

Design and Modulation of Novel Chemical Probes for RNA SHAPE Analysis: Water  
Soluble Isatoic Anhydrides and Nicotinic Acid Imidazolides

by

Adam Bradley Fessler

A dissertation submitted to the faculty of  
The University of North Carolina at Charlotte  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in  
Nanoscale Science

Charlotte

2021

Approved by:

---

Dr. Craig Ogle

---

Dr. Markus Etzkorn

---

Dr. Jerry Troutman

---

Dr. Christopher Bejger

---

Dr. Todd Steck

©2021  
Adam Bradley Fessler  
ALL RIGHTS RESERVED

## ABSTRACT

ADAM BRADLEY FESSLER. Design and Modulation of Novel Chemical Probes for RNA SHAPE Analysis: Water Soluble Isatoic Anhydrides and Nicotinic Acid Imidazolides (Under the direction of Dr. CRAIG A. OGLE)

The growing interest in RNA biology, along with an expansion in computational capabilities, throughput, and sensitivity has provided researchers new tools to probe RNA. RNA selective 2'-hydroxy acylation analyzed by primer extension (SHAPE) utilizes the unique reactivity of the 2'-OH on the ribose sugar at sites of flexibility within an RNA structure to probe the secondary structures of RNA. This work sought to address the limitations of current SHAPE probes including poor aqueous solubility. Several iterations of designs were tested and these results were used to direct the continued development of new probes. A strategy for modulation of probe reactivity was developed and utilized for isatoic anhydride-based probes and nicotinic acid imidazolidine-based probes. The work demonstrated the capability of new RNA probes to provide improved signal to noise of RNA SHAPE both in vivo and in vitro.

## ACKNOWLEDGEMENTS

Over the course of this degree, many hands and minds have helped to push me along and to keep me on track. I would like to thank all of the friends, colleagues, collaborators, and mentors who have participated in my research or contributed their time to coursework and my curiosities.

I would like to thank my Advisor, Dr. Craig Ogle, for his guidance. There are many lessons to learn during the course to degree, many more than simply the field in which you are studying. I hope to retain these lessons and academic curiosity in my future journeys. It has been a real pleasure; I hope retirement will treat you and your family well. Enjoy it Doc.

I certainly must thank my family, without your encouragement through this long journey, there is certainly no way I would have made it through this. In particular, I want to thank Brittany for her patience and her stabilizing presence which was, and remains, invaluable. I am excited for what is to follow but it is all in good hands, adventure awaits.

## TABLE OF CONTENTS

LIST OF ABBREVIATIONS	vii
LIST OF TABLES	ix
LIST OF FIGURES	xii
CHAPTER 1: INTRODUCTION	1
1.1 BIOCONJUGATION	1
1.2 RNA	4
1.3 PROBING	5
1.4 RNA SHAPE	6
1.5 ISATOIC ANHYDRIDE	8
1.6 NICOTINIC ACID IMIDAZOLIDES	9
1.7 SHAPE MaP	10
1.8 BARRIERS TO SUCCESS	11
1.9 DISSERTATION SUMMARY	13
1.10 INTRODUCTION REFERENCES	14
CHAPTER 2: WATER-SOLUBLE ISATOIC ANHYDRIDES: A PLATFORM FOR RNA-SHAPE ANALYSIS AND PROTEIN BIOCONJUGATION	19
2.1 ABSTRACT	19
2.2 INTRODUCTION	19
2.3 RESULTS AND DISCUSSION	22
2.3.1 Synthesis of IPIA	22
2.3.2 Ion Exchange	25
2.3.3 Reactivity of prepared reagents	25
2.3.4 Determination of aqueous stability	27
2.3.5 RNA-SHAPE analysis	28
2.4 CONCLUSIONS	32
2.5 EXPERIMENTAL	33
2.5.1 Chemicals	33
2.5.2 Instrumentation	33
2.5.3 Synthesis	34
2.5.4 Bioconjugation of BSA	34
2.5.5 PCR amplification and in vitro Transcription	34
2.5.6 In vitro SHAPE modification	35
2.5.7 cDNA synthesis and RNA-SHAPE analysis by denaturing Polyacrylamide Gel Electrophoreses (PAGE)	35
2.5.8 Signal-to-background ration calculation	36
2.6 REFERENCES	37
2.7 APPENDIX I: SUPPLEMENTARY INFORMATION	39
CHAPTER 3: INNATELY WATER SOLUBLE ISATOIC ANHYDRIDES WITH MODULATED REACTIVITIES FOR RNA SHAPE ANALYSIS	48
3.1 ABSTRACT	48
3.2 BODY	48
3.3 REFERENCES	58

3.4 APPENDIX II: SUPPLEMENTARY INFORMATION	61
CHAPTER 4: A NOVEL SHAPE REAGENT ENABLES THE ANALYSIS OF RNA STRUCTURE IN LIVING CELLS WITH UNPRECEDENTED ACCURACY	69
4.1 ABSTRACT	69
4.2 INTRODUCTION	69
4.3 MATERIALS AND METHODS	72
4.3.1 Synthesis of SHAPE reagents	72
4.3.2 Chemical characterization	73
4.3.3 Half-life determination	73
4.3.4 Extraction of native deproteinized <i>E. coli</i> rRNAs	73
4.3.5 Ex vivo Probing of <i>E. coli</i> rRNAs	74
4.3.6 In vivo probing of <i>E. coli</i> cells	74
4.3.7 In vivo probing of <i>B. subtilis</i> cells	75
4.3.8 In vivo probing of HEK293 cells	76
4.3.9 Total RNA extraction	76
4.3.10 SHAPE-MaP library preparation	77
4.3.11 Analysis of SHAPE-MaP data	78
4.3.12 Reference structure of rRNAs	79
4.3.13 Generation of Receiver Operating Characteristics (ROC) curves	79
4.3.14 Signal-to-noise estimation	79
4.4 RESULTS	80
4.4.1 Comparison of RT enzymes for SHAPE-MaP	80
4.4.2 Synthesis of six novel candidate SHAPE reagents	81
4.4.3 2A3 outperforms NAI at probing RNA in vivo	85
4.4.4 2A3 markedly improves the accuracy of RNA structure prediction	91
4.4.5 Characterization of 2A3	92
4.5 DISCUSSION	93
4.6 REFERENCES	96
4.7 APPENDIX III: SUPPLEMENTARY INFORMATION	99
CHAPTER 5: DISSERTATION SUMMARY	136

## LIST OF ABBREVIATIONS

$\epsilon$	molar extinction coefficient
$\sigma_m$	Hammett value for meta substituted benzoic acid derivatives
$\sigma_p$	Hammett value for para substituted benzoic acid derivatives
1M7	1-Methyl-7-nitroisatoic anhydride
2A3	2-Aminonicotinic acid imidazolid
3AIA	3-Azaisatoic anhydride
6A3	6-Aminopyridine-3-carboxylic acid imidazolid
Ab	antibody
ACN	acetonitrile
ADC	antibody drug conjugate
amu	atomic mass unit
B5	Benzotriazole-5-carboxylic acid imidazolid
BSA	bovine serum albumin
CDI	carbonyldiimidazole
cDNA	complimentary deoxyribonucleic acid strand
CMCT	N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate
CUAAC	copper catalyzed azide alkyne cycloaddition
DCM	Dichloromethane
DNA	deoxyribonucleic acid
DMF	Dimethylformamide
DMIA	N-(N,N-dimethylaminoethylene) isatoic anhydride
DMS	Dimethylsulfate
DMSO	Dimethyl sulfoxide
DOL	degree of labeling
dsRNA	double stranded RNA
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
ESI	electrospray ionization
Et <sub>2</sub> O	Diethyl ether
EtOAc	Ethyl acetate
EtOH	Ethanol
Fab	antigen binding fragment
FAI	2-Methyl-3-furoic acid imidazolid
GPC	gel permeation chromatography
HABA	4'-Hydroxyazobenzene-2-carboxylic acid
HPLC	high-performance liquid chromatography
HRP	Horseradish peroxidase
HRMS	high resolution mass spectrometry
I5	Indoline-5-carboxylic acid imidazolid
I6	Isoquinoline-6-carboxylic acid imidazolid
IPA	Isopropyl alcohol
IPIA	N-(3-iodopropyl) isatoic anhydride

LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
MeOH	Methanol
MFE	minimum free energy
MS	mass spectrometry
NaI	Sodium iodide
NAI	2-Methylnicotinic acid imidazoline
ncRNAs	non-coding RNAs
NGS	next generation sequencing
NHS	N-hydroxysuccinimide
NIC	Nicotinic acid imidazolidine
NMIA	N-methyl isatoic anhydride
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PE	primer extension
PEG	Polyethylene glycol
PI	isoelectric point
PPV	positive predictive value
SASA	solvent accessible surface area
snoRNA	small nucleolar RNA
SPR	surface plasmon resonance
Sulfo-NHS	N-hydroxysulfosuccinimide
SHAPE MaP	selective 2'-hydroxy acylation analyzed by primer extension with mutational profiling
ssDNA	single stranded DNA
ssRNA	single stranded RNA
SSU	small subunit
RBP	RNA binding proteins
RNA	ribonucleic acid
RNA SHAPE	RNA selective 2'-hydroxy acylation analyzed by primer extension
RNP	Telomerase ribonucleoprotein complex
rt	room temperature
RT	reverse transcription
tbTR	<i>Trypanosoma brucei</i> Telomerase RNA
TEA	Triethylamine
TGIRT-III	thermostable group II intron RT
TR	telomerase RNA
THF	Tetrahydrofuran
UV-vis	ultraviolet visible



## LIST OF TABLES

Table 2-1	Preparation of IPIA Derivatives at 40 °C.	26
Table 2-2	List of primers used to amplify PfU3 snoRNA	35
Table 4-1	SHAPE reagents used in this study. Chemical structure, full name and acronym of the SHAPE reagents tested in this study. NAI, used as a standard, is reported for comparison.	82

