UNIVERSITY OF CALIFORNIA SAN DIEGO

# Diverse Chemical Modification of Nucleic Acids with Sulfinate Salts and Structural Studies of Plant Viroids

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Chemistry

by

Anastassia Bella Hirlinger

Committee in charge:

Professor Navtej Toor, Chair
Professor Steve Briggs
Professor Gourisankar Ghosh
Professor Thomas Hermann
Professor Ulrich Muller

2020

The dissertation of Anastassia Hirlinger is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California San Diego

2020

# DEDICATION

This doctoral dissertation is dedicated to those who have been my cheerleaders and support system throughout my long educational career: my husband Michael Hirlinger, my daughter Isabel Hirlinger, my mother Amanda Anaya, my siblings Giovanna Anaya, O'shea Anaya and Lance Anaya, my mother-in-law JoAnne Arthur and father-in-law Chris Hirlinger. I wouldn't be where I am without your faith in me and infinite love and support.

TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| COVID-19 | Coronavirus disease of 2019 |
| mRNA | Messenger RNA |
| NTPs | Nucleotide triphosphates |
| DAAS | Sodium (difluoroalkylazido)sulfinate |
| IPS | Zinc isopropylsulfinate |
| PSMS | Zinc bis(phenylsulfonylmethanesulfinate) |
| TFMS | Zinc trifluoromethanesulfinate |
| MES | 2-ethanesulfonic acid |
| LC-MS | Liquid chromatography-mass spectrometry |
| A | Adenosine |
| U | Uridine |
| G | Guanosine |
| C | Cytosine |
| SPAAC | Strain-promoted azide alkyne cycloaddition |
| CuAAC | Copper (I)-catalyzed azide alkyne cycloaddition |
| GFP | Green fluorescent protein |
| tRNA | Transfer RNA |
| NMR | Nuclear magnetic resonance |
| EM | Electron microscopy |

| | |
|---|---|
| PLMVd | Peach latent mosaic viroid |
| CChMVd | Chrysanthemum chlorotic mottle viroid |
| EDTA | Ethylenediaminetetraacetic acid |
| DI-DVI | Domain 1 – domain 6 |
| FPLC | Fast protein liquid chromatography |
| PAGE | Polyacrylamide gel electrophoresis |

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Professor Navtej Toor for his invaluable mentoring, guidance and encouragement. I can't go without expressing my gratitude to former and current lab mates Aaron Robart, Russell Chan, Jessica Peters, Dan Haack, and Timothy Wiryaman who have all been mentors and vital resources to me. I'd also like to thank the members of my committee, Professor Ulrich Muller, Professor Gourisankar Ghosh, Professor Thomas Hermann and Professor Steven Briggs for their excellent scientific critiques and guidance.

This work was funded by the HHMI Gilliam Fellowship, the UCSD Molecular Biophysics Training Grant and the NIH grant R01GM123275.

Chapter 1 has been submitted for publication as it may appear in Nature Chemical Biology, 2020. Hirlinger, Anastassia; Vantourout, Julien; Grayson, Leah; Toor, Navtej; Springer Nature, 2020. The dissertation author was the primary investigator and author of this paper.

## VITA

| | |
|---|---|
| 2013 | Bachelor of Science, University of California Santa Cruz |
| 2013-2014 | Teaching Assistant, University of California San Diego |
| 2015 | Master of Science, University of California San Diego |
| 2018 | Teaching Assistant, University of California San Diego |
| 2020 | Teaching Assistant, University of California San Diego |
| 2020 | Doctor of Philosophy, University of California San Diego |

## PUBLICATIONS

[Submitted for publication] Hirlinger, A.; Vantourout, J.; Grayson, L.; Toor, N. Diverse chemical modification of RNA and DNA with sulfinate salts. **2020**.

Toor, N.; Gomez, A. Diverse and flexible chemical modification nucleic acids. International patent # WO2020069179A1. **2020**.

Hirlinger, A. Basics of Pharmacovigilance. *AMWA Journal* **2019**, *34* (1), e29-e30.

Gomez, A.; Toor, N. Selecting New RNA Crystal Contacts. *Structure* **2018**, 26 (9), 1166-1167.

Akiyama, B.; Gomez, A.; Stone, M.D. A Conserved Motif in Tetrahymena thermophila Telomerase Reverse Transcriptase Is Proximal to the RNA Template and Is Essential for Boundary Definition. *Journal of Biological Chemistry* **2013**, *288* (30), 22141-22149.

ABSTRACT OF THE DISSERTATION

Diverse Chemical Modification of Nucleic Acids with Sulfinate Salts and Structural Studies of Plant Viroids

by

Anastassia Bella Hirlinger

Doctor of Philosophy in Chemistry

University of California San Diego, 2020

Professor Navtej Toor, Chair

My co-authors and I have devised a simple, low-cost method to modify RNA with sulfinate salts that can directly add almost any desired functional group under mild conditions. Existing methods of RNA modification have relatively limited applicability due to constraints on the

size of the RNA and the lack of diversity of possible modifications. This chemistry modifies the Hoogsteen edge of nucleobases and is done as a single pot reaction. It can be applied to RNA or DNA of any size, as well as to individual nucleotides. Sulfinate salts can modify RNA with a broad range of functional groups, such as fluorophores, biotin and medicinally relevant small molecules such as trifluoromethyl groups. This methodology enables the exploration of diverse chemical groups on RNA that can potentially confer protection from nucleases, allow for efficient delivery of nucleic acids into cells, or act as new tools for the investigation of nucleic acid structure and function.

Prior to working on the chemical modification of RNA, I attempted to solve the structure of a viroid using cryo-electron microscopy (cryo-EM) and x-ray crystallography. Viroids are infectious RNAs that target plants, including important food crops such as potatoes, apples and avocados. They are 240 to 400 nucleotides long, single stranded, covalently closed circular RNAs. Viroids, incredibly, do not encode protein and have no DNA replication intermediate. Of the secondary structures that have been experimentally determined, two are predicted to fold into higher order tertiary structures: the peach latent mosaic viroid (PLMVd) and the chrysanthemum chlorotic mottle viroid (CChMVd). Biochemical evidence suggests that these tertiary interactions are essential for infectivity. Since

no 3D structure of any viroid exists, I decided to try to determine the PLMVd structure at high resolution. Many attempts were made to solve a viroid structure with both x-ray crystallography and cryo-EM, but I was ultimately unsuccessful due to the viroid's persistent aggregation.

# CHAPTER 1

## Introduction

There are currently over 170 different RNA modifications that are known to exist in prokaryotes and eukaryotes (1). These RNA modifications affect all aspects of gene expression and regulation with effects upon both transcription and translation. These modifications occur on the nucleobase aromatic rings and are incorporated through the action of specific enzymes for each functional group that is added to the RNA (1). One of the most prevalent modifications is pseudouridine in mRNA. This modification greatly increases the half-life of pseudouridine-containing RNA *in vivo* compared to unmodified RNA (2). Remarkably, pseudouridine is able to stabilize RNA *in vivo* with a random distribution of this residue throughout the mRNA sequence. Pseudouridine and its derivatives are now the basis for a multi-billion dollar industry that is currently developing mRNA therapeutics and vaccines for COVID-19 (3). mRNA is also modified with other functional groups in the cell, such as N6-methyladenosine, to regulate its translation by the ribosome (4). A recent report of acetyl modifications to cytidine nucleobases in mRNA

showed that these modifications promoted translation efficiency and mRNA stability *in vivo* (5). Therefore, RNA modification is an emerging area of study that is revealing itself to be a significant mechanism for modulating biological pathways and conferring new properties upon RNA.

Given the impact of nucleobase modifications upon the biochemical properties of RNA, there has been great interest in the incorporation of novel synthetic functional groups to enable additional functionality. Reagents such as dimethyl sulfate and kethoxal have been historically used to modify the Watson-Crick edge of nucleobases in RNA for structure probing (6); however, it is not possible to add a wide variety of functional groups using this type of chemistry. The synthesis and chemical diversity of synthetically modified RNA is limited by the following problems:

1. T7 RNA polymerase does not efficiently incorporate modified nucleobases during *in vitro* transcription. Long RNAs (>100 nucleotides) are typically synthesized using T7 RNA polymerase in the presence of modified nucleotide triphosphates (NTPs). However, T7 RNA polymerase is not able to accommodate highly modified NTPs within the restrictive confines of its active site (7,8). As a result, the diversity of modified RNA is limited by the enzymatic mechanism of T7 RNA polymerase. The

majority of possible modifications cannot be incorporated at all by the polymerase.

2. The low efficiency of incorporation of modified NTPs during *in vitro* transcription results in exorbitant costs of synthesis. Modified NTPs are extremely expensive and the low efficiency of incorporation greatly increases the cost of synthesis for large-scale production of therapeutics (8).

3. Chemical synthesis to incorporate modified nucleobases is largely limited to the production of short nucleic acids. Solid phase oligonucleotide (oligo) synthesis is limited to the modification of small oligos; generally, no more than 200 deoxyribonucleotides with the phosphoramidite method and no more than 60 ribonucleotides for RNA (9). For example, direct synthesis of modified messenger RNAs (typically >1,000 nucleotides) is currently not possible using chemical methods. A method to bioconjugate DNA oligos was recently published but the technology requires a solid support and is limited to short DNA oligos (10).

RNA poses unique challenges for chemical modification due to its relative lability compared to DNA (2,11,12). The modification of RNA has generated interest in medicinal chemistry due to its potential use as a therapeutic agent for a range of diseases (2,11,13–15). Unfortunately,

many reactions used in synthetic organic chemistry would result in the RNA being destroyed due to the harsh conditions. The modification of nucleic acids is also broadly useful as a research tool via bioconjugation tags for affinity purification or the addition of fluorophores for cellular tracking of labeled RNAs (11,12,16). Further development of RNA-based therapeutics would greatly benefit from a low-cost and simple method of introducing desired functional groups to RNA nucleobases. For example, mRNA-based vaccines are currently in clinical trials for COVID-19 and this requires the rapid production of billions of doses in a short period of time. This vaccine could be engineered with novel functional groups to exhibit greater stability within the human body for a stronger immune response to the encoded viral spike protein. This necessitates the development of chemistry that could easily scale to the large quantities of synthetic mRNA required with a low cost for the incorporation of modified nucleotides.

Here my co-authors and I have demonstrated the chemical modification of RNA and DNA with different sulfinate salt reagents that attach R groups to the Hoogsteen edge of the nucleobases. Specifically, I tested zinc trifluoromethanesulfinate (TFMS), zinc bis(phenylsulfonylmethanesulfinate) (PSMS), zinc isopropylsulfinate (IPS), and sodium (difluoroalkylazido)sulfinate (DAAS). Each reagent adds a unique functional group onto the substrate. TFMS adds a trifluoromethyl

group, PSMS adds a phenylsulfonylmethyl group, IPS adds an isopropyl group, and DAAS adds a difluoroalkylazido group in which the terminal azide moiety serves as a click chemistry substrate (17–19). Collectively, these reagents can modify all nucleobases and each nucleobase has a different reactivity level. In addition, the reaction chemistry is flexible with respect to its substrates, including both small and large oligonucleotides, as well as large sequences of RNA and DNA. Importantly, the nucleic acid functionality enabled by sulfinate salt chemistry is extremely diverse, considering the number of sulfinate salt reagents and the vast number of commercially available “clickable” moieties readily available to scientists.

## Results

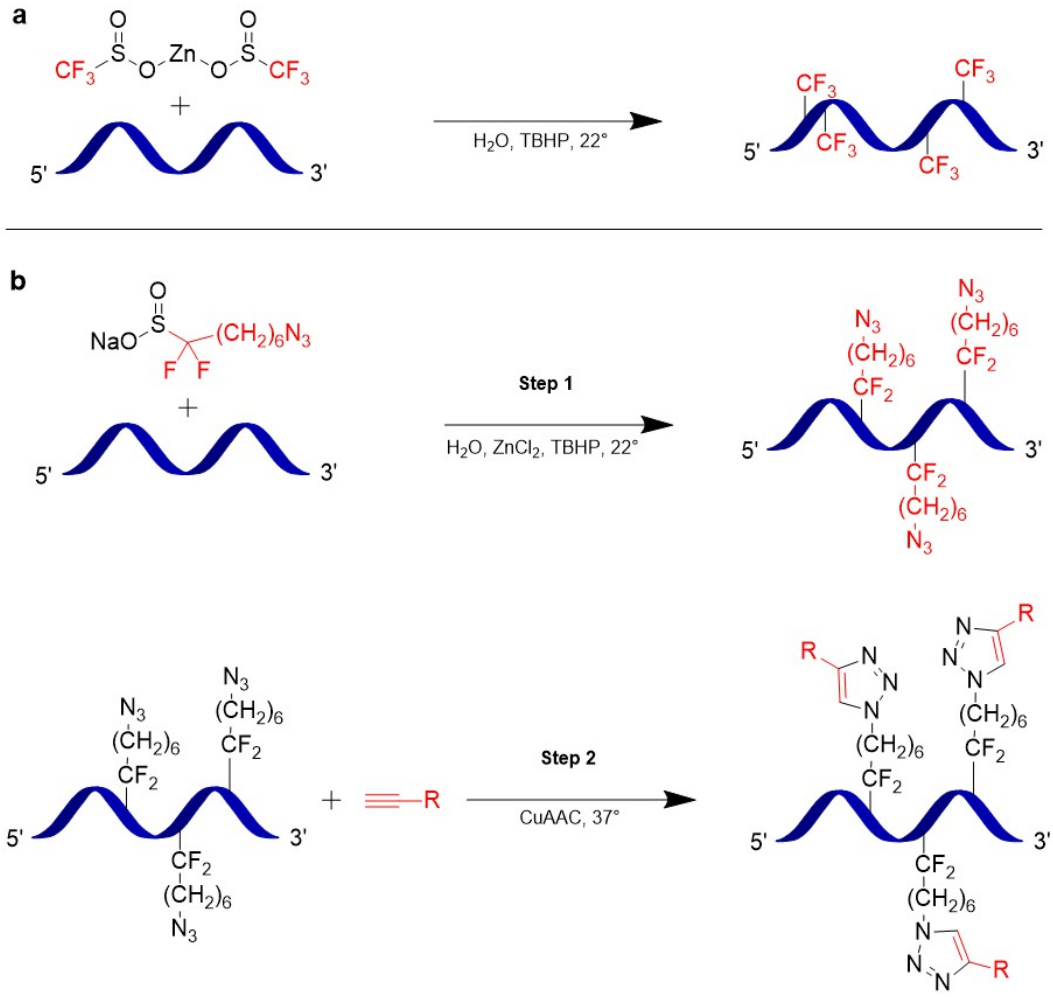
### A simple methodology for modification of nucleic acids

Sulfinate salts have been used to functionalize CH groups found in heteroarene-containing small molecule compounds. I hypothesized that these salts could be adapted to introduce functional groups onto the nucleobases of RNA since they are also heteroarenes. In contrast to small molecules, RNA is a long polymer that is prone to cleavage under

harsh conditions such as high pH, high temperature, and the presence of certain divalent cations. I therefore optimized reaction conditions to keep the RNA intact during the sulfinate modification reaction; this included an essential buffer component to control the pH of the reaction. The sulfinate salts TFMS, PSMS and IPS react with nucleic acids at room temperature, with water as the solvent and the pH maintained at 6.5 with 50 mM MES buffer (Figure 1.1a). Preliminary experiments suggested that reaction temperatures between 4°C and 40°C are suitable for nucleic acid modification. I also confirmed that the RNA remains intact after treatment with sulfinate salts (Supplementary Figure 1.1).

While all the sulfinate salts (TFMS, PSMS, IPS and DAAS) modify nucleic acids in a single step, DAAS adds a bioconjugation tag which can react with an alkyne in a second step to generate many possible final functional group additions (Figure 1.1b). The reactivity of sulfinate salts is specific to carbon-hydrogen sites on heteroarenes and there is one site of modification on all nucleobases (17–20) (Figure 1.1c). Julien Vantourout used fluorine-19 NMR on TFMS-modified DNA bases, which showed a single reactive position on each nucleobase (Supplementary Fig. 2). Notably, all these positions are on the Hoogsteen edge of nucleobases and I predict Watson-Crick edge base pairing would be minimally affected.

