was cleaved from the CPG and deprotected following the standard protocol described in section **2.6** Materials and Methods**)**.

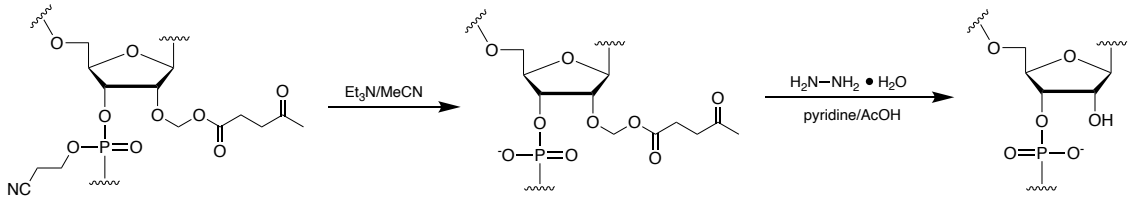


***Figure 2.5**. Cyanoethyl deprotection precedes ALE removal to prevent 2′↔3′ isomerization.*

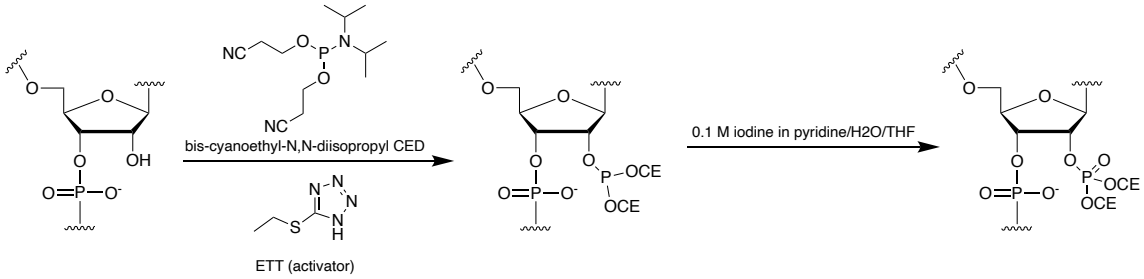


***Figure 2.6.** Phosphitylation followed by oxidation provides the protected phosphate group at the 2′ position.*

Using this method, we were able to successfully synthesize the desired 15mer with an internal 2′-PO_4. However, LCMS characterization of the final purified product revealed that the major product of this synthesis was in fact a 15mer lacking a phosphate group (**Figure 2.7**). We had a few hypotheses as to why the phosphate group may have failed to couple efficiently: (i) a low concentration and coupling time of the phosphitylating reagent were used, (ii) a poor batch of hydrazine hydrate may have led to incomplete removal of the ALE group, subsequently preventing the phosphate group from coupling, and/or (iii) the sterics (folding) of the 15mer, once synthesized, are such that they prevent the deprotection and coupling reagents from effectively accessing the internal site of reaction. Consequently, we tried repeating the synthesis with double the concentration of bis(2-cyanoethyl)-N,N-diisopropylphosphoramidite (0.2 M in MeCN), with a doubled coupling time (30 minutes). While this did have a small impact on the overall yield of the synthesis, the major product remained the 2′-OH 15mer (lacking 2′-PO_4 group). Once we exhausted a few more optimizations throughout the synthesis (increased concentration and coupling times, fresh reagents, more vigorously anhydrous conditions, etc.) and found that they did very little to improve the ratio of desired PO_4 product, we decided to try a new approach to our synthesis.

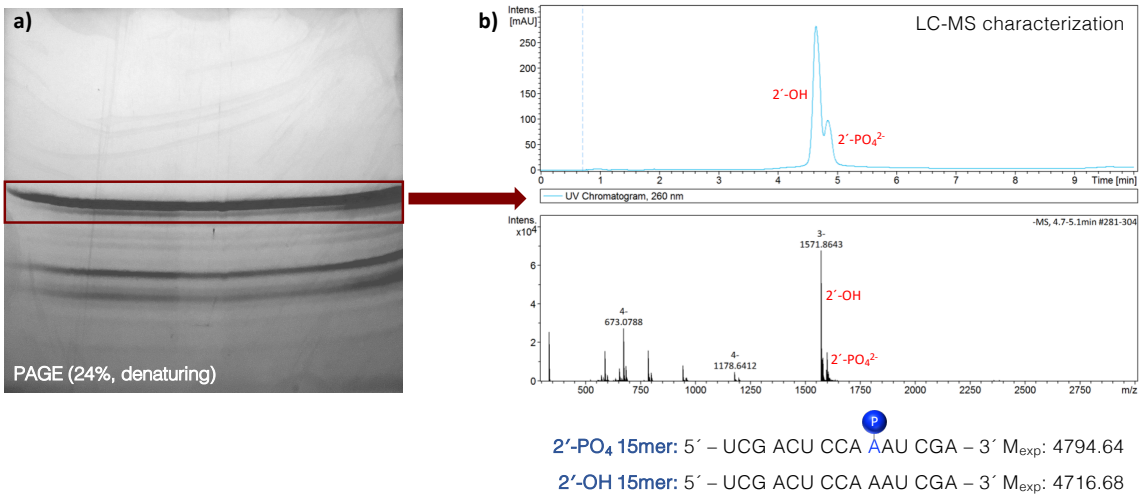


Figure 2.7. *LCMS characterization reveals that major product of synthesis is the 2′-OH product. a) PAGE of the synthesized oligo b) LCMS of the oligo extracted from the top PAGE band (box).*

2.2.3 Optimized synthetic route for RNAs with internal PO_4

To investigate the possibility that steric hindrance was preventing the reagents from reacting at an internal site, we decided to split our synthesis into two parts, so that all of the deprotection and phosphitylation chemistry can occur at the most exposed, terminal position. To do this, we synthesized the oligonucleotide (5′-UCGACUCCAA$^{(2'\text{-PO4})}$AUCGA-3′) from 3′ to 5′, *via* conventional solid phase synthesis, only up until the desired 2′-PO_4 site: 5′-A$^{(2'\text{-ALE})}$AUCGA-3′. With the ALE monomer now at the 5′ terminus, the 2-cyanoethyl and ALE groups were sequentially removed as previously described. Phosphitylation was performed at the optimized concentration and coupling time (0.2M, 30 min) for improved yield. Then, we grew the rest of the oligo (5′-UCGACUCCA-3′) from there *via* conventional solid phase synthesis (**Figure 2.8**). Following synthesis, the oligonucleotide was cleaved from the CPG and all base and backbone protecting groups were removed as described in section 2.6 Materials and Methods).

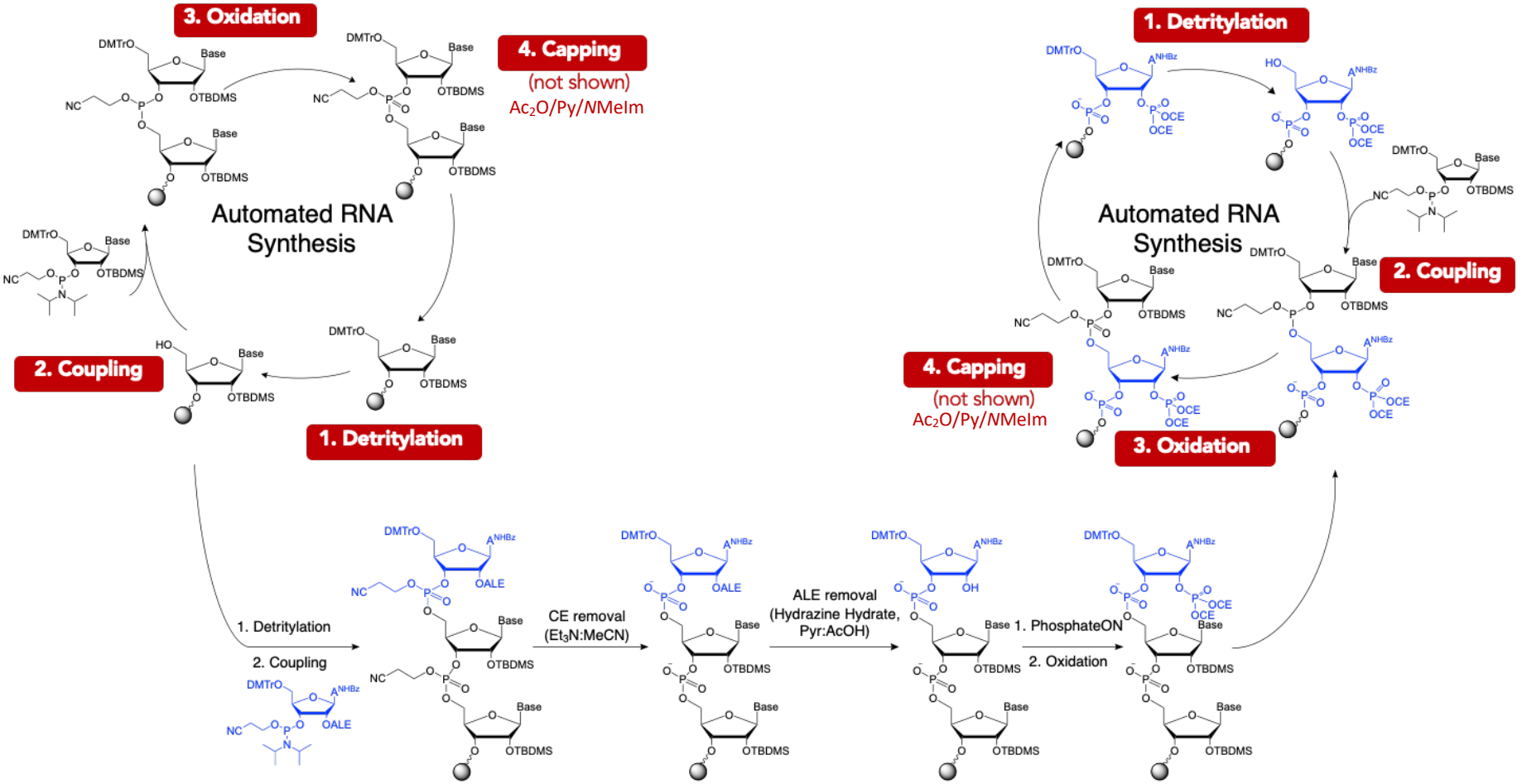


Figure 2.8. *Optimized synthetic route for the synthesis of internal 2'-phosphate-containing RNA.*

Following purification *via* PAGE (24 % acrylamide), LCMS showed that the major product of this synthesis was indeed the desired 2′-PO_4 RNA. However, a small amount of RNA lacking the internal 2′-PO_4 group was also produced, and separated from the desired product by Ion Exchange (IE HPLC) (**Figure 2.9**).

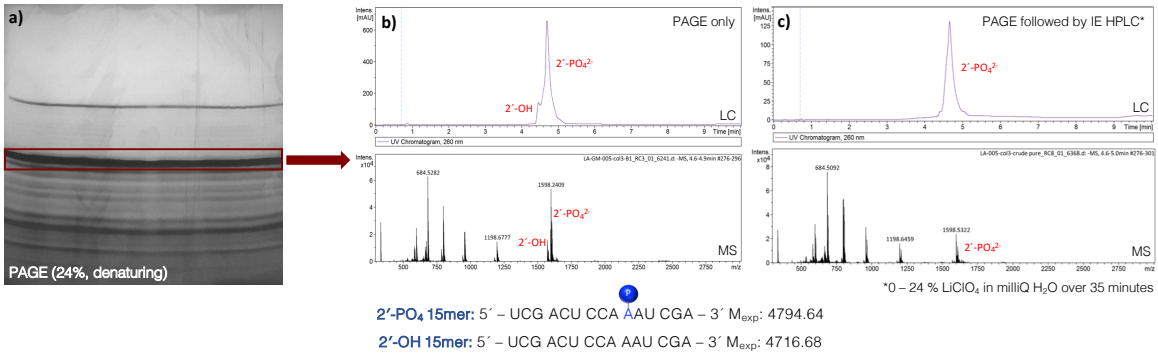


Figure 2.9. *Improved results suggest that the branchpoint 2′-OH is sterically hindered by the rest of the oligo in the 15mer, rendering phosphitylation less efficient. Purity is improved upon IE HPLC. a) PAGE of the synthesized oligo. b) LC-MS of the oligo extracted from the top PAGE band (box).c) LC-MS of the oligo extracted from PAGE band and further purified by IE HPLC.*

All subsequent oligonucleotides used in these studies were synthesized following this optimized procedure. It is important to note that, as we are working towards a minimal substrate inhibitor, we also carried out the synthesis of shorter oligonucleotides (6mer, 3mer) (**Table 2.1**). This would allow us to determine the minimum substrate requirements for binding to Tpt1, and eventually develop something more closely resembling a small-molecule drug. While the synthesis of shorter substrates (6mer and 3mer) followed that of the synthesis of the 15mer, the purification procedures had to be modified to allow for the separation of shorter oligos. Even if we were able to purify these short sequences by IE HPLC, we would lose the majority of the samples during the desalting step, typically via size-exclusion chromatography (SEC) on a Sephadex matrix. This method is more suitable for desalting oligonucleotides ≥10 nucleotides long. Thus, we turned to reverse phase (RP) HPLC which allowed for better and faster purification of the samples, without necessitating a desalting step.

Table 2.1. *Oligonucleotides synthesized as substrate mimics of Tpt1 to study the enzymatic system. LCMS characterization on right.*

Sample	Sequence	M_{exp}	M_{obs}
LA-R15	5'-**UCG ACU CCA AAU CGA** -3'	4794.64	4793.58
LA-R6	5'- **CCA AAU** -3'	1919.27	1919.28
LA-R3	5'-**AAA** -3'	1003.16	1004.22

N = 2'-OH (RNA) **N** = 2'-PO_4

2.3 Results

Once synthesized, our substrates were tested for activity against Tpt1 through a variety of assays including: (i) activity assays to determine whether the substrates are cleaved by Tpt1; (ii) titration assays which measure 2′-PO_4 removal with increasing concentrations of Tpt1 to determine its specific activity; and (iii) kinetic assays which measure 2′-PO_4 removal over a 30 minute time course with constant enzyme concentration to measure reaction rates. The results from these studies are summarized below.

2.3.1 RslTpt1 has RNA 2′-phosphotransferase activity in vitro

Having affirmed that the RslTpt1 is biologically active in RNA processing, our collaborators produced RslTpt1 in *E. coli* and purified the protein as described in Materials and Methods. Our aim was to elucidate the biochemical activities of RslTpt1 and its reaction mechanism. To assay RslTpt1 activity in the present study, we first used the 15-mer RNA oligonucleotide containing an internal 2′ -PO_4 (**Table 2.1**). The 15-mer 2′-PO_4 RNA was 5′ ^{32}P-labeled and gel-purified prior to use in Tpt1 activity assays. We expected to resolve the 2′-PO_4 RNA substrate from the 2′-OH product via denaturing PAGE, but were frustrated to find that this could not be achieved for an RNA of this length. Thus, we had to improvise a new approach to detect removal of the 2′-PO_4 from the 15-mer RNA.

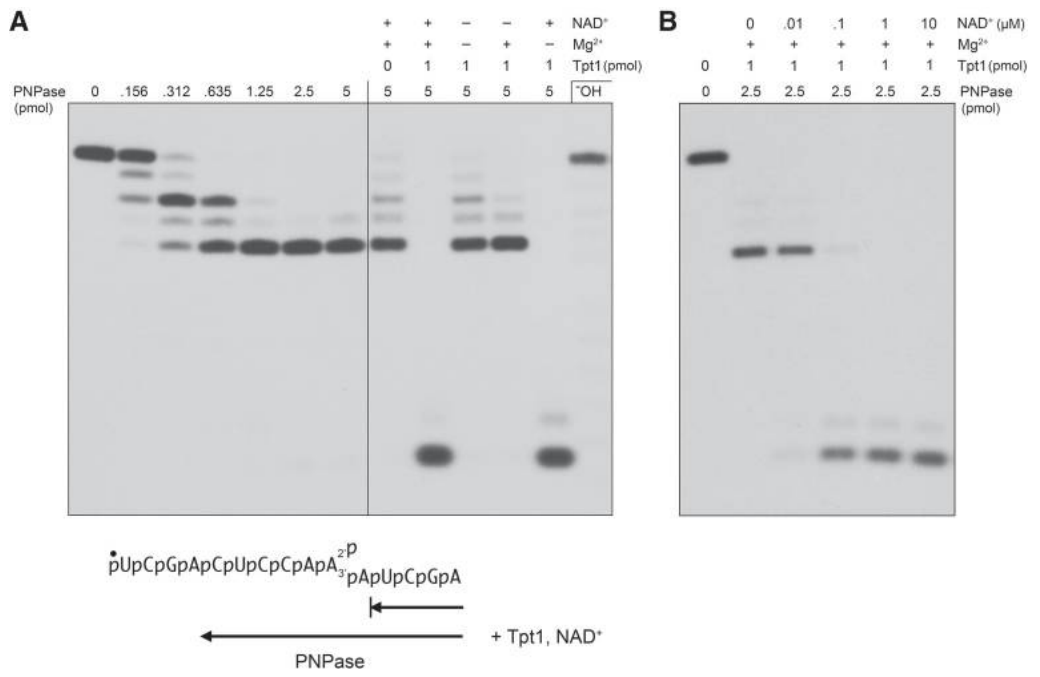


Figure 2.10. *Recombinant RslTpt1 has RNA 2′-phosphotransferase activity in vitro. (A, left side) An internal 2′-PO_4 is an obstacle to 3′ resection by polynucleotide phosphorylase. PNPase reaction mixtures (10 μL) containing 20 mM Tris-HCl (pH 7.5) 5 mM $MgCl_2$, 30 μM $(NH_4)_2HPO_4$, 20 nM (0.2 pmol) 5′ ^{32}P-labeled 15-mer 2′-PO_4 RNA (sequence shown at bottom), and increasing amounts of M. smegmatis PNPase as specified were incubated at 37°C for 15 min. The PNPase reactions were quenched by adding an equal volume of 90% formamide, 50 mM EDTA. (Right side of panel A) Removal of the 2′-PO_4 by RslTpt1 eliminates the block to PNPase digestion. Tpt1 reaction mixtures (10 μL) containing 100 mM Tris-HCl (pH 7.5), 20 nM 5′ ^{32}P-labeled 15-mer 2′-PO_4 RNA, 1 mM NAD^+ (where indicated by +), 5 mM $MgCl_2$ (where indicated by +), and 0 or 1 pmol RslTpt1 as specified were incubated at 37°C for 30 min, then adjusted to contain 30 μM $(NH4)2HPO_4$ and 5 pmol PNPase and incubated at 37°C for 15 min. The PNPase digestions were quenched by adding an equal volume of 90% formamide, 50 mM EDTA. The products were analyzed by electrophoresis (at 55 W constant power) through a 40-cm 20% polyacrylamide gel containing 7 M urea in 45 mM Tris-borate, 1 mM EDTA and visualized by autoradiography. A partial alkaline hydrolysis ladder of the 5′ ^{32}P-labeled 15-mer 2′-PO_4 RNA was analyzed in parallel in the rightmost lane; the alkaline ladder, when overexposed, provided markers to assign the size of the tetranucleotide PNPase reaction end product. The site in the RNA sequence at which PNPase halts 3′ digestion of the 15-mer 2′-PO_4 RNA substrate to yield an 11-mer 2′-PO_4 RNA is indicated at bottom by (|——). The tetranucleotide generated by PNPase digestion after removal of the 2′-PO_4 by RslTpt1 is indicated at bottom by the leftward arrow. (B) NAD^+ titration. Tpt1 reaction mixtures (10 μL) containing 100 mM Tris-HCl (pH 7.5), 20 nM 5′ ^{32}P-labeled 15-mer 2′-PO_4 RNA, 5 mM $MgCl_2$, and 0 or 1 pmol RslTpt1 as specified, and 0, 0.01, 0.1, 1, or 10 μM NAD^+ were incubated at 37°C for 30 min, then adjusted to contain 30 μM $(NH_4)_2HPO_4$ and 2.5 pmol PNPase and incubated for at 37°C for 15 min. The PNPase digestions were quenched by adding an equal volume of 90% formamide, 50 mM EDTA. The products were analyzed by urea-PAGE and visualized by autroradiography.*

The solution to the problem of discriminating a 2′-PO_4 from a 2′-OH was to treat the 5′ ^{32}P-labeled RNA with purified recombinant Mycobacterium smegmatis polynucleotide phosphorylase (MsmPNPase) prior to RNA analysis by denaturing PAGE. MsmPNPase catalyzes iterative Mg - dependent phosphorolysis of RNA 3′ ends to liberate a nucleotide diphosphate at each step[20]. Processive phosphorolysis of 5′ ^{32}P-labeled RNA by MsmPNPase generates a short radiolabeled oligonucleotide product (e.g., a trimer) derived from the 5′ end of the substrate RNA[21]. PNPase spearheads a major pathway of 3′

RNA decay in bacteria. It was shown previously that internal tracts of deoxynucleotides within an RNA substrate elicit a kinetic obstacle to phosphorolysis by MsmPNPase[21]. We suspected that MsmPNPase might also be affected by an internal 2′-PO_4 group. Indeed, we found that 3′ resection of the 15-mer 2′-PO_4 RNA was arrested after the removal of only four 3′ - terminal nucleotides to yield an 11-mer 2′-PO_4 RNA, i.e., MsmPNPase was unable to cleave the 3′–*O*–P bond vicinal to the 2′-PO_4 (**Figure 2.10A**).

The salient finding was that PNPase treatment of the RNA after it had been reacted in vitro with RslTpt1 in the presence of 1 mM NAD^+ and 5 mM $Mg2^+$ yielded a more rapidly migrating 5′ ^{32}P-labeled tetranucleotide as the limit product of PNPase digestion under these conditions (**Figure 2.10A**). This result indicates that Tpt1 effected quantitative removal of the 2′-PO_4 obstacle to PNPase digestion. By systematic omission of components from the Tpt1 reaction mixture, we found that removal of the PNPase-blocking 2′-PO_4 was strictly dependent on Tpt1 and NAD+ (**Figure 2.10A**). To roughly gauge the dependence of Tpt1 activity on NAD^+ concentration, we reacted 100 nM RslTpt1 with 20 nM 5′-^{32}P-labeled 15-mer 2′-PO_4 RNA substrate in the absence of added NAD^+ or in the presence of 0.01, 0.1, 1, or 10 μM NAD^+, prior to treating the reaction mixtures with PNPase (**Figure 2.10B**). The RNA was dephosphorylated completely in the presence of 10 or 1 μM NAD^+. That fact that the Tpt1 reaction was 89% complete at 0.1 μM NAD^+ (i.e., fivefold molar excess over the input RNA substrate and equimolar to input enzyme) attests to the efficiency with which NAD^+ is used as the phosphate acceptor by RslTpt1. Even at 0.01 μM NAD^+ (when NAD^+, at half the concentration of the input RNA substrate, was limiting), 16% of the substrate was dephosphorylated, i.e., one-third of the input NAD^+ was consumed by Tpt1.

To provide additional proof that Tpt1 removed the 2′-PO_4 group, the substrate and the reaction product were subjected to partial alkaline hydrolysis, which generated a ladder of 5′ ^{32}P-labeled RNAs by virtue of chemical transesterification at every phosphodiester with a vicinal ribose 2′-OH (**Figure 2.11A**).

The alkaline ladder derived from the 15-mer 2′-PO_4 RNA substrate has a “skipped” step corresponding to the position of the internal 2′-PO_4 group. Reaction of the substrate with Tpt1 and NAD^+ resulted in a gain of signal at this step in the alkaline ladder, to a level of hydrolysis commensurate with the flanking phosphodiesters (i.e., consistent with

complete removal of the 2′-PO_4). Restoration of the skipped step by Tpt1 depended on the inclusion of NAD^+ in the reaction mixture but did not require an added divalent cation (**Figure 2.11A**). Collectively, the results of these experiments establish that RslTpt1 is a bona fide NAD^+-dependent RNA 2′-phosphotransferase.

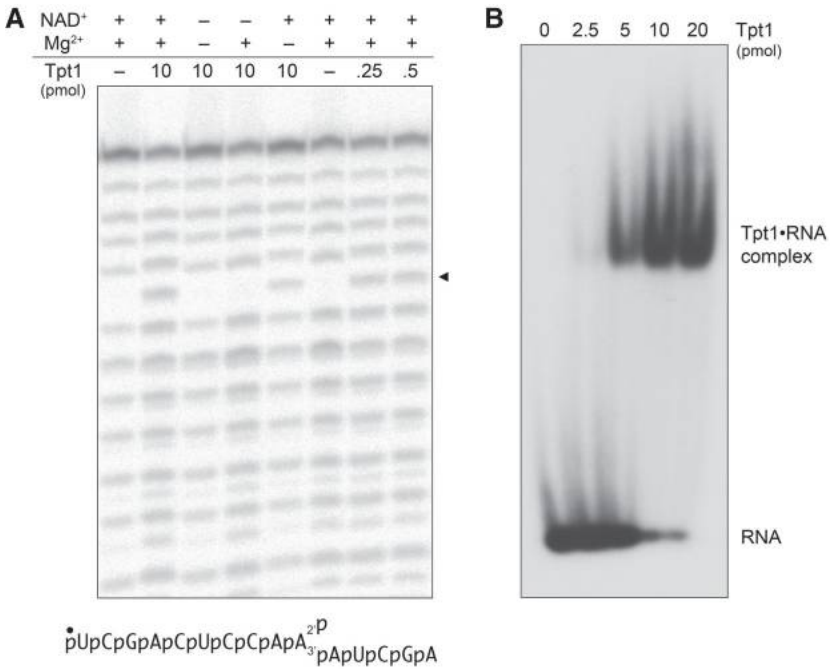


Figure 2.11. *RslTpt1 product analysis and assay of binding to 2′-PO_4 RNA. (A) Reaction mixtures (10 μL) containing 100 mM Tris acetate (pH 6.0), 20 nM (0.2 pmol) 5′ 32P-labeled 15-mer 2′-PO_4 RNA (shown at bottom), 1 mM NAD^+ (where indicated by +), 5 mM $MgCl_2$ (where indicated by +), and RslTpt1 as specified were incubated at 37°C for 30 min. Reaction components were omitted where indicated by –. For partial alkaline hydrolysis, the mixtures were adjusted to 40 mM $NaHCO_3$ (pH 8.0) and 60 mM Na_2CO_3 (pH 10.5) and heated at 95°C for 5 min. The samples were mixed with an equal volume of 90% formamide, 50 mM EDTA and analyzed by urea-PAGE. A scan of the gel is shown. The arrowhead at right denotes the position of the alkaline-sensitive internucleotide phosphodiester resulting from removal of the internal 2′-PO_4 by RslTpt1. (B) Reaction mixtures (10 μL) containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.1 μM (1 pmol) 5′ ^{32}P-labeled 15-mer 2′-PO_4 RNA, 5% (v/v) glycerol, and RslTpt1 as specified were incubated at 22°C for 10 min. The samples were supplemented with glycerol to 15% final concentration and then analyzed by electrophoresis (for 2 h at 110 V) through a 15-cm native 6% polyacrylamide gel in 22.5 mM Tris-borate buffer. An autoradiograph of the dried gel is shown. The positions of the free 15-mer RNA and a Tpt1•RNA complex of retarded mobility are indicated on the right.*

To query whether RslTpt1 binds to the 15-mer 2′-PO_4 RNA substrate in the absence of NAD^+, we used a native gel electrophoretic mobility shift assay. Incubation of RslTpt1 with 0.1 μM labeled RNA resulted in the formation of a discrete protein–RNA complex that migrated more slowly than the free RNA during native PAGE (**Figure 2.11B**). The extent of complex formation increased with input protein and was nearly complete at 1 and 2 μM RslTpt1. In separate gel-shift experiment comparing 5′ ^{32}P-labeled 15-mer synthetic RNAs that had either a 2′-PO_4 or a 2′-OH at the position corresponding to the anticodon-

loop splice junction, we observed that formation of a gel-shifted RslTpt1•RNA complex depended on the 2′-PO_4 group (**Supplementary Figure 2.1**).

2.3.2 Direct assay of RslTpt1 activity with 6-mer and 3-mer 2′-PO_4 RNA substrates

To simplify the Tpt1 assay format, we synthesized shorter RNA substrates with an internal 2′-PO_4 (6-mer and 3-mer; see **Figure 2.12**) that we predicted could be resolved by denaturing PAGE from their respective 2′-dephosphorylated derivatives. As shown in **Figure 2.12**, reaction of 100 nM Tpt1 with 100 nM 5′ ^{32}P-labeled 6-mer and 3-mer 2′-PO_4 RNA substrates resulted in their quantitative conversion to more slowly migrating species corresponding to the 2′-OH product. Removal of the 2′-PO_4 from the 6-mer and 3-mer substrates required added NAD^+ but not divalent cation (**Figure 2.12**). Activity of RslTpt1 on a "minimal" trinucleotide RNA substrate with an internal 2′-PO_4 is consistent with previous studies showing that *E. coli* KptA and yeast Tpt1 can efficiently dephosphorylate this trinucleotide[4, 17].

The extent of conversion of 6-mer 2′ -PO_4 RNA substrate into a 2′-OH product was proportional to the amount of RslTpt1 added (**Figure 2.12B**). From the slope of the enzyme titration curve, we calculated a specific activity of 470 fmol of 2′-OH RNA formed per fmol of RslTpt1 in a 30 min reaction, which translates into a turnover number of 15.7 min^{-1} (**Figure 2.12B**).

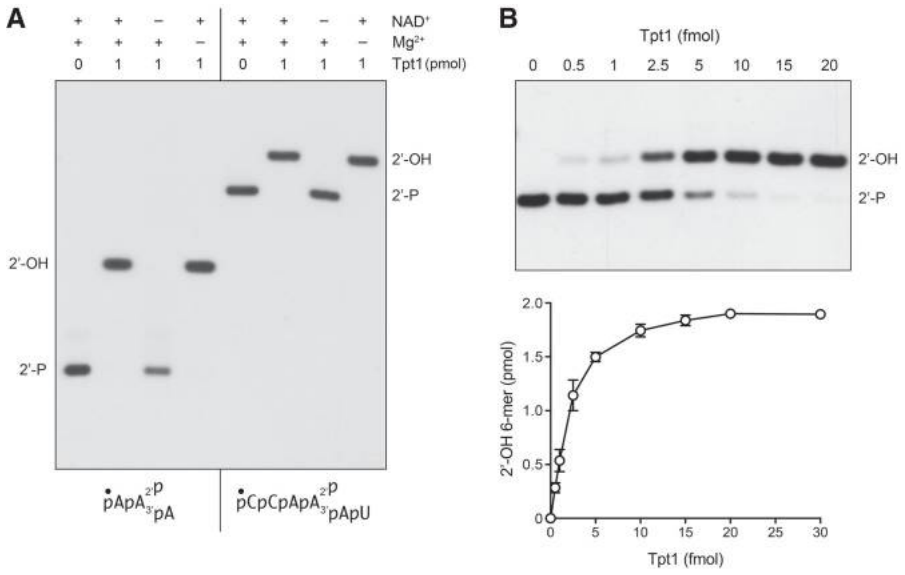


Figure 2.12. Direct assay of RslTpt1 activity with 6-mer and 3-mer 2'-PO4 RNA substrates. *(A) Reaction mixtures (10 µL) containing 100 mM Tris-HCl (pH 7.5), 0.1 µM (1 pmol) 5′ ^{32}P-labeled 6-mer 2′-PO$_4$ RNA or 3-mer 2′-PO$_4$ RNA (shown at bottom), 1 mM NAD+ (where indicated by +), 5 mM $MgCl_2$ (where indicated by +), and 0 or 1 pmol RslTpt1 as specified were incubated at 37°C for 30 min, then mixed with an equal volume of 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE and visualized by autoradiography. (B) Reaction mixtures (10 µL) containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM NAD^+, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 6-mer 2′-PO$_4$ RNA, and RslTpt1 as specified were incubated at 37°C for 30 min, then mixed with three volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE and visualized by autoradiography (top panel). The extent of 2′-OH RNA product formation was quantified by scanning the gel with a Fujifilm FLA-7000 imaging device and is plotted as a function of input RslTpt1 (bottom panel). Each datum in the graph is the average of three separate enzyme titration experiments ±SEM.*

2.3.3 Identification of conserved amino acids essential for RslTpt1 activity in vivo

Our RNA substrates were also useful for identifying essential enzyme-substrate contacts. Prior alanine-scanning mutational analysis of 14 amino acids of yeast Tpt1 that are conserved in homologous proteins from fungi, metazoa, protozoa, bacteria, and archaea identified four residues – Arg23, His24, Arg71, and Arg138 – as essential for Tpt1 function in vivo[9]. The activity of *E. coli* KptA in complementing tpt1Δ was abolished by alanine substitutions at the equivalent side chains: Arg21, His22, Arg69, and Arg125[9]. To gauge whether RslTpt1 relies on the same constellation of putative active site residues, alanines in lieu of RslTpt1 Arg16, His17, Arg64, and Arg119 were introduced and tested by plasmid shuffle for tpt1Δ complementation. All four mutant alleles were lethal in vivo in yeast, i.e., they failed to give rise to FOA-resistant colonies at 37°C, 30°C, or 20°C (**Figure 2.13A** and data not shown).

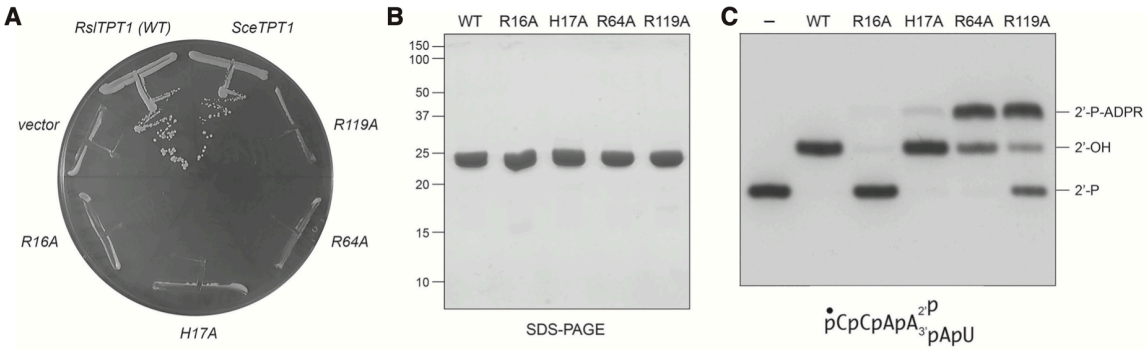


Figure 2.13. *Four conserved amino acids are essential for RslTpt1 activity in vivo. (A) Test of RslTPT1-Ala mutants for S. cerevsiae tpt1Δ complementation by plasmid shuffle. Yeast tpt1Δ p[CEN URA3 TPT1] cells were transformed with CEN plasmids bearing wild-type SceTPT1 (positive control), empty CEN vector (negative control), and wild-type RslTPT1 or the indicated RslTPT1-Ala mutant alelles. Single transformants were selected and streaked on agar medium containing 5-FOA. The plate was photographed after incubation for 7 days at 37˚C. (B) Aliquots (5 µg) of recombinant RslTpt1 were analyzed by SDS-PAGE. The Coomassie Blue-stained gel is shown. The positions and sizes (kDa) of marker polypeptides are indicated on the left. (C) Reaction mixtures (10 µl) containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM NAD+ , 0.1 µM (1 pmol) 5′ ^{32}P-labeled 6-mer 2′-PO4 RNA, and 0.05 µM (0.5 pmol) RslTpt1 (wild-type or mutant) were incubated at 37˚C for 30 min, then mixed with an equal volume of 90% formamide, 50 mM EDTA. The products were resolved by urea-PAGE and visualized by autoradiography.*

2.3.4 Mutational effects on RslTpt1 activity in vitro

To gauge the biochemical impact of the lethal alanine mutations, RslTpt1 proteins R16A, H17A, R64A, and R119A (**Figure 2.13B**) were assayed in a reaction of 0.5 pmol of RslTpt1 with 1 pmol of 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA. These experiments revealed distinctive effects of each mutation on reaction outcome. For instance, whereas wild-type RslTpt1 converted all of the 2′-PO_4 substrate into a single 2′-OH product of slower mobility, the R64A and R119A mutants generated a novel species, migrating above the 2′-OH product (**Figure 2.13C**), that we will show corresponds to the 2′-phospho-ADP-ribosylated RNA intermediate in the Tpt1 reaction pathway[4]. The 2′-P-ADPR species was more abundant than the 2′-OH product in the R64A and R119A reactions (**Figure 2.13C**). In contrast, the R16A mutant was virtually unreactive with the 6-mer 2′-PO_4 RNA at this level of input enzyme. These results suggest that the three essential arginines might act at different steps in the Tpt1 reaction pathway. Although the H17A mutant converted most of the input 2′-PO_4 substrate to 2′-OH product, residual 2′ -P-ADPR RNA was detectable in this single-point assay.

2.3.5 Mutational effects on the kinetics of 2′-PO_4 removal under conditions of enzyme excess

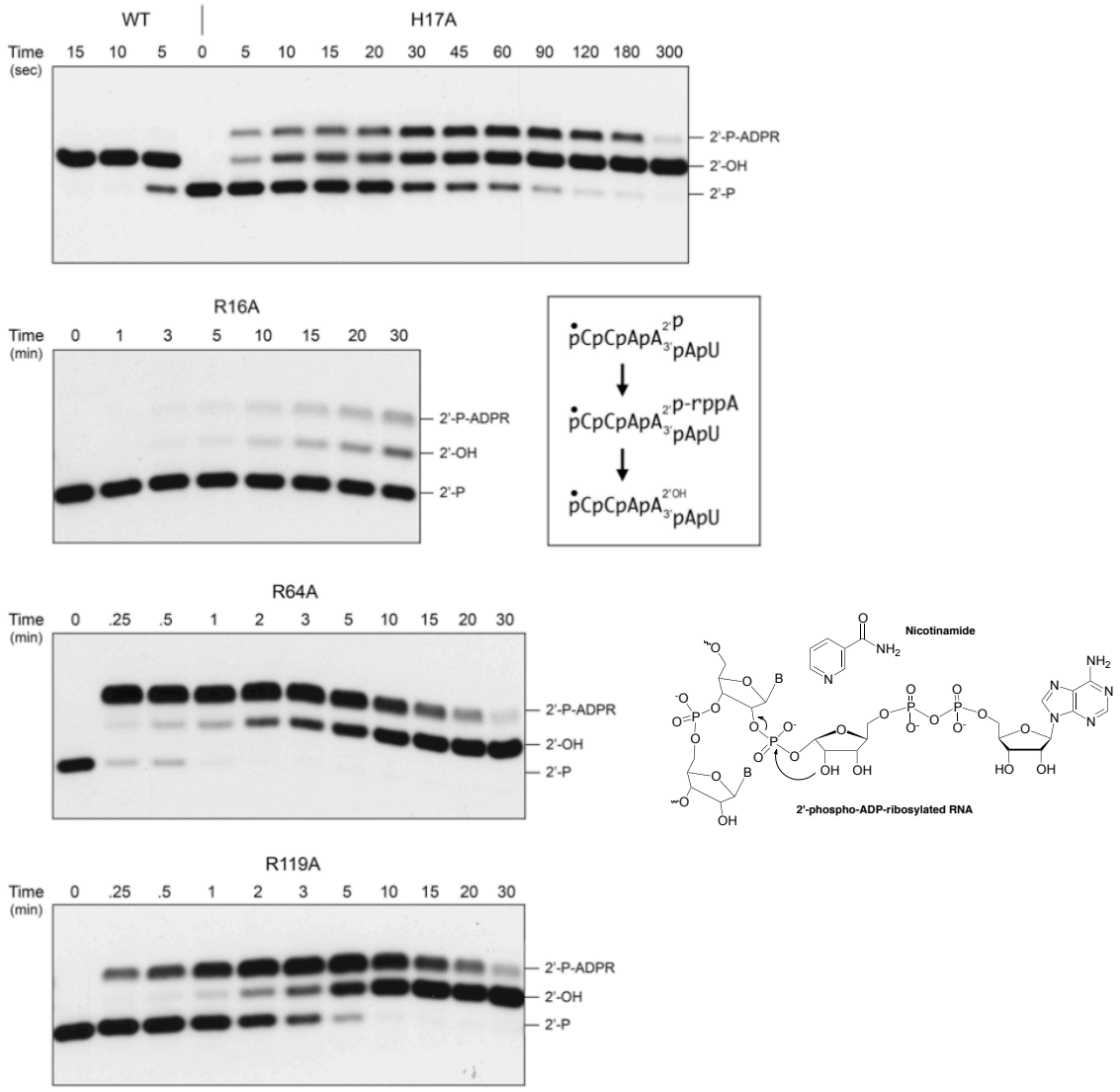
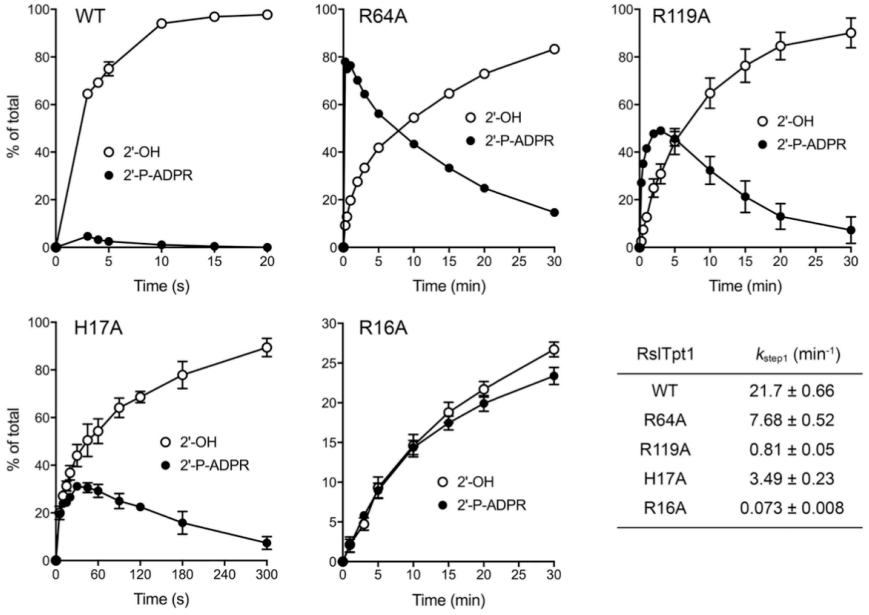


Figure 2.14. *Mutational effects on the kinetics of 2′-PO4 removal. Reaction mixtures containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM NAD^+ , 0.2 μM 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA, and 0.5 μM RslTpt1 (wild-type or mutant as specified) were incubated at 37˚C and quenched at the times specified (above the lanes) by addition of 3 volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed urea-PAGE and visualized by autoradiography. The inset box depicts the 6-mer 2′-PO_4 RNA substrate, 2′-P-ADPR RNA intermediate, and 2′-OH RNA product of the two-step Tpt1 reaction.*

We tracked the time course of the reaction of 0.5 μM RslTpt1 (wild-type and mutants) with 0.2 μM 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA. The product analyses by PAGE are shown in **Figure 2.14**. The distribution of 2′-OH product and 2′-P-ADPR RNAs (expressed as the percentage of total labeled RNA) as a function of time are plotted in **Figure 2.15**, with each data point being the average of three independent experiments. Wild-type RslTpt1 converted 75% of the 2′- PO_4 substrate to 2′-OH product in 5 sec (the earliest time sampled in the experiment shown in **Figure 2.14**) and conversion was

complete in 10 sec. In reactions sampled at 3 sec, wild-type RslTpt1 generated 65% 2′-OH RNA and 4.6% 2′-P-ADPR RNA (**Figure 2.15**). Thus, it seems that step 2 of the 2′ -phosphotransferase pathway (**Figure 2.14**) catalyzed by wild-type RslTpt1 is faster than step 1. An apparent step 1 rate constant of 21.7 min^{-1} for wild-type RslTpt1 (**Figure 2.15**, bottom right panel) was obtained by plotting the sum of the 2′-OH and 2′ -P-ADPR RNAs as a function of time and fitting the data by nonlinear regression to a one-phase association.



RslTpt1	k_{step1} (min^{-1})
WT	21.7 ± 0.66
R64A	7.68 ± 0.52
R119A	0.81 ± 0.05
H17A	3.49 ± 0.23
R16A	0.073 ± 0.008

Figure 2.15. Step-specific effects of mutations on the kinetics of 2'-PO4 removal. *Assays were performed as described in the legend to Figure 2.14. The reaction products were analyzed by ureaPAGE and quantified by scanning the gels. The distributions of the 2'-P-ADPR RNA intermediate and 2'-OH RNA product (expressed as percent of total RNA) are plotted as a function of reaction time. Each datum in the graphs is the average of three independent time course experiments ±SEM. The apparent rate constants for step1 shown in the table at bottom right were derived by plotting the sum of the 2'-P-ADPR RNA intermediate and 2'-OH RNA product as a function of time and fitting the data by non-linear regression to a one-phase association in Prism.*

The kinetic profile of the R64A enzyme was profoundly different. R64A converted 78% of the 2′ -PO_4 substrate to 2′-P-ADPR RNA in 15 sec, after which the 2′-P-ADPR RNA was transformed slowly into 2′-OH RNA over the remainder of a 30 min reaction (**Figure 2.14** and **Figure 2.15**). These data establish a clear precursor-product relationship between the 2′-P- ADPR RNA reaction intermediate and the 2′-OH RNA end product (as depicted in the two-step reaction pathway in **Figure 2.14**) and that replacing Arg64 with alanine selectively slows step 2 of the pathway. The apparent step 1 rate constant of 7.68 min^{-1} for the R64A mutant (**Figure 2.15**, bottom right panel) was threefold slower than that of the wild-type RslTpt1.

The H17A mutant was faster at end-product formation than R64A, insofar as virtually all of the input 2′-PO_4 RNA was converted to 2′-OH RNA in 5 min (**Figure 2.14**). H17A accumulated the 2′-P-ADPR intermediate, which comprised 31% of the total RNA at 30 and 45 sec (**Figure 2.14** and **Figure 2.15**). The H17A step 1 rate constant of 3.49 min^{-1} was sixfold slower than the wild-type rate.

The R119A mutant also displayed a step 2 kinetic defect, seen as the early accumulation of the 2′-P-ADPR intermediate (peaking at 48%–49% of total RNA at the 2 and 3 min sampling times) and its subsequent slow conversion to 2′- OH product over 30 min, with the additional feature that the rate of consumption of the substrate was slowed compared to wild-type, R64A, and H17A (Figs. 6 and 7). The R119A step 1 rate constant of 0.81 min^{-1} was 27-fold slower than the wild-type rate.

The R16A enzyme exhibited a much different defect, whereby it effected a very slow and incomplete consumption of the 2′-PO_4 substrate over a 30 min reaction that yielded a mixture of 2′ -P-ADPR RNA and 2′ -OH product (**Figure 2.14** and **Figure 2.15**). Thus, the R16A mutation drastically affects the first step of the Tpt1 reaction pathway. The R16A step 1 rate constant of 0.073 min^{-1} was 300-fold slower than the wild-type rate.

2.3.6 Isolation of the 2 -P-ADPR RNA intermediate allows study of step 2 per se

We scaled up the reaction of RslTpt1-R64A with the 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA substrate and then gel-purified the 2′-P-ADPR RNA intermediate. The experiment in **Figure 2.16**A shows that a 1-min incubation of 500 nM wild- type RslTpt1 with 50 nM isolated 2′-P-ADPR RNA in the absence of NAD^+ elicited its quantitative conversion to a 2′-OH RNA product. The H17A mutant also effected complete formation of 2′-OH product in 1 min (**Figure 2.16**A). In contrast, the R119A and R64A proteins achieved 56% and 22% conversion of 2′-P-ADPR RNA to 2′-OH RNA in 1 min, which increased to 97% and 75% conversion after 10 min. The R16A protein converted one-fourth of the input 2′-P-ADPR RNA to 2′-OH RNA in 10 min (**Figure 2.16**A).

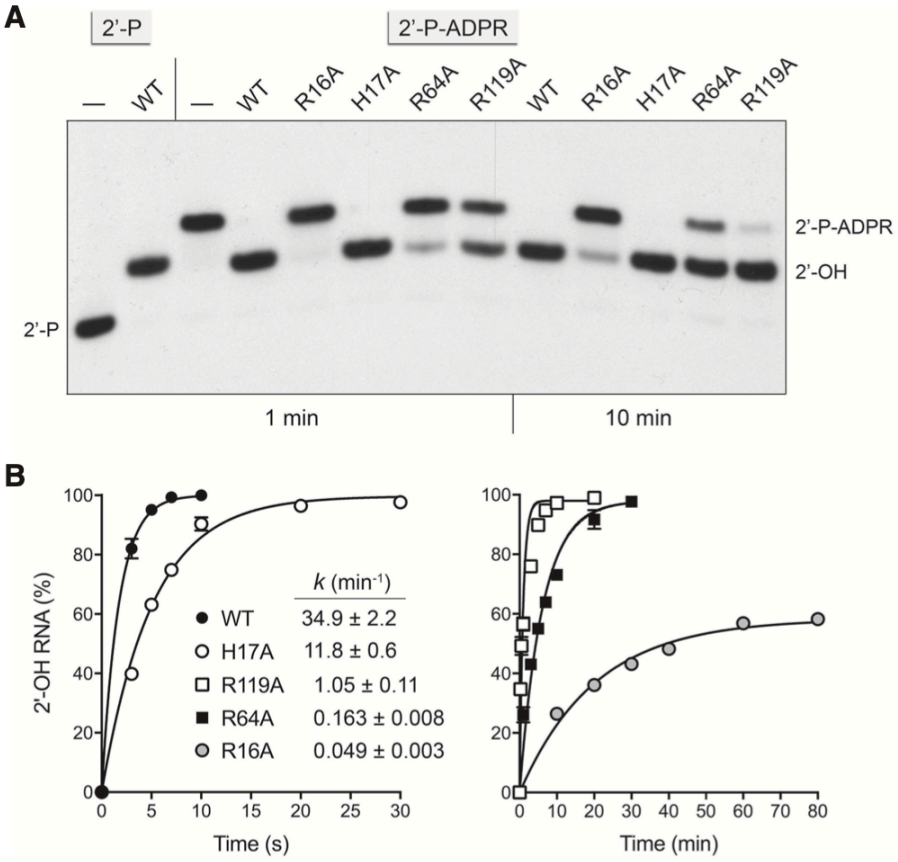


Figure 2.16. NAD$^+$-independent conversion of 2'-P-ADPR RNA intermediate to 2'-OH product. *(A) Reaction mixtures containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.05 μM isolated 5′ ^{32}P-labeled 6-mer 2′-P-ADPR RNA (lanes 2′-P-ADPR), and 0.5 μM RslTpt1 (wild-type or mutant as specified) were incubated at 37˚C and quenched at the times specified (1 min or 10 min) by addition of 3 volumes of cold 90% formamide, 50 mM EDTA. RslTpt1 was omitted from the reaction in lane –. The reaction products were analyzed urea-PAGE and visualized by autoradiography. Control reactions containing 0.05 μM 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA substrate, 1 mM NAD$^+$, and either 0.5 μM RslTpt1 (2′-P, lane WT) or no added enzyme (2'P, lane –) were analyzed in parallel in the leftmost lanes. (B) Reaction mixtures containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.05 μM 5′ ^{32}P-labeled 6-mer 2′-P-ADPR RNA, and 0.5 μM RslTpt1 (wild-type or mutant as specified) were incubated at 37˚C and quenched at the times specified by addition of 3 volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE and quantified by scanning the gels. The extents of 2′-OH RNA product formation (expressed as percent of total RNA) are plotted as a function of reaction time with x-axis units in seconds (left panel) or minutes (right panel). Each datum in the graphs is the average of three independent time course experiments ±SEM. The curves and apparent step 2 rate constants were obtained by fitting the data by non-linear regression to a one-phase association in Prism.*

A finer kinetic analysis of the isolated step 2 reaction of the wild-type and mutant enzymes is shown in **Figure 2.16**B. The wild-type, H17A, R119A, and R64A reactions all proceeded to completion on a time-scale ranging from 7 sec (wild- type) to 30 min (R64A). The R16A reaction reached an apparent endpoint of 58% conversion of 2′-P-ADPR RNA to 2′-OH RNA after 80 min. The kinetic profiles were fit to a one-phase association to derive apparent step 2 rate constants as follows: WT (34.9 min^{-1}); H17A (11.8 min^{-1}); R119A (1.05 min^{-1}); R64A (0.163 min^{-1}); R16A (0.049 min^{-1}). Thus, the H17A, R119A,

R64A, and R16A mutations slowed the rate of the isolated step 2 reaction by factors of 3, 33, 210, and 710, respectively.

2.3.7 Mutational effects on 2′-phophotransferase specific activity

The studies of mutational effects on RslTpt1 activity in the preceding sections were performed under conditions in which RslTpt1 was present in molar excess over the RNA substrate, a state unlikely to apply in vivo when the RslTpt1 mutants were tested for tpt1Δ complementation in yeast and found to be lethal. In the experiments shown in **Figure 2.17**A–D, the impact of the four lethal alanine changes on 2′-phosphotransferase specific activity was gauged by enzyme titration across a broad range spanning substoichiometeric to super-stoichiometeric RslTpt1:RNA conditions. The H17A, R119A, R64A, and R16A proteins generated substantial amounts of 2′-P-ADPR RNA intermediate at the lower range of input enzyme tested. The intermediate declined and the 2′-OH RNA product accumulated steadily as the H17A, R119A, and R64A protein concentrations were increased, with nearly complete conversion to 2′-OH product being achieved at the highest RslTpt1 levels tested. The specific activities of the mutants in forming the 2′-OH RNA product were calculated from the slopes of the titration curves in the linear range of enzyme dependence, as determined by linear regression in Prism. The specific activity values, expressed as pmol of 2′-OH RNA generated per pmol of input RslTpt1 in a 30 min reaction, were as follows: 3.59 ± 0.22 for H17A; 0.751 ± 0.056 for R119A; 0.90 ± 0.035 for R64A; and 0.098 ± 0.0059 for R16A. Compared to the wild-type RslTpt1 specific activity of 470 (Fig. 4B), the H17A, R119A, R64A, and R16A mutants exhibited specific activity decrements of 130-fold, 625-fold, 520-fold, and 4800-fold, respectively.

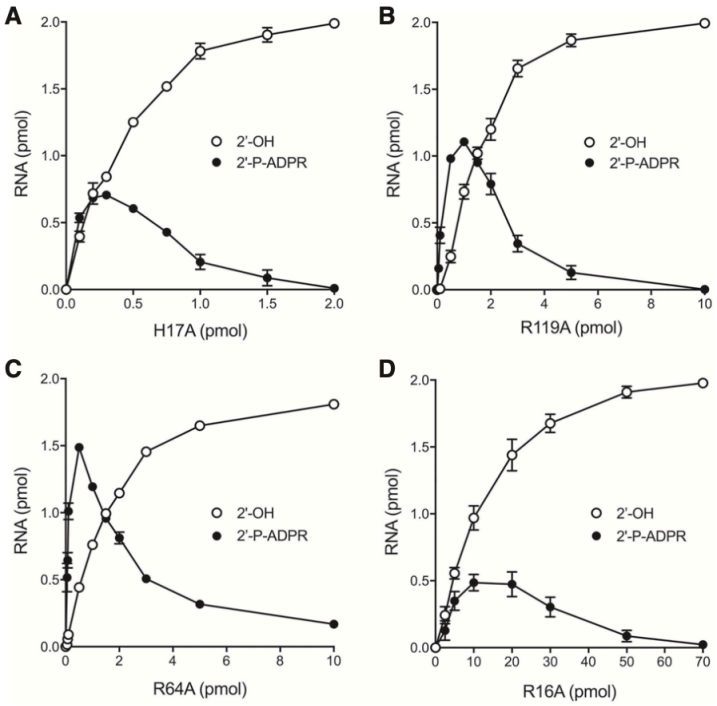


Figure 2.17. Mutational effects on 2′-phosphotransferase specific activity. *Reaction mixtures (10μl) containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM NAD^+, 0.2 μM (2 pmol) 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA, and RslTpt1 mutants H17A (panel A), R119A (panel B), R64A (panel C), and R16A (panel D) as specified were incubated at 37˚C for 30 min, then mixed with three volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE. The extents of 2′-P-ADPR RNA intermediate and 2′-OH RNA product formation were quantified by scanning the gels and are plotted as a function of input RslTpt1. Each datum in the graphs is the average of three separate enzyme titration experiments ±SEM.*

2.4 Uncovering Other Roles for Various Tpt1 Homologues

A longstanding mystery surrounds the role that Tpt1 enzymes might play in taxa that have no fungal/plant-type RNA ligases. One possibility is that bacteria, archaea, or metazoa do have a capacity to install internal RNA 2′-PO_4 groups, conceivably during RNA repair pathways that are as yet uncharacterized, and that such RNAs are potential substrates for Tpt1. An alternative scenario is that Tpt1 enzymes can catalyze reactions other than internal RNA 2′-PO_4 removal, via their unique NAD^+-dependent transferase mechanism.

2.4.1 Tpt1 facilitates NAD^+-dependent synthesis of a 5′-phospho-ADP-ribosylated RNA/DNA cap

Support for the latter scheme emerged from our recent findings that a subset of Tpt1 enzymes can transfer ADP-ribose from NAD^+ to a 5′-monophosphate end of RNA or DNA to install a 5′-phospho-ADP-ribose cap structure[10]. *Aeropyrum pernix* Tpt1 (ApeTpt1) was

particularly adept in this respect. Two other archaeal Tpt1s (from *Pyrococcus horikoshii* and *Archaeoglobus fulgidis),* one bacterial Tpt1 (from *Clostridium thermocellum*) and one fungal Tpt1 (from *Chaetomium thermophilum*) also exhibited 5′-phospho-ADP-ribose capping activity, albeit less effectively than ApeTpt1. In contrast, human Tpt1 and *Runella slithyformis* Tpt1 were inactive in 5′-phospho- ADP-ribose capping. The one-step synthesis of a 5′-phospho-ADP-ribosylated cap structure by a subset of Tpt1 enzymes (with no subsequent 5′-phosphotransferase step) extended the repertoire of the Tpt1 family and raised the prospect that Tpt1 enzymes might perform additional phospho-ADP-ribosylation reactions, with or without sub- sequent removal of the phosphate moiety.

2.4.2 Crystal structure of Tpt1 illuminates the mechanism of RNA 2′-PO_4 recognition and ADP-ribosylation

The lack of a structural snapshot of Tpt1 at relevant steps along the reaction pathway was a major knowledge gap in fungal tRNA metabolism and an impediment to inhibitor discovery and design. Valuable insights into substrate recognition and the mechanism of the transesterification step emerged from a crystal structure of *Clostridium thermocellum* Tpt1 (CthTpt1) in a product-mimetic complex with ADP-ribose-1″-PO_4 in the NAD^+ site and pAp in the RNA site.[22] Our collaborators attained a 1.4 Å crystal structure of Tpt1, conveniently captured in a complex with ADP-ribose-1″- phosphate in the NAD^+ site (generated during protein production in *E. coli*) and the pAp moiety of coenzyme A in the RNA site (**Figure 2.18**). This structure, which mimics the step 2 product complex of the Tpt1 reaction, yields keen insights into how Tpt1 recognizes a 2′-PO_4 RNA splice junction. It also provides the first evidence that a bacterium has an endogenous phosphorylated substrate with which a Tpt1 enzyme can react. We postulated that if the endogenous substrate for CthTpt1 is an RNA, it is either the case that (i) *E. coli* can form an RNA with an internal 2′-PO_4 modification, via an as yet undefined pathway; or (ii) CthTpt1 can catalyze removal of an RNA terminal 2′-monophosphate or 3′-monophosphate moiety.

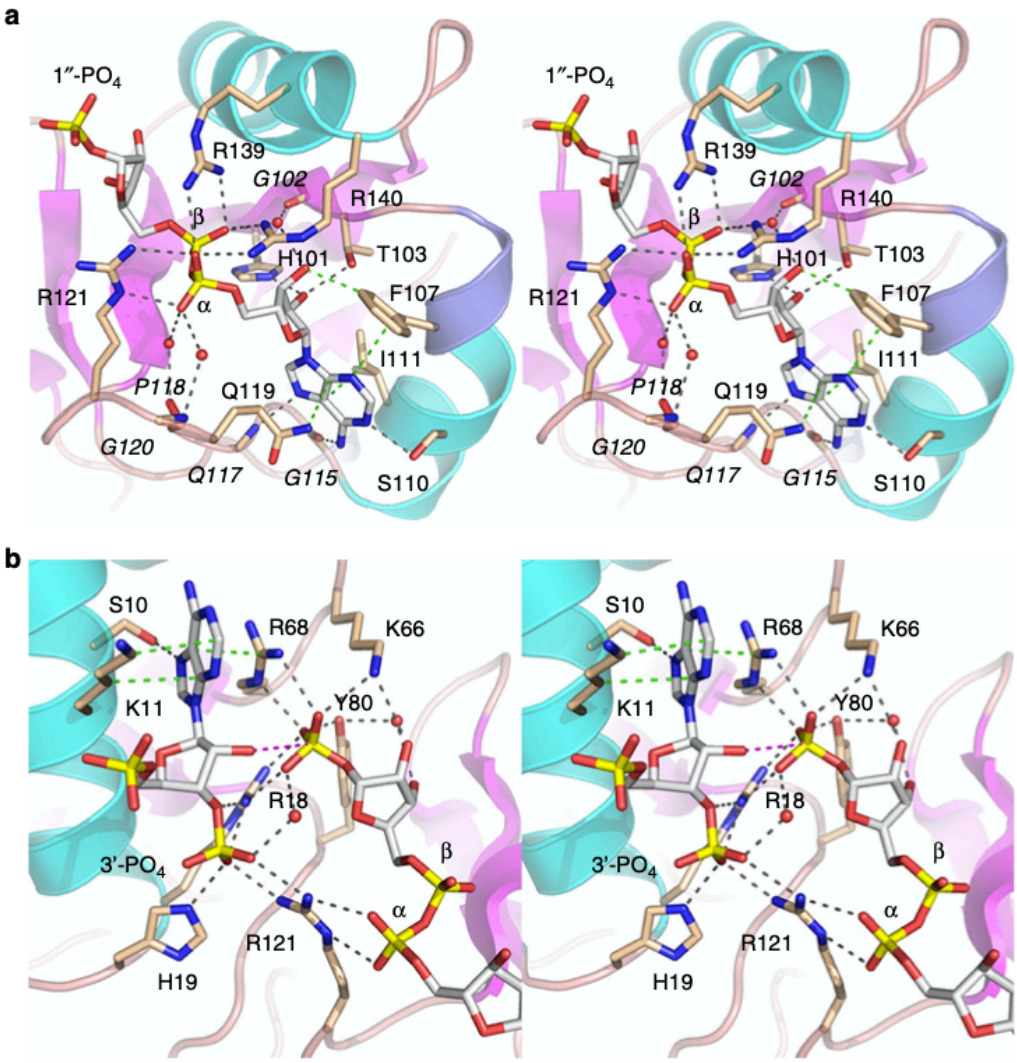


Figure 2.18. *Tpt1 active site. Stereo views of the atomic interactions of Tpt1 with the ADP-ribose moiety of the ADP-ribose-1″-PO4 product (a) and the pAp and 1″- PO4 moieties (b). Selected Tpt1 amino acids are depicted as stick models with beige carbons. The ligands are depicted as stick models with gray carbons. Waters are depicted as red spheres. Hydrogen bonds are denoted by black dashed lines and van der Waals contacts by green dashed lines. The distance (3.5 Å) from the ribose 2′-O to the 1″-P in (b) is indicated by a magenta dashed line. Amino acids that make side-chain contacts to the ligand are labeled in plain font; those that make main-chain contacts to the ligands are labeled in italics*

2.4.3 A subset of Tpt1 enzymes facilitates NAD^+-dependent RNA terminal 2′ and 3′ phosphomonoesterase activity

To test the latter hypothesis, we synthesized a pair of 10mer RNA oligonucleotides with either a terminal 2′-PO_4 or 3′-PO_4 moiety, and subsequently, a 2′-3′-bis phosphorylated 10mer. To our content, these substrates were dephosphorylated by a series of Tpt1 homologues to varying degrees: CthTpt1 and ApeTpt1 catalyze the NAD^+-dependent conversion of 2′-PO_4 and 3′-PO_4 ends to 2′-OH and 3′-OH ends, respectively,

whereas *Chaetomium* Tpt1 selectively coverts a 2′-PO_4 ends to a 2′-OH. In contrast, *Runella slithyformis* Tpt1 and human Tpt1 are not adept at either terminal phosphate removal reaction. These findings are elaborated upon in **Appendix I.**

2.5 Discussion and Summary of Current Findings

The present study enhances our understanding of the two-step pathway of NAD^+ -dependent RNA 2′-phosphate removal by Tpt1 enzymes, whereby attack of the 2′-phosphate on the nicotinamide ribose of NAD^+ leads to the formation of a 2′-phospho-ADP-ribosylated RNA intermediate that is subsequently converted to an RNA 2′-OH product via transesterification of the phosphate to the ribose O2′ [4, 5]. In functional studies carried out by the Shuman lab, Tpt1 enzyme from the bacterium *Runella slithyformis* (RslTpt1) was characterized as a bona fide NAD^+-dependent RNA 2 -phospho-transferase. It performed this essential reaction in vivo in budding yeast and in vitro using our short synthetic 2′-PO_4 RNA substrates. The biological activity of RslTpt1 relies on a conserved histidine and three conserved arginines that were shown to be essential for the in vivo activity of *S. cerevisiae* Tpt1 and *E. coli* KptA[9]. However, the earlier structure-function studies lacked a biochemical correlate, because an idoneous RNA substrate for SceTpt1 and EcoKptA was not in hand.