## LIST OF FIGURES AND SCHEMES

Figure 1-1	Simplified overview of RNA SHAPE compared to MaP. The cDNA truncation of SHAPE is measured via gel electrophoresis and nucleotide reactivity can be correlated to band darkness. SHAPE MaP uses sequencing libraries and computational software to determine nucleotide modification and individual nucleotide reactivity.	7
Figure 1-2	Common electrophiles for RNA SHAPE probing	8
Scheme 2-1	Preparation of IPIA (1) from isatoic anhydride	23
Figure 2-1	ORTEP diagram of 1, space group P212121, cell parameters $a=5.1602(3)$ Å, $b=6.9659(4)$ Å, $c=31.110(2)$ Å.	23
Scheme2-2	Competition Reaction of N,N-dimethyl octyl-amine for IPIA and n-BuI	24
Figure 2-2	Reaction Progress for Competition Assay Measured by $^1\text{H}$ NMR. Open squares ( $\square$ ) = n-BuI, Open triangles ( $\Delta$ ) = 1.	24
Figure 2-3	a) ORTEP diagram of 3b, space group P21/c, cell parameters $a=14.1392(11)$ Å, $b=9.2466(7)$ Å, $c=13.3238(9)$ Å, $\beta=99.099(7)^\circ$ b) ORTEP diagram of 3c, space group P21/c $a=15.21142(9)$ Å, $b=8.7274(7)$ Å, $c=13.3470(8)$ Å, $\beta=99.549(6)^\circ$ .	25
Figure 2-4	Scheme for labeling of BSA (0.15 mM) at 25 °C in 25 mM bicarbonate buffer (pH 8.4). b) Data table used in calculation of DOL. c) Triplicate data for BSA labeling reactions using 3b and 3c for 1 h at 25 °C.	27
Scheme 2-3	Preparation of 4 from DMIA	29
Figure 2-5	SHAPE-RT probing by various SHAPE electrophiles and secondary structure of U3 snoRNA (A) Cy5 labeled denaturing gel comparing modification in the synthetic pfU3 snoRNA using various SHAPE electrophiles at 15 mM (left panel) and 50 mM (right panel) concentration. Lower panel represents the signal-to-background (S/B) ratios calculated for each lane. (B) Secondary structure of U3 snoRNA depicting different domains. (C) Secondary structure of Stem II', Box C and Stem III of PfU3 snoRNA with the SHAPE reactivities mapped on	31

the unpaired residues. An enlarged version for nucleotide sequence found in supplemental (Figure S3A and S3B).

Figure 3-1	Graphical representation of the current coverage of chemical space for RNA SHAPE reagents	50
Figure 3-2	Synthesis of 4a & 4b A) NaH, DMF B) Br(CH <sub>2</sub> ) <sub>3</sub> Cl C) Me <sub>3</sub> N, DMF	52
Figure 3-3	Synthesis of 6, A) NaH, DMF B) MeI (1 eq) C) MeBr, acetone	54
Figure 3-4	In vitro SHAPE-RT modification of synthetic TbTR catalytic core, (A) Structure and reagent comparison. Cy5 labeled denaturing gel indicates modification in the unpaired nucleotide of (B) Template and template boundary element (C) Template distal helix. (D) Secondary structure of the TbTR catalytic core based on in vitro SHAPE modification of RNA using compound 6. Unpaired RNA nucleotides are shown in red and orange color based on their reactivities, where high reactivity is signified by red color and low reactivity by orange with the SHAPE reagent. Nucleotides denoted by grey color were not modified in this study.	56
Figure 4-1	Comparison of SHAPE reagents under ex vivo conditions. (A) Boxplot of SHAPE–MaP mutation frequencies for E. coli 16S and 23S rRNAs probed ex vivo after deproteinization. Box plots span the interquartile range (from Q1 to Q3). The grey area spans from the median in the DMSO sample (control) to the median in the NAI sample (reference). (B) ROC curves for all tested SHAPE reagents, calculated with respect to the accepted phylogenetically-inferred 16S and 23S rRNA structures from the Comparative RNA Web (30). The inset reports the area under the curve (AUC) for each compound. (C) Sample of SHAPE–MaP mutation frequencies for all tested compounds across a region spanning nucleotides 567 to 884 of E. coli 16S rRNA. Colored bases are those whose mutation frequencies exceed by 2-fold the median mutation frequency in the analyzed region. The accepted structure is reported as an arc plot.	85
Figure 4-2	Comparison of SHAPE reagents under in vivo conditions. (A) Boxplot of SHAPE–MaP mutation frequencies for E. coli 16S and 23S rRNAs probed in vivo. Box plots span the interquartile range (from Q1 to Q3). The grey area spans from the median in the DMSO sample (control) to the median in the NAI sample (reference). (B) ROC curves for all tested SHAPE reagents, calculated on solvent-exposed residues in the crystal	87

structure of the *E. coli* ribosome (PDB: 5IT8), with respect to the accepted phylogenetically-inferred 16S and 23S rRNA structures from the Comparative RNA Web (30). The inset reports the AUC for each compound. (C) Median in vivo SHAPE-MaP mutation frequencies across all stem-loops in the accepted *E. coli* 16S and 23S rRNA structures. Bases are numbered relatively to the loop. Positions -3 to -1 correspond to stem bases, while positions 0 and +1 correspond to loop bases. (D) Ratio between the median loop mutation frequency and the median stem mutation frequency, calculated on all stem-loops from C.

- |            |   |    |
|------------|---|----|
| Figure 4-3 | 2A3 successfully queries regions of the ribosome that are blind to other reagents. (A) Sample of SHAPE-MaP mutation frequencies for all tested compounds across a region spanning nucleotides 990 to 1162 of <i>E. coli</i> 23S rRNA. Colored bases are those whose mutation frequencies exceed by 2-fold the median mutation frequency in the analyzed region. The accepted structure is reported as an arc plot. The hyperreactive residue U1032 and the unreactive stretch of 19 nucleotides are respectively marked in red and blue. (B) Three-dimensional model of the domain depicted in A, colored by B factor (PDB: 5IT8). The inset zooms on the helical region containing the 2A3-hyperreactive residue U1032.  | 90 |
| Figure 4-4 | 2A3 outperforms NAI at experimentally-driven RNA structure modeling. (A) Grid search (jackknifing) of optimal slope/intercept value pairs for <i>E. coli</i> 16S and 23S rRNAs in vivo probing data for NAI and 2A3. Values represent the geometric mean of sensitivity and PPV for the secondary structures predicted using each slope/intercept value pair. The chosen value pair is boxed in green. (B) Arc plot comparison of <i>E. coli</i> 16S rRNA reference structure (top), and structure inferred using either NAI-derived or 2A3-derived restraints (bottom). Black/green arcs correspond to correctly predicted base-pairs, violet arcs to non-predicted base-pairs and red arcs to mispredicted base-pairs. PPV and sensitivity for each prediction are indicated. | 92 |
| Figure 4-5 | Mechanism of 2A3 reaction with RNA. Reaction of 2A3 with the 2'-OH of structurally-flexible RNA residues, resulting in the formation of a 2'-O-adduct.  | 95 |

## CHAPTER 1: INTRODUCTION

### 1.1 BIOCONJUGATION.

The addition of new functionality to a biomolecule, known as bioconjugation, is a common technique which provides researchers the capacity to alter the native characteristics of a material.<sup>1</sup> The end use for modified biologics spreads a diverse swath of applications, including therapeutic and diagnostic techniques.<sup>2-5</sup> Along with the interest to modify biomaterials for applications within medicine, many important fundamental research questions can be addressed using non-native biomaterials and bioconjugates. The breadth of bioconjugate strategies continues to expand to include new click chemistries and new probing methods. Bioconjugate chemistries will continue to expand for years to come.

Protein bioconjugation is commonly applied for medicinal purposes, either through modifications which enhance therapeutic outcomes or by conjugation of functional materials such as chemotherapeutics to protein carriers. Insulin bioconjugates using lysine modifications are applied to increase the circulation half-life of the active protein, helping reduce the frequency and volume of doses needed for diabetic patients. Antibody drug conjugates (ADC) use the addition of chemotherapeutics to antibodies for the design of targeted therapies, minimizing off target effects.<sup>6, 7</sup> These advances rely on the ability to modify a biomolecule in a controlled manner with strategies that enhance or modulate the native materials structure or function.

A commonality to many bioconjugation protocols is the use of an activated electrophile for reaction with a biological nucleophile. Due to this synthetic strategy, the reagents tend to be cross reactive with other nucleophiles including water, making the

completion of the bioconjugation process a competitive one.<sup>1</sup> The cross reactivity of bioconjugation reagents with water has been used as an argument/justification of the use of water miscible organic cosolvents for the delivery of bioconjugation reagents to solutions for the reaction. For this delivery, dimethylsulfoxide (DMSO) is the most commonly applied solvent due to its ability to dissolve a wide variety of molecules and its biological inertness.<sup>3, 8</sup> Although the solvent by a reactivity perspective is biologically “inert”, the widespread use of DMSO has been questioned by several studies which have shown the disruptive effects of the solvent.<sup>9-13</sup>

Numerous protein bioconjugation strategies rely on the use of the amine handle of lysine residues for the formation of the bond between the installed functionality and the biomolecule.<sup>4, 14, 15</sup> Formation of the bioconjugate then takes a pH responsive residue which actively participates in the solvation of the macromolecule and turns it into a non-responsive amide. This loss of the amine for formation of the amide bond can result in a decreased solubility of the biomolecule and modifies the macromolecules characteristic charge at physiological pH. Additionally, the conjugated payload can also change the local environment of the macromolecule by alteration of the surface character when a hydrophobic moiety such as a chemotherapeutic drug is added. These small changes can drastically alter the characteristics of a biomolecule in solution as well as in a cell.

The loss of a charged residue and conjugation of hydrophobic functionality into its place is a major issue bioconjugation strategies frequently fail to adequately address. The effects of these changes have been noted with antibody drug conjugates, where the ideal number of added functionalities has been claimed to be four.<sup>16</sup> At a degree of labeling (DOL) of four there is the highest chance of conjugation occurring on most of antibodies

in the distribution in innocuous locations and enough payload to remain an effective option for treatment. Moving towards higher degrees of labeling there is a dramatic decrease in the antibody solubility, hindering the applicability of the higher bioconjugates. Despite the inherent heterogeneity and reduction in solubility, lysine conjugation remains the most common route for the formation of bioconjugates.