**Figure 1.1: Modification of nucleic acids of any size with diverse functional groups.** **a**, Schematic of the single-step modification reaction between a nucleic acid and the zinc trifluoromethanesulfinate (TFMS). **b**, Schematic of the two-step bioconjugation reaction which enables a diverse set of functional groups to be added to a nucleic acid. In the first step, sodium difluoroalkylazidosulfinate (DAAS) reacts with a nucleic acid to generate a modification with a terminal azide. In the second step, any functional group with a terminal alkyne (alkyne with terminal R group in red) can react with the terminal azide moiety to produce a nucleic acid with various modifications. **c**, Sulfinate salt reactive positions indicated by the R group on nucleobases. The R group is the moiety added by TFMS, PSMS, IPS or DAAS.


a
$CF_3$
S
O
Zn
$CF_3$
5'
3'
$H_2O$, TBHP, 22°
b
NaO
$(CH_2)_6N_3$
F F
Step 1
$H_2O$, $ZnCl_2$, TBHP, 22°
$N_3$
$(CH_2)_6$
$CF_2$
Step 2
CuAAC, 37°
R
c
Uracil
Cytosine
Adenine
Guanine
Thymine
ribose
deoxyribose

**LC-MS analysis reveals modification efficiency of each nucleobase**

With the help of Leah Grayson, I reacted a 388-nucleotide group II intron RNA from *Oceanobacillus iheyensis* (*O.i.*) with sulfinate salts for different reaction times, followed by enzymatic digestion to obtain individual nucleotides. This nucleotide pool was analyzed by Julien Vantourout with liquid chromatography-mass spectrometry (LC-MS) to confirm the presence of modifications on the nucleobases after reaction with each sulfinate salt (Figure 1.2). The data shows that all sulfinate salt reagents modify each nucleobase similarly: uridine had the greatest conversion to the modified product, adenosine had the second greatest conversion, cytidine had the third greatest conversion and guanosine had the lowest conversion (Figure 1.2, Supplementary Table 1.1). The reaction time I used for most other sulfinate salt experiments was 24 hours, however, the reaction time can be reduced to 8 hours for 67 - 83% of the modification levels seen at 24 hours (Figure 1.2). Fluorine-19 NMR analysis of a mixture of TFMS-modified DNA bases, performed by Julien Vantourout, revealed highly similar levels of modification to the TFMS-modified RNA bases evaluated in the LC-MS experiment (Supplementary Figure 1.3).

Sulfinate modified 20-nucleotide DNA and RNA oligonucleotides were also analyzed by LC-MS to determine the extent of modification for a given length of sequence (Supplementary Table 1.2). This revealed that each sulfinate salt reagent made 1-2 modifications per 20 nucleotides. TFMS continues to be have the greatest conversion rate and modifies above 40% of the overall oligo population (Supplementary Table 1.2). PSMS has the second greatest conversion, modifying 35% of the overall oligo population and DAAS has the lowest conversion of the three salts, modifying at least 6% of the overall oligo population. Leah Grayson also helped me prepare and purify the oligos for this analysis.

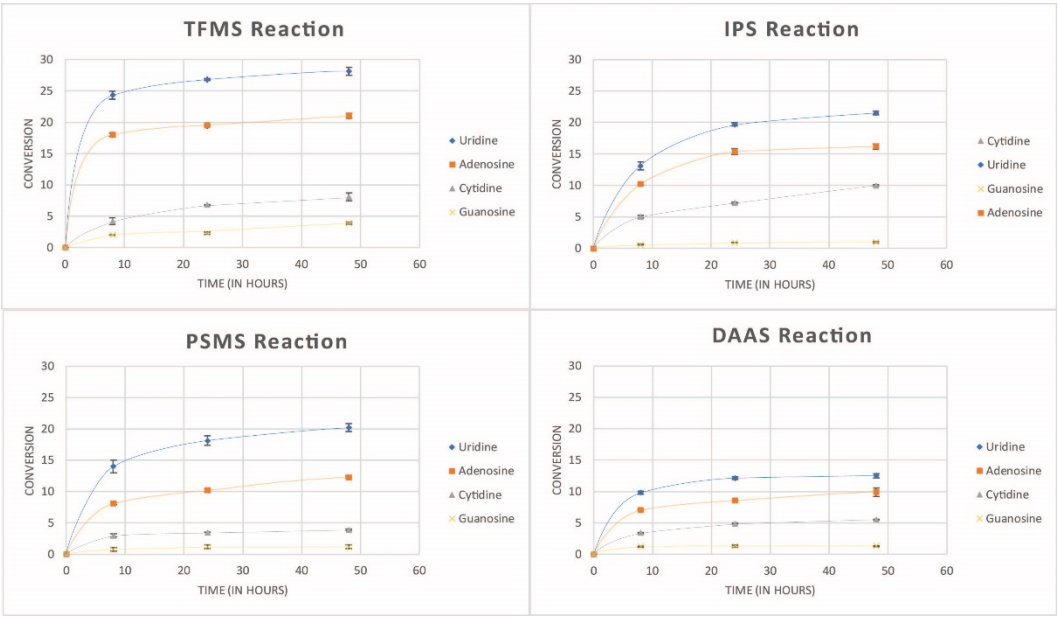



**Figure 1.2: Modified nucleoside time course study.** *O.i.* RNA at 1 mg/mL was modified with various sulfinate salts in separate reactions for 8 hours, 24 hours or 48 hours. RNAs were digested into nucleosides and LC-MS analysis was used to evaluate the fraction of nucleosides that were modified. The average conversion to the modified product from three replicate experiments is plotted on the graphs.

## A large TFMS-modified RNA has a fluorine-19 NMR chemical shift

Fluorine-19 NMR was used to confirm the presence of the trifluoromethyl groups on the nucleobases of an intact group II intron RNA. Natural, native RNAs do not contain fluorine, therefore the existence of a chemical shift by fluorine-19 NMR suggests the RNA was modified with a trifluoromethyl group. The observed chemical shifts for the *O.i.* RNA are -61.2 ppm and -61.4 ppm (Figure 1.3). I also modified an even larger 622-nucleotide group II intron RNA from *Pylaiella littoralis* (*P.li.*) with TFMS and

observed fluorine-19 NMR chemical shifts at -60.9 ppm and -62.7 ppm (Supplementary Figure 1.4). The two chemical shifts observed for each RNA may be due to the two different geometries of purines and pyrimidines, since NMR chemical shifts are affected by the local geometry and chemical environment. The TFMS-modified RNA chemical shifts are within the range for most alkyl- or aryl-bound trifluoromethyl groups, which is between about -60 and -80 ppm (21,22). The observed chemical shifts are also consistent with that of an aryl-bound trifluoromethyl group(21) and with the reported chemical shifts for trifluoromethyl-modified uracil, -65.1 ppm, and trifluoromethyl-modified caffeine, -62.7 ppm (23). To contrast, the observed chemical shift for the sulfinate salt TFMS in the same buffer as the RNAs is -87.4 ppm (Supplementary Figure 1.5).

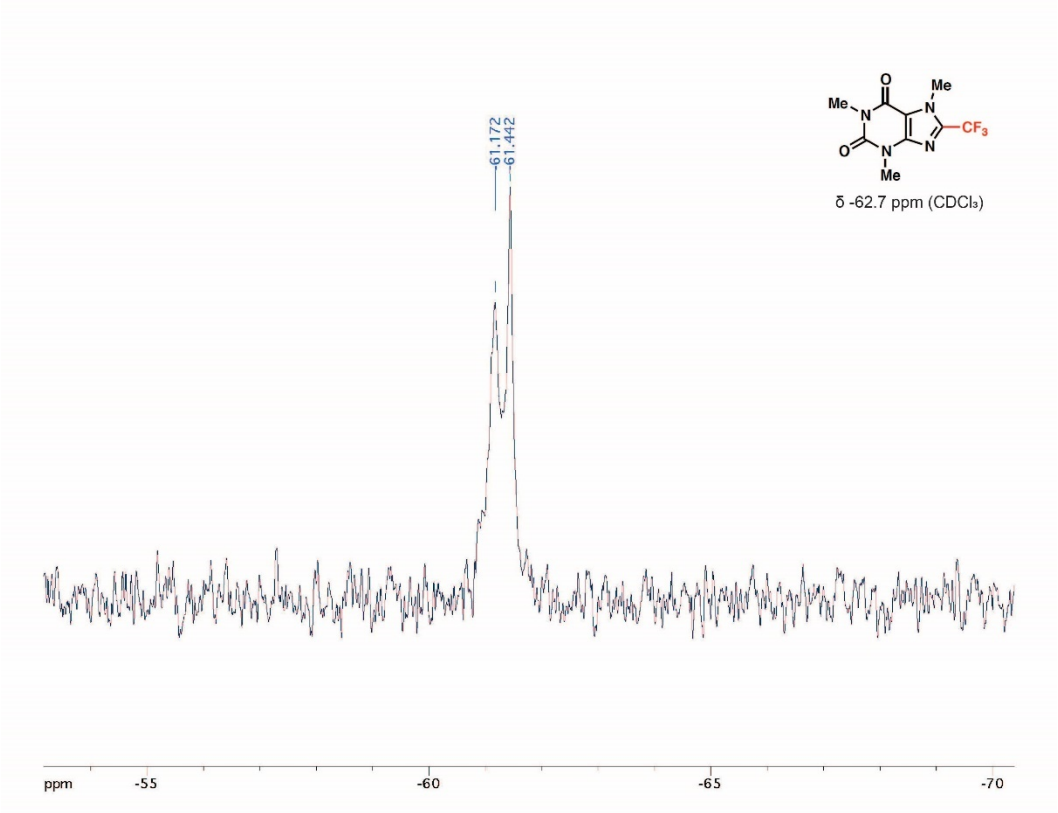
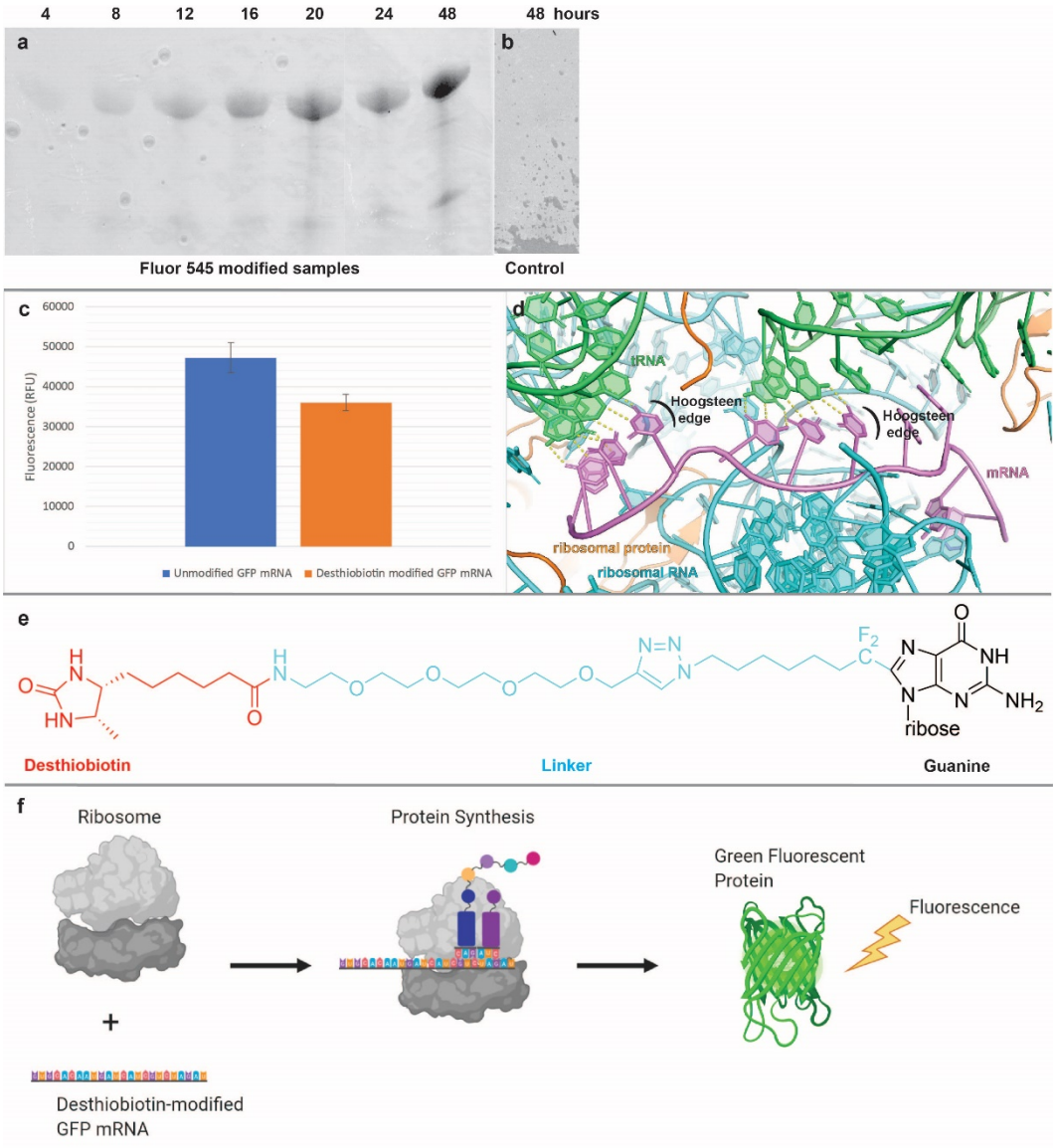



**Figure 1.3: Fluorine-19 NMR spectrum for a large trifluoromethyl-modified RNA.** The observed fluorine-19 NMR chemical shifts for TFMS-modified *O.i.* RNA are -61.2 ppm and -61.4 ppm in 10% $D_2O$. Inset is trifluoromethyl-modified caffeine, which has a reported chemical shift of -62.7 ppm in $CDCl_3$ (23).

**DAAS enables nucleic acid bioconjugation with diverse functional groups**

The azide group added by DAAS is compatible with both copper-free click chemistry and copper (I)-catalyzed azide alkyne cycloaddition (CuAAC) click chemistry reactions. Copper-free click chemistry is primarily used with metal-sensitive systems such as *in vivo* tagging of nucleic acids. Specifically, I employed strain-promoted azide alkyne cycloaddition (SPAAC) to react a DBCO labeled reagent with an azide-tagged group II intron RNA. I successfully added a DBCO labeled fluorophore, Fluor 545, to the *O.i.* RNA using copper-free click chemistry (Figure 1.4a) and the resulting modified RNA exhibited fluorescence. I also confirmed that CuAAC click chemistry is possible with these sulfinate-modified RNAs. CuAAC click chemistry is preferred for use in cell-free systems since it has higher selectivity and increased reactivity for its substrate. I added an alkyne labeled fluorophore, Fluor 488, to both RNA and DNA 20-nucleotide oligos and to the *O.i.* RNA using CuAAC (Supplementary Table 1.2, Supplementary Figure 1.6). I was similarly able to add Fluor 488 to plasmid DNA that was longer than 3000 base pairs (data not shown).

**Figure 1.4: Fluorescent modified RNA and *in vitro* translated modified mRNA. a**, Fluorescence scan of DAAS-modified *O.i.* RNA treated for different lengths of time with Fluor 545 on a 4% denaturing polyacrylamide gel. **b**, Fluorescence scan of unmodified *O.i.* RNA which underwent the same treatment as experimental samples except that there was no DAAS added in the first step of the reaction. **c**, GFP mRNA was modified with desthiobiotin and purified from the reaction mixture with streptavidin agarose resin. Desthiobiotin-GFP mRNA was added to HeLa lysate in vitro translation reactions and the fluorescence of translated GFP protein was detected on a plate reader. The average fluorescence from three replicate experiments is plotted on the bar graph. **d**, Structure of the *Thermus thermophilus* 70S ribosome in complex with mRNA and tRNAs in the A-, P- and E-sites. There is ample space for modifications on the Hoogsteen edge of mRNA (pink) during translation by the ribosome (ribosomal protein in orange and ribosomal RNA in blue). mRNA forms Watson-Crick base pairs (dashed lines) with A- and P-site tRNAs (green, left and center) and base stacks with the E-site tRNA (green, right). Black arcs indicate the Hoogsteen edge of nucleobases. (PDB accession 4V5D (24)). **e**, Structure of the desthiobiotin moiety (red) and its linker (blue) that was used to modify mRNA coding for GFP. **f**, Overview of the *in vitro* translation of the desthiobiotin modified mRNA into active green fluorescent protein. Fluorescence is only detected upon successful translation of the desthiobiotin-modified RNA by the ribosome to produce functional GFP protein.