The properties of wild-type RslTpt1 highlight its efficiency. Under conditions of RNA substrate excess, RslTpt1 performed 470 rounds of 6-mer RNA 2′-dephosphorylation per enzyme during a 30 min reaction at 37°C, without detectable accumulation of a 2′-phospho-ADP-ribosylated RNA intermediate in the linear range of the enzyme titration. Under conditions of enzyme excess, RslTpt1 converted the 6-mer substrate to product in just a few seconds. After 3 sec (the earliest time we could sample manually), the 2′-OH product and 2′-P-ADPR intermediate comprised 65% and 4.6% of the total RNA, respectively, suggesting that step 1 is likely to be rate-limiting. Indeed, the kinetic data for wild-type RslTpt1 (in **Figure 2.15**) fit well by linear regression to a unidirectional two-step mechanism with step 1 and step 2 rate constants of 21.0 ± 0.62 min^{-1} and 221 ± 51 min^{-1}, respectively (i.e., step 2 is 10-fold faster than step 1). Consistent with formation of the 2′-P-ADPR RNA intermediate being rate-limiting, the single-turnover step 1 rate constant of 21 min^{-1} was similar to the turnover number of 15.7 min^{-1} observed by enzyme

titration under steady-state conditions. Note that the turnover number of RslTpt1 with the 6-mer substrate pCCAA$^{2'PO4}$AU was similar to the k_{cat} value of 19 min^{-1} reported for SceTpt1 with the 8-mer substrate pGUAA$^{2'PO4}$AUCU [17].

The intermediacy of the 2′-P-ADPR species was affirmed by the ability of wild-type RslTpt1 to effect rapid and complete conversion of the gel-purified 6-mer 2′-P-ADPR RNA to the 2′-OH RNA product in the absence of NAD^+. However, the apparent single-turnover rate constant of 34.9 min^{-1} for step 2 in isolation was six-fold slower than the step 2 rate derived for the reaction of RslTpt1 with 2′-PO_4 RNA substrate at the same concentration of input enzyme (0.5 μM). We suspect that the isolated step 2 reaction with 2′-P-ADPR RNA in solution may be subject to a rate-limiting RNA binding (or precatalytic conformational) step that does not apply when the 2′-P-ADPR intermediate is formed in situ on the enzyme.

The four RslTpt1-Ala mutants studied (R16A, H17A, R64A, and R119A) display distinctive biochemical defects that are mechanistically informative. They had grossly reduced specific activities as 2′-phosphotransferases in vitro, as gauged by enzyme titrations. Given the likelihood that Tpt1 must act iteratively in vivo to handle the flux of 2′-phosphate-containing splice junctions formed by yeast tRNA ligase, we would attribute the lethality of the RslTpt1-Ala mutants in yeast to their reduced specific activities, which could reflect a defect in substrate recognition, chemical catalysis, or a combination thereof. To focus on mutational effects on the chemical steps, we tracked the kinetics of the reaction of the RslTpt1-Ala proteins with the 6-mer 2′-PO_4 RNA substrate under conditions of enzyme excess. We thereby found that, unlike wild-type RslTpt1, the H17A, R119A, and R64A mutants accumulated substantial levels of the 2′-P- ADPR intermediate, which comprised 31%, 50%, and 78% of the total RNA at their peak times, before slowly converting the intermediate to 2′-OH product. Deriving apparent step 1 rate constants by fitting the sum of the intermediate and product RNAs versus time to a one phase association showed that the rates of step 1 for H17A, R64A, and R119A were slower than wild-type RslTpt1 by factors of 6, 3, and 27, respectively. A more severe defect was seen for the R16A mutant, which slowly generated intermediate and product in roughly equal amounts over a 30 min period and for which the apparent step 1 rate constant was 300-fold slower than wild-type.

It is noteworthy that none of the temporal reaction profiles of the RslTpt1-Ala mutants fit well to a simple unidirectional two-step kinetic mechanism, the problem being that the observed decay of the 2′-P-ADPR species and conversion to product at later reaction times was too slow for a good fit. The "least poor" of the fits were seen for the R64A (k_1 of 5.4 ± 0.9 min^{-1} and k_2 of 0.089 ± 0.005 min^{-1}) and R119A (k_1 of 0.62 ± 0.07 min^{-1} and k_2 of 0.15 ± 0.01 min^{-1}). The step 2 rates for R64A and R119A calculated from fitting to the unidirectional two-step scheme are twofold and sixfold slower than the observed rates of the isolated step 2 reaction with the purified 2′-P-ADPR RNA intermediate (0.163 min^{-1} and 1.05 min^{-1}, respectively). The results are more in keeping with a branched kinetic mechanism for the mutant enzymes, in which a fraction of the 2′-P-ADPR RNA formed during step 1 reversibly enters an off-pathway state that is most simply postulated to entail dissociation of 2′-P- ADPR RNA from the mutant enzyme and rebinding from solution in order to proceed through step 2. A key point is that the single turnover reaction with the 2′-PO_4 RNA substrate (0.2 μM) is performed in the presence of 1 mM NAD^+, a sensible condition given that the intracellular concentration of NAD^+ in E. coli is 2.6 mM[23]. Consequently, any 2′-P-ADPR RNA that dissociates from Tpt1 after step 1 is competing with a ≥5000-fold excess of NAD^+ for binding to the enzyme active site, wherein it is likely that the ADP-ribose moieties of the 2′-P-ADPR RNA and NAD^+ occupy the same or closely overlapping positions, such that their binding is mutually exclusive.

By focusing on mutational effects on the step 1 rate and the rate of the isolated step 2 reaction, we can conclude that Arg64 is selectively required for the step 2 transesterification reaction: viz., the R64A mutation slows the step 1 rate by only threefold while slowing the step 2 rate by a factor of 214 *vis à vis* wild-type RslTpt1. In contrast, the other mutants studied here have relatively less skew in their impacts on step 1 versus step 2. To wit, H17A is slowed sixfold and threefold; R119A is slowed 27-fold and 33-fold; and R16A is slowed 300-fold and 710-fold in steps 1 and 2, respectively. We hypothesize that Arg64 makes atomic contacts in the active site that are uniquely pertinent to the chemistry of step 2, e.g., contacts to the RNA 2′-phosphate moiety of the 2′-phospho-ADP-ribosylated RNA intermediate. A bidentate interaction of Arg64 with the 2′-phosphate oxygens would stabilize the extra negative charge in the putative penta-coordinated phosphorane transition-state of the step 2 transesterification reaction. (This model is

appealing given that we find that RslTpt1 does not require a divalent cation for activity in vitro.) In contrast, the chemistry of step1 occurs at the ribose C1 carbon of NAD^+ and, although Arg64 might promote the binding of the 2′-PO_4 RNA to the enzyme via interactions with the 2′-phosphate, there is no implicit need for such contacts in order for the 2′-phosphate to act as a nucleophile in step 1. Previously, Steiger et al (2005)[5] identified a double-mutant K69A-R71S of SceTpt1 that was lethal in vivo and had ~1000-fold lower specific activity in dephosphorylating a trinucleotide 2′-PO_4 RNA substrate in vitro, yet had the distinctive property of accumulating substantial amounts of the 2′-P-ADPR intermediate. Arg71 of SceTpt1 (which results in lethality in vivo when mutated singly to alanine) is the equivalent of Arg64 in RslTpt1.

The observation that the R16A, H17A, and R119A mutations affect both steps of the Tpt1 pathway with varying but balanced severity suggests that Arg16, His17, and Arg119 either make particular contacts to the RNA or NAD^+ substrates that are important for both steps or make different atomic contacts during steps 1 and 2, both of which are important. A priori, in a metal-independent mechanism, there is a likely need for a general base catalyst to deprotonate the ADP-ribose O2″ position so it can act as the nucleophile during step 2 transesterification and a histidine side chain is a good candidate for this role (as well as potentially contributing to NAD^+ binding during step 1). Similarly, there is a presumptive need for a general acid catalyst to donate a proton to the RNA ribose O2′ leaving group during step 2 transesterification. Clarification of the basis for substrate recognition and reaction chemistry by Tpt1 obviously hinges on obtaining an ensemble of atomic structures of an exemplary Tpt1 enzyme in complexes with substrates, reaction intermediates, and products. At present, a crystal structure has been published of the apoenzyme form of a biochemically uncharacterized Tpt1 homolog from the archeaon *Aeropyrum pernix*[24]. In the ApeTpt1 enzyme, shown in **Figure 2.19**, the conserved side chains corresponding to RslTpt1 Arg16, His17, Arg64, and Arg119 are clustered together in a groove between the two lobes of the enzyme in the vicinity of two chloride anions modeled in the refined structure and rendered in **Figure 2.19** as yellow spheres. The chlorides, which are separated by 6.4 Å, are potential mimetics of RNA phosphates. The ApeTpt1 structure is consistent with our conclusion that Arg16, His17, Arg64, and Arg119 are indeed constituents of the Tpt1 active site.

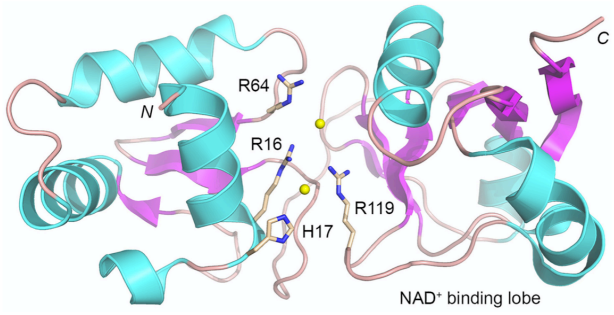


Figure 2.19. Conserved essential Tpt1 amino acids colocalize at the putative active site*. The tertiary structure of Aeropyrum pernix Tpt1 (pdb 1WFX) is depicted as a cartoon trace with cyan α helices and magenta β strands. The amino (N) and carboxl (C) termini of the protein are indicated. The C-terminal lobe is implicated in NAD^+ binding on the basis of its homolog to the NAD^+ -binding module of Diphtheria toxin (pdb 1TOX). The four conserved amino acid side chains that are essential for the activity of SceTpt1, EcoKptA, and RslTpt1 are rendered as stick models with beige carbons and numbered according to their position in RslTpt1. Two chloride anions modeled in the ApeTpt1 structure are depicted as yellow spheres.*

Finally, it is worth highlighting our observation that an internal 2′-PO_4 modification arrests 3′ RNA resection by PNPase, which is a biologically relevant agent of RNA 3′ decay in bacteria. Kierzek *et al.* (2000)[16] reported that an internal 2′ -PO_4 group did not prevent cleavage of the vicinal RNA 3′-5′ phosphodiester by snake venom phosphodiesterase. If our findings regarding PNPase extend to the 3′ exonuclease activity of the RNA exosome (which has structural homology with PNPase), then the question is raised as to whether Tpt1 enzymes might facilitate the turnover of cellular RNAs that are inappropriately spliced or repaired by a yeast-like ligase. A related question is whether and how yeast tRNAs that retain a 2′-PO_4 post-splicing (e.g., when Tpt1 is conditionally inactivated) are handled in vivo.

2.6 Materials and Methods

2.6.1 Plasmids for expression of RslTPT1 in yeast

The RslTPT1 ORF was amplified by PCR from Runella slithyformis (ATCC 49304) genomic DNA with primers that introduced a BamHI site at the translation start codon and an XhoI site flanking the stop codon. The resulting PCR product was digested with BamHI and XhoI and then inserted into yeast expression vectors p413-TPIB (CEN HIS3) and p423-TPIB (2μ HIS3), wherein expression of RslTPT1 is driven by the constitutive yeast TPI1 promoter. Alanine mutations (R16A, H17A, R64A, and R119A) were introduced into the RslTPT1 gene via a two-stage overlap extension polymerase chain

reaction using Phusion polymerase. The mutated PCR products were digested with BamHI and XhoI and inserted into BamHI/XhoI-cut p423-TPIB plasmids. The RslTPT1 ORF was sequenced completely in the p413-RslTPT1 and p423-RslTPT1 plasmids to confirm that no unwanted coding changes were introduced during PCR amplification and cloning.

2.6.2 Test of RslTPT1 function by plasmid shuffle

The S. cerevisiae tpt1Δ haploid strain YBS501 (MATa ura3–1 ade2–1 trp1–1 his3–11,15 leu2–3,11–2 can1–100 tpt1::LEU2 p360-TPT1), in which the TPT1 ORF was deleted and replaced by LEU2, is dependent for viability on the p360-TPT1 plasmid (CEN URA3 SceTPT1) (Schwer et al. 2004). YBS501 was transformed with: (i) the p413-RslTPT1 and p423-RslTPT1 plasmids carrying either a WT or Ala-mutant allele of RslTPT1; (ii) a p413-SceTPT1 plasmid (CEN HIS3 SceTPT1) plasmid as a positive control; and (iii) the empty CEN HIS3 vector as negative control. Transformants were selected at 30°C on His^- agar medium. Three individual His^+ colonies were patched to His^- agar medium and cells from each isolate were then streaked on agar medium containing 0.75 mg/mL 5-FOA (5-fluoroorotic acid). The plates were incubated at 20°C, 30°C, and 37°C. The p413-SceTPT1 and the p413-RslTPT1 and p423- RslTPT1 plasmids with wild-type RslTPT1 alleles all supported the formation of FOA-resistant colonies. In contrast, the vector and the p423-RslTPT1-Ala plasmids did not allow formation of FOA-resistant colonies after 7 d at any of the temperatures tested. Thus, the R16A, H17A, R64A, and R119A mutations were deemed lethal in vivo. Viable FOA-resistant tpt1Δ p413-SceTPT1, tpt1Δ p413-RslTPT1, and tpt1Δ p423-RslTPT1 colonies were grown in YPD-Ad (yeast, peptone, 2% dextose, 0.1 mg/mL adenine) liquid medium at 30°C to mid-log phase (A600 0.6), then diluted to attain A_{600} of 0.1, and aliquots (3 μL) of serial 10-fold dilutions were spotted on YPD agar plates and incubated at 20°C, 30°C, and 37°C.

2.6.3 Recombinant RslTpt1 proteins

The RslTPT1 ORF was amplified by PCR from R. slithyformis genomic DNA with primers designed to introduce an NdeI site immediately 5′ of the translation initiation codon and XhoI site immediately 3′ of the stop codon. The resulting PCR product was digested with NdeI and XhoI and then inserted into pET28a expression vector so as to fuse RslTpt1

to an N-terminal His6 tag. Alanine mutations (R16A, H17A, R64A, and R119A) were introduced into the expression vector either by a two-step overlap extension PCR with mutagenic primers and insertion of the mutated genes into the pET28a plasmid or by quick-change PCR. The RslTPT1 ORF was sequenced completely in the pET plasmids to confirm that no unwanted coding changes were introduced during PCR amplification and cloning. The pET28a-His_6RslTpt1 and pET28a-His_6RslTpt1-Ala plasmids were transformed into Escherichia coli BL21(DE3)-CodonPlus. Cultures (1 liter) derived from single kanamycin and chloramphenicol resistant transformants were grown at 37°C in Luria–Bertani medium containing 50 μg/mL kanamycin and 35 μg/mL chloramphenicol until the A_{600} reached 0.5 to 0.6, at which time the cultures were adjusted to 0.5 mM isopropyl-β-D-thiogalactoside and 2% (v/v) ethanol and then incubated for 16 h at 17°C with continuous shaking. Cells were harvested by centrifugation, and the pellets were stored at −80°C. All subsequent procedures were performed at 4°C. Thawed bacteria were suspended in 50 mL buffer A (50 mM Tris- HCl, pH 7.5, 1 M NaCl, 20 mM imidazole, 0.05% Triton X-100, 10% glycerol) and lysozyme was added to a final concentration of 1 mg/mL along with one protease inhibitor tablet (Roche). After mixing for 50 min, the resulting lysates were sonicated to reduce viscosity and insoluble material was removed by centrifugation for 45 min at 14,000 rpm. The soluble extracts were mixed for 1 h with 5 mL of His60 Ni Superflow resin (Clontech) that had been equilibrated in buffer A. The resin was washed twice with 50 mL of buffer A, then serially with 30 mL of buffer B (50 mM Tris- HCl, pH 7.5, 3 M KCl), and 50 mL of 40 mM imidazole in buffer C (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10% glycerol). The resin was then poured into a column and the bound material was serially step-eluted with 100 mM, 200 mM, 300 mM, and 500 mM imidazole in buffer C. The elution profiles were monitored by SDS-PAGE. The peak His_6RslTpt1-containing fractions were pooled and adjusted to 10 mM ethylenediaminetetraacetic acid (EDTA) and 10 mM dithiothreitol (DDT). The His_6RslTpt1 protein preparations were then concentrated by centrifugal ultrafiltration (to ∼15–20 mg/ mL in 10 mL) and gel-filtered through a 120-ml Superdex 200 column equilibrated with buffer D (50 mM Tris-HCl, pH 7.0, 300 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol) at a flow rate of 1 mL/min, while collecting 2-mL fractions. Peak His_6RslTpt1 fractions were concentrated by centrifugal ultrafiltration and stored at −80°C. Protein concentrations were determined by using the BioRad dye

reagent with bovine serum albumin as the standard. The yields of His_6RslTpt1 proteins were: WT, 50 mg; R16A mutant, 80 mg; H17A mutant, 65 mg; R64A mutant, 60 mg; and R119A mutant, 65 mg.

2.6.4 General methods for solid phase synthesis of oligonucleotides

Oligoribonucleotide syntheses were carried out using an ABI 3400 DNA synthesizer (Applied Biosystems) on a Unylinker (ChemGenes) solid support at a 1 μmol scale. Conventional 2′-*tert*-butyldimethylsilyl (TBDMS) ribonucleoside and 2′-acetyl levulinyl (ALE) riboadenosine 3′-phosphoramidites (0.15 M in MeCN) (ChemGenes) were used. For 2′-phosphitylation at the internal rA position, bis-cyanoethyl,-N,N-diisopropyl-phosphoramidite (0.20 in MeCN) was used. Phosphoramidites were dissolved in MeCN and activated with 5-ethylthio-1H-tetrazole (ETT) (0.25 M in MeCN). Capping was carried out by the simultaneous delivery of acetic anhydride in pyridine/tetrahydrofuran (THF) and N-methylimidazole (16% in THF) and contacting the solid support for 6 sec. Oxidation of the phosphite triester intermediates was effected with 0.1 M iodine in pyridine/H_2O/THF (20 sec); a solution of 3% trichloroacetic acid in DCM, delivered over 1.8 min, was used to deprotect DMTr groups. For 2′-phosphate-containing substrates, a solution of anhydrous TEA/MeCN (2:3 v/v) was used to remove cyanoethyl phosphate protecting groups, while a 0.5 M solution of hydrazine hydrate in pyridine/AcOH (3:2 v/v) was used to remove ALE protecting groups. All oligonucleotides were deprotected and cleaved from the solid support using an ammonium hydroxide and ethanol solution, as detailed below. TBDMS groups were removed using TREAT-HF. Crude oligonucleotides were purified via HPLC and characterized by LCMS.

2.6.5 Synthesis of 2′-PO_4 branchpoint-containing oligonucleotides

Synthesis of RNAs containing internal 2′-PO_4 moieties was carried out in two parts, as per the synthesis of branched RNAs[25]. For the sequence: 5′-UCGACUCCAA$^{2'\text{-}PO4}$ AUCGA, the first section of the oligonucleotide, 5′-A$^{2'\text{-}ALE}$ AUCGA, was synthesized on the solid support in the conventional 3′ to 5′ solid phase synthesis. The A$^{2'\text{-}ALE}$ unit in this sequence was introduced by coupling a 5′-DMTr-2′-ALE-3′-OCE phosphoramidite (0.15

M in MeCN) for 15 min (all other nucleotides were introduced using the 2′-TBDMS monomers). Subsequently, to remove cyanoethyl phosphodiester protecting groups, anhydrous TEA/MeCN (2:3 v/v) solution was passed through the solid support (20 min, repeated four times). This step ensures that all phosphate 3′-5′ linkages including that vicinal to the 2′-ALE group is in the diester form, thus preventing chain cleavage in the ensuing removal of the ALE group. After washing with MeCN (20 min) and drying the solid support over an argon stream (10 min), the synthesis columns were temporarily removed from the synthesizer and dried in vacuo for 30 min. To remove the 2′-ALE groups, the columns were returned to the synthesizer and a freshly prepared solution of 0.5 M hydrazine hydrate in pyridine/AcOH (3:2 v/v) was flowed through the columns (20 sec flow + 3.75 min sleep, repeated four times). After washing (MeCN, 10 min) and drying (Ar gas, 10 min), the solid supports were dried again in vacuo (30 min). To phosphitylate at the newly exposed 2′-OH, bis(2-cyanoethyl)-N,N-diisopropylphosphoramidite (0.20 M in MeCN) was coupled for 30 min, and then further oxidized using 0.1 M iodine in pyridine/H_2O/THF (20 sec). To complete the oligonucleotide, standard 3′ to 5′ synthesis was continued on the 5′ terminus of the growing oligonucleotide using the sequence 5′-UCGACUCCA to yield the desired 15mer oligonucleotide substrate. This synthetic route was repeated to provide the 5′-CCAA$^{2'\text{-PO4}}$AU hexamer and 5′-AA$^{2'\text{-PO4}}$A trimer substrates.

2.6.6 Oligonucleotide Deprotection

Deprotection and cleavage of oligonucleotides from the solid support was achieved by treatment with 1 mL of cold 29% aqueous ammonia/ethanol (3:1, v/v) for 48 h at room temperature. Samples were centrifuged and the supernatant was transferred to a clean 1.5 mL Eppendorf tube and vented for 30 min, chilled on dry ice, and evaporated to dryness. Removal of the 2′-TBDMS protecting groups was achieved by treatment with a 300 µL solution of NMP/Et_3N/TREAT-HF (3:4:6, v/v) for 90 min at 65°C, followed by quenching with 3 M NaOAc buffer (50 µL; pH 5.5) and precipitation of the crude oligonucleotide from cold butanol (1 mL, −20°C). Samples were chilled on dry ice for 30 min and then centrifuged. After removing the supernatant, the remaining pellet (containing RNA) was evaporated to dryness, taken up in autoclaved milli-Q water (1 mL), filtered, and quantitated by UV spectroscopy.

2.6.7 Oligonucleotide purification

Crude oligonucleotides were purified by ion exchange (IE) or reverse phase (RP) HPLC. RNAs six nucleotides and longer were purified by IE-HPLC using a Waters Protein-Pak DEAE 5PW anion exchange column (21.5 × 150 mm). A mobile phase of 1 M aqueous $LiClO_4$ in milli-Q water was used for analysis and purification (for 15mers: 0%–24% $LiClO_4$ over 35 min, 4 m/min, 60°C; for hexamers: 0%–20% $LiClO_4$ over 30 min, 4 mL/min, 60°C). Following collection of the desired peaks, fractions were combined and excess $LiClO_4$ salts were removed using Gel Pak 2.5 size exclusion columns (Glen Research). Trimers were purified by RP-HPLC using a Hamilton PRP-1 column. A mobile phase of ACN in 0.1 M triethyl-ammonium acetate (TEAA) at a flow rate of 4 mL/min was used (0%–11% over 15 min, 60°C). Purified oligonucleotides were characterized by electrospray ionization-mass spectrometry (the HPLC elution profiles and MS analyses are shown in **Supplementary Figure 2.2**) and quantitated by UV spectroscopy. Extinction coefficients were determined using the IDT OligoAnalyzer tool (www.idtdna.com/analyzer/Applications/OligoAnalyzer).

2.6.8 5′ ^{32}P-labeled 2′-PO_4 RNA substrates

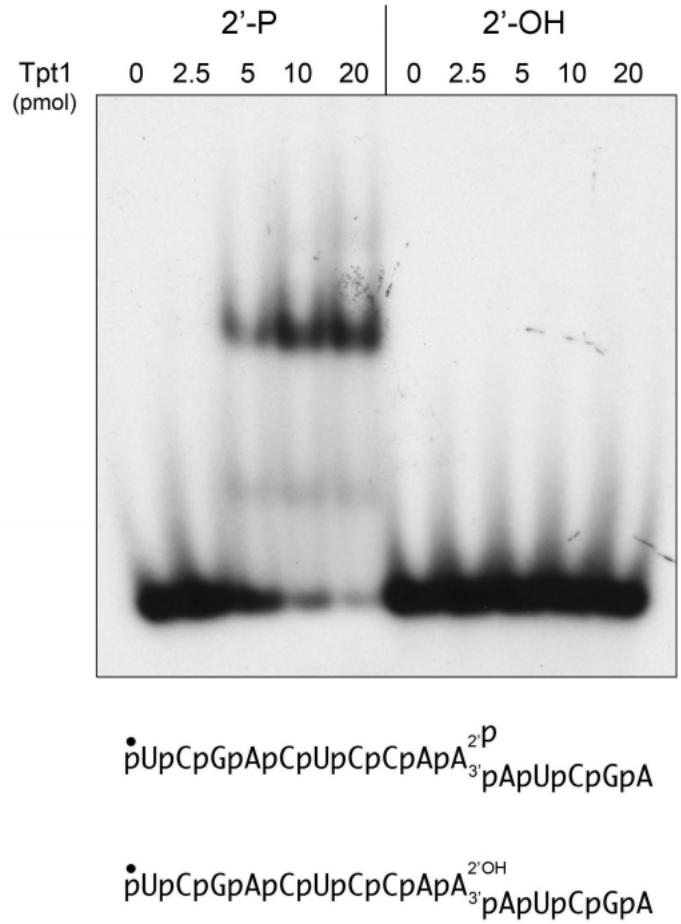
The 15-mer, 6-mer, and 3-mer RNA oligonucleotides containing an internal 2′-phosphate were 5′ ^{32}P-labeled by reaction with phosphatase-dead T4 polynucleotide kinase (Pnkp-D167N) in the presence of [γ^{32}P]ATP. The reaction was quenched with 90% formamide, 50 mM EDTA, 0.01% xylene cyanol and the radiolabeled RNAs were purified by electrophoresis through a 40-cm 20% polyacrylamide gel containing 7 M urea in 45 mM Tris-borate, 1 mM EDTA. The radiolabeled RNAs were eluted from excised gel slices, recovered by ethanol precipitation, and resuspended in 10 mM Tris-HCl, pH 6.8, 1 mM EDTA (TE) and stored at −20°C.

2.6.9 Isolation of 5′ ^{32}P-labeled 2′-P-ADPR RNA intermediate

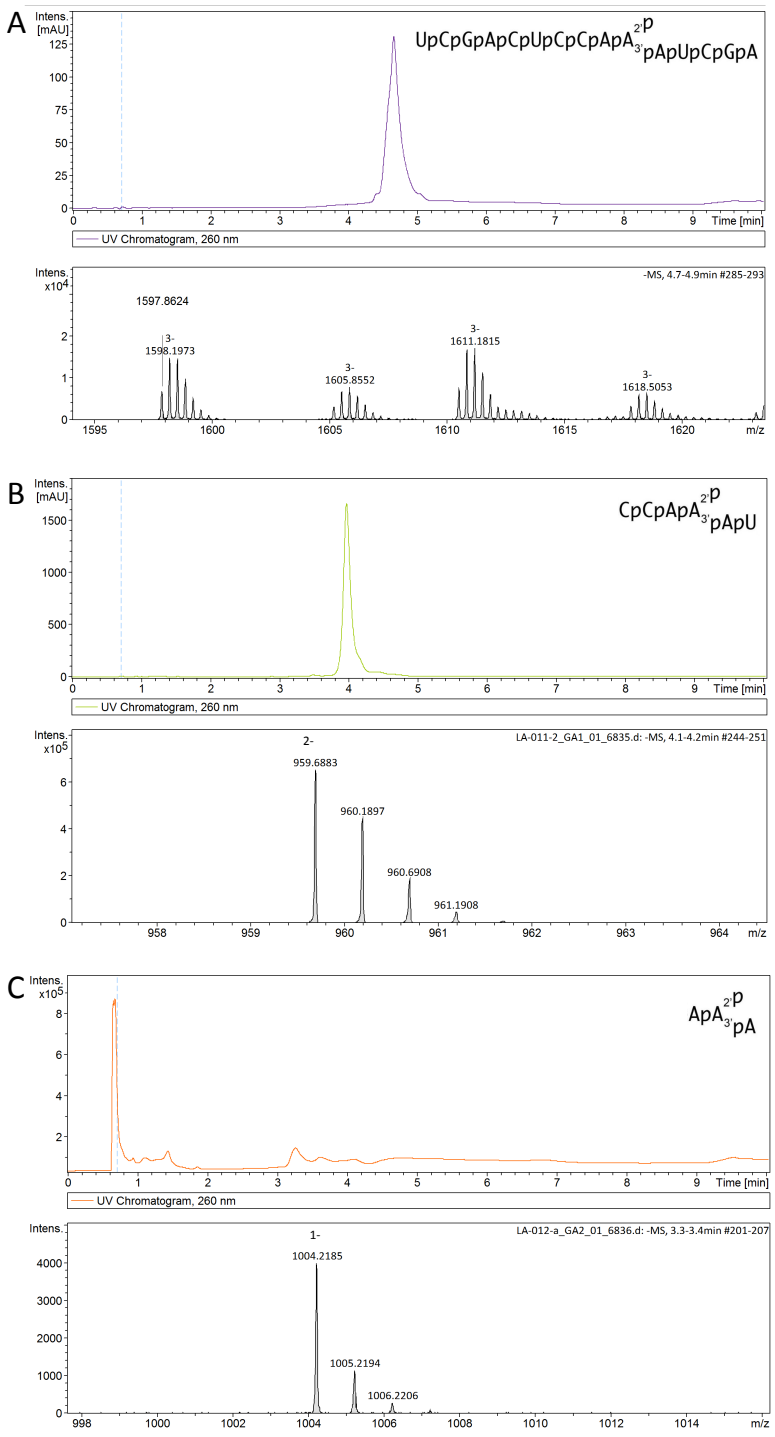
Reaction mixtures (4 × 20 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+ (Sigma), 5 mM $MgCl_2$, 100 pmol 5′ ^{32}P-labeled 6-mer 2′-PO_4 RNA, and 100 pmol RslTpt1-R64A were incubated for 20 sec at 37°C. The reactions were quenched with 3 volumes of cold 90% formamide, 50 mM EDTA, 0.01% xylene cyanol. The radiolabeled 2′-P-ADPR

RNA species was separated from substrate and product RNAs by urea-PAGE (as described above), eluted from excised gel slices, ethanol precipitated, resuspended in TE buffer, and stored at −20°C.

2.7 Supplemental Figures



Supplementary Figure 2.1. Native gel shift assay of RNA binding by RslTpt1*. Reaction mixtures (10 µl) containing 100 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 0.1 µM (1 pmol) 5′ ^{32}P-labeled 15-mer 2′-PO_4 RNA or 15-mer 2′-OH RNA (shown at bottom), 5% (v/v) glycerol, and RslTpt1 as specified were incubated at 22˚C for 10 min. The mixtures were supplemented with glycerol to 15% final concentration and then analyzed by electrophoresis (for 2 h at 110 V) through a 15-cm native 6% polyacrylamide gel in 22.5 mM Tris-borate buffer. An autoradiograph of the dried gel is shown.*



Supplementary Figure 2.2. LCMS (-ve mode) chromatograms of synthetic 2'-PO$_4$ RNA substrates. *(A) 5'-UCGACUCCAA(2'-PO$_4$)AUCGA mass calculated: 4794.64; mass observed: 4793.58 (B) 5'-CCAA(2'-PO$_4$)AU mass calculated: 1919.27; mass observed: 1919.37 (C) 5'-AA(2'-PO$_4$)A mass calculated: 1003.16; mass observed: 1004.22.*

2.8 References

1. Culver, G. M.; McCraith, S. M.; Consaul, S. A.; Stanford, D. R.; Phizicky, E. M., A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. *J Biol Chem* **1997,** *272* (20), 13203-10.
2. McCraith, S. M.; Phizicky, E. M., An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. *Journal of Biological Chemistry* **1991,** *266* (18), 11986-11992.
3. Culver, G. M.; McCraith, S. M.; Zillmann, M.; Kierzek, R.; Michaud, N.; LaReau, R. D.; Turner, D. H.; Phizicky, E. M., An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate. *Science* **1993,** *261* (5118), 206.
4. Spinelli, S. L.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Transient ADP-ribosylation of a 2′-Phosphate Implicated in Its Removal from Ligated tRNA during Splicing in Yeast. **1999,** *274* (5), 2637-2644.
5. Steiger, M. A.; Jackman, J. E.; Phizicky, E. M., Analysis of 2'-phosphotransferase (Tpt1p) from Saccharomyces cerevisiae: evidence for a conserved two-step reaction mechanism. *RNA* **2005,** *11* (1), 99-106.
6. Munir, A.; Abdullahu, L.; Damha, M. J.; Shuman, S., Two-step mechanism and step-arrest mutants of Runella slithyformis NAD+-dependent tRNA 2′-phosphotransferase Tpt1. **2018,** *24* (9), 1144-1157.
7. Spinelli, S. L.; Malik, H. S.; Consaul, S. A.; Phizicky, E. M., A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proceedings of the National Academy of Sciences* **1998,** *95* (24), 14136-14141.
8. Popow, J.; Schleiffer, A.; Martinez, J., Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences* **2012,** *69* (16), 2657-2670.
9. Sawaya, R.; Schwer, B.; Shuman, S., Structure-function analysis of the yeast NAD+-dependent tRNA 2'-phosphotransferase Tpt1. *RNA (New York, N.Y.)* **2005,** *11* (1), 107-113.
10. Munir, A.; Banerjee, A.; Shuman, S., NAD+-dependent synthesis of a 5′-phospho-ADP-ribosylated RNA/DNA cap by RNA 2′-phosphotransferase Tpt1. *Nucleic Acids Research* **2018**, gky792-.
11. Harding, H. P.; Lackey, J. G.; Hsu, H.-C.; Zhang, Y.; Deng, J.; Xu, R.-M.; Damha, M. J.; Ron, D., An intact unfolded protein response in Trpt1 knockout mice reveals phylogenic divergence in pathways for RNA ligation. *RNA (New York, N.Y.)* **2008,** *14* (2), 225-232.
12. Spinelli, S. L.; Consaul, S. A.; Phizicky, E. M., A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction. *Rna* **1997,** *3* (12), 1388-400.

13. Schwer, B.; Sawaya, R.; Ho, C. K.; Shuman, S., Portability and fidelity of RNA-repair systems. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (9), 2788-2793.

14. Cherry, P. D.; White, L. K.; York, K.; Hesselberth, J. R., Genetic bypass of essential RNA repair enzymes in budding yeast. *Rna* **2018,** *24* (3), 313-323.

15. Munir, A.; Shuman, S., Characterization of Runella slithyformis HD-Pnk, a Bifunctional DNA/RNA End-Healing Enzyme Composed of an N-Terminal 2',3'-Phosphoesterase HD Domain and a C-Terminal 5'-OH Polynucleotide Kinase Domain. *J Bacteriol* **2017,** *199* (3).

16. Kierzek, R.; Steiger, M. A.; Spinelli, S. L.; Turner, D. H.; Phizicky, E. M., The Chemical Synthesis of Oligoribonucleotides with Selectively Placed 2′-O-Phosphates. *Nucleosides, Nucleotides & Nucleic Acids* **2000,** *19* (5-6), 917-933.

17. Steiger, M. A.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Substrate recognition by a yeast 2'-phosphotransferase involved in tRNA splicing and by its Escherichia coli homolog. *Biochemistry* **2001,** *40* (46), 14098-105.

18. Lackey, J. G.; Damha, M. J., The acetal levulinyl ester (ALE) group for the 2'-hydroxyl protection of ribonucleosides and the synthesis of oligoribonucleotides. *Nucleic Acids Symp Ser (Oxf)* **2008,** (52), 35-6.

19. Lackey, J. G.; Mitra, D.; Somoza, M. M.; Cerrina, F.; Damha, M. J., Acetal Levulinyl Ester (ALE) Groups for 2′-Hydroxyl Protection of Ribonucleosides in the Synthesis of Oligoribonucleotides on Glass and Microarrays. *Journal of the American Chemical Society* **2009,** *131* (24), 8496-8502.

20. Unciuleac, M.-C.; Shuman, S., Distinctive Effects of Domain Deletions on the Manganese-Dependent DNA Polymerase and DNA Phosphorylase Activities of Mycobacterium smegmatis Polynucleotide Phosphorylase. *Biochemistry* **2013,** *52* (17), 2967-2981.

21. Unciuleac, M.-C.; Shuman, S., Discrimination of RNA from DNA by polynucleotide phosphorylase. *Biochemistry* **2013,** *52* (38), 6702-6711.

22. Banerjee, A.; Munir, A.; Abdullahu, L.; Damha, M. J.; Goldgur, Y.; Shuman, S., Structure of tRNA splicing enzyme Tpt1 illuminates the mechanism of RNA 2'-PO4 recognition and ADP-ribosylation. *Nat Commun* **2019,** *10* (1), 218.

23. Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D., Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* **2009,** *5* (8), 593-9.

24. Kato-Murayama, M.; Bessho, Y.; Shirouzu, M.; Yokoyama, S., Crystal structure of the RNA 2'-phosphotransferase from Aeropyrum pernix K1. *J Mol Biol* **2005,** *348* (2), 295-305.

25. Katolik, A.; Johnsson, R.; Montemayor, E.; Lackey, J. G.; Hart, P. J.; Damha, M. J., Regiospecific solid-phase synthesis of branched oligoribonucleotides that mimic intronic lariat RNA intermediates. *J Org Chem* **2014,** *79* (3), 963-75.

- page intentionally blank -

Chapter 3

Towards the rational design of Tpt1 inhibitors

3.1 Introduction

tRNA 2'-phosphotransferase (Tpt1) is an essential enzyme in the fungal and plant tRNA splicing pathways that removes the 2'-PO_4 at the splice junction generated by fungal and plant tRNA ligases[1]. Tpt1 catalyzes a two-step reaction whereby: (i) the internal RNA 2'-PO_4 attacks NAD^+ to form an RNA-2'-phospho-(ADP-ribose) intermediate; and (ii) transesterification of the ribose O2" to the 2'-phosphodiester yields 2'-OH RNA and ADP-ribose-1",2"-cyclic phosphate[2-6] (**Chapter 2, Figure 1**). Whereas Tpt1 homologs with verified RNA 2'-phosphotransferase activity in vitro and in vivo are present in taxa other than fungi and plants[6-9], their roles and their potential endogenous substrates in these taxa (bacteria, metazoa, and archaea) are unknown. Some Tpt1 orthologs catalyze additional NAD^+-dependent ADP-ribosyltransferase reactions – such as ADP-ribose capping of RNA and DNA 5'-phosphate ends[9] and removal of nucleic acid 2' or 3' terminal monophosphates[10] – albeit less vigorously than the canonical internal RNA 2'-phosphate removal reaction. Such findings raise interesting questions about Tpt1 substrate recognition and how it might vary among taxa.

We view Tpt1 as a promising target for discovery of antifungals, based on the fact that the enzymatic mechanism of tRNA splicing in metazoa – mediated by the RNA ligase RtcB – is entirely different from that of fungi and does not result in a junction 2'-PO_4[11]. Whereas mammals do have a homolog of Tpt1, it plays no essential role in mammalian physiology, i.e., a homozygous *tpt1-KO* mouse is viable, develops normally, and has no defects in protein synthesis[12]. This is in contrast to fungi, where Tpt1 is essential for viability, for example, in the model fungi *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* and in the pathogenic fungus *Candida albicans*[1, 13].

One can envision two ways to interdict Tpt1 activity: (i) by inhibiting transfer of ADP-ribose from NAD^+ to 2'-PO_4 RNA, thereby resulting in the accumulation of spliced tRNAs that retain the 2'-PO_4 in the anticodon loop and are thus non-functional in protein synthesis; or (ii) by inhibiting the transesterification step and trapping the ADP-ribosylated RNA intermediate, thereby generating tRNAs with bulky lesions in the anticodon loop. Valuable insights into substrate recognition and the mechanism of the transesterification step emerged from a crystal structure of *Clostridium thermocellum* Tpt1 in a product-mimetic complex with ADP-ribose-1"-PO_4 in the NAD^+ site and pAp in the RNA site[14]

and from kinetic and mutational analyses of *Runella slithyformis* Tpt1 that identified a Arg-His-Arg-Arg catalytic tetrad in the active site and pinpointed one of the arginines as specifically essential for step 2 of the Tpt1 pathway[6]. Mutating this arginine slowed the step 1 rate by only threefold while slowing the step 2 rate by a factor of 214, thus resulting in the transient accumulation of very high levels of the normally evanescent ADP-ribosylated RNA intermediate.

Having created a situation in which the intermediate can be trapped by mutating the Tpt1 enzyme, we then sought to achieve a similar outcome by chemical modifications of the 2'-PO_4 and NAD^+ RNA substrates. To gauge the tolerance of the system for modifications at positions flanking the internal 2'-PO_4, we first tested a variety of well-studied modifications, known to increase the pharmacokinetic properties of oligonucleotide drugs at those positions (**Figure 3.1,** Top). To further improve cell uptake and biodistribution, we synthesized the same series of oligonucleotides, but containing phosphorothioate (PS) backbone linkages, instead of phosphate (PO). We then focused our attention to the 2'-phosphate position itself – keeping in mind that this functional group is involved in both steps of the mechanism: first as a nucleophile when binding the RNA to NAD^+, as well as the second transesterification step which releases the mature RNA from the ADPR intermediate (**Figure 3.1,** Bottom) We also explored modifications at the 2' position of the sugar ring that would allow the RNA substrate to bind Tpt1 and trap the enzyme in an intermediary state. Additionally, the cofactor (NAD^+) also provided a unique handle for inhibition of the enzyme and was further explored.

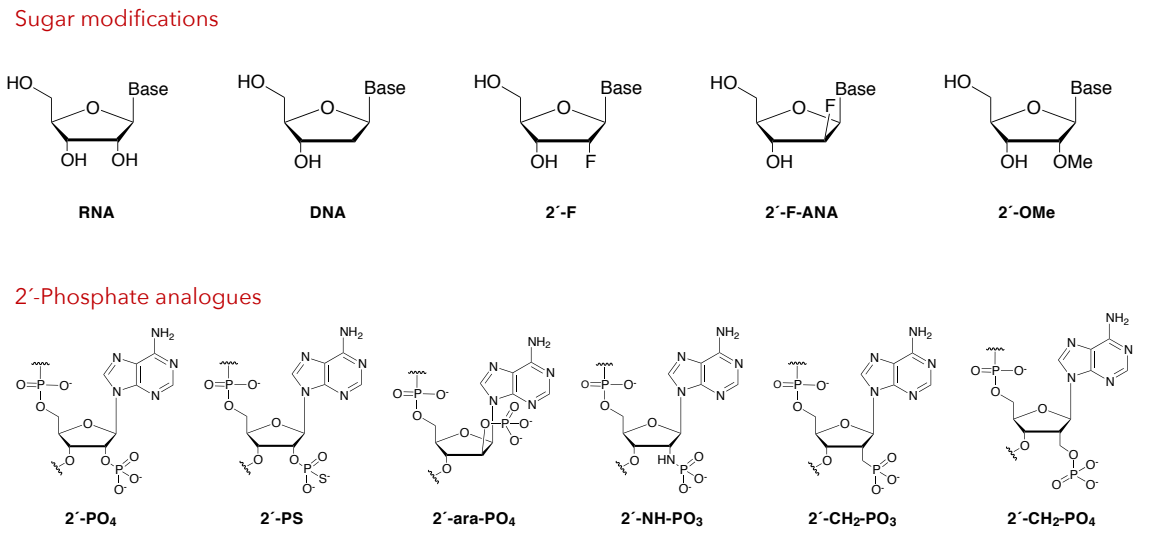


Figure 3.1. *Overview of the modifications explored in this chapter. Top Panel: Ribofuranose sugar modifications explored at positions flanking the 2'-PO₄. Bottom Panel: Unique 2'-PO₄ analogues explored (or candidates to explore) for their inhibitory effects against Tpt1.*

Our results highlight that replacement of the nucleosides flanking the 2'-PO_4 have a negligible effect on the activity of the enzyme. Substitution of the PO backbone for PS linkages reduces the activity. Replacement of the RNA ribose-2'-PO_4 nucleotide with unique 2' modifications such as 2'-PS and 2'-arabinoPO_4 result in transient inhibition. Whereas substitution of the nicotinamide mononucleotide (NMN) ribose of NAD^+ with 2'-fluoroarabinose results in permanent trapping of the Tpt1 reaction intermediate. These findings are expanded upon in this chapter. Experiments towards the synthesis of the 2'-NH-PO_3 and 2'-CH_2-PO_3 analogues are described at the end of this Chapter.

3.2 Modifications of the Sugar Showcase Tolerance of Tpt1 Substrate Recognition

The modification of the ribofuranose (sugar) in nucleic acids is a widely used method for manipulating the activity of nucleic acids. These modifications impact the local conformation and chemical reactivity of the sugar. Changes in the conformation and dynamics of the sugar moiety alter the local and potentially global structure and plasticity of nucleic acids, which in turn contributes to recognition, binding of ligands and enzymatic activity of proteins[15, 16]. In a therapeutic context, sugar modifications were designed to

reduce the entropic cost of duplex formation, increasing the thermodynamic stability of oligonucleotide duplexes, and ultimately allowing our oligonucleotides to better bind to their targets. Furthermore, such analogues are typically more resistant to biodegradation and can therefore be used advantageously as tools in antisense and siRNA research[15, 17] and CRISPR gene editing systems[18] (and as we will show in Chapters 4 and 5).

Here, we synthesized and tested a variety of Tpt1 substrate analogues containing well-studied RNA modifications at positions flanking the 2'-PO_4 position. These substrates were 6 nucleotides long, as this was determined to be an ideal size for analyzing substrate recognition and enzyme activity *via* direct assays.

3.2.1 *Runella* Tpt1 efficiently removes an internal 2'-phosphate from a DNA substrate

Previous studies of RNA repair enzymes that modify or join RNA ends have underscored the theme that their "RNA specificity" is dictated by a requirement for ribose at only a limited number of nucleotide positions within their polynucleotide substrates, as gauged by the ability to replace many or most of the ribonucleotides by deoxyribonucleotides. For example, the RNA specificity of the RNA nick-sealing enzyme T4 RNA ligase 2 arises from a requirement for at least two ribonucleotides immediately flanking the 3'-OH of the nick; the rest of the nicked duplex can be replaced by DNA[19]. In the case of the plant tRNA ligase (AtRNL), which acts *via* the same chemical pathway of healing and sealing as fungal tRNA ligase Trl1, a single ribonucleoside-2',3'-cyclic-PO_4 moiety enables AtRNL to efficiently splice an otherwise all-DNA strand[20]. *Clostridium thermocellum* Hen1 (CthHen1) is an RNA ribose-2'-O-methyltransferase that marks the 3' terminal nucleoside of broken RNAs and protects ligated repair junctions from iterative damage by transesterifying endonucleases[21]. CthHen1 is adept at methylating a polynucleotide composed of 23 deoxyribonucleotides and one 3' terminal ribonucleotide, signifying that it has no strict RNA specificity beyond the 3' nucleoside[22].

To query the extent of the RNA requirement for Tpt1, we synthesized and then 5' ^{32}P-labeled an analog of the 6-mer 2'-PO_4 oligonucleotide substrate in which the five nucleotides flanking the 2'-PO_4 branched ribonucleotide were replaced by deoxynucleotides (**Figure 3.2**). The 2'-PO_4 RNA and DNA substrates (at 0.2 μM concentration) were reacted for 30 min with increasing amounts of *Runella* Tpt1 in the

presence of 1 mM NAD^+. The product distributions were quantified and are plotted as a function of input RslTpt1 (**Figure 3.2**A,B). The salient points were that *Runella* Tpt1 is virtually equally adept at removing an internal 2'-PO_4 from RNA and DNA polynucleotides, without accumulating significant amounts of the 2'-phospho-ADP-ribosylated intermediate at limiting enzyme concentrations.

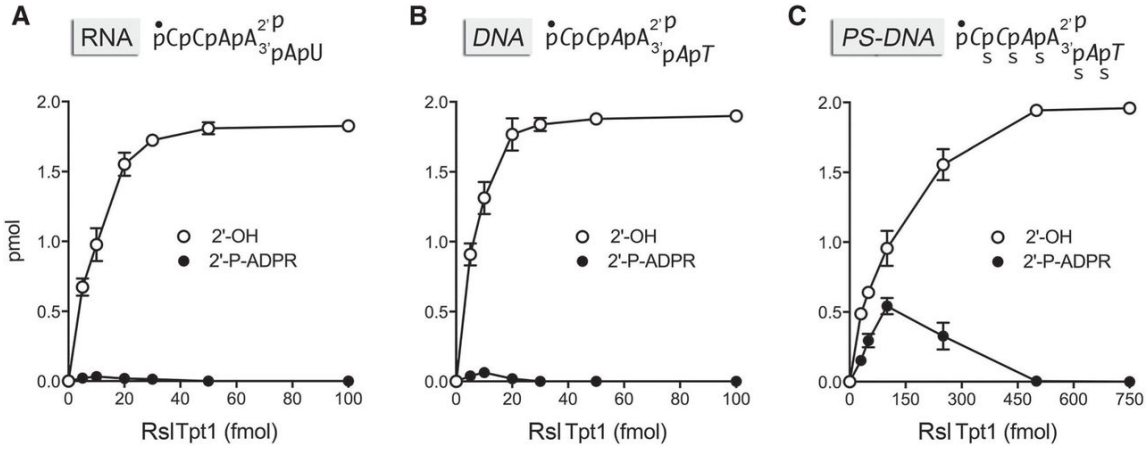


Figure 3.2. *Runella Tpt1 efficiently removes an internal DNA 2'-phosphate. Reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM (2 pmol) 5'-^{32}P-labeled 6-mer 2-PO_4 RNA (panel A), DNA (panel B), or PS-DNA (panel C) substrates, and Runella Tpt1 as specified were incubated at 37˚C for 30 min. The reaction products were analyzed by urea-PAGE. The extents of formation of the 2'-OH product and the ADP-ribosylated intermediate are plotted as a function of input RslTpt1. Each datum is the average of three independent titration course experiments ±SEM.*

3.2.2 Effect of 2'F-arabinose (FANA) sugars flanking the 2'-phosphate RNA branchpoint

Arabinonucleic acid (ANA) is the C2'-epimer of RNA in which the chirality of the 2' position of the pentose sugar is switched so that the C2'–OH group points up, i.e., in the direction opposite that of the 2' hydroxyl of a ribose. 2'-Fluoroarabinonucleic acid (2'-FANA) is the C2'–F analog of ANA, and their structure (e.g., sugar pucker) closely mimics that of DNA[23]. We tested *Runella* Tpt1 activity on a 2'-FANA version of the 6-mer 2'-PO_4 oligonucleotide substrate in which the five nucleotides flanking the 2'-PO_4 branched ribonucleotide were replaced by 2'-fluoro-arabinonucleotides. RslTpt1 removed the 2'-PO_4 from the 2'-FANA substrate, albeit with ~7-fold lower specific activity compared to the 6-mer 2'-PO_4 RNA control substrate (**Figure 3.3**A,B). The 2'-phospho-ADP-ribosylated intermediate accumulated to an extent of 4% of total labeled RNA at limiting enzyme

concentrations (**Figure 3.3**B). Thus, RslTpt1 activity is tolerant of DNA in lieu of RNA adjacent to the 2'-PO_4 branch but is less tolerant of a 2'-FANA nucleic acid.

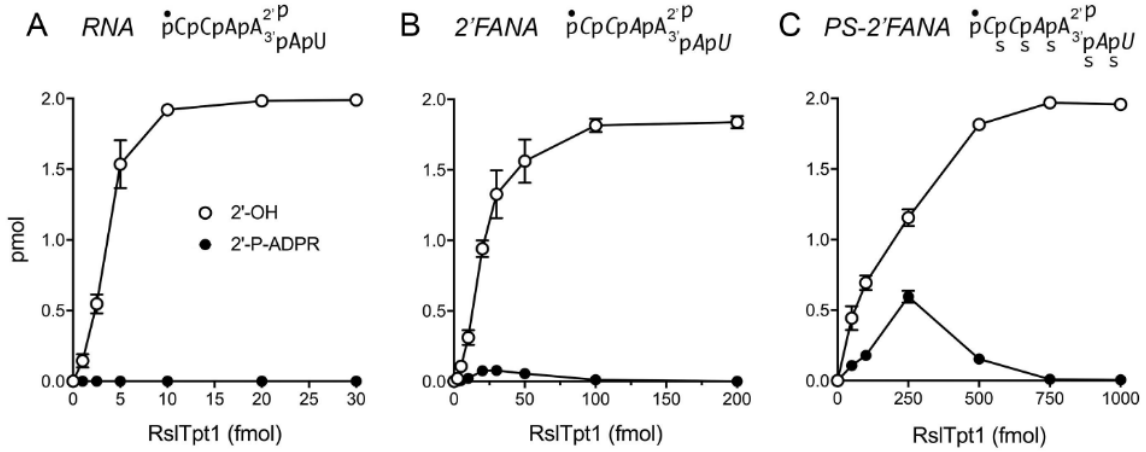


Figure 3.3. *Effect of 2'F-arabinose sugars flanking the 2'-phosphate RNA branchpoint. Reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM (2 pmol) 5'-^{32}P-labeled 6-mer 2-PO_4 RNA (panel A), 2'FANA (panel B), or PS-2'FANA (panel C) substrates, and Runella Tpt1 as specified were incubated at 37°C for 30 min. The reaction products were analyzed by urea-PAGE. The extents of formation of the 2'-OH product and the ADP-ribosylated intermediate are plotted as a function of input RslTpt1. Each datum is the average of three independent titration course experiments ±SEM.*

We also synthesized 6mer substrates with 2'-fluoro as well as 2'-OMe modifications at positions flanking the 2'-phosphate (**Table 3.1**, Section 3.8). While enzyme experiments for these substrates against Tpt1 are still in progress, we do not expect much different results than the ones shown above. Since the DNA 6mer performs equally as well as the RNA 6mer, we understand that Tpt1 does not discriminate the conformation of the nucleotides and thus both RNA and DNA-like modifications should be tolerated. The synthesis and characterization of these substrates is described in the experimental.

3.3 Modifications of the Phosphate Backbone Slightly Reduce Tpt1 Activity

3.3.1 Effect of 3'-5' phosphorothioate backbone linkages flanking the 2'-phosphate RNA branchpoint

Phosphorothioate (PS) modifications of the 3'-5' phosphodiester backbone of DNA and 2'-FANA oligonucleotides provide stability against nuclease digestion while allowing for RNase H-mediated cleavage of complementary RNAs[24, 25]. As such, PS-DNA and PS-

2'FANA oligos are effective tools for antisense interference manipulations of gene expression *in vivo*[26]. Here we introduced diastereomeric PS modifications at all five 3'-5' phosphodiesters of the 6-mer 2'-PO_4 DNA and 2'FANA substrates and tested them for activity with *Runella* Tpt1 (as each PS linkage results in a pair of diastereoisomer, each 6-mer is a mixture of 256 PS diastereoisomers). These PS-modified 2'-PO_4 oligos could be completely converted to 2'-OH products by (**Figure 3.2**C and **Figure 3.3**C). However, the enzyme titration curves were shifted significantly to the right *vis-à-vis* the DNA and 2'-FANA substrates with unmodified phosphodiester backbones (**Figure 3.2**B and **Figure 3.3**B). *Runella* Tpt1 specific activity was reduced by 11-fold and 7-fold, respectively, by the PS modifications of DNA and 2'-FANA substrates. Moreover, there was substantial accumulation of the 2'-phospho-ADP-ribosylated intermediate at sub-saturating enzyme levels, to an extent of 27% and 30% of total nucleic acid for the PS-DNA and PS-2'-FANA substrates, respectively (**Figure 3.2**C and **Figure 3.3**C).

3.4 Modifications of the 2'-PO_4 Moiety Directly Impact Tpt1 Activity

To be able to impact the activity of the enzyme directly, it was imperative to look at other modifications of the 2'-phosphate moiety. We designed a variety of modifications at this position which would influence both the sterics (2'-araPO) and the electronics (2'-PS) of the reaction mechanism. These results are summarized in this section. We would like to note that we synthesized a substrate with an additional nucleotide at the 2'-phosphate moiety (LA-2'-riboG6) by adding a reverse amidite following ALE deprotection. However, Tpt1 did not show any activity against this substrate, supposedly because the additional nucleotide was too bulky to allow the substrate to bind to the active site.

3.4.1 Effect of C2'-PS at the 2'-phosphate RNA branchpoint on Tpt1 dephosphorylation

We first started with the simplest modification and most obvious PO_4 substitution. Incorporating a single non-bridging phosphorothioate at the 2'-branchpoint of a sixmer RNA required just a simple modification of the synthetic method described in Chapter 2, where we sulfurize following the addition of the phosphitylating reagent, instead of

oxidizing. Our initial thoughts on this modification led us to believe that this substrate would react faster with NAD^+, given the nucleophilicity of S^- is greater than that of O^-. However, given that the van der Waals radius of S is larger than O (180 pm vs. 152 pm), we also expected this modification to perturb the system to some extent. A kinetic assay of the reaction with this substrate under conditions of enzyme excess (0.2 µM 2'-PS RNA; 1 µM RslTpt1) reveals an apparent step 1 rate constant for 2'-P(S) RNA of 20.6 ± 1.44 min^{-1}, which is identical to that for the unmodified 2'P RNA of 21.7 ±0.66 (**Figure 3.4**). Perhaps because the 2'-PS monoester is not chiral (in contrast to the PS diester backbone linkages), it is possible for the C2'-O-P bond to rotate and find a rotamer where the sulfur does not clash with anything in the active site.

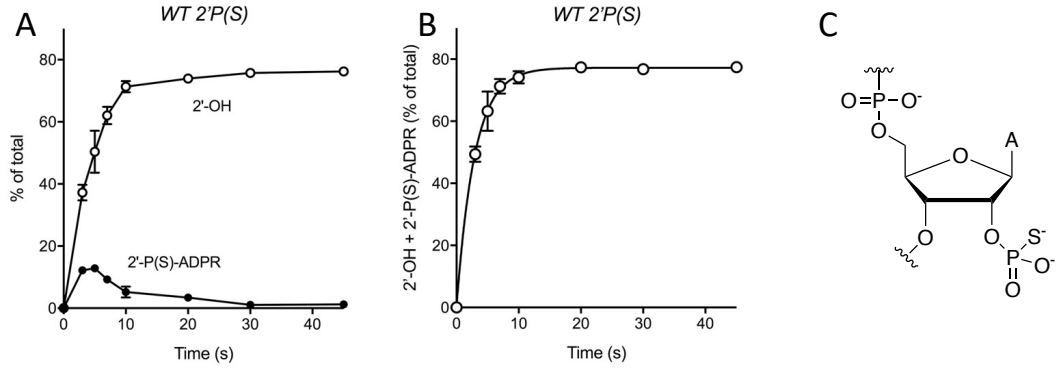


Figure 3.4. *Effect of 2'-phosphorothioate (PS) analogue at internal position. Reaction mixtures (100 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM 5' ^{32}P-labeled 6-mer 2'-PO_3S substrate, and 1 µM wild-type RslTpt1 were incubated at 37˚C. The reactions were initiated by adding enzyme to a pre-warmed reaction mixture. Aliquots (10 µl, containing 2 pmol of RNA) were withdrawn at the times specified and quenched immediately with 3 volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE and quantified by scanning the gels. The extents of formation of the 2'-OH product and the ADP-ribosylated intermediate, separately (A) or combined (B), are plotted as a function of reaction time. Each datum is the average of three independent titration or time-course experiments (±SEM). (C) Chemical structure of the 2'-PS RNA branchpoint.*

3.4.2 Effect of an arabinose sugar at the 2'-phosphate RNA branchpoint

We synthesized a 6-mer RNA with an internal arabinose-2'-phosphate moiety (ANA 2'-PO_4) shown in **Figure 3.5**D. A 30 min reaction of 0.2 µM (2 pmol) 5' ^{32}P-labeled ara-2'-PO_4 RNA with increasing concentrations of *Runella* Tpt1 in the presence of 1 mM NAD^+ resulted in the initial accumulation of very high levels of 2'-phospho-ADP-ribosylated RNA intermediate, to a peak level of 70% of total RNA at 0.05 to 0.1 µM RslTpt1 (0.5 to 1 pmol), followed by its progressive conversion to 2'-OH RNA product as enzyme concentration was increased up to 0.1 µM (10 pmol) (**Figure 3.5**A). A kinetic analysis of the reaction under conditions of enzyme excess (0.2 µM ara-2'-PO_4 RNA; 1 µM

RslTpt1) is shown in **Figure 3.5**B and revealed a clear precursor-product relationship whereby the ADP-ribosylated RNA intermediate accumulated steadily at early times and comprised 61% of total RNA at 1.5 min before declining steadily thereafter as all of the RNA was converted to 2'-OH product. The data fit well by nonlinear regression in Prism to a unidirectional two-step mechanism with apparent step 1 and step 2 rate constants of 1.50 ± 0.07 min^{-1} and 0.35 ± 0.01 min^{-1}, respectively. Comparison of these data to reaction rates of *Runella* Tpt1 with the 6-mer ribose-2'-PO_4 branched RNA (step 1 and step 2 rate constants of 21 min^{-1} and 221 min^{-1}, respectively; Chapter 2) highlights the profound impact of an arabinose sugar at the 2'-PO_4 branchpoint, which slows step 1 by a factor of 14 but exerts a much stronger 630-fold effect on step 2.