Bioconjugation of macromolecules other than proteins have a few commonplace applications but have been limited compared to the rapid expansion of their amino acid counterparts. This reduced number of strategies is due to the increased difficulty in utilizing natural handles available for the modification.<sup>3, 17</sup> Proteins have numerous amino acids, of which, several can be utilized for bioconjugation strategies. Lipid, carbohydrate and nucleic acid bioconjugates are typically modified from their native state in order to use click strategies, including sulfhydryl maleimide chemistries and CuAAC strategies.<sup>18, 19</sup> Nucleic acid bioconjugation has seen a growth in recent years as made to order deoxy-ribonucleic acid (DNA) strands have been designed with functional ends, amino or thiol terminated.<sup>20-22</sup> The 3'-OH of DNA offers a modest nucleophile for targeted modification of double stranded DNA (dsRNA), but only one is only freely available at one end of the strand. Single stranded DNA (ssDNA), with exposed Watson-Crick faces of the nucleotides, can be more difficult to target as the moderately reactive amine groups of the bases can compete for modification with the moderate nucleophilic terminal 3'-OH. Due to this, purchase of pre-functionalized DNA strands is a convenient and reproducible strategy for the preparation of bioconjugates.

## 1.2 RNA.

Ribonucleic acids (RNA) contain a ribose sugar which retains the hydroxy in the 2' position, this hydroxyl distinguishes RNA from DNA and is responsible for the drastic difference in functionality and reactivity.<sup>23,24</sup> DNA is utilized for the storage of information within the cell, this information can be called upon at any time by the cell for transcription into RNA followed by translation into proteins. This process and hierarchy of information is the central dogma of biology, and this is also the reason that modification of DNA rightfully should be more difficult. The material is intentionally difficult to alter, which allows for life to work and for information to be passed from one cell to its daughter cells, and from one organism to its offspring. While mutations can and do occur, it is ideal for DNA of many organisms to be robust for survival and proliferation.

The hierarchy of information storage plays a direct role in the features of the biomolecules. Proteins are the final step of information transfer and serve a diverse role in the body, responsible to maintenance and functional preservation of the cell.<sup>25</sup> Due to this, proteins are more transient than DNA, being synthesized and applied to more damaging tasks then finally recycled at a much higher rate.<sup>26</sup> They serve as a vital, but renewable resource to the cell. In between the storage and the worker biomolecule is the information transfer material, RNA. Once thought of as simply a messenger, we are slowly learning of the diverse roles in which RNAs can serve.<sup>27-30</sup>

RNA has a unique responsibility of being an information carrier, coding RNA, while simultaneously possessing the necessary features to be functional as is the case in many non-coding RNAs. This knowledge gave rise to the RNA world hypothesis which describes early life possessing functional RNA which served catalytic function and could



be used as information storage. The idea of functional RNA is not entirely new but for years the ribosome has been known but treated as a one-off success by the cell. There is a resurgence in interest in RNA within the scientific community as its use in cellular function and responsibilities is not fully understood but the roles are clearly there.

### **1.3 PROBING.**

Another kind of bioconjugation, although less commonly thought about as bioconjugation, is the reaction of chemical probes with biomolecules. Probing of biological structures provides an important window of insight into structure-function relationships. These relationships play a key role in elucidation of strategies for determining druggable targets, methods for modulation of activity, and numerous complex research questions.<sup>31-33</sup> Probing of biological samples is inherently difficult due to several factors including complex structural dynamics, biological interactomes, and physiological conditions. Due to the significance of biological probing, numerous methods for the determination of biological structures have been pursued for all four biological macromolecules.

The determination of structures has been done using many different techniques including microscopy, spectroscopy, and many other chemical methodologies.<sup>34-36</sup> Characterization of structure using reactive probing compounds can provide a snapshot of complex roles and structures a biomolecule may have within a living cell.<sup>37-39</sup> The use of chemical probes can be highly selective in their target reactivity or have general reactivity which can be understood through a secondary determination step. Chemical probing provides interesting structural characteristics such as solvent exposed surfaces, hyper reactive moieties, and atypical residues. The probing of RNA and protein structures has been a significant interest in structural biology as well as chemical biology. The ability to

determine and understand structural outputs is key for the continuing growth and greater understanding into biological processes.

#### **1.4 RNA SHAPE**

RNA selective 2'-hydroxy acylation analyzed by prime extension (RNA SHAPE) relies on the unique reactivity of the 2'-OH to react with an electrophile, to generate a 2'-O-ester.<sup>40</sup> This reaction is only able to occur at the site of flexible residues on RNA as these residues can adapt geometries which are atypical for paired RNAs and helps to activate the ribose sugar towards reaction. Nucleotide flexibility is only found in nucleotides which are not found in base pairing interactions. SHAPE can then probe the secondary structure within RNA. The formation of 2'-O-ester can be determined through the use of primer extension which relied on reverse transcriptase (RT) to turn RNA into a complementary DNA strand (cDNA). At the site of chemical modification, the enzyme is unable to process the RNA strand any further and truncates the produced cDNA. The produced library of cDNA is then separated by denaturing gel electrophoresis and the labeled nucleotides can be determined by the banding pattern on the gel, Figure 1. Where darker banding correlate to a more commonly truncated cDNA, and therefore a reactive flexible nucleotide. This method of structural analysis has been used for over a decade to determine the nucleotide flexibility and reactivity. The data has been able to demonstrate highly reactive architectures found within RNA structures which may demonstrate RNA's capacity to perform catalytic functions within the cell.<sup>41</sup>

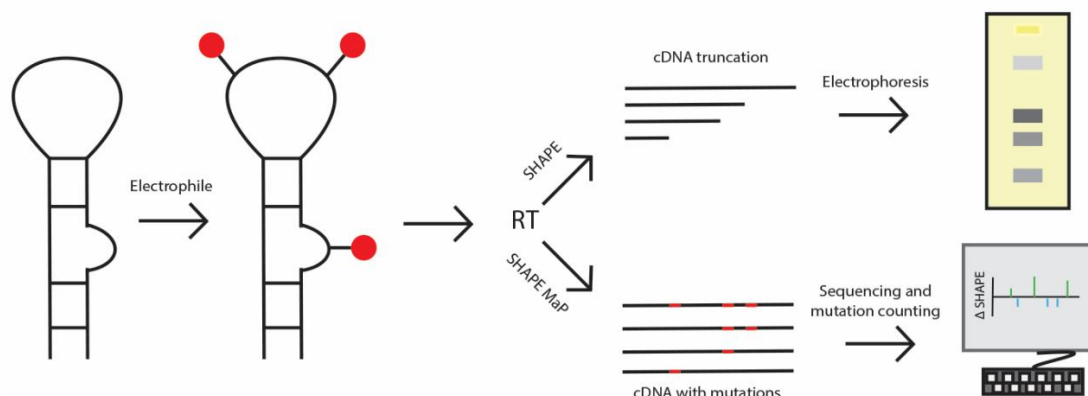


Figure 1-1: Simplified overview of RNA SHAPE compared to MaP. The cDNA truncation of SHAPE is measured via gel electrophoresis and nucleotide reactivity can be correlated to band darkness. SHAPE MaP uses sequencing libraries and computational software to determine nucleotide modification and individual nucleotide reactivity.

Many electrophiles have been developed for the use in RNA SHAPE, but only a select few have been highly successful for probing in cell RNA. The most common reagents are 1-methyl-7-nitroisatoic anhydride (1M7) and 2-methylnicotinic acid imidazolid (NAI), Figure 2. These two reagents have drastically different properties, where 1M7 is poorly water soluble but highly reactive and NAI is much less reactive but has some water solubility. The highly reactive nature of 1M7 has been shown to be an ideal characteristic for SHAPE analyses due to the self-quenching hydrolysis and rapid reaction that reduces the number of conformations the RNA can sample during probing.<sup>42</sup> The solubility and cell permeability of NAI has also been shown to be advantageous for in vivo RNA analysis, where 1M7 has been shown to fail.<sup>37</sup>

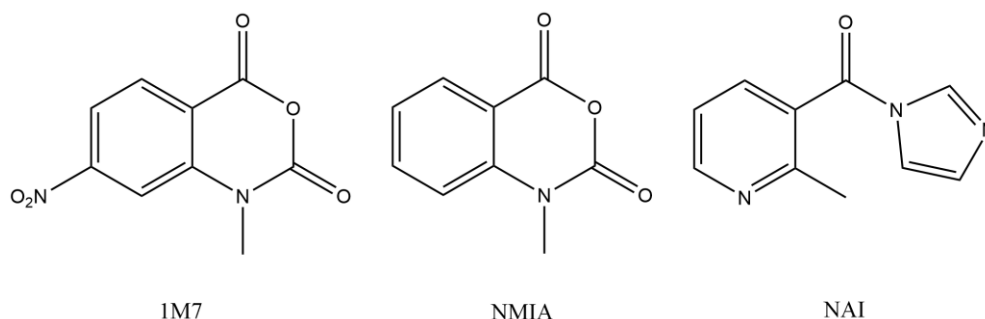


Figure 2: Common electrophiles for RNA SHAPE probing

RNA SHAPE experiments once relied on the use of PAGE based analysis, resulting in the ability to investigate windows of around 100 nucleotides at a time.<sup>43</sup> This then requires the use of different primers for different regions of an RNA which were to be investigated, becoming particularly laborious and difficult on long RNAs. The use of next generation sequencing allows for improved workflow but the nature of SHAPE strategies result in the termination of the cDNA at a modification site, meaning each read only provides the first modification that is reached, artificially skewing the signal towards early modifiable residues.<sup>39, 44, 45</sup>

### 1.5 ISATOIC ANHYDRIDE.

N-methyl isatoic anhydride (NMIA) was the first reagent used for the determination of RNA structure that relied on the reaction with the 2'-OH of the ribose sugar for SHAPE chemistry. This reagent was found principally to react with nucleotides not found in base pairing interactions.<sup>41</sup> This is due to ribose sugars with greater flexibility than those found in helical or based paired arrangements. This flexibility leads to the exposure of the 2'-OH for reaction with the electrophilic anhydride forming a stable adduct. Several new isatoic anhydride-based reagents have been prepared to react with RNA. Of these, 1M7 has repeatedly been demonstrated as an important and capable electrophile for the formation of 2'-O-adducts. This reagent relies on the addition of a nitro group para to the electrophilic

carbonyl of IA, resulting in a very electron poor carbonyl which activates the reagent to reaction with the ribose sugars and other nucleophiles. Isatoic anhydride-based electrophiles led the expansion of RNA structural analysis from the use of dimethylsulfate (DMS) which reacted with N2 of uracil and N1 of guanine to structural probing of flexible 2'-OH.<sup>46, 47</sup>

## 1.6 NICOTINIC ACID IMIDAZOLIDES.

As SHAPE chemistries expanded and the interest grew, researchers pushed towards the use of probes in living cells. This was first demonstrated by Chang et al. using a novel nicotinic acid imidazolidine derivative, NAI.<sup>48</sup> They argued the impressive result was made possible through the use of more soluble, lesser reactive probing compounds than NMIA or 1M7. This gave the reagent time to accumulate in the cell with less intrusion of the deactivating hydrolysis. From there, many groups have successfully probed RNA structures in living cells, including work from several groups with 1M7.<sup>37, 49-52</sup> One important note of the nicotinic acid imidazolidine based probing is the increased reagent amount (mole to mole) used for probing.