4 8 12 16 20 24 48 48 hours
a
b
Fluor 545 modified samples
Control
c
60000
50000
40000
30000
20000
10000
0
Fluorescence (RFU)
Unmodified GFP mRNA
Desthiobiotin modified GFP mRNA
d
tRNA
Hoogsteen edge
Hoogsteen edge
mRNA
ribosomal protein
ribosomal RNA
e
Desthiobiotin
Linker
Guanine
ribose
f
Ribosome
Protein Synthesis
Green Fluorescent Protein
Fluorescence
Desthiobiotin-modified GFP mRNA

To further demonstrate that bioconjugation with DAAS allows for nearly any alkyne-labeled target molecule to be added to a nucleic acid, I modified a 20-nucleotide DNA oligo with biotin (Supplementary Table 1.2) and modified mRNA coding for green fluorescent protein (GFP) with desthiobiotin. Figure 1.4e shows that the resulting functional group on mRNA consists of a relatively long linker with a desthiobiotin moiety attached at the terminus. Biotin and desthiobiotin are well-known affinity tags for streptavidin and are widely used for purification of biomolecules. Desthiobiotin-modified GFP mRNA was purified away from unmodified GFP mRNA using streptavidin agarose resin and was then tested for biochemical activity using an in vitro translation assay (Figure 1.4f). This assay involves incubation of the modified RNA with HeLa cell lysate that contains ribosomes and fluorescence is detected upon successful translation of the mRNA into functionally active GFP protein. The modified GFP mRNA produced 76% of the fluorescence of the unmodified GFP mRNA (Figure 1.4c), which suggests that the sulfinate salt protocol yields modified mRNA that is biochemically active and also suggests that the ribosome can tolerate large, bulky substitutions on the Hoogsteen edge of RNA nucleobases (Figure 1.4d).

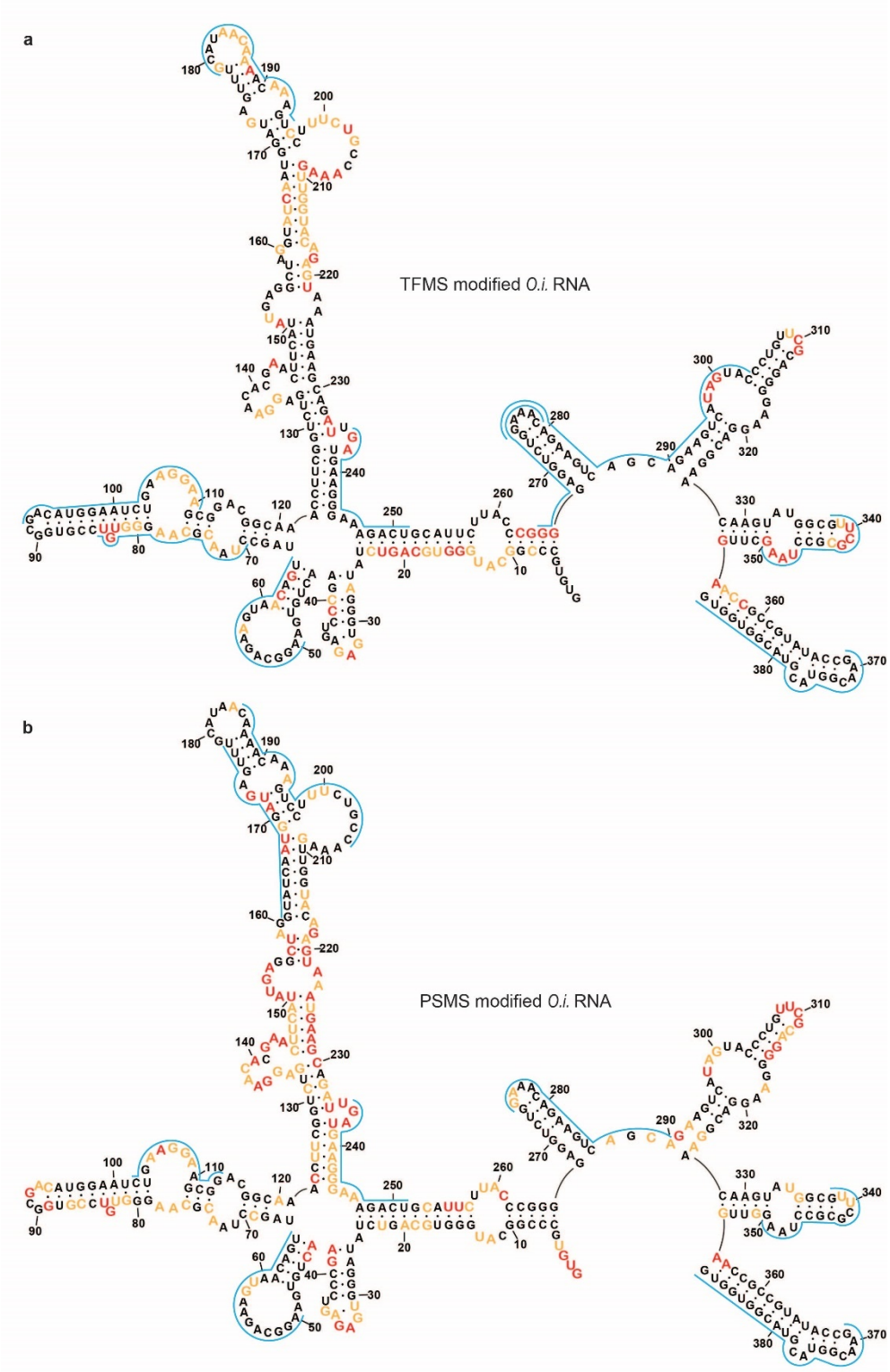
**Sulfinate salts for structural probing of the *O.i.* RNA**

I had initially found that modifications were made to RNA using a primer extension assay on PSMS-modified *O.i.* RNA. Primer extension assays are used for structural probing based on the idea that a modification made to an RNA which generates a primer stop in a reverse transcription reaction is potentially sampling a structural feature of the RNA. I also found that TFMS modifications generated strong primer stops and that DAAS modifications generated very weak primer stops, suggesting TFMS, like PSMS, may have a structural probing use and that DAAS likely did not (Supplementary Figure 1.7).

The 3D structure of the *O.i.* RNA is already known, so it is a useful model RNA to test the structural probing capacity of the PSMS and TFMS sulfinate salts. I used several primers to probe the TFMS modified *O.i.* structure from G1 up to C358 (Figure 1.5a), as well as the PSMS modified structure from G1 up to A356 (Figure 1.5b). The primer extension data was normalized to create a reactivity range from 0 to about 1.5, the same as is done for SHAPE structural analysis (25). Nucleotides considered highly reactive toward the sulfinate salt are in red (reactivity >0.85), those considered moderately reactive are in yellow (reactivity 0.4 - 0.85) and nucleotides considered unreactive are in black (reactivity <0.4). The

sulfinate salt reactions were done at 40 °C for 24 hours with splicing buffer and TBHP. All controls contained splicing buffer and the *O.i.* RNA but did not contain zinc chloride or TBHP. In experiments following the completion of my primer extension data, I tested whether the addition of zinc chloride and TBHP (Supplementary Figure 1.7a) or the addition of zinc chloride alone (Supplementary Figures 1.7b-d) to the control samples could impact the primer extension analysis. Zinc is present in PSMS and TFMS experimental samples due to the sulfinate salt formulations [zinc bis(phenylsulfonylmethanesulfinate) and zinc trifluoromethanesulfinate]. It appears that the addition of zinc chloride does generate a few new primer stops to the control samples and that the addition of TBHP beyond the addition of zinc chloride does not further increase this difference.

**Figure 1.5: PSMS and TFMS evaluation as structural probing reagents. a**, TFMS modified *O.i.* RNA and **b**, PSMS modified *O.i.* RNA primer extension data. Nucleotides in black, yellow and red have low, medium and high reactivity for the sulfinate salt, respectively. The blue lines show the location of the primers used in the primer extension assays. No data was recorded for C358 – G388 in **a** and no data was recorded for A356 – G388 in **b**.


a
TFMS modified 0.i. RNA
b
PSMS modified 0.i. RNA

## Discussion

I have presented a simple, low-cost method for the modification of both RNA and DNA. It satisfies the need for a technique to modify both RNA and DNA, large or small, under mild conditions that allow the nucleic acid to remain intact. These modification reactions can be easily accomplished without the need for purchasing enzymes or specialized equipment. Sulfinate salt chemistry greatly expands the functional landscape of nucleic acids. To my knowledge, this is the first demonstration of large nucleic acid modification and of RNA modification using sulfinate salts. The modifications made by TFMS, PSMS and IPS are done in a single step reaction and can add small polar trifluoromethyl groups, bulky aromatic phenylsulfonylmethyl groups, and branched nonpolar isopropyl groups, respectively. There are other sulfinate salts commercially available which can add a variety of other functional moieties to nucleic acids.

The modification of nucleobases with sulfinate salts is supported by mass spectrometry and NMR analysis. The presence of the expected modifications by TFMS, PSMS, IPS and DAAS are confirmed with LC-MS on RNA nucleosides (Figure 1.2, Supplementary Table 1.1). At the oligo

level, the expected modifications from TFMS, PSMS, and DAAS are confirmed on DNA and the expected modification by DAAS is confirmed on RNA (Supplementary Table 1.2). Fluorine-19 NMR analysis provided evidence of trifluoromethyl modifications on both large RNAs and on DNA nucleosides (Figure 1.3, Supplementary Figures 1.2 – 1.4). Trifluoromethyl modifications on *O.i.* and *P.li.* RNAs treated with TFMS are supported by fluorine-19 NMR chemical shifts (Figure 1.3, Supplementary Figure 1.4).

The different electronic properties of the nucleobases are likely the primary reason for the different reactivities of nucleobases with sulfinate salt reagents and for the modification sites on nucleobases. All reactive positions are on the Hoogsteen edge of nucleic acids, which I predict enables nucleic acids to fold and function normally (Figure 1.1c, Supplementary Figure 1.2). These positions are likely solvent exposed and unlikely to affect the ability of the RNA to engage in Watson-Crick pairing.

I modified RNA with an azide group that can be used as a bioconjugation site for the further addition of diverse functional groups using click chemistry. Using this methodology, I successfully added biotin and fluorophores to the Hoogsteen edge of nucleobases (Figure 1.4a,

Supplementary Table 1.2, Supplementary Figure 1.6). I determined that a desthiobiotin-modified mRNA coding for GFP was translated into active GFP protein exhibiting fluorescence, which suggests mRNA modified by sulfinate salts can be efficiently translated by the ribosome (Figure 1.4c). This biochemical activity is supported by the crystal structure of the ribosome with bound mRNA captured in the process of synthesizing protein (24) (Figure 1.4d). This ribosome structure contains all three bound tRNAs in the A-, P-, and E-sites and therefore represents a catalytically relevant snapshot of protein synthesis in action. In this structure there is ample space on the Hoogsteen edges of the tRNA-bound mRNA nucleotides to accommodate the incorporation of large functional groups. The desthiobiotin is attached to the nucleobases via a relatively long linker and is therefore flexible, which likely reduces the chances of this large modification causing possible steric interference that could inhibit translation. In addition, the known natural acetylation of cytidine occurs in close proximity to the Hoogsteen edge and also does not affect translation efficiency (5). Taken together, the data is consistent with the hypothesis that mRNAs can tolerate large, bulky substitutions on the Hoogsteen edge of RNA nucleobases and still undergo efficient translation by the ribosome to produce protein.

The use of sulfinate salts to modify RNA results in the random incorporation of functional groups throughout the sequence. Our experiments with the DNA/RNA oligonucleotides showed a conversion level of 1 to 2 modifications per 20 nucleotides. Therefore, it is likely that the 1369-nt GFP mRNA contains between 69 to 137 modifications.

I was not able to correlate the PSMS or TFMS modification sites on *O.i.* RNA with any particular feature of the 3D structure (Figure 1.5). The lack of zinc chloride in the control samples in Figure 1.5 and the affect that zinc has on the control primer extension (Supplementary Figure 1.7b-d) suggests that some of the reactivity observed for TFMS and PSMS could be inflated. However, I believe that a careful repeat of these experiments with zinc in the control samples would show that the nucleotide reactivity would be slightly reduced overall and would continue to show a similar pattern of modification. The ability of the reverse transcriptase to tolerate bulky modifications, like the phenylsulfonylmethyl group added by PSMS, to the Hoogsteen edge of nucleobases suggests sulfinate modification has the potential to serve as a methodology for probing RNA structure in a manner similar to the well-known SHAPE protocol (25), but will likely require further optimization.

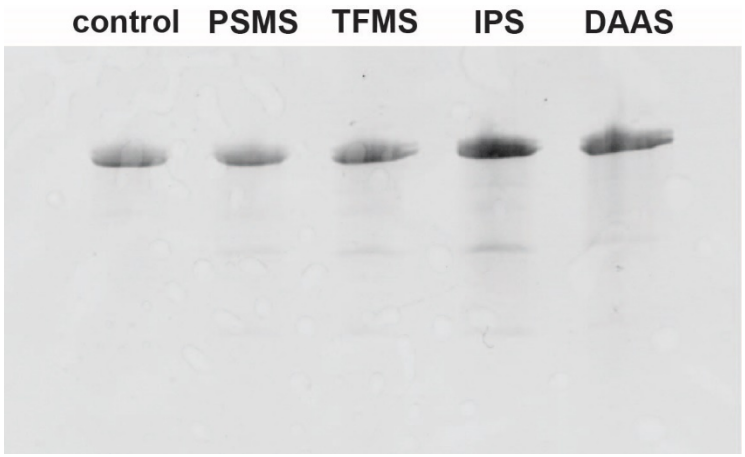
Modification of nucleic acids with sulfinate salts has the potential to aid development of therapeutics through the identification of moieties that can more efficiently deliver and more effectively protect therapeutic gene products. Some modifications, such as pseudouridinylation and N4-acetylcytidine, have been shown to extend the half-life of mRNA in mammalian cells (2,5). The simplicity of making modifications to RNA using sulfinate salt chemistry enables a low-cost, efficient search for modifications that could further extend the half-life of RNA *in vivo*. Similarly, this method facilitates the search for modifications that could allow RNA to more readily cross cellular membranes for the delivery of nucleic acid therapeutics. For example, covalent lipid-RNA conjugates could be generated by reacting the DAAS-modified RNA with an alkyne-labeled lipid that would result in greater hydrophobicity to enable cell entry. This technique could also be used to generate novel nucleoside analogs as viral inhibitors. Modified nucleosides such as remdesivir and NHC are proposed drug therapies for SARS-CoV-2, the cause of a current and ongoing pandemic (26,27). Modifications to the Hoogsteen edge of such nucleoside analogs are relatively unexplored and may be a source of future drugs. This is facilitated by the fact that there are a very large number of alkyne-labeled compounds that could be used in such screening experiments for both large RNAs as well as nucleoside analogs.

Sulfinate salt chemistry is also useful as a tool for basic research of nucleic acid biochemistry. Fluorescent probe attachment can be helpful for tracking nucleic acid interactions on gels and for nucleic acid localization within cells. The attachment of affinity probes, such as biotin, can make any nucleic acid purification simple and is also useful for identification of specific nucleic acid-protein interactions. A variety of other functional groups can be added to nucleic acids, such as sugars or glycans, amino acids, magnetic beads, gold nanoparticles, etc. In this regard, it has been recently reported that some RNAs are glycosylated within cells (28). Sulfinate chemistry could be used to generate large quantities of glycosylated RNAs to probe this new class of modified RNAs and their biological function.

In summary, the modification of RNA and DNA with sulfinate salts is a simple and inexpensive method for the addition of novel functional moieties to nucleobases. Recently, effective mRNA vaccines have been developed for the SARS-CoV-2 virus that contain modified nucleosides (29,30). This method enables a diverse set of nucleic acid modifications and in contrast to several other modification methods, all types of nucleic acids can be substrates.