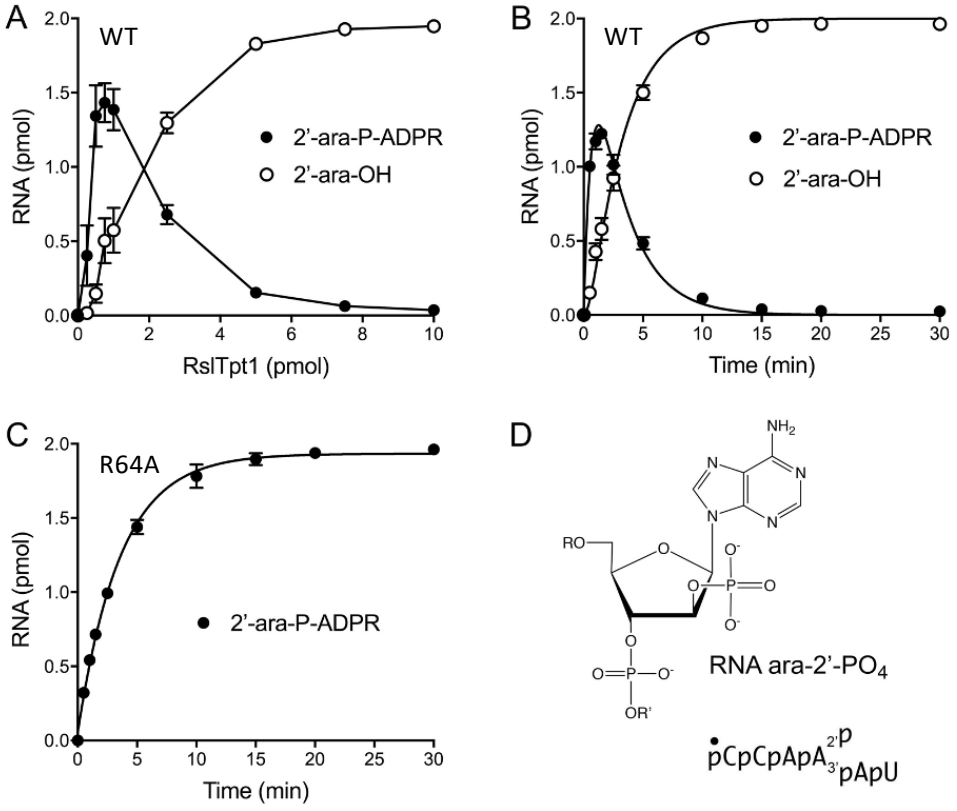


Figure 3.5. Effect of an arabinose sugar at the 2'-phosphate RNA branchpoint. *(A) Reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD$^+$, 0.2 µM (2 pmol) 5'-^{32}P-labeled 6-mer ara-2'-PO_4 substrate (shown in panel D), and 0, 0.25, 0.5, 0.75, 1 2.5, 5, 7.5, or 10 pmol RslTpt1 were incubated at 37˚C for 30 min. (B,C) Reaction mixtures (100 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD$^+$, 0.2 µM 5'-^{32}P-labeled 6-mer ara-2'-PO_4 substrate, and 1 µM wild-type RslTpt1 (panel B) or RslTpt1-R64A (panel C) were incubated at 37˚C. The reactions were initiated by adding enzyme to a pre-warmed reaction mixture. Aliquots (10 µl, containing 2 pmol of RNA) were withdrawn at the times specified and quenched immediately with 3 volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE and quantified by scanning the gels. The extents of formation of the 2'-OH product and the ADP-ribosylated intermediate are plotted as a function of input RslTpt1 (panel A) or reaction time (panels B and C). Each datum is the average of three independent titration or time-course experiments (±SEM). (D) Chemical structure of the arabinose-2'-PO_4 RNA branchpoint.*

Because we do not yet have a structure of a Tpt1 enzyme in complex with an RNA containing a 2'-phosphate, we are unable to convincingly model how the arabinose-2'-PO_4 modification affects substrate binding and reaction chemistry. However, we can guess from the available Tpt1 structure with product-mimetic ligands that the extensive network of electrostatic interactions of essential Tpt1 amino acid side chains with the transferred phosphate moiety[14] would be perturbed if the 2'-phosphate were oriented "up" (beta-face) from the arabinose sugar instead of "down" (alpha-face) from ribose. We predict this would misalign the 2'-phosphate nucleophile with respect to the ribose C1" atom of NAD^+ and the departing nicotinamide during step 1 catalysis. From the rate effects, we suspect that an "up" orientation of the arabinose-2'-phosphodiester to ADP-ribose even more deeply misalign the geometry of the ADP-ribose O2" nucleophile and the RNA arabinose O2' leaving group during step 2 catalysis.

3.4.3 Synthesis of 2'-NH-PO_3 adenosine derivative

Inspired by the previous work done in our lab by Tago *et al.* regarding the synthesis of inhibitors of the RNA lariat debranching enzyme (Dbr1), we decided to pursue a bridging phosphoramidite branchpoint (2'-NH-PO_3)[27]. In line with what we noted above, a crystal structure of the Dbr1 in complex with 2',5'-PS-branched RNA suggested that the sulfur atom of PS, which has a larger van der Waals radius than that of oxygen, sterically hinders the accommodation of the phosphorus linkage within the active site[27, 28]. Naturally, this prompted them to move towards a phosphoramidite bridge instead, as the size of the N more closely resembles that of the natural O atom. Furthermore, *and highly relevant to the Tpt1 reaction mechanism*, the PN is expected to be more resistant to hydrolysis (step 2), since under neutral or basic conditions, RNH^- is a poorer leaving group relative to RO^-. Thus, we expect this modification to still act as a substrate for Tpt1, albeit significantly more resistant to cleavage than the natural substrate. At the very least, it will compete with the natural substrate for the active site of the enzyme, thus deactivating it. These types of inhibitors (an alternate inhibitor as opposed to a full inhibitor) can be appealing for therapeutic applications where complete inhibition of particular enzymes may be lethal[29]. As an aside, replacing the O with N should steer the sugar pucker towards C2'-endo, as

opposed to the C3'-endo orientation of the natural RNA substrate[30]. It will be interesting to see if this will play a role in the mechanistics of the reaction.

The scheme below summarizes how we were able to obtain an adenosine C2'-NH-Fmoc 3'-*O*-phosphoramidite that we can then incorporate into an automated solid phase synthesis cycle to obtain the desired RNA sixmer with the desired C2'-phosphoramidate ($NH\text{-}PO_3$) moiety (**Figure 3.6**). We start with commercially available Ara-Adenosine (**1**) and simultaneously protect the 5' and 3' hydroxyls using 1,1,3,3-tetraisopropyl-1,3-dichlorodisiloxane (otherwise known as Markiewicz reagent) under basic conditions. Then the remaining 2' hydroxyl was activated using triflate chloride to produce a very effective leaving group for the subsequent reaction. Lithium azide was used as a nucleophile to displace the triflate and give compound **4**. While the free amino group on the adenosine base had not caused any problems until this point, it was now important to protect it to avoid undesired side reactions further on in our scheme, as we will be generating another free NH_2 group in the subsequent step. This protection was carried out using benzoyl chloride to give the N6-benzoyl adenosine derivative. Then, the azide of compound **5** was reduced to a secondary amine (**6**) via hydrogenolysis. At this point, we can protect the amine with an orthogonal protecting group like fluorenylmethoxycarbonyl (Fmoc), which is very amenable to amines, in order to create a handle for downstream phosphitylation. Following N-Fmoc and N-Bz protections, the silyl bridge was cleaved using triethylamine trihydrogen fluoride (TREAT-HF), to yield compound **8**, which upon dimethoxytritylation (DMTr-Cl) afforded compound **9**. Finally, using 2-cyanoethyl N,N-diisopropylchlorophosphoramidite, we were able to generate the desired C2'-NH-Fmoc 3'-phosphoramidite derivative (**10**).

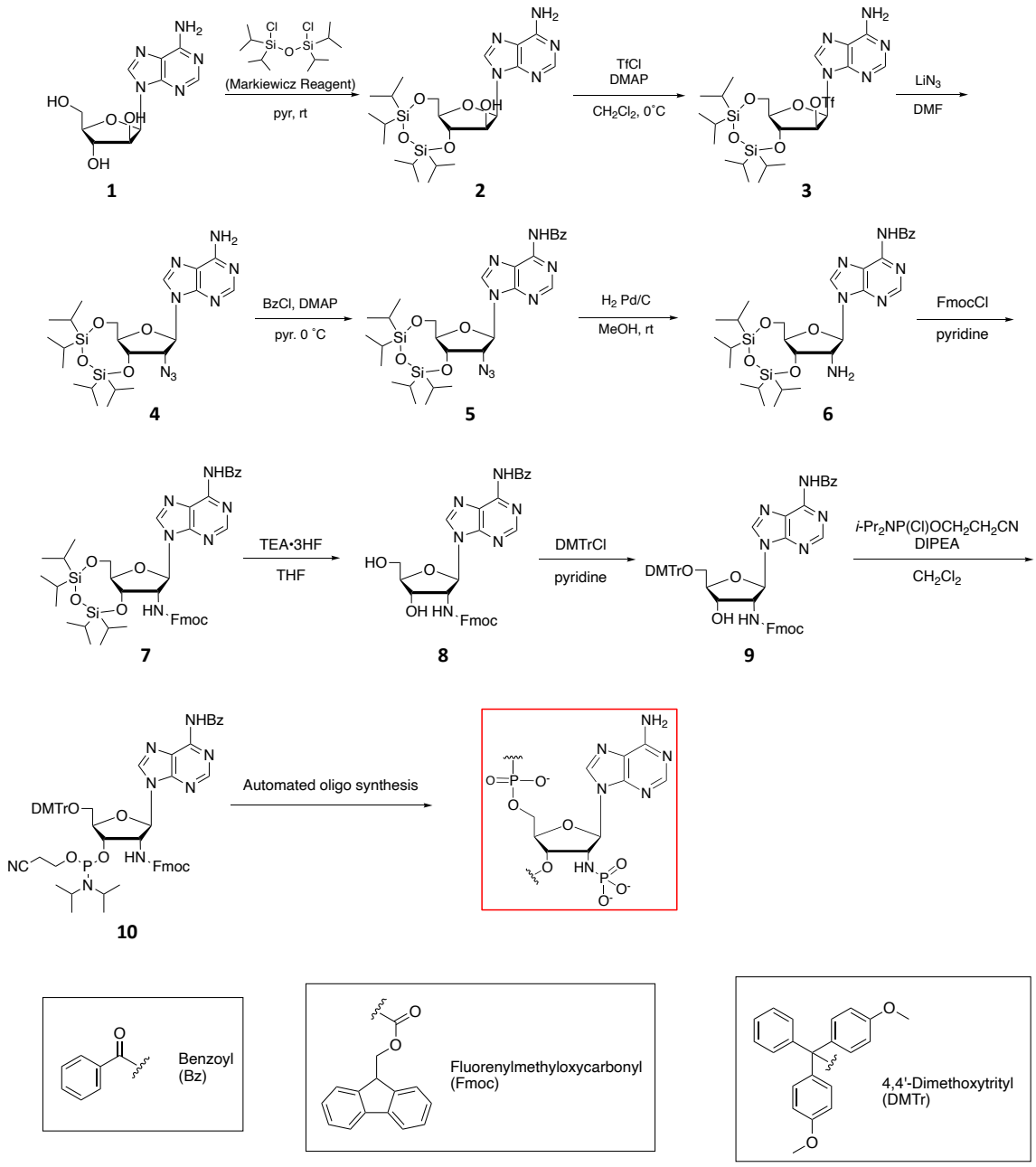


***Figure 3.6**. Scheme of synthetic route towards C2'-NH-PO$_3$ phosphoramidite.*

3.4.4 Synthesis of 2'-CH$_2$-PO$_3$ adenosine derivatives

Beyond the 2'-PO analogues that we have tested so far, we sought to design an analogue that would inhibit Tpt1 activity entirely. While the mind is free to envision a multitude of potential functionalities at the 2' that would inhibit the Tpt1 reaction mechanism, we are limited to a few reasonable functionalities for which we can outline a feasible synthetic route. We identified a methylene phosphonate (C2'-CH$_2$PO$_3$) analogue

which, when incorporated into RNA, we hypothesize will inhibit Tpt1 activity. This is a very unique analogue which has not been explored as a nucleic acid modification, and it is ideal for this purpose because it contains the same number of atoms as the native RNA substrate. However, simply replacing the C2' oxygen atom with CH_2 will eradicate the possibility for the transesterification reaction to occur, thus trapping the intermediate and blocking the enzyme. Additionally, as the methylene group is electron donating, it should make the PO_3 group more nucleophilic, thereby increasing the rate of Step 1 (Chapter 2, Figure 2.1), allowing this potential inhibitor to outcompete the natural substrate for the enzyme.

The synthesis of such an analogue proved very difficult, and several potential synthetic routes were explored before arriving at the one we adopted (**Figure 3.7**). Similar to the synthesis of compound **10**, we protected ara-adenosine with Markiewicz reagent to afford (**2**). Reaction of (**2**) with triflate chloride afforded the C2'-triflate derivative (**3**), which readily undergoes substitution with a variety of nucleophiles – a feature that we will continue to exploit when preparing other potential substrates (or inhibitors) of Tpt1 or similar enzymes. Compound (**3**) proved to be stable to silica gel column chromatography; furthermore it could be stored under argon at room temperature for several weeks without any detectable degradation. For the purpose of synthesizing the C2'-phosphonate (**13**), we reacted (**3**) with organolithium species (**12**), which we synthesized from commercially available diethyl methylphosphonate (**11**).

It is important to note that upon characterization of compound (**13**) by MS, we observed a mass of [M+39]. Initially, we thought that this was a potassium adduct [M+K] (+39). In fact, thorough NMR spectral characterization was fully consistent with the assigned structure. (**Supplementary Figure 3.1 to Supplementary Figure 3.6**). However, high resolution MS analysis is not consistent with the formation of an [M+39] adduct, but rather a compound with the formula $C_{27}H_{50}N_5NaO_8PSi_2$ (M = 682.2842). The desired product's chemical formula is $C_{27}H_{50}N_5O_7PSi_2K$ (M = 682.2623), indicating the presence of one additional oxygen atom, and a sodium (Na) adduct. Nevertheless, we decided to carry this compound (**13**) forward keeping the inconsistent MS vs HRMS data. Thus, the supposed compound (**13**) was protected at N-6 of the nucleobase using benzoyl chloride to give **14**. Removal of the 3',5' silyl bridge, followed by the installation of the 5'-

DMTr and 3'-phosphoramidite groups afforded (**17**). We are planning to use (**17**) in our synthesis methods to produce the desired 2'-CH_2-PO_3 RNA analogue. We remain confident that this modification will yield a suitable substrate for Tpt1, thus trapping the enzyme and yielding RNA-Tpt1 adducts suitable for the co-crystallization.

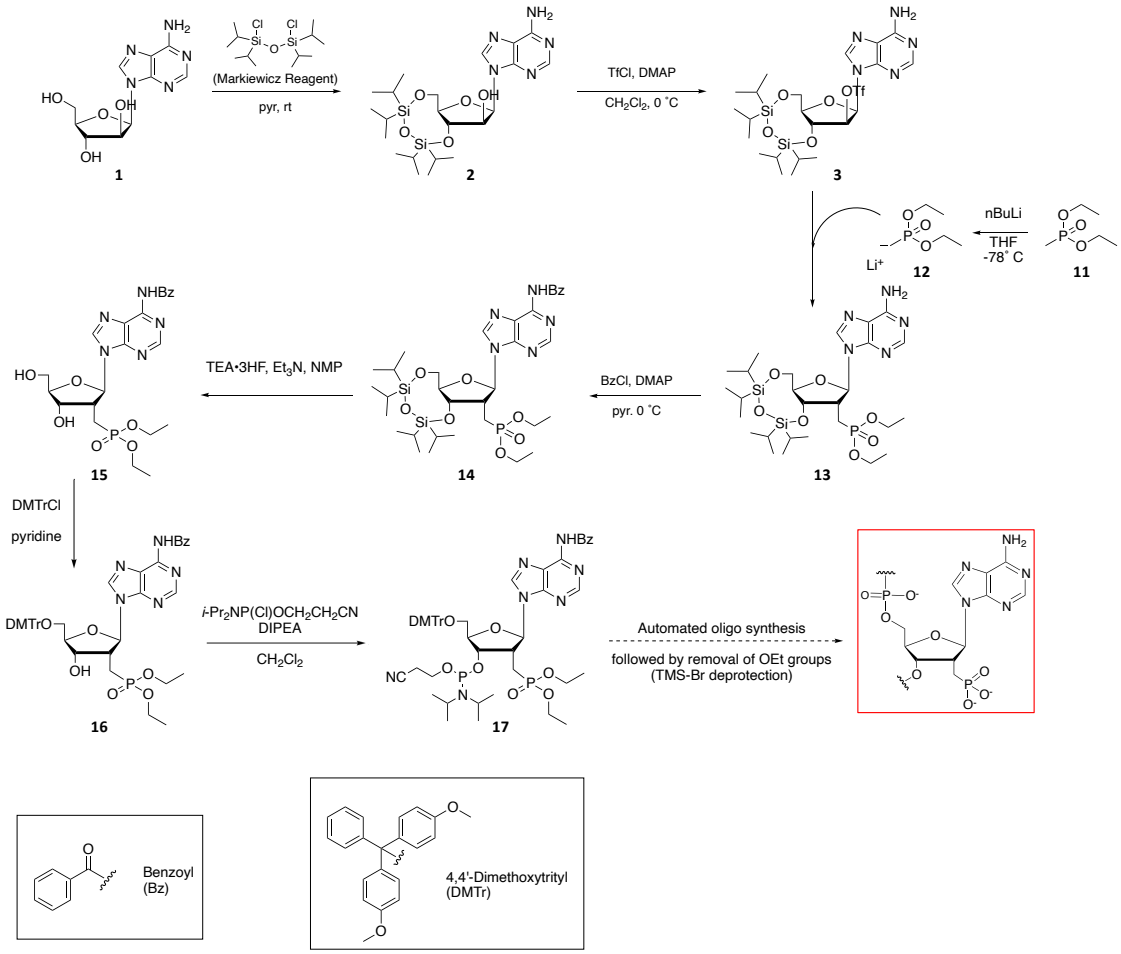


Figure 3.7. *Scheme for synthetic route towards C2'-CH_2-PO_3 phosphoramidite.*

3.5 Modifications of NAD^+ Affect Tpt1 Activity

3.5.1 Utilization of ara-2''-fluoro NAD^+ by Tpt1 enzymes traps the ADP-ribosylated RNA

The ara-2"F analog of NAD^+ (shown in **Figure 3.8**A) could, in principle, be able to support the first step in the Tpt1 pathway but would be unable to undergo transesterification in the second step for the lack of a O2" nucleophile. Because the effect of an arabinose sugar in lieu of ribose on the nicotinamide nucleoside moiety of NAD^+ on substrate utilization by Tpt1 had not been queried previously, we surveyed four different Tpt1

enzymes (each at 0.5 μM concentration) for activity in the presence of 0.2 μM 5' ^{32}P-labeled 6-mer 2'-PO_4-branched RNA oligonucleotide (shown in **Figure 3.8**B) and either 50 μM NAD^+ or ara-2"F-NAD^+. The Tpt1 proteins were from *Runella slithyformis* (a bacterium), *Clostridium thermocellum* (a bacterium), *Chaetomium thermophilum* (a fungus), and *Homo sapiens* (human). After a 30 min incubation at 37˚C, the reaction products were resolved by urea-PAGE and visualized by autoradiography (**Figure 3.8**A). Control reactions showed that each of the Tpt1 enzymes converted all of the input 2'-PO_4 RNA substrate to a slower-migrating 2'-OH RNA product in the presence of 50 μM NAD^+. When provided with 50 μM ara-2"F-NAD^+, the *Runella* and *Clostridium* enzymes converted nearly all of the substrate into an even more slowly migrating species that corresponds to an ADP-ribosylated RNA[6], which in this case is an RNA-2'-phospho-(ADP-fluoroarabinose) dead-end product of step 1 of the Tpt1 pathway (**Figure 3.8**A). *Chaetomium* Tpt1 also formed the dead-end product in the presence of 50 μM ara-2"F-NAD^+, though the extent of substrate conversion was lower. By contrast, human Tpt1 effected no detectable reaction of the 2'-PO_4 RNA substrate in the presence of 50 μM ara-2"F-NAD^+ (**Figure 3.8**). We surmise that Tpt1 enzymes from different sources may vary in their sensitivity to the arabinose sugar modification of NAD^+, but the utilization of ara-2"F-NAD^+ as a substrate by Tpt1 does indeed result in trapping of the step 1 reaction product.

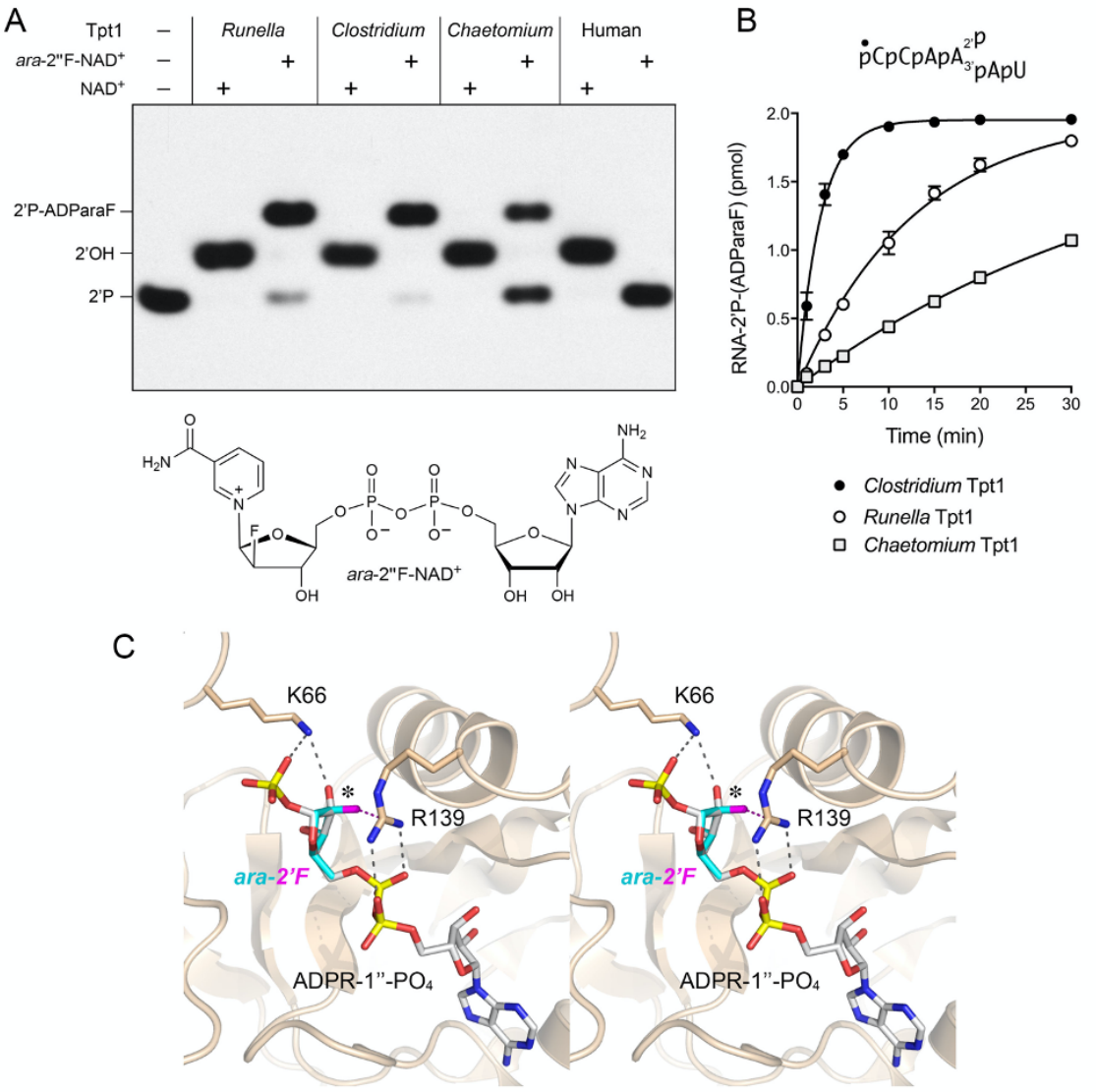


Figure 3.8. Utilization of ara-2"-fluoro NAD+ by Tpt1 enzymes traps the ADP-ribosylated RNA. *(A) Tpt1 reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 6-mer 2′-PO$_4$ RNA substrate (shown in panel B), either 50 µM NAD+ or ara-2"F-NAD+ (where indicated by +), and 0.5 µM (5 pmol) Runella slithyformis, Clostridium thermocellum, Chaetomium themophilum, or human Tpt1 were incubated at 37˚C for 30 min. The products were analyzed by urea-PAGE and visualized by autoradiography. The identities of the radiolabeled RNAs are indicated on the right. The chemical structure of ara-2"F-NAD+ is shown at bottom. (B) Reaction mixtures (100 µl) containing 100 mM Tris-HCl, pH 7.5, 50 µM ara-2"F-NAD+ 0.2 µM 5′ ^{32}P-labeled 6-mer ara-2′-PO$_4$ substrate, and 1 µM Clostridium, Runella, or Chaetomium Tpt1 were incubated at 37˚C. Aliquots (10 µl, containing 2 pmol of RNA) were withdrawn at the times specified and quenched immediately with 3 volumes of cold 90% formamide, 50 mM EDTA. The reaction products were analyzed by urea-PAGE and quantified by scanning the gels. The extents of formation of the RNA-2'P-(ADParaF) product are plotted as a function of reaction time. Each datum the average of three independent time-course experiments (±SEM). The data were fit to a one-phase association model in Prism. (C) Stereo view of the structure of Clostridium Tpt1 (beige cartoon trace) in complex with ADPR-1"-PO$_4$ (stick model with gray carbons) in the NAD+ site (from pdb 6E3A). The figure highlights atomic contacts of Lys66 with the ADP-ribose-O2′ and Arg139 with the a and b phosphates of ADPR/NAD+ (denoted by black dashed lines). A structure of ara-2'F (stick model with cyan carbon and magenta fluorine atoms; excerpted from pdb 3I9M) is superimposed on the ribose of ADPR-1"-PO$_4$. The asterisk denotes the opposite orientations of the ribose hydroxyl and arabinose fluorine substituents. Close contact (2.7 Å) of the modeled 2'F with Arg139 is indicated by the magenta dashed line.*

The temporal profiles of the reactions of *Runella*, *Clostridium*, and *Chaetomium* Tpt1 (0.5 μM) with 2'-PO_4 RNA substrate (0.2 μM) and 50 μM ara-2"F-NAD^+ are shown in **Figure 3.8**B. The data revealed a kinetic hierarchy whereby the apparent rate constant of the *Clostridium* enzyme for transfer of ADP-fluoroarabinose to the RNA 2'-phosphate (0.40 ± 0.022 min^{-1}) was 5-fold faster than that of *Runella* Tpt1 (0.077 ± 0.0062 min^{-1}) and 16-fold faster than *Chaetomium* Tpt1 (0.025 ± 0.0028 min^{-1}).

Although the yield of 2'-phospho-(ADP-fluoroarabinose) 6-mer RNA was high for *Runella* and *Clostridium* Tpt1, the apparent step 1 rate constant of *Runella* Tpt1 with 50 μM ara-2"F-NAD^+ was at least two orders of magnitude slower than the step 1 rate constant observed with 1 mM NAD^+ [6]. Clues to the slowing effect of the ara-2"F modification were gleaned by inspection of the crystal structure of *Clostridium* Tpt1 in complex with ADP-ribose-1"-PO_4, a derivative of the ADP-ribose-1",2"-cyclic-PO_4 reaction product that had undergone cyclic phosphate hydrolysis in situ[14] (**Figure 3.8**C). The ADP-ribose moiety occupies the NAD^+ substrate site in the C-terminal lobe of the Tpt1 tertiary structure and the 1"-PO_4 moiety is the α anomer, consistent with stereochemical inversion at the nicotinamide ribose C1 of β-NAD^+ during step1 of the Tpt1 pathway. The ribose O2' is hydrogen bonded to Lys66, a residue conserved among Tpt1 family enzymes; Lys66 also makes a hydrogen bond to the 1"-PO_4 (**Figure 3.8**C). The ribose C2 atom makes van der Waals contact to the nearby Arg139 side chain (conserved among Tpt1 family enzymes) that engages the α and β phosphates of ADP-ribose/NAD^+ (**Figure 3.8**C). By superimposing a structure of ara-2'F (extracted from the ara-2"F-ADPR ligand in the 1.75 Å crystal structure of human CDC38;[31] on the ribose of ADP-ribose-1",2"-cyclic-PO_4 in the Tpt1 complex, we can see the "up" configuration of the ara-2'F moiety versus the "down" configuration of the ribose-2'OH (highlighted by the asterisk in **Figure 3.8**C). In addition to eliminating the Lys66 contact to the NAD^+ substrate, the ara-2'F modification introduces a steric clash between the 2'F atom and the Arg139 side chain. Note that the equivalent of Arg139 in *Runella* Tpt1 (Arg137) is critical for its RNA 2'-phosphotransferase activity in vitro and in vivo, as gauged by the effect of its replacement with alanine[14]. We envision that the structural changes modeled in **Figure 3.8**C account, at least in part, for the slowing of the Tpt1 step 1 reaction with ara-2"F-NAD^+.

3.6 Additive Effect of an Arabinose-2'-Phosphate and the *Runella* Tpt1 R64A Mutation

Runella Tpt1 amino acids Arg16, His17, Arg64, and Arg119 are conserved in all Tpt1 homologs and essential for activity[6]. Kinetic analysis of *Runella* Tpt1 alanine mutants implicated Arg64 and Arg16 as catalysts of the transesterification step. The *Clostridium* Tpt1 structure showed that the equivalent two arginines make bidentate contacts to the transferred 2'-PO_4 in the ADP-ribose-1"-PO_4 product ligand, consistent with the arginines stabilizing the transition state of the transesterification step. Whereas R16A mutation of *Runella* Tpt1 Arg16 also profoundly affects step 1 of the pathway, the R64A mutation is highly selective in its effect on step 2, leading to transient accumulation of very high levels of the ADPR intermediate[6]. Because the R64A effect resembles that of the arabinose-2'-PO_4 RNA modification, we tested the reaction of the R64A mutant with the ara-2'-PO_4 RNA substrate (**Figure 3.5**C). We found that RslTpt1-R64A catalyzed quantitative conversion of the ara-2'-PO_4 RNA substrate to 2'-phospho-ADP-ribosylated RNA over a 30 min time course with an apparent rate constant of 0.28 ± 0.013 min^{-1}. There was no detectable formation of a 2'-OH RNA product. The rate constant of the R64A mutant for transfer of ADP-ribose to the RNA arabinose-2'-phosphate was 5-fold slower than that of wild-type *Runella* Tpt1. The effect of the R64A mutation on step 1 catalysis at the arabinose-2'-phosphate was similar to the 3-fold rate decrement in step 1 catalysis by the R64A mutant at a ribose-2'-phosphate[6].

3.7 Conclusions

The present study illuminates the distinctive effects of non-ribose sugars in the 2'-PO_4 nucleic acid and NAD^+ substrates on Tpt1 activity. Whereas replacement of each of the ribose sugars flanking the internal 2'-PO_4 with a deoxynucleotide did not affect the efficiency of 2'-PO_4 removal by *Runella* Tpt1, their replacement with 2'-fluoroarabinose reduced Tpt1 specific activity, albeit without accumulation of the 2'-phospho-ADP-ribosylated RNA intermediate. By contrast, replacing the ribose of the 2'-PO_4 nucleotide with arabinose selectively and severely reduced the rate of the transesterification step and thereby resulted in the build-up of very high levels of the reaction intermediate. Replacing the NMN ribose of NAD^+ with 2'-fluoroarabinose (which eliminates the step 2 ADP-ribose

O2" nucleophile) results in trapping of RNA-2'-phospho-(ADP-fluoroarabinose) as a dead-end step 1 product. We find that Tpt1 orthologs differ in their ability to use ara-2"F-NAD^+.

Our identification of substrate analogs that trap the ADP-ribosylated RNA intermediate has implications for the development of Tpt1 "poisons" as anti-fungals. A poison elicits the formation of a reaction intermediate that is potentially more deleterious (in this case by virtue of the large bulk of 2'-phospho-ADP ribose at the splice junction compared to the 2'-phosphate) than an inhibitor that results in the mere accumulation of 2'-phosphate substrate. Indeed, DNA topoisomerase poisons that trap covalent topoisomerase-DNA intermediates are among the most successful and widely used anti-bacterial and anti-cancer agents in clinical practice[32].

The 2'-phosphate analogues designed and synthesized in this chapter (2'-NH-PO_3, 2'-CH_2-PO_3), once incorporated into oligonucleotides, will facilitate the co-crystallization of Tpt1 enzymes with RNA reaction intermediates and possibly transition-state mimetics. Our ultimate goal would be to use the crystal structures to guide further synthesis of modified RNAs or RNA-like ligands as candidate inhibitors that we could then study for anti-fungal drug discovery.

3.8 Experimental

3.8.1 Recombinant Tpt1 proteins

Tpt1 enzymes from *Clostridium thermocellum*, *Homo sapiens*, *Runella slithyformis*, and *Chaetomium thermophilum* were produced in *E. coli* and purified as described previously[6, 9].

3.8.2 Solid-phase synthesis of oligonucleotides with internal 2′-phosphates

General methods

Oligonucleotide syntheses were carried out using an ABI 3400 DNA synthesizer (Applied Biosystems) on a Unylinker (ChemGenes) solid support at a 1 μmol scale. Conventional deoxyribonucleoside, 2′-*tert*-butyl-dimethylsilyl (TBDMS) ribonucleoside, 2′-fluoro-arabinonucleoside (2′-FANA) and 2′-acetyl levulinyl (ALE) ribonucleoside phosphoramidites (0.15 M in MeCN) (ChemGenes) were used. Additionally, 2′-arabino levulinyl (Lev) phosphoramidite was synthesized as described previously[33]. For

phosphitylation, *bis*-cyanoethyl,-*N,N*-diisopropyl-phosphoramidite (0.20 in MeCN) was used. Phosphoramidites were dissolved in MeCN and activated with 5-ethylthio-1H-tetrazole (0.25 M in MeCN). Capping was carried out by the simultaneous delivery of acetic anhydride in pyridine/THF and *N*-methylimidazole (16% in THF) and contacting the solid support for 6 sec. Oxidation of the phosphite triester intermediates was effected with 0.1 M iodine in pyridine/H_2O/THF (20 sec); a solution of 3% trichloroacetic acid in THF, delivered over 1.8 min, was used to deprotect DMTr groups. For 2′-phosphate-containing substrates, a solution of anhydrous TEA/MeCN (2:3 v/v) was used to remove cyanoethyl phosphate protecting groups, while a 0.5 M solution of hydrazine hydrate in pyridine/AcOH (3:2 v/v) was used to remove ALE protecting groups. All oligonucleotides were deprotected and cleaved from the solid support using an ammonium hydroxide and ethanol solution. TBDMS groups were removed using TREAT-HF. Crude oligonucleotides were purified *via* HPLC and characterized by LC–MS.

Synthesis of 2′ phosphate-containing oligonucleotides

Synthesis of RNA sequences containing 2′-phosphate moieties was carried out as previously described[6, 9] and the sequences are characterized in Table 3.1. The equivalent sequences containing DNA, 2′-FANA, 2′-FRNA, or 2′-OMe at positions flanking the 2′-phosphate RNA branchpoint were also synthesized in this manner using the respective phosphoramidites (DNA, 2′-FANA, 2′-FRNA, 2′-OMe). For the arabinose-2′-phosphate (ANA 2′-PO_4)-containing sequence: 5′-rCrCrAaraA$^{(2'\text{-P})}$rArU-3′, the first section of the oligonucleotide: 5′-araA$^{(2'\text{-Lev})}$rArU-3′, was synthesized on the solid support in the conventional 3′ to 5′ solid phase synthesis. The araA$^{2'\text{-Lev}}$ unit in this sequence was introduced by coupling a 5′-DMTr-2′-O-Lev-3′-OCE arabinose phosphoramidite (0.15 M in MeCN) for 15 min. In the case of arabinonucleosides, the removal of the backbone cyanoethyl phosphodiester protecting groups is not necessary, since the *trans* orientation of the free 2′-OH to the 3′-phosphate triester will prevent chain cleavage at the arabinose position. To remove the 2′-ALE groups, the columns were returned to the synthesizer and a freshly prepared solution of 0.5 M hydrazine hydrate in pyridine/AcOH (3:2 v/v) was flowed through the columns (20 sec flow + 3.75 min sleep, repeated 4×). After washing (MeCN, 10 min) and drying (Ar gas, 10 min), the solid supports were dried again in vacuo

(30 min). To phosphitylate at the newly exposed 2′-OH, bis-cyanoethyl-*N,N*-diisopropyl CED phosphoramidite (0.20 M in MeCN) was coupled for 30 min, and then further oxidized using 0.1 M iodine in pyridine/H_2O/THF (20 sec). To complete the oligonucleotide, standard 3′ to 5′ synthesis was continued on the 5′ terminus of the growing oligonucleotide using the sequence 5′rCrCrA-3′ to yield the desired hexamer oligonucleotide substrate.

Deprotection and cleavage of oligonucleotides from the solid support was achieved by treatment with 1 mL of cold 29% aqueous ammonia/ethanol (3:1 v/v) for RNA and 2′-FANA oligonucleotides, and 1 mL of cold 29% aqueous ammonia for DNA oligonucleotides for 16 h at 65°C. Samples were centrifuged and the supernatant was transferred to a clean 1.5 mL Eppendorf tube and vented for 30 min, chilled on dry ice, and evaporated to dryness. Removal of the 2′-silyl protecting groups for the RNA oligonucleotides was achieved by treatment with a 300 μL solution of NMP/Et_3N/TREAT-HF (3:4:6 v/v/v) for 90 min at 65°C, followed by quenching with 3 M NaOAc buffer (50 μL; pH 5.5) and precipitation of the crude oligonucleotide from cold butanol (1 mL, −20°C). Samples were chilled on dry ice for 30 min and then centrifuged. After removing the supernatant, the remaining pellet (containing oligonucleotide) was evaporated to dryness, taken up in autoclaved milliQ water (1 mL), filtered, and quantified by UV spectroscopy.

Crude oligonucleotides were HPLC-purified using a Waters Protein-Pak DEAE 5PW anion exchange column (21.5 × 150 mm). A mobile phase of 1 M aqueous $LiClO_4$ in milli-Q water was used for analysis and purification (0%–20% $LiClO_4$ over 30 min, 4 mL/min, 60°C). Following collection of the desired peaks, fractions were combined and excess $LiClO_4$ salts were removed using Gel Pak 2.5 size exclusion columns (Glen Research). Purified oligonucleotides were characterized by electrospray ionization-mass spectrometry and quantified by UV spectroscopy. Extinction coefficients were determined using the IDT OligoAnalyzer tool (www.idtdna.com/analyzer/Applications/OligoAnalyzer). The oligonucleotides were stored at −20°C.

Table 3.1. Oligonucleotides synthesized for probing Tpt1 enzyme activity. LCMS characterization on right.

Sample	Sequence	M_{exp}	M_{obs}
LA-RNA	5'- **CCA AAU** -3'	1919.27	1919.28
LA-DNA	5'- CCA AAT -3'	1854.31	1853.35
LA-DNA-PS	5'- C*C*A* A*A*T -3'	1934.31	1933.21
LA-F-ANA	5'- CCA AAU -3'	1941.26	1943.27
LA-F-ANA-PS	5'- C*C*A* A*A*U -3'	2021.25	2023.15
LA-OMe	5'- CCA AAT -3'	1990.36	1989.32
LA-OMe- PS	5'- C*C*A* A*A*T -3'	2070.36	2069.22
LA-F-RNA	5'- CCA AAT -3'	1930.26	1929.22
LA-RNA-2′-PS	5'- **CCA AAU** -3'	1936.33	1935.19
LA-ARA	5'- **CCA AAU** -3'	1919.27	1919.29
LA-2'-rG	5'- **CCA A$^{2'\text{-rG}}$ AU** -3'	2185.36	2184.38

N = 2'-OH (RNA) N = 2'-PO_4 N = 2'-H (DNA) N = 2'-F-ANA
N = 2'-OMe N = 2'-F-RNA N = 2'-PSO_3 N = 2'-araPO_4 * = PS linkage

3.8.3 5′ ^{32}P-labeled oligonucleotide substrates

Synthetic 6-mer oligonucleotides 5′-CCAA$^{2'P}$AU containing an internal 2′-PO_4 were 5′ ^{32}P-labeled by reaction with phosphatase-dead T4 polynucleotide kinase (Pnkp-D167N) in the presence of [γ^{32}P]ATP. The reactions were quenched with 90% formamide, 50 mM EDTA, 0.01% xylene cyanol and the radiolabeled RNAs were purified by electrophoresis through a 40-cm 20% polyacrylamide gel containing 7 M urea in 45 mM Tris-borate, 1 mM EDTA. The radiolabeled oligonucleotides were eluted from excised gel slices, recovered by ethanol precipitation, and resuspended in 10 mM Tris-HCl, pH 6.8, 1 mM EDTA, and stored at −20°C.

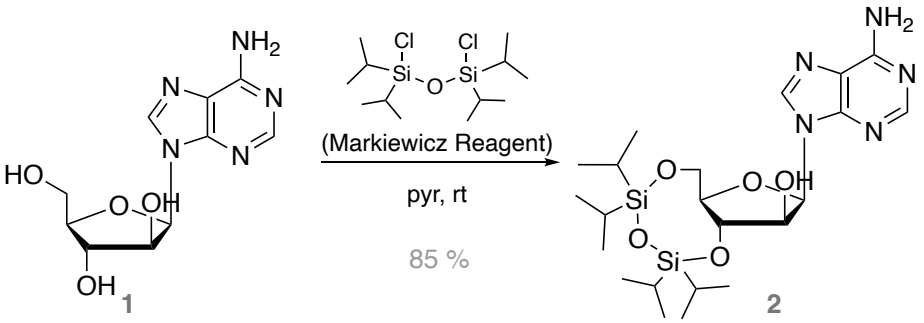
3.8.4 Assay of Tpt1 activity

Reaction mixtures containing 100 mM Tris-HCl (pH 7.5), 0.2 μM 5′ ^{32}P-labeled nucleic acid substrates, NAD^+ or ara-2″F-NAD^+ as specified, and Tpt1 as specified in the figure legends were incubated at 37°C. The reactions were quenched at the times specified in the figure legends by addition of 3 volumes of cold 90% formamide, 50 mM EDTA. The products were analyzed by electrophoresis (at 55 W constant power) through a 40-cm 20%

polyacrylamide gel containing 7 M urea in 45 mM Tris-borate, 1 mM EDTA and visualized by autoradiography and/or scanning the gel with a Fujifilm FLA-7000 imaging device. The products were quantified by analysis of the gel scans in ImageQuant. Ara-2′F-NAD$^+$ was purchased from BIOLOG (Bremen, Germany; Cat. No. D148).

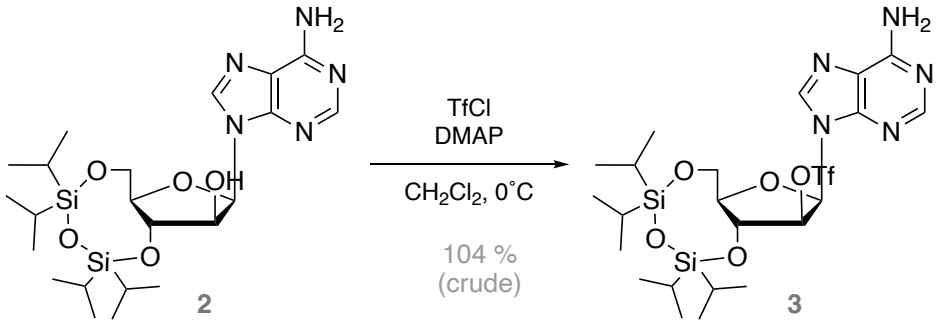
3.8.5 Synthesis of C2′-NH-PO$_3$ phosphoramidite

3',5'-O-[1,1,3,3-Tetrakis(1-methylethyl)-1,3-disiloxanediyl] adenosine (2)



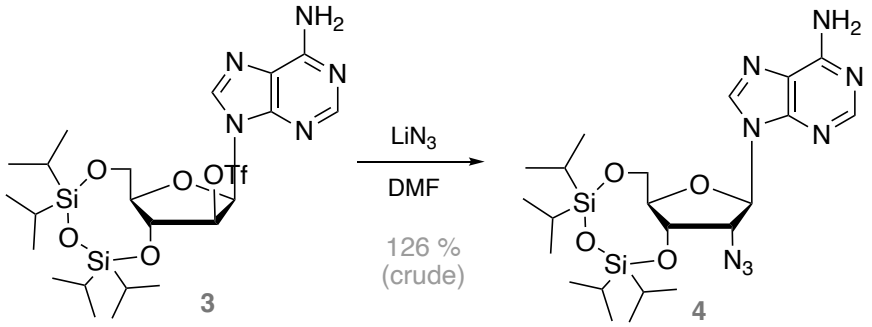
Arabino Adenosine (2.000 g, 7.484 mmol) was suspended in 60 mL pyridine in a round bottom flask with a stir bar, under argon. Markiewicz reagent (1,1,3,3-tetraisopropyl-1,3-dichlorodisiloxane) (3.844 mL, 8.606 mmol) was added, and the mixture was stirred at room temperature. Over time, the white suspension begins to dissolve. After 3 hours, once the reaction mixture was almost clear, the reaction was monitored by TLC and appeared to be complete. The pyridine was evaporated off, and the product was extracted with EtOAc. The organic layer was washed with water, 1M $NaHSO_4$ (3x), saturated NaHCO3, and brine. The crude product, an off white solid, was purified via column chromatography (100 % EtOAc), and the desired product (Rf = 0.3) was concentrated in vacuo to yield a white solid (3.254 g, 85 %). ^{1}H NMR (500 MHz, DMSO) δ 8.10 (s, 1H), 8.03 (s, 1H), 7.26 (s, 2H), 6.20 (d, *J* = 6.6 Hz, 1H), 5.77 (d, *J* = 5.9 Hz, 1H), 4.57 (t, *J* = 8.0 Hz, 1H), 4.53 – 4.46 (m, 1H), 4.10 (dd, *J* = 12.6, 4.4 Hz, 1H), 3.92 (dd, *J* = 12.6, 3.0 Hz, 1H), 3.79 (ddd, *J* = 7.7, 4.4, 3.1 Hz, 1H), 1.17 – 0.98 (m, 28H). ^{13}C NMR (126 MHz, DMSO) δ 156.47, 152.86, 150.01, 140.05, 118.97, 82.24, 80.00, 75.54, 75.30, 61.93, 17.85, 17.70, 17.43, 17.38, 17.32, 17.28, 13.29, 12.93, 12.71, 12.45. HRMS **(ES$^+$)** calculated: 509.2490, found: 510.2548.

3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]-, 2'-(trifluoromethanesulfonate) Adenosine (3)



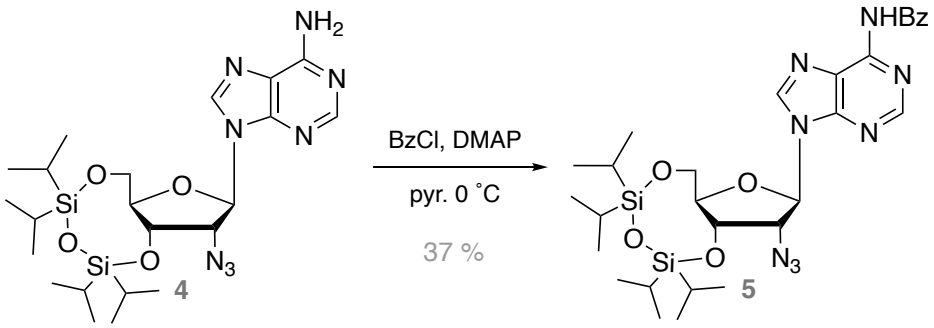
In an oven dried flask equipped with a stir bar, 3′,5′-siloxane protected arabino adenosine (3.000 g, 5.885 mmol) and DMAP (2.157 g, 17.656 mmol) were added. The flask was evacuated and refilled with argon three times. Dry CH_2Cl_2 (30 mL) was added and the mixture was cooled to 0 ˚C. Then, trifluoromethanesulfonyl chloride (TfCl) (0.751 mL, 7.062 mmol) was added to 10 mL anhydrous CH_2Cl_2 (10 mL) and the combined mixture was added dropwise to the stirring reaction at 0˚C, which turned yellow immediately upon addition of TfCl. The reaction was stirred and gradually warmed to room temperature for 1 hr. To work up, the reaction was diluted with CH_2Cl_2 and washed once with Na_2SO_4, once with $NaHCO_3$, and once with brine, extracting with CH_2Cl_2 each time. The organic layer was dried over $NaHSO_4$ and concentrated *in vacuo*. The crude product (3.912, 104 %) was a foamy off white solid, and was carried forward to the next step without further purification. HRMS **(ES$^+$)** calculated: 641.1983, found: 642.2048.

2'-Azido-2'-deoxy-3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]adenosine (4)



Compound 3 (2.400 g, 3.739 mmol) was dissolved in 30 mL of anhydrous DMF and the flask was evaporated and refilled with argon. NaN_3 solution (20% aqueous solution) (0.842 mL, 18.697 mmol) was added and the reaction was stirred at room temperature overnight. The next day, EtOAc was used to dilute the reaction, followed by extraction with H_2O (3x), brine. The organic phase was dried over Na_2SO_4 and the solution was filtered and evaporated to dryness. The reaction had gone to completion and the crude product (pale yellow solid) was carried over to the next step without further purification. HRMS **(ES^+)** calculated: 534.2555, **[M+Na]** found: 557.2440

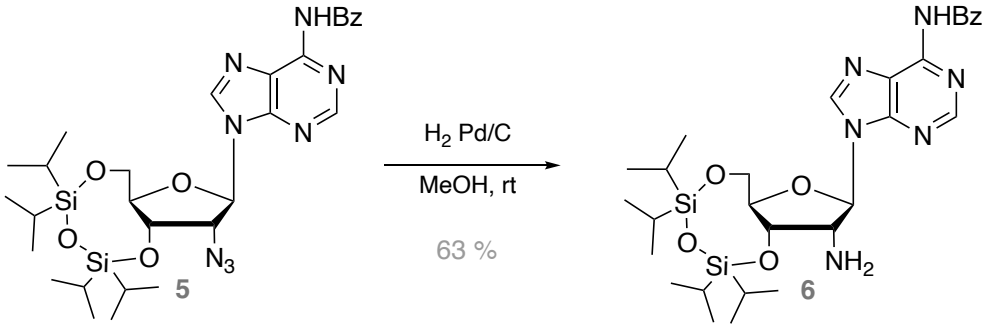
<u>*N*^{6}-Benzoyl-2'-azido-2'-deoxy-3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl] adenosine (5)</u>



Compound 4 (2.000 g, 3.740 mmol) and DMAP (0.014 g, 0.112 mmol) were dissolved in 30 mL of anhydrous pyridine. Benzoyl chloride (0.478 mL, 4.114 mmol) was added dropwise at 0 ˚C. The reaction turned darker yellow immediately upon this addition. The reaction was left stirring at room temperature overnight. Pyridine was evaporated off and the reaction was redissolved in EtOAc and washed with $NaHSO_4$ (3x) and brine, then dried over Na_2SO_4 and the solvent evaporated. The crude product was a foamy white solid and upon inspection by TLC, two products had appeared (R_f 0.8 and 0.5, in 1:1 hexanes:EtOAc). The two products were purified via flash chromatography (30 % to 50 % EtOAc in hexanes). NMR and MS confirm that the product at R_f 0.8 was doubly benzoylated. For future repetitions of this reaction, in order to prevent double benzoylation, BzCl was first dissolved in pyridine and added slowly to the reaction. The desired product (R_f 0.5) was isolated as a white foam (0.876, 37 %). Both products were carried forward to the next step, as the second N^6 benzoyl group is expected to be very labile. ^{1}H NMR (500 MHz, DMSO) δ 11.25 (s, 1H), 8.68 (s, 1H), 8.55 (s, 1H), 8.08 – 8.03 (m, 2H), 7.69 – 7.62 (m, 1H), 7.56 (dd, *J* = 8.4, 7.0 Hz, 2H), 6.02 (d, *J* = 1.4 Hz, 1H), 5.79 (s, 1H), 5.70 (d, *J* =

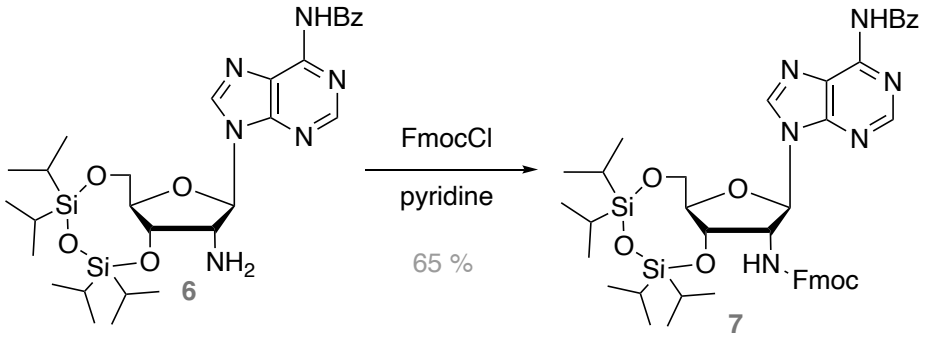
4.7 Hz, 1H), 4.84 (dd, *J* = 8.2, 5.2 Hz, 1H), 4.66 (td, *J* = 5.0, 1.6 Hz, 1H), 4.11 – 4.00 (m, 2H), 3.96 (dd, *J* = 12.4, 2.5 Hz, 1H), 1.06 – 0.90 (m, 28H). HRMS **(ES$^-$)** calculated: 638.2817, found: 637.2715.

<u>*N*^{6}-Benzoyl-2'-amino-2'-deoxy-3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl] adenosine (6)</u>



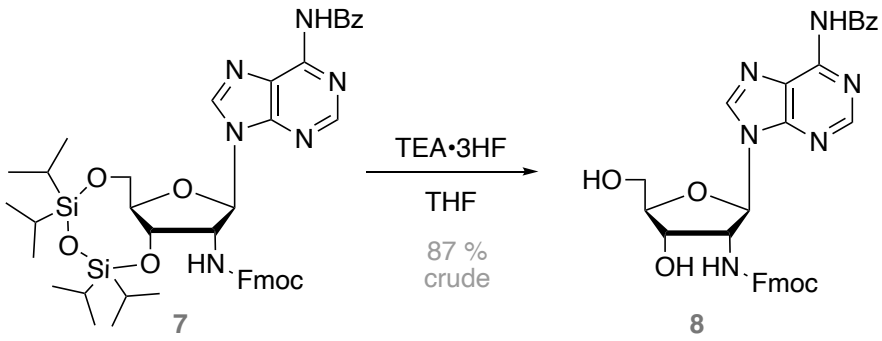
A round bottom flask equipped with a stir bar was charged with compound 5 (0.805 g, 1.083 mmol) and Pd/C 20 wt% (0.160 g) dissolved in MeOH. The flask was evacuated and refilled with hydrogen gas three times, then left to stir at room temperature overnight. The next day, the reaction progression was checked by TLC (R_f 0.15, 100% EtOAc) and appeared to have gone to 90%. The crude mixture was filtered over celite and washed several times with MeOH and CH_2Cl_2. The solvent was evaporated in vacuo to yield an off-white foam. The product was purified *via* flash chromatography (100 % EtOAc to 20 % MeOH in EtOAc) to yield an off-white foam (0.420 g, 63 %). ^{1}H NMR (500 MHz, DMSO) δ 11.21 (s, 1H), 8.70 (s, 1H), 8.61 (s, 1H), 8.08 – 8.03 (m, 2H), 7.69 – 7.62 (m, 1H), 7.56 (dd, *J* = 8.4, 7.0 Hz, 2H), 5.86 (d, *J* = 4.1 Hz, 1H), 4.82 (t, *J* = 6.5 Hz, 1H), 4.12 – 3.93 (m, 4H), 2.42 – 2.02 (s, broad, 2H), 1.16 – 0.95 (m, 28H). HRMS **(ES$^+$)** calculated: 612.2912, **[M+Na]** found: 635.2782.

<u>*N*^{6}-Benzoyl-2'-deoxy-2'-[[(9H-fluoren-9-ylmethoxy)carbonyl]amino]-3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl] adenosine (7)</u>



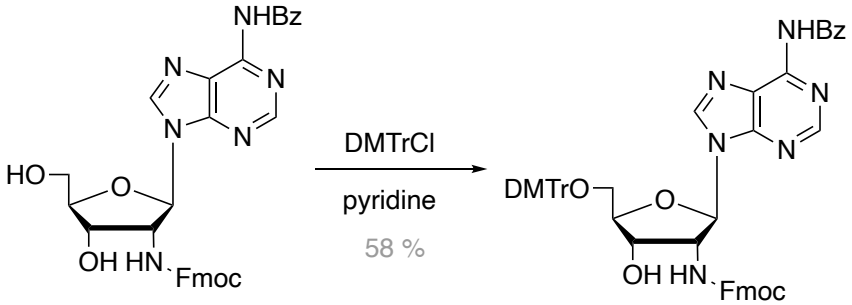
Compound 6 (0.400 g, 0.653 mmol) was co-evaporated with anhydrous pyridine 3 times, then dissolved in anhydrous pyridine (8 mL), and the reaction was stirred at ambient temperature under argon. 9-Fluorenylmethyloxycarbonyl chloride (0.219 g, 0.848 mmol) was dissolved in 2 mL pyridine (turned yellow in the pyridine) and added to the reaction. The reaction was quenched 1 hour later by the addition of H_2O (2 mL). CH_2Cl_2 and $NaHCO_3$ (sat. aq.) were subsequently added, and the organic phase was washed with $NaHCO_3$ (3x). The combined aqueous layers were extracted with CH_2Cl_2 once, and the combined organic layers were dried over Na_2SO_4, filtered, and concentrated. The residue was purified by flash chromatography (30 % EtOAc in hexanes) to yield compound 7 (0.355 g, 65 %) as an amorphous yellow solid. ^{1}H NMR (500 MHz, DMSO) δ 11.25 (s, 1H), 8.71 (d, J = 5.1 Hz, 2H), 8.08 – 8.03 (m, 2H), 7.91 – 7.80 (m, 3H), 7.75 – 7.62 (m, 3H), 7.56 (t, J = 7.7 Hz, 2H), 7.41 (t, J = 7.5 Hz, 2H), 7.34 – 7.25 (m, 2H), 6.08 (d, J = 4.7 Hz, 1H), 5.03 (td, J = 8.7, 4.8 Hz, 1H), 4.93 (t, J = 7.3 Hz, 1H), 4.31 (qd, J = 10.6, 6.8 Hz, 2H), 4.20 (t, J = 6.6 Hz, 1H), 4.09 – 3.95 (m, 3H), 1.16 – 0.94 (m, 29H). HRMS **(ES$^+$)** calculated: 834.3593, **[M+Na]** found: 857.3513.

<u>*N*^{6}-Benzoyl-2′-deoxy-2′-[(9-fluorenylmethyloxycarbonyl)amino] adenosine (8)</u>



A Teflon round bottom flask equipped with a stir bar was charged with compound 7 (0.335 g, 0.401 mmol) and THF (3 mL). TREAT-HF (0.043 mL, 0.269 mmol) was added to the reaction, which was then left stirring at room temperature overnight. The next day, the reaction appeared to have gone to near completion by TLC (5 % MeOH in EtOAc) and the reaction was quenched with methanol and further diluted with EtOAc. The organic layer was washed with brine (with some added $NaHSO_4$ to avoid basic conditions) and dried over Na_2SO_4. The crude product was concentrated *in vacuo* carried forward onto the next step without purification (0.244 g, 87 %).

<u>N^6-Benzoyl-2′-deoxy-5′-O-[4,4′-dimethoxytrityl]-2′-[(9-fluorenyl-methyloxycarbonyl)amino]adenosine (9)</u>

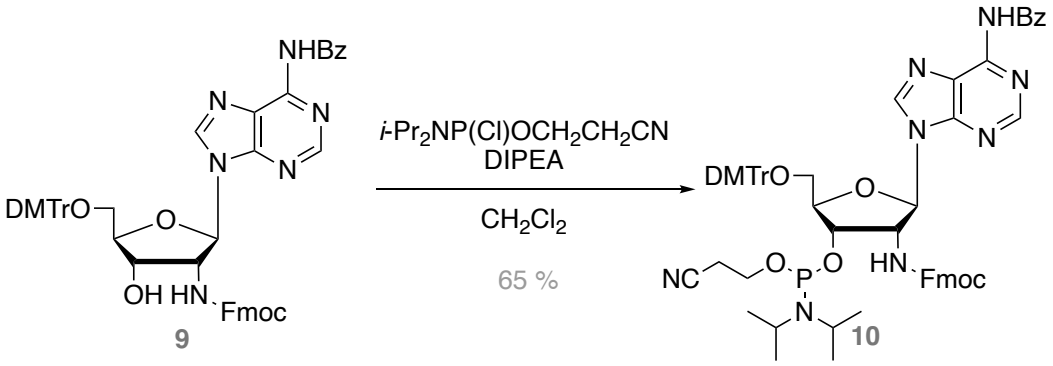


8 **9**

Compound 8 (0.244 g, 0.350 mmol) was co-evaporated with dried pyridine three times, then dissolved in dried pyridine (4.0 mL), and the mixture was stirred at ambient temperature under argon. 4,4′-O,O-dimethoxytrityl chloride (0.136 g, 0.403 mmol) was then added into the reaction mixture, and this was stirred for 4 h and quenched through the addition of excess methanol (2 mL). Next, CH_2Cl_2 and $NaHCO_3$ (sat. aq.) were added, and the organic phase was washed three $NaHCO_3$ (3x). The combined aqueous layers were extracted with CH_2Cl_2 once. The combined organic layers were dried over Na_2SO_4, filtered, and concentrated. The residue was purified by flash chromatography (80% EtOAc in hexanes) to yield compound 9 (0.181 g, 52 %) as an amorphous solid. ^{1}H NMR (500 MHz, DMSO) δ 11.23 (s, 1H), 8.63 (s, 1H), 8.54 (s, 1H), 8.08 – 8.02 (m, 2H), 7.87 (d, J = 7.6 Hz, 2H), 7.68 – 7.62 (m, 3H), 7.56 (dd, J = 8.4, 7.1 Hz, 2H), 7.39 (ddt, J = 9.8, 6.4, 1.9 Hz, 5H), 7.28 – 7.22 (m, 6H), 6.85 (dd, J = 8.8, 4.3 Hz, 5H), 6.18 (d, J = 7.8 Hz, 1H), 5.81 (d, J = 4.7 Hz, 1H), 5.17 (q, J = 7.6 Hz, 1H), 4.39 (d, J = 6.7 Hz, 1H), 4.29 (dd, J = 10.0, 6.8 Hz, 1H), 4.20 (ddd, J = 13.4, 10.1, 6.8 Hz, 3H), 4.04 (q, J = 7.1 Hz, 1H), 3.73 (s, 6H), 3.28

(h, J = 4.8 Hz, 2H), 2.00 (s, 2H), 1.18 (t, J = 7.1 Hz, 2H). HRMS **(ES$^+$)** calculated: 894.3377, **[M+Na]** found: 917.3258.