The probing work done with isatoic anhydride probes is commonly done at a final reagent concentration of 10 mM, wherein the DMSO concentration is 100 mM. When using the imidazolidine probes like NAI, final reagent concentration is 100 mM and DMSO concentration is 100 mM. This increase in the reagent concentration certainly plays an important role in the improved signal in vitro as well as in vivo. Additionally, background and control experiments are very important when the reaction time is increased as well as when the DMSO concentration is high. Under these conditions, many deleterious side effects are possible, including cell lysis or induced membrane fluidity, RNA degradation,

and other strange effects which reduce confidence in results and can contribute to experimental noise.

### **1.7 SHAPE MaP.**

One limitation regarding the use of traditional RNA SHAPE experiments for the probing of RNA structures is the limited output of data which is obtained from this labor-intensive experiment. The use of computationally based massively parallel sequencing approaches improves the output by collection of increased quantities of data. SHAPE with mutational profiling (MaP) measures large and transcriptome-wide systems and data is measured using computer software.<sup>53</sup> This digital output of data measures both modified and unmodified sites allowing for the variability of individual nucleotide reactivity to be determined. This technique produces much more quantitative data than traditional SHAPE experiments and is being adopted by numerous laboratories around the world for RNA SHAPE analysis, Figure 1.<sup>52, 54, 55</sup>

The modified strategy of RNA SHAPE uses altered buffer conditions for the RT production of cDNA. In this buffer, the RT does not stall at the labeled nucleotides, instead, the modification prevents the RT from reading the base and results in the incorporation of a random nucleotide or deletes the nucleotide at the site of each modification. This leads to a large library of cDNA, where each may contribute several probing reads from mutated or deletions at labeled sites.<sup>53</sup> The sequences are aligned computationally using similarities within the produced cDNA. This large quantity of data is then used to calculate individual nucleotide reactivities and provides additional depth to the experiment.<sup>55</sup> The use of computational libraries to accurately analyze the RNA secondary structure provides comparable results to traditional SHAPE experiments without the labor intensive and user

dependent workflow of the traditional experiment. SHAPE MaP is typically performed using NMIA, 1M7, or NAI as the electrophile, with 1M7 being described as ideal due to the rapid reaction and self-quenching nature which allows for isolation of intact RNA.<sup>49</sup>

## **1.8 BARRIERS TO SUCCESS.**

Biological buffers have a range of compositions based on their applications. Among the mix are monovalent and divalent cations, the respective anions, phosphates, and many others. One component which can be overlooked during bioconjugation processes is the water in the system. Water is not an impressive nucleophile, but in high concentration the reactivity of water, especially with electrophilic reagents cannot be ignored. The deactivating hydrolysis of reactive probes can play a major role in experimental failures, low conjugation efficiency and other deleterious outcomes.

An additional consideration is an organic compounds solubility in water. Frequently conjugation reagents and probes alike are poorly soluble in aqueous buffers, which tends to arise as a consequence of their reactive design or through their functionality. This issue tends to be addressable using DMSO for solvation, which can allow for more concentrated solutions to be prepared and diluted into aqueous buffers. Often, after initial dissolution the reagents can star in solution at useable concentrations and possibly more reasonable levels of DMSO. The frequent and indiscriminate use of DMSO is an area of growing dispute. Numerous works have shown direct involvement of the solvent in a reaction as well as leading to structural melt. Due to this, additional controls are important when using DMSO in biological samples.

The specific sequence of nucleotides in RNA leads to their capability to form highly intricate hierarchal structures much more complex than the primary sequence may appear.

The secondary structures: hairpins, bulges, and loops within the RNA structure lead to the final tertiary “shape” in which the RNA is stable. The structure is based on the interaction of segments within the individual RNA strand to form base pairing interactions within the single strand. These base pairing interactions within the RNA lead to sections of ssRNA and sections of dsRNA. RNA can adopt numerous structures from base pairing events but the conformation possessing the highest number of base pairings is assumed to be the most thermodynamically favorable state. The use of thermodynamic minimization alone is unable to accurately predict the structure of RNA due to an incomplete knowledge of RNA folding rules.<sup>56</sup> The secondary structures can be calculated using advanced computational software, but these computational models require a secondary validation as the complex nature of structures within the cell is well known to protein and nucleic acid biologists. Much information about the complex nature of RNA structures has been elucidated using in vitro RNA probing.<sup>57</sup> Issues with relying on the refolding of extracted RNA ex vivo without intracellular proteins have been brought to attention of structural biologist, meaning the in vitro data may not be the true in vivo structure.<sup>58</sup> For this reason, the future of RNA probing in vivo is of great importance within the field.

Cell membranes also play an important role in the success of probing techniques in vivo. Difficulties associated with membrane permeability, retention, or rejection can result in drastic reduction of the probe output. A number of general guidelines can be followed to make a general assessment of permeability for eukaryotic cell lines.<sup>59, 60</sup> Increased cell surface glycosylation and additional membranes further complicate probing of prokaryotic structure in cell.<sup>61</sup> In general, in cell probing raise a new set of complex problems which require time and patience to address and attempt to overcome.



## 1.9 DISSERTATION SUMMARY.

The early advancements made in RNA SHAPE largely have come from the use of 1M7 by Weeks et al.<sup>24, 41, 42, 53, 54, 62</sup> Recent, compelling results from other laboratories have drawn into question the accuracy of data collected using 1M7 due to inconsistencies when using this compound in vivo.<sup>37, 48, 63</sup> A review, published after the work demonstrated in Chapter 2 of this dissertation, from the Weeks laboratory highlighted the ability of new reagents to take the field of RNA SHAPE to maturation.<sup>51</sup> The proposed electrophiles of this recent review were focused on combining the high reactivity of 1M7 and true water solubility. This work addressed this highlighted need through the use of a water soluble isatoic anhydride platform in Chapter 3. Tuning of SHAPE electrophiles may provide researchers with greater control of the conditions and provide access to more optimal outcomes for in vivo experiments.

The three subsequent publications detail the work on improving the output of RNA SHAPE experiments. This includes attempts for compatibilization of the water soluble isatoic anhydride core RNA probes for improved output in vitro and in vivo through the use of modified strategies. Additionally, a new nicotinic acid imidazolidine reagent is introduced and detailed. The improved probing output of the new nicotinic acid imidazolidine introduced in chapter 4 details of the drastic improvement of in vivo probing output were described using multiple cell lines.

These chapters describe the synthesis and design of new probing reagents. Demonstrating the advantageous use of water-soluble compounds for biological applications is described and demonstrated through the experimental outcomes. The use of electrophile activation is demonstrated by the tunable nature of reagent strategies. An

additional strategy of intramolecular activation of an electrophile is demonstrated and the results of an in vivo experiment detailed. The work highlights the use of smart chemistries to improve and expand the foundations that have been laid with earlier probing reagents.

## 1.10 INTRODUCTION REFERENCES.

1. Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M., RNA Structure Analysis at 1. Hermanson, G. T., *Bioconjugate techniques*. 3rd ed. ed.; Academic Press: London, 2013.
2. Hein, C. D.; Liu, X.-M.; Wang, D., Click chemistry, a powerful tool for pharmaceutical sciences. *Pharmaceutical research* **2008**, *25* (10), 2216-2230.
3. Kalia, J.; Raines, R. T., Advances in Bioconjugation. *Current organic chemistry* **2010**, *14* (2), 138-147.
4. Boylan, N. J.; Zhou, W.; Proos, R. J.; Tolbert, T. J.; Wolfe, J. L.; Laurence, J. S., Conjugation site heterogeneity causes variable electrostatic properties in Fc conjugates. *Bioconjugate chemistry* **2013**, *24* (6), 1008-1016.
5. Chudasama, V.; Maruani, A.; Caddick, S., Recent advances in the construction of antibody–drug conjugates. *Nature Chemistry* **2016**, *8*, 114.
6. Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H.-C., Selective N-terminal functionalization of native peptides and proteins. *Chemical science* **2017**, *8* (4), 2717-2722.
7. Boga, S. B.; Krska, S. W.; Lin, S.; Pissarnitski, D.; Yan, L.; Kecec, A.; Tang, W.; Pierson, N. A.; Strulson, C. A.; Streckfuss, E.; Zhu, X.; Zhang, X.; Kelly, T.; Parish, C. A., Site-Selective Synthesis of Insulin Azides and Bioconjugates. *Bioconjugate Chemistry* **2019**, *30* (4), 1127-1132.
8. Lee, H.; Park, J.-B., Evaluation of the effects of dimethylsulphoxide on morphology, cellular viability, mRNA, and protein expression of stem cells culture in growth media. *Biomed Rep* **2017**, *7* (4), 291-296.
9. Cheng, C.-Y.; Song, J.; Pas, J.; Meijer, L. H. H.; Han, S., DMSO induces dehydration near lipid membrane surfaces. *Biophysical journal* **2015**, *109* (2), 330-339.
10. De Abreu Costa, L.; Henrique Fernandes Ottoni, M.; Dos Santos, G. M.; Meireles, B. A.; Gomes de Almeida, V.; De Fátima Pereira, W.; Alves de Avelar-Freitas, B.; Eustáquio Alvim Brito-Melo, G., Dimethyl Sulfoxide (DMSO) Decreases Cell Proliferation and TNF- $\alpha$ , IFN- $\gamma$ , and IL-2 Cytokines Production in Cultures of Peripheral Blood Lymphocytes. *Molecules* **2017**, *22* (11).
11. de Ménorval, M.-A.; Mir, L. M.; Fernández, M. L.; Reigada, R., Effects of dimethyl sulfoxide in cholesterol-containing lipid membranes: a comparative study of experiments in silico and with cells. *PloS one* **2012**, *7* (7), e41733-e41733.
12. Notman, R.; den Otter, W. K.; Noro, M. G.; Briels, W. J.; Anwar, J., The Permeability Enhancing Mechanism of DMSO in Ceramide Bilayers Simulated by Molecular Dynamics. *Biophysical Journal* **2007**, *93* (6), 2056-2068.