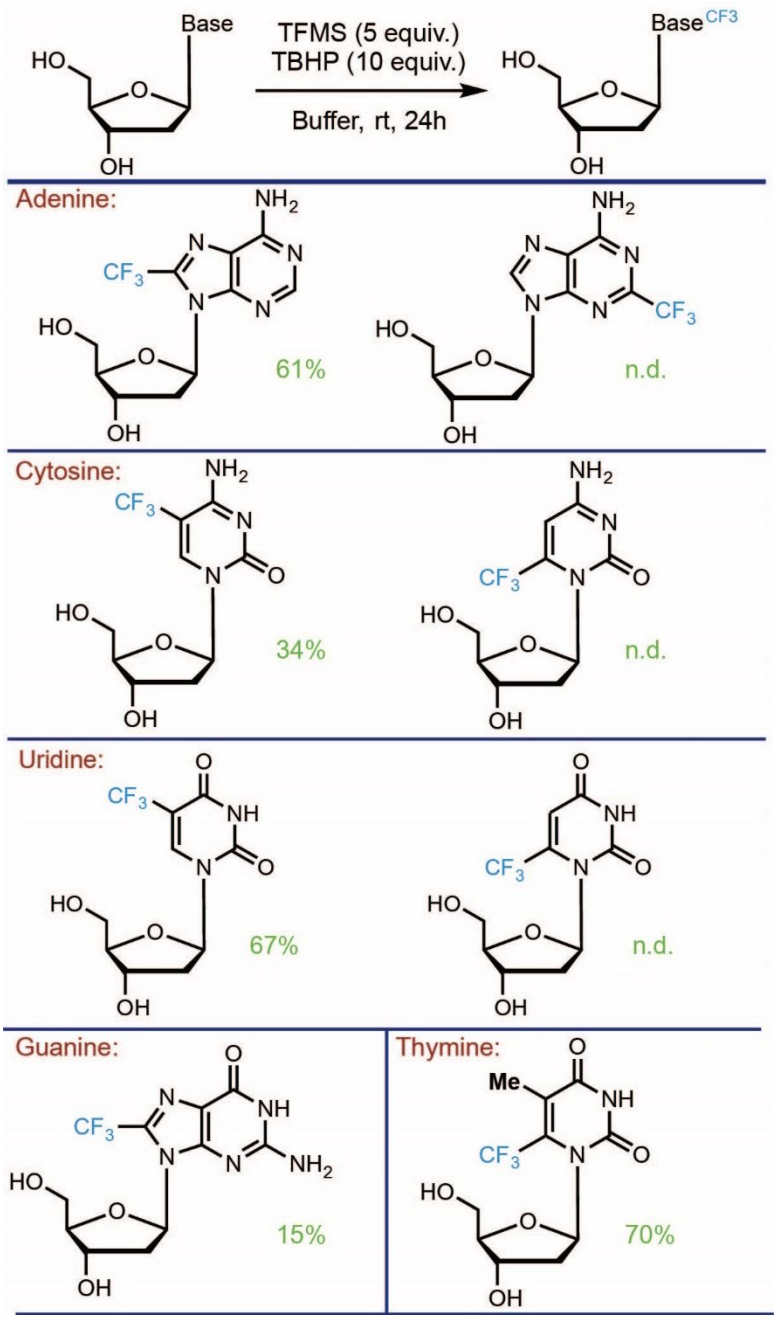
Acknowledgements: Chapter 1 has been submitted for publication as it may appear in Nature Chemical Biology, 2020. Hirlinger, Anastassia; Vantourout, Julien; Grayson, Leah; Toor, Navtej; Springer Nature, 2020. The dissertation author was the primary investigator and author of this paper.

## Supplementary Information

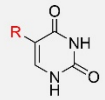
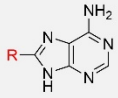
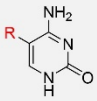
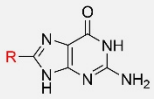


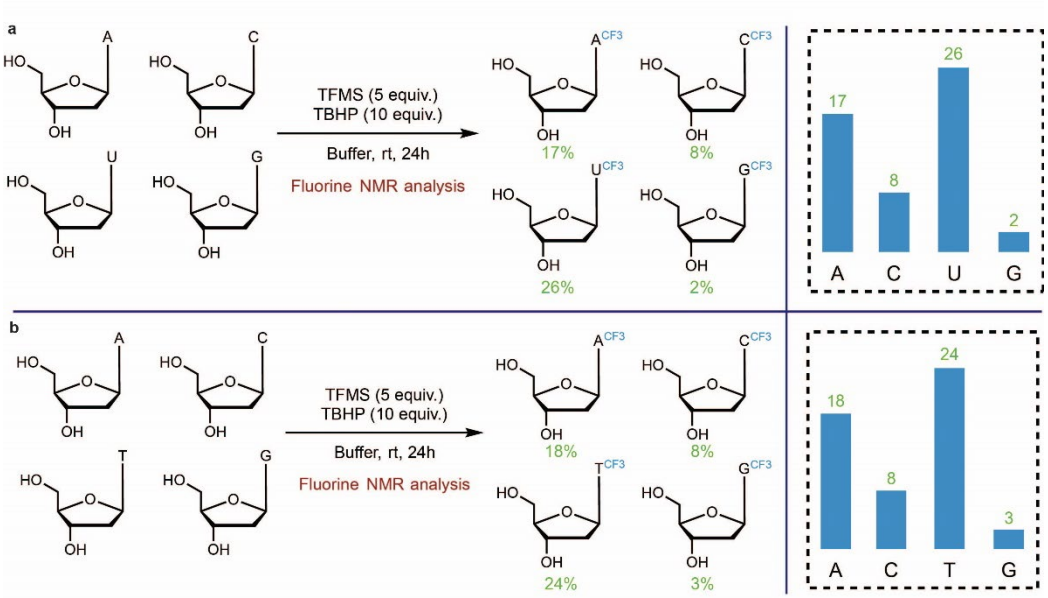

**Supplementary Figure 1.1: Modified RNA is intact post-treatment.** *O.i.* RNA was treated with sulfinate salts in 8-hour modification reactions and run on a denaturing polyacrylamide gel. The control did not have any sulfinate salt added to the reaction.

**Supplementary Figure 1.2: Nucleobase modification sites evaluated by fluorine-19 NMR.** Nucleosides were modified by TFMS in separate reactions and fluorine-19 NMR was used to identify the sites of trifluoromethylation on each nucleobase. The $CF_3$-thymine $^{19}F$ NMR yield was 70% and the $^{19}F$-NMR chemical shift was δ -63.1 ppm (376 MHz, $CDCl_3$). The $CF_3$-adenine $^{19}F$ NMR yield was 61% and the $^{19}F$-NMR chemical shift was δ -62.4 ppm (376 MHz, $CDCl_3$). The $CF_3$-cytosine $^{19}F$ NMR yield was 34% and the $^{19}F$-NMR chemical shift was δ -61.9 ppm (376 MHz, $CDCl_3$). The $CF_3$-guanine $^{19}F$ NMR yield was 15% and the $^{19}F$-NMR chemical shift was δ -62.1 ppm (376 MHz, $CDCl_3$). The $CF_3$-uridine $^{19}F$ NMR yield was 67% and the $^{19}F$-NMR chemical shift was δ -62.7 ppm (376 MHz, $CDCl_3$).

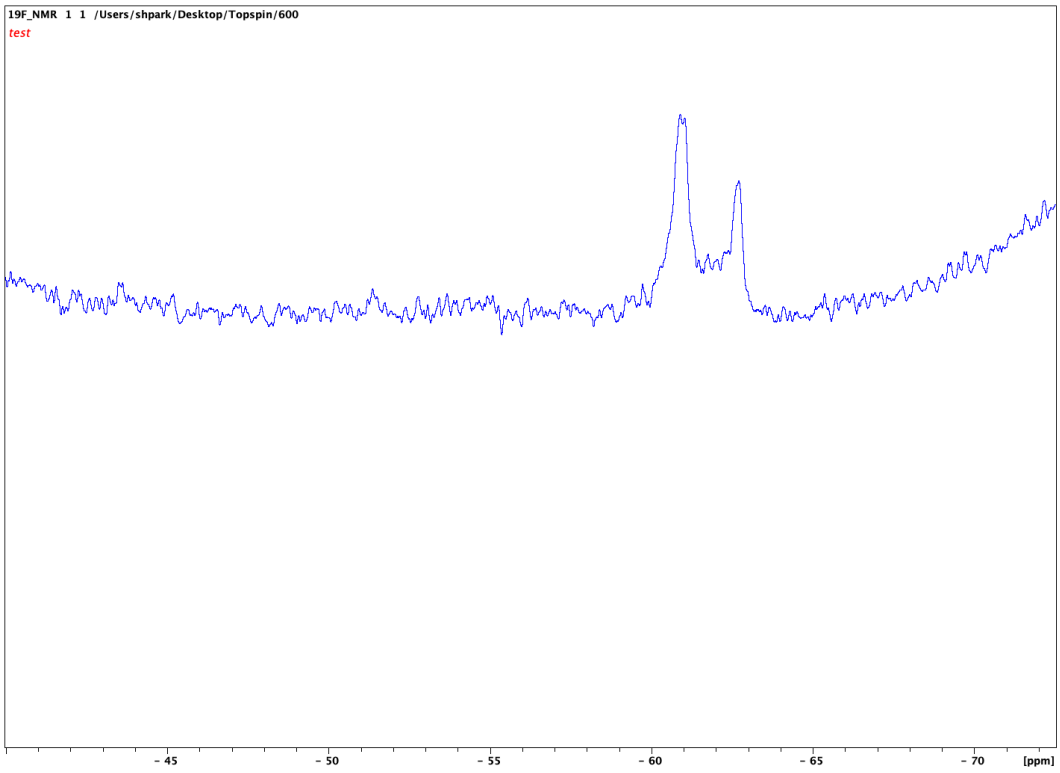

Base
HO
O
OH
TFMS (5 equiv.)
TBHP (10 equiv.)
Buffer, rt, 24h
Base CF3
Adenine:
NH2
CF3
61%
n.d.
Cytosine:
34%
n.d.
Uridine:
67%
n.d.
Guanine:
15%
Thymine:
Me
70%

**Supplementary Table 1.1: Nucleoside modification efficiency reported by LC-MS.** *O.i.* RNA was modified in separate reactions with sulfinate salts for 48 hours and then enzymatically digested into nucleotides and dephosphorylated. The results from the LC-MS analysis on the modified ribonucleosides mixtures is shown in the table. Percentages indicate the number of nucleosides with the expected modification out of total nucleosides (modified and unmodified) and are averages from three replicate experiments.

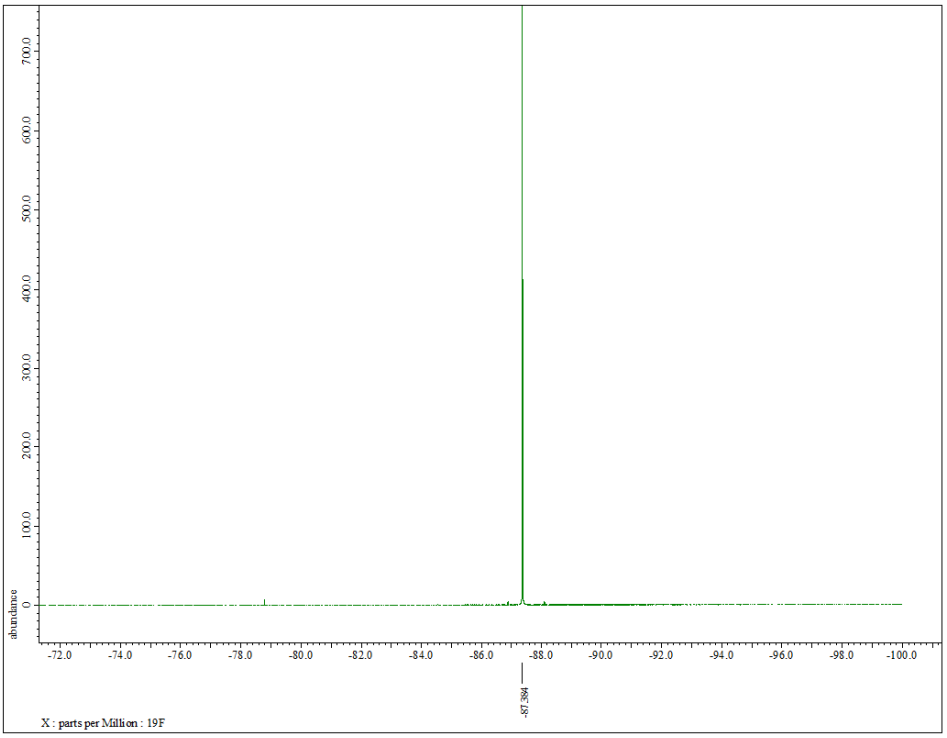
| Nucleobase | Sulfinate Salt | | | |
|---|---|---|---|---|
| | TFMS | IPS | PSMS | DAAS |
| Uracil  | 28.1% | 21.5% | 20.2% | 12.5% |
| Adenosine  | 21.0% | 16.2% | 12.3% | 9.9% |
| Cytosine  | 4.2% | 10.0% | 3.9% | 5.5% |
| Guanine  | 2.0% | 1.0% | 1.2% | 1.3% |




**Supplementary Figure 1.3: Competitive modification between different nucleosides.** Fluorine-19 NMR was used to evaluate the relative levels of TFMS modification on a mixture of DNA nucleosides. For each nucleobase, the percentage of the total found to be modified is shown in a bar graph on the right. The chemical shifts for $CF_3$-modified nucleosides is the same as in Supplementary Figure 1.2. **a**, Competitive TFMS modification between A, C, G and U nucleosides. **b**, Competitive TFMS modification between A, C, G and T nucleosides.




**Supplementary Figure 1.4: Fluorine-19 NMR spectrum for a large trifluoromethyl-modified RNA.** The 622-nucleotide *P.li.* RNA was modified with TFMS and analyzed by fluorine-19 NMR. Chemical shifts are observed at -60.9 ppm and -62.7 ppm.

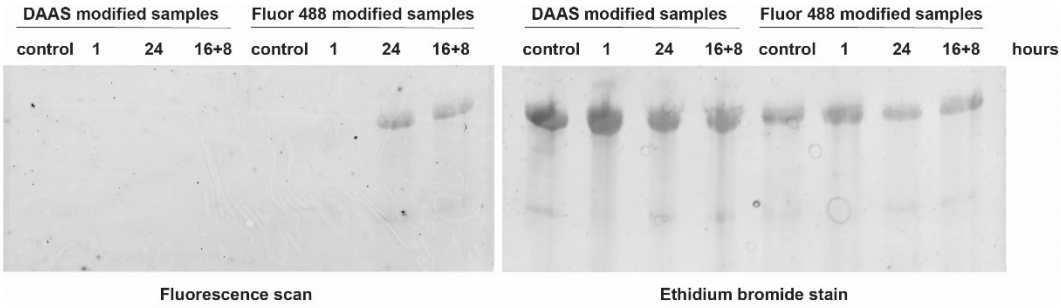



**Supplementary Figure 1.5: Fluorine-19 NMR spectrum for zinc trifluoromethanesulfinate.** The fluorine-19 NMR spectrum for zinc trifluoromethanesulfinate (TFMS) shows an observed chemical shift at -87.4 ppm.

**Supplementary Table 1.2: Oligonucleotide modification efficiency reported by LC-MS.** 20-nucleotide oligos were modified with sulfinate salts and analyzed by LC-MS. RNA and DNA oligos have the same sequence.

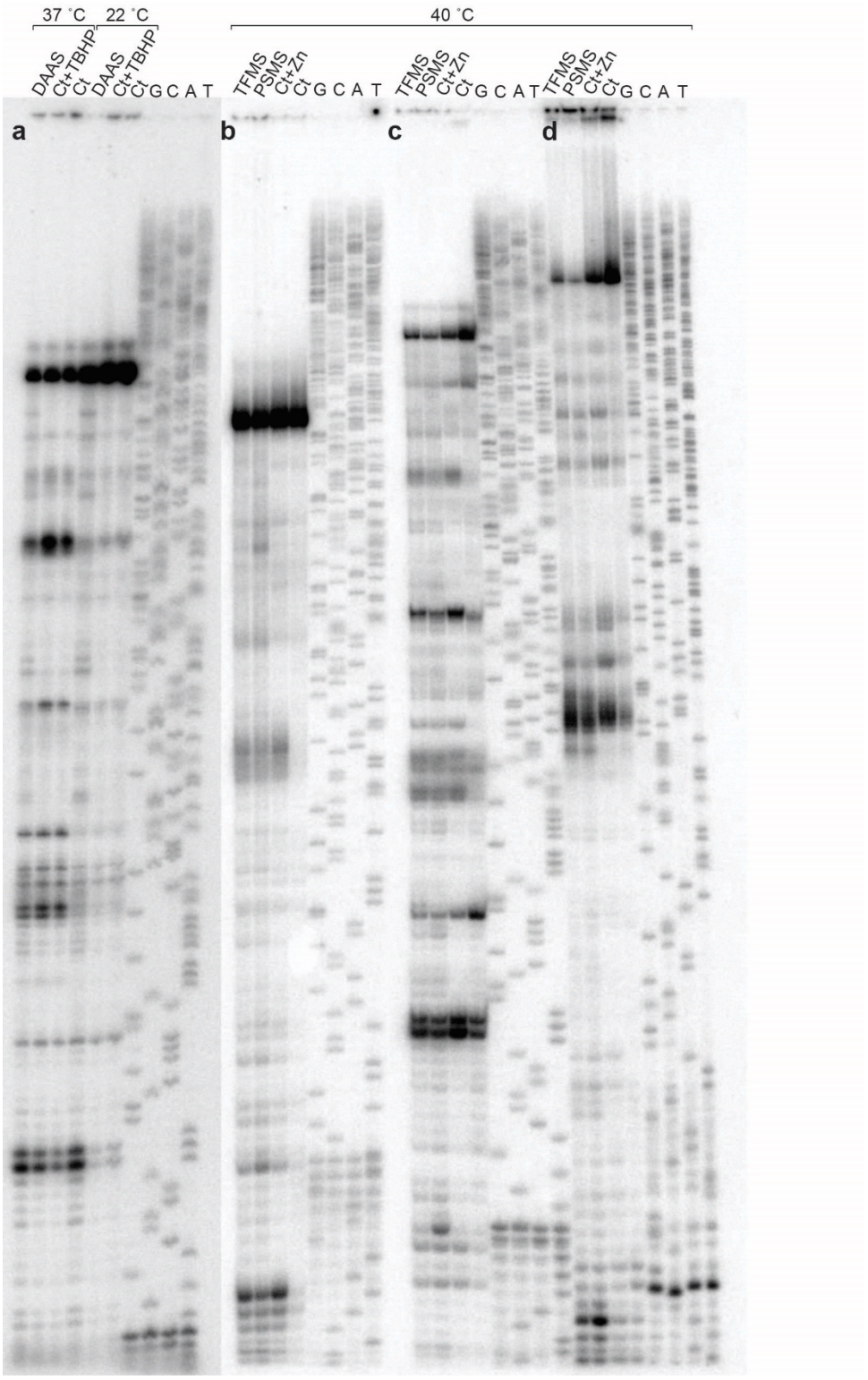
| TFMS modified ssDNA | PSMS modified ssDNA |
|---|---|
| 23.1% 1 modification per oligo  | 23.4% 1 modification per oligo  |
| 20.6% 2 modifications per oligo  | 11.7% 2 modifications per oligo  |
| Overall 43.7% oligos modified | Overall 35.1% oligos modified |

| DAAS modified ssDNA | DAAS modified ssRNA |
|---|---|
| 4.0% 1 modification per oligo  | 4.1% 1 modification per oligo  |
| 2.7% 2 modifications per oligo  | 1.8% 2 modifications per oligo  |
| Overall 6.7% oligos modified | Overall 5.9% oligos modified |




**Supplementary Figure 1.6: Fluorescent RNA produced from copper-catalyzed azide-alkyne cycloaddition.** The *O.i.* RNA was modified with DAAS in a first step reaction and then with Fluor 488-alkyne in a second step CuAAC reaction. The control reaction components did not contain DAAS. The time refers to the hours of DAAS modification of the RNA, where the 16 + 8 sample had a second addition of DAAS reagent at the 16-hour mark and the reaction proceeded for another 8 hours. Right and left images show the same 4% denaturing polyacrylamide gel: left, a fluorescence scan of the gel and right, an ethidium bromide stain of the gel.