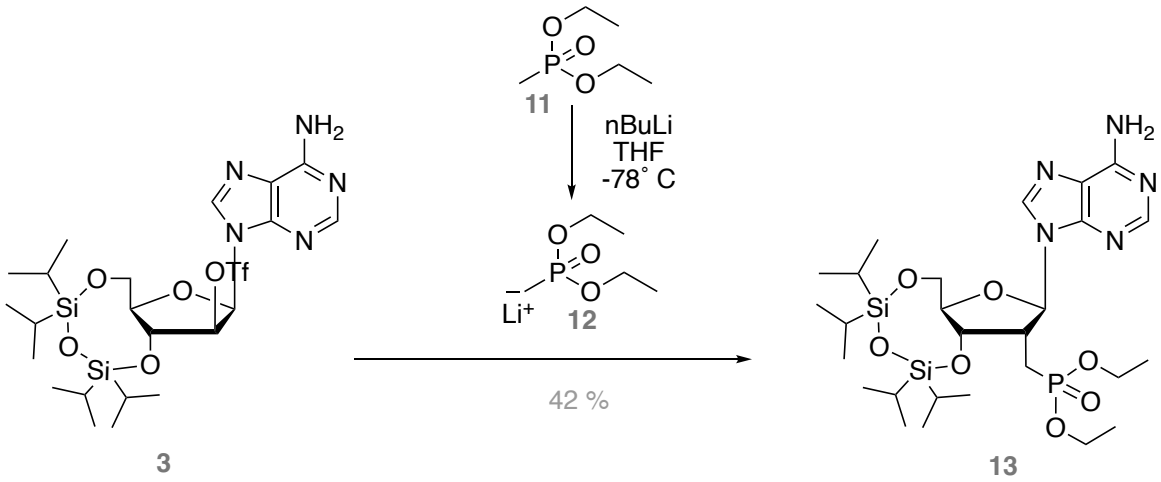
N^6-Benzoyl-2′-deoxy-5′-O-[4,4′-dimethoxytrityl]-2′-[(9-fluorenyl-methyloxycarbonyl)amino]adenosine 3′-(2-cyanoethyl N,N-diiso-propylphosphoramidite) (10)



Compound 9 (0.160 g, 0.179 mmol) was co-evaporated with dried pyridine three times and dried under vacuum overnight. Anhydrous CH_2Cl_2 was dried over molecular sieves overnight. The dried compound was dissolved in anhydrous CH_2Cl_2 (3.5 mL), under argon, and the reaction was cooled to 0 ˚C. *N,N*-diisopropylethylamine (0.156 mL, 0.894 mmol) was added and allowed to stir in the reaction for 10 minutes. Then, 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (0.160 mL, 0.715 mmol) was added and the reaction was allowed to warm to room temperature by stirring for another 2.5 hours (to completion, as monitored by TLC). The reaction mixture was concentrated *in vacuo* and immediately purified by flash chromatography (1:1 EtOAc/hexanes to 4:1 EtOAc/hexanes) to yield compound 10 as an amorphous solid (0.128 g, 65 %). ^{1}H NMR (500 MHz, DMSO) δ 11.26 (s, 1H), 8.64 – 8.53 (m, 2H), 8.08 – 8.03 (m, 2H), 7.91 – 7.84 (m, 2H), 7.77 (t, J = 7.5 Hz, 1H), 7.71 – 7.53 (m, 5H), 7.43 – 7.35 (m, 4H), 7.25 (ddd, J = 16.7, 9.0, 4.5 Hz, 10H), 6.87 – 6.78 (m, 4H), 6.25 (dd, J = 8.0, 6.2 Hz, 1H), 5.44 – 5.29 (m, 1H), 5.19 (t, J = 5.4 Hz, 0H), 4.69 – 4.51 (m, 1H), 4.45 – 4.28 (m, 3H), 4.24 – 4.01 (m, 2H), 3.58 (dq, J = 15.1, 7.3 Hz, 3H), 2.73 – 2.55 (m, 3H), 1.26 – 1.10 (m, 12H), 1.03 – 1.00 (m, 6H). ^{31}P NMR (203 MHz, DMSO) δ 149.48, 147.85. HRMS **(ES$^+$)** calculated: 1094.4456, **[M+Na]** found: 1117.4326.

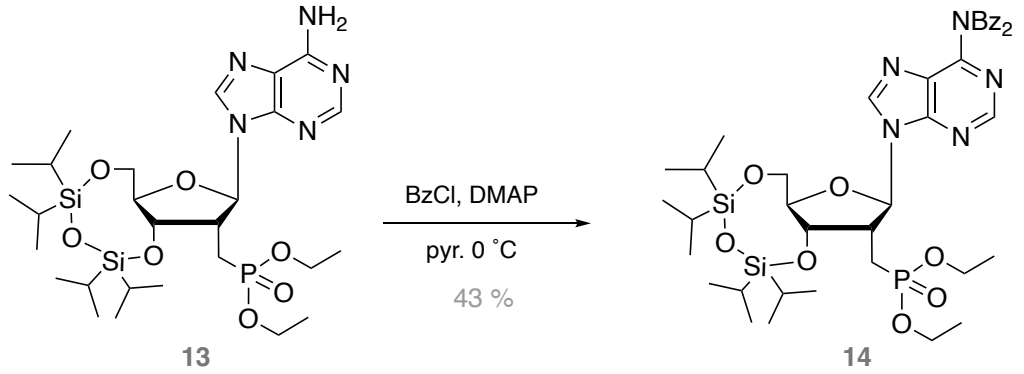
3.8.6 Synthesis of intermediates towards C2′-CH_2-PO_3 phosphoramidite

2'-deoxy-2'-C-methyl(diethoxyphosphinyl)-3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]- adenosine (13)



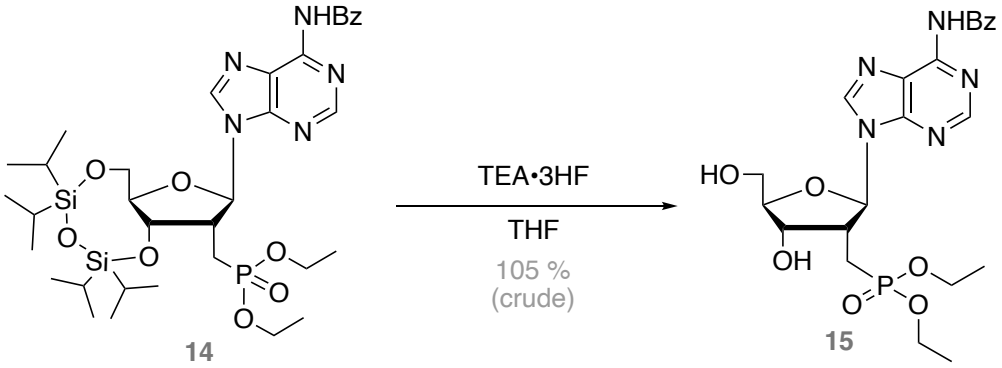
An oven-dried 100 mL round bottom flask equipped with a stir bar was evacuated and refilled with argon 3x. Anhydrous THF (10 mL) was added to the flask and cooled to -78 ˚C. Diethyl methylphosonate (2.630 mL, 17.996 mmol) was then added. nBuli (2.5 M solution in hexane, 8.227 mL, 20.567 mmol) was carefully added to the flask. The nBuli was a pale yellow liquid. This mixture was allowed to stir at -78 ˚C for 1.5 hours. The triflate, compound 3 (3.300 g, 5.142 mmol) was dissolved in 10 mL anhydrous THF and added slowly to the reaction. The reaction turned dark yellow/green, and then dark orange upon this addition. The reaction was stirred at -78 ˚C for another 3 hours. Then, the reaction was quenched by adding brine to the mixture, and the organic product was extracted with EtOAc, and washed with Brine two times. The crude product (0.320 g) was a pale, yellow oil. The product was purified via column chromatography (100% EtOAc (R_f = 0.1) to 10 % MeOH in EtOAc and concentrated in vacuo to yield a white solid (1.400 g, 42 %). ^{1}H NMR (500 MHz, DMSO) δ 8.12 (s, 1H), 8.11 (s, 1H), 7.33 (s, 2H), 6.25 (s, 1H), 4.69 – 4.63 (m, 1H), 4.13 – 4.06 (m, 1H), 4.06 – 3.88 (m, 6H), 2.44 – 2.25 (m, 2H), 1.21 (dt, *J* = 9.8, 7.1 Hz, 6H), 1.13 – 0.98 (m, 26H), 0.88 (td, *J* = 7.1, 3.7 Hz, 1H) **(Supplementary Figure 3.1**). ^{13}C NMR (126 MHz, DMSO) δ 156.55, 152.81, 149.94, 140.69, 118.97, 86.08, 80.64, 77.61, 62.51, 61.73, 17.82, 17.27, 16.54, 13.33, 12.62 (**Supplementary Figure 3.2**). ^{31}P NMR (203 MHz, DMSO) δ 28.07. HRMS **(ES$^+$)** calculated: 643.2986, **[M+K]** found: 682.2848. Sample NMR assignments provided in Supplemental Figures (Section 3.9).

N^6-Dibenzoyl-2'-deoxy-2'-C-methyl(diethoxyphosphinyl)-3',5'-O-[1,1,3,3-tetrakis(1-methylethyl)-1,3-disiloxanediyl]- adenosine (14)



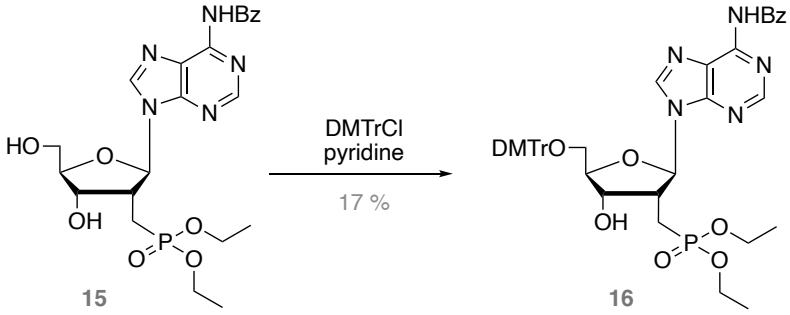
Compound 13 (1.200 g, 1.864 mmol) and DMAP (0.007 g, 0.056 mmol) were dissolved in 10 mL anhydrous pyridine and the reaction placed under argon. Benzoyl chloride (0.476 mL, 4.100 mmol), dissolved in 5 mL anhydrous pyridine, was added slowly at 0˚C (added two equivalents this time because usually end up needing to add more). The reaction was left stirring at room temperature overnight. Finally, pyridine was evaporated off, and the sample was dissolved in EtOAc and washed with $NaHSO_4$ (3x), then brine, then dried over Na_2SO_4, and the solvent evaporated. The crude product was a foamy yellow solid. Double benzoylation product was observed as in compound 5. The two products were purified via column chromatography (30 % - 50 % EtOAc in hexanes) and the desired compound was isolated as a white solid (1.394 g, 43 %). ^{1}H NMR (500 MHz, DMSO) δ 8.25 (s, 1H), 7.99 – 7.93 (m, 1H), 7.78 (ddd, *J* = 8.4, 4.6, 1.3 Hz, 3H), 7.66 – 7.58 (m, 2H), 7.54 – 7.43 (m, 4H), 6.39 (s, 1H), 6.05 (s, 1H), 4.16 (tdd, *J* = 12.7, 9.8, 5.8 Hz, 1H), 4.07 – 3.85 (m, 5H), 3.76 (dd, *J* = 13.5, 2.6 Hz, 1H), 3.71 (dt, *J* = 9.4, 2.1 Hz, 1H), 2.68 – 2.57 (m, 1H), 2.48 – 2.38 (m, 1H), 1.27 (t, *J* = 7.1 Hz, 4H), 1.19 (t, *J* = 7.0 Hz, 3H), 1.11 – 0.94 (m, 31H), 0.67 – 0.61 (m, 6H). ^{31}P NMR (203 MHz, DMSO) δ 25.61. HRMS **(ES$^+$)** calculated: 851.3511, **[M+K]** found: 890.3351.

N^6-Benzoyl-2'-deoxy-2'-C-methyl(diethoxyphosphinyl) adenosine (15)



Compound 14 (0.550 g, 0.735 mmol) was dissolved in 5 mL anhydrous THF and transferred into a Teflon round bottom flask equipped with a stir bar. The flask was evacuated and refilled with argon. Triethylamine trihydrogen fluoride (TREAT-HF) (0.080 mL, 0.493 mmol) was added to the flask and the reaction was then left stirring at room temperature overnight. The next day, the reaction appeared to have gone to near completion by TLC (5 % MeOH in EtOAc) and the reaction was quenched with methanol and further diluted with EtOAc. The organic layer was washed with brine and dried over Na_2SO_4. The crude product was carried over to the next step without further purification (0.390 g, 105 %). HRMS **(ES$^+$)** calculated: 505.1726, **[M+K]** found: 544.1559.

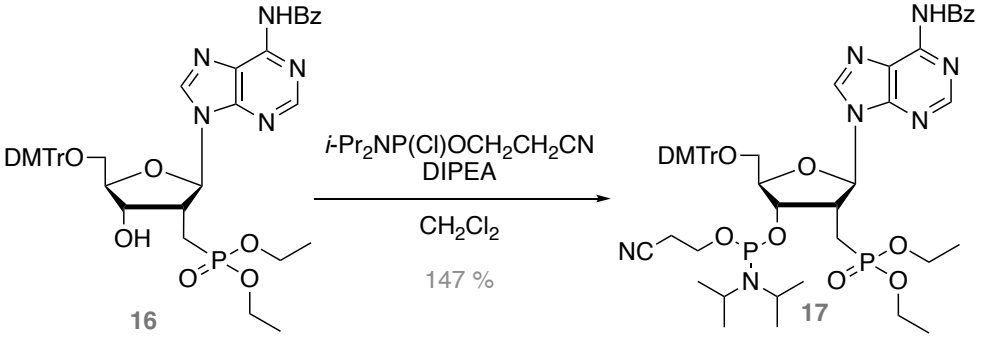
<u>*N*^{6}-Benzoyl-2'-deoxy-5′-O-[4,4′-dimethoxytrityl]-2'-C-methyl(diethoxyphosphinyl) adenosine (16)</u>



Compound 15 (crude, 0.370 g, 0.732 mmol) was co-evaporated with pyridine (6 mL), and the flask was equipped with a stir bar and refilled with argon. DMTrCl (0.298 g, 0.878 mmol) was dissolved in 2 mL pyridine and added dropwise to the reaction, under argon. The reaction was allowed to stir at room temperature overnight. The next day the reaction appeared complete by TLC (5% MeOH in EtOAc) and was quenched with MeOH. The reaction was evaporated and redissolved in CH_2Cl_2 and extracted from $NaHCO_3$, two times, followed by brine, and dried over Na_2SO_4. The crude product was purified via flash

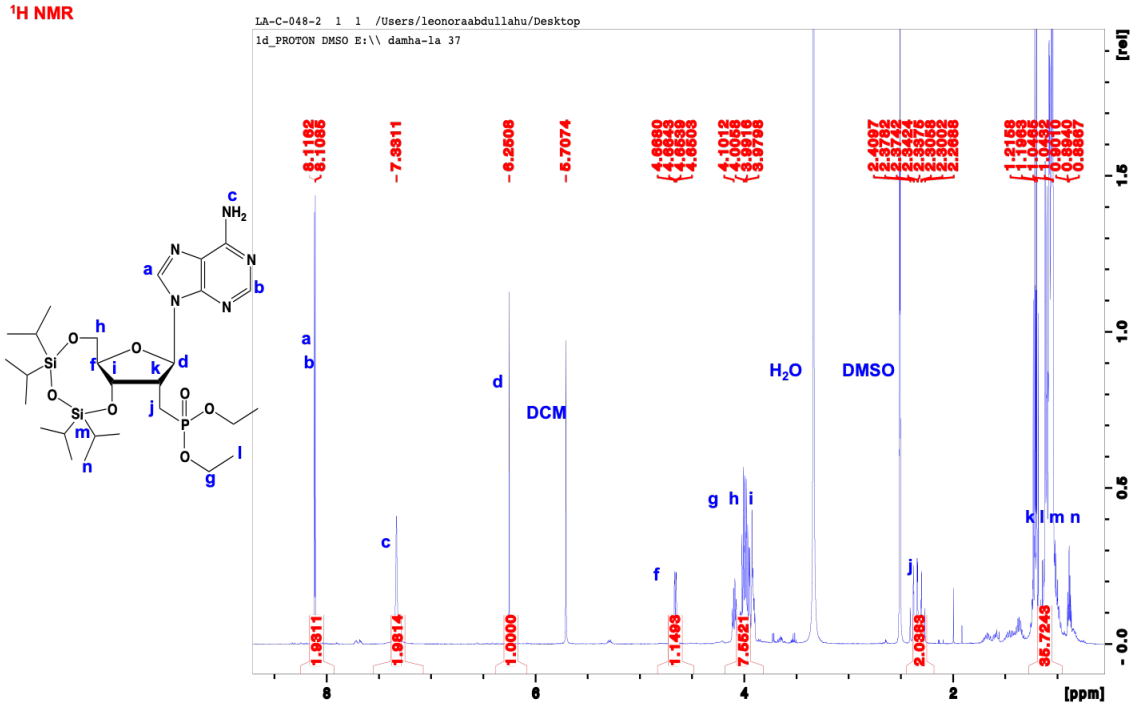
chromatography (0-5% MeOH in EtOAc) and the desired product was isolated as a white foam (0.100 g, 17 %). ^{31}P NMR (203 MHz, DMSO) δ 28.24. HRMS **(ES$^+$)** calculated: 807.8401, **[M+K]** found: 846.2886

<u>*N*^{6}-Benzoyl-2′-deoxy-5′-O-[4,4′-dimethoxytrityl]-2-C-methyl(diethoxyphosphinyl) adenosine 3′-(2-cyanoethyl N,N-diiso-propylphosphoramidite) (17)</u>

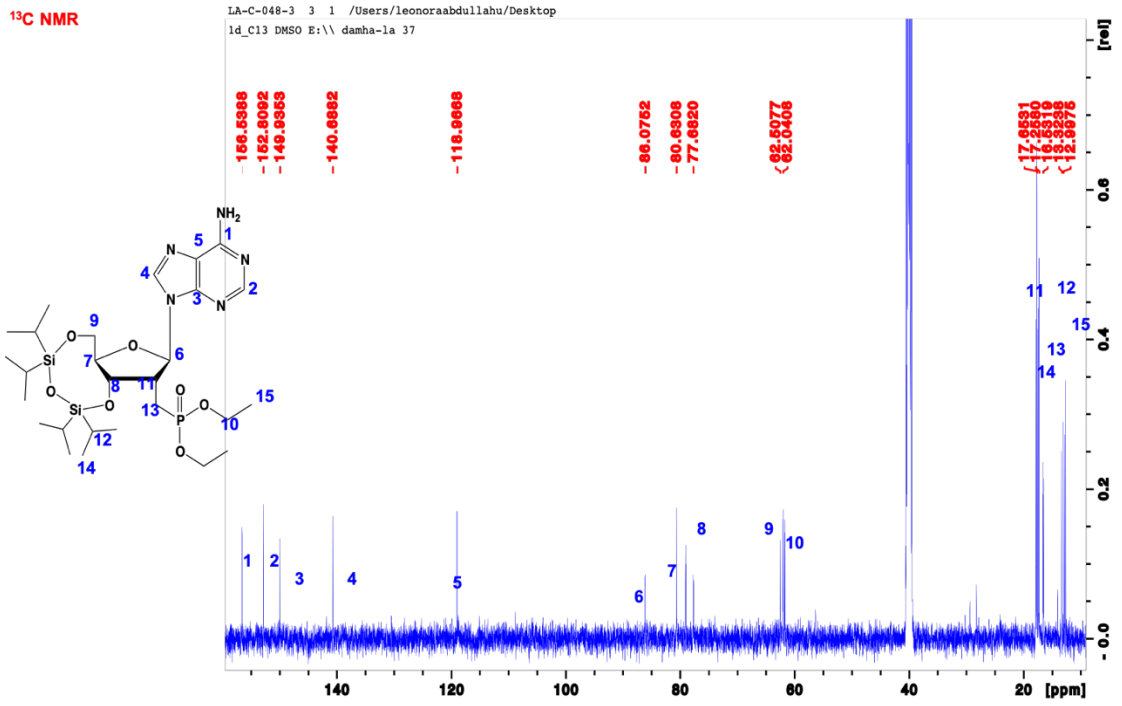


Compound 16 (0.070 g, 0.087 mmol) was co-evaporated with pyridine (3x) and dried under vacuum overnight. Anhydrous CH_2Cl_2 was allowed to dry over molecular seives overnight. The starting material was dissolved in CH_2Cl_2 (2 mL) and cooled to 0˚C. Diisopropylethylamine (DIPEA) was added and the reaction was stirred for 10 minutes. Then, 2-Cyanoethyl N,N-diisopropyl chlorophosphoramidite (0.077 mL, 0.347 mmol) was added and the reaction was warmed to room temperature and stirred for another 2.5 hours. The reaction was monitored by TLC, and the reaction mixture was concentrated in vacuo and immediately purified by flash chromatography under argon (50-100 % EtOAc in hexanes). The desired product was isolated as a sticky white solid (0.128 g, 147 % - presumably not dry, and containing unreacted P-reagent). ^{31}P NMR (203 MHz, DMSO) δ 150.40, 26.82. HRMS **(ES$^+$)** calculated: 1008.063, **[M+K]** found: 1046.3972.

3.9 Supplementary Figures

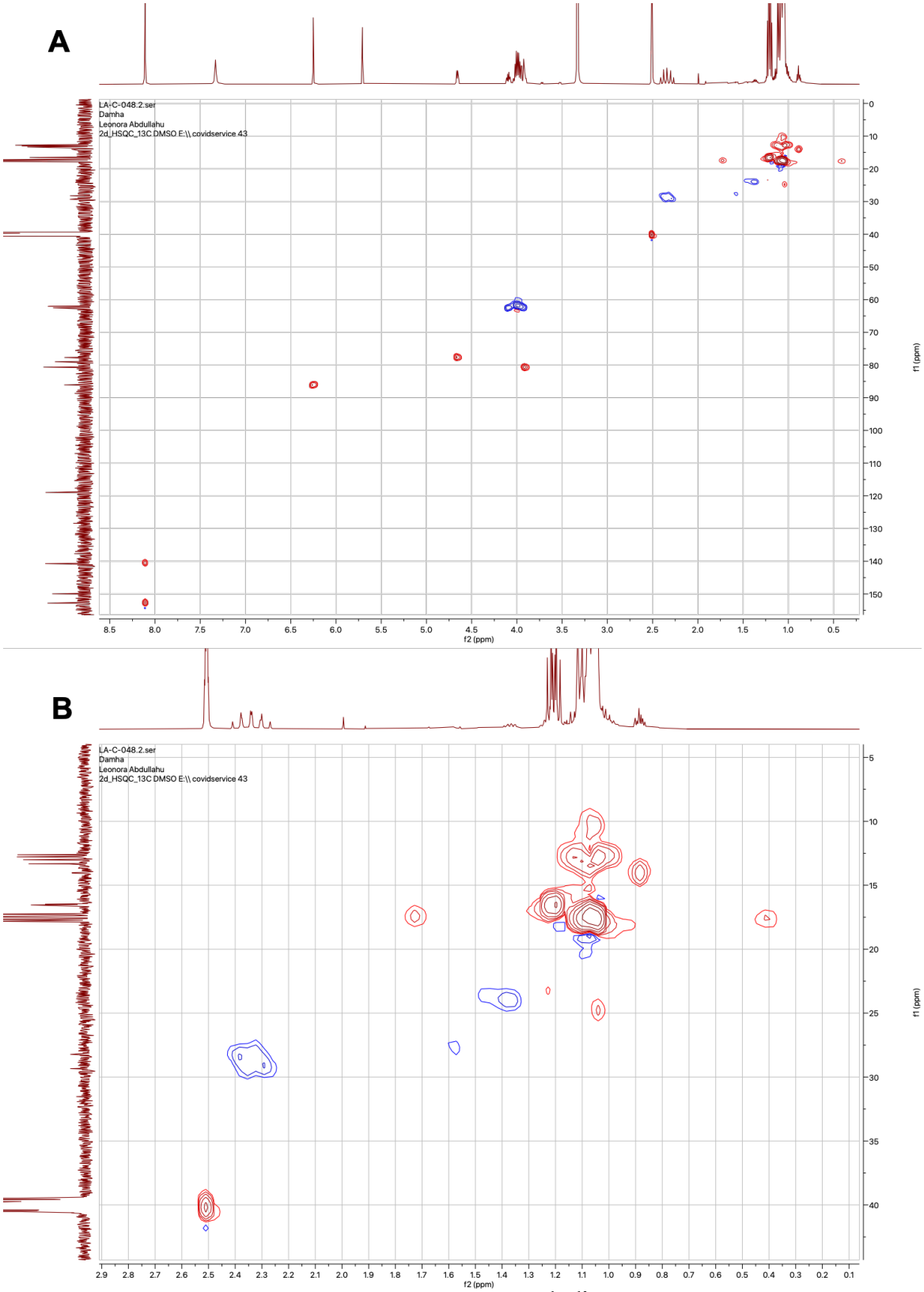


Supplementary Figure 3.1. *^{1}H NMR spectrum of compound 13 with protons assigned.*

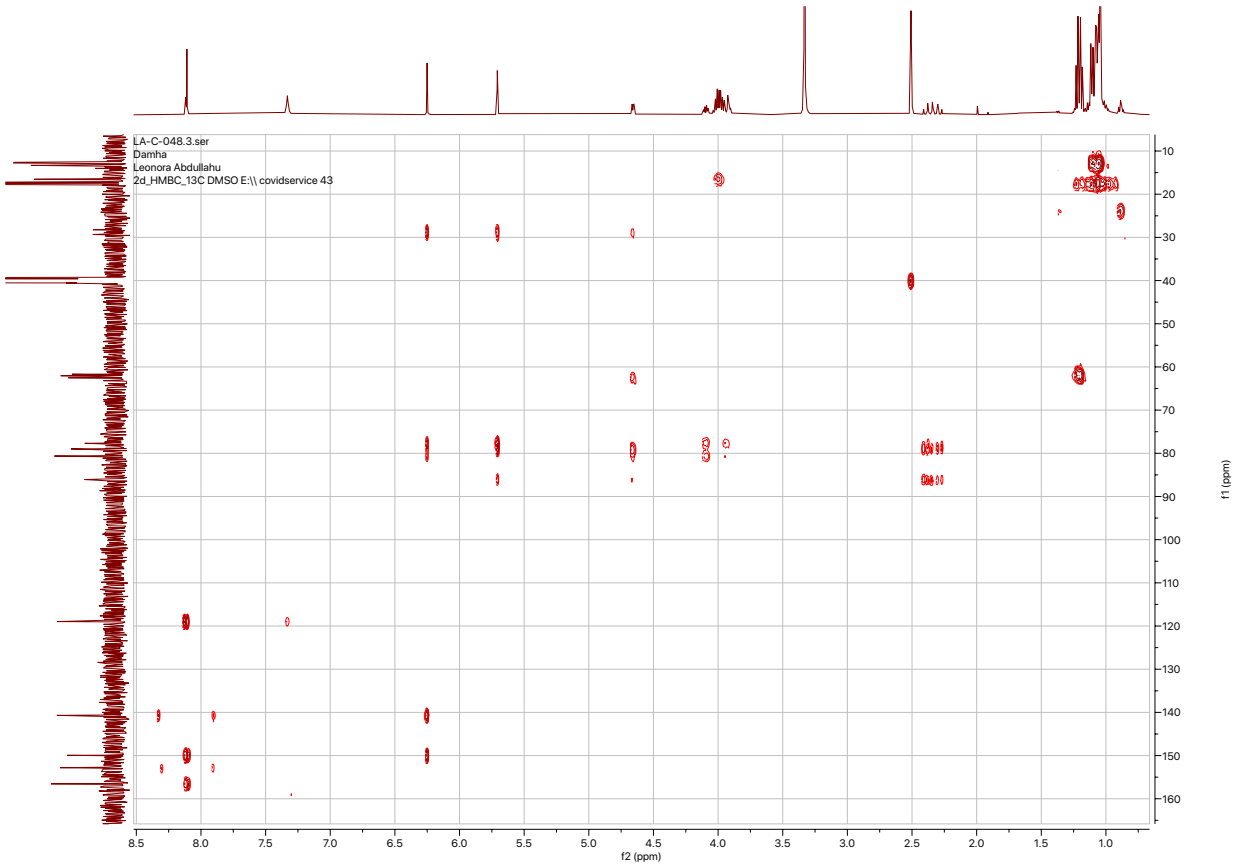


Supplementary Figure 3.2. ^{13}C NMR spectrum of compound 13, with carbons assigned.

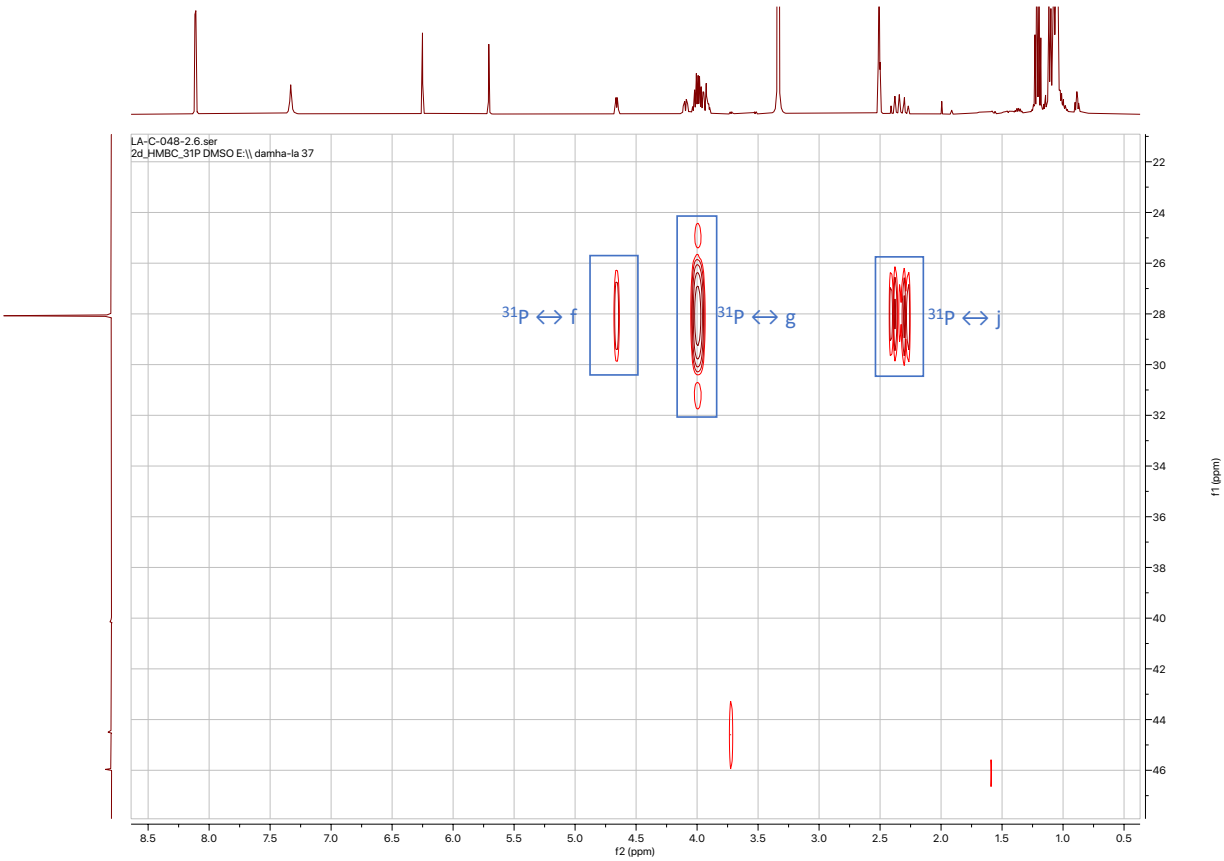
***Supplementary Figure 3.3**. COSY spectrum of Compound 13 with visible ^{1}H-^{1}H coupling assigned.*



Supplementary Figure 3.4.*HSQC spectrum of Compound 13 showing ^{1}H-^{13}C direct coupling. (A) Entire spectrum (B) Close up of 0-3 ppm (^{1}H) and 0-45 ppm (^{13}C).*



Supplementary Figure 3.5. *HMBC spectrum of compound 13 showing ^{1}H-^{13}C through-bond coupling.*



Supplementary Figure 3.6. *^{31}P HMBC spectrum of compound 13 showing ^{1}H-^{31}P through-bond coupling, with peaks assigned.*

3.10 References

1. Culver, G. M.; McCraith, S. M.; Consaul, S. A.; Stanford, D. R.; Phizicky, E. M., A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. *J Biol Chem* **1997,** *272* (20), 13203-10.
2. McCraith, S. M.; Phizicky, E. M., An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. *Journal of Biological Chemistry* **1991,** *266* (18), 11986-11992.
3. Culver, G. M.; McCraith, S. M.; Zillmann, M.; Kierzek, R.; Michaud, N.; LaReau, R. D.; Turner, D. H.; Phizicky, E. M., An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate. *Science* **1993,** *261* (5118), 206.
4. Spinelli, S. L.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Transient ADP-ribosylation of a 2′-Phosphate Implicated in Its Removal from Ligated tRNA during Splicing in Yeast. **1999,** *274* (5), 2637-2644.
5. Steiger, M. A.; Jackman, J. E.; Phizicky, E. M., Analysis of 2'-phosphotransferase (Tpt1p) from Saccharomyces cerevisiae: evidence for a conserved two-step reaction mechanism. *RNA* **2005,** *11* (1), 99-106.
6. Munir, A.; Abdullahu, L.; Damha, M. J.; Shuman, S., Two-step mechanism and step-arrest mutants of Runella slithyformis NAD+-dependent tRNA 2′-phosphotransferase Tpt1. **2018,** *24* (9), 1144-1157.
7. Spinelli, S. L.; Malik, H. S.; Consaul, S. A.; Phizicky, E. M., A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proceedings of the National Academy of Sciences* **1998,** *95* (24), 14136-14141.
8. Sawaya, R.; Schwer, B.; Shuman, S., Structure-function analysis of the yeast NAD+-dependent tRNA 2'-phosphotransferase Tpt1. *RNA (New York, N.Y.)* **2005,** *11* (1), 107-113.
9. Munir, A.; Banerjee, A.; Shuman, S., NAD+-dependent synthesis of a 5′-phospho-ADP-ribosylated RNA/DNA cap by RNA 2′-phosphotransferase Tpt1. *Nucleic Acids Research* **2018**, gky792-.
10. Munir, A.; Abdullahu, L.; Banerjee, A.; Damha, M. J.; Shuman, S., NAD(+)-dependent RNA terminal 2' and 3' phosphomonoesterase activity of a subset of Tpt1 enzymes. *Rna* **2019,** *25* (7), 783-792.
11. Popow, J.; Schleiffer, A.; Martinez, J., Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences* **2012,** *69* (16), 2657-2670.
12. Harding, H. P.; Lackey, J. G.; Hsu, H.-C.; Zhang, Y.; Deng, J.; Xu, R.-M.; Damha, M. J.; Ron, D., An intact unfolded protein response in Trpt1 knockout mice reveals phylogenic divergence in pathways for RNA ligation. *RNA (New York, N.Y.)* **2008,** *14* (2), 225-232.

13. Segal, E. S.; Gritsenko, V.; Levitan, A.; Yadav, B.; Dror, N.; Steenwyk, J. L.; Silberberg, Y.; Mielich, K.; Rokas, A.; Gow, N. A. R.; Kunze, R.; Sharan, R.; Berman, J., Gene Essentiality Analyzed by In Vivo Transposon Mutagenesis and Machine Learning in a Stable Haploid Isolate of Candida albicans. *mBio* **2018,** *9* (5).

14. Banerjee, A.; Munir, A.; Abdullahu, L.; Damha, M. J.; Goldgur, Y.; Shuman, S., Structure of tRNA splicing enzyme Tpt1 illuminates the mechanism of RNA 2'-PO4 recognition and ADP-ribosylation. *Nat Commun* **2019,** *10* (1), 218.

15. McKenzie, L. K.; El-Khoury, R.; Thorpe, J. D.; Damha, M. J.; Hollenstein, M., Recent progress in non-native nucleic acid modifications. *Chemical Society Reviews* **2021,** *50* (8), 5126-5164.

16. El-Khoury, R.; Damha, M. J., 2′-Fluoro-arabinonucleic Acid (FANA): A Versatile Tool for Probing Biomolecular Interactions. *Accounts of Chemical Research* **2021,** *54* (9), 2287-2297.

17. Deleavey, G.; Damha, M. J., Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing. *Chemistry & biology* **2012,** *19*, 937-54.

18. O'Reilly, D.; Kartje, Z. J.; Ageely, E. A.; Malek-Adamian, E.; Habibian, M.; Schofield, A.; Barkau, C. L.; Rohilla, K. J.; DeRossett, L. B.; Weigle, A. T.; Damha, M. J.; Gagnon, K. T., Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity. *Nucleic Acids Res* **2019,** *47* (2), 546-558.

19. Nandakumar, J.; Shuman, S., How an RNA ligase discriminates RNA versus DNA damage. *Mol Cell* **2004,** *16* (2), 211-21.

20. Remus, B. S.; Shuman, S., Distinctive kinetics and substrate specificities of plant and fungal tRNA ligases. *Rna* **2014,** *20* (4), 462-73.

21. Chan, C. M.; Zhou, C.; Huang, R. H., Reconstituting bacterial RNA repair and modification in vitro. *Science* **2009,** *326* (5950), 247.

22. Jain, R.; Shuman, S., Active site mapping and substrate specificity of bacterial Hen1, a manganese-dependent 3' terminal RNA ribose 2'O-methyltransferase. *Rna* **2011,** *17* (3), 429-38.

23. Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; González, C., The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res* **2012,** *40* (18), 9329-39.

24. Damha, M. J.; Wilds, C. J.; Noronha, A.; Brukner, I.; Borkow, G.; Arion, D.; Parniak, M. A., Hybrids of RNA and Arabinonucleic Acids (ANA and 2′F-ANA) Are Substrates of Ribonuclease H. *Journal of the American Chemical Society* **1998,** *120* (49), 12976-12977.

25. Wilds, C. J.; Damha, M. J., 2'-Deoxy-2'-fluoro-beta-D-arabinonucleosides and oligonucleotides (2'F-ANA): synthesis and physicochemical studies. *Nucleic Acids Res* **2000,** *28* (18), 3625-35.

26. Kalota, A.; Karabon, L.; Swider, C. R.; Viazovkina, E.; Elzagheid, M.; Damha, M. J.; Gewirtz, A. M., 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (2'F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing. *Nucleic Acids Res* **2006,** *34* (2), 451-61.

27. Tago, N.; Katolik, A.; Clark, N. E.; Montemayor, E. J.; Seio, K.; Sekine, M.; Hart, P. J.; Damha, M. J., Design, Synthesis, and Properties of Phosphoramidate 2',5'-Linked Branched RNA: Toward the Rational Design of Inhibitors of the RNA Lariat Debranching Enzyme. *J Org Chem* **2015,** *80* (20), 10108-18.

28. Montemayor, E. J.; Katolik, A.; Clark, N. E.; Taylor, A. B.; Schuermann, J. P.; Combs, D. J.; Johnsson, R.; Holloway, S. P.; Stevens, S. W.; Damha, M. J.; Hart, P. J., Structural basis of lariat RNA recognition by the intron debranching enzyme Dbr1. *Nucleic Acids Res* **2014,** *42* (16), 10845-55.

29. Wang, H.; Hill, K.; Perry, S. E., An Arabidopsis RNA lariat debranching enzyme is essential for embryogenesis. *J Biol Chem* **2004,** *279* (2), 1468-73.

30. Guschlbauer, W.; Jankowski, K., Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Research* **1980,** *8* (6), 1421-1433.

31. Liu, Q.; Graeff, R.; Kriksunov, I. A.; Jiang, H.; Zhang, B.; Oppenheimer, N.; Lin, H.; Potter, B. V. L.; Lee, H. C.; Hao, Q., Structural basis for enzymatic evolution from a dedicated ADP-ribosyl cyclase to a multifunctional NAD hydrolase. *The Journal of biological chemistry* **2009,** *284* (40), 27637-27645.

32. Pommier, Y., Drugging topoisomerases: lessons and challenges. *ACS Chem Biol* **2013,** *8* (1), 82-95.

33. Katolik, A.; Johnsson, R.; Montemayor, E.; Lackey, J. G.; Hart, P. J.; Damha, M. J., Regiospecific solid-phase synthesis of branched oligoribonucleotides that mimic intronic lariat RNA intermediates. *J Org Chem* **2014,** *79* (3), 963-75.

Chapter 4

2'-Amino-modified crRNA for studying CRISPR Cas9 and Cas12a systems

4.1 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and their CRISPR-associated (Cas) enzymes represent an ancient bacterial immune system evolved to fight invading viral pathogens[1-8]. The core enzymatic component is a ribonucleoprotein (RNP) comprised of one or more Cas proteins, with at least one being a nuclease, that are associated with one or two CRISPR RNAs (crRNAs)[5, 6, 9-18]. After recognition of a short protospacer adjacent motif (PAM) sequence by the Cas endonuclease, base-pairing of crRNA nucleotides (nucleotides) to target DNA then guides sequence-specific and site-specific endonucleolytic cleavage of DNA phosphodiester bonds[19-22].

Cas9 from Streptococcus pyogenes (SpCas9) and Cas12a, also known as Cpf1, from *Acidominococcus* species (AsCas12a) are among the most heavily co-opted and engineered CRISPR-Cas systems for creation of model organisms and development of human therapeutics[23-34]. They both efficiently induce DNA double-strand breaks and guide genome editing in mammalian cells[17, 24, 29, 34-39], utilize a single Cas enzyme and can function with a single guide RNA (sgRNA), although Cas9 functions as a dual-guide RNA in nature[6, 17, 29, 34]. While these enzymes can ultimately result in similar gene editing outcomes, such as genetic knock out or donor DNA sequence knock-in, they are quite distinct. Cas12a is a smaller protein that naturally utilizes a single crRNA guide (39 nucleotides) less than half the size of the commonly used artificial Cas9 sgRNA (99 nucleotides) [17, 22, 40]. Cas12a naturally recognizes thymine (T)-rich PAMs whereas SpCas9 uses guanine (G)-rich PAMs[6, 17, 20, 41]. The Cas12a crRNA adopts a pseudoknot structure at its 5' end, as opposed to a series of stem structures for SpCas9, which creates a ‘handle’ for the Cas12a to bind[22, 40].

CRISPR-Cas systems hold tremendous promise for revolutionizing medicine, particularly through improved gene therapy or gene-targeted approaches[25, 42-51]. To achieve this potential, however, various challenges must be overcome. These include long-term safety, delivery, and efficacy[25, 38, 47, 49-54]. Lipid nanoparticles are a leading technology being developed for delivery of Cas enzymes, either as protein or as mRNA that encode the Cas protein, along with their cognate CRISPR RNA guides[51, 55-59]. In principle, this

may allow greater spatial and temporal control over tissue and cell targeting[55] and provide a better window of safety when compared to therapeutic delivery approaches that involve genetically encoded expression of the CRISPR-Cas system, such as adeno-associated viruses[56]. However, RNA is not an ideal drug due to its size and lability – it is quickly degraded in the body. Chemical modification has emerged as a practical necessity to stabilize and control the half-life, bioavailability, and activity of RNA and nucleic acid drugs[60, 61]. Chemical modification has been a pillar in the growing success of other nucleic acid-based therapeutics, including antisense oligonucleotides (ASOs) and small interfering RNAs (siRNAs)[53, 60, 62]. The lessons learned from these technologies can be applied to CRISPR-based therapeutics. The degree of chemical modification required for crRNA guides is unclear and may depend on delivery or target tissue. Nonetheless, ultimate control over stability, protein interaction, enzyme activity, and pharmacology would ideally be achieved by understanding what chemical modifications are tolerated at every position within a crRNA guide[44, 59, 60, 63-66].

Previous investigations of CRISPR-Cas guide RNA chemical modification have primarily focused on improving therapeutic gene editing by increasing nuclease resistance, understanding biochemical limitations, and reducing off-target editing. Minimal addition of modified nucleotides or inter-nucleotide linkages at the guide RNA 5' and 3' termini have been shown to maintain or increase editing, presumably through reduced nuclease sensitivity. These have included 2'-*O*-methyl with phosphorothioate (MS) or 2'-*O*-methyl-3'-thio PACE (MSP) for Cas9 [67, 68]. In the Cas12a system, 3' end capping has been explored with 2'-F, 2'-*O*-methyl, unlocked nucleic acid (UNA), locked nucleic acid (LNA) and phosphorothioate (PS) [69] while 5' end capping has been explored by addition of short MS or DNA-PS modified sequences onto the 5' handle [70]. Interestingly, extension of the 3' end with 2'-*O*-methyl nucleotides [71] or the 5' handle with DNA-PS [72] has also been reported to improve editing or increase the kinetics of *trans* cleavage activity for Cas12a, respectively.

More complete or systematic modification of Cas9 and Cas12a guide RNAs has been explored using a variety of chemistries. DNA (2'-deoxy) has been reported to alter enzyme activity or specificity while helping to probe 2'-hydroxyl requirements of the enzyme [64, 73-75]. A broader array of modifications, like 2'-F, 2'-*O*-methyl, bridged nucleic

acid (BNA), UNA, LNA, 2'F-arabinonucleic acid (FANA), and PS have been tested alone or in combination as well [59, 65, 66, 76-79]. These studies largely demonstrated that much of the guide could be modified, with moderate to high editing preserved, but particular positions, such as seed regions or individual residues, required RNA nucleotides. Specificity could also be improved with careful placement of modifications in the guide region [73, 74, 78, 79]. For Cas9, retention of editing activity has been suggested to correlate with conserved polar contacts between the protein and the hydroxyl at the ribose 2' position and the need to retain A-form-like helical architecture [64-66, 75]. For these apparently conserved positions between Cas9 and its guide, RNA cannot be readily replaced.

For Cas12a, attempts to fully modify the crRNA guide have proven unsuccessful [69, 73, 76]. However, the 5' pseudoknot handle posed the greatest challenge as it appeared to tolerate very little modification to the ribose, suggesting that successful modification of this unique structure would be the major limitation to greater modification of Cas12a guides for therapeutic development. To address this bottleneck, we took a rational, structure-guided approach with respect to predicted 2'-hydroxyl (2'-OH) polar contacts. Utilizing chemical modifications that specifically explore ribose and 2'-OH properties, we successfully modified the 5' pseudoknot and obtained very robust gene editing when as few as 6 out of 19 residues remained unmodified. The key positions that remained resistant to modification correlated with predicted 2'-OH polar contacts within the Cas12a-crRNA crystal structure[22, 40]. The bridging phosphodiester bonds at these positions could further be modified to PS and retain high *cis* and *trans* cleavage activity, as well as gene editing. When crRNAs with chemically modified 5' pseudoknot handles were screened for *trans* cleavage activity, some induced differential activities like high *trans* activity with little or no *cis* dsDNA target cleavage. These results were likely due to the high catalytic turnover of *trans* activity even when very modest *cis* cleavage occurred. Together, these results advance efforts toward chemical modification of AsCas12a guides for therapeutic editing and potential diagnostic applications.

4.2 History of 2′-Amino-2′-Deoxynucleosides

The utility of oligonucleotide therapeutics is considerably enhanced by chemical modifications that lend resistance to nuclease attack. In particular, substitution at the 2′ position of ribonucleotides with 2′-amino (2′-NH_2), 2′-fluoro (2′-F), or a variety of 2′-O-alkyl moieties confers resistance to ribonucleases that utilize the 2′-OH group for cleavage of the adjacent phosphodiester bond[80]. As modifications of the 2′ of the pentose sugar of nucleotides continue to be further developed and better studied, the 2′-NH_2 modification remains relatively uncommon, despite having been well studied since the 1970s by Fritz Eckstein and Krysztof Jankowski[81, 82]. Guschlbauer and Jankowski first discovered that nucleoside conformation is influenced predominantly by the electronegativity of the sugar substituents[82]. Their extensive NMR studies on ribosyl nucleosides revealed a linear relationship between the electronegativity of the 2′-substituent and the carbon-13 chemical shift of C2′. Detailed analysis of proton NMR spectra allowed the proton-proton coupling constants to be used to compute the conformational equilibria between the canonical N (C3′-*endo*) and S (C2′-*endo*) conformers of uridines (and later adenosines) with varying 2′-substituents. They found that the N-form conformation increases in contribution with the electronegativity of the 2′-substituent. Thus, the most electronegative the substituent pulls the pucker to its side. This is apparently true for both C2′ and C3′-substituents. They then verified the same line of reasoning for arabinosyl nucleosides, where electronegative substituents at the C2′ favour the S conformation. One observation that stood out amongst the trend, however, was the 2′-NH_2 substitution. Despite being more electronegative than the 2′-deoxy (2′-H) ($^{\varepsilon}$R dUn = 2.9 vs dU = 2.2), it consistently preferred a conformation even more S than 2′-H (preference for N conformer = 0.25 vs 0.40) (**Figure 4.1**). From this observation, it is clear that electronegativity is not solely responsible for the conformational effects observed.

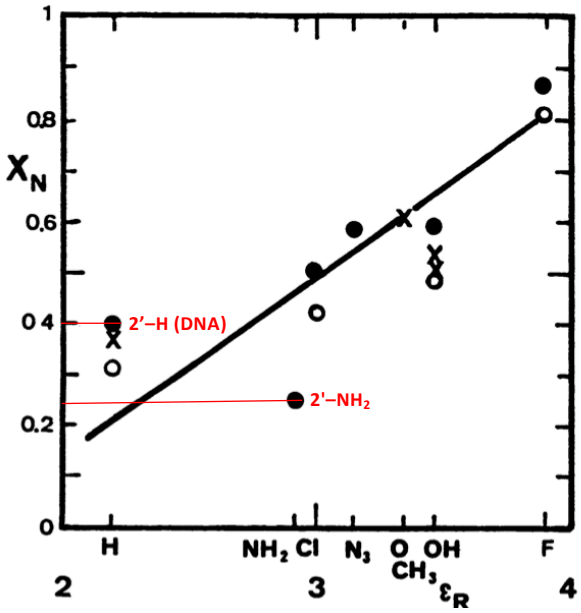


Figure 4.1. *Mole fraction of N conformation (x_N) of 2'-substituted uridines as a function of the electronegativity (^{e}R) of the substituent. o in DMSO; • in H_2O; x data from the literature. Image adapted from reference 87.*

For the CRISPR studies described in this chapter, we focus heavily on 2'-amino-2'-deoxy- or 2'-fluoro-2'-deoxynucleotides at selected positions. These analogues are especially interesting when comparing to the native ribonucleosides for a couple of reasons. First, as mentioned above, they differ significantly in their preference for C3'-*endo* (N), which is characteristic for ribonucleosides, over C2'-*endo* (S) conformation, with 2'-F displaying a N:S ratio greater than that of ribonucleosides, and 2'-NH_2 favouring the S conformation even more than 2'-deoxynucleosides (2'-H). Second, these analogues differ in their ability to participate in hydrogen bonding. The 2'-F atom can only serve as a hydrogen-bond acceptor, however, the 2'-NH_2 group can act as both a donor and acceptor, much like the 2'-OH group of ribose, making it an ideal analogue for assessing hydroxyl contact requirements. While the pK_a of a typical primary alkyl amine is around 10.6, the pK_a of 2'-amino-2'-deoxyuridine is reported to be 6.2 for the amino group, due to inductive effects of the sugar ring and the nucleobase[83-86]. Thus, the 2'-NH_2 is expected to be mostly non-protonated during the experiments carried out at physiological pH.

Another aspect of the 2'-NH_2 modification incorporated into an oligonucleotide that has previously been studied is its influence on thermal stability of the resultant duplexes

with either DNA or RNA. Eckstein was the first to report on the thermal stability of duplexes consisting of oligodeoxy- or oligoribonucleotides containing 2'-amino-2'-deoxycytidines in one strand and natural 2'-deoxy- or ribonucleotides in the other strand. The influence of the 2'-amino substituent can be summarised as having a strong destabilising effect on thermal stability. For the RNA/RNA duplex the T_m was decreased from 52.5 °C for the unmodified duplex to 48.2 °C by the incorporation of one single 2'-aminocytidine into a strand and decreased further to 39.9°C by the incorporation of three 2'-aminocytidines. The effect was even more dramatic for the DNA/DNA duplexes where the T_m was decreased from 30.1 °C to around 5 °C upon the incorporation of three 2'-aminocytidines. In a duplex where the triple 2'-aminocytidine-containing strand was in an RNA context and the other strand was the DNA strand, the T_m was lowered from 26.8 °C to <0 °C. From even earlier literature, it was known that a homopolynucleotide of 2'-amino-2'-deoxyuridylic acid did not form a duplex with poly(rA)[81].

Based on the pK_a data mentioned above, one would expect that the thermal stability of duplexes containing 2'-aminocytidines would be increased by decreasing the pH below the pK_a value so that an ion pair could form between the protonated amino group and the phosphate. However, Eckstein found that the T_m was even lower at pH 5.0 than at pH 7.0 for the RNA/RNA duplex (23.6 ˚C vs 39.9 ˚C). Thus, ion pair formation had the opposite effect to what was expected. Eckstein further notes that if it was purely conformation of the sugar that was responsible for the stability of a duplex, one should expect the presence of 2'-aminonucleosides to destabilise an RNA/RNA but not a DNA/DNA duplex. These nucleosides exist predominantly in the 2'-endo conformation, typical for the B-helix of DNA and would thus be expected not to distort a DNA helix. However, the 2'-amino group, although conferring the 2'-endo conformation on the 2'-aminonucleoside, is similar in size to the hydroxyl group and will thus also exert a steric barrier inhibiting the adoption of a 2'-endo conformation when present in a B-helix. From this, it can be understood why 2'-aminonucleosides destabilise even DNA/DNA duplexes. Despite the supposed drawback of the destabilizing effect of 2'-aminonucleosides, they remain an attractive analogue due to other advantages gained, such as nuclease stability, H-bonding abilities, and the presence of a unique functional group.

4.2.1 2′-Amino-2′-deoxynucleosides for CRISPR-Cas systems

While the majority of this chapter focusses on our study of Cas12a, it is important to highlight some of our initial studies that were carried out on the Cas9 system, as this system laid the groundwork for subsequent Cas12a studies. Intrigued by the characteristics of the 2′-amino analogue, we decided use this modification to build upon a study that was previously ongoing in our lab by O'Reilly and colleagues[66]. To briefly summarize their findings, they were using chemical modification of the guide RNA of CRISPR-Cas9 to characterize structure-activity relationships within CRISPR ribonucleoprotein (RNP) enzymes and identifying compatible chemistries for controlling activities. They found that ribose sugars that promoted or accommodated A-form helical architecture in and around the crRNA seed region were tolerated best. Cas9 activity in vitro tolerated most chemical modifications at predicted 2′-hydroxyl contact positions, whereas editing activity in cells was much less tolerant (**Figure 4.2**). We wanted to explore the requirements for 2′-hydroxyl contacts further, and thus decided to incorporate 2′-amino-2′-deoxynucleotides at those 6 contact positions.

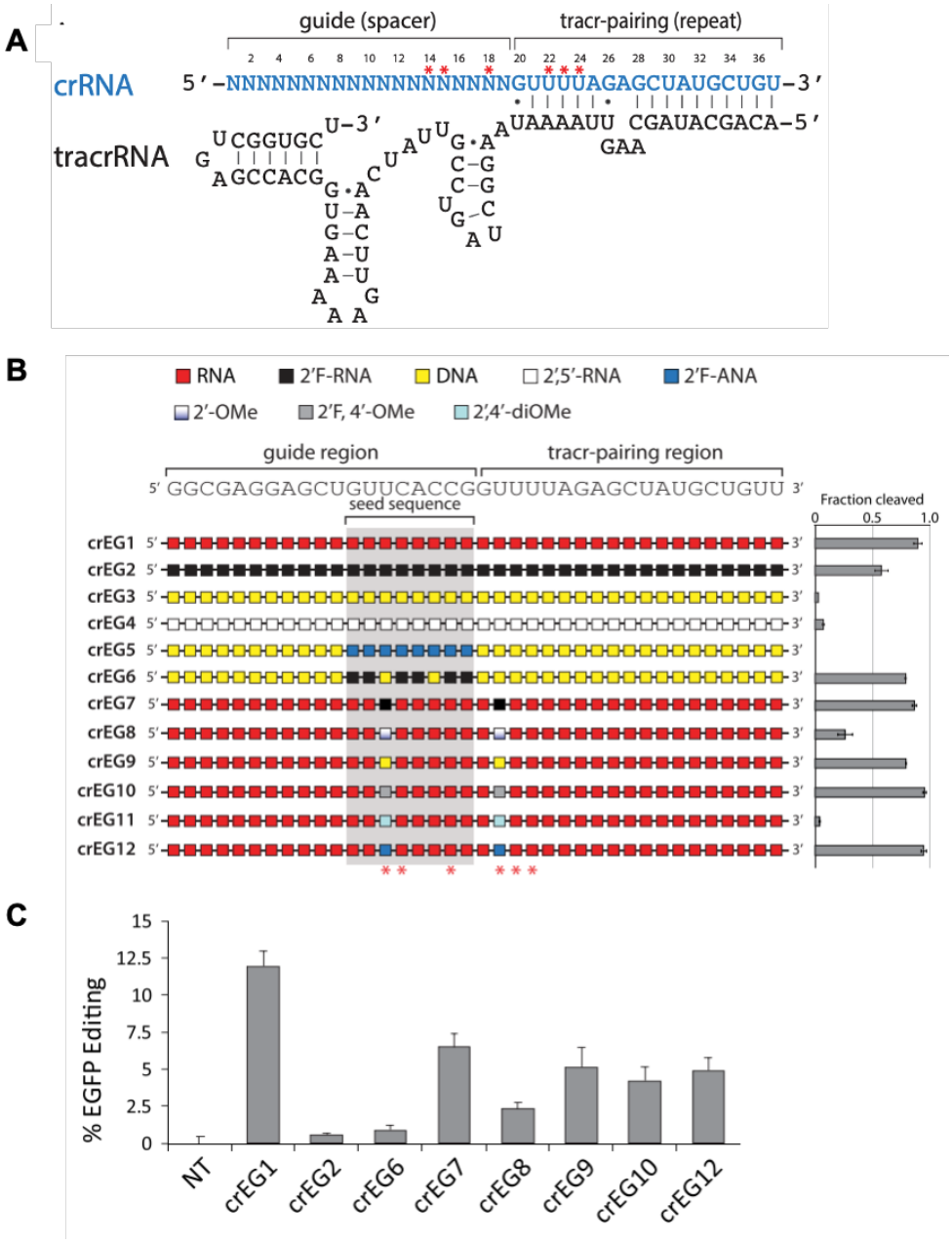
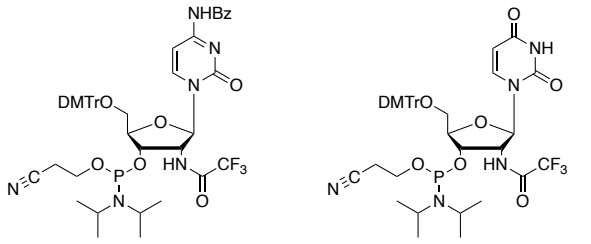


Figure 4.2. *(A) Sequence and secondary structure of a dual RNA guide for Streptococcus pyogenes Cas9. Asterisks indicate structure-predicted 2-OH contacts with Cas9. (B) In vitro cleavage activity of Cas9 RNPs assembled with chemically modified crRNAs targeting an EGFP gene. crRNA sequence is shown above and structure-predicted 2-OH contacts with Cas9 are indicated with asterisks below. Enzyme activity is reported to the right. Error is reported as s.e.m. for three or more replicates.(C) Cell-based editing activity of crRNAs cotransfected with tracrRNA into HEK 293T cells stably expressing EGFP and Cas9. Gene editing efficiency was measured as loss of EGFP fluorescenceby flow cytometry 5 days post-transfection. Error is reported as s.e.m. for four replicate treatments.Figure adapted from reference 66.*

We initially set out to synthesize the crRNA guide strand used in the previous studies but replacing all 2′-OH contact positions with 2′-NH_2, by incorporating commercially available 2′-amino phosphoramidites (C & U) (ChemGenes Corporation)

into standard solid-phase synthesis to make crEG-N1 (**Figure 4.3**). We were disappointed to observe that this sequence yielded a very poor synthesis, and the Ion Exchange (IEX) HPLC trace of the resulting oligonucleotide was indicative of a degraded sample (**Figure 4.4**). To gauge whether the amino residues were the causing the degradation issue, we then synthesized crEG-N3, to contain only two 2′-NH_2 residues (one in the seed region and one outside of it), and the outstanding four hydroxyl contacts remained 2′-OH (RNA). While the IEX HPLC trace of this crude oligo also showed evidence of degradation, there was one peak which stood out amongst the others, which we were able to purify in hopes of further characterization (**Figure 4.4**). We were once again disappointed to observe that LCMS characterization of the single peak we isolated *via* HPLC purification had apparently degraded at some point between purification and characterization, as seen in the LCMS trace in **Figure 4.4** (bottom). This degradation was confirmed when we reinjected the purified sample into the HPLC and saw complete degradation there as well (not shown).



crEG-N1: 5'-GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU-3'

crEG-N3: 5'-GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU-3'

N = 2'-OH (RNA) N = OH contact position N = 2′NH_2 N = seed region

Figure 4.3. *Commercially available 2'-Amino CED phosphoramidites (top) and Cas9 crRNA guide sequences made to incorporate 2'-NH_2 at hydroxyl contact positions (bottom) with the seed region highlighted in cyan.*

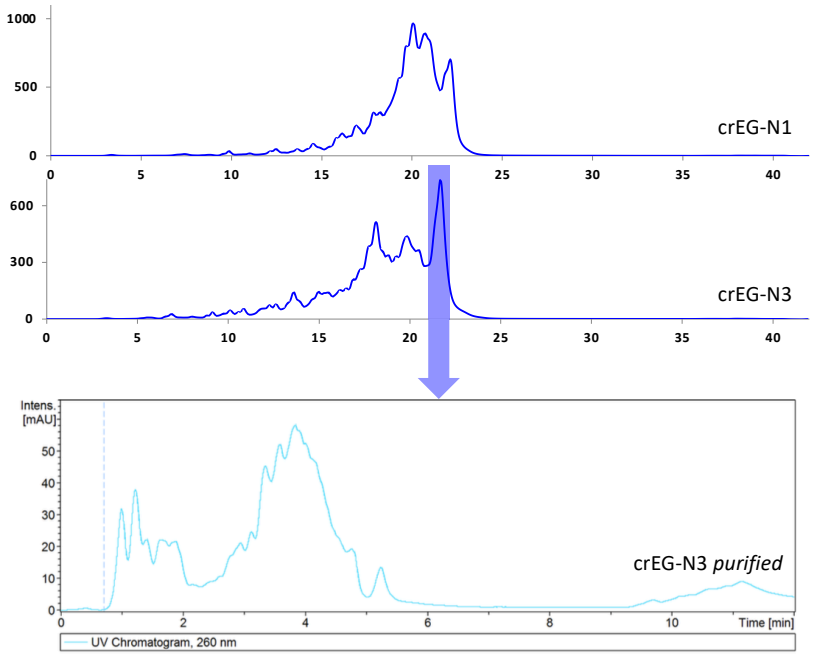


Figure 4.4. *IEX HPLC spectra of crude samples crEG-N1 and crEG-N3 (top). LCMS chromatogram of purified peak from crEG-N3 sample showing degradation of the sample (bottom).*

4.2.2 Optimizing the deprotection of 2′-amino 3′-O-phosphoramidites

At this point, we hypothesized that perhaps our deprotection conditions were causing the observed degradation issue. The supplier's deprotection conditions for the trifluoroacetate protecting group on the amino amidites indicated standard ammonia deprotection conditions, which means that the 2′-NH_2 were being deprotected during ammonolysis, at the same time as the oligo was being cleaved from the solid support, the base protecting groups were being removed, and the cyanoethyl backbone protecting groups were being removed. One area for concern we identified was that if the 2′-NH_2 was being deprotected faster than the cyanoethyl phosphate backbone protecting group, perhaps the exposed 2′-NH_2 was cleaving the backbone phosphotriester (which we know is more susceptible to cleavage than the resulting phosphodiester). To test this hypothesis, we designed a quick analysis of various deprotection conditions. We synthesized a 10mer polyU with a single internal 2′-NH_2 residue:

5′-U UUU UUU UUU-3′

Our test conditions included: Test 1. Regular ammonolysis (CPG in in NH_4OH/EtOH, 65˚ C 16 hrs); Test 2. Removal of cyanoethyl protecting groups while on solid support (3:2 Et_3N:MeCN, 4x 20 min), followed by ammonolysis (NH_4OH/EtOH, 65˚ C 16 hrs); and Test 3. Fast deprotection (40 % aqueous methylamine, 65˚ C 10 minutes). All of these methods were followed by standard desilylation conditions to remove the remaining RNA silyl protecting groups. Crude HPLC traces of the resulting oligonucleotides are shown in **Figure 4.5**. Already, we saw that a single amino insert (as opposed to two or more), resulted in much cleaner syntheses, with more defined peaks. In order to identify what each peak corresponds to, we isolated the individual peaks and characterized them by LCMS. The desired product, with an observed mass of 2996.3286, came out at 15 minutes, and was observed in all three deprotection condition tests. Tests 1 and 2 gave almost identical traces, signifying that the cyanoethyl phosphate protecting group is most likely cleaved before the trifluoroacetate amino protecting group under standard ammonolysis. The peak at 14 minutes, observed in all three test conditions (Δ = -308.0352) is indicative of an N-1 peak (one uridine residue failing to couple during the synthesis), and speaks more to the quality of the synthesis, as opposed to the deprotection conditions being tested. The peak at 16 minutes, observed only in Test 3 (Δ = +31.9620) indicated incomplete deprotection, so the fast deprotection was most likely not long enough to ensure complete deprotection. It is not recommended to perform this deprotection on RNA for longer than 10 minutes, as it can lead to degradation of the oligonucleotide.