13. Verheijen, M.; Lienhard, M.; Schrooders, Y.; Clayton, O.; Nudischer, R.; Boerno, S.; Timmermann, B.; Selevsek, N.; Schlapbach, R.; Gmuender, H.; Gotta, S.; Geraedts, J.; Herwig, R.; Kleinjans, J.; Caiment, F., DMSO induces drastic changes in human cellular processes and epigenetic landscape in vitro. *Scientific Reports* **2019**, *9* (1), 4641.
14. Hassani, L.; Nourozi, R., Modification of Lysine Residues of Horseradish Peroxidase and Its Effect on Stability and Structure of the Enzyme. *Applied Biochemistry and Biotechnology* **2014**, *172* (7), 3558-3569.
15. Jain, N.; Smith, S. W.; Ghone, S.; Tomczuk, B., Current ADC Linker Chemistry. *Pharmaceutical Research* **2015**, *32* (11), 3526-3540.
16. McCombs, J. R.; Owen, S. C., Antibody Drug Conjugates: Design and Selection of Linker, Payload and Conjugation Chemistry. *The AAPS Journal* **2015**, *17* (2), 339-351.
17. De Benedetto, G.; Salvini, L.; Gotta, S.; Cescutti, P.; Micoli, F., Investigation on Sugar-Protein Connectivity in Salmonella O-Antigen Glycoconjugate Vaccines. *Bioconjugate Chemistry* **2018**, *29* (5), 1736-1747.
18. Raouane, M.; Desmaële, D.; Urbinati, G.; Massaad-Massade, L.; Couvreur, P., Lipid Conjugated Oligonucleotides: A Useful Strategy for Delivery. *Bioconjugate Chemistry* **2012**, *23* (6), 1091-1104.
19. van der Put, R. M. F.; Kim, T. H.; Guerreiro, C.; Thouron, F.; Hoogerhout, P.; Sansonetti, P. J.; Westdijk, J.; Stork, M.; Phalipon, A.; Mulard, L. A., A Synthetic Carbohydrate Conjugate Vaccine Candidate against Shigellosis: Improved Bioconjugation and Impact of Alum on Immunogenicity. *Bioconjugate Chemistry* **2016**, *27* (4), 883-892.
20. Patel, P. L.; Rana, N. K.; Patel, M. R.; Kozuch, S. D.; Sabatino, D., Nucleic Acid Bioconjugates in Cancer Detection and Therapy. *ChemMedChem* **2016**, *11* (3), 252-269.
21. Wang, T.-P.; Chiou, Y.-J.; Chen, Y.; Wang, E.-C.; Hwang, L.-C.; Chen, B.-H.; Chen, Y.-H.; Ko, C.-H., Versatile Phosphoramidation Reactions for Nucleic Acid Conjugations with Peptides, Proteins, Chromophores, and Biotin Derivatives. *Bioconjugate Chemistry* **2010**, *21* (9), 1642-1655.
22. Zhang, X.; Xu, Z.; Moumin, D. S.; Ciulla, D. A.; Owen, T. S.; Mancusi, R. A.; Giner, J.-L.; Wang, C.; Callahan, B. P., Protein-Nucleic Acid Conjugation with Sterol Linkers Using Hedgehog Autoprocessing. *Bioconjugate Chemistry* **2019**, *30* (11), 2799-2804.
23. Giegerich, R.; Voss, B.; Rehmsmeier, M., Abstract shapes of RNA. *Nucleic acids research* **2004**, *32* (16), 4843-4851.
24. Wilkinson, K. A.; Merino, E. J.; Weeks, K. M., Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature Protocols* **2006**, *1*, 1610.
25. Alberts, B., *Essential cell biology*. 2nd ed. ed.; Garland Science Pub.: New York, NY, 2004.
26. Badawi, A. H.; Siahaan, T. J., Suppression of MOG- and PLP-induced experimental autoimmune encephalomyelitis using a novel multivalent bifunctional peptide inhibitor. *Journal of Neuroimmunology* **2013**, *263* (1-2), 20-27.
27. Barta, A.; Jantsch, M. F., RNA in Disease and development. *RNA Biol* **2017**, *14* (5), 457-459.

28. Clancy, S., RNA Functions. *Nature Education* **2008**, *1* (1), 102.
29. Perez-Perri, J. I.; Rogell, B.; Schwarzl, T.; Stein, F.; Zhou, Y.; Rettel, M.; Brosig, A.; Hentze, M. W., Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nature Communications* **2018**, *9* (1), 4408.
30. Rich, A., The Era of RNA Awakening: Structural biology of RNA in the early years. *Quarterly Reviews of Biophysics* **2009**, *42* (2), 117-137.
31. Ehresmann, C.; Baudin, F.; Mougél, M.; Romby, P.; Ebel, J. P.; Ehresmann, B., Probing the structure of RNAs in solution. *Nucleic Acids Research* **1987**, *15* (22), 9109-9128.
32. Kwok, C. K.; Tang, Y.; Assmann, S. M.; Bevilacqua, P. C., The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends in Biochemical Sciences* **2015**, *40* (4), 221-232.
33. Piao, M.; Sun, L.; Zhang, Q. C., RNA Regulations and Functions Decoded by Transcriptome-wide RNA Structure Probing. *Genomics Proteomics Bioinformatics* **2017**, *15* (5), 267-278.
34. Ding, F.; Lavender, C. A.; Weeks, K. M.; Dokholyan, N. V., Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nature Methods* **2012**, *9*, 603.
35. Barnwal, R. P.; Yang, F.; Varani, G., Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys* **2017**, *628*, 42-56.
36. Köster, T.; Reichel, M.; Staiger, D., CLIP and RNA interactome studies to unravel genome-wide RNA-protein interactions in vivo in *Arabidopsis thaliana*. *Methods* **2019**.
37. Lee, B.; Flynn, R.; Kadina, A.; Guo, J.; Kool, E.; Chang, H., Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA* **2017**, *23* (2), 169-174.
38. Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y., Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnology* **2018**, *36*, 880.
39. Flynn, R. A.; Zhang, Q. C.; Spitale, R. C.; Lee, B.; Mumbach, M. R.; Chang, H. Y., Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nature protocols* **2016**, *11* (2), 273-290.
40. Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M., RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *Journal of the American Chemical Society* **2005**, *127* (12), 4223-4231.
41. McGinnis, J. L.; Dunkle, J. A.; Cate, J. H. D.; Weeks, K. M., The Mechanisms of RNA SHAPE Chemistry. *Journal of the American Chemical Society* **2012**, *134* (15), 6617-6624.
42. Mortimer, S. A.; Weeks, K. M., A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *Journal of the American Chemical Society* **2007**, *129* (14), 4144-4145.
43. Weeks, K. M.; Mauger, D. M., Exploring RNA structural codes with SHAPE chemistry. *Accounts of chemical research* **2011**, *44* (12), 1280-1291.

44. Poulsen, L. D.; Kielpinski, L. J.; Salama, S. R.; Krogh, A.; Vinther, J., SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA (New York, N.Y.)* **2015**, *21* (5), 1042-1052.
45. Watters, K. E.; Yu, A. M.; Strobel, E. J.; Settle, A. H.; Lucks, J. B., Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods (San Diego, Calif.)* **2016**, *103*, 34-48.
46. Rouskin, S.; Zubradt, M.; Washietl, S.; Kellis, M.; Weissman, J. S., Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **2014**, *505* (7485), 701-705.
47. Tijerina, P.; Mohr, S.; Russell, R., DMS footprinting of structured RNAs and RNA-protein complexes. *Nature protocols* **2007**, *2* (10), 2608-2623.
48. Spitale, R. C.; Crisalli, P.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y., RNA SHAPE analysis in living cells. *Nature chemical biology* **2013**, *9* (1), 18-20.
49. McGinnis, J. L.; Liu, Q.; Lavender, C. A.; Devaraj, A.; McClory, S. P.; Fredrick, K.; Weeks, K. M., In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. *Proceedings of the National Academy of Sciences* **2015**, *112* (8), 2425.
50. Smola, M. J.; Calabrese, J. M.; Weeks, K. M., Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **2015**, *54* (46), 6867-6875.
51. Busan, S.; Weidmann, C. A.; Sengupta, A.; Weeks, K. M., Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **2019**, *58* (23), 2655-2664.
52. Martin, S.; Blankenship, C.; Rausch, J. W.; Sztuba-Solinska, J., Using SHAPE-MaP to probe small molecule-RNA interactions. *Methods* **2019**, *167*, 105-116.
53. Smola, M. J.; Rice, G. M.; Busan, S.; Siegfried, N. A.; Weeks, K. M., Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols* **2015**, *10*, 1643.
54. Smola, M. J.; Weeks, K. M., In-cell RNA structure probing with SHAPE-MaP. *Nature Protocols* **2018**, *13*, 1181.
55. Incarnato, D.; Morandi, E.; Simon, L. M.; Oliviero, S., RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic acids research* **2018**, *46* (16), e97-e97.
56. Mathews, D. H.; Moss, W. N.; Turner, D. H., Folding and finding RNA secondary structure. *Cold Spring Harbor perspectives in biology* **2010**, *2* (12), a003665-a003665.
57. Spitale, R. C.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y., RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdisciplinary Reviews: RNA* **2014**, *5* (6), 867-881.
58. Schroeder, R.; Grossberger, R.; Pichler, A.; Waldsich, C., RNA folding in vivo. *Current Opinion in Structural Biology* **2002**, *12* (3), 296-300.
59. Yang, N. J.; Hinner, M. J., Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods Mol Biol* **2015**, *1266*, 29-53.
60. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Journal of medicinal chemistry* **1997**, *40* (3), 3-13. The article was

originally published in *Advanced Drug Delivery Reviews* 23 (1997) 3–25.1. *Advanced Drug Delivery Reviews* **2001**, 46 (1), 3-26.

61. Richter, M. F.; Drown, B. S.; Riley, A. P.; Garcia, A.; Shirai, T.; Svec, R. L.; Hergenrother, P. J., Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **2017**, 545 (7654), 299-304.

62. Weeks, K. M., Advances in RNA structure analysis by chemical probing. *Current Opinion in Structural Biology* **2010**, 20 (3), 295-304.