**Supplementary Figure 1.7: Primer extension assays on *O.i.* RNA.** Four primer extension assays are shown for the *O.i.* RNA modified with DAAS (**a**), TFMS and PSMS (**b-d**). **a** and **c** assays used the same primer while **b** and **d** assays used different primers. G, C, A, and T reactions are sequencing reactions on the *O.i.* plasmid DNA. **a**, *O.i.* RNA modified with DAAS was compared to controls with zinc chloride (Ct) and with zinc chloride and TBHP (Ct+TBHP). The sulfinate salt experimental and control reactions were incubated at 37° or 22 °C. **b**, *O.i.* RNA modified with TFMS and PSMS was compared to controls with zinc chloride (Ct+Zn) or without zinc chloride (Ct). The sulfinate salt experimental and control reactions were incubated at 40 °C.


37 °C
22 °C
40 °C
DAAS
Ct+TBHP
Ct
G C A T
TFMS
PSMS
Ct+Zn
a
b
c
d

## Methods

*Modification of nucleic acids with zinc sulfinate salts*

Before starting the modification, I prepared RNA substrates in 50 mM MES pH 6.5 (optionally with 10 mM $MgCl_2$). DNA substrates were usually in water or in 10 mM Tris HCl pH 7.5 and 0.1 mM EDTA. The following reaction components were added together (final concentrations): RNA or DNA at 1 mg/mL, 10 mM zinc sulfinate salt (TFMS, PSMS, or IPS), 50 mM MES pH 6.5, 10 mM $MgCl_2$ and water [minus the volume of 70% tert-butyl hydroperoxide (TBHP)]. The reaction tube was placed on ice for at least 2 minutes prior to the addition of TBHP. 70% TBHP (*aq*) was added to the reaction at a final concentration of approximately 15 mM. The reaction tube was kept on ice for at least 2 minutes after TBHP was added and was then moved to room temperature (22 °C) for 24 hours. The reaction was stopped by filtering out TBHP and salts from the nucleic acid with buffer exchange on centrifugal filters with a 30 kDa or 50 kDa molecular weight cut off (MWCO) membrane. I used these centrifugal filters for large RNAs like the *Oi.* group II intron. The solution used in the buffer exchange was 50 mM MES pH 6.5 or 50 mM MES pH 6.5 and 10 mM $MgCl_2$. To purify the 20-mer DNA and RNA oligos

I used Mini Quick Spin RNA Columns or Mini Quick Spin Oligo Columns (Roche) and followed with ethanol precipitation. Small nucleic acid samples were resuspended in water for mass spectrometry analysis.

*Modification of nucleic acids with DAAS*

Before starting the modification protocol, I prepared RNA substrates in 50 mM MES pH 6.5, 10 mM $MgCl_2$ buffer. DNA substrates were in water. The first step reactions contained 10 mM DAAS, 50 mM MES pH 6.5, 10mM $MgCl_2$, 4.6 mM $ZnCl_2$, 1 mg/mL nucleic acid and water to the final volume (minus that of TBHP). The reaction tube was placed on ice for at least 2 minutes prior to the addition of TBHP. 70% TBHP (*aq*) was added to the reaction at a final concentration of approximately 15 mM. The reaction tube was kept on ice for at least 2 minutes after TBHP addition. The reaction tube was moved to the lab bench to stay at room temperature (22 °C) for 24 hours. The first step reaction was stopped either by buffer exchange with centrifugal filters or by purification with Mini Quick Spin Columns (Roche) followed by ethanol precipitation. Large nucleic acids used centrifugal filters with a 30 kDa MWCO (Corning) or 50 kDa MWCO (Millipore) and small nucleic acids (20-mer oligos) were

purified with mini spin columns and ethanol precipitation. The solution used in the buffer exchange was 50 mM MES pH 6.5 and 10 mM $MgCl_2$.

*Modification with copper (I)-catalyzed azide alkyne cycloaddition (CuAAC) click chemistry*

Each reaction contained components at these final concentrations: 1uM of an alkyne-labeled molecule (concentrations as low as 15 nM also worked), 20% DMSO, 6 mM THPTA, 6 mM sodium ascorbate, 3 mM $CuSO_4$ and a DAAS-modified nucleic acid at 1 mg/mL. Control reactions contained an unmodified nucleic acid instead of DAAS-modified nucleic acid. The reaction tubes were placed in a 37 °C water bath for 30 minutes. For large nucleic acids, the reaction was stopped by applying the samples to 30 kDa or 50 kDa MWCO centrifugal filter columns and purifying by buffer exchange with 50 mM MES pH 6.5 (optionally with 10 mM $MgCl_2$) buffer. Small nucleic acids were purified with Roche Mini Quick Spin Columns followed by ethanol precipitation. The modified nucleic acids were resuspended in either water, 50 mM MES at pH 6.5, 10 mM $MgCl_2$ buffer or 50 mM Tris HCl pH 7.5, 0.1 mM EDTA buffer. Fluorescence scans were made on a Typhoon FLA 9500 using the 495 nm excitation wavelength, 519 nm emission wavelength and LPG filter.

*Modification with strain-promoted azide alkyne cycloaddition (SPAAC)*

Each 40 µL reaction contained 15.7 nmol of DBCO-PEG4-Fluor 545, 1X PBS pH 7.4 and a DAAS-modified nucleic acid at a final concentration of 1 mg/mL. Control reactions contained an unmodified nucleic acid instead of DAAS-modified nucleic acid. The reaction tubes were at room temperature (22 °C) for a total of 19 hours. The reaction was stopped by applying the samples to 30 kDa MWCO centrifugal filter columns and purifying by buffer exchange with 50 mM MES pH 6.5, 10 mM $MgCl_2$ buffer. Fluorescence scans were made on a Typhoon FLA 9500 using the 542 nm excitation wavelength, 568 nm emission wavelength and LPG filter.

*Sample prep for electrospray ionization mass spectrometry*

After RNA was modified it was digested into nucleotides and dephosphorylated. In a 25 µL reaction volume, 5 µg of RNA was added to 0.5 µL nuclease P1, 25 mM NaCl, and 2.5 mM $ZnCl_2$. The reaction was incubated at 37° C for 2 hours and then 3 µL of 1 M $NH_4HCO_3$ and 0.5 µL of shrimp alkaline phosphatase (rSAP) were added. This reaction was incubated at 37° C for 2 hours and then diluted with $H_2O$ to a 50 µL final volume. The modification status of the ribonucleosides were

next analyzed with electrospray ionization mass spectrometry (ESI-MS) in positive ion mode. A Waters I-Class HPLC was used with a Waters BEH C18 column (2.1 x 55 mm, 1.7 µm, 130 Å) and a gradient of 114 mM hexafluoroisopropanol and 14 mM triethylamine in water (A) and acetonitrile (B) (0.3 mL/min, 10-90% B over 10 minutes) at 60 °C.

*Fluorine-19 NMR analysis of individual $CF_3$-modified deoxynucleosides*

A solution of a DNA building block (0.1 mmol, 1 equiv, 0.1 M) and TFMS (0.5 mmol, 5 equiv) in Tris HCl pH 7.5 (containing 100 mM EDTA) was stirred at room temperature for 5 minutes, followed by a slow addition of tert-butyl hydrogen peroxide (70% solution in water, 1 mmol, 10 equiv) by an eppendorf pipette (metal needles should not be used as they decompose the reagent) with vigorous stirring. The reaction was stirred at room temperature for 24 hours and analyzed by $^{19}F$ NMR using hexafluorobenzene as an internal standard. The regiochemistry is consistent with that observed in previous studies reported in the literature for trifluoromethylation of nucleobases in harsh organic solvents (31,32).

*Competitive fluorine-19 NMR analysis of $CF_3$-modified deoxynucleosides*

A solution of the four DNA building blocks (4 x 0.025 mmol, 1 equiv, 0.1 M) and TFMS (0.125 mmol, 5 equiv) in Tris HCl pH 7.5 (containing 100 mM EDTA) was stirred at room temperature for 5 minutes, followed by a slow addition of tert-butyl hydrogenperoxide (70% solution in water, 0.25 mmol, 10 equiv) by an eppendorf pipette (metal needles should not be used as they decompose the reagent) with vigorous stirring. The reaction was stirred at room temperature for 24 hours and the ratio was analyzed by $^{19}F$ NMR.

*RNA sample prep for fluorine-19 NMR*

The modified RNA was highly concentrated in a solution of 50 mM MES pH 6.5, 10 mM $MgCl_2$, and 10% $D_2O$ was added such that the final concentration was at least 10 mg/mL RNA. The RNA fluorine-19 NMR data was captured with a 300 MHz Bruker.

*In vitro transcription of RNA*

Transcription reactions contained the following components: 40 µg linearized plasmid DNA, 40 mM Tris HCl pH 7.5, 25 mM $MgCl_2$, 2 mM

spermidine, 2.5 mM NTPs, 0.05% Triton X-100, 5 mM dithiothreitol, 5 µL T7 polymerase, 0.05 units of inorganic pyrophosphatase and water to a 1 mL final volume. Transcription reactions proceeded at 37 °C for a minimum of 3 hours and reactions were stopped with addition of 12 µL of 0.1 M $CaCl_2$ and 20 units of TURBO DNase (Invitrogen). Incubation with DNase proceeded for approximately 40 minutes at 37 °C. 6 units of Proteinase K are added to reactions and the reactions were incubated at 37 °C for an hour. Transcription reaction tubes were spun down for 5 minutes at 21,130 RCF and the supernatant was filtered through a 0.2 µm membrane. The RNA was purified by buffer exchange in Amicon Ultra centrifugal filter units (Millipore-Sigma).

*In vitro translation of desthiobiotin-modified GFP mRNA in HeLa lysate*

125 µg of desthiobiotin-modified GFP mRNA was added to 50 µL of streptavidin coated agarose resin (100 µL of the 50% suspension) in 50 mM Tris HCl pH 7.5, 0.1 mM EDTA buffer. In a binding step, desthiobiotin-GFP mRNA was left on the resin for 20 minutes at room temperature. The resin was washed 9 times in batch format with 1 mL washes of 50 mM Tris HCl pH 7.5, 300 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 buffer. I added 50 µL elution buffer (8 mM biotin, 50 mM Tris HCl pH 7.5, 0.1 mM EDTA)

to the washed resin and this elution step proceeded at room temperature for 30 minutes. A second and third addition of elution buffer were added to the resin and all eluate was combined after measuring the RNA concentration. The eluted mRNA was ethanol precipitated and resuspended in water.

I used the Thermo Scientific 1-Step Human Coupled IVT Kit – DNA for the in vitro translation reactions of the desthiobiotin-modified GFP mRNA. I transcribed the pCFE-GFP DNA provided in this kit into RNA and modified it with desthiobiotin using DAAS in a first step reaction and desthiobiotin-alkyne in a CuAAC second step reaction. Each in vitro translation reaction contained components from the IVT kit, including 12.5 µL of HeLa lysate, 2.5 µL accessory proteins and 5 µL reaction mix. 15 nM of the unmodified GFP mRNA (167 ng) was added to control reactions and water was added to the final volume, 25 µL. 167 ng of the purified desthiobiotin-modified GFP RNA was added to experimental reactions. All reactions were incubated at 30 °C for 90 minutes, transferred to a microplate and had fluorescence measured by the Tecan Spark microplate reader. Fluorescence signals were recorded every 5 minutes using excitation wavelength 485 nm and emission wavelength 535 nm. This experiment was performed in triplicate.

*Primer extension assays*

The sulfinate salt reactions for primer extension (Figure 1.5) included a splicing buffer component (100 mM $MgCl_2$, 500 mM $NH_4Cl$, 50 mM Tris HCl pH 7.5) along with the standard conditions for either the zinc sulfinate salts (TFMS and PSMS) or for the sodium sulfinate salt DAAS (see method descriptions). These reactions were incubated at 40 °C for 24 hours. The control reactions included the RNA and splicing buffer.

γ-ATP labeled primers (5000 cpm) were incubated with the modified RNA, DTT, dNTPs, Superscript III reverse transcriptase and 5X First Strand Buffer for 60 minutes at 55 °C, then at 70 °C for 15 minutes. Sequencing reactions included the *O.i.* plasmid DNA, Affymetrix Thermo Sequenase Polymerase, Thermo Sequenase Reaction Buffer, γ-ATP labeled primer and one of the ddNTP termination mixes (either ddGTP, ddATP, ddTTP or ddCTP). The sequencing reactions followed a cycling program with 40 cycles of 30 seconds at 95 °C, 30 seconds at 55 °C and 2 minutes at 72 °C. I loaded approximately 600 cpm of the reverse transcription reactions and 1000 cpm of the sequencing reactions on a denaturing 8% acrylamide gel. The resulting gel was exposed to a phosphor screen overnight.

# CHAPTER 2

## Introduction

Viroids are autonomous, pathogenic RNAs that do not encode for proteins or have a DNA intermediate and yet cause devastating plant diseases which affect food crops such as peach, apple and avocado, among others (33,34). They also have the distinction of being the smallest known infectious agents (35). As nutrition becomes increasingly important for the world's growing population, the importance of combating food crop diseases will increase as well. Plants infected with viroids are generally disfigured with a stunted and mottled phenotype (Figure 2.1). An exceptional example of viroid pestilence occurred in the 1950s, when viroids were responsible for nearly eliminating the chrysanthemum industry (36).


**Figure 2.1: Viroid infected food crops. a,** A healthy Russet Burbank potato and right, viroid infected potatoes (37). **b,** A healthy tomato plant is on the far left and to its right are increasingly stunted plants infected with severe strains of the potato spindle tuber viroid (34).

More than 30 species of viroids are known, and they are organized into two families: *Pospiviroidae* and *Avsunviroidae*. The *Pospiviroidae* family has a central conserved region, they typically have rod-shaped secondary structures, replicate by the asymmetric rolling circle mechanism in the nucleus and are not known to exhibit ribozyme activity. The peach latent mosaic viroid (PLMVd) and the chrysanthemum chlorotic mottle viroid (CChMVd), subjects of my study, belong to the *Avsunviroidae* family. Members of the *Avsunviroidae* family have no central conserved domain, have highly branched secondary structures and contain hammerhead ribozyme within their sequences. *Avsunviroidae* viroids replicate in the chloroplast by the symmetric rolling circle mechanism, in which the hammerhead ribozyme cleaves the concatemeric sequence into monomeric sequences (Figure 2.2). Previous evidence has suggested certain enzymes in replication roles, but there is little direct support for the identity of enzymes involved in PLMVd and CChMVd replication (33,38).