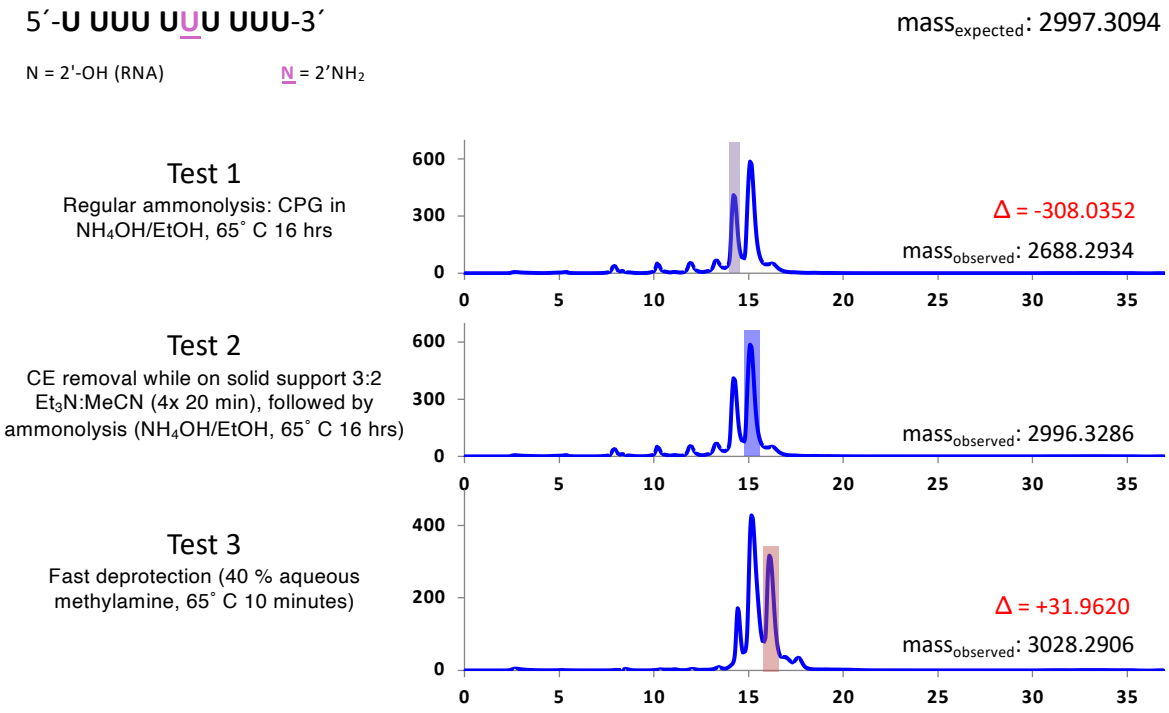


Figure 4.5. *Testing different deprotection conditions for the most effective removal of 2'-NH$_2$ triflouroacetal protecting group.*

4.2.3 Synthesis of 2′-NH$_2$-containing Cas9 crRNA

At this point, it became clear to us that if we wanted to successfully incorporate 2′-NH$_2$ analogues into our crRNA, we should only incorporate one 2′-NH$_2$ per oligonucleotide. With that in mind, we synthesized a series of oligonucleotides, walking a single 2′-NH$_2$ residue throughout the six 2′-hydroxyl contact positions (**Table 4.1**). While these syntheses also showed large amounts of degradations as well, we were able to purify and confirm that we had isolated our desired oligonucleotides. A sample HPLC trace along with LCMS characterization of isolated peaks (from the LA-N-14-6) is provided in **Supplementary Figure 4. 1**.

Table 4.1. Cas9 crRNA series containing single 2′-NH_2 inserts at predicted 2′-hydroxyl contact positions.

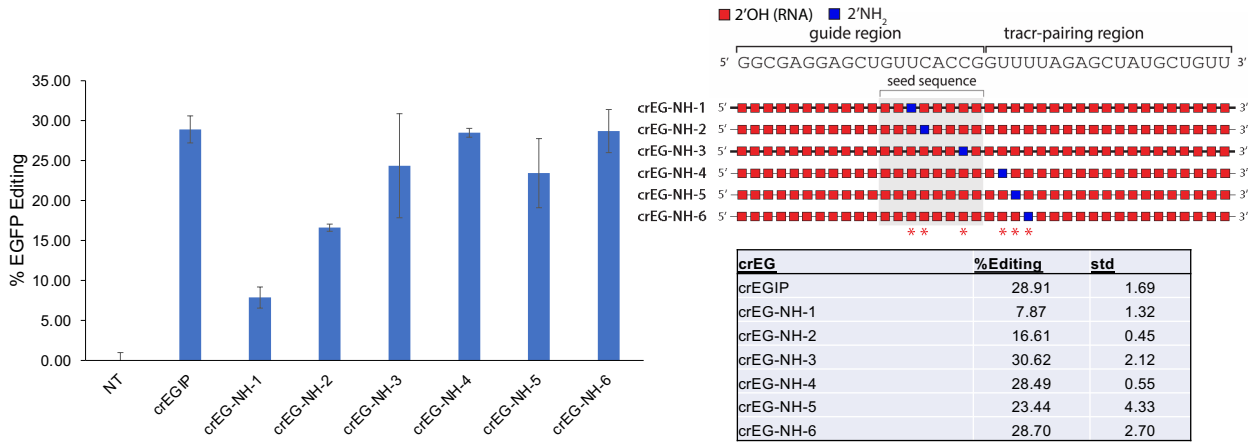
Sample	Sequence	M_{exp}	M_{obs}
crEG-NH-1	5'-**GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU**-3'	12210.2105	12209.6580
crEG-NH-2	5'-**GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU**-3'	12210.2139	12209.6753
crEG-NH-3	5'-**GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU**-3'	12210.2139	12209.4910
crEG-NH-4	5'-**GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU**-3'	12210.2105	12210.0585
crEG-NH-5	5'-**GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU**-3'	12210.2105	12209.8044
crEG-NH-6	5'-**GG CGA GGA GCU GUU CAC CGG UUU UAG AGC UAU GCU GUU**-3'	12210.2105	12209.7900

guide region | tracr-pairing region

N = 2'-OH (RNA) **N** = OH contact position N = 2′NH_2 **N** = seed region

4.2.4 In cell editing activity of 2′-NH_2-containing Cas9 crRNA

The chemically modified 2′-NH_2 crRNA series was tested for cell editing activity *via* cotransfection with tracrRNA into HEK293T cell expressing EGFP and Cas9. The results summarized in **Figure 4.6** indicate that a single 2′-NH_2 insert is tolerated throughout the crRNA, but slightly less so in the seed region. This falls in agreement with previous modifications that were tested which also showed increased sensitivity in the seed region. Interestingly, however, the 2′-NH_2 modification is not traditionally A-form-like, as it adopts a C2′-endo conformation at the nucleoside level, and thus favours the B-form. Yet, it is still highly compatible with the system. As indicated by O'Reilly *et al.*, flexibility then may be an important factor in conformational transitions of Cas9, particularly in the guide region[66]. This is expected, as the guide region is what must bind to the DNA target and must assume an A-form-like conformation for efficient catalysis.



crEG	%Editing	std
crEGIP	28.91	1.69
crEG-NH-1	7.87	1.32
crEG-NH-2	16.61	0.45
crEG-NH-3	30.62	2.12
crEG-NH-4	28.49	0.55
crEG-NH-5	23.44	4.33
crEG-NH-6	28.70	2.70

Figure 4.6. *Cell editing activity of chemically modified 2'-NH_2 crRNAs cotransfected with tracrRNA into HEK 293T cells stably expressing EGFP and Cas9. Gene editing efficiency was measured as loss of EGFP fluorescenceby flow cytometry 5 days post-transfection. Error is reported as s.e.m. for four replicate treatments.*

As stated above, these initial studies of the 2′-NH_2 modification were used to study the Cas9 system, as it was the one being studied in our lab at the time. These studies allowed us to troubleshoot major hurdles in the synthesis of 2′-NH_2-containing oligonucleotides and helped us establish guidelines for the amount of 2′-NH_2 we could incorporate into a sequence. We then turned our attention back to the Cas12a system, which we will heavily discuss in the remainder of this chapter.

4.3 Results for Gene Editing with Cas12a Guides Possessing Ribose-Modified 5' Pseudoknots

Acidominococcus species Cas12a (AsCas12a) and a 39-nt crRNA were used for this study. The crRNA can be structurally and functionally divided into two domains: a 19-nt 5' handle with a pseudoknot structure that anchors crRNA binding to Cas12a and a 20-nt 3' guide region that base-pairs to complementary target strand DNA (**Figure 4.7**A). A crystal structure of AsCas12a has enabled prediction of potentially important polar contacts or hydrogen bonds between AsCas12a and the RNA via 2'-hydroxyl (2'-OH) groups in the pseudoknot structure, as well as intramolecular 2'-OH RNA-RNA contacts[22]. The importance of RNA A-form helical structure and 2'-OH chemistry has only been partially investigated for pseudoknot stability and protein interaction[87, 88] but not in the context of CRISPR-Cas systems. The unique non-canonical nature of the Cas12a 5' pseudoknot, as

well as previous attempts at modification[69, 76], suggested that it would be challenging to modify for structure-function probing and therapeutic development. To focus on the pseudoknot structure and uncouple it from guide region effects, we maintained an entirely native RNA ribose (2'-OH) guide region throughout most of this study. This was accomplished through splint ligation of a 5' phosphorylated guide RNA and chemically modified pseudoknot RNA using T4 DNA ligase[89] (Figure S1) or direct solid-phase chemical synthesis of the entire crRNA when modification schemes prevented efficient enzymatic ligation. Enzymatic ligation facilitated more cost-effective and modular production of full-length crRNAs. To investigate the role of the ribose 2'-OH and A-form structural preferences in 5' pseudoknot structure-function for AsCas12a, we substituted RNA nucleotides with 2'-deoxy (DNA), 2'-fluoro (2'-F), oxepane (Ox), 2'-arabino (2'-araOH), and 2'-amino (2'-NH_2) residues (**Figure 4.7**B).

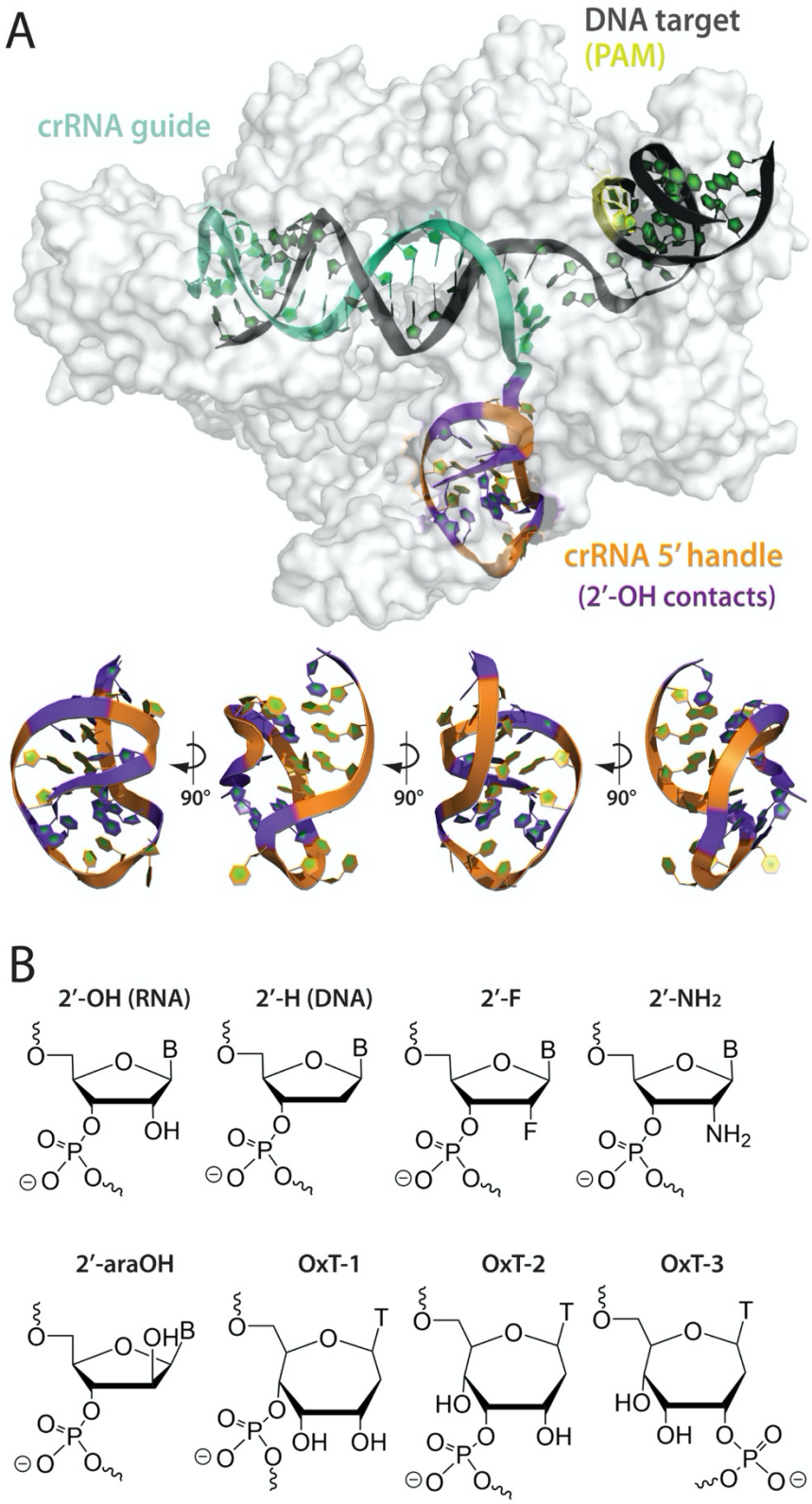


Figure 4.7. *(**A**) AsCas12a 5' handle pseudoknot structure (PDB ID 5B43). Cas12a protein is rendered as a surface (white), DNA target as dark grey, PAM sequence in yellow, crRNA guide in teal, 5' handle in orange, and 5' handle nucleotides that make 2'-OH polar contacts within the structure in purple. (**B**) Chemically modified ribose analogs and substitutes used in this study.*

4.3.1 Probing 2'-Hydroxyl Contacts and A-Form Helical Preference with 2'-Deoxyribose Nucleotides

While DNA nucleotides can probe the importance of 2'-OH contacts, they introduce conformational flexibility and can therefore also probe the importance of A-form structural preferences[64, 90, 91]. Pseudoknots are non-canonical RNA structures that have also been generated with single-stranded DNA[92, 93]. Thus, we initially synthesized a crRNA with all nucleotides in the pseudoknot substituted with DNA (cpEGIP-D1). No sequence- specific (*cis*) cleavage activity was observed for this design in vitro, indicating that either 2'-OH contacts, A-form helical structure, or both were necessary for Cas12a ribonucleoprotein (RNP) assembly or enzyme activity (**Figure 4.8**A). Based on an AsCas12a crystal structure[22], we identified putative critical 2'-OH contacts in the pseudoknot structure at positions -1, -6, -10, -13, -14, -17, -18, and -19. Converting these positions back to RNA, except -19, (cpEGIP-N1) completely rescued in vitro cleavage activity. These results, along with inspection of the Cas12a crystal structure, suggested to us that the 2'-OH at residue -19 may not be a critical polar contact that impacts activity. Conversely, an all-RNA pseudoknot with the same seven potentially critical 2'-OH contact positions (omitting position -19) replaced with DNA (cpEGIP-N2) only provided a quarter of the normal AsCas12a activity. Thus, these results supported the critical role of 2'-OH contacts for either pseudoknot structure or protein interaction. However, the observation that cpEGIP-N2 still conferred cleavage activity with no 2'-OH groups at the predicted critical positions also suggested that A-form helical structure or C3'-endo sugar pucker plays an important, albeit potentially lesser role. Indeed, adding additional DNA nucleotides to cpEGIP-N2 to create cpEGIP-A1, which possessed only 6 RNA residues, abrogated cleavage activity. The activity of cpEGIP-N2, although low, also suggested that pseudoknot 2'-OH contacts are not obligatory for activity.

To further test the role of individual positions and explore RNA-DNA combinations, we synthesized and tested several more crRNAs. Starting with a design that preserved the putative critical 2'-OH contacts and included more RNA residues (cpEGIP-B1), we replaced -19, -18, and -17 positions individually (cpEGIP-B1a, -B1b, and -B1c).

None of these substitutions were detrimental to activity (**Figure 4.8**A). We then added more DNA residues, including at positions -14 and -1. This resulted in substantially reduced activity (cpEGIP-B1d). Adding RNA back to position -14 and -15 improved activity (cpEGIP-B1e). Placing a DNA nt at the putative critical 2'-OH position -6, as well as -3, showed even greater enzyme activity (cpEGIP-B1f), suggesting that -6 is more tolerant of 2'-OH loss. Comparing two pseudoknots with the same DNA substitution pattern but with or without DNA at position -1 (cpEGIP- B1g and -B1h) revealed that DNA at -1 was not only well tolerated but seemed to enhance activity. These results suggested that no specific critical 2'-OH position was necessarily more important than another and instead that their cumulative loss is additive and negatively impacts enzyme activity.

To test both specific positions and cumulative DNA effects, we systematically walked overlapping blocks of 4 to 6 DNA nucleotides across the pseudoknot sequence (cpEGIP-E1 through -E5). AsCas12a exhibited high activity with all these crRNAs (**Figure 4.8**A). This result supports the notion that no individual RNA position is essential and high activity can be achieved in vitro with DNA substitutions if sufficient RNA nt content is preserved. Indeed, a new series of crRNAs where we added increasing numbers of DNA residues, including at potentially critical 2'-OH contact positions (cpEGIP-F1 through -F6), abrogated AsCas12a activity. These results support the important role of A-form helical preference or the reduced flexibility introduced by C3'-endo sugar pucker.

Previous studies by us and others have determined that specific native 2'-OH ribose chemistry (RNA nucleotides) is not necessary for in vitro cleavage but is required for efficient gene editing in cells by CRISPR-Cas9 from *Streptococcus pyogenes*[65, 66, 75]. Critical 2'-OH contact positions lie in the SpCas9 crRNA guide (spacer) seed region and the repeat region proximal to the guide[65, 66]. Thus, they are localized in the center of the SpCas9 crRNA. Furthermore, RNA-DNA chimeric Cas9 crRNAs composed of DNA with only RNA at the critical 2'- OH positions were highly active in vitro[64]. To determine if a similar phenomenon existed for the 5' pseudoknot of Cas12a, we selected several RNA-DNA chimeric pseudoknot crRNAs with high activity in vitro and tested their ability to knock-out, and therefore edit, an EGFP gene. Our collaborators in the Gagnon lab started with HEK293T cells stably expressing EGFP[74] and transduced them with a 3xNLS-

AsCas12a-expressing lentivector followed by selection. crRNAs were lipid-transfected into these custom HEK293T cells constitutively expressing EGFP and AsCas12a. For cell-based editing experiments, 5' pseudoknots were ligated to a different guide sequence (hence the designation cpEGIPe) designed to target the integrated EGFP gene. Our native all-RNA control routinely generated about 80% knockout of EGFP when measured by flow cytometry 5 days after transfection (**Figure 4.8**B). However, the RNA-DNA chimeric pseudoknots generated little or no editing activity. Notably, cpEGIPe-N1 was inactive despite preserving RNA at the putative critical 2'-OH positions. Only cpEGIPe-B1a provided modest editing (approximately 10%), perhaps due to better conservation of A-form structure by additional RNA nucleotides. Thus, as had been observed with SpCas9 crRNAs, retaining 2'-OH at potentially critical positions appeared to be necessary but not sufficient to provide gene editing activity[65, 66]. These results highlight the more complex nature of editing in cells and support the complementary role of both A-form helical structure and 2'-OH contacts within the Cas12a crRNA 5' pseudoknot.

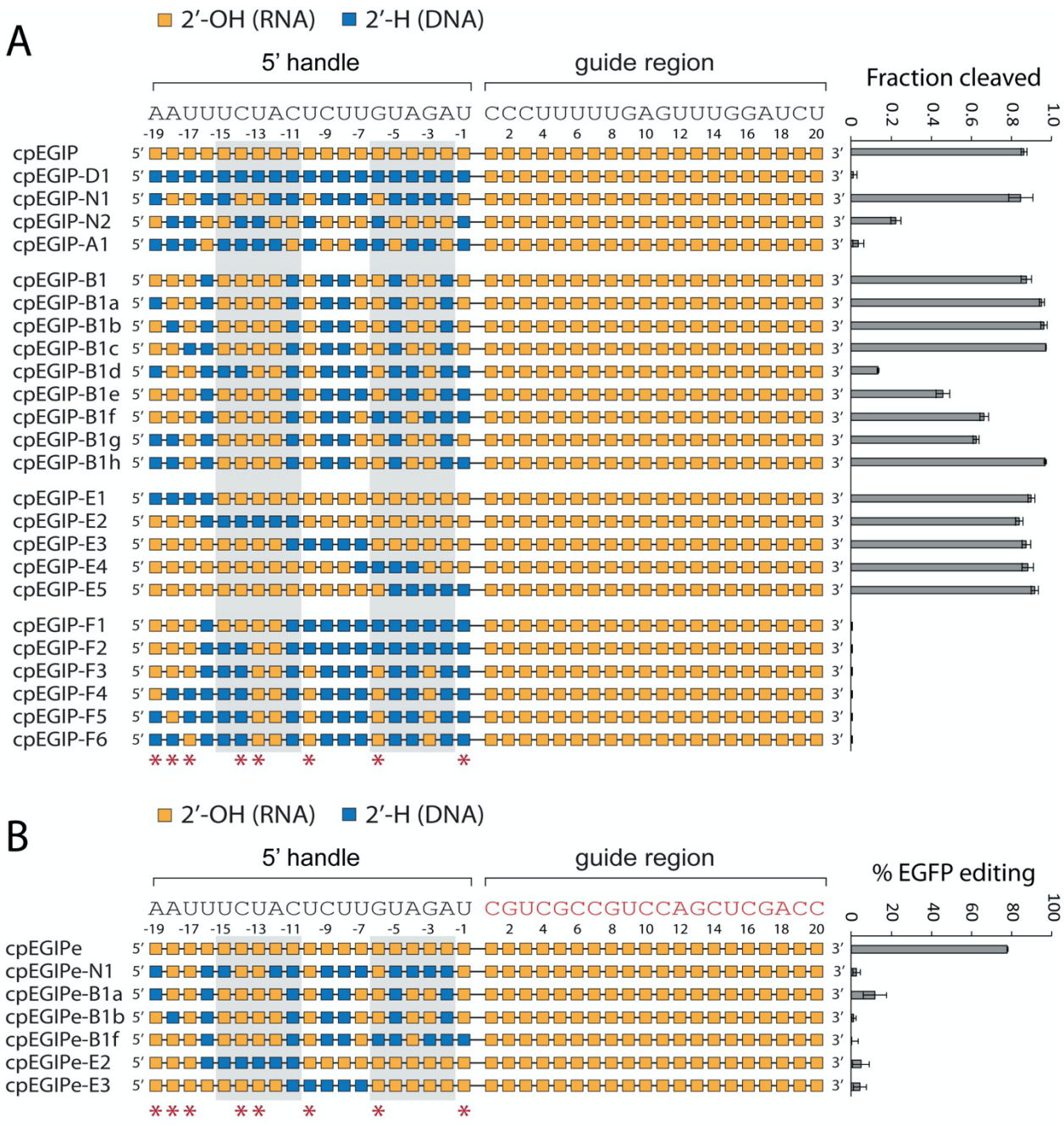


Figure 4.8. In vitro cleavage and cell-based editing activity of AsCas12a using crRNA guides modified with 2'-deoxyribose (DNA) nucleotides. *DNA modification patterns are illustrated, with corresponding in vitro cleavage activity (**A**) or cell-based editing activity (**B**) shown to the right. Error is reported as standard deviation. Nucleotides of the 5' handle predicted from structural data to make polar contacts through 2'-OH hydrogen bonding are indicated with asterisks below. The base-paired stem nucleotides of the 5' handle pseudoknot are enclosed with gray boxes. In vitro cleavage and cell-based editing assays utilize the distinct guide sequences that are shown.*

4.3.2 2'-Hydroxyl Substitutes and RNA Mimics in the 5' Pseudoknot

To further investigate the nature of A-form helical preference and ribose chemical compatibility in 5' pseudoknot structure-activity, we substituted RNA nucleotides with additional modified ribose nucleotides or sugar replacements. The first design was a complete substitution of the pseudoknot with 2'-fluoro-ribose (2'-F) nucleotides (cpEGIP-SJ01). This fully modified pseudoknot provided very high cleavage activity, similar to or higher than the native all-RNA control (**Figure 4.9**A). The robust activity of this modified pseudoknot indicated that the hydroxyl chemistry at the 2' position was completely dispensable for intrinsic enzyme activity. Importantly, 2'-F is known to stabilize C3'-endo sugar pucker and strongly favor A-form helical structure[94, 95]. However, when tested for cell-based editing activity cpEGIP-SJ01 was completely inactive (**Figure 4.9**B). 2'-F nucleotides are potentially able to accept hydrogen bonds but not donate them, making them an incomplete replacement for RNA where 2'-OH contacts may be critical. While the combination of stable structure, mimicry of RNA properties, and maintenance of sufficient 2' contacts[94] have enabled 2'-F to substitute for 2'-OH in vitro, cell-based editing is apparently more sensitive and likely requires retention of specific 2'-OH contacts.

To specifically investigate positions with predicted 2'-OH contacts, a member of our group (Dr. Sunit Jana) synthesized crRNAs with RNA or 2'-F pseudoknots containing oxepane nucleic acid (ONA) and 2'-araOH in putative critical 2'-OH positions. ONA is a 7-membered ring structure that can be synthesized with multiple hydroxyl groups and a phosphodiester linkage at different ring positions[96]. Dr. Jana synthesized thymidine (OxT) nucleotide replacements with the phosphodiester linkage at three different positions, designated OxT-1, - 2, and -3. These were then incorporated into the pseudoknot sequence at putative critical 2'-OH positions -13 and -17 during solid-phase synthesis. The incorporation of ONA (cpEGIP-SJ03, -SJ04, and - SJ05) completely abrogated AsCas12a activity, indicating that certain ONA properties, such as the bulkiness of the 7-membered ring or conformational restraints, may simply be incompatible despite providing hydroxyl groups at various positions (**Figure 4.9**A). Based on these results we chose to not proceed further with ONA.

Arabinose nucleotides (2'-araOH) are stereoisomers of RNA that place the hydroxyl group at the 2' position on the opposite face of the ribose ring. While arabinonucleosides retain a 2'-OH group, the "up" orientation steers the arabinose sugar to adopt a C2'-endo (DNA-like) conformation[97]. Dr. Jana prepared RNA strands containing 2'-araOH at -13 and -17 positions (cpEGIP-SJ06), and some enzyme activity was retained. Next, additional 2'-araOH substitutions were explored but in the background of a 2'-F pseudoknot, reasoning that the strong A-form preference of 2'-F might offset structural preferences of 2'-araOH. Incorporating more than two 2'-araOH (cpEGIP-SJ09, -SJ12, and -SJ13) severely reduced or eliminated enzyme activity (**Figure 4.9**A). Only cpEGIP-SJ10, with 2'-araOH at positions -13 and -17, supported high in vitro AsCas12a cleavage activity. These results suggest that 2'-F may be successfully combined with other modifications that would otherwise be too detrimental to support high activity. Despite the promising in vitro activity of cpEGIP-SJ10, it provided no editing in cells (**Figure 4.9**B).

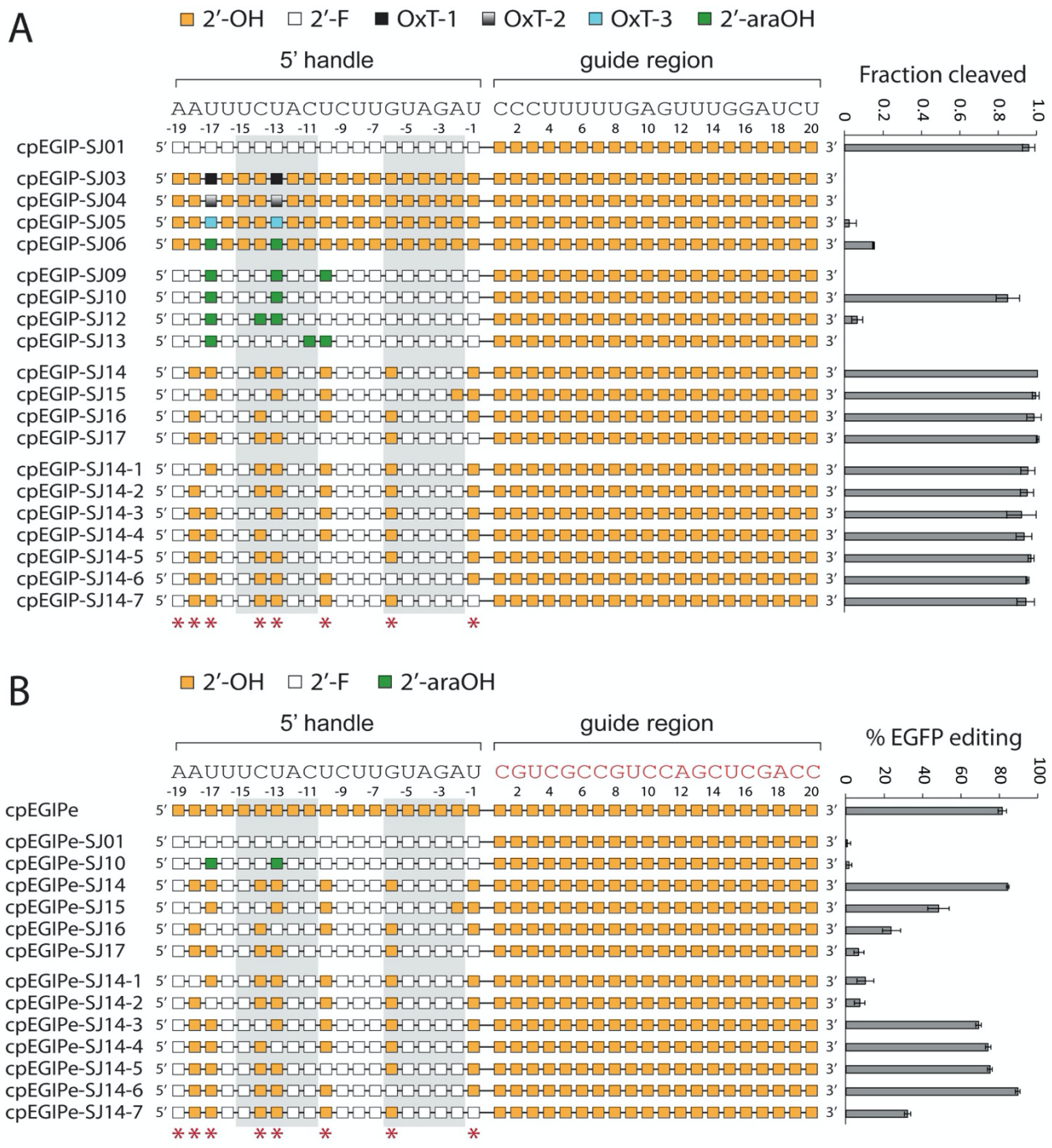


Figure 4.9. In vitro cleavage and cell-based editing activity of AsCas12a using crRNA guides modified with 2'-F, oxepane nucleotides, and 2'-araOH. *Chemical modification patterns are illustrated with corresponding in vitro cleavage activity (**A**) or cell-based editing activity (**B**) shown to the right. Error is reported as standard deviation. Nucleotides of the 5' handle predicted from structural data to make polar contacts through 2'-OH hydrogen bonding are indicated with asterisks below. The base-paired stem nucleotides of the 5' handle pseudoknot are enclosed with gray boxes. In vitro cleavage and cell-based editing assays utilize the distinct guide sequences that are shown.*

4.3.3 2'-Fluoro, 2'-Hydroxyl, and 2'-Amino Combinations in the 5' Pseudoknot

The high in vitro activity of 2'-F and its ability to counteract detrimental effects of 2'-araOH incorporation prompted us to create pseudoknots with a combination of 2'-F and RNA nucleotides that might improve cell-based editing. We made an all-2'-F pseudoknot with RNA nucleotides at putative critical 2'-OH positions (except -19). This design, cpEGIP-SJ14, supported very high *in vitro* cleavage (*cis*) activity (**Figure 4.9**A). Importantly, when we transfected this same design (cpEGIPe-SJ14) into HEK293T cells stably expressing EGFP and AsCas12a, we observed robust editing of EGFP at levels as high or greater than our native RNA control cpEGIPe (**Figure 4.9**B). Other designs that further reduced the number of RNA nucleotides to 5 and incorporated 2'-F at putative critical positions (cpEGIP-SJ15, -16, an -17) did not have any effect on in vitro cleavage. However, the analogous designs targeting cellular EGFP showed varying degrees of reduced editing efficiency **Figure 4.9**). This result supports the critical role of the putative 2'-OH positions and the value of modifications that retain strong A-form-like helical structure for high editing efficiency.

Different degrees of editing among RNA and 2'-F combinations suggested that some 2'-OH interactions were more critical than others, but clear patterns were difficult to discern. Therefore, we created a new series of cpEGIP-SJ14 designs where each of the seven remaining RNA nucleotides were individually substituted with a 2'-F residue. While all designs were highly active in vitro (**Figure 4.9**A), clearer trends were observed for cell-based editing. Placing 2'-F at -18 and -17 in the SJ14 design (cpEGIPe-SJ14-1 and -SJ14-2) was very detrimental to editing, reducing it to below 10% (**Figure 4.9**B). Positions -14, -13, and -10 could be individually replaced by 2'-F (cpEGIPe-SJ14-3, -SJ14-4, and -SJ14-5) with only mild reductions in editing compared to the RNA control. Surprisingly, replacement at position - 6 (cpEGIPe-SJ14-6) resulted in very robust editing of ~90%, suggesting that this position is less sensitive and can be replaced readily with 2'-F. This result reinforced our initial finding with DNA replacements that suggested position -6 does not depend heavily on 2-OH contacts or chemistry (**Figure 4.8**A). Finally, replacing position -1 with 2'-F created strong reductions in gene editing activity. Walking 2'-F through the remaining putative 2'-OH positions identified a crRNA with a pseudoknot

having only 6 RNA residues out of 19 (cpEGIPe-SJ14-6) that produced AsCas12a editing activity as high or greater than a native crRNA. It also identified higher sensitivity to 2'-F substitution at the terminal ends of the pseudoknot sequence.

The positive performance of 2'-F but generally poor performance of 2'-araOH and Ox nucleotides indicated that smaller structural perturbations of the ribose and the ability to provide hydrogen-bonding potential at the 2' position were desirable. To determine if further RNA reduction at putative critical 2'-OH positions might be possible, we identified 2'-amino (2'-NH_2) as a promising modification. The relatively small amino moiety and its ability to both donate and accept hydrogen bonds made it an attractive replacement for RNA. In addition, the pKa of the 2'-amino should be 6.2, as determined by Eckstein and colleagues[98], making the major species non-protonated at physiological pH. 2'-NH_2 nucleotides have been used for aptamer and ribozyme development[80, 99, 100], but have not been applied to CRISPR guide RNA modification previously, and overall, remains an understudied modification. Starting with an all-2'-F pseudoknot, we sought to replace the putative critical positions individually with 2'-NH_2 to ensure that in vitro activity was not affected. Because 2'-NH_2 is an uncommon modification and only pyrimidine monomers were commercially available, we did not substitute positions -18 and -6. All the crRNAs containing a 2'-NH_2 substitution in the 2'-F pseudoknot (cpEGIP-LA-N14-1 through -LA-N14-5) were highly active in vitro (**Figure 4.10**A), indicating no negative impact on intrinsic enzyme activity. However, none of these designs targeting EGFP in cells showed significant gene editing activity (**Figure 4.10**B).

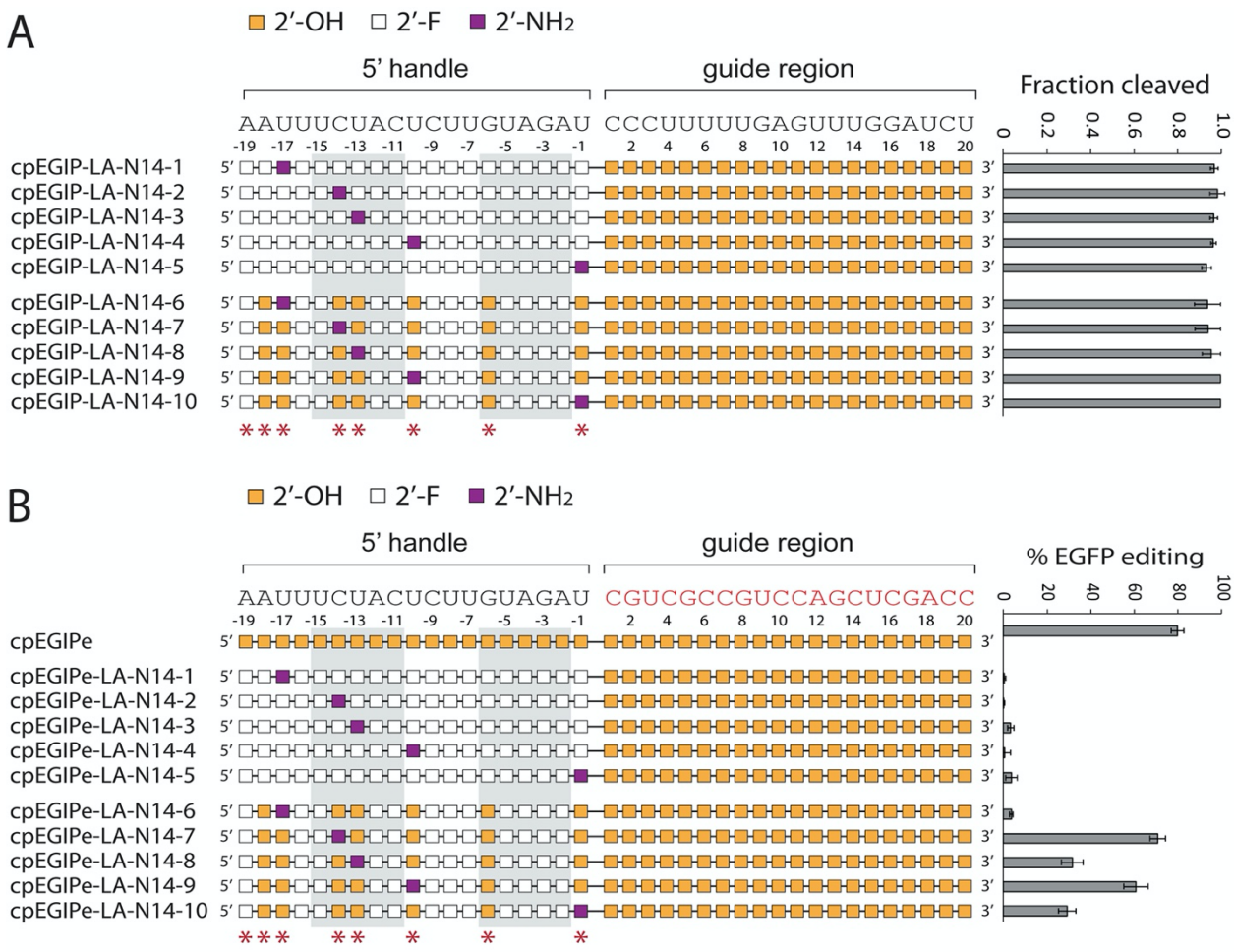


Figure 4.10. In vitro cleavage and cell-based editing activity of AsCas12a using crRNA guides modified with 2'-F and 2'-NH₂. *Chemical modification patterns are illustrated with corresponding in vitro cleavage activity (**A**) or cell-based editing activity (**B**) shown to the right. Error is reported as standard deviation. Nucleotides of the 5' handle predicted from structural data to make polar contacts through 2'-OH hydrogen bonding are indicated with asterisks below. The base-paired stem nucleotides of the 5' handle pseudoknot are enclosed with gray boxes. In vitro cleavage and cell-based editing assays utilize the distinct guide sequences that are shown.*

To begin with modified pseudoknots more likely to generate gene editing activity, we created the equivalent of cpEGIP-SJ14 (~80% editing) but walked a 2'-NH_2 through each remaining native RNA nt position (except -18 and -6 since A-2'-NH_2 and G-2'-NH_2 phosphoramidites are not commercially available). These new designs, cpEGIP-LA-N14-6 through -LA-N14-10, all exhibited robust in vitro cleavage activity (Figure **4.10**A). When targeting the stably expressed EGFP gene in cells, these designs presented varying degrees of gene editing. Substitution at position -17 with 2'-NH_2 abrogated gene editing activity (cpEGIPe-LA-N14-6). Substitution at positions -13 and -1 resulted in approximately 30% gene editing (cpEGIPe-LA-N14-8 and -LA-N14-10). In contrast, substitutions with 2' NH_2

at positions -10 and -14 (cpEGIPe-LA-N14-9 and -LA-N14-7) provided ~60-70% editing, which is substantial considering that the native RNA pseudoknot typically provides 80% editing. Interestingly, substitution at position -14 resulted in similar activities whether 2'-F (cpEGIPe-SJ14-3) or 2'-NH_2 (cpEGIPe-LA-N14-7) were used, suggesting relatively flexible accommodation at this position. 2'-NH_2 was not suitable for replacing RNA at positions -17 or -1, which was also the case with 2'-F (cpEGIPe-SJ14-2 and cpEGIPe-SJ14-7). However, position -13 was differentially affected, with 2'-F being very well-tolerated (cpEGIPe-SJ14-4) whereas 2'-NH_2 was poorly tolerated (cpEGIPe-LA-N14-8). Amines do not form hydrogen bonds as stably as alcohols and are bulkier at the 2' position than OH and F, which may explain differential activity effects.

4.4 Differential *Cis* and *Trans* Cleavage Activity

Differences in sequence-specific *cis* cleavage *in vitro* versus cell-based gene editing prompted us to investigate whether chemically modified 5' pseudoknots impacted non-sequence-specific *trans* cleavage. We employed a similar fluorescence-based assay as that originally reported by Doudna and coworkers[101]. We systematically screened all crRNAs with modified pseudoknots and plotted their maximum *trans* activity after 1 h of reaction time. We found that levels of *trans* and *cis* activity closely mirrored one another for most crRNAs (**Figure 4.11**). This would be expected since *trans* cleavage activity relies on the same active site as *cis* cleavage activity[102, 103]. However, a few crRNAs with modified 5' pseudoknots generated differential *cis* versus *trans* activity. For example, cpEGIP-B1d and cpEGIP-F6 both had little or no *cis* activity but exhibited about 50% *trans* activity (**Figure 4.11**A). After incubation for 2 h both reached similar activity levels as the all-RNA control crRNA. In contrast, cpEGIP-SJ10 and cpEGIP-LA-N14-1 both exhibited very high *cis* cleavage but low *trans* cleavage activity (~10%) (**Figure 4.11**B-C). Incubation for over 2 h resulted in no additional cleavage for cpEGIP-SJ10 and less than 40% for -LA-N14-1.

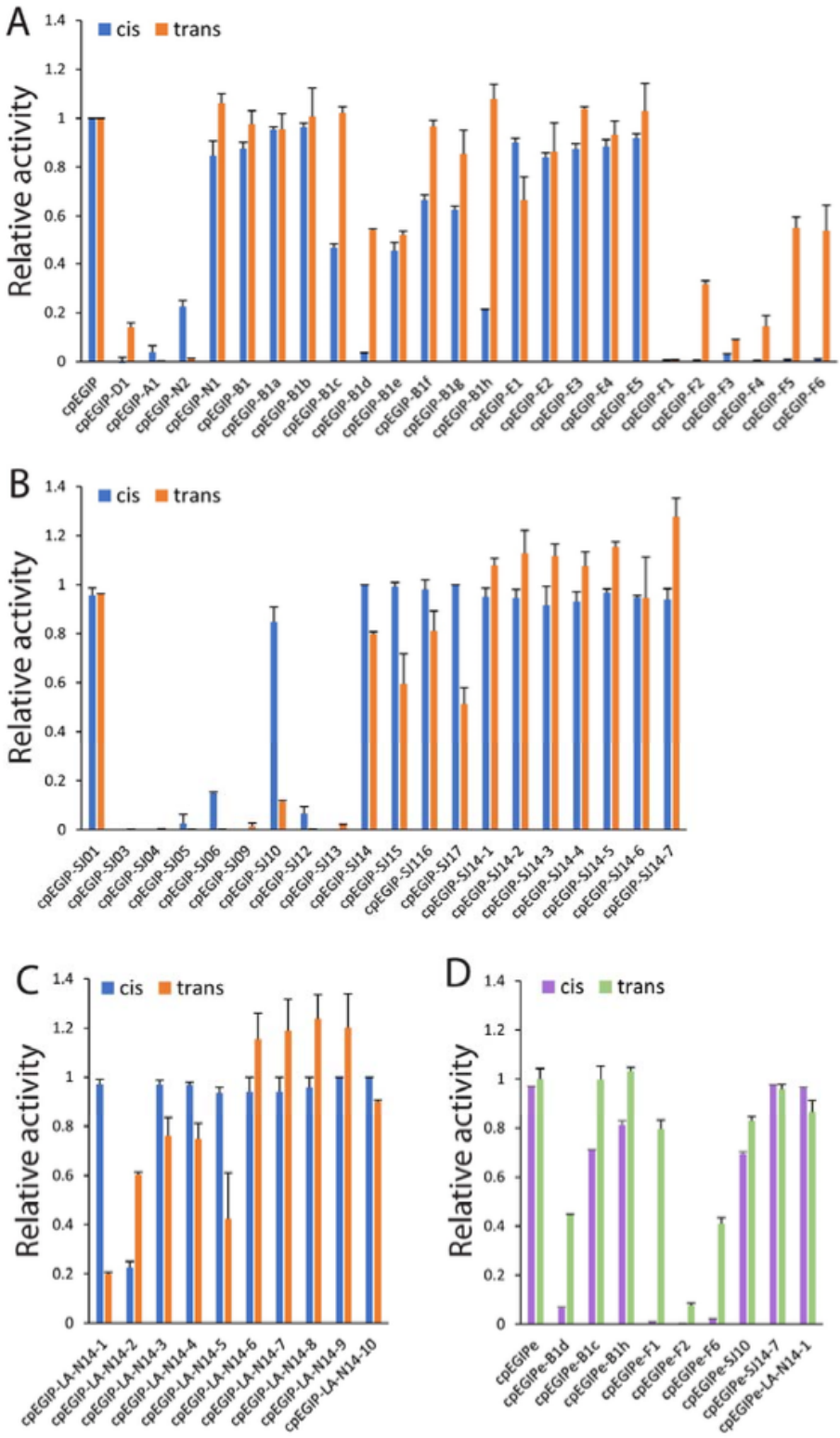


Figure 4.11. Chemical modification of the AsCas12a crRNA 5' handle modulates non-sequence- specific trans ssDNase activity. (A-D) *Comparing sequence-specific cis and non-specific trans activity of AsCas12a with crRNAs containing modified 5' handles. Error is reported as standard deviation. The guide used in Panels A-C (cpEGIP) is*

designed to target an EGFP-containing plasmid for in vitro cis cleavage while the guide used in Panel D (cpEGIPe) is designed to target a chromosomally-integrated EGFP gene for cell-based editing.

To determine whether differential *cis-trans* activity was guide sequence-dependent, we ligated the modified 5' pseudoknot to the cpEGIPe guide sequence used for editing in cells. With this guide sequence, both cpEGIPe-B1d and cpEGIPe-F6 still produced markedly low *cis* but about 50% *trans* activity (**Figure 4.11**D), which reached full *trans* activity after an additional 60 min of incubation. In addition, cpEGIPe-F1 still produced almost no *cis* activity but now generated very high *trans* activity. Two modified crRNAs that had previously shown high *cis* but low *trans* activity, cpEGIPe-SJ10 and cpEGIPe-LA-N14-1, now exhibited similar *cis-trans* activity profiles. Chemical modifications may enable conformational changes to Cas12a that allow *trans* cleavage without target binding. However, *trans* cleavage assays lacking target DNA for several crRNAs did not produce any significant *trans* activity . An alternative explanation for *trans* activity with little or no *cis* activity could potentially be "nicking" of one of the target DNA duplex strands. This could conceivably activate Cas12a *trans* activity but not be readily observed in typical *cis* cleavage assays, which use non-denaturing gel electrophoresis. Therefore, we individually 5' radiolabeled non-target (sense) and target (antisense) strands and performed *trans* cleavage assays on duplexes using cpEGIPe and cpEGIPe-B1d, -F1, -F6, and -SJ14-7 crRNAs, followed by resolution of reaction products on denaturing polyacrylamide gels. The cleavage of individual strands very closely matched that of our previous *cis* assays, indicating that differential strand nicking was not a likely explanation for high *trans* but low *cis* activity. To further investigate the mechanism of apparent differential *cis-trans* activity, we generated hybrid dsDNA targets that incorporated a partially phosphorothioate (PS) modified non-target (sense) or target (antisense) strand[103]. The PS modified target strand was observed to consistently reduce *trans* activity by 60% for both native and chemically modified pseudoknot-containing crRNAs while a PS-modified non-target strand had little or no effect. These results are consistent with *trans* activity first requiring *cis* cleavage of the target strand[103]. Thus, we conclude that high catalytic turnover of *trans* activity[101] versus the low turnover of *cis* activity is the most likely explanation for apparent differential *cis-trans* activity. The modified crRNAs that displayed differential *cis-trans* activity likely fell into an activity window or regime that made this effect easily observable.

The crRNA possessing the maximum number of modifications and exhibiting little or no loss in editing was cpEGIPe-SJ14-6. However, this modification scheme retained 6 RNA residues, which may be subject to RNase attack in therapeutic contexts. Thus, to further stabilize these final positions, we replaced the phosphodiester linkage 3' to each remaining RNA residue with phosphorothioate (PS) linkages. This newly modified pseudoknot was ligated to the cpEGIPe guide possessing two terminal 3' PS linkages and 2'-O-methyl nucleotides to further stabilize against 3' exonucleases. This new modified crRNA, cpEGIPe-SJ14-6PSfull, exhibited full *cis*, *trans* and editing activities (**Figure 4.12)**. Thus, *As*Cas12a crRNAs can be generated with heavily modified pseudoknots that should protect every nucleotide from nucleases and retain robust editing activity. To test the nuclease resistance of several pseudoknots and ligated crRNAs, the Gagnon lab also performed serum stability assays and found that native crRNA (cpEGIPe) and an RNA-DNA chimeric pseudoknot (cpEGIPe-F6) were completely digested under the conditions tested. However, when testing some modified crRNA, their results then showed that select 2'-F and PS modification of the 5' pseudoknot can stabilize against common serum nucleases.

To determine whether pseudoknot modifications may be impacting enzyme activity through large conformational effects, we performed limited trypsin hydrolysis on *As*Cas12a RNP complexes. *As*Cas12a by itself, assembled as a native RNP with cpEGIPe, or further bound to its cognate dsDNA target revealed unique limited trypsin digestion patterns. However, no difference in the trypsin digestion pattern was observed when comparing native RNP complexes with RNPs assembled from heavily modified pseudoknot guides (cpEGIPe-SJ14-6 and cpEGIPe-SJ14-6PSfull) previously shown to possess robust *in vitro* cleavage and editing activity (**Figure 4.9** and **Figure 4.12**). Further addition of a cognate dsDNA target to modified RNP complexes also showed no differences when compared to the native RNP complex with bound target. Thus, crRNAs with modified pseudoknots that confer high editing activity do not perturb gross *As*Cas12a conformation.

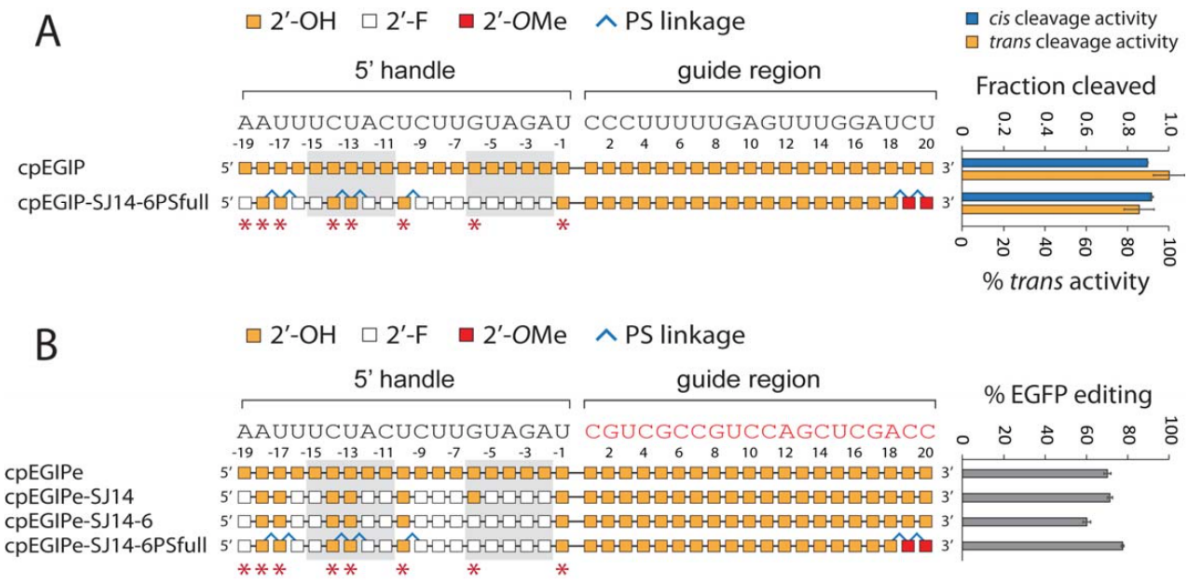


Figure 4.12. In vitro cis and trans cleavage activity and cell-based editing of AsCas12a using heavily modified pseudoknot crRNA guides. *Chemical modification patterns are illustrated with corresponding in vitro activity* ***(A)*** *or cell-based editing* ***(B)*** *shown to the right. Error is reported as standard deviation. Nucleotides of the 5' handle predicted from structural data to make polar contacts through 2'-OH hydrogen bonding are indicated with asterisks below. The base-paired stem nucleotides of the 5' handle pseudoknot are enclosed with light gray boxes. In vitro cleavage and cell-based editing assays utilize the distinct guide sequences that are shown.*

4.5 Discussion

Chemical modification of RNA is a method to probe the role of functional groups in RNA and RNP structure and biochemical activity[66, 104]. It is also a critical step in converting RNA into drug-like molecules[60]. In this study, we have used chemical modification to investigate the necessity of the ribose 2'-OH group in the AsCas12a crRNA 5' pseudoknot structure for enzyme activity. The primary question was whether the pseudoknot structure was too complex or sensitive for efficient chemical modification, which previous studies had suggested[69, 76]. The rationale for our approach is the high likelihood that therapeutic applications of Cas12a that require direct delivery of the crRNA will also require heavy or complete chemical modification, as is the case for siRNAs. In particular, the lability of RNA is directly linked to the presence of a 2'-OH that can help catalyze intramolecular phosphodiester bond self-cleavage or attract ribonucleases[105]. The 5' pseudoknot anchors crRNA binding to Cas12a and is highly conserved across the CRISPR-Cas12a family[17, 40, 106]. The conservation of pseudoknot sequence and structure suggests that the 5' pseudoknot architecture is fine-tuned for Cas12a function and cannot

be easily perturbed[17, 106]. However, if successful chemical modification schemes can be identified, they should translate to other Cas12a proteins if RNA-protein contacts are relatively well-preserved, which appears to be the case from recent structural investigations[22, 40, 107].

Here we focused primarily on 2'-deoxyribose, or DNA, and 2'-F as substitutions but also included modifications designed to potentially retain hydrogen bonding between pseudoknot nucleotides or with AsCas12a protein. We have demonstrated that as many as 13 out of 19 nucleotides can be converted to 2'-F with no impact on efficiency of EGFP editing in human cells. Attempting to systematically modify remaining putative critical 2'-OH positions with 2'-F, Ox, 2'-araOH, and 2'-NH_2 resulted in moderate success. However, it was clear that a few positions are resistant to modification and seem to rely heavily on the 2'-OH group. For example, our results suggested that the more centrally positioned critical 2'-OH residues, especially at residues -14, -10, and -6, were more tolerant to substitution while terminal residues (-18, -17, and -1) were less tolerant. Although ONA and 2'-araOH were not sufficient substitutes for ribose, it appears that a few 2'-NH_2 and 2'-deoxy could be incorporated in the context of 2'-F. Positions most amenable to these modifications are position -19 and more central positions like -6. In summary, removal of all but a handful of RNA nucleotides is achievable in the 5' pseudoknot structure of AsCas12a crRNAs while still maintaining very high gene editing activity in cells. In addition, selective placement of PS linkages 3' to remaining RNA nucleotides further stabilized the RNA while not compromising editing activity. A similar approach succeeded in stabilizing remaining RNA positions in a Cas9 guide RNA, although editing in the final heavily modified RNA guide was significantly reduced [65]. Previously, full PS modification of the native 5' pseudoknot of the *As*Cas12a crRNA resulted in a substantial loss of editing in cells [76].

Previous attempts to modify the 5' pseudoknot sequence replaced entire blocks of contiguous nucleotides, several at a time, with 2'-F or 2'-O-methyl[69]. This approach is typically suitable for oligonucleotides that function almost exclusively through Watson-Crick base-pairing, such as antisense oligonucleotides (ASOs) or small interfering RNAs (siRNAs)[60, 108]. However, it resulted in no editing activity, likely due to loss of critical 2'-

OH contacts. The lability of remaining 2'-OH positions might be best stabilized by phosphodiester backbone modifications, such as phosphorothioate (PS) linkages[65]. Indeed, complete conversion of the 5' pseudoknot to PS linkages was reportedly tolerated in a previous study, providing up to 70% of unmodified crRNA activity [76]. Overall, these results reinforce the need to incorporate molecular and structural data into modification design and the value of nucleotide analogs that can mimic or substitute the properties of native ribose.

Ultimately, complete ribose chemical modification of the entire crRNA would be desirable for precise control over therapeutic development. Attempts to modify the guide region of Cas12a crRNAs explored 2'-O-methyl, 2'-F, DNA, locked nucleic acid (LNA), unlocked nucleic acid (UNA), and PS linkages[69, 76]. The general conclusion from these studies was that modifications at or near the 3' end of the guide region were relatively well-tolerated for most modifications. However, modification at the 5' end of the guide, and therefore within or near the 'seed' region, were very poorly tolerated. A notable feature of the seed region of both Cas12a and Cas9 CRISPR systems is the concentrated presence of conserved 2'-OH contacts with the Cas protein[14, 22]. In addition, the ability to recognize and bind DNA targets, as well as catalyze cleavage, has been shown to be dependent on the formation of an A-form-like duplex between the guide and target DNA[64, 66]. Together, the 2'-OH and A-form-like architectural requirements of crRNAs reflect on the unique co-evolution of these bacterial systems to use RNA as guides[12, 109, 110]. Complete chemical modification of guide RNAs, such as for therapeutic CRISPR-based editing, will require addressing the 2'-OH bottleneck that has finally become apparent for Cas 9 and Cas12a, and probably for other Cas enzymes. Nonetheless, our results combined with previous studies now provide a roadmap to near complete chemical modification across the entire Cas12a crRNA. We propose that a primarily 2'-F and 2'-O-methyl oligonucleotide, with some DNA or 2'-NH2 nucleotides, combined with selective PS modification, may be sufficient to provide drug-like profiles with favorable editing activity. Exploring a greater diversity of modifications at well-placed positions may further unlock efficient editing with little or no native RNA residues.

A recurring theme in crRNA chemical modification studies, especially with Cas9, is that heavy modification schemes successful for in vitro cleavage often do not translate into high editing efficiency in cells[59, 65, 66, 75]. For example, all RNA nucleotides can be replaced with 2'-F in the Cas9 crRNA[66], and now the Cas12a pseudoknot, and provide very high cleavage in vitro. However, these crRNAs provide little or no editing in cells. Converting predicted 2'-OH polar contact positions back to RNA usually rescues gene editing[65, 66]. Thus, it appears that intrinsic biochemical activity is not dependent on 2'-OH contacts in bulk reactions in vitro but these become critical within cells, perhaps due to low biochemical RNP concentrations, more complex chromosomal DNA targets, or other unexplored biochemical or cellular factors. Experiments like precision biochemical, thermodynamic, and structural studies, as well as sophisticated single-molecule or gene editing assays in cells, may be necessary to fully understand this phenomenon and unlock 'RNA- free' CRISPR-Cas editing for advanced therapeutic control.