63. Spitale, R. C.; Flynn, R. A.; Zhang, Q. C.; Crisalli, P.; Lee, B.; Jung, J.-W.; Kuchelmeister, H. Y.; Batista, P. J.; Torre, E. A.; Kool, E. T.; Chang, H. Y., Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **2015**, 519 (7544), 486-490.

## CHAPTER 2: WATER-SOLUBLE ISATOIC ANHYDRIDES: A PLATFORM FOR RNA-SHAPE ANALYSIS AND PROTEIN BIOCONJUGATION

Adapted with permission from *Bioconjugate Chem.* 2018, 29, 9, 3196–3202.  
Copyright (2018) American Chemical Society.

### 2.1 ABSTRACT:

N-(3-iodopropyl) isatoic anhydride (IPIA) has been demonstrated to serve as an efficient substrate for the development of an extended bioconjugation platform. Derivatives of IPIA are water soluble, adaptable, and share a common chromophore rendering them easily quantifiable. We demonstrate the preparation of the readily diversified bioconjugation platform technology and application of the reagents in RNA SHAPE analysis.

### 2.2 INTRODUCTION

Bioconjugation is a cornerstone technology with wide ranging applications. The fundamental basis of bioconjugation is the formation of a stable, generally covalent linkage, between a biomolecule and a functionality. The ability to prepare bioconjugates led to the preparation of numerous, functional materials which have solidified bioconjugate technologies as a critical field. Bioconjugates are a major component in many biochemical assays, and are essential in several biopharmaceutical based applications such as Antibody Drug Conjugates (ADCs),<sup>1-3</sup> and in the burgeoning field of nanotherapeutics.<sup>4-8</sup> Continued development of bioconjugate technologies will rely on both research into applications and into the preparation of the bioconjugation reagents to fulfil the increasing technical demands.

Traditionally, bioconjugates have been produced using a select handful of chemistries which can react under physiological conditions.<sup>9</sup> These chemistries rely on reaction with native features of a biomolecule such as a nucleophilic amino acid or through engineered,

non-native handles.<sup>10-11</sup> The bioconjugation reagent is typically designed to be heterobifunctional, with a reactive centre which is covalently linked to a functionality. This strategy has been successful in the generation of many useful bioconjugates; despite this, the development of the bioconjugation reagent has been largely ignored. Generally, the bioconjugation reagent is viewed simply as a tool for the installation of functionality. Consequently, the bioconjugation reagent does not necessarily require additional functionality besides the reactive handle. Designing the bioconjugation reagent with intrinsic functionality such as a chromophore immediately offers several distinct advantages. The bioconjugate is then directly observable using the absorbance or fluorescent signature if the chromophore falls outside the biological window. This provides a quantitative ability to determine a successful reaction and can provide a trackable tag for purification.

The modification of a biomolecule often alters the solution dynamics of the native material;<sup>12</sup> frequently altering the iso-electric point (PI) and hydrophobicity or hydrophilicity. Modification of the lysine residue results in the loss of this protonated side chain residue, directly altering the PI and often altering the solubility. Regardless, the nucleophilic lysine residue remains the most common target for bioconjugation despite the lack of site specificity with most biomolecules. Additionally, it has been shown that preparation of biomolecules with organic co-solvents such as dimethyl sulfoxide (DMSO) can alter the biomolecules 3D, folded structure.<sup>13</sup> Many bioconjugation strategies rely on the use of DMSO to solvate the bioconjugation reagent. Upon conjugation, the solubility of the bioconjugate is largely depends on the biomolecule's innate solubility to overcome the insolubility of installed payload. These inherent issues led us to the development of a



bioconjugation platform with innate water solubility for the installation of functionality and the direct replacement of the lysine residue's positive charge under physiological conditions.

Recently, we have reported on the development of N-(N,N-dimethylaminoethylene) isatoic anhydride (DMIA) as the basis for a new platform for the construction of bioconjugation reagents.<sup>14</sup> N-alkylated derivatives of DMIA have several practical benefits for bioconjugation applications. Firstly, these DMIA derivatives are inherently water soluble due to the presence of a quaternary ammonium salt. Secondly, the DMIA platform has an intrinsic chromophore for single point quantification. Finally, the DMIA platform is adaptable and can readily include a variety of chemical linkers enabling optimization of linker composition. DMIA incorporates an electrophilic anhydride with a nucleophilic tertiary amine into a single chemical platform. Utilizing the difference in reactivity inherent to these two reactive centres, we have been able to successfully develop an assortment of bioconjugate reagents with broad application in a variety of fields. Our primary limitation in expanding the scope of this new platform has been the need for relatively reactive alkyl bromides or iodides with the desired functionality to react with the tertiary amine of DMIA. A general procedure for the incorporation of activated halides onto desirable chemical functionality for quaternization with DMIA looks unwieldy. This suggested that we flip the sense of the chemistry around, envisioning the incorporation of a reactive halide into the isatoic anhydride platform. The incorporation of the halide onto the isatoic anhydride backbone would provide a material with two electrophilic centres of differential reactivity.

We envisaged that incorporation of an active halide directly onto isatoic anhydride provided fewer difficulties when looking towards incorporation of a broad variety of

functionalities onto the platform. We reasoned that the neighbouring carbonyl would provide anchimeric assistance for the ensuing quaternization reaction. Additionally, many organic building blocks possess amines, alcohols or carboxylic acids. These functionalities all allow for the incorporation of a tertiary amine through differing synthetic pathways; although frequently each are done through simple and high yielding reactions. The ability to readily prepare multiple functionalities from commercially available building blocks which could then be quaternized with the installed halide provided justification to modify the existing platform to explore a complimentary strategy to expand the existing technology. We therefore saw the design of an isatoic anhydride prepared with the activated halide directly attached as much more applicable for the generation of bioconjugation reagents including nucleic acid applications.

## **2.3 RESULTS AND DISCUSSION**

### **2.3.1 Synthesis of IPIA.**

We envisioned a simple two-step synthesis starting with the alkylation of isatoic anhydride using commercially available 1-chloro-3-iodopropane followed by an *in situ* Finkelstein reaction to install the desired iodide (Scheme 2-1). The primary alkylation was carried out in anhydrous dimethylformamide (DMF) according to the procedure described in Scheme 1. During the formation of 1, the initial alkylation proceeds via displacement of the iodide resulting in quantitative formation of the chloride product and in-situ generation of sodium iodide (NaI). Contrary to the traditional requirements for the Finkelstein reaction, we observed that heating during the removal of DMF from our reaction mix facilitated clean conversion of the chloride product to the desired product, 1, via reaction

with the NaI generated from step 1. This simple one-pot synthesis has been used to generate **1** on a multi-gram scale yielding crystals suitable for x-ray crystallography (Figure 2-1).

Scheme 2-1. Preparation of IPIA (**1**) from isatoic anhydride

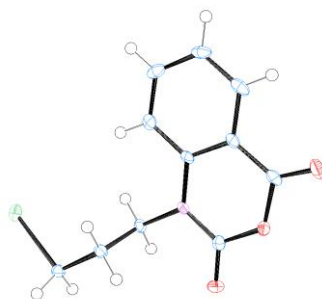
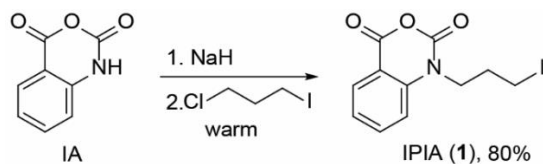
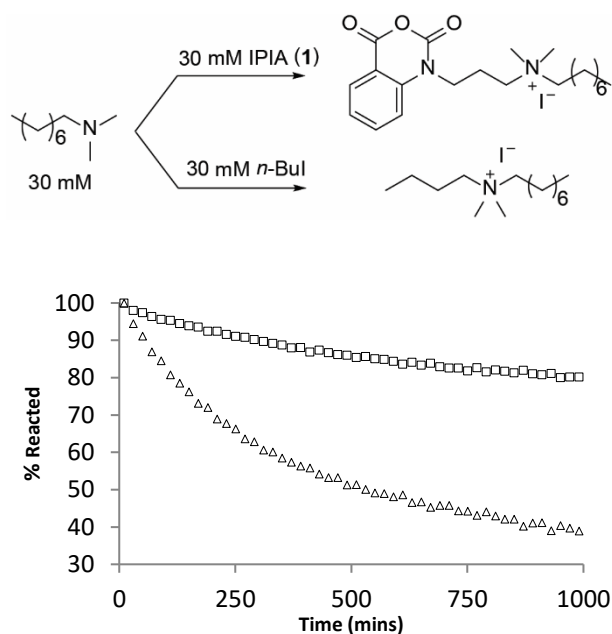


Figure 2-1: ORTEP diagram of **1**, space group  $P2_12_12_1$ , cell parameters  $a=5.1602(3)$  Å,  $b=6.9659(4)$  Å,  $c=31.110(2)$  Å.

Rate of reaction of **1** with tertiary amines. The primary iodide on **1** provides a reactive handle for the quaternization of a tertiary amine. The rate was measured in a competition assay of **1** vs *n*-BuI, a prototypical primary alkyl iodide, using proton NMR spectroscopy. N,N-dimethyl octylamine was chosen as a model tertiary amine for the reaction (Scheme 1-2). The disappearance of the two substrates, in acetone-*d*<sub>6</sub>, was monitored by integrating the peaks of the methylene hydrogens *alpha* to the iodine for the two substrates vs time and over 17 hrs (Figure 2-2). Compound **1** reacts 4-5 times faster than *n*-BuI at 25 °C.

Scheme 2-2. Competition Reaction of N,N-dimethyl octyl-amine for IPIA and *n*-BuIFigure 2-2: Reaction Progress for Competition Assay Measured by  $^1\text{H}$  NMR. Open squares (□) = *n*-BuI, Open triangles (△) = 1.

Under these conditions **1** easily out-competes iodobutane for N,N-dimethyloctylamine indicating a clear rate enhancement for the electrophilicity of **1** likely due to anchimeric assistance. Quaternization of **1** can be performed in acetone or tetrahydrofuran (THF) at 40 °C in a manner consistent with the procedure outlined for DMIA as previously reported.<sup>14</sup> This quaternization process occurs cleanly and in high yields. In many instances, the resulting quaternary ammonium salts precipitate readily from the reaction solution resulting in high purity products in a single step. Alternatively, when the quaternary ammonium salt derivatives have exhibited unwanted solubility in either acetone or THF we have been able to facilitate the precipitation of the resulting product by addition of Et<sub>2</sub>O without loss in product purity or yield. A variety of substrates were generated using the quaternization strategy (Table 2-1). The precipitation of the products from acetone often

gave x-ray quality crystals (Figure 1-3), **3b** gives crystals as the iodide where as **3c** gives crystals as the bromide.