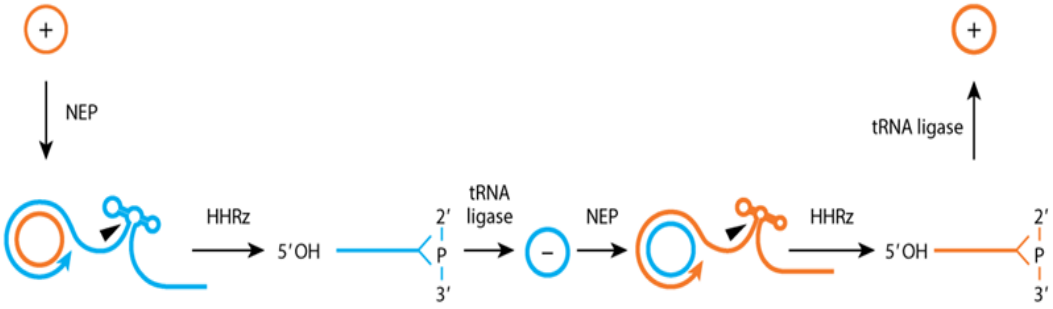



**Figure 2.2: Mechanism of replication for the *Avsunviroidae* family.** The symmetric rolling-circle mechanism of replication, proposed for the *Avsunviroidae* family (39). Circular positive polarity RNA is proposed to be transcribed by the nuclear encoded polymerase (NEP) into a concatemeric negative polarity RNA that self-cleaves by the hammerhead ribozyme. A linear negative polarity RNA with a 2',3'-cyclic phosphodiester terminus results which is proposed to be ligated and circularized by tRNA ligase. The same enzymes would then process the circular negative polarity RNA into the circular positive polarity RNA.

Only a few viroid RNA secondary structures have been determined and of those, two are known to contain higher order tertiary interactions: the peach latent mosaic viroid, PLMVd, and the chrysanthemum chlorotic mottle viroid, CChMVd (35,40,41). More importantly, biochemical evidence has suggested that forming the proper tertiary structure is essential for infectivity (41,42). For both CChMVd and PLMVd, sequences with mutations disrupting pseudoknot formation were unable to infect plants whereas sequences that maintained the pseudoknot were able to infect their respective plants. In other words, the viroid RNA forms an ordered three-dimensional structure that is responsible for its pathogenicity. Solving the 3D structure of a viroid would provide valuable insight into the relationship between its structure and infectivity and would also represent the first structure of any viroid.

Structural characterization of a viroid provides a framework with which I could test the importance and role of conserved RNA tertiary contacts in viroid infectivity. Currently, little is understood about the mechanism of viroid infection and there are few treatments with limited success to prevent these diseases (43). The following aims were to be addressed in the course of my research: (1) determine a high resolution three-dimensional structure of a viroid, (2) test the effects of RNA tertiary

interactions upon folding and (3) upon plant infection; and (4) determine the identity of protein partners of the viroid RNA *in vivo*.

The tertiary contacts observed in the viroid structure would be tested for their effects upon RNA folding using mutagenesis and native gel assays. Viroid sequences with tertiary interactions inactivated would be assayed for infectivity in plants to investigate the relationship between RNA structure and pathogenesis. The expectation is that abolishing tertiary contacts which are important for viroid replication and pathogenesis would both destabilize folding of the RNA and prevent widespread infection within the plant. This would provide insight into regions of the RNA structure required for viroid infection and further elucidate the disease mechanism. This information can serve as a general model for the relationship between structure and function in all viroids and their resulting diseases. The structure-infectivity relationship could potentially unveil a mechanism of inactivation by targeted small molecules or anti-sense RNA therapy that would be tested on viroid infected food crops.

The mechanism of viroid infection is largely unknown; however, viroid RNAs must interact with multiple proteins within plant cells for replication, infection, and pathogenesis. I could isolate and capture viroid protein partners by attaching a MS2 tag on viroid RNA and using mass

spectrometry to determine the identities of associated proteins. The MS2 tag is a tandem hairpin RNA sequence that tightly binds the MS2 protein and can be engineered on to RNAs of interest for an affinity tag pull down of associated proteins (44). In particular, identification of the RNA polymerase and any accessory proteins are of interest, as well as the presence or absence of an RNA ligase. Uncovering the identity of PLMVd associated proteins provides insight into viroid replication and pathogenesis and the processes involved should have parallels to a majority of known viroids. PLMVd associated proteins could then be used in subsequent co-crystallization trials with the RNA for structure determination.

The long-term goal of this investigation was to find new ways to prevent viroid infection in plants, therefore contributing to better crop yields, which has a direct impact on human health. To accomplish my aims, I used several viroid sequences with both x-ray crystallography and cryo-electron microscopy (cryo-EM). I employed a native purification procedure, in conjunction with *in vitro* T7 transcription, to prepare large quantities of viroid RNA for crystallization trials. As for cryo-EM, I made several new viroid ribonucleoprotein constructs and extensively optimized a purification procedure to prepare samples for screening by negative stain EM. I acquired GF305 *Prunus persica* seedlings, a peach viral

indicator cultivar, for the in vivo aims of my investigation; however, I did not find a growth chamber space where I could grow the plants. Viroid RNA aggregation was a persistent issue in my attempts to solve a structure with x-ray crystallography and cryo-EM and I also had difficulty getting time on a microscope for screening EM samples. Ultimately the combined roadblocks led to the termination of this investigation.

## Results

### Constructs of PLMVd.034 and PLMVd.282 for crystallization

Before any fusion constructs were designed, I attempted to solve the structure of two variants of PLMVd, PLMVd.034 and PLMVd.282, by x-ray crystallography. Sequence names of the variants are from the Subviral RNA database (45).

The PLMVd.282 sequence has a cruciform structure on its longest paired stem, P11, which distinguishes it from PLMVd.034 (Figure 2.3). The design of the PLMVd.282 construct for crystallization removed the lower P11b loop sequence so that a T7 RNA Polymerase promoter sequence could be added at the 5’ end and a restriction site could be

added on the 3’ end. For another PLMVd.282 construct, I engineered crystal contacts into P11 by adding a GNRA tetraloop- receptor. GNRA tetraloop-receptors are RNA folding motifs that are used in crystallography to encourage ordered crystal packing, which can help with crystal formation (46). A GAAA tetraloop was added to the upper stem, P11a, and the tetraloop receptor was added to the lower stem, P11b. In order to increase the chance that the receptor and tetraloop reach each other, it’s important to vary the length of the paired stems that carry the receptor and loop sequences so several different constructs are made from a GNRA tetraloop-receptor template in crystallization trials.

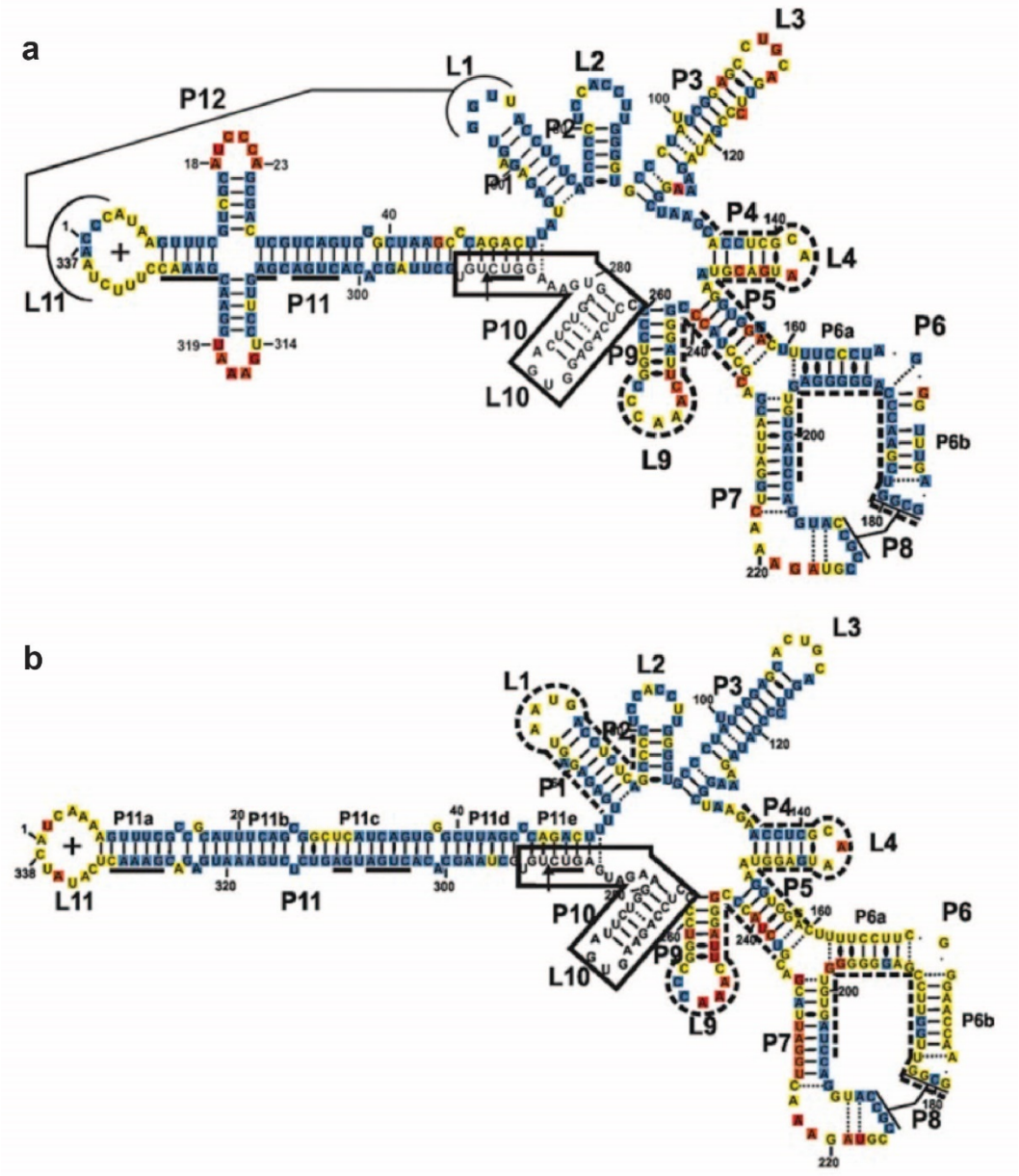



**Figure 2.3: Secondary structures of PLMVd.282 and PLMVd.034.** Proposed secondary structures for the positive polarity of PLMVd variants PLMVd.282 (**a**) and PLMVd.034 (**b**), as indicated by SHAPE data(40). The nucleotides marked blue are not accessible, the nucleotides marked yellow are intermediately accessible, and those in red are the most accessible.

As for PLMVd.034, I designed a construct which had L11 removed so that the T7 promoter and an enzyme restriction site could be added at the end of P11. To this same construct I added a GAAA tetraloop to L3 of PLMVd.034 and the receptor to P4. For reasons I'll discuss next, no other PLMVd.034 constructs were made for crystallization trials.

In order to get a crystal structure, an RNA must to be soluble at a high concentration. The in vitro transcription protocol I use to make RNA is designed to produce highly concentrated RNA by first making milligram amounts of RNA and then by using concentrators to get the RNA into as small a volume as possible. Last, I used a crystallization robot to set crystal trays of my highly concentrated RNA. After transcribing the viroid RNA but before concentrating it, I noticed hydrogel formation in the tubes containing the RNA. The hydrogel was an indication of RNA aggregation and the RNA appeared to aggregate more as I tried concentrating it and if there was $Mg^{2+}$ in the solution (Figure 2.4). Changing the buffer solution I normally used in the concentration step to one containing no magnesium and including EDTA reduced some of the aggregation but not nearly enough for crystallization purposes. I aimed to concentrate the RNA to near 10 mg/mL or higher to set crystallization plates of a construct and the PLMVd constructs rarely concentrated higher than 1 mg/mL RNA. Possibly due to the aggregation of the viroid RNA, no x-ray

crystallography structure resulted from my efforts. Another possible explanation for why the RNAs did not crystallize is that the viroid structure is highly dynamic and flexible, which I attempted to address with the design of group II intron-viroid chaperone constructs.

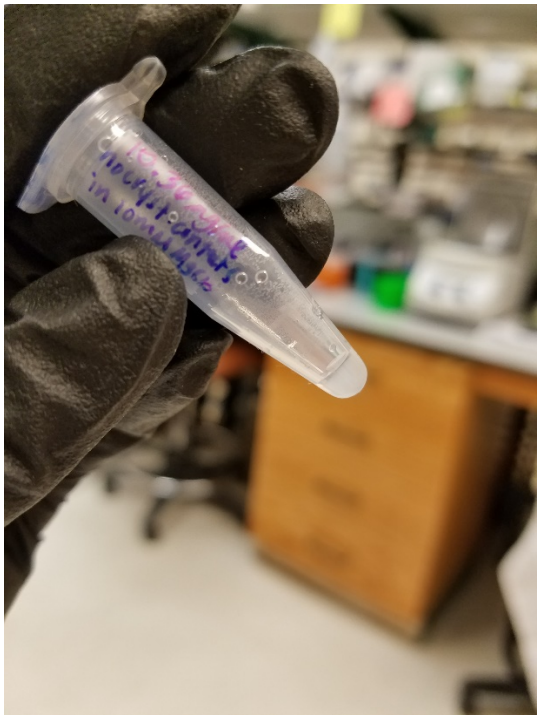


**Figure 2.4: PLMVd.282 hydrogel formation when in solution with $MgCl_2$.** PLMVd.282 RNA at 10 mg/mL in solution with 10 mM $MgCl_2$ formed an insoluble hydrogel.

**Group II intron-viroid chaperone constructs for crystallization**

My efforts to solve the structure of a viroid heavily utilized various group II introns as structural chaperones, both with x-ray crystallography and with electron microscopy. Group II introns are large catalytic RNAs that self-splice. The group II intron secondary structure contains six domains, DI – DVI, which are organized around a central hub. I used three group II introns as structural chaperones in my research, each one isolated from a different bacterium and are so named: *Oceanobacillus iheyensis* (*O.i.*), *Pylaiella littoralis* (*P.li.*) and *Thermosynechococcus elongatus* (*T.el.*) (Figure 2.5). Members of the Toor lab have previously solved the structure of these group II introns using x-ray crystallography and cryo-electron microscopy (cryo-EM) (47–49) so it was relatively simple to design viroid fusion constructs.

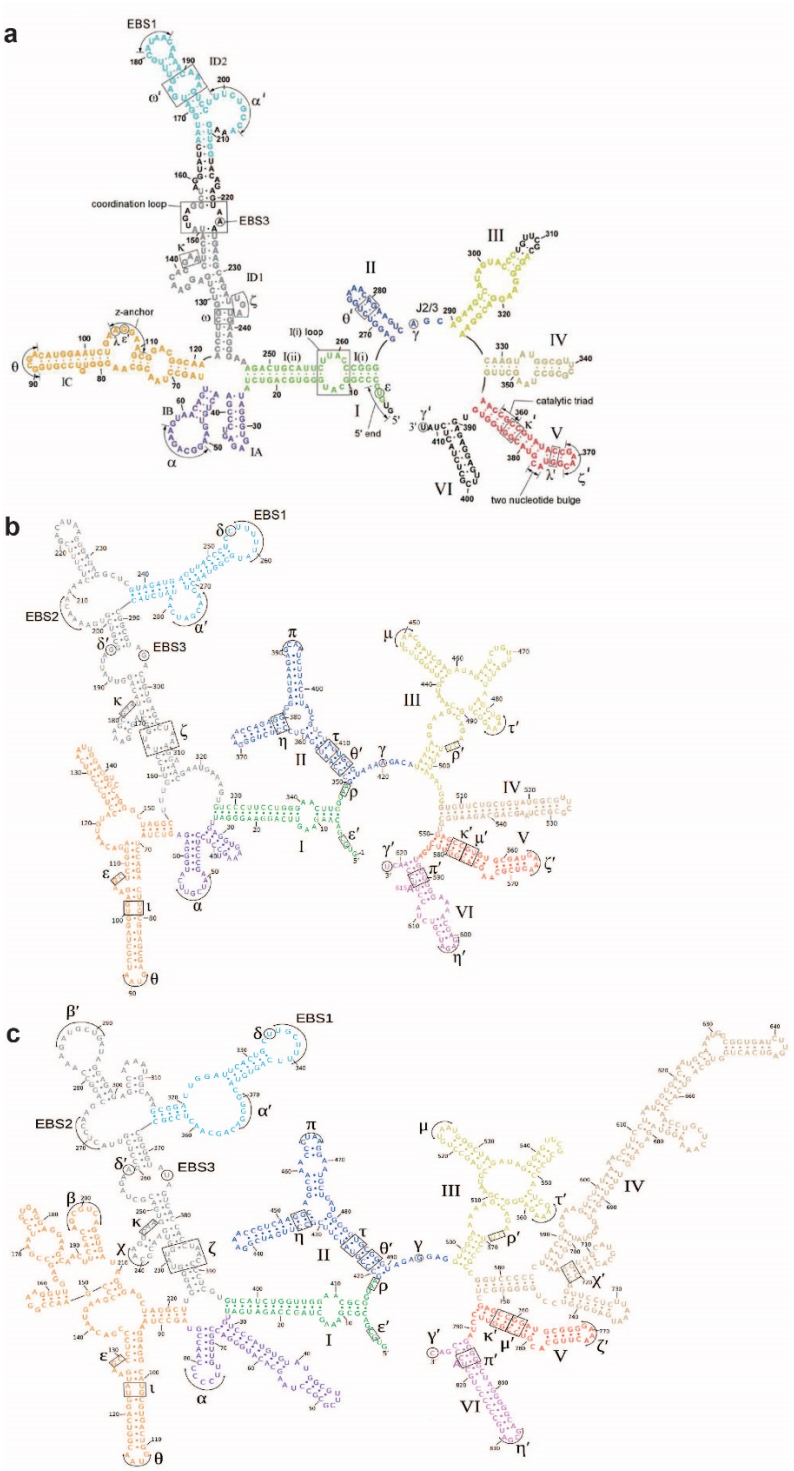



**Figure 2.5: Secondary structures of group II introns *O.i., P.li.* and *T.el.* a**, *O.i.* secondary structure. **b**, *P.li.* secondary structure. **c**, *T.el.* 3c secondary structure.