The discovery of non-sequence-specific ssDNase cleavage, or *trans* activity, by Cas12a was quickly converted into an amplification-based diagnostic method[101]. After sequence-specific binding and cutting (*cis* cleavage) of a DNA target, the catalytic domain of Cas12a is proposed to become solvent exposed such that ssDNA is accessible to the active site[102, 103]. This results in high turnover degradation of ssDNA. Once DNA target paired to the guide RNA is released, ssDNase cleavage activity should cease as Cas12a returns to a pre-catalytic structural state[102]. Thus, *trans* cleavage activity of Cas12a appears to be intrinsically tied to its *cis* cleavage activity, turnover of *cis* cleavage products, and conformational state transitions.

In this study, we initially found what appeared to be differences in *cis* versus *trans* cleavage activity for a few crRNAs possessing chemically modified pseudoknots. Certain modification patterns or schemes appear to have contributed to this effect. Ultimately, differential *cis-trans* activity may be a consequence of the ratio of *cis* to *trans* activity within a certain activity regime established by the modified guide. However, *trans* activity by modified crRNAs still required cleavage of the target antisense DNA strand in a dsDNA target. The pseudoknot is proposed to influence catalysis by structural modulation of AsCas12a conformational transitions, which may explain how various activity profiles

were achieved by modified pseudoknots. If differential *cis*-*trans* cleavage activity could be controlled predictably, it might prove useful for some applications. For example, binding to a specific on-target DNA without cleaving (*cis* activity) but still allowing *trans* activity, could provide *trans* activity for long durations and possibly improve diagnostics. Conversely, *cis* cleavage without activation of *trans* cleavage might improve the specificity or safety of other applications, like gene editing. Recently, it was reported that extending the 5' or 3' ends of the LbCas12a crRNA could increase on-target activity and especially enhance *trans* cleavage rates[111]. Thus, multiple types of crRNA modification may be useful for modulating *cis* specificity and *trans* activity turnover.

4.6 Materials and Methods

4.6.1 RNA Synthesis

DNA oligonucleotides, crRNAs, and DNA-RNA chimeric 5′-handle RNAs were synthesized and purified by Integrated DNA Technologies (IDT). Full-length crRNAs, including DNA-RNA and chemically modified crRNA, were ligated using splint ligation. Chemically modified 5'-handles were custom synthesized and mass confirmed by mass spectrometry. Single-stranded fluorophore-quencher DNA substrate was synthesized commercially by IDT.

Oligonucleotide chemical syntheses were carried out using either an ABI 3400 DNA synthesizer (Applied Biosystems) or a MerMade 12 Oligonucleotide synthesizer (BioAutomation) on a Unylinker CPG (ChemGenes) solid support at a 1 μmol scale. Conventional 2′-tert-butyl-dimethylsilyl (TBDMS) ribonucleoside, 2′-fluoro-ribonucleoside (2′-FRNA), and 2′-trifluoroacetyl amino phosphoramidites (ChemGenes) were used, along with newly-synthesized oxepane phosphoramidites. Phosphoramidites were dissolved in MeCN (0.15 M) and activated with 5-ethylthio-1H-tetrazole (0.25 M in MeCN). Capping of failed couplings was carried out by the simultaneous delivery of acetic anhydride in pyridine/THF and N-methylimidazole (16 % in THF) and contacting the solid support for 6 seconds. Oxidation of the phosphite triester intermediates was affected with

0.1 M iodine in pyridine/H_2O/THF (20 seconds). A solution of 3% trichloroacetic acid in THF, delivered over 95 sec, was used to deprotect DMTr groups.

Deprotection and cleavage of oligonucleotides from the solid support was achieved by treatment with 1 mL of cold 29% aqueous ammonia/ethanol (3:1, v/v) for 16 h at 65˚C (or 48 h at room temperature for oxepane-containing oligonucleotides). The samples were centrifuged, and the supernatant was transferred to a clean 1.5 mL microcentrifuge tube and vented for 30 min, chilled on dry ice, and evaporated to dryness. Removal of the 2′-TBDMS protecting groups for the RNA-containing oligonucleotides was achieved by treatment with a 300 µL solution of NMP/Et_3N/TREAT-HF (3:4:6, v/v) for 90 min at 65˚C (samples in the SJ series were desilylated by exposure to neat TREAT-HF for 48 h), followed by quenching with 3M NaOAc buffer (50 µL; pH 5.5) and precipitation of the crude oligonucleotide from cold butanol (1 mL, -20˚C). Samples were chilled on dry ice for 30 min and then centrifuged. After removing the supernatant, the remaining pellet was evaporated to dryness, taken up in autoclaved milliQ water (1 mL), and filtered.

Crude oligonucleotides were purified by ion exchange (IE) HPLC. cpEGIP-LA oligonucleotides were purified on a Waters 1525 instrument using a Protein-Pak DEAE 5PW anion exchange column (21.5 mm × 150 mm). A buffer system consisting of buffer A (10 mM NaOAc, 20% MeCN in milli-Q water) and buffer B (0.5 M $LiClO_4$, 10mM NaOAc, 20% MeCN in milli-Q water) was used for analysis and purification. Using a gradient of 0 – 50 % $LiClO_4$ (buffer B) over 50 min (10 mL/min, 60˚C), the desired compounds eluted at around 30 min. cpEGIP-SJ oligonucleotides were purified on an Agilent 1200 Series Instrument using a Protein-Pak DEAE 5PW column (7.5 × 75 mm) at a flow rate of 1 ml/min, with a buffer system consisting of buffer A (milli-Q water) and buffer B (1 M $LiClO_4$ in milli-Q water) using a method of 0−60% buffer B over 37 min at 60˚C. Following collection of the desired peaks, excess $LiClO_4$ salts were removed using Gel Pak 2.5 size exclusion columns (Glen Research). Purified oligonucleotides were characterized by electrospray ionization-mass spectrometry (Table 4.1**,** Table 4.2) and quantitated by UV spectroscopy. Extinction coefficients were determined using the IDT OligoAnalyzer tool. Extinction coefficients for RNA were used for oligonucleotides containing oxepane modifications.

Table 4.2. Cas12a crRNA series synthesized with either 2'-F or 2'-NH_2 inserts at predicted 2'-hydroxyl contacts, in an otherwise 2'-F background.

Sample	Sequence	M_{exp}	M_{obs}
LA-F14-1	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.7851	6008.8672
LA-F14-2	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.8016	6008.8672
LA-F14-3	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.7859	6008.8672
LA-F14-4	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.8016	6008.8672
LA-F14-5	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.8016	6008.8672
LA-F14-6	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.7860	6008.8672
LA-F14-7	5' – AAU UUC UAC UCU UGU AGA U – 3'	6008.8016	6008.8672
LA-N14-1	5' – AAU UUC UAC UCU UGU AGA U – 3'	6017.8295	6017.7734
LA-N14-2	5' – AAU UUC UAC UCU UGU AGA U – 3'	6017.8486	6014.6875
LA-N14-3	5' – AAU UUC UAC UCU UGU AGA U – 3'	6017.8295	6017.7734
LA-N14-4	5' – AAU UUC UAC UCU UGU AGA U – 3'	6017.8295	6017.7734
LA-N14-5	5' – AAU UUC UAC UCU UGU AGA U – 3'	6017.8295	6017.7734
LA-N14-6	5' – AAU UUC UAC UCU UGU AGA U – 3'	6005.8047	6005.6875
LA-N14-7	5' – AAU UUC UAC UCU UGU AGA U – 3'	6005.8081	6005.6875
LA-N14-8	5' – AAU UUC UAC UCU UGU AGA U – 3'	6005.8047	6005.6875
LA-N14-9	5' – AAU UUC UAC UCU UGU AGA U – 3'	6005.8047	6005.6875
LA-N14-10	5' – AAU UUC UAC UCU UGU AGA U – 3'	6005.8047	6005.6875

N = 2'-OH (RNA) N = 2'-F N = 2'NH_2 **N** = OH contact position **N** = base pairing region

4.6.2 Splint Ligation

crRNA guide sequence (200 pmol), synthesized with a 5' phosphate, was ligated to 200 pmol of 5'-handle RNA (synthesized with a 3'-OH) using T4 DNA ligase[89] (**Table S1**) For DNA ligase, 220 pmol splint DNA, complementary to both guide and 5'-handle RNA, was annealed to 200 pmol of each crRNA portion in RNA resuspension buffer (5 mM Tris, pH 7.4, 0.5 mM EDTA) at 65°C, then slow-cooled to room temperature. Ligation was performed using 1 µL of concentrated DNA ligase (30 unit/µL) (Thermo Scientific, EL0013), 1x ligation buffer (400 mM Tris-HCL, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP), and 0.5 µL of SUPERase-In (Invitrogen) in a final volume of 100 µL. The reaction was incubated for 90 min at 37°C and stopped by adding 1 µL of 0.5 M EDTA pH 8 and phenol-chloroform extracted. Ligation products were visualized by resolving on 15% denaturing polyacrylamide gels and staining with methylene blue. Full-length ligation products were gel-purified by crush-and-soak elution, phenol-chloroform extracted,

ethanol precipitated, and quantified by measuring absorbance at 260 nm and calculated extinction coefficients using nearest neighbor approximations with Beer's Law. Electrospray ionization-mass spectrometry was used to confirm the mass of a few full length, ligated crRNAs to ensure ligation and gel-purification was proceeding properly.

4.6.3 *As*Cas12a Enzyme Expression and Purification

Plasmid encoding of an *As*Cas12a was obtained from Addgene (79007). *As*Cas12a proteins were prepared similarly to that previously described for *Sp*Cas9[18]. Briefly, protein expression was induced in Rosetta (DE3) cells grown in Luria-Bertani (LB) broth with 0.2 mM isopropyl thiogalactopyranoside (IPTG) at 18°C for 16 h. Cell pellets were resuspended in 6 mL chilled binding buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM PMSF, 5 mM imidazole) per 0.5 L of culture. Resuspended cells were sonicated and clarified by centrifugation. For the 0.5 L of culture, 5 mL His-Pur Cobalt-CMA resin (Thermo Scientific) was equilibrated with binding buffer and the supernatant added to the equilibrated resin and incubated at 4°C for 1 h with inversion to mix every 15 min. The column was washed sequentially with at least 10 bed volumes of increasing concentrations of NaCl in wash buffer (20 mM Tris-HCl, pH 8.0, 10 mM imidazole, 0.25/0.5/0.75/1.0 M NaCl). Protein was eluted with 130 mM imidazole buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 200 mM imidazole). Purified *As*Cas12a enzyme was concentrated and buffer exchanged using centrifugal concentrators (Sartorius, 30,000 MWCO) into 2x storage buffer (40 mM Tris, pH 7.5, 300 mM KCl, 1 mM EDTA, and 2 mM DTT). Then one volume of glycerol added to obtain a final of 50% glycerol. Protein stocks were then stored at -80°C. Concentration of *As*Cas12a was determined by UV absorbance at 280 nm using a calculated extinction coefficient and Beer's law.

4.6.4 *In Vitro As*Cas12a *Cis* Cleavage Activity Assays

A 1 kb fragment of the target EGFP gene was PCR-amplified using "cpEGIP_vitro targ" primers from plasmid DNA (Addgene, 26777) and purified by phenol-chloroform extracted and ethanol precipitation. Target DNA (185 ng) was spotted into tubes and combined with the *As*Cas12a (0.75 μM final) and crRNA (0.3 μM final) in 1x cleavage

buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT, 0.5 mM EDTA, 2 mM $MgCl_2$) supplemented with 0.1 mg/mL of purified yeast tRNA. Molar excess of the *As*Cas12a over crRNA allows RNP concentration approximation by using crRNA concentration. Although a fraction of the Cas enzyme remains unbound, we have previously found this approach to be an accurate method for predicting the concentration of actual assembled RNP complexes[64, 66]. The 40 µL reaction was incubated at 37°C for 2 h. The reaction was then treated with 10 µg of RNase A (Thermo Scientific) for 15 min followed by 20 µg of Proteinase K (Thermo Scientific) for 15 min at room temperature. The reaction products were then precipitated in 10 volumes of 2% $LiClO_4$ in acetone for >1 h at -20°C. Precipitated reactions were centrifuged and washed with acetone, air dried, and resuspended in gel-loading dye (10% glycerol, 1x TBE, orange G dye) and resolved on 1.5% agarose gels. Agarose gels were stained with ethidium bromide and visualized using a UV imager. The fractions of target cleaved versus uncleaved were quantified using ImageJ software. Error bars for all quantified data represent experimental replicates, not technical replicates.

For cleavage time-course assays, *As*Cas12a and crRNA (cpEGIP or cpEGIP-SJ14) were assembled at room temperature for 10 min. The RNP was then added to a target DNA in at 37°C in a final of 1x cleavage buffer and reactions stopped at indicated time points by precipitation with 2% $LiClO_4$ in acetone and incubated at -20°C. Samples were pelleted by centrifugation then washed with acetone. After samples were air dried, they were resuspended in 1x loading dye (10% glycerol, 1x TBE, orange G dye), treated with RNase A and proteinase K as described above, and cleavage products resolved on 1.5% agarose gels. Quantification was performed as described above.

4.6.5 Generation of HEK293T Cells Stably Expressing EGFP and *As*Cas12a

HEK293T cells stably expressing EGFP were a kind of gift from Dr. Wen Xue[59]. We subcloned *As*Cas12a from pY010 (pcDNA3.1-hAsCpf1) (Addgene, 69982) into pLJM1-EGFP lentivector[112] (Addgene, 19319) such that the EGFP gene was replaced with *As*Cas12a. Subcloning was performed with *Nhe*I and *Eco*RI restriction enzymes and resulted in retention of the C-terminal NLS and HA tag on *As*Cpf1 in the new vector,

pLJM1-*As*Cas12a. The PGK-driven puromycin resistance gene was replaced with a hygromycin resistance gene downstream of an internal ribosomal entry site. HEK293T cells were transfected with pLJM1-*As*Cas12a, pCMV-VSVG (Addgene, 8454) and pCMV-dR8.2 dvpr (Addgene, 8455)[113] using the calcium phosphate technique. Medium was replaced 18 h after transfection. Lentiviral containing medium was collected 48 h later and centrifuged at 1000 × g for 5 min. Hexadimethrine bromide was added to the medium (8 μg/mL) and the medium was used to infect HEK293T-EGFP cells on 6 well plates for 8 h.

Transfection of cells with pLJM1-*As*Cas12a and staining with anti-HA (Santa Cruz Biotechnology, 1:1000) revealed primarily cytoplasmic localization. Therefore, we cloned an additional 3x NLS tag onto the N-terminus of *As*Cas12a to generate pLJM1-3xNLS-*As*Cas12a. Staining with anti-HA revealed exclusively nuclear staining. Stably integrated cells were selected using 100 μg/mL hygromycin and a clonal cell line was expanded. Expression and nuclear localization of *As*Cas12a was confirmed in the stable cell line.

4.6.6 Cell-Based Editing Measured by Flow Cytometry

HEK293T cells expressing EGFP and *As*Cas12a were grown in Dulbecco's modified eagle's medium (DMEM) with 1x non-essential amino acids (NEAA), 5% cosmic calf serum (CCS) and 2.5% fetal bovine serum (FBS) without antibiotics. Cells were reverse transfected (40,000 cells) in six experimental replicates in 96-well plates with 20 pmols of crRNA and 0.3 μL RNAiMAX lipid (Invitrogen) in a final reaction of 200 μL of OptiMEM. After 8 h, one volume of media containing 5% FBS and 5% CCS was added to cells and further incubated overnight. Media was then replaced with full media and cells grown for an additional 4 days. For flow cytometry, cells were washed with 200 μL PBS and trypsinized by adding 70 μL of trypsin-EDTA solution. 100 μL of media was added to the cell. The cells were spun for 5 min at 300 x g at room temperature. Cells were washed again with 200 μL PBS, resuspended in 200 μL PBS and counted in an Attune flow cytometer. EGFP was detected using the blue laser (BL1 channel). At least 20,000 events were collected and analyzed by Attune software. The cells were gated based on forward and side scattering (FSC-A/SSC-A) to remove cell debris, gated to select single cells, and gated to select EGFP positive cells. The quadrant gate was established using the signal

from non-EGFP expressing control cells. Untreated HEK293T cells expressing EGFP and *As*Cas12a contained ~5% non-fluorescent cells. The average from six replicates was used for background subtraction to determine the extent of cell-based editing after treatment. Two crRNAs with different guide sequences, cpEGIP and cpEGIPe, were initially screened. cpEGIPe provided substantial editing activity (75-80%) and was chosen for subsequent modification and editing experiments in cells.

4.6.7 Fluorophore-Quencher (FQ) *Trans* Cleavage Reporter Assay

*As*Cas12a and crRNA were preassembled into RNP complex by incubating 500 nM *As*Cas12a with 550 nM crRNA and 25 nM DNA target in 1X cleavage buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT, 0.5 mM EDTA, 2 mM $MgCl_2$) on ice for 15 min. Certain experiments either omitted target DNA, crRNA, or used phosphorothioate-modified target at the same concentrations as indicated above. Reactions were initiated by adding fluorophore-quencher (FQ) ssDNA substrate to a final of 3 μM and placing the plate in a Bio-Rad CFX96 instrument with the block set at 4°C. The block was cycled to 37°C and the plate was read using the SYBR-only channel every 53 sec (reading takes 7 sec) such that fluorescence readings were collected at 1 min intervals. Fluorescence was collected for up to 150 min. Maximum fluorescence values at 1 h were used to compare non-sequence-specific *trans* cleavage activity of varying crRNAs containing modified 5' pseudoknots. Error was calculated by standard deviation, representing three or more experimental replicates.

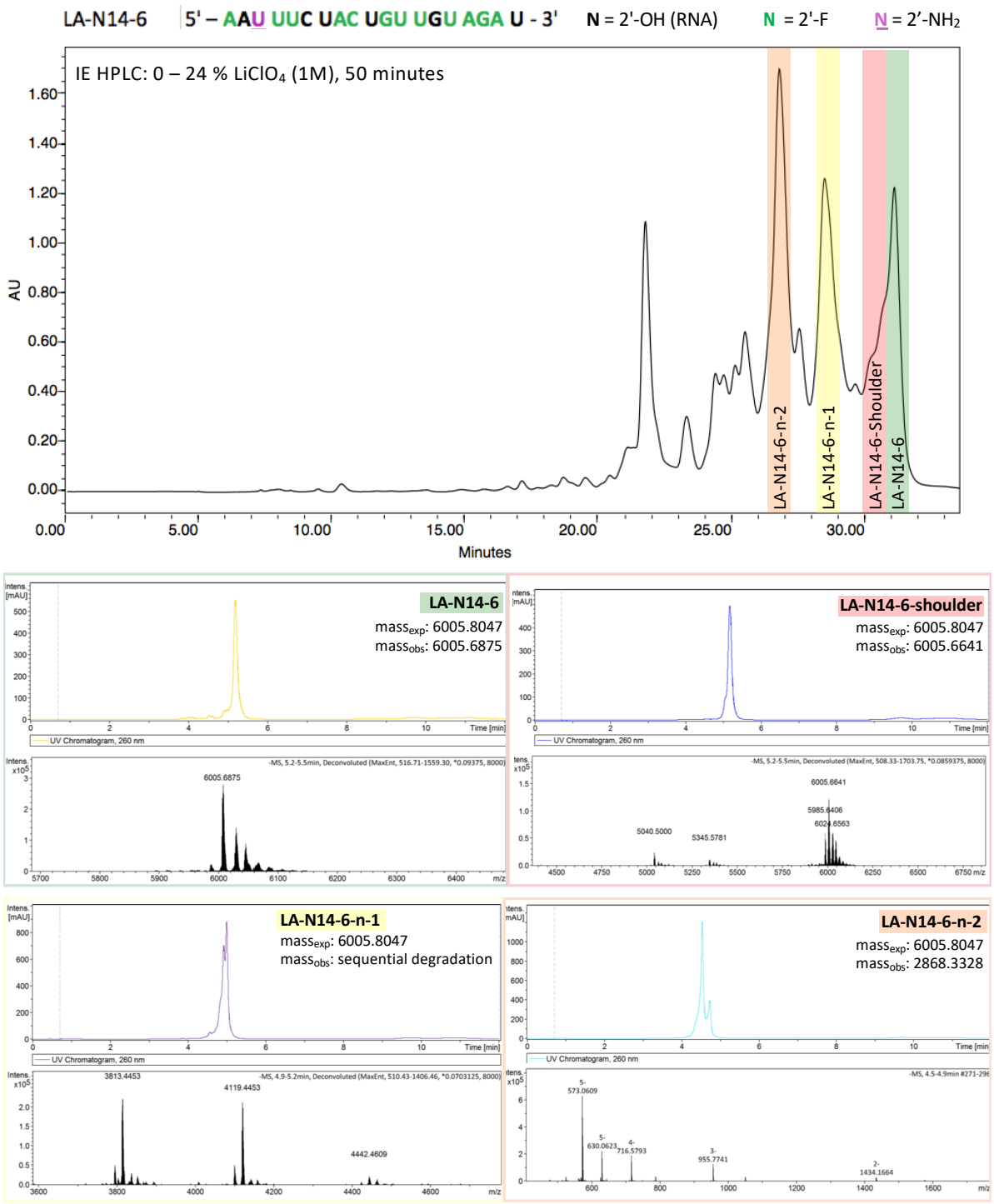
For strand-nicking cleavage assays utilizing radiolabeled target DNA strands, individual DNA strands (cpEGIPe target sequence) were 5' radiolabeled with T4 polynucleotide kinase and ^{32}P-γ-ATP as previously described[114]. Radiolabeled DNA strands were then gel-purified from a 15% denaturing Urea-PAGE by crush-and-soak elution after visualizing radioactive bands by autoradiography. Duplexes with differentially radiolabeled sense (non-target) or antisense (target) strands were annealed and slow-cooled with a ~1.2-fold molar excess of unlabeled complementary strand. Reaction conditions then utilized those of a standard *trans* cleavage assay but lacking a fluorophore-quencher labeled ssDNA substrate. Reactions proceeded for 1 h before being

stopped by phenol-chloroform extraction. Cleavage products were resolved on a 15% denaturing PAGE and exposed to phosphorimager for visualization.

4.6.8 Limited Trypsin Hydrolysis of AsCas12a Protein and RNP

Trypsin proteolysis was performed using 30 μg of *As*Cas12a in the presence and absence of crRNA and double-stranded target DNA (1:1.5:2 molar ratio)[115]. The RNP was incubated at room temperature for 5 min. The resulting samples were incubated with Trypsin-EDTA solution (0.05%, Invitrogen) at a mass ratio of 100:1 and the partial proteolysis was conducted at 37°C for 15 min. The reaction was stopped by the addition of SDS-PAGE loading buffer and heating the samples for 5 min at 95°C. The reaction products were analyzed by 12% SDS-PAGE and stained with Coomassie brilliant blue G-250 in 50% (v/v) methanol and 10% (v/v) acetic acid then destained in the same solution without dye.

4.7 Supplementary Figures



Supplementary Figure 4. 1. *Exemplary IE HPLC trace for LA-N14-6 (oligonucleotide containing one 2'-NH_2 insert. The highlighted peaks were isolated and LCMS characterizations are shown below. The desired product appears in the ultimate peak, with previous peaks containing degraded oligonucleotides.*

4.8 References

1. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007,** *315* (5819), 1709-12.
2. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D., Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading)* **2005,** *151* (Pt 8), 2551-2561.
3. Brouns, S. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J.; Snijders, A. P.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J., Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **2008,** *321* (5891), 960-4.
4. Garneau, J. E.; Dupuis, M.; Villion, M.; Romero, D. A.; Barrangou, R.; Boyaval, P.; Fremaux, C.; Horvath, P.; Magadán, A. H.; Moineau, S., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **2010,** *468* (7320), 67-71.
5. Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V., Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **2012,** *109* (39), E2579-E2586.
6. 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.
7. Mojica, F. J.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E., Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **2005,** *60* (2), 174-82.
8. Tyson, G. W.; Banfield, J. F., Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **2008,** *10* (1), 200-7.
9. Abudayyeh, O. O.; Gootenberg, J. S.; Konermann, S.; Joung, J.; Slaymaker, I. M.; Cox, D. B.; Shmakov, S.; Makarova, K. S.; Semenova, E.; Minakhin, L.; Severinov, K.; Regev, A.; Lander, E. S.; Koonin, E. V.; Zhang, F., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **2016,** *353* (6299), aaf5573.
10. 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.
11. Kim, E.; Koo, T.; Park, S. W.; Kim, D.; Kim, K.; Cho, H. Y.; Song, D. W.; Lee, K. J.; Jung, M. H.; Kim, S.; Kim, J. H.; Kim, J. H.; Kim, J. S., In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **2017,** *8*, 14500.

12. Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J.; Backofen, R.; Koonin, E. V., An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **2015,** *13* (11), 722-36.

13. Nishimasu, H.; Cong, L.; Yan, W. X.; Ran, F. A.; Zetsche, B.; Li, Y.; Kurabayashi, A.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal Structure of Staphylococcus aureus Cas9. *Cell* **2015,** *162* (5), 1113-26.

14. Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **2014,** *156* (5), 935-49.

15. Nuñez, J. K.; Kranzusch, P. J.; Noeske, J.; Wright, A. V.; Davies, C. W.; Doudna, J. A., Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **2014,** *21* (6), 528-34.

16. Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A., RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012,** *482* (7385), 331-8.

17. Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **2015,** *163* (3), 759-71.

18. Anders, C.; Jinek, M., In vitro enzymology of Cas9. *Methods Enzymol* **2014,** *546*, 1-20.

19. Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M., Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **2014,** *513* (7519), 569-73.

20. Gao, L.; Cox, D. B. T.; Yan, W. X.; Manteiga, J. C.; Schneider, M. W.; Yamano, T.; Nishimasu, H.; Nureki, O.; Crosetto, N.; Zhang, F., Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol* **2017,** *35* (8), 789-792.

21. Shah, S. A.; Erdmann, S.; Mojica, F. J.; Garrett, R. A., Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol* **2013,** *10* (5), 891-9.

22. Yamano, T.; Nishimasu, H.; Zetsche, B.; Hirano, H.; Slaymaker, I. M.; Li, Y.; Fedorova, I.; Nakane, T.; Makarova, K. S.; Koonin, E. V.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **2016,** *165* (4), 949-62.

23. Conboy, I.; Murthy, N.; Etienne, J.; Robinson, Z., Making gene editing a therapeutic reality. *F1000Res* **2018,** *7*.

24. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013,** *339* (6121), 819-23.

25. Dai, W. J.; Zhu, L. Y.; Yan, Z. Y.; Xu, Y.; Wang, Q. L.; Lu, X. J., CRISPR-Cas9 for in vivo Gene Therapy: Promise and Hurdles. *Mol Ther Nucleic Acids* **2016,** *5* (8), e349.

26. Doudna, J. A.; Charpentier, E., Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014,** *346* (6213), 1258096.

27. Hur, J. K.; Kim, K.; Been, K. W.; Baek, G.; Ye, S.; Hur, J. W.; Ryu, S.-M.; Lee, Y. S.; Kim, J.-S., Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nature biotechnology* **2016,** *34* (8), 807-808.

28. Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016,** *533* (7603), 420-424.

29. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M., RNA-guided human genome engineering via Cas9. *Science* **2013,** *339* (6121), 823-6.

30. Monteys, A. M.; Ebanks, S. A.; Keiser, M. S.; Davidson, B. L., CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. *Mol Ther* **2017,** *25* (1), 12-23.

31. Pineda, M.; Moghadam, F.; Ebrahimkhani, M. R.; Kiani, S., Engineered CRISPR Systems for Next Generation Gene Therapies. *ACS Synth Biol* **2017,** *6* (9), 1614-1626.

32. Qi, Lei S.; Larson, Matthew H.; Gilbert, Luke A.; Doudna, Jennifer A.; Weissman, Jonathan S.; Arkin, Adam P.; Lim, Wendell A., Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013,** *152* (5), 1173-1183.

33. Vojta, A.; Dobrinić, P.; Tadić, V.; Bočkor, L.; Korać, P.; Julg, B.; Klasić, M.; Zoldoš, V., Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* **2016,** *44* (12), 5615-28.

34. Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; DeGennaro, E. M.; Winblad, N.; Choudhury, S. R.; Abudayyeh, O. O.; Gootenberg, J. S.; Wu, W. Y.; Scott, D. A.; Severinov, K.; van der Oost, J.; Zhang, F., Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* **2017,** *35* (1), 31-34.

35. Kim, S.; Kim, D.; Cho, S. W.; Kim, J.; Kim, J. S., Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **2014,** *24* (6), 1012-9.

36. Kim, D.; Kim, J.; Hur, J. K.; Been, K. W.; Yoon, S. H.; Kim, J. S., Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* **2016,** *34* (8), 863-8.

37. Kleinstiver, B. P.; Tsai, S. Q.; Prew, M. S.; Nguyen, N. T.; Welch, M. M.; Lopez, J. M.; McCaw, Z. R.; Aryee, M. J.; Joung, J. K., Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* **2016,** *34* (8), 869-74.

38. O'Geen, H.; Henry, I. M.; Bhakta, M. S.; Meckler, J. F.; Segal, D. J., A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res* **2015,** *43* (6), 3389-404.

39. Wang, S.; Ren, S.; Bai, R.; Xiao, P.; Zhou, Q.; Zhou, Y.; Zhou, Z.; Niu, Y.; Ji, W.; Chen, Y., No off-target mutations in functional genome regions of a CRISPR/Cas9-generated monkey model of muscular dystrophy. *The Journal of biological chemistry* **2018,** *293* (30), 11654-11658.

40. Dong, D.; Ren, K.; Qiu, X.; Zheng, J.; Guo, M.; Guan, X.; Liu, H.; Li, N.; Zhang, B.; Yang, D.; Ma, C.; Wang, S.; Wu, D.; Ma, Y.; Fan, S.; Wang, J.; Gao, N.; Huang, Z., The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **2016,** *532* (7600), 522-6.

41. Leenay, R. T.; Maksimchuk, K. R.; Slotkowski, R. A.; Agrawal, R. N.; Gomaa, A. A.; Briner, A. E.; Barrangou, R.; Beisel, C. L., Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol Cell* **2016,** *62* (1), 137-47.

42. Beaudet, A. L.; Meng, L., Gene-targeting pharmaceuticals for single-gene disorders. *Hum Mol Genet* **2016,** *25* (R1), R18-26.

43. Fellmann, C.; Gowen, B. G.; Lin, P. C.; Doudna, J. A.; Corn, J. E., Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat Rev Drug Discov* **2017,** *16* (2), 89-100.

44. Gagnon, K. T.; Corey, D. R., Stepping toward therapeutic CRISPR. *Proceedings of the National Academy of Sciences of the United States of America* **2015,** *112* (51), 15536-15537.

45. Hynes, R. O.; Coller, B. S.; Porteus, M., Toward Responsible Human Genome Editing. *Jama* **2017,** *317* (18), 1829-1830.

46. Karimian, A.; Azizian, K.; Parsian, H.; Rafieian, S.; Shafiei-Irannejad, V.; Kheyrollah, M.; Yousefi, M.; Majidinia, M.; Yousefi, B., CRISPR/Cas9 technology as a potent molecular tool for gene therapy. *J Cell Physiol* **2019,** *234* (8), 12267-12277.

47. Shim, G.; Kim, D.; Park, G. T.; Jin, H.; Suh, S. K.; Oh, Y. K., Therapeutic gene editing: delivery and regulatory perspectives. *Acta Pharmacol Sin* **2017,** *38* (6), 738-753.

48. Shin, J. W.; Kim, K. H.; Chao, M. J.; Atwal, R. S.; Gillis, T.; MacDonald, M. E.; Gusella, J. F.; Lee, J. M., Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum Mol Genet* **2016,** *25* (20), 4566-4576.

49. Skipper, K. A.; Mikkelsen, J. G., Toward In Vivo Gene Therapy Using CRISPR. *Methods Mol Biol* **2019,** *1961*, 293-306.

50. Song, M., The CRISPR/Cas9 system: Their delivery, in vivo and ex vivo applications and clinical development by startups. *Biotechnol Prog* **2017,** *33* (4), 1035-1045.

51. Wilson, R. C.; Gilbert, L. A., The Promise and Challenge of In Vivo Delivery for Genome Therapeutics. *ACS Chem Biol* **2018,** *13* (2), 376-382.

52. Chew, W. L., Immunity to CRISPR Cas9 and Cas12a therapeutics. *Wiley Interdiscip Rev Syst Biol Med* **2018,** *10* (1).

53. Dowdy, S. F., Overcoming cellular barriers for RNA therapeutics. *Nature Biotechnology* **2017,** *35* (3), 222-229.

54. O'Geen, H.; Yu, A. S.; Segal, D. J., How specific is CRISPR/Cas9 really? *Curr Opin Chem Biol* **2015,** *29*, 72-78.

55. Cheng, Q.; Wei, T.; Farbiak, L.; Johnson, L. T.; Dilliard, S. A.; Siegwart, D. J., Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nat Nanotechnol* **2020,** *15* (4), 313-320.

56. Kowalski, P. S.; Rudra, A.; Miao, L.; Anderson, D. G., Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. *Mol Ther* **2019,** *27* (4), 710-728.

57. Miller, J. B.; Zhang, S.; Kos, P.; Xiong, H.; Zhou, K.; Perelman, S. S.; Zhu, H.; Siegwart, D. J., Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA. *Angew Chem Int Ed Engl* **2017,** *56* (4), 1059-1063.

58. Wang, M.; Zuris, J. A.; Meng, F.; Rees, H.; Sun, S.; Deng, P.; Han, Y.; Gao, X.; Pouli, D.; Wu, Q.; Georgakoudi, I.; Liu, D. R.; Xu, Q., Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proceedings of the National Academy of Sciences* **2016,** *113* (11), 2868.

59. Yin, H.; Song, C. Q.; Suresh, S.; Wu, Q.; Walsh, S.; Rhym, L. H.; Mintzer, E.; Bolukbasi, M. F.; Zhu, L. J.; Kauffman, K.; Mou, H.; Oberholzer, A.; Ding, J.; Kwan, S. Y.; Bogorad, R. L.; Zatsepin, T.; Koteliansky, V.; Wolfe, S. A.; Xue, W.; Langer, R.; Anderson, D. G., Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat Biotechnol* **2017,** *35* (12), 1179-1187.

60. Khvorova, A.; Watts, J. K., The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol* **2017,** *35* (3), 238-248.

61. Meade, B. R.; Dowdy, S. F., The road to therapeutic RNA interference (RNAi): Tackling the 800 pound siRNA delivery gorilla. *Discov Med* **2009,** *8* (43), 253-6.

62. Smith, C. I. E.; Zain, R., Therapeutic Oligonucleotides: State of the Art. *Annu Rev Pharmacol Toxicol* **2019,** *59*, 605-630.

63. Filippova, J.; Matveeva, A.; Zhuravlev, E.; Stepanov, G., Guide RNA modification as a way to improve CRISPR/Cas9-based genome-editing systems. *Biochimie* **2019,** *167*, 49-60.

64. Kartje, Z. J.; Barkau, C. L.; Rohilla, K. J.; Ageely, E. A.; Gagnon, K. T., Chimeric Guides Probe and Enhance Cas9 Biochemical Activity. *Biochemistry* **2018,** *57* (21), 3027-3031.

65. Mir, A.; Alterman, J. F.; Hassler, M. R.; Debacker, A. J.; Hudgens, E.; Echeverria, D.; Brodsky, M. H.; Khvorova, A.; Watts, J. K.; Sontheimer, E. J., Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. *Nat Commun* **2018,** *9* (1), 2641.

66. O'Reilly, D.; Kartje, Z. J.; Ageely, E. A.; Malek-Adamian, E.; Habibian, M.; Schofield, A.; Barkau, C. L.; Rohilla, K. J.; DeRossett, L. B.; Weigle, A. T.; Damha, M. J.; Gagnon, K. T., Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity. *Nucleic Acids Res* **2019,** *47* (2), 546-558.

67. Hendel, A.; Bak, R. O.; Clark, J. T.; Kennedy, A. B.; Ryan, D. E.; Roy, S.; Steinfeld, I.; Lunstad, B. D.; Kaiser, R. J.; Wilkens, A. B.; Bacchetta, R.; Tsalenko, A.; Dellinger, D.; Bruhn, L.; Porteus, M. H., Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology* **2015,** *33* (9), 985-9.

68. Basila, M.; Kelley, M. L.; Smith, A. V. B., Minimal 2'-O-methyl phosphorothioate linkage modification pattern of synthetic guide RNAs for increased stability and efficient CRISPR-Cas9 gene editing avoiding cellular toxicity. *PloS one* **2017,** *12* (11), e0188593.

69. Li, B.; Zhao, W.; Luo, X.; Zhang, X.; Li, C.; Zeng, C.; Dong, Y., Engineering CRISPR-Cpf1 crRNAs and mRNAs to maximize genome editing efficiency. *Nat Biomed Eng* **2017,** *1* (5).

70. Park, H. M.; Liu, H.; Wu, J.; Chong, A.; Mackley, V.; Fellmann, C.; Rao, A.; Jiang, F.; Chu, H.; Murthy, N.; Lee, K., Extension of the crRNA enhances Cpf1 gene editing in vitro and in vivo. *Nature communications* **2018,** *9* (1), 3313.

71. Ha, D. I.; Lee, J. M.; Lee, N. E.; Kim, D.; Ko, J. H.; Kim, Y. S., Highly efficient and safe genome editing by CRISPR-Cas12a using CRISPR RNA with a ribosyl-2'-O-methylated uridinylate-rich 3'-overhang in mouse zygotes. *Experimental & molecular medicine* **2020,** *52* (11), 1823-1830.

72. Nguyen, L. T.; Smith, B. M.; Jain, P. K., Enhancement of trans-cleavage activity of Cas12a with engineered crRNA enables amplified nucleic acid detection. *Nature communications* **2020,** *11* (1), 4906.

73. Kim, H.; Lee, W. J.; Oh, Y.; Kang, S. H.; Hur, J. K.; Lee, H.; Song, W.; Lim, K. S.; Park, Y. H.; Song, B. S.; Jin, Y. B.; Jun, B. H.; Jung, C.; Lee, D. S.; Kim, S. U.; Lee, S. H., Enhancement of target specificity of CRISPR-Cas12a by using a chimeric DNA-RNA guide. *Nucleic acids research* **2020,** *48* (15), 8601-8616.

74. Yin, H.; Song, C. Q.; Suresh, S.; Kwan, S. Y.; Wu, Q.; Walsh, S.; Ding, J.; Bogorad, R. L.; Zhu, L. J.; Wolfe, S. A.; Koteliansky, V.; Xue, W.; Langer, R.; Anderson, D. G., Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat Chem Biol* **2018,** *14* (3), 311-316.

75. Rueda, F. O.; Bista, M.; Newton, M. D.; Goeppert, A. U.; Cuomo, M. E.; Gordon, E.; Kröner, F.; Read, J. A.; Wrigley, J. D.; Rueda, D.; Taylor, B. J. M., Mapping

the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat Commun* **2017,** *8* (1), 1610.

76. McMahon, M. A.; Prakash, T. P.; Cleveland, D. W.; Bennett, C. F.; Rahdar, M., Chemically Modified Cpf1-CRISPR RNAs Mediate Efficient Genome Editing in Mammalian Cells. *Mol Ther* **2018,** *26* (5), 1228-1240.

77. Rahdar, M.; McMahon, M. A.; Prakash, T. P.; Swayze, E. E.; Bennett, C. F.; Cleveland, D. W., Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. *Proc Natl Acad Sci USA* **2015,** *112* (51), E7110-7.

78. Ryan, D. E.; Taussig, D.; Steinfeld, I.; Phadnis, S. M.; Lunstad, B. D.; Singh, M.; Vuong, X.; Okochi, K. D.; McCaffrey, R.; Olesiak, M.; Roy, S.; Yung, C. W.; Curry, B.; Sampson, J. R.; Bruhn, L.; Dellinger, D. J., Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic acids research* **2018,** *46* (2), 792-803.

79. Cromwell, C. R.; Sung, K.; Park, J.; Krysler, A. R.; Jovel, J.; Kim, S. K.; Hubbard, B. P., Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat. commun.* **2018,** *9* (1), 1448.

80. Jellinek, D.; Green, L. S.; Bell, C.; Lynott, C. K.; Gill, N.; Vargeese, C.; Kirschenheuter, G.; McGee, D. P.; Abesinghe, P.; Pieken, W. A.; et al., Potent 2'-amino-2'-deoxypyrimidine RNA inhibitors of basic fibroblast growth factor. *Biochemistry* **1995,** *34* (36), 11363-72.

81. Hobbs, J.; Sternbach, H.; Sprinzl, M.; Eckstein, F., Polynucleotides containing 2'-amino-2'-deoxyribose and 2'-azido-2'-deoxyribose. *Biochemistry* **1973,** *12* (25), 5138-5145.

82. Guschlbauer, W.; Jankowski, K., Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Research* **1980,** *8* (6), 1421-1433.

83. Moffatt, J. G.; Verheyden, J. P. H.; Wagner, D., Synthesis of some pyrimidine 2'-amino-2'-deoxynucleosides. *The Journal of Organic Chemistry* **1971,** *36* (2), 250-254.

84. Svoboda, M., Jonas, J., Sicher, J. , Stereochemical studies. X. The effect of ring size and configuration on the basicity of the 2-aminocyclanols. *Collect. Czech. Chem. Commun.* **1958,** *23*, 1551-1558.

85. Inouye, S., On the Prediction of pK_a Values of Amino Sugars. *CHEMICAL & PHARMACEUTICAL BULLETIN* **1968,** *16* (6), 1134-1137.

86. Barlow, C. B.; Guthrie, R. D.; Prior, A. M., Nitrogen-containing carbohydrate derivatives: Part XX. Ionisation constants of amino sugars. *Carbohydrate Research* **1969,** *10* (4), 481-485.

87. Ferré-D'Amaré, A. R.; Doudna, J. A., RNA folds: insights from recent crystal structures. *Annu Rev Biophys Biomol Struct* **1999,** *28*, 57-73.

88. Liu, F.; Theimer, C. A., Telomerase activity is sensitive to subtle perturbations of the TLC1 pseudoknot 3' stem and tertiary structure. *Journal of molecular biology* **2012,** *423* (5), 719-735.

89. Kershaw, C. J.; O'Keefe, R. T., Splint ligation of RNA with T4 DNA ligase. *Methods Mol Biol* **2012,** *941*, 257-69.

90. Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R., The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry* **1993,** *32* (16), 4207-15.

91. Ui-Tei, K.; Naito, Y.; Zenno, S.; Nishi, K.; Yamato, K.; Takahashi, F.; Juni, A.; Saigo, K., Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. *Nucleic Acids Research* **2008,** *36* (7), 2136-2151.

92. Baker, E. S.; Dupuis, N. F.; Bowers, M. T., DNA Hairpin, Pseudoknot, and Cruciform Stability in a Solvent-Free Environment. *The Journal of Physical Chemistry B* **2009,** *113* (6), 1722-1727.

93. Fu, L.; Zhuang, J.; Tang, D.; Que, X.; Lai, W.; Chen, G., DNA pseudoknot-functionalized sensing platform for chemoselective analysis of mercury ions. *Analyst* **2012,** *137* (19), 4425-4427.

94. Forconi, M.; Schwans, J. P.; Porecha, R. H.; Sengupta, R. N.; Piccirilli, J. A.; Herschlag, D., 2'-Fluoro substituents can mimic native 2'-hydroxyls within structured RNA. *Chemistry & biology* **2011,** *18* (8), 949-954.

95. Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M., 2'-Fluoro RNA shows increased Watson-Crick H-bonding strength and stacking relative to RNA: evidence from NMR and thermodynamic data. *Angew Chem Int Ed Engl* **2012,** *51* (47), 11863-6.

96. Habibian, M.; Martínez-Montero, S.; Portella, G.; Chua, Z.; Bohle, D. S.; Orozco, M.; Damha, M. J., Seven-Membered Ring Nucleoside Analogues: Stereoselective Synthesis and Studies on Their Conformational Properties. *Organic Letters* **2015,** *17* (21), 5416-5419.

97. Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; González, C., The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res* **2012,** *40* (18), 9329-39.

98. Aurup, H.; Tuschl, T.; Benseler, F.; Ludwig, J.; Eckstein, F., Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: thermal stability and chemical reactivity. *Nucleic Acids Res* **1994,** *22* (1), 20-4.

99. Williams, D. M.; Pieken, W. A.; Eckstein, F., Function of specific 2'-hydroxyl groups of guanosines in a hammerhead ribozyme probed by 2' modifications. *Proceedings of the National Academy of Sciences of the United States of America* **1992,** *89* (3), 918-921.

100. Yan, X.; Gao, X.; Zhang, Z., Isolation and characterization of 2'-amino-modified RNA aptamers for human TNFalpha. *Genomics Proteomics Bioinformatics* **2004,** *2* (1), 32-42.

101. Chen, J. S.; Ma, E.; Harrington, L. B.; Da Costa, M.; Tian, X.; Palefsky, J. M.; Doudna, J. A., CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **2018,** *360* (6387), 436-439.

102. Stella, S.; Mesa, P.; Thomsen, J.; Paul, B.; Alcón, P.; Jensen, S. B.; Saligram, B.; Moses, M. E.; Hatzakis, N. S.; Montoya, G., Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **2018,** *175* (7), 1856-1871.e21.

103. Swarts, D. C.; Jinek, M., Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Mol Cell* **2019,** *73* (3), 589-600.e4.

104. Nilsen, T. W., Nucleotide analog interference mapping. *Cold Spring Harb Protoc* **2015,** *2015* (6), 604-8.

105. Westheimer, F. H., Why nature chose phosphates. *Science (New York, N.Y.)* **1987,** *235* (4793), 1173-1178.

106. Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; Severinov, K.; Zhang, F.; Koonin, E. V., Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **2017,** *15* (3), 169-182.

107. Swarts, D. C.; van der Oost, J.; Jinek, M., Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol Cell* **2017,** *66* (2), 221-233.e4.

108. Deleavey, G. F.; Damha, M. J., Designing chemically modified oligonucleotides for targeted gene silencing. *Chem Biol* **2012,** *19* (8), 937-54.

109. Chylinski, K.; Makarova, K. S.; Charpentier, E.; Koonin, E. V., Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **2014,** *42* (10), 6091-105.

110. Hüttenhofer, A.; Schattner, P., The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat Rev Genet* **2006,** *7* (6), 475-82.

111. Nguyen, L. T.; Smith, B. M.; Jain, P. K., Enhancement of trans-cleavage activity of Cas12a with engineered crRNA enables amplified nucleic acid detection. *Nature Communications* **2020,** *11* (1), 4906.

112. Sancak, Y.; Peterson, T. R.; Shaul, Y. D.; Lindquist, R. A.; Thoreen, C. C.; Bar-Peled, L.; Sabatini, D. M., The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **2008,** *320* (5882), 1496-501.

113. Stewart, S. A.; Dykxhoorn, D. M.; Palliser, D.; Mizuno, H.; Yu, E. Y.; An, D. S.; Sabatini, D. M.; Chen, I. S.; Hahn, W. C.; Sharp, P. A.; Weinberg, R. A.; Novina, C. D., Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **2003,** *9* (4), 493-501.

114. Habibian, M.; Harikrishna, S.; Fakhoury, J.; Barton, M.; Ageely, E. A.; Cencic, R.; Fakih, H. H.; Katolik, A.; Takahashi, M.; Rossi, J.; Pelletier, J.; Gagnon, K. T.; Pradeepkumar, P. I.; Damha, M. J., Effect of 2'-5'/3'-5' phosphodiester linkage heterogeneity on RNA interference. *Nucleic Acids Res* **2020,** *48* (9), 4643-4657.

115. Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A., A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **2015,** *348* (6242), 1477-81.

- page intentionally blank -

- page intentionally blank -

Chapter 5

RNA mimics for CRISPR Cas12a gene editing

5.1 Introduction

CRISPR-Cas is a prokaryotic adaptive immunity system used to cleave invading nucleic acids. In recent years, researchers have been able to repurpose these programmable systems in a number of impactful genome engineering applications[1-7]. CRISPR technologies promise to revolutionize therapeutic medicines in the very near future. An assortment of CRISPR–Cas systems exist across diverse species of bacteria and archaea, which differ in their components and mechanisms of action. The most comprehensively studied example of CRISPR-Cas systems is type II (falling under the Class 2 systems), which is characterized by the signature gene *cas9*[8]. Cas9 is capable of programmable RNA-guided DNA interference without the need for additional proteins[1], a property that has been used to repurpose Cas9 for genome editing. Cas enzymes can be programmed with a guide RNA to form a ribonucleoprotein complex (RNP) which is then guided to partially unwind and target duplex DNA. Once Cas9 is bound to the target DNA, a nuclease domain causes double stranded breaks cleaving the DNA, which cellular systems are quick to repair, often incorporating insertions or deletions, ultimately leading to genome editing. Cas12a (previously named Cpf1) has recently emerged as an alternative for Cas9.

Cas9 and Cas12a have distinct evolutionary origins and exhibit different structural architectures, resulting in distinct molecular mechanisms (**Figure 5.1**) [9]. They differ in their target recognition mechanisms by identifying and binding to different protospacer-adjacent motif (PAM) sequences in the target DNA; Cas9 recognizes a G-rich PAM (5'-NGG-3') downstream of the cleavage site while Cas12a recognizes a T-rich PAM (5'-TTTN-3') upstream of the cleavage site (**Figure 5.1** Middle). For Cas9, dsDNA cleavage is accomplished by the concerted activation of its two nuclease domains, HNH and RuvC, which respectively cut the target and non-target strands, upstream of the protospacer adjacent motif (PAM) sequence, generating blunt ends (**Figure 5.1** Bottom). For Cas12a, the RuvC nuclease is independently responsible for cleaving both the target strand and the non-target strand downstream of PAM, but the precise cleavage mechanism remains unknown. It does, however, generates staggered ends with 5' nucleotide overhangs. Importantly, for therapeutic development purposes, the guide RNA of Cas12a is 39 nucleotides long, making it much shorter than the 99mer single guide RNA (sgRNA) generally used for the Cas9 protein. A shorter oligonucleotide simplifies chemical

synthesis and allows us to incorporate a greater variety of modifications, as yields are less jeopardized.

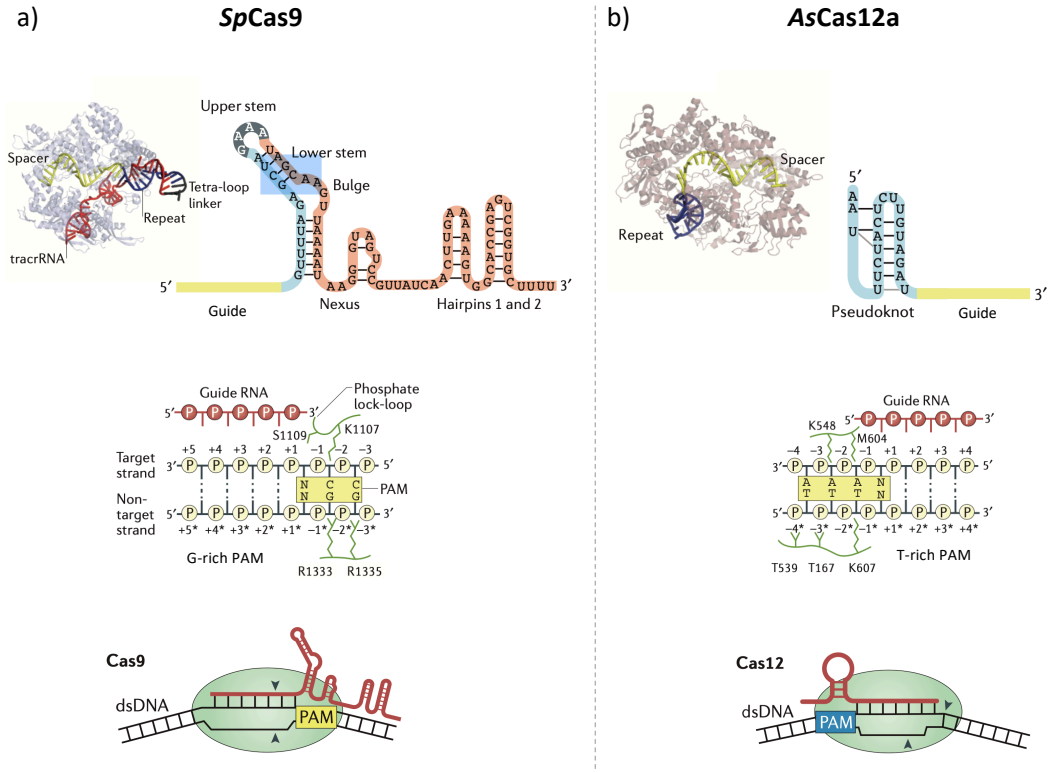


Figure 5.1. Distinct properties differentiating SpCas9 (a) from AsCas12a (b). *(Top) The full-length single guide crRNA (~99nts) required for Cas9 activity vs. 39-nucleotide crRNA required for Cas12a activity. (Middle) G-rich PAM sequence recognized by Cas9 downstream of the cleavage site vs. T-rich PAM sequence recognized by Cas12a upstream of the cleavage site. (Bottom) Identical cleavage sites on both DNA strands, generated by Cas9, upstream of PAM vs. Staggered cleavage sites on the DNA strands, generated by Cas12a, downstream of PAM. Figure adapted with permission from reference 10.*

While shorter crRNAs should be easier to design and synthesize, the atypical 5' pseudoknot handle of the guide RNA makes it difficult to predict which nucleotide modifications are best tolerated. In fact, the 5' pseudoknot handle has so far posed great challenges as it appeared to tolerate very little modification to the ribose sugar, suggesting that successful modification of the pseudoknot structure would be the major limitation. In Chapter 4, we described a variety of modifications that allowed us to identify modifications compatible with editing activity in cells. We also identified critical crRNA/Cas12 and crRNA/Cas9 interactions by using a series of crRNAs that are modified at the predicted 2'-OH contact positions. In this chapter, we aim to probe these interactions further by examining additional RNA analogues. We also examine the impact of these modifications on the structure of the pseudoknot.

5.2 RNA Mimics Which Retain Hydroxyl Contacts Provide Unique RNA Modifications

The main conclusion from the studies described in Chapter 4 is the definite requirement for conserved polar contacts between the protein and the 2'OH of ribose for observing robust in-cell editing activity[11-14]. Here we focused our efforts on a regioisomer of RNA, namely 2',5'-linked RNA, as well as the 2'-stereoisomer of RNA, namely arabinonucleic acids (ANA) (**Figure 5.2**). The appealing aspect of these isomers is their similar ionic character, hydrophobicity, and mass to the native RNA. Additionally, monomer availability and affordability make synthesis of these analogues very attractive RNA mimics.

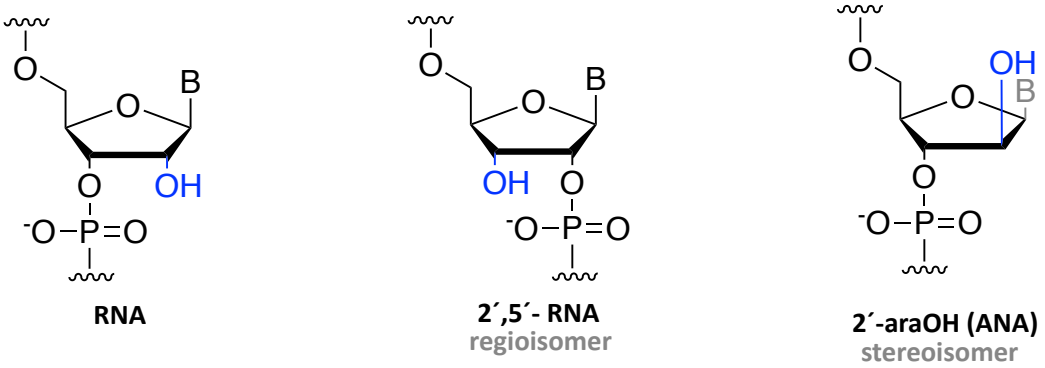


Figure 5.2*. The structure of RNA compared to its regioisomer, 2',5'-RNA (middle), and stereoisomer, 2'-araOH (ANA) (right).*

5.2.1 2',5'-linked RNA as an attractive RNA modification

2',5'-Linked RNA (**Figure 5.3**) are more resistant towards nuclease mediated degradation compared to RNA, which makes it an attractive RNA mimic for therapeutic applications[15]. Previous studies from our lab demonstrated that 2',5'-linkages had the added benefit of abrogating immunostimulatory activities of siRNA[16]. Recent high-resolution crystallographic data show that RNA duplexes containing a few 2',5'-linkages share the same global A-like structure as native duplexes, suggesting that RNA duplexes can accommodate perturbations caused by these linkages[17, 18]. Remarkably, RNA helical structures are well retained even with 40 % backbone heterogeneity (i.e., 2',5'-/3',5'-mixmers)[17]While the presence of 2',5'-linkages destabilizes RNA duplexes,[19,20], it is not

unknown how this modification may impact RNA pseudoknot structure and stability. All these features made 2',5'-RNA an excellent system to study.

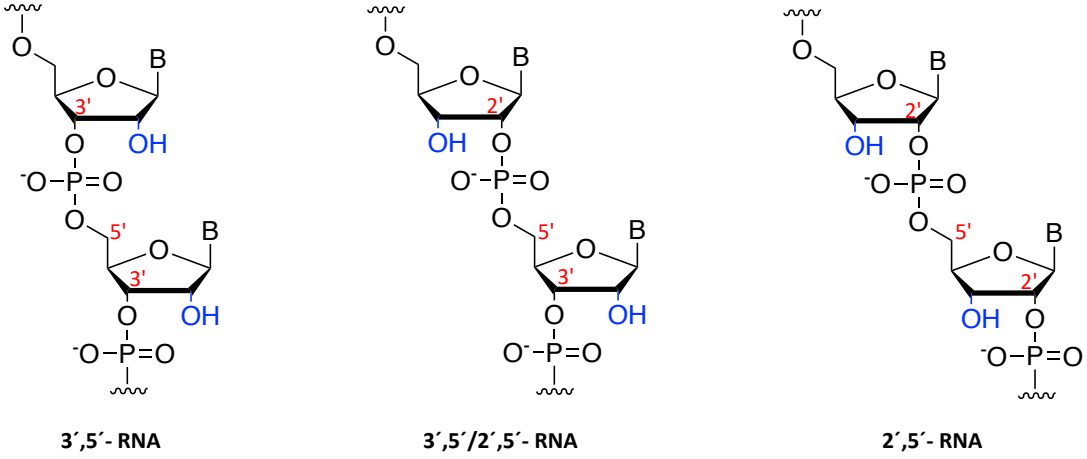


Figure 5.3. *Structures of regioisomeric RNAs linkages. Figure adapted with permission from reference 16.*

5.2.2 2'-Arabinonucleic acid (ANA) as an attractive RNA modification

Arabinonucleic acid (ANA) is the 2'-epimer of RNA. A seemingly simple switch in the orientation of the 2'OH group from "down" to "up" (**Figure 5.2**) makes the backbone resistant to chemical and enzymatic cleavage, and yet without affecting the ability of ANA to hybridize to complementary RNA (albeit with some loss of affinity). NMR and molecular dynamic studies of ANA:RNA hybrids revealed a C2'-O-H--O(5') bonding interaction which steers the sugar toward the C2'-endo (South) sugar conformation. Furthermore, they revealed unfavourable steric interactions between the 2'-araOH and the nucleobase, in part explaining the lower stability of ANA:RNA hybrids compared to pure RNA duplexes.[25, 26]. These interactions are alleviated when the ANA nucleotides form part of a single stranded loop structure. We hypothesized that this could also apply to the folded Cas12a pseudoknot structure. Of note, ANA:RNA hybrids are substrates of RNase H and the RNA-induced silencing complex[21-24].

5.2.3 Synthesis of modified crRNAs containing 2',5'-RNA and ANA residues

We set out to synthesize a series of RNA strands containing either 2'-F/2'-OH/ 2',5'-RNA (LA-T14 series; **Table 5.1.a**) or 2'-F/2'-OH/2'-araOH residues (LA-A14 series; **Table 5.1**,**b**). These RNAs were chosen based on the finding that 2'F-RNA strands that

retain critical 2'-OH contacts (cpEGFP-SJ14) exhibit excellent in-cell gene editing activity (Chapter 4 – Section 4.3.2). The ANA and 2'-5'-RNA residues were systematically introduced to replace each of the predicted 2' OH contact positions (-18, -17, -14, -13, -10, -6, and -1) (**Table 5.1**).

To ensure efficient coupling of ANA 3'-amidites and RNA 2'-amidites during our automated synthesis cycle, these monomers were allowed to react for 900 seconds. A detailed procedure for the synthesis of this series is summarized in Section **5.5.15.5.1 Oligonucleotide Synthesis**. It is important to note that in most cases, the pseudoknot 19-nt sequences were enzymatically ligated to an unmodified guide 20-nt RNA sequence to yield the entire 39-nt Cas12a crRNA sequences. The exceptions were sequences LA-T14-7-1 and LA-T14-7-2, LA-A14-7-1, and LA-A14-7-2. In these cases, the entire 39-nt strand was assembled chemically via solid phase synthesis to avoid possible complications encountered during enzymatic ligation.

Table 5.1. *LA-T14 (a) and LA-A14 (b) series of oligonucleotides synthesized, with ESI mass spectroscopy characterization.*

a) 2′,5′- linkages (3′-OH) at critical contacts			
Name	Sequence	$Mass_{exp}$	$Mass_{obs}$
LA-T14-1	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.82
LA-T14-2	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.78
LA-T14-3	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.78
LA-T14-4	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.78
LA-T14-5	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.69
LA-T14-6	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.69
LA-T14-7-1	5′ – AAU UUC UAC UCU UGU AGA U CCC UUU UUG AGU UUG GAU CU - 3′	12291.53	12290.65
LA-T14-7-2	5′ – AAU UUC UAC UCU UGU AGA U CGU CGC CGU CCA GCU CGA CC - 3′	12324.66	12345.65
b) 2′-araOH at critical contacts			
Name	Sequence	$Mass_{exp}$	$Mass_{obs}$
LA-A14-1	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.31
LA-A14-2	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5965.37
LA-A14-3	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.31
LA-A14-4	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.31
LA-A14-5	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.31
LA-A14-6	5′ – AAU UUC UAC UCU UGU AGA U - 3′	5966.78	5966.31
LA-A14-7-1	5′ – AAU UUC UAC UCU UGU AGA U CCC UUU UUG AGU UUG GAU CU - 3′	12285.49	12289.67
LA-A14-7-2	5′ – AAU UUC UAC UCU UGU AGA U CGU CGC CGU CCA GCU CGA CC - 3′	12318.61	12322.60

N – 2′-F N – 2′-OH N – 2′,5′ linkage N – 2′-araOH

5.2.4 Gene editing activity of 2',5'-RNA and 2'-araOH series

Acidominococcus species Cas12a (*As*Cas12a) and the 39-nucleotide crRNA shown above were used in the following studies. It is important to note that for this series of oligonucleotides, we moved directly to cell-based editing, as we learned from our previous studies (Chapter 4) that *in vitro* cleavage activity is not always predictive of cell-based (*in vivo*) activity. RNA guides LA-T14-7-1 and LA-A14-7-1 served as negative controls.