### 2.3.2 Ion exchange.

During the use of these compounds, **3a-3e**, it was noted that the products from the DMIA platform were typically more soluble than the products from IPIA. Suspecting this was likely due the iodide counter ion, we exchanged the iodide for a chloride or bromide after synthesis using ion exchange resins. This resulted in a notable increase in solubility

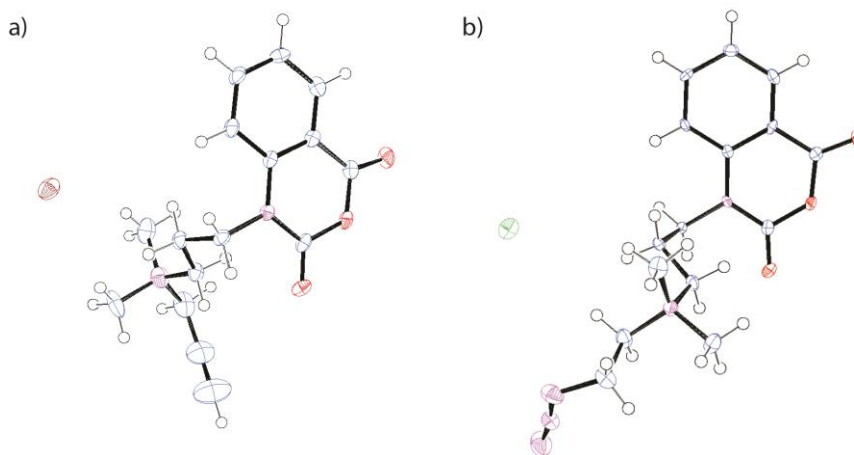


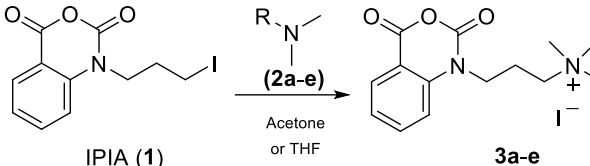


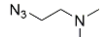
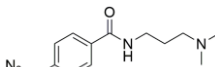
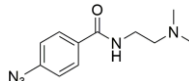
Figure 2-3: a) ORTEP diagram of **3b**, space group  $P2_1/c$ , cell parameters  $a=14.1392(11)$  Å,  $b=9.2466(7)$  Å,  $c=13.3238(9)$  Å.  $\beta=99.099(7)^\circ$  b) ORTEP diagram of **3c**, space group  $P2_1/c$   $a=15.21142(9)$  Å,  $b=8.7274(7)$  Å,  $c=13.3470(8)$  Å,  $\beta=99.549(6)^\circ$ .

of all derivatives, rendering all derivatives soluble at greater than 65 mM for the chloride exchanged derivatives.

### 2.3.3 Reactivity of prepared reagents.

The relative reactivities of chloride exchanged **3b** and **3c** were demonstrated by reaction with bovine serum albumin (BSA). BSA served as a model protein which

Table 2-1. Preparation of IPIA Derivatives at 40 °C.

					
	IPIA (1)	R-N(Me) <sub>2</sub>	$\epsilon^*$	Yield <sup>†</sup>	Solubility <sup>+</sup>
1		3600	68 (3a)	25.6	
2		3600	78 (3b)	2.4	
3		3600	65 (3c)	2.2	
4		4200	59 (3d)	0.3	
5		4300	66 (3e)	0.3	

\* Measured in M<sup>-1</sup>cm<sup>-1</sup>, uncertainty was  $\pm 100$   
<sup>†</sup> Reported as percent isolated yield  
<sup>+</sup> Reported in mM as iodide counter anions

possesses a high content of surface exposed lysine residues for reaction with the anhydride of the quaternary compounds (Figure 2-4) The modification of BSA proceeds rapidly and efficiently. Quantification of the degree of labeling (DOL) is performed using a laboratory spectrophotometer as described in the provided supplemental information. The DOL represents an average of the resulting bioconjugate, which possess inherent heterogeneity due to the many lysine residues available for reaction. It was noted that under identical reaction conditions the two compounds had the same reactivity, the functionality is well removed from the reactive center of the molecule providing similar reactivity for the anhydride. This separation results in the ability to develop a reproducible procedure for a given biomolecule to prepare a bioconjugate with controllable degree of labeling. This is a distinct advantage for the development of platform based bioconjugation. We have

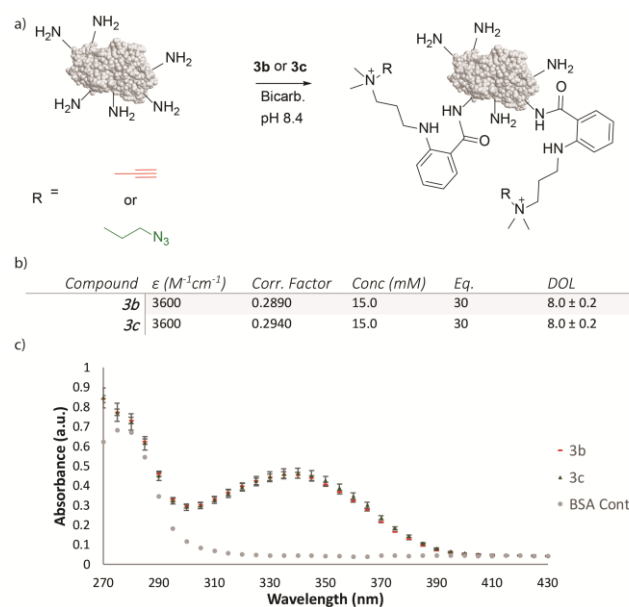


Figure 2-4: a) Scheme for labeling of BSA (0.15 mM) at 25 °C in 25 mM bicarbonate buffer (pH 8.4). b) Data table used in calculation of DOL. c) Triplicate data for BSA labeling reactions using 3b and 3c for 1 h at 25 °C.

previously demonstrated the effectiveness of the water-soluble isatoic anhydride platform in the CuAAC reaction.<sup>14</sup>

### 2.3.4 Determination of aqueous stability.

The half-life of the reagents was tested using **3a** as a representative compound. The reagents were monitored using <sup>1</sup>HNMR, watching the disappearance of a downfield aromatic signal and its reappearance upfield due to the hydrolysis of the anhydride ring (Figure S1). The half-life of the reactive anhydride was determined in D<sub>2</sub>O and 25 mM bicarbonate buffer pH 8.4 D<sub>2</sub>O for the bioconjugation reactions. The half-lives were found to be 50 h in D<sub>2</sub>O and 237 min at pH 8.4 respectively at room temperature. The stability of the anhydride provides necessary time to prepare the reagents in aqueous solutions for bioconjugation reactions.

### 2.3.5 RNA-SHAPE analysis.

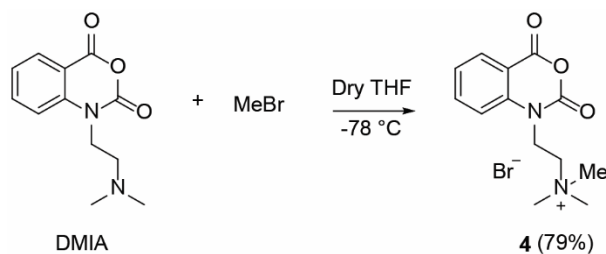
RNA has an inherent ability to form diverse secondary and tertiary structure either by base pairing with itself or interacting with other biomolecules inside the cell. By folding into complex structure, different classes of RNAs are able to perform essential biological functions. Determining the RNA structure thus provide important information about their evolution and key functions they are executing inside the cell. Chemical and enzymatic probing are the most common assays available to determine the structure of RNA at nucleotide resolution.<sup>15-18</sup> Chemical probing can be performed either by using dimethylsulfate (DMS) which causes RNA methylation of unpaired adenosine or cytosine nucleotides<sup>19</sup> or by Selective 2'-Hydroxyl Acylation analysed by Primer Extension (SHAPE). SHAPE involves targeting of the 2'-OH group of nucleotides present in single stranded or flexible RNA using chemical probes. Electrophiles like N-methylisatoic anhydride (NMIA), 2-methylnicotinic acid imidazolid (NAI), 1-methyl-7-nitro-isatoic anhydride (1M7) are used to form a 2'-O-adduct to probe secondary structures of RNA. These adducts can then be identified by the primer extension reaction where they block the Reverse Transcriptase enzyme from carrying out cDNA elongation, followed by electrophoretic separation by running denaturing gel.<sup>15</sup> Most of the commonly used RNA-SHAPE reagents are sparingly water soluble, making it challenging to perform RNA structure probing in true aqueous conditions. In addition, these chemicals have very short half-lives varying from  $t_{1/2} \approx 17s$  for 1M7 to  $t_{1/2} \approx 30min$  for NAI.<sup>20</sup> DMSO, the most commonly used solvent for the chemical probes is known to have small but significant effect on RNA structures<sup>21</sup> resulting in non-specific background data. DMSO is used for the delivery of the probe due to the insolubility of the reagent as well as the short half-life.



The half-life of NMIA under SHAPE conditions<sup>22</sup> is reported to be 260 s, intermediate between 1M7 and NAI. For comparison, the half-life for **3a** was determined under SHAPE conditions. To this end 3X SHAPE buffer, pH 8.0 was prepared in H<sub>2</sub>O. This was diluted to 1x with D<sub>2</sub>O, giving final concentrations of 111 mM HEPES, 111 mM NaCl, and 6.7 mM MgCl<sub>2</sub> and equilibrated to 37 °C. **3a** was dissolved into the 1X buffer and <sup>1</sup>H NMR was used to monitor the half-life. Under the SHAPE conditions the half-life of **3a** was found to be 4.5 min, not surprisingly, very similar to NMIA. A short half-life has been previously shown as advantageous for reagents for SHAPE analysis.<sup>23</sup> These differing reaction conditions demonstrate the wide applicability of the reagents to both in bioconjugation and RNA SHAPE analysis.