To address the possibility that viroids have a flexible structure, I decided to use group II introns *O.i.* and *P.li.* as chaperones for crystallization. The ordered packing of group II intron RNAs creates a greater number of crystal contacts and encourages uniform crystal packing. Chaperone constructs also provide the opportunity to get a more complete structure of a viroid, since it's unnecessary to excise part of the PLMVd sequence to engineer in a transcription start site and an enzyme restriction site.

I designed two chaperone constructs for crystallization that utilized the GNRA tetraloop-receptor folding motif. The first was the Oi-PLMVd chaperone construct in which I placed a GAAA tetraloop in domain II (DII) and the tetraloop receptor in DIV. The second was Pli-PLMVd, in which I added the GAAA tetraloop to DIa and the receptor to DIV. Both constructs used the PLMVd.034 variant and fused a group II intron paired stem to P3. Oi-PLMVd connected to the viroid P3 on DIII and Pli-PLMVd connected on DII. I usually used a magnesium-containing buffer for concentrating and purifying the RNAs since I was able to get a high concentration of the RNA even with 10 mM $MgCl_2$ present. The presence of magnesium is important in splicing steps, where the group II intron folds into a post-catalysis structure and is therefore more structurally homogeneous. No crystals

resulted from Oi-PLMVd and a few crystals formed for Pli-PLMVd that did not diffract to a high resolution (Figure 2.6).

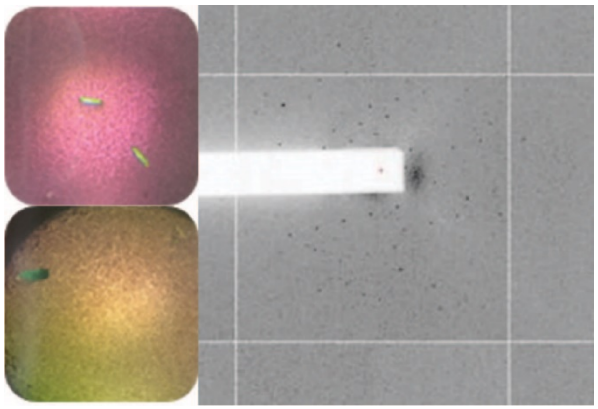


**Figure 2.6: Pli-PLMVd crystals and diffraction pattern.** Left, Pli-PLMVd crystals in different crystallization conditions. Right, 20 Å diffraction pattern for a Pli-PLMVd crystal.

I also attempted to solve the structure of a Pli-PLMVd-U1A ribonucleoprotein, in which a U1A protein binding site was engineered into DIV of *P.li*. The U1A protein tightly binds to a 10-nucleotide loop in a sequence-specific manner ($K_D$ = 2 x $10^{-11}$ M) (50) and can be useful for crystallization of RNAs due to its high number of protein-protein contacts in previous crystal structures (51). Crystals of Pli-PLMVd-U1A nucleated but also did not diffract to a high resolution (Figure 2.7).

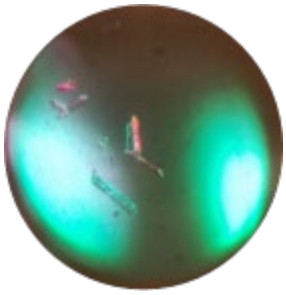


**Figure 2.7: Crystals from the Pli-PLMVd-U1A ribonucleoprotein complex.** Crystals of Pli-PLMVd in complex with a U1A protein.

The same solubility issues I had with the PLMVd.034 and PLMVd.282 constructs returned to a lesser degree with the chaperone constructs. My efforts to crystallize a viroid structure were stymied by the difficulty with getting high concentrations of viroid RNA and the solubility difficulties in magnesium (Figure 2.8).

**Figure 2.8: Pli-PLMVd hydrogel formation when in solution with $MgCl_2$.** Pli-PLMVd RNA at 8.3 mg/mL in solution with 10 mM $MgCl_2$ forms an insoluble hydrogel.

## Group II intron-viroid chaperone constructs for cryo-EM

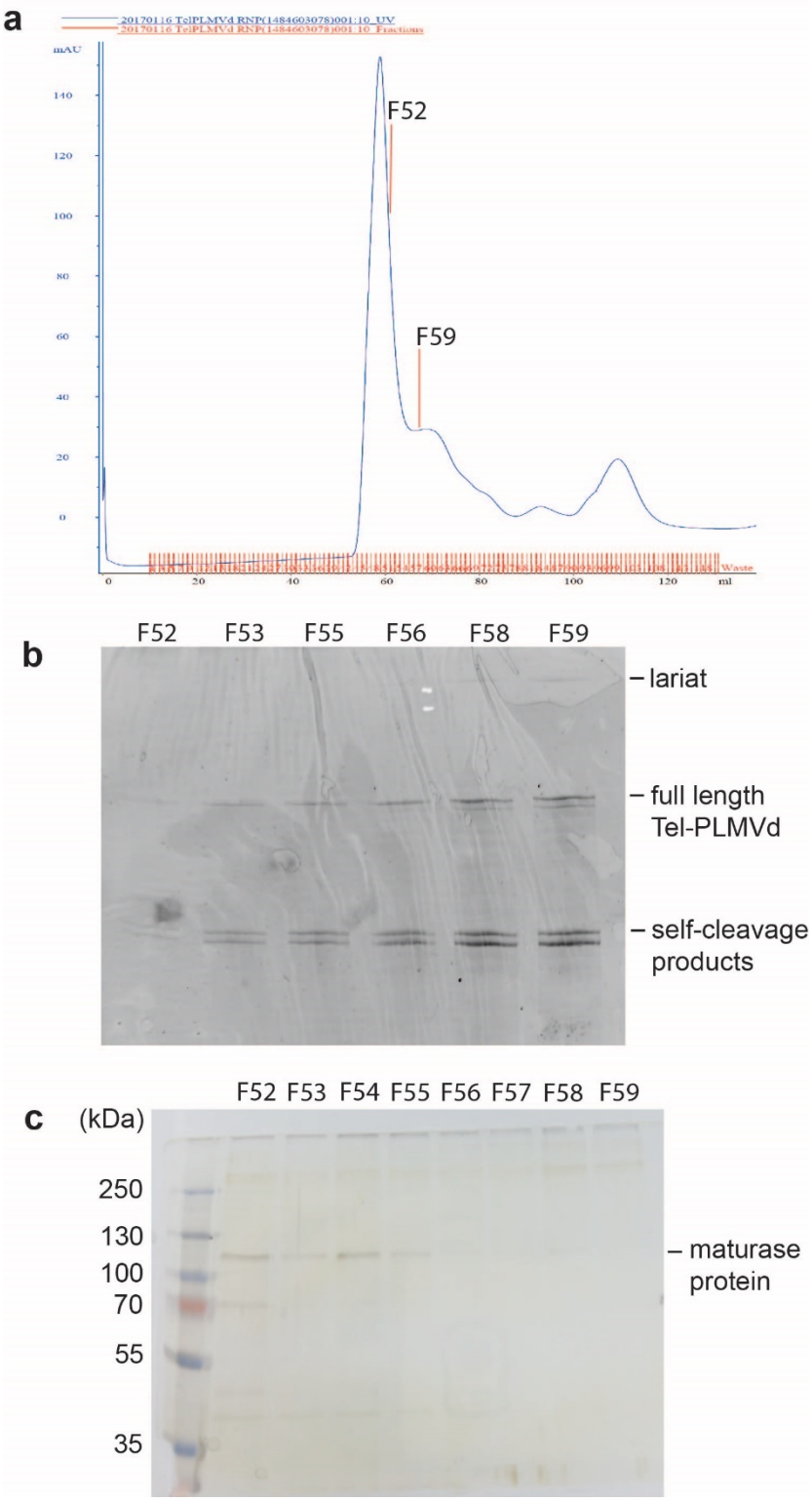
I moved on to trying to solve a structure with cryo-EM since this method does not necessarily require a high concentration of RNA and larger molecular weight samples, like the chaperone constructs, are ideal for cryo-EM. Prior to freezing samples for cryo-EM data collection, samples must be screened by negative stain electron microscopy to ensure proper concentration and the uniformity of particles. I learned with

the Pli-PLMVd construct that RNA alone does not stain well in uranyl acetate and is practically invisible with negative stain EM. As a consequence, our following RNA constructs included protein for easy visualization with negative stain EM screening.

The first chaperone construct I used that included protein was a group II intron from *T.el.* that forms a ribonucleoprotein complex with a maturase protein for splicing. I used *T.el.* 3c and *T.el.* 4h ribonucleoproteins and both fused DIIa to P3 of PLMVd.034. An additional *T.el.* 4h construct, Tel4h-D4-PLMVd, fused DIV to the PLMVd.034 sequence. I assembled *T.el.* RNA and maturase protein complexes in a simultaneous binding and splicing step at 45°C in 10 mM $MgCl_2$ and purified it over a size exclusion column by fast protein liquid chromatography (FPLC) before visualizing fractionated samples by negative stain EM. Magnesium is unavoidable in the assembly step as it's a necessary component for the maturase protein to properly bind and splice the RNA. Samples that appeared to have homogenous, appropriately sized single particles and were visually dense by negative stain were considered promising and saved for a potential cryo-EM data collection.

I found some fractions from Tel3c-PLMVd and Tel4h-D4-PLMVd that appeared to have the appropriate characteristics for this ribonucleoprotein, but these did not result in a structure. Fractions from a representative FPLC elution profile for Tel3c-PLMVd show that both the RNA and maturase protein co-eluted in several fractions and were relatively pure (Figure 2.9).

**Figure 2.9: Purification of Tel3c-PLMVd ribonucleoprotein. a**, The FPLC elution profile for Tel3c-PLMVd. Red lines indicate where fraction 52 and 59 eluted in the chromatogram. **b**, Tel3c-PLMVd RNA was observed in fractions 52 – 59. Lariat from the spliced complex is observed at the top of the gel, full length Tel3c-PLMVd is mid-gel, and at the bottom are the PLMVd hammerhead self-cleavage products. **c**, *T.el.* 4c maturase protein (106 kDa) was observed in fractions 52 – 55.


a
mAU
140
120
100
80
60
40
20
0
F52
F59
0
20
40
60
80
100
120
ml
Waste
b
F52
F53
F55
F56
F58
F59
– lariat
– full length Tel-PLMVd
– self-cleavage products
c
F52 F53 F54 F55 F56 F57 F58 F59
(kDa)
250
130
100
70
55
35
– maturase protein

The CChMVd.022 variant was also used in chaperone constructs with *T.el.* 4h. TelD4-CChMVd and TelhydraCChMVd both fused to P11 of CChMVd by DIV and DI, respectively. These constructs were transcribed into RNA but the maturase protein did not splice the RNA, so they were not used for further structural analysis.

## Discussion

Initial studies of PLMVd suggested the RNA was prone to aggregation, particularly in solution with magnesium. It was unclear if this property alone could explain the difficulty growing crystals of the viroid. Another likely explanation for why the RNA may resist crystallization is that the viroid structure is highly flexible and dynamic. To address that possibility, group II introns were used as chaperones. Crystallization chaperones can increase the number of crystal contacts, which aids crystal nucleation. Although I had a few hits from my crystal trays containing Pli-PLMVd and Pli-PLMVd-U1A, these crystals did not result in a structure (Figures 2.6 and 2.7). The absence of a crystal structure did not answer whether there was an inherent issue with flexibility in the viroid structure or if the aggregation of the viroid was the primary problem, so I moved on to trying to use group II introns as chaperones in cryo-EM constructs.

Small biological samples are more difficult to resolve for cryo-EM structure determination, so the group II intron chaperones are useful for increasing the size of particles, making the viroid more optimally sized. The chaperones also aided EM structure determination by add protein-binding sites without interrupting the viroid structure or sequence, which makes particles easier to visualize by negative stain EM. Analysis of my purification procedure by PAGE and negative stain EM suggested that I was able to isolate FPLC fractions with the correctly assembled Tel-PLMVd complex and that these complexes produced homogeneous single particles (data not shown). These cryo-EM samples did not reveal the expected Tel-PLMVd structure and I soon after discontinued my pursuit of solving the structure of a viroid. The *T.el.* chaperone constructs may yet yield the structure of a viroid; however, the low solubility of the RNA and the purification analysis should be carefully addressed.

Among the reasons for why I was unable to solve a structure of a viroid, I believe viroid RNA aggregation was a significant factor in disrupting complex purification and structure visualization. Magnesium is present and necessary to transcribe the RNA, for the group II intron RNAs to splice into homogenous structures and for the *T.el.* intron to assemble and splice with the maturase protein. Additionally, nearly every purification and sample preparation for crystallography or cryo-EM involves some

degree of RNA concentration, and this almost certainly resulted in small amounts of viroid aggregation and possibly even hydrogel formation. Although this feature of viroid RNA likely impeded structure determination, aggregation potentially plays a biological role in the viroid's lifecycle. Plant chloroplasts, where PLMVd has been suggested to replicate (38), have an approximate magnesium concentration of 5 mM (52). I have not seen reports of viroid aggregation in the literature so it's possible this investigation is the first to learn of it.

Without the insight from a high-resolution viroid structure, it did not make sense to test the impact of mutations to the tertiary structure. Sadly, the infection and viroid-binding pull down bioassays were impossible to complete since I did not have a growth chamber for the GF305 *Prunus persica* seedlings I had acquired. My investigation into the structure of a viroid did not have the desired results, but important lessons were learned about navigating a difficult research project.

Acknowledgements: I would like to thank Professor Thomas M. Gradziel at UC Davis, who gifted me 100 GF305 *Prunus persica* seedlings.

## Methods

*Cloning and in vitro transcription of crystallization and cryo-EM RNA constructs*

DNA was synthesized in gBlocks gene fragments (IDT) and inserted in the EcoRV site of the pUC57 plasmid. Plasmid DNA was linearized with an appropriate restriction enzyme. Transcription reactions contained the following components: 40 µg linearized plasmid DNA, 40 mM Tris HCl pH 7.5, 25 mM $MgCl_2$, 2 mM spermidine, 2.5 mM NTPs, 0.05% Triton X-100, 5 mM dithiothreitol, 5 µL T7 polymerase, 0.05 units of inorganic pyrophosphatase and water to a 1 mL final volume. Transcription reactions proceed at 37 °C for a minimum of 3 hours and reactions are stopped with addition of 12 µL of 0.1 M $CaCl_2$ and 20 units of TURBO DNase (Invitrogen). Incubation with DNase proceeds for approximately 40 minutes at 37 °C. 6 units of Proteinase K are added to reactions and the reactions are incubated at 37 °C for an hour. Transcription reaction tubes are spun down for 5 minutes at 21,130 RCF and the supernatant is filtered through a 0.2 µm membrane. The RNA was purified by buffer exchange in Amicon Ultra centrifugal filter units (Millipore-Sigma).