Oligonucleotides were transfected into HEK293T cells stably expressing EGFP and AsCas12a and the results are summarized below.

The 2',5' series showed editing activity ranging from very low to considerably high (**Figure 5.4**). LA-T14-1 and LA-T14-7-2 showed editing activity below 10%, in line with our previous finding that terminal contact positions (-18 and -1) are less tolerant to modifications. Unsurprisingly, LA-T14-7-1 also showed no editing activity, acting as a negative control for this assay, as it targeted a different guide region. LA-T14-6, with a 2',5'-linkage at position -6 appears to tolerate the modification rather well, as it showed very impressive editing activity (~70 %), which is almost comparable to that of native RNA (cpEGIPe). The 2'-ara series, on the other hand, showed moderate editing activity across the board, except for LA-A14-1, which showed a slightly higher editing efficiency of 50% (**Figure 5.5**). We were a bit surprised by this result since terminal hydroxyl contact positions tend to be more sensitive to modifications. LA-A14-7-1 again showed negligibly low activity, as observed for the negative control.

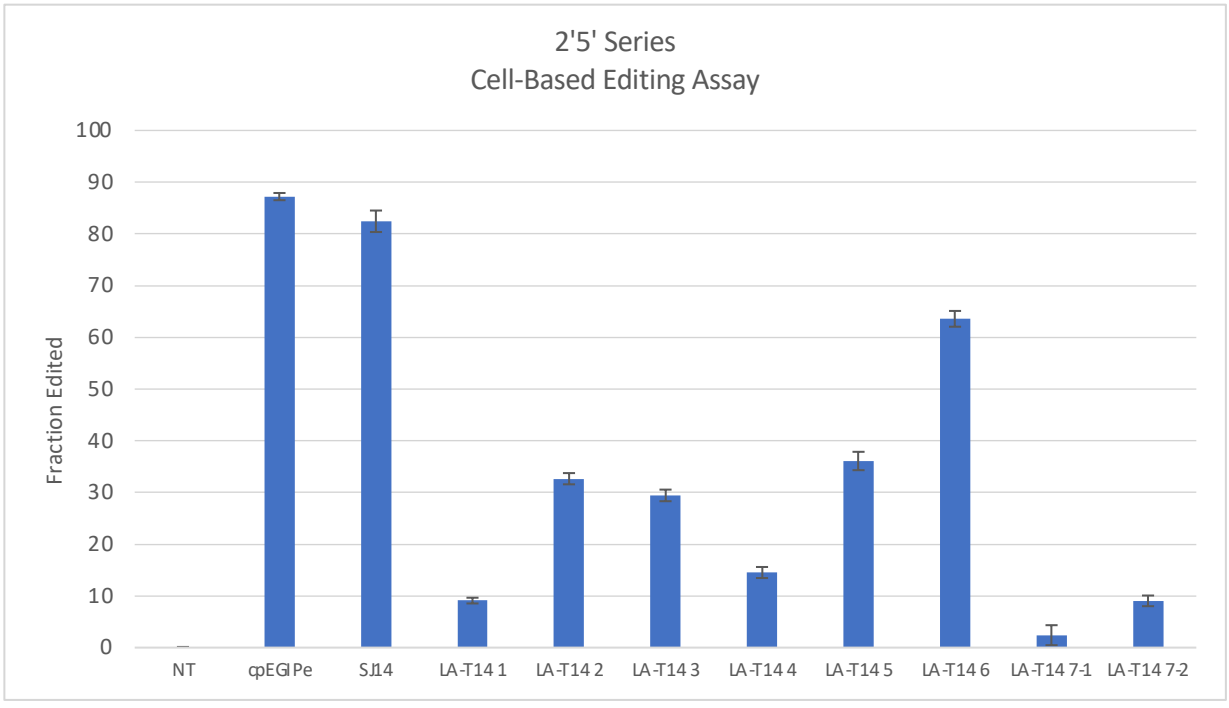


***Figure 5.4**. Cell-based editing activity for 2'5'-RNA series.*

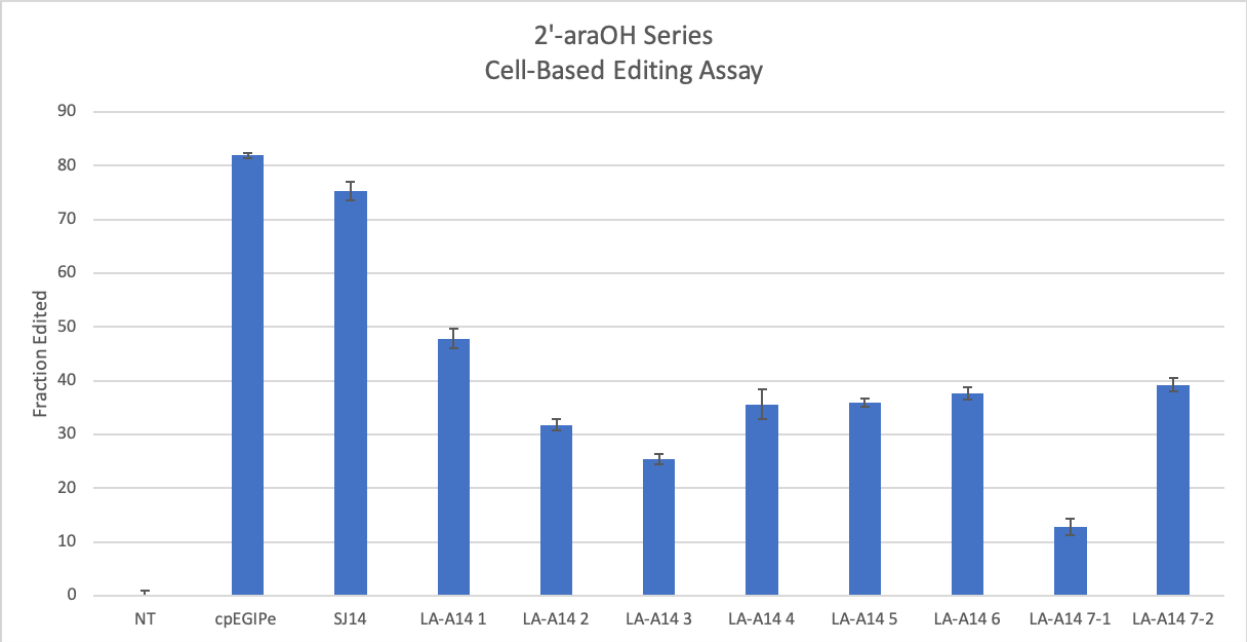


Figure 5.5. *Cell-based editing activity for 2'-araOH series*

In general, it appears that 2'-ara modifications are better tolerated than 2',5'-linked RNA. Going forward, it would be interesting to combine the modifications of LA-T14-6 and LA-A14-1 to assess potential synergistic effects between ANA and 2',5'-RNA analogues. Since modifications at position the 3' end of the pseudoknot are not tolerated, it would be interesting to incorporate an acyclic nucleotide modification to introduce flexibility in this location. Ultimately, our goal is to design an oligonucleotide that incorporates our most effective modifications.

5.3. The Importance of the Base Pairing Region of Cas12a 5'Pseudoknot on Structure Formation and Stability

*As*Cas12a crRNA contains a 20-nt pseudoknot linked to the 5' end of a 23-nt spacer (guide sequence). The pseudoknot structure is stabilized by five Watson–Crick base pairs, a noncanonical U–U base pair, one UCUU tetraloop, one reverse Hoogsteen A–U base pair, and three 5′-overhangs (**Figure 5.6**)[27]. Until now, we had been mainly been focussing on the putative 2'-OH contacts, however we felt it was also important to determine how

modifications incorporated in the 5-bp stem (highlighted in grey; positions -2 to -6 and -11 to -15) affected pseudoknot structure and stability.

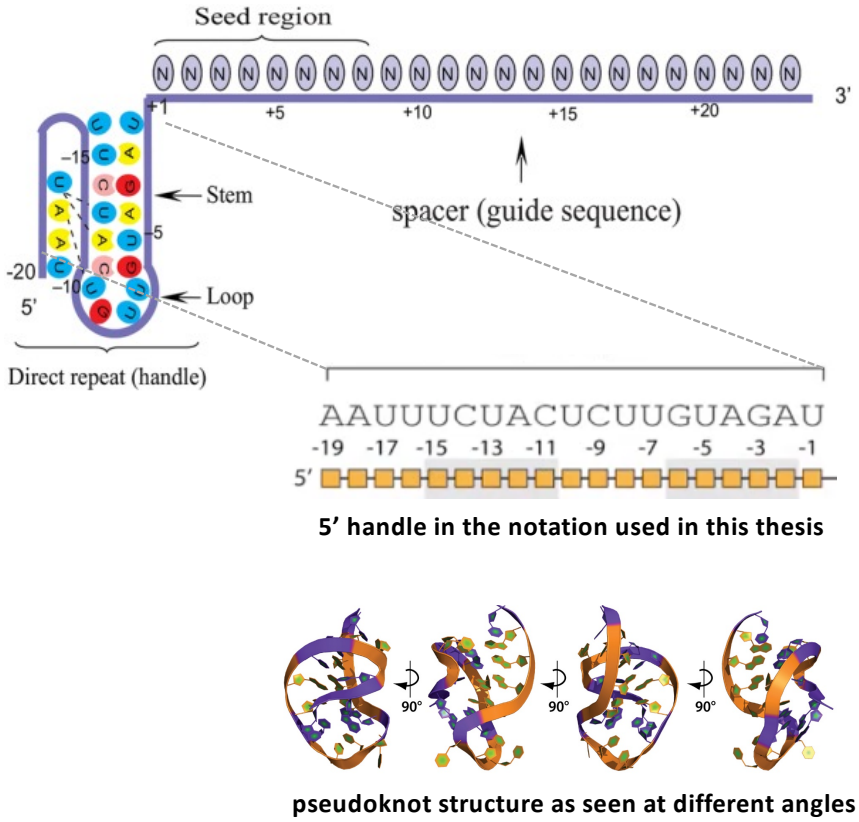


Figure 5.6. *Schematic representation of mature Cas12a crRNA. Figure adapted from reference* [28]

5.3.1 Incorporating DNA modifications in the base pairing region of the pseudoknot

For Cas9, retention of editing activity has been suggested to correlate with conserved polar contacts between the protein and the hydroxyl at the ribose 2' position and the need to retain A-form-like helical architecture[11-14]. We hypothesized that Cas12a also required an A-like helical pseudoknot structure and decided to probe this. The presence (or absence) of the 2'-hydroxyl on the sugar that distinguishes RNA from DNA determines the conformational preference for A form (C-3' endo) or B form (C-2' endo) conformations. RNA duplexes prefer to adopt an A-form helical structure[29], whereas DNA is polymorphic and can adopt B or A form, with preference of the B form under physiological conditions[29]. We therefore introduced DNA residues in the pseudoknot stem (LA-pk-1, 2, and 3; **Table**

5.2). The all-RNA and all-DNA sequences served as controls (LA-RNA-pk and LA-DNA-pk) (**Table 5.2**).

Table 5.2. Initial series of sequences designed to probe pseudoknot structure and stability. *Red asterisks represent 2'-hydroxyl contact position. Mass spectrometry characterization on right.*

Sample	Sequence	M_{exp}	M_{obs}
LA-RNA-pk	5' – AAU UUC UAC UCU UGU AGA U - 3'	5942.7512	5942.7813
LA-DNA-pk	5' – AAT TTC UAC TCT TGT AGA T - 3'	5764.9922	5764.9375
LA-pk-1	5' – AAU UTC TAC UCU UGT AGA U - 3'	5824.8512	5824.8438
LA-pk-2	5' – AAU UTC TAC UCU UGU AGA U - 3'	5890.8082	5890.7500
LA-pk-3	5' – AAU UUC UAC UCU UGT AGA U - 3'	5876.7942	5875.7500
	** * * * * *		

N – 2′-OH (RNA) N – 2′-H (DNA) **N – base pairing region**

5.3.2 Testing for the formation of other secondary structures of the 5'-handle

The sequences shown in Table 5.2 were analyzed by polyacrylamide gel electrophoresis (PAGE) under native conditions, so as not to disturb formation of a folded (pseudoknot-like) structure. One batch of the samples were loaded directly on the gel. A second batch was first annealed by first heating at 95˚ C for 5 minutes, and then allowing them to cool to room temperature before loading into the gel. Only a single species was observed for all sequences (**Figure 5.7**), suggesting that they fold spontaneously with or without pre-annealing. Lane 12 was loaded with an oligonucleotide known to form a 1:1 duplex/hairpin mixture (the slower migrating band (top) is the bimolecular duplex, whereas the faster moving band is the unimolecular hairpin). Lane 11 contains a control DNA/RNA duplex structure formed by mixing LA-RNA-pk (the native RNA pseudoknot) with its complementary DNA strand. The control migrating in lane 13 is the RNA pseudoknot (LA-RNA-pk) loaded in 50 % formamide after first boiling at 95 ˚C for 5 minutes. The absence of a slower migrating band in all the modified LA-pk sequences (lanes 2-10) suggest exclusive formation of a stable, unimolecular pseudoknot-like structure that forms spontaneously and resists denaturation.

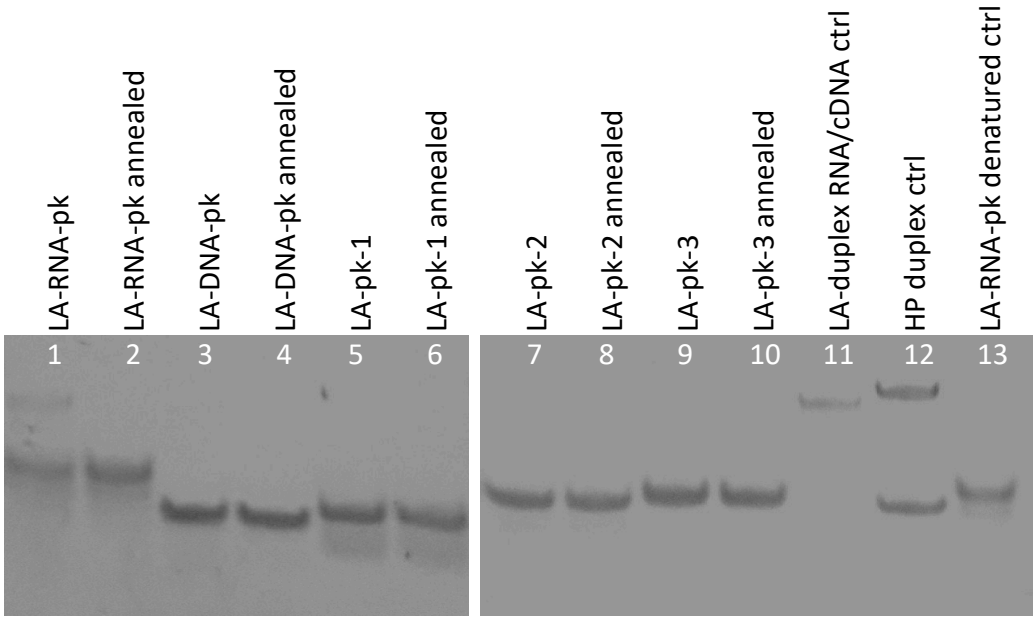


Figure 5.7. Native PAGE results showing robustness of the 5' handle pseudoknot. *Gel conditions: 20% acrylamide, 1x MOPS, 50 mM KCl, 2 mM $MgCl_2$, 20 mM NaP_i, pH 7.3. Sample concentration: 50 µM in 15 µL with final glycerol concentration of 7 % (except denatured control (13), 55:45 formamide:water).*

5.3.2.i The use of MOPS buffer for structural studies of the Cas12a pseudoknot

3-(Morpholin-4-yl)propane-1-sulfonic acid (MOPS; pKa 7.20[30]) is an excellent buffer for studying biomolecules at near-neutral pH. We chose this buffer over the more common Tris buffer (pK_a of 8.06[30]) since it requires a considerable amount of acid (HCl) to be added to achieve neutrality. The high concentration of Cl^- anions impact the salt concentration in the buffers, which in turn, may impact structure and gel analysis.

5.3.3 UV thermal melting studies reveal stability of Cas12a pseudoknot

To test the thermal stability of the pseudoknot structures, we produced thermal denaturation curves by measuring the UV absorbance of our single strand pseudoknots while heating the samples from 5 ˚C to 95 ˚C and then cooling back down to 5 ˚C. The full experiment is described in section 5.5.5. The T_m curves revealed that the native RNA pseudoknot exhibits two transitions centered at 14 ˚C and 62 ˚C (**Table 5.3, Figure 5.12**). We postulate that the 14 ˚C transition results from the intermolecular association of this sequence, i.e., the 5-nt UCUAC segment of one strand hybridizes with the 5-nt GUAGA segment of another strand. The presence of this duplex is is further supported by the presence of a very faint band in lane 1 of the gel shown in **Figure 5.7.** The rest of the

sequences show no evidence of bimolecular duplex formation. The T_m values of all LA-pk sequences were in the 45 – 62 ˚C range indicating the formation of stable folded species (**Table 5.3**). The lower T_m value of LA-pk-1 relative to LA-RNA-pk is consistent with the observation that DNA:RNA hybrids are generally less stable than RNA:RNA duplexes. The larger T_m value of LA-pk-3 (47 ˚C) vs LA-pk-2 (50 ˚C) is consistent with the observation that RNA (Purine-rich):DNA(Pyrimidine-rich) hybrids are more stable than RNA (Pyrimidine-rich):DNA(Purine-rich) hybrids[31].

Table 5.3. *UV Thermal stability data of pseudoknot control sequences.*

Sample	Sequence	T_m	ΔT_m
LA-RNA-pk	5' – **AAU UUC UAC UCU UGU AGA U** – 3'	62 ˚C (and 14 ˚C)	
LA-DNA-pk	5' – **AAT TTC UAC TCT TGT AGA T** – 3'	52 ˚C	-10 ˚C
LA-pk-1	5' – **AAU UTC TAC UCU UGT AGA U** – 3'	45 ˚C	-17 ˚C
LA-pk-2	5' – **AAU UTC TAC UCU UGU AGA U** – 3'	50 ˚C	-12 ˚C
LA-pk-3	5' – **AAU UUC UAC UCU UGT AGA U** – 3'	47 ˚C	-15 ˚C

** * * * * *

N – 2´-OH (RNA) **N – 2´-H (DNA)** **N – base pairing region**

5.3.4 Circular dichroism studies reveal helical structure of base pairing regions of Cas12a pseudoknot

To the best of our knowledge, there are no structural studies reported on the *As*Cas12a RNA pseudoknot guided by Circular Dichroism (CD). To gauge the importance of A-form helical structure on pseudoknot activity, we performed CD experiments on our pseudoknot structures. Circular dichroism of nucleic acid helices arises from the interaction of polarized light with the chiral and helical arrangement of its constituents[32]. Traditionally, an A-form (RNA) helix exhibits a positive band at 270 nm and a negative band at 210 nm[32]. B-form (DNA) helices exhibit a broad positive band in the 260 – 280 nm region, and a negative band at 245 nm[32]. DNA:RNA hybrids of mixed base sequence exhibit CD traces characteristic of "A-like" helices but also of both pure A and pure B helices[21]. These duplexes generally have more A-form character, with a reduced negative band at 210 nm and a broader positive band at 270 nm[21].

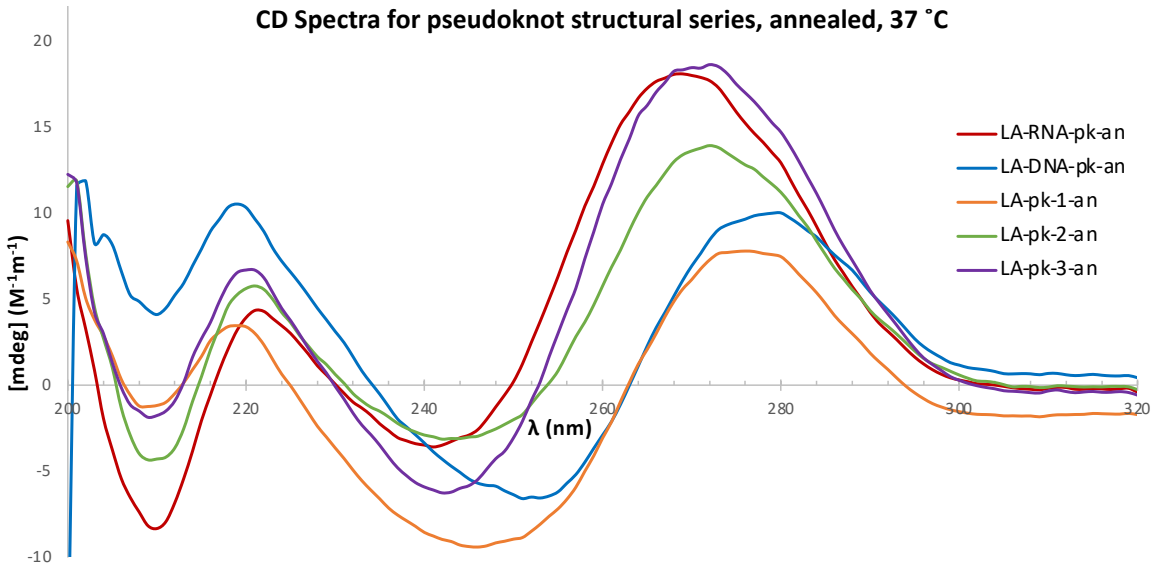


Figure 5.8. CD spectra of pseudoknot structural series. *Samples were slow annealed the day before and experiment was conducted at 37 ˚C. Pseudoknot concentration is 50 μM in 1x MOPS buffer, 50 mM KCl, 2 mM $MgCl_2$, 20 mM NaP_i, at pH 7.3.*

LA-RNA-pk exhibited a CD spectrum representative of an A-form helix and LA-DNA-pk exhibited B-form helicity (**Figure 5.8**). Unsurprisingly, LA-pk-1, which contains DNA in both base pairing regions, also shows a B-form helical character, although to a lesser extent than the all-DNA strand. The more rigid RNA in the residues outside of the base pairing region in LA-pk-1 may restrict the stem (base pairing region) in a way such that the B-form helicity is slightly reduced. LA-pk-2 exhibits a CD spectrum representative of a hybrid, with a positive maximum peak at 270, and a negative peak at 210, implying stronger A-form helicity. The spectrum for LA-pk-3 very closely resembles the that of the all-RNA, indicating that this pseudoknot folds to form a pure A-form structure.

Of particular interest to us was to monitor pseudoknot folding by heat annealing experiments. First, we compared the CD spectra of annealed *versus* not annealed samples. These are shown in **Figure 5.9**. As noted above, we did not observe any major differences among these samples, suggesting a fast (kinetic) and stable (thermodynamic) folding for the pseudoknot structures.

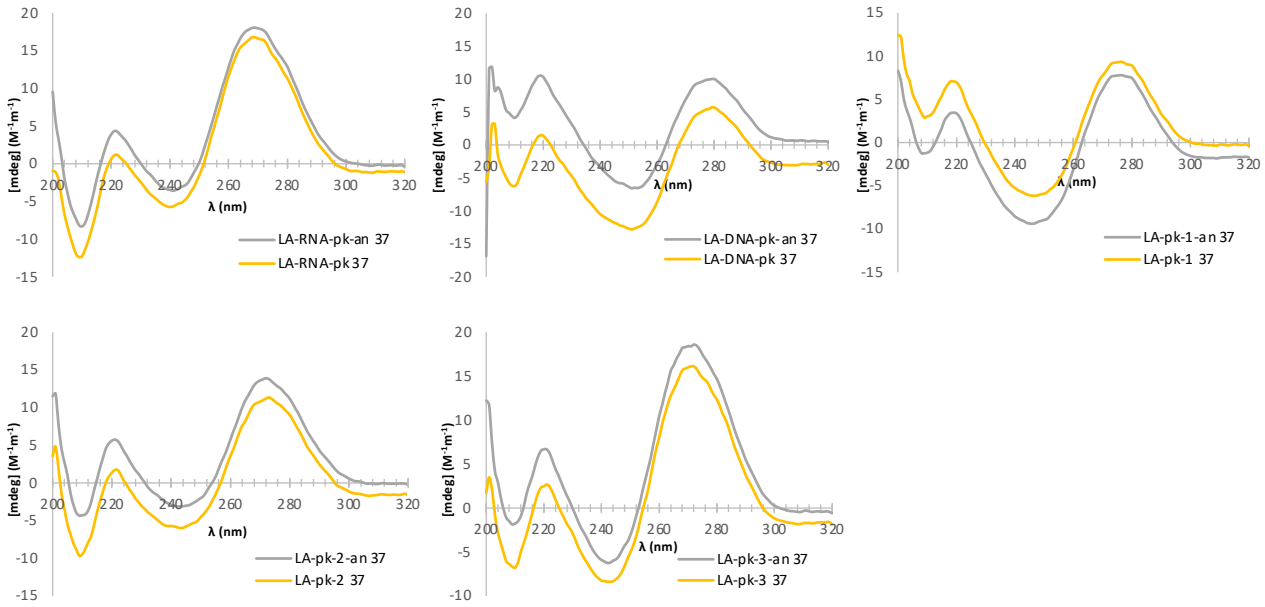


Figure 5.9. CD spectra of pseudoknot structural series comparing the effects of annealing on pseudoknot helicity. *Experiments were conducted at 37 ˚C. Pseudoknot concentration is 50 µM in 1x MOPS buffer, 50 mM KCl, 2 mM $MgCl_2$, 20 mM NaP_i, at pH 7.3.*

Next, we compared the CD spectra at 22 ˚C, 37 ˚C and 75 ˚C of previously annealed samples. We observed no apparent difference between the spectra taken at 22 ˚C and 37 ˚C (**Figure 5.10**), consistent with the thermal stability experiments which indicate that the pseudoknot structures are stable at temperatures up to 62 ˚C. As expected, the CD spectra of pseudoknots at 75 ˚C resembled those of unfolded, single strands (data not shown).

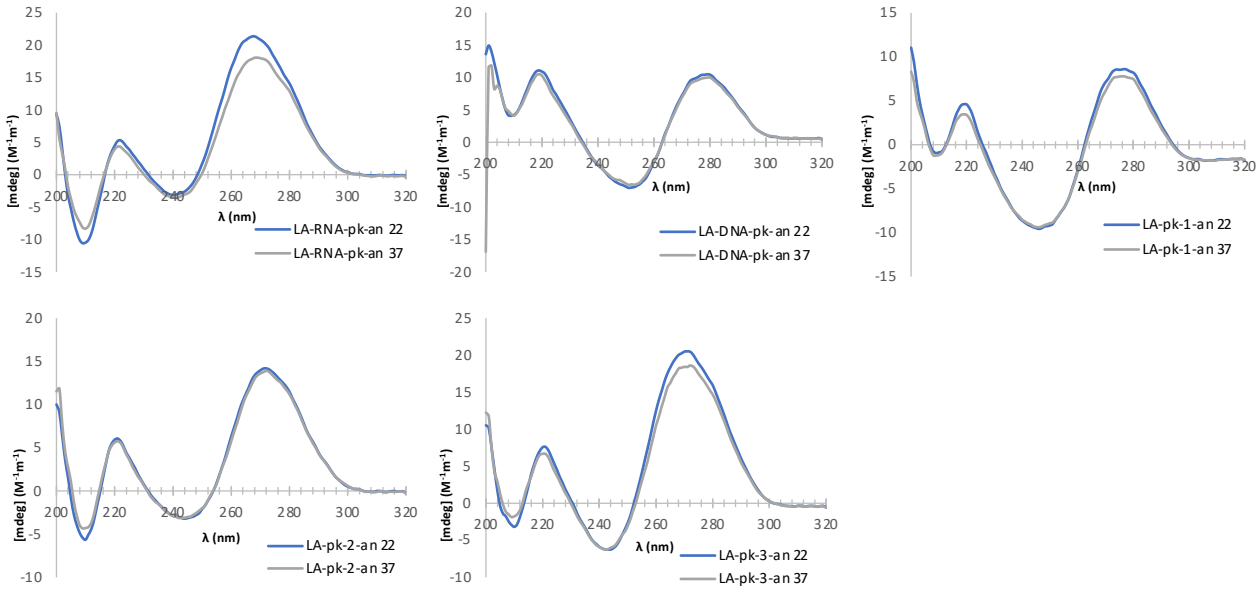


Figure 5.10. CD spectra of pseudoknot structural series comparing the effects of temperature on pseudoknot helicity. *Experiments were conducted at 22 ˚C and 37 ˚C. Pseudoknot concentration is 50 µM in 1x MOPS buffer, 50 mM KCl, 2 mM $MgCl_2$, 20 mM NaP_i, at pH 7.3.*

5.3.5 Modifications in the base pairing region affect editing activity of Cas12a

From the experiments described in Chapter 4, we learned that an all-DNA guide strand is unable to cleave a target DNA duplex. This is due to a combination of factors, i.e., requirement for an A-form RNA pseudoknot structure, and critical Cas12-RNA 2'-OH contacts required for cleavage. Will the same be true for crRNAs having a chimeric DNA-RNA pseudoknot such as LA-pk-2 and LA-pk-3 (**Table 5.3**)? Our CD studies suggested that the chimeric pseudoknots adopt an A-like conformation very similar to that of the all-RNA pseudoknot. Hence, it was of interest to see how these modified pseudoknots performed in our gene editing assays.

First, the modified pseudoknots were splint ligated to the guide RNA (Chapter 4) thus providing the required full-length crRNA for testing. As shown in **Figure 5.11**, LA-pk-3 shows an editing efficiency of ~45 %, while LA-pk-2 barely shows any activity. This correlates with the finding that the CD spectrum of LA-pk-3 most closely resembled that of the A-form structure of native RNA strand. Furthermore, LA-pk-3 lacks only one critical hydroxyl contact (position -6), whereas for LA-pk-2, lacks two (-13, -14) **(Table 5.3**). Together, these findings are consistent with the notion that an A-form pseudoknot

structure and critical 2'-OH/Cas12 contacts are important, at least in part, for observing efficient gene editing activity.

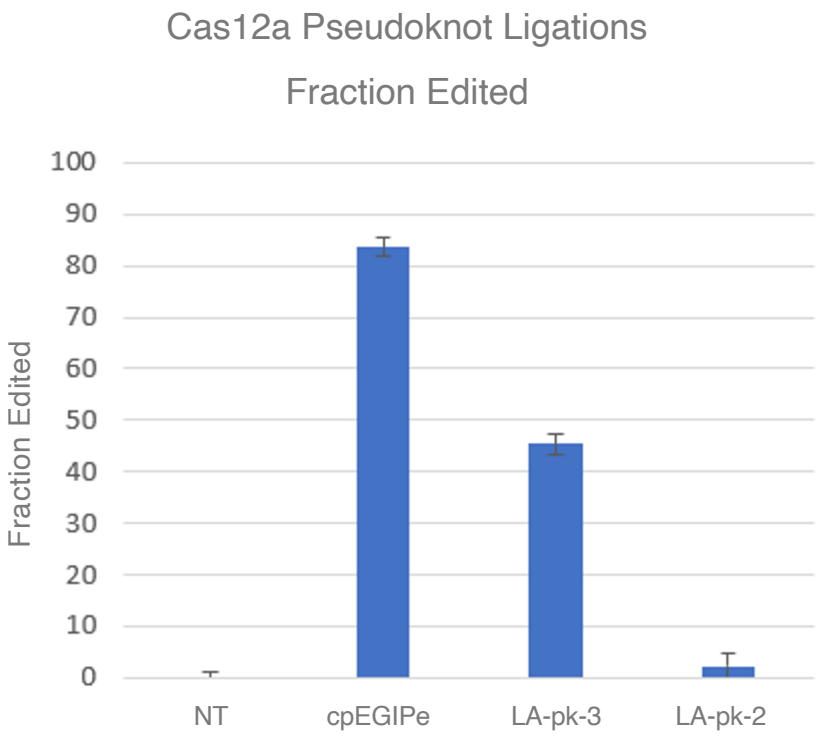


Figure 5.11. *Cell-based editing activity for Cas12a crRNA with modified base pairing regions.*

5.4 Conclusion

Our preliminary results provide useful insights into the structure of the crRNA 5' pseudoknot. Structural studies performed on a variety of modified and unmodified pseudoknots confirm the robustness and stability of their structure. Our studies also provide guidance as to which modifications may be introduced into the pseudoknot without affecting activity. The role of an A-form helical structure and presence of critical 2'-OH in the structure appear to be important for both Cas9 and Cas12 systems. Our studies provide guidance for further exploring CRISPR-Cas systems. For example, the finding that DNA is tolerated in the 5'-handle (pseudoknot) structure provides an opportunity to test DNA mimics, such as ANA and 2'-F-ANA, in the base pairing region of the pseudoknot, and beyond, while retaining the RNA residues required for optimal activity.

5.5 Experimental

5.5.1 Oligonucleotide Synthesis

Oligonucleotide chemical syntheses were carried out using an ABI 3400 DNA synthesizer (Applied Biosystems) on a Unylinker CPG (ChemGenes) solid support at a 1 μmol scale. Conventional 2′-tert-butyl-dimethylsilyl (TBDMS) ribonucleoside, 2′-fluoro-ribonucleoside (2′-FRNA) phosphoramidites were used, in conjunction with 3'-TBDMS-2'-CE (2'-5' Linked RNA) and 2'-arabinoside phosphoramidites (ChemGenes). Phosphoramidites were dissolved in MeCN (0.10 M for DNA, 0.15 M else) and activated with 5-ethylthio-1H-tetrazole (0.25 M in MeCN). Capping of failed couplings was carried out by the simultaneous delivery of acetic anhydride in pyridine/THF and N-methylimidazole (16 % in THF) and contacting the solid support for 6 seconds. Oxidation of the phosphite triester intermediates was affected with 0.1 M iodine in pyridine/H_2O/THF (20 seconds). A solution of 3% trichloroacetic acid in THF, delivered over 95 sec, was used to deprotect DMTr groups.

Deprotection and cleavage of oligonucleotides from the solid support was achieved by treatment with 1 mL of cold 29% aqueous ammonia/ethanol (3:1, v/v) for 16 h at 65˚C. The samples were centrifuged, and the supernatant was transferred to a clean 1.5 mL microcentrifuge tube and vented for 30 min, chilled on dry ice, and evaporated to dryness. Removal of the 2′-TBDMS protecting groups for the RNA-containing oligonucleotides was achieved by treatment with a 300 μL solution of NMP/Et_3N/TREAT-HF (3:4:6, v/v) for 90 min at 65˚C, followed by quenching with 3M NaOAc buffer (50 μL; pH 5.5) and precipitation of the crude oligonucleotide from cold butanol (1 mL, -20˚C). Samples were chilled on dry ice for 30 min and then centrifuged. After removing the supernatant, the remaining pellet was evaporated to dryness, taken up in autoclaved milliQ water (1 mL), and filtered.

Crude oligonucleotides were purified by ion exchange (IE) HPLC. Oligonucleotides were purified on a Waters 1525 instrument using a Protein-Pak DEAE 5PW anion exchange column (21.5 mm × 150 mm) as described in Chapter 4. Following collection of the desired peaks, excess $LiClO_4$ salts were removed using Gel Pak 2.5 size

exclusion columns (Glen Research). Purified oligonucleotides were characterized by electrospray ionization-mass spectrometry and quantitated by UV spectroscopy. Extinction coefficients were determined using the IDT OligoAnalyzer tool. Extinction coefficients for RNA were used for oligonucleotides containing 2'-5'-RNA and 2'-ANA modifications.

5.5.2 Splint Ligation

crRNA guide sequence (200 pmol), synthesized with a 5' phosphate, was ligated to 200 pmol of 5'-handle RNA (synthesized with a 3'-OH) using T4 DNA ligase[33]. For DNA ligase, 220 pmol splint DNA, complementary to both guide and 5'-handle RNA, was annealed to 200 pmol of each crRNA portion in RNA resuspension buffer (5 mM Tris, pH 7.4, 0.5 mM EDTA) at 65°C, then slow-cooled to room temperature. Ligation was performed using 1 μL of concentrated DNA ligase (30 unit/μL) (Thermo Scientific, EL0013), 1x ligation buffer (400 mM Tris-HCL, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP), and 0.5 μL of SUPERase-In (Invitrogen) in a final volume of 100 μL. The reaction was incubated for 90 min at 37°C and stopped by adding 1 μL of 0.5 M EDTA pH 8 and phenol-chloroform extracted. Ligation products were visualized by resolving on 15% denaturing polyacrylamide gels and staining with methylene blue. Full-length ligation products were gel-purified by crush-and-soak elution, phenol-chloroform extracted, ethanol precipitated, and quantified by measuring absorbance at 260 nm and calculated extinction coefficients using nearest neighbor approximations with Beer's Law. Electrospray ionization-mass spectrometry was used to confirm the mass of a few full length, ligated crRNAs to ensure ligation and gel-purification was proceeding properly.

5.5.3 *As*Cas12a Enzyme Expression and Purification

Plasmid encoding of an *As*Cas12a was obtained from Addgene (79007). *As*Cas12a proteins were prepared similarly to that previously described for *Sp*Cas9[34]. Briefly, protein expression was induced in Rosetta (DE3) cells grown in Luria-Bertani (LB) broth with 0.2 mM isopropyl thiogalactopyranoside (IPTG) at 18°C for 16 h. Cell pellets were resuspended in 6 mL chilled binding buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM PMSF, 5 mM imidazole) per 0.5 L of culture. Resuspended cells were sonicated and

clarified by centrifugation. For the 0.5 L of culture, 5 mL His-Pur Cobalt-CMA resin (Thermo Scientific) was equilibrated with binding buffer and the supernatant added to the equilibrated resin and incubated at 4°C for 1 h with inversion to mix every 15 min. The column was washed sequentially with at least 10 bed volumes of increasing concentrations of NaCl in wash buffer (20 mM Tris-HCl, pH 8.0, 10 mM imidazole, 0.25/0.5/0.75/1.0 M NaCl). Protein was eluted with 130 mM imidazole buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 200 mM imidazole). Purified *As*Cas12a enzyme was concentrated and buffer exchanged using centrifugal concentrators (Sartorius, 30,000 MWCO) into 2x storage buffer (40 mM Tris, pH 7.5, 300 mM KCl, 1 mM EDTA, and 2 mM DTT). Then one volume of glycerol added to obtain a final of 50% glycerol. Protein stocks were then stored at -80°C. Concentration of *As*Cas12a was determined by UV absorbance at 280 nm using a calculated extinction coefficient and Beer's law.

5.5.4 Cell-Based Editing Measured by Flow Cytometry

HEK293T cells expressing EGFP and *As*Cas12a were grown in Dulbecco's modified eagle's medium (DMEM) with 1x non-essential amino acids (NEAA), 5% cosmic calf serum (CCS) and 2.5% fetal bovine serum (FBS) without antibiotics. Cells were reverse transfected (40,000 cells) in six experimental replicates in 96-well plates with 20 pmols of crRNA and 0.3 µL RNAiMAX lipid (Invitrogen) in a final reaction of 200 µL of OptiMEM. After 8 h, one volume of media containing 5% FBS and 5% CCS was added to cells and further incubated overnight. Media was then replaced with full media and cells grown for an additional 4 days. For flow cytometry, cells were washed with 200 µL PBS and trypsinized by adding 70 µL of trypsin-EDTA solution. 100 µL of media was added to the cell. The cells were spun for 5 min at 300 x g at room temperature. Cells were washed again with 200 µL PBS, resuspended in 200 µL PBS and counted in an Attune flow cytometer. EGFP was detected using the blue laser (BL1 channel). At least 20,000 events were collected and analyzed by Attune software. The cells were gated based on forward and side scattering (FSC-A/SSC-A) to remove cell debris, gated to select single cells, and gated to select EGFP positive cells. The quadrant gate was established using the signal from non-EGFP expressing control cells. Untreated HEK293T cells expressing EGFP and *As*Cas12a contained ~5% non-fluorescent cells. The average from six replicates was used

for background subtraction to determine the extent of cell-based editing after treatment. Two crRNAs with different guide sequences, cpEGIP and cpEGIPe, were initially screened. cpEGIPe provided substantial editing activity (75-80%) and was chosen for subsequent modification and editing experiments in cells.

5.5.5 UV Thermal melting studies

UV thermal denaturation data were obtained on a Varian CARY 100 UV-visible spectrophotometer equipped with a Peltier temperature controller. The amount of oligo used per strand was 0.4 OD, determined by quantitation on nanodrop. Samples were dissolved in 1 mL of 10 mM NaP_i and 2mm $MgCl_2$ solution. Sample collections were started at 95 °C and lowered to 5 ˚C by 0.5°C/min. Absorbance was measured at 260 nm every minute. Nitrogen was used for samples that went below 15°C to prevent condensation. The thermal melting temperature (T_m) were determined by calculating the midpoint of dissociation curves by first derivatives (**Figure 5.12**).

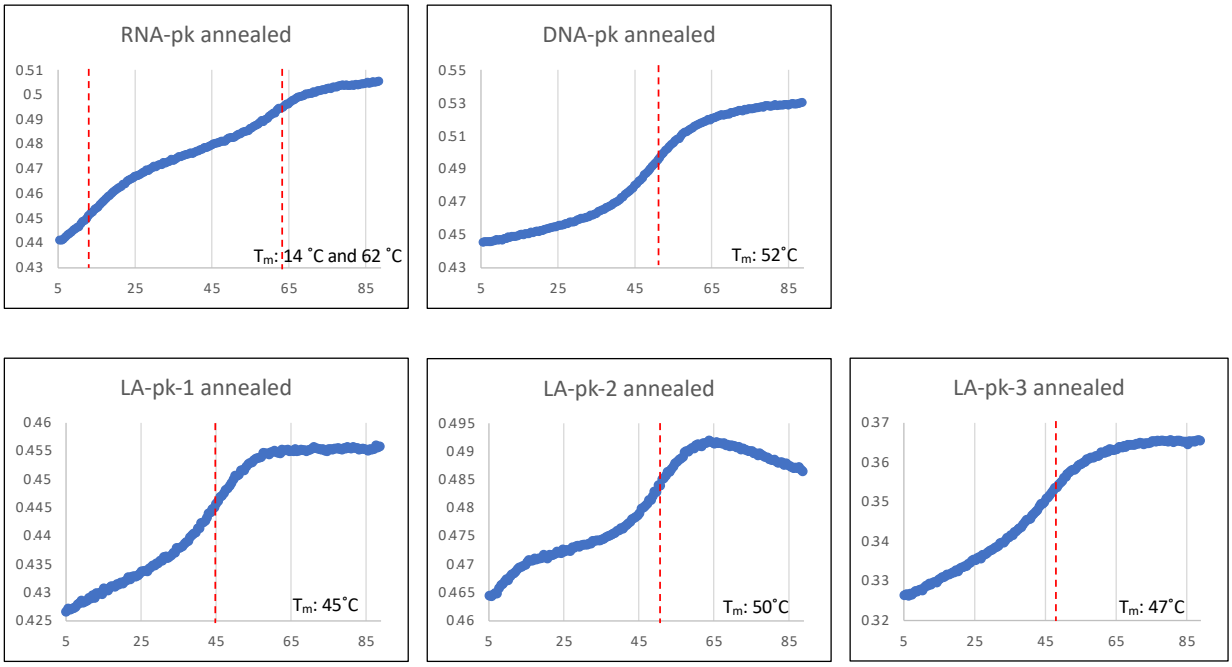


Figure 5.12. *UV Thermal melting curves of the Cas12a pseudoknot structural series.*

5.5.6 Circular dichroism studies

Circular Dichroism (CD) studies were performed on an Applied Photophysics CD (Chirascan) spectrophotometer using a 1 mm path length cuvette. Temperature was maintained, at either 22 ˚C, 37˚C or 75 ˚C, using an external temperature control unit. Pseudoknot concentration of 50 μM in 1x MOPS buffer, 50 mM KCl, 2 mM $MgCl_2$, 20 mM NaP_i, at pH 7.3. Spectra were recorded from 200-300 nm at a scan rate of 100 nm min-1 and a response time of 2.0 s with three acquisitions recorded for each spectrum. The background was automatically subtracted using Chirascan software. All three acquisitions were averaged and smoothed using the Chirascan software.

5.6 References

1. 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.
2. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F., Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013,** *339* (6121), 819-23.
3. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M., RNA-guided human genome engineering via Cas9. *Science* **2013,** *339* (6121), 823-6.
4. Rath, D.; Amlinger, L.; Hoekzema, M.; Devulapally, P. R.; Lundgren, M., Efficient programmable gene silencing by Cascade. *Nucleic Acids Res* **2015,** *43* (1), 237-46.
5. O'Connell, M. R.; Oakes, B. L.; Sternberg, S. H.; East-Seletsky, A.; Kaplan, M.; Doudna, J. A., Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **2014,** *516* (7530), 263-6.
6. Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; Qi, L. S.; Kampmann, M.; Weissman, J. S., Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **2014,** *159* (3), 647-61.
7. Larson, M. H.; Gilbert, L. A.; Wang, X.; Lim, W. A.; Weissman, J. S.; Qi, L. S., CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols* **2013,** *8* (11), 2180-2196.
8. Chylinski, K.; Makarova, K. S.; Charpentier, E.; Koonin, E. V., Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **2014,** *42* (10), 6091-105.
9. Swarts, D. C.; Jinek, M., Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *WIREs RNA* **2018,** *9* (5), e1481.

10. Chen, J. S.; Doudna, J. A., The chemistry of Cas9 and its CRISPR colleagues. *Nature Reviews Chemistry* **2017,** *1* (10), 0078.

11. O'Reilly, D.; Kartje, Z. J.; Ageely, E. A.; Malek-Adamian, E.; Habibian, M.; Schofield, A.; Barkau, C. L.; Rohilla, K. J.; DeRossett, L. B.; Weigle, A. T.; Damha, M. J.; Gagnon, K. T., Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity. *Nucleic Acids Res* **2019,** *47* (2), 546-558.

12. Kartje, Z. J.; Barkau, C. L.; Rohilla, K. J.; Ageely, E. A.; Gagnon, K. T., Chimeric Guides Probe and Enhance Cas9 Biochemical Activity. *Biochemistry* **2018,** *57* (21), 3027-3031.

13. Rueda, F. O.; Bista, M.; Newton, M. D.; Goeppert, A. U.; Cuomo, M. E.; Gordon, E.; Kröner, F.; Read, J. A.; Wrigley, J. D.; Rueda, D.; Taylor, B. J. M., Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat Commun* **2017,** *8* (1), 1610.

14. Mir, A.; Alterman, J. F.; Hassler, M. R.; Debacker, A. J.; Hudgens, E.; Echeverria, D.; Brodsky, M. H.; Khvorova, A.; Watts, J. K.; Sontheimer, E. J., Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. *Nat Commun* **2018,** *9* (1), 2641.

15. Prakash, T. P.; Kraynack, B.; Baker, B. F.; Swayze, E. E.; Bhat, B., RNA interference by 2',5'-linked nucleic acid duplexes in mammalian cells. *Bioorg Med Chem Lett* **2006,** *16* (12), 3238-40.

16. Habibian, M.; Harikrishna, S.; Fakhoury, J.; Barton, M.; Ageely, E. A.; Cencic, R.; Fakih, H. H.; Katolik, A.; Takahashi, M.; Rossi, J.; Pelletier, J.; Gagnon, K. T.; Pradeepkumar, P. I.; Damha, M. J., Effect of 2'-5'/3'-5' phosphodiester linkage heterogeneity on RNA interference. *Nucleic Acids Res* **2020,** *48* (9), 4643-4657.

17. Shen, F.; Luo, Z.; Liu, H.; Wang, R.; Zhang, S.; Gan, J.; Sheng, J., Structural insights into RNA duplexes with multiple 2′-5′-linkages. *Nucleic Acids Res* **2017,** *45* (6), 3537-3546.

18. Sheng, J.; Li, L.; Engelhart, A. E.; Gan, J.; Wang, J.; Szostak, J. W., Structural insights into the effects of 2'-5' linkages on the RNA duplex. *Proc Natl Acad Sci U S A* **2014,** *111* (8), 3050-5.

19. Giannaris, P. A.; Damha, M. J., Oligoribonucleotides containing 2',5'-phosphodiester linkages exhibit binding selectivity for 3',5'-RNA over 3',5'-ssDNA. *Nucleic Acids Res* **1993,** *21* (20), 4742-9.

20. Wasner, M.; Arion, D.; Borkow, G.; Noronha, A.; Uddin, A. H.; Parniak, M. A.; Damha, M. J., Physicochemical and biochemical properties of 2',5'-linked RNA and 2',5'-RNA:3',5'-RNA "hybrid" duplexes. *Biochemistry* **1998,** *37* (20), 7478-86.

21. Noronha, A. M.; Wilds, C. J.; Lok, C. N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J., Synthesis and biophysical properties of arabinonucleic acids (ANA): circular dichroic spectra, melting temperatures, and ribonuclease H susceptibility of ANA.RNA hybrid duplexes. *Biochemistry* **2000,** *39* (24), 7050-62.

22. Kalota, A.; Karabon, L.; Swider, C. R.; Viazovkina, E.; Elzagheid, M.; Damha, M. J.; Gewirtz, A. M., 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (2'F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing. *Nucleic Acids Res* **2006,** *34* (2), 451-61.

23. Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M., 2'-Fluoroarabino- and arabinonucleic acid show different conformations, resulting in deviating RNA affinities and processing of their heteroduplexes with RNA by RNase H. *Biochemistry* **2006,** *45* (13), 4141-52.

24. Denisov, A. Y.; Noronha, A. M.; Wilds, C. J.; Trempe, J. F.; Pon, R. T.; Gehring, K.; Damha, M. J., Solution structure of an arabinonucleic acid (ANA)/RNA duplex in a chimeric hairpin: comparison with 2'-fluoro-ANA/RNA and DNA/RNA hybrids. *Nucleic Acids Res* **2001,** *29* (21), 4284-93.

25. Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; González, C., The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res* **2012,** *40* (18), 9329-39.

26. Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J., Differential stability of 2'F-ANA*RNA and ANA*RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res* **2010,** *38* (7), 2498-511.

27. Yamano, T.; Nishimasu, H.; Zetsche, B.; Hirano, H.; Slaymaker, I. M.; Li, Y.; Fedorova, I.; Nakane, T.; Makarova, K. S.; Koonin, E. V.; Ishitani, R.; Zhang, F.; Nureki, O., Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **2016,** *165* (4), 949-62.

28. Safari, F.; Zare, K.; Negahdaripour, M.; Barekati-Mowahed, M.; Ghasemi, Y., CRISPR Cpf1 proteins: structure, function and implications for genome editing. *Cell Biosci* **2019,** *9*, 36.

29. Saenger, W., *Principles of Nucleic Acid Structure*. Springer-Verlag: New York, 1984; p 556.

30. Arnstein, H., Data for biochemical research (Third Edition) : by R.M.C. Dawson, D.C. Elliott, W.H. Elliott and K.M. Jones Oxford University Press; Oxford, 1986 xii + 580 pages. £35.00. *FEBS Letters* **1988,** *234*, 506-506.

31. Roberts, R.; Crothers, D., Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **1992,** *258* (5087), 1463-1466.

32. Kypr, J.; Kejnovská, I.; Renciuk, D.; Vorlícková, M., Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res* **2009,** *37* (6), 1713-25.

33. Kershaw, C. J.; O'Keefe, R. T., Splint ligation of RNA with T4 DNA ligase. *Methods Mol Biol* **2012,** *941*, 257-69.

34. Anders, C.; Jinek, M., In vitro enzymology of Cas9. *Methods Enzymol* **2014,** *546*, 1-20.

- page intentionally blank -

Chapter 6

Contributions to Knowledge and Outlook

6.1 Contributions to Knowledge and Outlook

6.1.1 Uncovering molecular mechanism of a repair enzyme using step-arrest mutants of Tpt 1 and substrate mimics of RNA (Chapter 2)

The studies described in this chapter relates to the molecular mechanism and structure of Tpt1, an essential component of the fungal and plant tRNA splicing machinery that catalyzes transfer of an internal RNA 2′-phosphate to NAD+ yielding RNA 2′-OH and ADP-ribose 1′,2′-cyclic phosphate (2′-P-ADPR) products. Specifically, we studied the Tpt1 ortholog from the bacterium *Runella slithyformis*. This enzyme performs 470 rounds of RNA dephosphorylation in 30 minutes, under conditions of RNA substrate excess, without detectable accumulation of an intermediate. Under conditions of enzyme excess, Tpt1 converted the RNA substrate to product in just a few seconds. After 3 sec (the earliest time we could sample manually), the 2′-OH product and 2′-P-ADPR intermediate comprised 65% and 4.6% of the total RNA, respectively, suggesting that step 1 is likely to be rate-limiting. The intermediacy of the 2′-P-ADPR species was affirmed by the ability of wild-type RslTpt1 to effect rapid and complete conversion of the gel-purified 6-mer 2′-P-ADPR RNA to the 2′-OH RNA product in the absence of NAD^+.

We observed that the biological activity of this enzyme relies on four conserved amino acids in the active site (Arg-His-Arg-Arg) and pinpointed one of the arginines as specifically essential for step 2 of the Tpt1 pathway. Mutating this arginine slowed the step 1 rate by only threefold while slowing the step 2 rate by a factor of 214, thus resulting in the transient accumulation of very high levels of the ADP-ribosylated RNA intermediate.

Valuable insights into substrate recognition and the mechanism of the transesterification step emerged from a crystal structure of Tpt1 in a product-mimetic complex with ADP-ribose-1″-PO_4 in the NAD^+ site and pAp in the RNA site. This structure, which mimics the step 2 product complex of the Tpt1 reaction, yielded keen insights into how Tpt1 recognizes a 2′-PO_4 RNA splice junction. It also provided the first evidence that a bacterium has an endogeneous phosphorylated substrate with which a Tpt1 enzyme can react. It appears that we have only begun to scratch the surface, and that other functions and molecular mechanisms may yet be discovered. By continuing to design new 2′-PO_4 RNA analogues, we may identify additional functions and the evolutionary significance of this enzyme.

6.1.2 Identifying substrate RNA analogs that trap intermediates of the Tpt1 as dead-end step 1 products - Towards the rational design of Tpt1 inhibitors (Chapter 3)

This chapter highlights the distinctive effects of RNA and NAD^+ mimetics on Tpt1 activity. Whereas replacement of each of the ribose sugars flanking the internal 2'-PO_4 with a deoxynucleotide did not affect the efficiency of 2'-PO_4 removal by *Runella* Tpt1, its replacement with 2'-fluoroarabinose reduced Tpt1 specific activity, albeit without accumulation of the 2'- ADP-ribosylated RNA intermediate. By contrast, replacing the ribose 2'-PO_4 nucleotide with arabinose 2'-PO_4 at the branchpoint selectively and severely reduced the rate of the transesterification step with buildup of the reaction intermediate. Replacing the ribose of NAD^+ with 2'-fluoroarabinose (which eliminates the step 2 ADP-ribose O2" nucleophile) results in trapping of RNA-2'-phospho-(ADP-fluoroarabinose) as a dead-end step 1 product. We also found that Tpt1 orthologs differ in their ability to use ara-2"F-NAD^+.

We designed and attempted to synthesize a novel RNA with a 2'- phosphonate moiety (2'-CH_2PO_3) as a potential selective inhibitor of step 2. Specifically, we outlined a synthetic scheme for introducing the CH_2PO_3 moiety at the C2' position of adenosine. The most challenging step was the formation of the C-C bond at the 2' position, which we carried out using an organolithiated species.

The identification of substrate analogs that trap the ADP-ribosylated RNA intermediate has implications for the development of Tpt1 “poisons” as anti-fungals agents. The 2'-phosphate analogues designed and synthesized described here (2'-NH-PO_3, 2'-CH_2-PO_3), will facilitate the co-crystallization of Tpt1 enzymes with RNA reaction intermediates and possibly with transition-state mimetics. The ultimate goal of this endeavor is to guide the synthesis of other modified RNAs that could be applied in drug discovery.

6.1.3 CRISPR-Cas Gene Editing with RNA mimetics (Chapter 4)

We explored the impact of ribose chemical modification to the 5' handle of *Acidominococcus species* Cas12a (AsCas12a) CRISPR RNA (crRNA). The 5' handle is a non-canonical pseudoknot structure that mediates conserved binding to Cas12a. Significant

2'-deoxy and full 2'-fluoro modification were well-tolerated in vitro. However, in-cell gene editing required retention of native RNA residues, especially at positions predicted to involve 2'-hydroxyl polar contacts. Replacing these critical positions with ribose-modified or replacement nucleotides designed to partially maintain hydrogen-bonding, including arabinonucleic acid, 2'- fluoro, and 2'-amino, revealed chemical sensitivity and compatibility. Modified 5' pseudoknots with as little as six out of nineteen native 2'-hydroxyl residues supported robust gene editing. The more centrally positioned critical 2'-OH residues, especially at -14, -10, and -6 positions, were more tolerant to substitution while terminal residues (-18, -17, and -1) were less tolerant. We also observed differential cis versus trans activity profiles that were due to low cis but high trans catalytic turnover. These results indicate that the 5' pseudoknot structure can tolerate significant modification and may influence Cas12a activity. These results advance efforts toward chemical modification of AsCas12a guides for therapeutic editing.

6.1.4 Evaluation of regio/stereoisomeric RNAs for CRISPR Cas12a gene editing (Chapter 5)

So far our findings indicate that nucleotides resistant to modification are those with predicted 2'-OH polar contacts within the Cas12a-crRNA complex. To investigate this further, attempts were made to modify the 5' pseudoknots of Cas12a crRNAs with 2'-arabinonucleotides (ANA) and 2',5'-linked ribonucleotides (2',5'-RNA). The general conclusion from these studies was that these modifications affected cleavage activity depending on the position of the modification, with some being relatively well-tolerated. However, the level of activity never surpassed that of the native crRNA or a cr-2'FRNA designed to potentially retain hydrogen bonding with AsCas12a protein.

RNA-DNA chimeric pseudoknots generated little or no editing activity. Several crRNA were inactive despite preserving RNA at the putative critical 2'-OH positions. Some provided modest editing, perhaps due to better conservation of A-form structure as assessed by circular dichroism studies. Thus, as had been observed with Cas9 crRNAs, retaining 2'-OH at potentially critical positions appeared to be necessary but not sufficient to provide gene editing activity. These results highlight the more complex nature of editing in cells

and support the complementary role of both A-form helical structure and 2'-OH contacts within the Cas12a crRNA 5' pseudoknot.

Overall, the results provided in chapters 4 and 5 indicate that the 5' pseudoknot of Cas12a crRNA can tolerate heavy modification when structural and chemical compatibility are used to guide design. Rules for successful chemical modification of the 5' pseudoknot should accelerate therapeutic development and may be valuable for CRISPR-Cas12a diagnostic applications.

6.2 List of Publications

1. Ageely, E.A., Chilamkurthy, R., Jana, S., Abdullahu, L., O'Reilly, D., Jensik, P.J., Damha, M.J., Gagnon, K.T. Gene Editing with Ribose-Modified CRISPR-Cas12a 5' Pseudoknots. *Nat. Commun*. **2021,** *accepted,* MS# NCOMMS-20-43072.
2. Dantuluri, S., Schwer, S., Abdullahu, L., Damha, M.J., Shuman, S. Activity and substrate specificity of Candida, Aspergillus, and Coccidioides Tpt1: essential tRNA splicing enzymes and potential antifungal targets. *RNA* **2021**, doi:10.1261/rna.078660.120.
3. Dantuluri, S., Abdullahu, L., Munir, A., Katolik, A., Damha M.J., Shuman S. Substrate analogs that trap the 2'- phospho-ADP-ribosylated RNA intermediate of the Tpt1 (tRNA 2'-phosphotransferase) reaction pathway. *RNA* **2020**, *26* (4), 373-381. doi:10.1261/rna.074377.119.
4. Munir, A., Abdullahu, L., Banerjee, A., Damha, M.J., Shuman, S. NAD+-dependent RNA terminal 2' and 3' phosphomonoesterase activity of a sub-set of Tpt1 enzymes. *RNA* **2019,** *25* (7), 783-792. doi: 10.1261/rna.071142.119
5. Banerjee, A., Munir, A., Abdullahu, L., Damha, M.J., Goldgur, Y., Shuman, S. Structure of tRNA splicing enzyme Tpt1 illuminates the mechanism of RNA 2'-PO4 recognition and splice junction ADP-ribosylation. *Nat. Commun.* **2019,** *10* (1), 218. doi: 10.1038/s41467-018-08211-9
6. Munir, A., Abdullahu, L., Damha, M. J., Shuman, S. Two-step mechanism and step-arrest mutants of Runella slithyformis NAD+-dependent tRNA 2′-phosphotransferase Tpt1. *RNA* **2018,** *24* (9), 1144-1157. doi: 10.1261/rna.067165.118.

6.3 List of Conference Presentations

1. "Small molecule-like oligonucleotides as novel antifungal reagents against tRNA 2' phosphotransferase (Tpt1)," Abdullahu, L., Damha, M.J. Annual PROMOTE Retreat, May 2021, Virtual, Short Talk.

2. “Small molecule-like oligonucleotides as novel antifungal reagents against tRNA 2' phosphotransferase (Tpt1),” Abdullahu, L., Damha, M.J. Oligonucleotide Therapeutics Society Annual Meeting, September 2020, Virtual, Short Talk.
3. “Synthesis of 2'-modified RNAs as substrates towards the inhibition of tRNA 2'-phosphotransferase (Tpt1),” Abdullahu, L., Munir, A., Shuman, S., Damha, M.J. Oligonucleotide Therapeutics Society Annual Meeting, October 2019, Munich, Germany. *Poster Award Recipient*
4. “Synthesis of 2'-modified RNAs as substrates for tRNA 2'-phosphotransferase (Tpt1),” Abdullahu, L., Munir, A., Shuman, S., Damha, M.J. Canadian Chemistry Conference and Exhibition, June 2019, Quebec City, Quebec, Canada.

- page intentionally blank -

- page intentionally blank -

Appendix 1

NAD^+-dependent RNA terminal 2' and 3' phosphomonoesterase activity of a sub-set of Tpt1 enzymes

Munir, A., Abdullahu, L., Banerjee, A.; Damha, M. J.; Shuman, S., NAD(+)-dependent RNA terminal 2' and 3' phosphomonoesterase activity of a subset of Tpt1 enzymes. *RNA* 2019, *25* (7), 783-792.

A.1 Introduction

The enzyme Tpt1 is an essential agent of fungal and plant tRNA splicing that removes the 2'-PO_4 at the splice junction generated by fungal and plant tRNA ligases[1]. Tpt1 catalyzes a unique two-step reaction whereby the internal RNA 2'-PO_4 attacks NAD^+ to form an RNA-2'-phospho-ADP-ribosyl intermediate that undergoes transesterification to yield 2'-OH RNA and ADP-ribose-1",2"-cyclic phosphate products[2-6]. Tpt1 homologs are distributed widely in archaeal, bacterial, and metazoan taxa[7]. Many of the bacterial species with Tpt1 homologs have no known intron-containing tRNAs and/or no known pathways to generate RNAs with internal 2'-PO_4 modifications. In metazoa and archaea, the mechanism of tRNA exon ligation is entirely different from that of fungi and plants and does not result in a junction 2'-PO_4[8]. Nevertheless, it is clear by the criteria of genetic complementation of yeast *tpt1*Δ and direct biochemical assay of recombinant Tpt1 proteins that Tpt1 enzymes from widely divergent organisms all possess NAD^+-dependent RNA 2'-phosphotransferase activity[6, 7, 9, 10].