*Crystallization trials*

RNA was concentrated to 10 mg/mL. An Art Robbins Gryphon crystallization robot was used to add RNA to a crystal tray with a Natrix screen (Hampton Research) and crystals were grown in sitting drops by vapor diffusion. The sitting drops were made with 0.5 µL of 10 mg/mL RNA, 0.5 µL of precipitant Natrix solution, and 0.5 µL spermine. Crystal trays were centrifuged at 1000 RPM for 1 minute and trays are then incubated at 22 °C or 30°C. Conditions which appeared to promote crystallization were replicated and optimized in hand-set crystal trays.

*Group II intron self-splicing assays*

*P.li.* intron-containing RNAs were spliced in 10 mM $MgCl_2$, 1 M $NH_4Cl$, 40 mM Tris HCl pH 7.5, 0.02% SDS at 45 °C for 30 minutes. *O.i.* intron-containing RNAs were spliced in 100 mM $MgCl_2$, 500 mM $NH_4Cl$, 50 mM Tris HCl pH 7.5 at 45 °C for 30 minutes. The RNAs were purified by buffer exchange in Amicon Ultra centrifugal filter units (Millipore-Sigma).

*T.el. maturase protein purification*

Cloning details for the maturase protein have been previously reported by my colleagues (49). The *T.el.* maturase protein plasmid sequence also encoded an MBP and a 6x-His tag for increased solubility and affinity purification, respectively. A 2-liter cell culture of 6x-His-MBP-Tel was inoculated with 20 mL of starter culture and grown in LB containing carbenicillin at 37 °C and shaken at 250 RPM. At an optical density of 0.6 – 0.8, 1 mM IPTG was added to the cell culture to induce protein expression and the temperature was lowered to 22 °C. Cells were incubated for 48 hours and then spun down at 5,000 RCF for 10 minutes at 20 °C. The cell pellet was resuspended in lysis buffer (2 M urea, 500 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, 5 mM 2-mercaptoethanol and 0.2 mM PMSF) and lysed with a probe sonicator at 60% amplitude for 80 seconds total. The lysate was then centrifuged and the supernatant was mixed with 2 mL $Ni^{2+}$ resin for an hour on a rotator at 4 °C. The mixture was centrifuged and then the resin was transferred to a gravity purification column. The resin was washed with lysis buffer and then with 7 different wash buffers. Wash 1 contained 2 M urea, 1.5 M KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, and 5 mM 2-mercaptoethanol. Wash 2 contained 2 M urea, 500 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, 10% glycerol and 5 mM 2-mercaptoethanol. Wash 3 contained

2 M urea, 500 mM KCl, 20 mM Tris HCl pH 7.5, 20 mM imidazole, 10% glycerol and 5 mM 2-mercaptoethanol. Wash 4 contained 1.5 M urea, 500 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, 10% glycerol and 5 mM 2-mercaptoethanol. Wash 5 contained 1 M urea, 500 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, 10% glycerol and 5 mM 2-mercaptoethanol. Wash 6 contained 0.5 M urea, 500 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, 10% glycerol and 5 mM 2-mercaptoethanol. Wash 7 contained 500 mM KCl, 20 mM Tris HCl pH 7.5, 10 mM imidazole, 10% glycerol and 5 mM 2-mercaptoethanol. The maturase protein was eluted from the resin with 250 mM imidazole, 500 mM KCl, 20 mM Tris HCl pH 7.5, 10% glycerol and 5 mM 2-mercaptoethanol. The eluate was collected in a conical tube and then transferred to a 50 kDa MWCO centrifugal filter unit (Millipore-Sigma) for a buffer exchange into filtration buffer (500 mM KCl, 20 mM Tris HCl pH 7.5, 10% glycerol and 2 mM DTT). The protein solution was concentrated to 0.5 – 1 mL and brought to 50% glycerol for storage at -80 °C.

*T.el. ribonucleoprotein assembly and purification*

*T.el.*-intron containing RNA was spliced with the *T.el.* maturase protein in 40mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 300 mM $NH_4Cl$ and 5 mM

DTT at 45 °C for 5 minutes. The assembled ribonucleoprotein was then concentrated in an Amicon Ultra centrifugal filter unit (Millipore-Sigma) and the concentrate was spun down to get rid of any precipitate. The sample was loaded onto a size exclusion column (Superdex 200) which was equilibrated into 300 mM $NH_4Cl$, 40 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT. Ran sample over the column and collected 1 mL fractions of the eluate.

## References


1. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. Science (80- ) [Internet]. 2018 Sep 28 [cited 2020 Sep 22];361(6409):1346–9. Available from: https://pubmed.ncbi.nlm.nih.gov/30262497/

2. Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Mol Ther. 2008;16(11):1833–40.

3. Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Phase 1/2 study of COVID-19 RNA vaccine BNT162b1 in adults. Nature [Internet]. 2020 [cited 2020 Oct 7]; Available from: https://pubmed.ncbi.nlm.nih.gov/32785213/

4. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation [Internet]. Vol. 29, Genes and Development. Cold Spring Harbor Laboratory Press; 2015 [cited 2020 Sep 22]. p. 1343–55. Available from: https://pubmed.ncbi.nlm.nih.gov/26159994/

5. Arango D, Sturgill D, Alhusaini N, Dillman AA, Sweet TJ, Hanson G, et al. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. Cell [Internet]. 2018 Dec 13 [cited 2020 Oct 5];175(7):1872-1886.e24. Available from: https://pubmed.ncbi.nlm.nih.gov/30449621/

6. Ziehler WA, Engelke DR. Probing RNA Structure with Chemical Reagents and Enzymes. Curr Protoc Nucleic Acid Chem [Internet]. 2000 Feb [cited 2020 Sep 22];00(1):6.1.1-6.1.21. Available from: https://pubmed.ncbi.nlm.nih.gov/18428862/

7. Sawant AA, Tanpure AA, Mukherjee PP, Athavale S, Kelkar A, Galande S, et al. A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. Nucleic Acids Res [Internet]. 2015

[cited 2020 Jun 1];44(2):16. Available from: https://academic.oup.com/nar/article-abstract/44/2/e16/2468197

8. Liu Y, Sousa R, Wang Y-X. Specific labeling: An effective tool to explore the RNA world. BioEssays [Internet]. 2016 Feb 1 [cited 2020 Jun 2];38(2):192–200. Available from: http://doi.wiley.com/10.1002/bies.201500119

9. Flamme M, McKenzie LK, Sarac I, Hollenstein M. Chemical methods for the modification of RNA. Vol. 161, Methods. Academic Press Inc.; 2019. p. 64–82.

10. Flood DT, Knouse KW, Vantourout JC, Kitamura S, Sanchez BB, Sturgell EJ, et al. Synthetic Elaboration of Native DNA by RASS (SENDR). 2020 [cited 2020 Nov 18]; Available from: https://dx.doi.org/10.1021/acscentsci.0c00680

11. Gillingham D, Rasale D. Direct and selective modification of RNA – An open challenge in nucleic acid chemistry. Chimia (Aarau). 2018;72(11):777–81.

12. George JT, Srivatsan SG. Posttranscriptional chemical labeling of RNA by using bioorthogonal chemistry. Vol. 120, Methods. Academic Press Inc.; 2017. p. 28–38.

13. Trepotec Z, Lichtenegger E, Plank C, Aneja MK, Rudolph C. Delivery of mRNA Therapeutics for the Treatment of Hepatic Diseases. Vol. 27, Molecular Therapy. Cell Press; 2019. p. 794–802.

14. Ni S, Yao H, Wang L, Lu J, Jiang F, Lu A, et al. Chemical modifications of nucleic acid aptamers for therapeutic purposes. Vol. 18, International Journal of Molecular Sciences. MDPI AG; 2017.

15. Komarova N, Kuznetsov A. Inside the black box: What makes Selex better? Vol. 24, Molecules. MDPI AG; 2019.

16. Gemmill D, D'souza S, Meier-Stephenson V, Patel TR. REVIEW Current approaches for RNA-labelling to identify RNA-binding

proteins 1. [cited 2020 Jun 1]; Available from: www.nrcresearchpress.com/bcb

17. Fujiwara Y, Dixon JA, O'hara F, Funder ED, Dixon DD, Rodriguez RA, et al. Practical and innate carbon hydrogen functionalization of heterocycles. Nature. 2012.

18. Zhou Q, Gui J, Pan CM, Albone E, Cheng X, Suh EM, et al. Bioconjugation by native chemical tagging of C-H bonds. J Am Chem Soc. 2013;

19. Gui J, Zhou Q, Pan CM, Yabe Y, Burns AC, Collins MR, et al. C-H methylation of heteroarenes inspired by radical SAM methyl transferase. J Am Chem Soc. 2014;

20. O'Hara F, Blackmond DG, Baran PS. Radical-based regioselective C-H functionalization of electron-deficient heteroarenes: Scope, tunability, and predictability. J Am Chem Soc. 2013 Aug 14;135(32):12122–34.

21. Dolbier WR. Guide to Fluorine NMR for Organic Chemists. Guide to Fluorine NMR for Organic Chemists. 2008.

22. Sloop J. 19-Fluorine nuclear magnetic resonance chemical shift variability in trifluoroacetyl species. Reports Org Chem. 2013;

23. Ji Y, Brueckl T, Baxter RD, Fujiwara Y, Seiple IB, Su S, et al. Innate C-H trifluoromethylation of heterocycles. Proc Natl Acad Sci U S A. 2011 Aug 30;108(35):14411–5.

24. Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat Struct Mol Biol [Internet]. 2009 May [cited 2020 Oct 5];16(5):528–33. Available from: https://pubmed.ncbi.nlm.nih.gov/19363482/

25. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc Natl Acad Sci U S A

[Internet]. 2013 Apr 2 [cited 2020 Dec 17];110(14):5498–503. Available from: https://www.pnas.org/content/110/14/5498

26. Gordon CJ, Tchesnokov EP, Woolner E, Perry JK, Feng JY, Porter DP, et al. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. J Biol Chem. 2020 May 15;295(20):6785–97.

27. Sheahan TP, Sims AC, Zhou S, Graham RL, Pruijssers AJ, Agostini ML, et al. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. Sci Transl Med [Internet]. 2020 Apr 6 [cited 2020 Apr 20]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/32253226

28. Flynn R, Smith B, Johnson A, Pedram K, George B, Malaker S, et al. Mammalian Y RNAs are modified at discrete guanosine residues with N-glycans. bioRxiv [Internet]. 2019 Sep 30 [cited 2020 Sep 29];787614. Available from: https://doi.org/10.1101/787614

29. Pfizer and BioNTech Choose Lead mRNA Vaccine Candidate Against COVID-19 and Commence Pivotal Phase 2/3 Global Study | Pfizer [Internet]. [cited 2020 Dec 13]. Available from: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-choose-lead-mrna-vaccine-candidate-0

30. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. N Engl J Med [Internet]. 2020 Nov 12 [cited 2020 Dec 13];383(20):1920–31. Available from: https://www.nejm.org/doi/full/10.1056/NEJMoa2022483

31. Musumeci D, Irace C, Santamaria R, Montesarchio D. Trifluoromethyl derivatives of canonical nucleosides: Synthesis and bioactivity studies. Medchemcomm [Internet]. 2013 Oct 25 [cited 2020 Dec 4];4(10):1405–10. Available from: www.rsc.org/medchemcomm

32. Heidelberger C, Parsons DG, Remy DC. Syntheses of 5-Trifluoromethyluracil and 5-Trifluoromethyl-2′-deoxyuridine. J Med Chem [Internet]. 1964 Jan 1 [cited 2020 Dec 4];7(1):1–5. Available from: https://pubs.acs.org/sharingguidelines

33. Ding B. The Biology of Viroid-Host Interactions. 2009 [cited 2020 Oct 31]; Available from: www.annualreviews.org

34. Hadidi A, Flores R, Randles J, Semancik J. Viroids: properties, detection, diseases and their control [Internet]. 2003 [cited 2020 Oct 23]. Available from: https://books.google.com/books?hl=en&lr=&id=ktYAYhLKOooC&oi=fnd&pg=PA3&dq=2003+hadidi+viroids&ots=z6DVQFw3LR&sig=3Dbti8oE32N7ChiMlZV3W044wvU

35. Hadidi A. Next-generation sequencing and CRISPR/Cas13 editing in viroid research and molecular diagnostics [Internet]. Vol. 11, Viruses. MDPI AG; 2019 [cited 2020 Oct 31]. Available from: https://pubmed.ncbi.nlm.nih.gov/30699972/

36. Diener TO. Discovering viroids — a personal perspective. Nat Rev Microbiol [Internet]. 2003 [cited 2020 Oct 23];1(1):75–80. Available from: https://pubmed.ncbi.nlm.nih.gov/15040183/

37. Potato spindle tuber viroid | Agriculture and Food [Internet]. [cited 2020 Oct 23]. Available from: https://www.agric.wa.gov.au/potatoes/potato-spindle-tuber-viroid

38. Ding B. Viroids: Self-replicating, mobile, and fast-evolving noncoding regulatory RNAs [Internet]. Vol. 1, Wiley Interdisciplinary Reviews: RNA. Wiley Interdiscip Rev RNA; 2010 [cited 2020 Oct 23]. p. 362–75. Available from: https://pubmed.ncbi.nlm.nih.gov/21956936/

39. Flores R, Gago-Zachert S, Serra P, Sanjuán R, Elena SF. Viroids: Survivors from the RNA world? [Internet]. Vol. 68, Annual Review of Microbiology. Annual Reviews Inc.; 2014 [cited 2020 Nov 1]. p. 395–414. Available from: www.annualreviews.org

40. Dubé A, Bolduc F, Bisaillon M, Perreault J-P. Mapping studies of the

Peach latent mosaic viroid reveal novel structural features. Mol Plant Pathol [Internet]. 2011 Sep [cited 2017 May 4];12(7):688–701. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21726370

41. Gago S, De La Peña M, Flores R. A kissing-loop interaction in a hammerhead viroid RNA critical for its in vitro folding and in vivo viability. RNA [Internet]. 2005 Jul [cited 2020 Oct 23];11(7):1073–83. Available from: https://pubmed.ncbi.nlm.nih.gov/15928342/

42. Dubé A, Baumstark T, Bisaillon M, Perreault JP. The RNA strands of the plus and minus polarities of peach latent mosaic viroid fold into different structures. RNA [Internet]. 2010 Mar [cited 2020 Oct 23];16(3):463–73. Available from: https://pubmed.ncbi.nlm.nih.gov/20089682/

43. Flores R, Navarro B, Kovalskaya N, Hammond RW, Di Serio F. Engineering resistance against viroids. Vol. 26, Current Opinion in Virology. Elsevier B.V.; 2017. p. 1–7.

44. Yoon JH, Srikantan S, Gorospe M. MS2-TRAP (MS2-tagged RNA affinity purification): Tagging RNA to identify associated miRNAs. Methods. 2012 Oct 1;58(2):81–7.

45. Rocheleau L, Pelchat M. The Subviral RNA database: A toolbox for viroids, the hepatitis delta virus and satellite RNAs research. BMC Microbiol [Internet]. 2006 Mar 6 [cited 2020 Oct 25];6. Available from: https://pubmed.ncbi.nlm.nih.gov/16519798/

46. Fiore JL, Nesbitt DJ. An RNA folding motif: GNRA tetraloop-receptor interactions. Q Rev Biophys [Internet]. 2013 Aug [cited 2020 Oct 29];46(3):223–64. Available from: https://doi.org/10.1017/S0033583513000048

47. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. Science (80- ) [Internet]. 2008 Apr 5 [cited 2020 Oct 27];320(5872):77–82. Available from: www.sciencemag.org/cgi/content/full/320/5872/72/DC1TablesS1and S2www.sciencemag.orgSCIENCEVOLhttp://science.sciencemag.org/

48. Robart AR, Chan RT, Peters JK, Rajashankar KR, Toor N. Crystal structure of a eukaryotic group II intron lariat. Nature [Internet]. 2014 Sep 24 [cited 2020 Oct 27];514(7521):193–7. Available from: https://www.nature.com/articles/nature13790

49. Haack DB, Yan X, Zhang C, Hingey J, Lyumkis D, Baker TS, et al. Cryo-EM Structures of a Group II Intron Reverse Splicing into DNA. Cell. 2019 Jul 25;178(3):612-623.e12.

50. Hall KB, Stump WT. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. Nucleic Acids Res [Internet]. 1992 Aug 25 [cited 2020 Oct 27];20(16):4283–90. Available from: https://pubmed.ncbi.nlm.nih.gov/1508720/

51. Ferré-D'Amaré AR, Doudna JA. Crystallization and structure determination of a hepatitis delta virus ribozyme: Use of the RNA-binding protein U1A as a crystallization module. J Mol Biol. 2000 Jan 21;295(3):541–56.

52. Tränkner M, Jamali Jaghdani S. Minimum magnesium concentrations for photosynthetic efficiency in wheat and sunflower seedlings. Plant Physiol Biochem. 2019 Nov 1;144:234–43.