A longstanding mystery surrounds the role that Tpt1 enzymes might play in taxa that have no fungal/plant-type RNA ligases. One possibility is that bacteria, archaea, or metazoa do have a capacity to install internal RNA 2'-PO_4 groups, conceivably during RNA repair pathways that are as yet uncharacterized, and that such RNAs are potential substrates for Tpt1. An alternative scenario is that Tpt1 enzymes can catalyze reactions other than internal RNA 2'-PO_4 removal, via their unique NAD^+-dependent transferase mechanism.

Support for the latter scheme emerged from the recent findings that that a sub-set of Tpt1 enzymes can transfer ADP-ribose from NAD^+ to a 5'-monophosphate end of RNA or DNA to install a 5'-phospho-ADP-ribose cap structure[10]. *Aeropyrum pernix* Tpt1 (ApeTpt1) was particularly adept in this respect. Two other archaeal Tpt1s (from *Pyrococcus horikoshii* and *Archaeoglobus fulgidis*), one bacterial Tpt1 (from *Clostridium thermocellum*) and one fungal Tpt1 (from *Chaetomium thermophilum*) also exhibited 5'-phospho-ADP-ribose capping activity, albeit less effectively than ApeTpt1. By contrast, human Tpt1 and *Runella slithyformis* Tpt1 were inactive in 5'-phospho-ADP-ribose capping. The one-step synthesis of a 5'-phospho-ADP-ribosylated cap structure by a sub-set of Tpt1 enzymes (with no subsequent 5'-phosphotransferase step) extended the

repertoire of the Tpt1 family and raised the prospect that Tpt1 enzymes might perform additional phospho-ADP-ribosylation reactions, with or without subsequent removal of the phosphate moiety.

The intrigue around Tpt1 was heightened by the 1.4 Å X-ray structure of *Clostridium thermocellum* Tpt1 (CthTpt1), produced as a recombinant protein in *Escherichia coli* and providentially crystallized with ADP-ribose-1"-phosphate in the NAD^+ site[11]. To our knowledge, there had been no prior description of ADP-ribose-1"-phosphate as a metabolite in *E. coli*. The presence of this molecule in the NAD^+ site of CthTpt1 crystals is most plausibly explained by its genesis during CthTpt1 expression *in vivo*, e.g., via CthTpt1-catalyzed ADP-ribosylation of an endogenous bacterial phospho-substrate, followed by phosphoryl transfer to yield ADP-ribose-1",2"-cyclic phosphate, which was hydrolyzed *in situ* to ADP-ribose-1"-phosphate. We construed this to be the first evidence that a bacterium has an endogenous substrate with which a Tpt1 enzyme can react. If the endogenous substrate for CthTpt1 is an RNA, it is either the case that: (i) *E. coli* can form an RNA with an internal 2'-PO_4 modification, via an as yet undefined pathway; or (ii) CthTpt1 can catalyze removal of an RNA terminal 2'-monophosphate or 3'-monophosphate moiety.

In the present study, we interrogate the reaction of Tpt1 enzymes with 5' ^{32}P-labeled $pRNA^{2'}p$ and $pRNA_{3'}p$ substrates. We report that CthTpt1 and ApeTpt1 catalyze the NAD^+-dependent conversion of 2'-PO_4 and 3'-PO_4 ends to 2'-OH and 3'-OH ends, respectively, whereas *Chaetomium* Tpt1 selectively coverts a 2'-PO_4 ends to a 2'-OH. By contrast, *Runella slithyformis* Tpt1 and human Tpt1 are not adept at either terminal phosphate removal reaction.

A.2 Results

A.2.1 NAD^+-dependent removal of RNA terminal 2' and 3' monophosphates by *Clostridium thermocellulm* Tpt1.

Tpt1 proteins from diverse taxa displayed varying abilities to perform the recently discovered nucleic acid 5'-phospho-ADP-ribose capping reaction[10]. In order to test if the Tpt1 repertoire might embrace phosphoryl transfer reactions at the opposite end of the polynucleotide, we synthesized a pair of 10-mer RNA oligonucleotides with either a terminal 2'-PO_4 or 3'-PO_4 moiety. The 10-mers were 5' ^{32}P-labeled and then gel-purified for use in assaying terminal phosphate removal. Control reactions of the labeled $pRNA^{2'}p$ and $pRNA_{3'}p$ substrates with bacteriophage T4 polynucleotide 2',3' phosphatase[12] resulted in quantitative removal of the 2'-PO_4 or 3'-PO_4 groups to yield 5' ^{32}P-labeled $pRNA^{2'OH}$ and $pRNA_{3'OH}$ products that migrated slower than the input substrate during denaturing PAGE (**Figure A1.1** (A), lanes T4). The instructive findings were that a 30 min reaction of 500 nM CthTpt1 with 200 nM $pRNA^{2'}p$ or $pRNA_{3'}p$ and 1 mM NAD^+ resulted in the conversion of the input radiolabeled RNA into species that comigrated with the $pRNA^{2'OH}$ and $pRNA_{3'OH}$ products of the T4 enzyme-catalyzed 2',3' dephosphorylation reactions (**Figure A1.1** (A)). Removal of the 2'-PO_4 and 3'-PO_4 groups by CthTpt1 depended on NAD^+ (**Figure A1.1** (A)). Additional slowly migrating NAD^+-dependent reaction products were generated by CthTpt1, especially in the case of the $pRNA_{3'}p$ substrate (**Figure A1.1** (A)). These species correspond to terminally phospho-ADP-ribosylated RNAs, as inferred from their susceptibility to post-reaction treatment with alkaline phosphatase (CIP; see below).

As expected, the ^{32}P-labeled $pRNA^{2'}p$ and $pRNA_{3'}p$ substrates were eliminated by CIP, as were the predominant $pRNA^{2'OH}$ and $pRNA_{3'OH}$ products (**Figure A1.1** (B)). By contrast, the most slowly migrating minor products were CIP-resistant, signifying that the 5' ^{32}P-labeled phosphate had been capped with ADP-ribose[10]. Because the mobility of the 5'-ADPR capped RNAs did not change after CIP treatment, we surmise that the original 2'-PO_4 and 3'-PO_4 groups were removed during the reaction with CthTpt1. The prominent labeled species in the $pRNA_{3'}p$ reaction that migrated between $pRNA_{3'OH}$ and ADPR-$pRNA_{OH}$ was effaced by CIP, suggesting that it corresponds to an anticipated pRNAp-ADPR intermediate in the terminal monophosphate removal reaction.

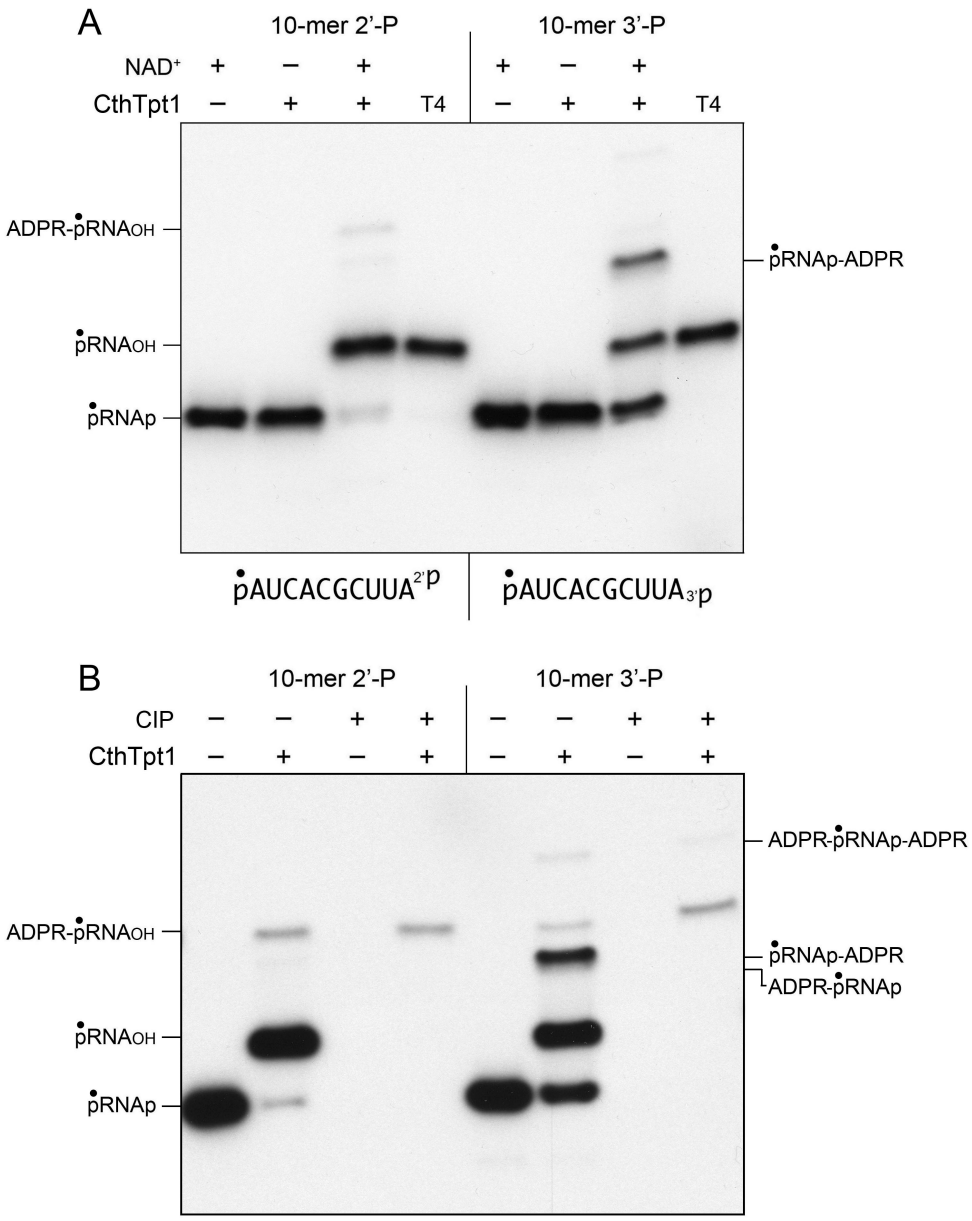


Figure A1.1. NAD$^+$-dependent removal of RNA terminal 2'- and 3'-monophosphates by CthTpt1. *(A) Tpt1 reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 10-mer pRNA$^{2'}$p or pRNA$_{3'}$p substrates (shown at the bottom), 1 mM NAD$^+$ (where indicated by +), and 0.5 µM (5 pmol) CthTpt1 (where indicated by +) were incubated at 37˚C for 30 min. T4 polynucleotide 2',3'-phosphatase reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5′ ^{32}P-labeled RNAs, 1 mM $MgCl_2$, and 0.5 µM (5 pmol) T4 Pnkp were incubated at 37˚C for 30 min. The products were analyzed by urea-PAGE and visualized by autoradiography. (B) Reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD$^+$, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 10-mer pRNA$^{2'}$p or pRNA$_{3'}$p substrates (shown at the bottom), and 0.5 µM (5 pmol) CthTpt1 (where indicated by +) were incubated at 37˚C for 30 min. The reaction mixtures were heated at 65˚C for 5 min and then either treated for 10 min at 37˚C with 10 U of calf intestine alkaline phosphatase (CIP; from NEB) in 1x Cutsmart buffer (50 mM potassium acetate, 20 mM Tris-acetate, pH 7.9, 10 mM magnesium acetate, 100 µg/ml BSA, pH 7.9) or mock-incubated without phosphatase. The reactions were quenched with three volumes of cold 90% formamide, 50 mM EDTA and the products were analyzed by urea-PAGE and visualized by autoradiography.*

The temporal profile of the CthTpt1 reaction with the 5' ^{32}P-labeled pRNA$^{2'}$p and pRNA$_{3'}$p substrates is shown in **Figure A1.2** (A and B). The kinetics of RNA terminal phosphate removal were quantified by plotting the sum of the pRNA$_{OH}$ and ADPR-pRNA$_{OH}$ species (expressed as the percent of total labeled RNA) as a function of reaction time (**Figure A1.2**, OH). The kinetics of 5' capping were evident as the sum of the ADPR-pRNA$_{OH}$ and ADPR-pRNAp species plotted as a function of time (**Figure A1.2**, 5'-ADPR). Also quantified was the pRNAp-ADPR intermediate. The key findings were as follows. First, the rate of 2'-PO$_4$ removal by CthTpt1 (2.94 ± 0.05 percent•min^{-1}) was 2.4-fold faster than the rate of 3'-PO$_4$ removal (1.24 ± 0.02 percent•min^{-1}). Second, the pRNAp-ADPR intermediate accumulated to a significant extent during 3'-PO$_4$ removal (up to 16% of total labeled RNA) and prior to the formation of the 3'-OH end product, but not during 2'-PO$_4$ removal (1.6% of total RNA). This result suggests that transesterification of the ADPR ribose-O2" to the 3'-PO$_4$ of the pRNAp-ADPR intermediate was uniquely rate-limiting for 3'-PO$_4$ removal. Third, the 2' and 3' terminal phosphate removal reactions were much more vigorous than 5'-phospho-ADPR capping reaction, insofar as the extents of 2'-PO$_4$ and 3'-PO$_4$ removal after 45 min were 40-fold and 24-fold greater than the extent of 5'-capping, respectively.

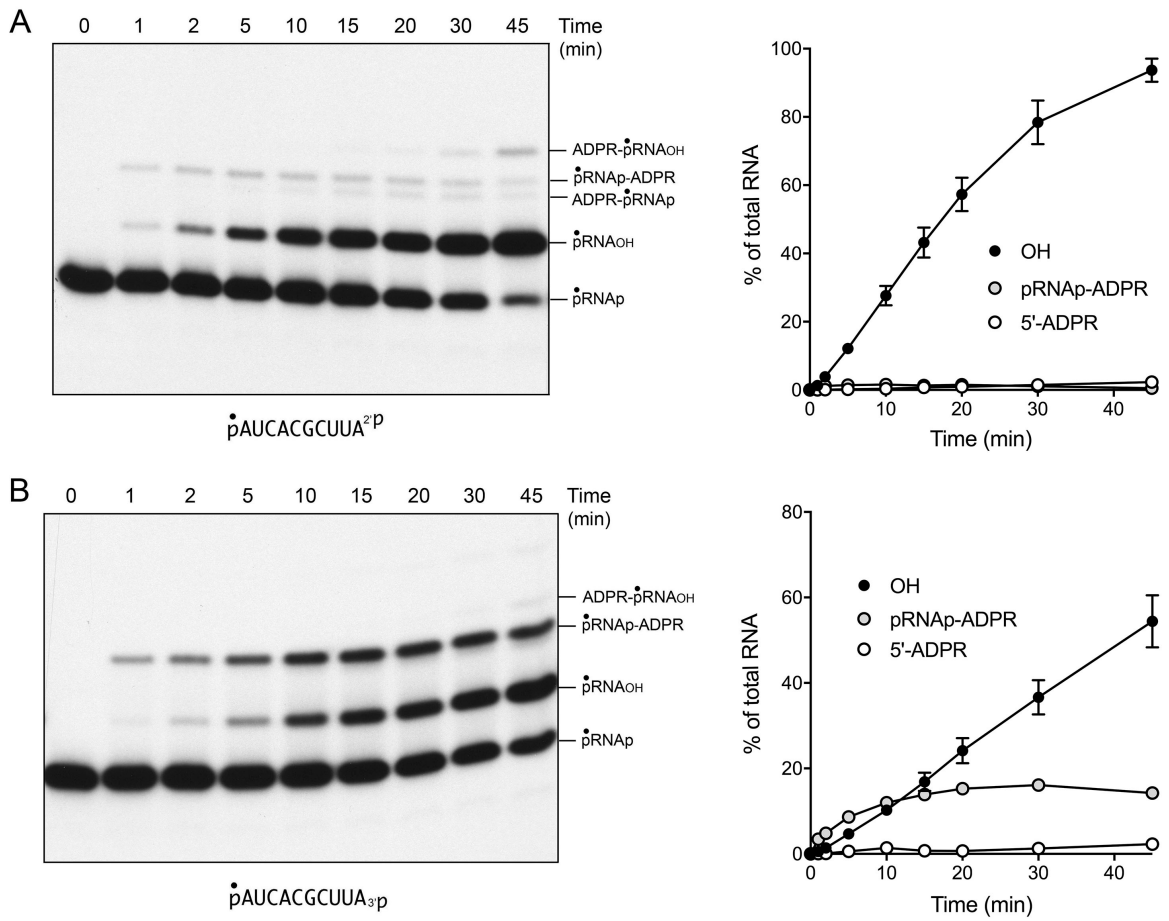


Figure A1.2. Temporal profile of RNA terminal 2'-PO$_4$ and 3'-PO$_4$ removal by CthTpt1. *Reaction mixtures (100 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD$^+$, 0.2 µM (20 pmol) 5' ^{32}P-labeled 10-mer pRNA$^{2'}$p or pRNA$_{3'}$p substrates (shown at the bottom), and 0.5 µM (50 pmol) CthTpt1 were incubated at 37°C. Aliquots (10 µl) were withdrawn at the times specified and quenched immediately with 3 volumes of of cold 90% formamide, 50 mM EDTA. The products were analyzed by urea-PAGE and visualized by autoradiography (A and B, left panels). The identities of the radiolabeled RNAs are indicated on the right. The kinetics of terminal phosphate removal were quantified in the right panels by plotting the sum of the pRNA$_{OH}$ and ADPR-pRNA$_{OH}$ species (expressed as the percent of total labeled RNA) as a function of reaction time (OH). The kinetics of 5' capping were evident as the sum of the ADPR-pRNA$_{OH}$ and ADPR-pRNAp species plotted as a function of time (5'-ADPR). Also quantified was the pRNAp-ADPR intermediate. Each datum in the graphs is the average of three independent time course experiments ±SEM.*

The dependence of product formation during a 30 min reaction on the amount of input CthTpt1 is shown in **Figure A1.3**. The specific activities for terminal phosphate removal, calculated from the slope of the titration curves at sub-saturating enzyme, were 0.19 pmol/pmol for pRNA$^{2'}$p and 0.10 pmol/pmol for pRNA$_{3'}$p. Only the pRNA$_{3'}$p reaction led to the accumulation of a pRNAp-ADPR intermediate at sub-saturating enzyme levels (**Figure A1.3**). Thus, CthTpt1 is acting in a stoichiometric fashion as a terminal phosphatase with the 10-mer substrates and is ~2-fold more effective at a terminal 2'-PO$_4$

versus a terminal 3'-PO_4. The specific activity of CthTpt1 in 5'-phospho-ADPR capping (0.0024 pmol/pmol for $pRNA^{2'}p$ and 0.0021 pmol/pmol for $pRNA_{3'}p$) was 80-fold less than its activity in terminal 2'-PO_4 removal. By contrast, titration of CthTp1 for removal of an internal 2'-PO_4, using a 5' ^{32}P-labeled 6-mer RNA substrate $pCCAA^{2'p}AU_{OH}$ that mimics a tRNA splice junction[6], indicated that it acted catalytically, generating 150 fmol of 2'-OH RNA product per fmol of input CthTpt1 (data not shown).

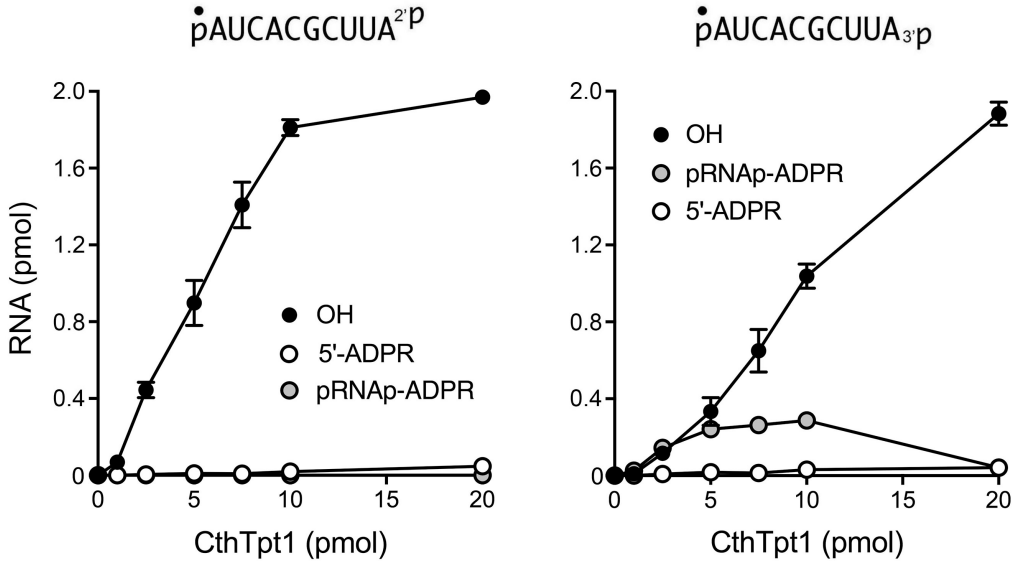


Figure A1.3. Dependence of 2'-PO_4 and 3'-PO_4 removal on CthTpt1 concentration. *(A) Tpt1 reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5' ^{32}P-labeled 10-mer $pRNA^{2'}p$ or $pRNA_{3'}p$ substrates (shown at the top), 1 mM NAD^+, and CthTpt1 as specified were incubated at 37°C for 30 min. The products were analyzed by urea-PAGE. The extents of product formation, quantified as described in Fig. 2, are plotted as a function of input enzyme. Each datum is the average of three independent titration experiments ±SEM.*

A.2.2 RNA terminal 2' and 3' phosphatase activity of *Aeropyrum pernix* Tpt1.

Reaction of 500 nM ApeTpt1 with 200 nM $pRNA^{2'}p$ or $pRNA_{3'}p$ and 1 mM NAD^+ resulted in the conversion of the input radiolabeled RNA into two major products: $pRNA_{OH}$ and ADPR-$pRNA_{OH}$ (**Figure A1.4** (A)). Product formation depended on NAD^+. The kinetic profile the ApeTpt1 reaction with the 5' ^{32}P-labeled $pRNA^{2'}p$ and $pRNA_{3'}p$ substrates is shown in Figs. 5A and B and quantified as described above for CthTpt1. The identity of the labeled species formed at early times in the $pRNA_{3'}p$ reaction was verified by their sensitivity to CIP (see **Figure A1.4** (B)), specifically insofar as the species designated as ADPR-pRNAp (which increased steadily at 1, 2, and 5 min) was converted by CIP into the more slowly migrating ADPR-$pRNA_{OH}$ capped RNA. As noted previously[10], ApeTpt1 was more adept than CthTpt1 at 5'-phospho-ADPR capping

(compare **Figure A1.2** and **Figure A1.5**). Nonetheless, the rate of 2'-PO_4 removal by ApeTpt1 (12.8 ± 0.6 percent•min^{-1}) was 6.7-fold faster than the rate of 5' capping of the $pRNA^{2'}p$ substrate (1.91 ± 0.03 percent•min^{-1}) (**Figure A1.5** (A)). Similarly, the rate of 3'-PO_4 removal by ApeTpt1 (5.01 ± 0.18 percent•min^{-1}) was 2.8-fold faster than the rate of 5' capping of the $pRNA_{3'}p$ substrate (1.82 ± 0.03 percent•min^{-1}) (**Figure A1.5** (B)). These data highlighted that ApeTpt1 was 2.6-fold faster at 2'-PO_4 removal than 3'-PO_4 removal.

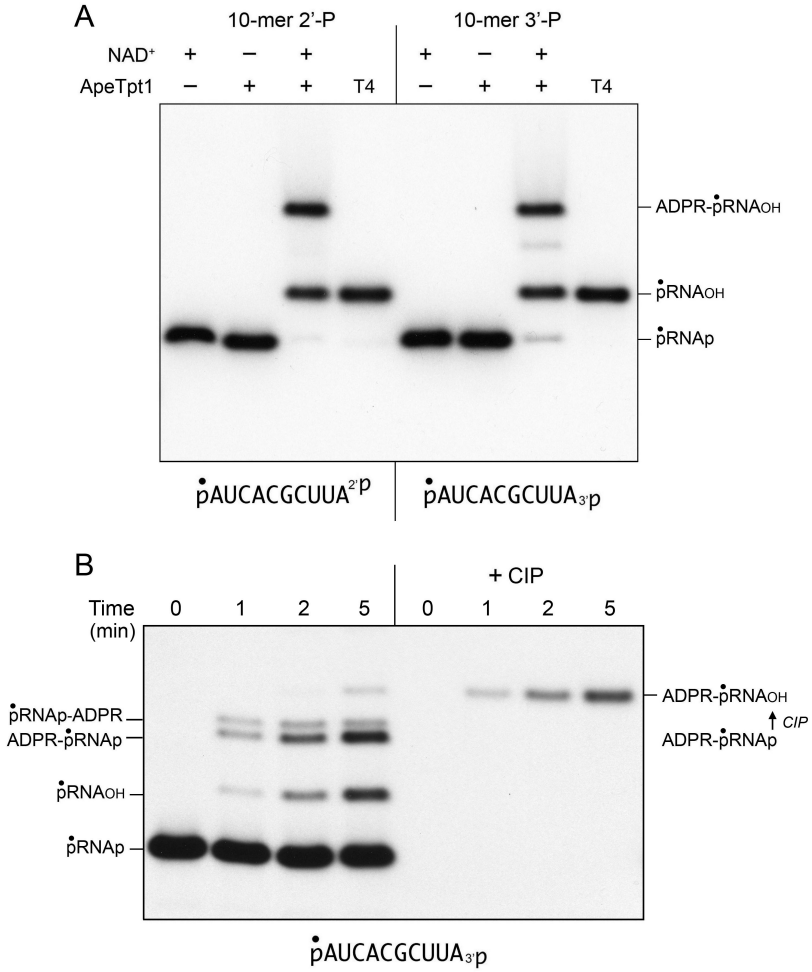


Figure A1.4. NAD^+-dependent RNA terminal 2'-PO_4 and 3'-PO_4 removal by ApeTpt1. *(A) Tpt1 reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 10-mer $pRNA^{2'}p$ or $pRNA_{3'}p$ substrates (shown at the bottom), 1 mM NAD^+ (where indicated by +), and 0.5 µM (5 pmol) ApeTpt1 (where indicated by +) were incubated at 37˚C for 30 min. T4 polynucleotide 2',3'-phosphatase reactions were performed as described in Fig. 1. The products were analyzed by urea-PAGE and visualized by autoradiography. The identities of the labeled RNAs are indicated on the right. (B) A reaction mixtures (100 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM (20 pmol) 5′ ^{32}P-labeled 10-mer $pRNA_{3'}p$ substrates (shown at the bottom), and 0.5 µM (50 pmol) ApeTpt1 was incubated at 37˚C. Aliquots (10 µl) were withdrawn at the times specified and quenched immediately with 3 volumes of of cold 90% formamide, 50 mM EDTA. Duplicate aliquots were withdrawn and heated at 65˚C for 5 min and then treated for 10 min at 37˚C with CIP as described in Fig. 1. The products were analyzed by urea-PAGE and visualized by autoradiography. The identities of the labeled RNAs are indicated on the left. The conversion of ADPR-pRNAp to ADPR-$pRNA_{OH}$ by CIP is indicated on the right.*

The specific activities of ApeTpt1 for terminal phosphate removal, calculated from the slope of the titration curves at sub-saturating enzyme, were 0.52 pmol/pmol for $pRNA^{2'}p$ and 0.23 pmol/pmol for $pRNA_{3'}p$ (not shown). The specific activity of ApeTpt1 in 5'-phospho-ADPR capping was 0.12 pmol/pmol for $pRNA^{2'}p$ and $pRNA_{3'}p$. The specific activity of ApeTpt1 in removing an internal 2'-PO_4, from the 6-mer RNA $pCCAA^{2'p}AU_{OH}$ was 15 fmol/fmol (not shown). These results indicate ApeTpt1 is 10-fold less effective than CthTpt1 in the canonical internal 2'-phosphotransferase reaction but 2-fold more effective than CthTpt1 in as a terminal 2'-phosphotransferase.

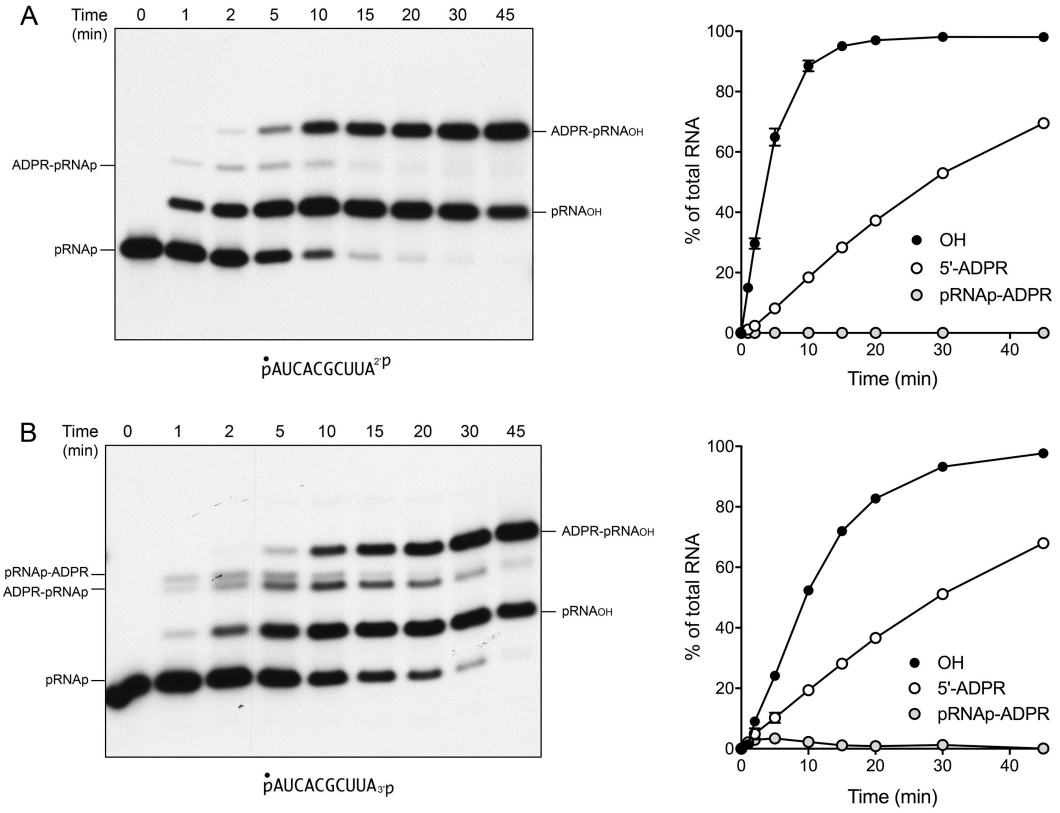


Figure A1.5. Temporal profile of RNA terminal 2'-PO_4 and 3'-PO_4 removal by ApeTpt1. *Reaction mixtures (100 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM (20 pmol) 5'-^{32}P-labeled 10-mer $pRNA^{2'}p$ or $pRNA_{3'}p$ substrates (shown at the bottom), and 0.5 µM (50 pmol) ApeTpt1 were incubated at 37°C. Aliquots (10 µl) were withdrawn at the times specified and quenched immediately with 3 volumes of of cold 90% formamide, 50 mM EDTA. The products were analyzed by urea-PAGE and visualized by autoradiography (A and B, left panels). The identities of the radiolabeled RNAs are indicated on the right. The temporal profiles of terminal phosphate removal, 5' capping, and appearance of the pRNAp-ADPR intermediate were plotted in the graphs at right as described in Fig. 2. Each datum in the graphs is the average of three independent time course experiments ±SEM.*

A.2.3 *Chaetomium thermophilum* is selective for NAD^+-dependent removal of a

terminal 2'-PO_4.

Reaction of 500 nM *Chaetomium thermophilum* Tpt1 (ChaetTpt1) with 200 nM $pRNA^{2'}p$ and 1 mM NAD^+ resulted in 2'-PO_4 removal to form a predominant $pRNA_{OH}$ product and a 5' capped species ADPR-$pRNA_{OH}$ (**Figure A1.6** (A)). By contrast, during a parallel reaction with the $pRNA_{3'}p$ substrate, ChaetTpt1 failed to remove the 3'-PO_4, and instead generated a capped product ADPR-pRNAp that retained the original 3'-PO_4. Treatment with CIP affirmed the product assignments, particularly the CIP-dependent conversion of ADPR-$pRNA_{3'}p$ to ADPR-$pRNA_{OH}$ (**Figure A1.6** (B)). The kinetic profile of the ChaetTpt1 reaction with $pRNA^{2'}p$ is shown in Fig. 6C and quantified in **Figure A1.6** D. The rate of 2'-PO_4 removal (9.5 ± 0.5 percent•min^{-1}) was 6.3-fold faster than the rate of 5' capping (1.5 ± 0.04 percent•min^{-1}). These results show that ChaetTpt1 is highly selective in its ability to ADP-ribosylate and remove a terminal RNA 2'-PO_4 but not a 3'-PO_4.

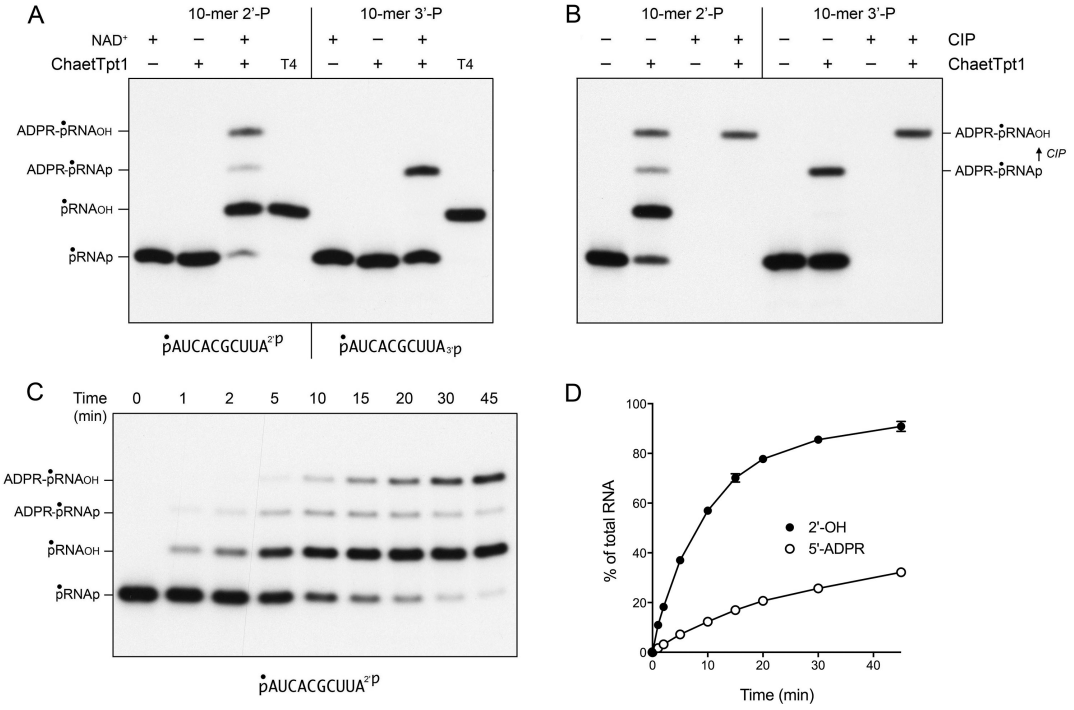


Figure A1.6. NAD^+-dependent RNA terminal 2'-PO_4 removal by Chaetomium Tpt1*. (A) Tpt1 reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 10-mer $pRNA^{2'}p$ or $pRNA_{3'}p$ substrates (shown at the bottom), 1 mM NAD^+ (where indicated by +), and 0.5 µM (5 pmol) ChaetTpt1 (where indicated by +) were incubated at 37˚C for 30 min. T4 polynucleotide 2',3'-phosphatase reactions were performed as described in Fig. 1. The products were analyzed by urea-PAGE and visualized by autoradiography. The identities of the radiolabeled RNAs are indicated on the left. (B) Reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM (2 pmol) 5′ ^{32}P-labeled 10-mer $pRNA^{2'}p$ or $pRNA_{3'}p$ substrates (shown at the bottom), and 0.5 µM (5 pmol) ChaetTpt1 (where indicated by +) were incubated at 37˚C for 30 min. The reaction mixtures were heated at 65˚C for 5 min and then either treated for 10 min at 37˚C with CIP or mock-incubated without CIP as described in Fig. 1. The products were analyzed by urea-PAGE and visualized by autoradiography. The conversion of ADPR-pRNAp to ADPR-$pRNA_{OH}$ by CIP is indicated on the right. (C) A reaction mixture (100 µl) containing 100 mM Tris-HCl, pH 7.5, 1 mM NAD^+, 0.2 µM (20 pmol) 5′ ^{32}P-labeled 10-mer $pRNA^{2'}p$ substrate (shown at bottom), and 0.5 µM (50 pmol) ChaetTpt1 was incubated at 37˚C. Aliquots*

(10 µl) were withdrawn at the times specified and quenched immediately with 3 volumes of of cold 90% formamide, 50 mM EDTA. The products were analyzed by urea-PAGE and visualized by autoradiography. The identities of the radiolabeled RNAs are indicated on the left. (D) The extents of terminal 2'-PO_4 removal and 5' capping calculated from kinetic assays in part C are plotted as a funtion of reaction time. Each datum the average of three independent time course experiments ±SEM.

Via enzyme titrations, we determined that the specific activity of ChaetTpt1 for terminal 2'-PO_4 removal was 0.84 pmol/pmol and that ChaetTpt1 was unable to remove a terminal 3'-PO_4 (**Figure A1.7**). The specific activity of ChaetTpt1 in 5'-phospho-ADPR capping was 0.15 pmol/pmol for $pRNA^{2'}p$ and 0.21 pmol/pmol for $pRNA_{3'}p$ (**Figure A1.7**). The specific activity of ApeTpt1 in removing an internal 2'-PO_4, from the 6-mer RNA $pCCAA^{2'p}AU_{OH}$ was 86 fmol/fmol (not shown).

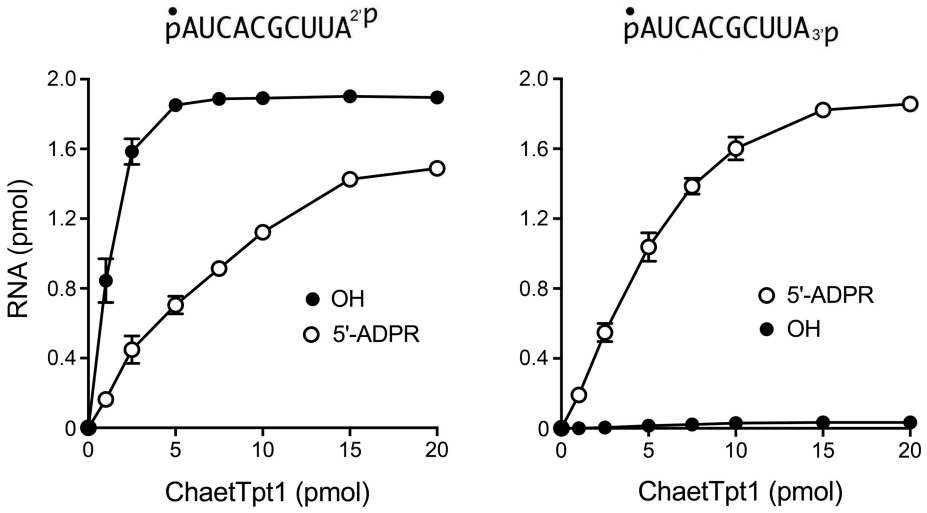


Figure A1.7. Dependence of 2'-PO_4 removal on ChaetTpt1 concentration. *Reaction mixtures (10 µl) containing 100 mM Tris-HCl (pH 7.5), 1 mM NAD^+, 0.2 µM (2 pmol) 5' ^{32}P-labeled 10-mer $pRNA^{2'}p$ or $pRNA_{3'}p$ substrates (shown at the top), and ChaetTpt1 as specified were incubated at 37°C for 30 min. The reaction mixtures were analyzed by urea-PAGE. The extents of product formation are plotted as a function of input enzyme. Each datum is the average of three independent titration experiments ±SEM.*

A.2.4 *Runella slithyformis* and human Tpt1 are inept at 2' and 3' terminal phosphate removal.

Runella slithyformis Tpt1 (RslTpt1), which is among the biochemically best understood Tpt1 enzymes with respect to the canonical Tpt1 reaction[6], and human Tpt1 (HsaTpt1) were both reported to be inactive in 5'-phospho-ADP-ribose capping[10]. Here we

found that RslTpt1 and HsaTpt1 (at 500 nM concentration) were unreactive with 200 nM pRNA$^{2'}$p or pRNA$_{3'}$p and 1 mM NAD$^+$ (**Figure A1.8** (A and B)), the same conditions that were permissive for the terminal phosphatase activities of other Tpt1 orthologs. Control reactions performed in parallel showed that RslTpt1 and HsaTpt1 completely removed the internal 2'-PO$_4$ from a 6-mer RNA mimetic of a tRNA splice junction (**Figure A1.8** (A and B)).

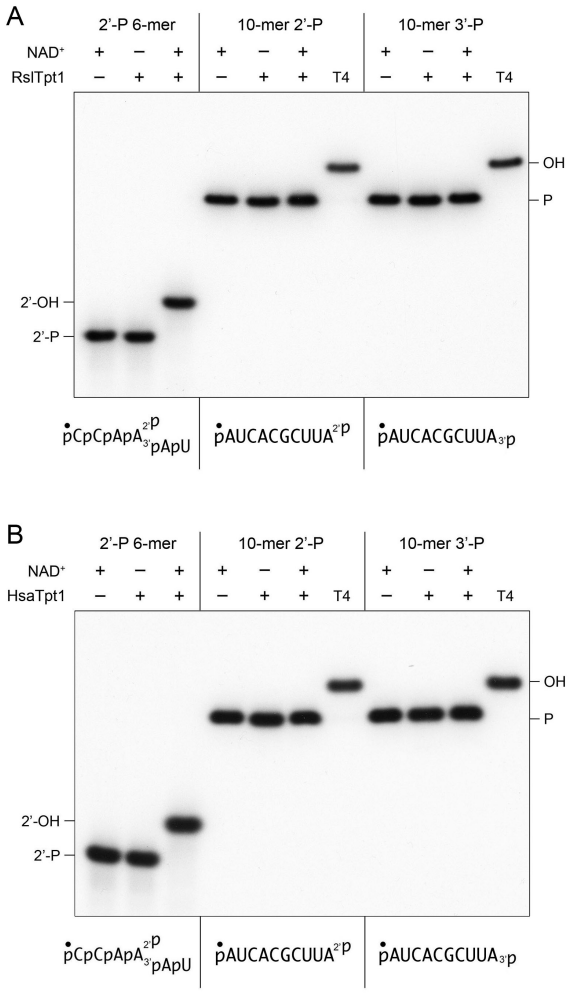


Figure A1.8. Runella and human Tpt1 are inept at RNA terminal 2'-PO$_4$ and 3'-PO$_4$ removal. *Tpt1 reaction mixtures (10 µl) containing 100 mM Tris-HCl, pH 7.5, 0.2 µM (2 pmol) 5'-^{32}P-labeled 6-mer pCCAA$^{2'p}$AU$_{OH}$ substrate or 10-mer pRNA$^{2'}$p or pRNA$_{3'}$p substrates (shown at the bottom), 1 mM NAD$^+$ (where indicated by +), and 0.5 µM (5 pmol) RslTpt1 (panel A) or human Tpt1 (panelB) (where indicated by +) were incubated at 37˚C for 30 min. T4 polynucleotide 2',3'-phosphatase reactions were performed as in Fig. 1. The products were analyzed by urea-PAGE.*

A.3 Discussion

The present study extends the catalytic repertoire of the Tpt1 enzyme family to include the NAD^+-dependent conversion of RNA terminal 2' and 3' monophosphate ends to 2'-OH and 3'-OH ends, respectively. The salient finding here is that different Tpt1 enzymes vary in their capacity and positional specificity for terminal phosphate removal. To wit, the *Clostridium* and *Aeropyrum* Tpt1 proteins are active on 2'-PO_4 and 3'-PO_4 ends, with a 2.4- to 2.6-fold kinetic preference for the former. *Chaetomium* Tpt1 acts specifically on a terminal 2'-PO_4 end and is virtually unreactive with a 3'-PO_4. By contrast, *Runella* Tpt1 and human Tpt1 are ineffective in removing either a 2'-PO_4 or 3'-PO_4 end. A previous study of *S. cerevisiae* Tpt1 employing a 5' ^{32}P-labeled trinucleotide RNA substrate with a terminal 2'-PO_4 ($pAAA^{2'P}$) indicated that the yeast enzyme was extremely feeble at terminal 2'-PO_4 removal, by a factor of 34,000 compared to its vigorous activity with a trinucleotide substrate with an internal 2'-PO_4 ($pApA^{2'P}pA$)[13]. In parallel studies with the *E. coli* Tpt1 ortholog KptA, Steiger et al. were unable to detect terminal 2'-PO_4 removal from $pAAA^{2'P}$.

Of the three Tpt1 enzymes that we find are active in terminal 2'-PO_4 removal, ApeTpt1 was the fastest, followed by ChaetTpt1 and then CthTpt1. On a per enzyme basis, ChaetTpt1 has the highest specific activity in terminal 2'-PO_4 removal, followed by ApeTpt1 and then CthTpt1. In each case, the rate and extent of the terminal 2'-PO_4 phosphate removal reaction was greater than the rate and extent of 5' ADPR capping of the RNA 5'-PO_4 end. There is a rough correlation among the enzymes studied here between their capacities for terminal phosphate removal and 5' ADPR capping, insofar as the *Runella* and human Tpt1 enzymes that are inept at terminal phosphate removal were the two enzymes found previously to be inactive in 5' ADPR capping[10]. We do not see a clear connection between the taxonomy of the Tpt1 enzymes and an expanded catalytic repertoire. For example, whereas Tpt1 from the bacterium *C. thermocellum* is adept at terminal phosphate removal, the ortholog from the bacterium *R. slithyformis* is not.

Our results here highlight RNA 2'-PO_4 and 3'-PO_4 ends as plausible endogenous substrates with which CthTpt1 reacted to form the ADP-ribose-1"-phosphate ligand retained in the NAD^+ site in the CthTpt1 crystal structure[11]. This is a more parsimonious scenario than invoking a hypothetical *E. coli* pathway for the formation of an internal RNA

2'-PO_4. However, we cannot at this point exclude the possibility that CthTpt1 reacts with a non-nucleic acid phospho-substrate in *E. coli*.

It will be of interest to understand why certain Tpt1 enzymes react with non-canonical phospho-substrates whereas others do not. The structure of CthTpt1 as a product complex mimetic with ADP-ribose-1"-phosphate in the NAD^+ site and pAp in the RNA site revealed that the essential active site amino acid chains (an Arg-His-Arg-Arg tetrad) form an elaborate network of contacts to the 1"-phosphate (equivalent to the splice junction 2'-PO_4 prior to its transfer to NAD^+) and the vicinal 3'-PO_4 of the splice junction[11]. The inference from the CthTpt1 structure that the junction 2'-PO_4 and 3'-PO_4 are the most critical determinants of substrate recognition accords with the biochemical insights of Steiger et al.[13] who showed that an RNA trinucleotide with a gem-diphospho 2'-PO_4,3'-PO_4 terminus sufficed for efficient 2'-phosphotransferase activity of *S. cerevisiae* Tpt1. With this in mind, it is relatively straightforward to imagine that Tpt1 might react with a terminal 2'-PO_4 via the same substrate binding mode seen in the CthTpt1 crystal – in which the terminal nucleobase makes a π-cation stack on an arginine and the 2'-PO_4 is coordinated by two catalytic arginines – except that the 3'-PO_4 is missing. We envision that: (i) the absence of the 3'-PO_4 contacts accounts (in large part) for the lower specific activity of CthTpt1 at a terminal 2'-PO_4 *versus* an internal 2'-PO_4; and (ii) the *Runella* and human Tpt1 enzymes that do not remove a terminal 2'-PO_4 are more stringently reliant on contacts to a vicinal 3'-PO_4 (and perhaps to 3'-flanking ribonucleotides) than are the Tpt1s than can execute the terminal 2'-phosphotransferase reaction. It would appear that said stringency is exerted during the first step of the Tpt1 reaction pathway, insofar as we do not detect RNA terminal 2'-phospho-ADP-ribosylation by RslTpt1 or human Tpt1.

Accounting for the ability of CthTpt1 and ApeTpt1 to remove a terminal 3'-PO_4 in light of the CthTpt1 crystal structure requires some conformational gymnastics at the terminal ribonucleotide, which would need to rotate about one or more bonds to place the 3'-PO_4 in roughly the same position *vis-à-vis* NAD^+ normally occupied by the 2'-PO_4. The accumulation of a terminal 3'-phospho-ADP-ribosylated RNA intermediate during the CthTpt1 3'-phosphotransferase reaction (Fig. 2B) suggests that the geometry of the 3'-p-ADPR adduct is not optimal for the ensuing transesterification step in which the ADPR ribose O2" attacks the 3'-phosphodiester and displaces the RNA 3'-OH. To our knowledge,

there has been no prior description of enzymatic ADP-ribosylation of an RNA 3'-phosphate end. Transfer of an ADP-ribose from NAD^+ to a DNA 3'-monophosphate end by human PAPR1-E998Q and PARP3 was reported recently[14, 15].

In sum, our results here fortify the theme that ADP-ribosyl transfer to a phosphorylated substrate is the unifying mechanistic feature of Tpt1-catalyzed reactions, which are not limited to the canonical activity of Tpt1 healing the 2'-PO_4, 3'-5' phosphodiester RNA splice junction formed during fungal and plant tRNA splicing. An enticing extrapolation of this idea is that some Tpt1 enzymes might catalyze NAD^+-dependent phosphoryl transfer reactions with non-nucleic acid phospho-substrates and that interrogation of such reactions in vitro might illuminate what Tpt1 is doing in its manifold biological niches.

A.4 Methods

A.4.1 Recombinant Tpt1 proteins

Tpt1 enzymes from *Clostridium thermocellum*, *Aeropyrum pernix*, *Homo sapiens*, *Runella slithyformis*, and *Chaetomium thermophilum* were produced in *E. coli* and purified as described previously[6, 10].

A.4.2 Solid phase synthesis of RNA oligonucleotides with 2'-PO_4 and 3'-PO_4 termini

Oligonucleotide syntheses were carried out using an ABI 3400 DNA synthesizer (Applied Biosystems) on either a Unylinker or 3′-Phosphate-ON (ChemGenes) solid support at a 1 μmol scale. Conventional 2′-*tert*-butyl-dimethylsilyl (TBDMS) ribonucleoside and 2′-acetyl levulinyl (ALE) ribonucleoside phosphoramidites (0.15 M in MeCN) (ChemGenes) were used. For phosphitylation, *bis*-cyanoethyl,-*N,N*-diisopropyl-phosphoramidite (0.20 in MeCN) was used. Phosphoramidites were dissolved in MeCN and activated with 5-ethylthio-1H-tetrazole (0.25 M in MeCN). Capping was carried out by the simultaneous delivery of acetic anhydride in pyridine/THF and *N*-methylimidazole (16% in THF) and contacting the solid support for 6 s. Oxidation of the phosphite triester intermediates to the phosphate triesters was affected with 0.1 M iodine in pyridine/H_2O/THF (20 s); a solution of 3% trichloroacetic acid in THF, delivered over 1.8 min, was used to deprotect DMTr groups. For 2′-phosphate-containing substrates, a solution of anhydrous TEA/MeCN (2:3 v/v) was used to remove cyanoethyl phosphate

protecting groups, while a 0.5 M solution of hydrazine hydrate in pyridine/AcOH (3:2 v/v) was used to remove 2′-ALE protecting groups. All oligonucleotides were deprotected and cleaved from the solid support using 29% aqueous ammonia/ethanol (3:1, v/v) solution. TBDMS groups were removed using TREAT-HF. Crude oligonucleotides were purified *via* HPLC and characterized by LC-MS.

The 3′-PO_4 10-mer 5′-AUCACGCUUA$_{3'\text{-P}}$ was synthesized on 3′-Phosphate-ON lcaa CPG 500Å solid support following standard 3′ to 5′ solid phase synthesis. The use of this solid support ensures a phosphate moiety remains on the 3′-terminus of the RNA upon cleavage from the support. The 2′-PO_4 10-mer 5′-AUCACGCUUA$^{2'\text{-P}}$ was synthesized on Universal Unylinker lcaa CPG 500Å solid support by first coupling a 5′-DMTr-2′-ALE adenosine 3′-O-cyanoethyl (CE) phosphoramidite (0.15 M in MeCN) for 15 min. Subsequently, to remove the cyanoethyl phosphotriester protecting group, anhydrous TEA/MeCN (2:3 v/v) solution was passed through the solid support (20 min, repeated 3x). This step ensures that the phosphate linkages vicinal to the 2′-ALE group is in the diester form, thus preventing chain cleavage in the ensuing removal of the ALE group[16]. After washing with MeCN (20 min) and drying the support over an argon stream (10 min), the synthesis columns were temporarily removed from the synthesizer and dried *in vacuo* for 30 min. To remove the 2′-ALE groups, the columns were returned to the synthesizer and a freshly prepared solution of 0.5 M hydrazine hydrate in pyridine/AcOH (3:2 v/v) was flowed through the columns (20 s flow + 3.75 min sleep, repeated 4x). After washing (MeCN, 10 min) and drying (Ar gas, 10 min), the solid supports were dried again *in vacuo* (30 min). To phosphitylate at the newly exposed 2′-OH, bis-cyanoethyl-*N,N*-diisopropyl-phosphoramidite (0.20 M in MeCN) was coupled for 30 min, and then further oxidized using 0.1 M iodine in pyridine/H_2O/THF (20 s). To complete the oligonucleotide, standard 3′ to 5′ synthesis was continued on the 5′ terminus of the growing oligonucleotide using the sequence 5′-AUCACGCUU to yield the desired 10-mer oligonucleotides.

Deprotection and cleavage of oligonucleotides from the solid support was achieved by treatment with 1 ml of cold 29% aqueous ammonia/ethanol (3:1, v/v) for 48 h at room temperature. Samples were centrifuged and the supernatant was transferred to a clean 1.5 ml Eppendorf tube and vented for 30 min, chilled on dry ice, and evaporated to dryness. Removal of the 2′-silyl protecting groups was achieved by treatment with a 300 µl solution

of NMP/Et_3N/TREAT-HF (3:4:6, v/v) for 90 min at 65˚C, followed by quenching with 3M NaOAc buffer (50 µl; pH 5.5) and precipitation of the crude oligonucleotide from cold butanol (1 ml, -20˚C). Samples were chilled on dry ice for 30 min and then centrifuged. After removing the supernatant, the remaining pellet (containing RNA) was evaporated to dryness, taken up in autoclaved milliQ water (1 ml), filtered, and quantitated by UV spectroscopy.

Crude oligonucleotides were purified by ion exchange (IE) HPLC using a Waters Protein-Pak DEAE 5PW anion exchange column (21.5 x 150 mm). A mobile phase of 1 M aqueous $LiClO_4$ in milli-Q water was used for analysis and purification of the RNA (0–24% $LiClO_4$ over 30 min, 4 mL/min, 60˚C). Following collection of the desired peaks, fractions were combined and excess $LiClO_4$ salts were removed using Gel Pak 2.5 size exclusion columns (Glen Research).

Purified oligonucleotides were characterized by electrospray ionization-mass spectrometry (**Supplementary Figure A.1**) and quantitated by UV spectroscopy. Extinction coefficients were determined using the IDT OligoAnalyzer tool (www.idtdna.com/analyzer/Applications/OligoAnalyzer).

A.4.3 5' ^{32}P-labeled oligonucleotide substrates

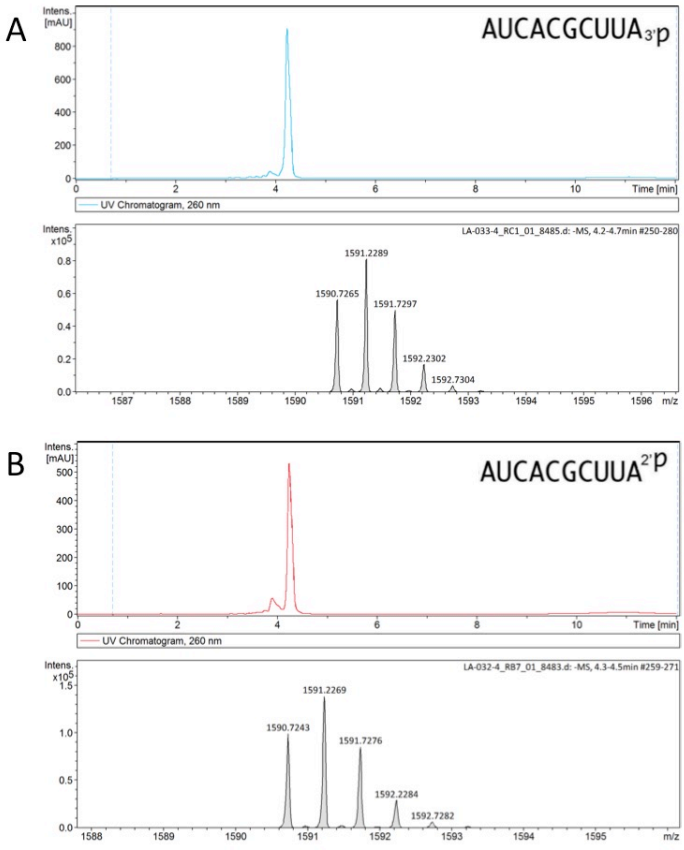
A synthetic 6-mer RNA oligonucleotide 5'-CCAA$^{2'P}$AU containing an internal 2'-PO_4 (Munir et al. 2018a) and the two synthetic 10-mer RNAs with 2'-PO_4 or 3'-PO_4 termini were 5' ^{32}P-labeled by reaction with phosphatase-dead T4 polynucleotide kinase (Pnkp-D167N) in the presence of [γ^{32}P]ATP. The reactions were quenched with 90% formamide, 50 mM EDTA, 0.01% xylene cyanol and the radiolabeled RNAs were purified by electrophoresis through a 40-cm 20% polyacrylamide gel containing 7 M urea in 45 mM Tris-borate, 1 mM EDTA. The radiolabeled oligonucleotides were eluted from excised gel slices, recovered by ethanol precipitation, and resuspended in 10 mM Tris-HCl, pH 6.8, 1 mM EDTA (TE) and stored at -20°C.

A.4.4 Assay of Tpt1 activity

Reaction mixtures containing 100 mM Tris-HCl (pH 7.5), 0.2 µM 5' ^{32}P-labeled RNA substrates, 1 mM NAD^+, and Tpt1 as specified in the figure legends were incubated at

37˚C. The reactions were quenched at the times specified in the figure legends by addition of three volumes of cold 90% formamide, 50 mM EDTA. The products were analyzed by electrophoresis (at 55 W constant power) through a 40-cm 20% polyacrylamide gel containing 7 M urea in 45 mM Tris-borate, 1 mM EDTA and visualized by autoradiography and/or scanning the gel with a Fujifilm FLA-7000 imaging device. The products were quantified by analysis of the gel scans in ImageQuant.

A.5 Supplementary Figures



Supplementary Figure A.1. LC-MS (-ve mode) chromatograms of synthetic RNA substrates. *(A) 5'-AUCACGCUUA$_{3'\text{-}P}$ – mass calculated: 3182.41, mass observed: 3182.46. (B) 5'-AUCACGCUUA$^{2'\text{-}P}$ – mass calculated: 3182.41, mass observed: 3182.45*

A.6 References

1. Culver, G. M.; McCraith, S. M.; Consaul, S. A.; Stanford, D. R.; Phizicky, E. M., A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. *J Biol Chem* **1997,** *272* (20), 13203-10.
2. McCraith, S. M.; Phizicky, E. M., An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. *Journal of Biological Chemistry* **1991,** *266* (18), 11986-11992.
3. Culver, G. M.; McCraith, S. M.; Zillmann, M.; Kierzek, R.; Michaud, N.; LaReau, R. D.; Turner, D. H.; Phizicky, E. M., An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate. *Science* **1993,** *261* (5118), 206.
4. Spinelli, S. L.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Transient ADP-ribosylation of a 2′-Phosphate Implicated in Its Removal from Ligated tRNA during Splicing in Yeast. **1999,** *274* (5), 2637-2644.
5. Steiger, M. A.; Jackman, J. E.; Phizicky, E. M., Analysis of 2'-phosphotransferase (Tpt1p) from Saccharomyces cerevisiae: evidence for a conserved two-step reaction mechanism. *RNA* **2005,** *11* (1), 99-106.
6. Munir, A.; Abdullahu, L.; Damha, M. J.; Shuman, S., Two-step mechanism and step-arrest mutants of Runella slithyformis NAD+-dependent tRNA 2′-phosphotransferase Tpt1. **2018,** *24* (9), 1144-1157.
7. Spinelli, S. L.; Malik, H. S.; Consaul, S. A.; Phizicky, E. M., A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proceedings of the National Academy of Sciences* **1998,** *95* (24), 14136-14141.
8. Popow, J.; Schleiffer, A.; Martinez, J., Diversity and roles of (t)RNA ligases. *Cellular and Molecular Life Sciences* **2012,** *69* (16), 2657-2670.
9. Sawaya, R.; Schwer, B.; Shuman, S., Structure-function analysis of the yeast NAD+-dependent tRNA 2'-phosphotransferase Tpt1. *RNA (New York, N.Y.)* **2005,** *11* (1), 107-113.
10. Munir, A.; Banerjee, A.; Shuman, S., NAD+-dependent synthesis of a 5′-phospho-ADP-ribosylated RNA/DNA cap by RNA 2′-phosphotransferase Tpt1. *Nucleic Acids Research* **2018**, gky792-.
11. Banerjee, A.; Munir, A.; Abdullahu, L.; Damha, M. J.; Goldgur, Y.; Shuman, S., Structure of tRNA splicing enzyme Tpt1 illuminates the mechanism of RNA 2'-PO4 recognition and ADP-ribosylation. *Nat Commun* **2019,** *10* (1), 218.
12. Das, U.; Shuman, S., Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res* **2013,** *41* (1), 355-65.
13. Steiger, M. A.; Kierzek, R.; Turner, D. H.; Phizicky, E. M., Substrate recognition by a yeast 2'-phosphotransferase involved in tRNA splicing and by its Escherichia coli homolog. *Biochemistry* **2001,** *40* (46), 14098-105.

14. Munnur, D.; Ahel, I., Reversible mono-ADP-ribosylation of DNA breaks. *Febs j* **2017,** *284* (23), 4002-4016.

15. Zarkovic, G.; Belousova, E. A.; Talhaoui, I.; Saint-Pierre, C.; Kutuzov, M. M.; Matkarimov, B. T.; Biard, D.; Gasparutto, D.; Lavrik, O. I.; Ishchenko, A. A., Characterization of DNA ADP-ribosyltransferase activities of PARP2 and PARP3: new insights into DNA ADP-ribosylation. *Nucleic Acids Res* **2018,** *46* (5), 2417-2431.

16. Lackey, J. G.; Mitra, D.; Somoza, M. M.; Cerrina, F.; Damha, M. J., Acetal Levulinyl Ester (ALE) Groups for 2′-Hydroxyl Protection of Ribonucleosides in the Synthesis of Oligoribonucleotides on Glass and Microarrays. *Journal of the American Chemical Society* **2009,** *131* (24), 8496-8502.

- page intentionally blank -

- END OF THESIS -