To investigate the ability of water-soluble isatoic anhydride derivatives to probe the secondary structures of RNA we chose to include a compound from the previously reported DMIA platform. We decided to alter the preparation from the previously reported synthesis to eliminate the need to perform ion exchange due to reduced solubility when prepared using methyl iodide. We chose to prepare the methyl derivative of DMIA using the simple synthetic scheme depicted below (Scheme 2-3). This synthesis took advantage of reactivity of methyl bromide to perform quaternization with DMIA whereas typical reactions with DMIA typically required the use of primary iodides or activated bromides.

Scheme 2-3: Preparation of **4** from DMIA



The quaternary ammonium isatoic anhydride derivatives are water-soluble, unlike the previous isatoic anhydride SHAPE reagents. NMIA has been used less frequently for SHAPE analysis than some of the newer generation materials which offer increased reactivity and solubility. We therefore chose to compare **4**, **3a** and **3c** as SHAPE reagents side-by-side with NAI, a commonly used chemical probe, to determine their reactivity in RNA-SHAPE (Figure 2-5a). For this purpose, we synthesized U3 snoRNA from malaria pathogen *Plasmodium falciparum*.<sup>24</sup>

U3 is a highly conserved snoRNA, which is required for processing of the small subunit (SSU) ribosomal RNA (rRNA) in eukaryotes.<sup>25-26</sup> A generic structure of U3 RNA contains a 5' and a 3' domain connected by a hinge region (Figure 2-5b). While 5' domain is required for processing of rRNA by U3, the 3' domain structures are important for protein binding and localization. Especially, the Stem-loop II, box B and C are known to interact with 15.5K and hU3-55K proteins in higher eukaryotes for proper localization of the RNA.<sup>27</sup> Stem III of U3 RNA has U-rich internal and terminal loops and required for proper folding of RNA. The Plasmodium U3 has a unique stem III, which is longer and lacks any sequence homology with yeast or vertebrate U3 RNAs.<sup>24</sup> We performed *in vitro* SHAPE using synthetic PfU3 snoRNA to map the above stem-loop II and III of PfU3 by adding desired concentration of electrophiles and incubating with the RNA for both 5 min and 15 min respectively. One major limitation of NAI is that the minimum amount required to detect modification on RNA structures is 20 mM,<sup>19</sup> which is what we observed as well. In our hand, 15 mM of NAI failed to efficiently modify the RNA when compared to unmodified DMSO lane (Figure 2-5a). However, our compounds at 15 mM concentration revealed distinct modification of unpaired RNA nucleotides compared to control in the denaturing

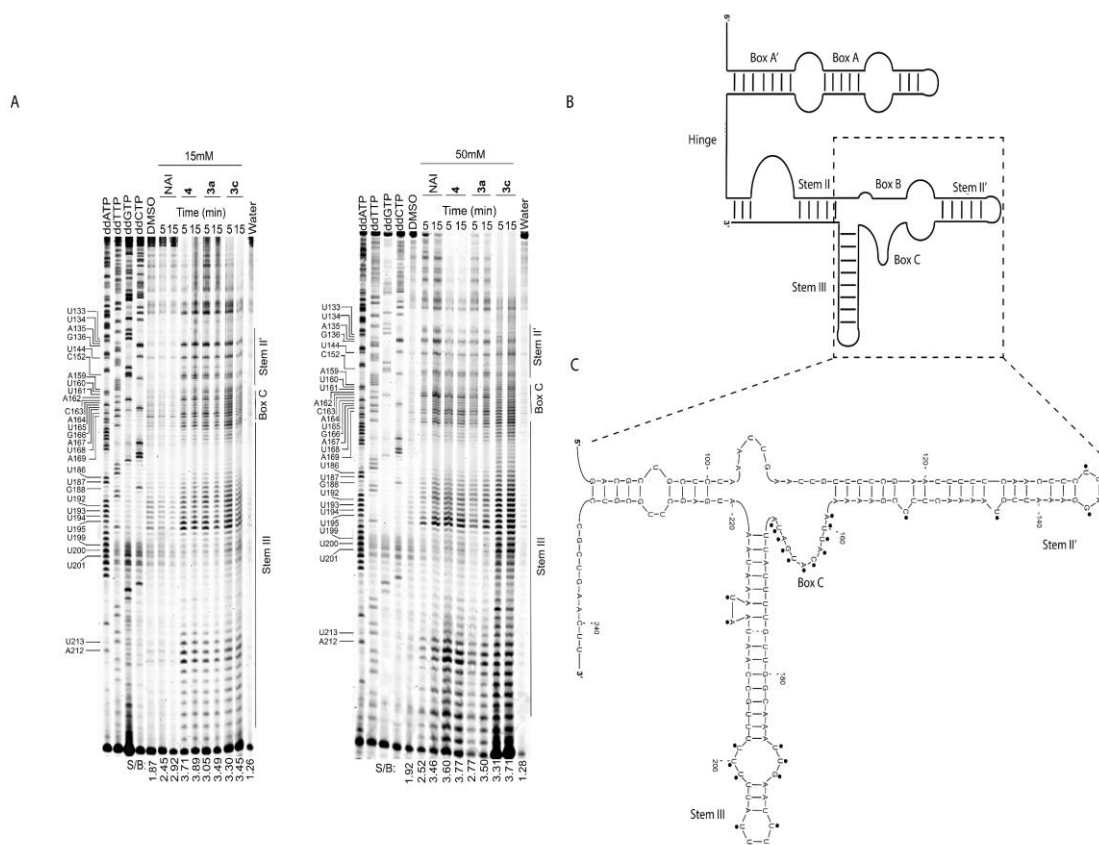


Figure 2-5: SHAPE-RT probing by various SHAPE electrophiles and secondary structure of U3 snoRNA (A) Cy5 labeled denaturing gel comparing modification in the synthetic pfU3 snoRNA using various SHAPE electrophiles at 15 mM (left panel) and 50 mM (right panel) concentration. Lower panel represents the signal-to-background (S/B) ratios calculated for each lane. (B) Secondary structure of U3 snoRNA depicting different domains. (C) Secondary structure of Stem II', Box C and Stem III of PfU3 snoRNA with the SHAPE reactivities mapped on the unpaired residues. An enlarged version for nucleotide sequence found in supplemental (Figure S3A and S3B).

gel (Figure 2-5a). Also, increased reactivity of these compounds compared to NAI was evident from high signal-to-background ratio calculated by measuring the band intensity. To determine if U3 RNA structure profile produced by our compounds are comparable to that of NAI, we further probed the U3 structure using 50 mM of our compounds and NAI. Our data shows that the stops or modifications generated by our compounds are well correlated with that of 50 mM NAI signatures, suggesting that all the three chemicals have

the ability to accurately determine the RNA structure *in vitro* (Figure 2-5a). However, we observed some degradation of RNA when **3c** was used at 50 mM concentration for modification. Nevertheless, the modifications generated by our water-soluble compounds were well-defined for quantitative measurement of the band from the denaturing gel. Based on the U3RNA modification from our SHAPE assay we derived a secondary structure model of PfU3 using Mfold.<sup>28</sup> As expected, most of these modifications lie on flexible residues that are not involved in base pairing interactions (Figure 5c). Thus, our reagents demonstrate robust and reproducible results in modifying RNA outside the cell and hold promise for SHAPE reactivity inside the cell due to its reasonably high stability in aqueous solution. Substantial interest has been shown in for the of SHAPE analysis for *in vivo* probing of structure,<sup>20</sup> these reagents may be ideal for work *in vivo* due to their inherent water solubility which negates the need for destructive co-solvents such as DMSO. *In vivo* application of these reagents are planned.

## 2.4 CONCLUSIONS

The synthesis of a **1** has been demonstrated. This reagent can easily be quaternized with a variety of tertiary amines to produce quaternary ammonium salts which render the compound water soluble and facilitates easy purification. These water-soluble derivatives are suitable for bioconjugation of proteins and RNA for SHAPE analysis. The preparation of complementary reagents, azide and alkyne allow for the use of these reagents for either half or both halves of the CuAAC click chemistry. All quaternary derivatives possess reasonable water solubility and absorb at 330 nm, outside the main biological window, allowing for use in quantifiable bioconjugation. Additionally, the isatoic anhydride backbone of the reagents fluoresces at 420 nm, giving a secondary handle which can be

used for tracking of the bioconjugate. Successful *in vitro* SHAPE reactivity of U3 snoRNA was demonstrated using the water soluble isatoic anhydride derivatives. The newly introduced compounds showed a high reactivity, having observably labeled the 2' OH of RNA within a minute (Figure S4) and providing a higher signal to background ratio than NAI. Additionally, the reagents can be prepared in aqueous conditions where as NAI required preparation in DMSO followed by dilution into RNA buffers. These reagents are suitable for numerous applications which can be performed without the need for cosolvents such as DMSO. Inherent water solubility of quaternary isatoic anhydride derivatives offer distinct advantages compared to traditional bioconjugation strategies.

## **2.5 EXPERIMENTAL**

### **2.5.1 Chemicals.**

All compounds were purchased commercially and used without further purification unless otherwise noted. Solvents were dried and stored over 4 Å molecular sieves. Buffers were prepared using 18 Ω water.

### **2.5.2 Instrumentation.**

NMR spectra was recorded on a JEOL-ECA 500 MHz NMR spectrometer using acetone-d<sub>6</sub>, DMSO-d<sub>6</sub>, and D<sub>2</sub>O. Spectra were referenced to the respective solvent peaks for <sup>1</sup>H and <sup>13</sup>C. UV-vis data was obtained using a Molecular Devices Spectramax M5 plate reader in 25 mM bicarbonate pH 8.4. Routine ESI MS data were collected on a Thermo-Fisher MSQ Plus single quadrupole mass spectrometer. High resolution MS data were analyzed on an Orbitrap XL in positive mode at a resolution of 60000, spray voltage of 5 kV, sheath 12, Aux 10, Sweep 0, Capillary Temp 275 °C, Capillary Voltage 32.00 V, Tube Lens of 85.00 V Compounds were dissolved in 1 mL ACN, then 8 μL was diluted in 1:1

H<sub>2</sub>O:ACN with 0.1% Formic acid. 2  $\mu$ L of this solution was injected using an Accela 1250 LC system, at a solvent flowrate of 200  $\mu$ L/min of 1:1 H<sub>2</sub>O:ACN with 0.1% Formic Acid. X-ray data were collected on an Agilent/Oxford Diffraction Gemini A Ultra diffractometer utilizing Mo radiation ( $\lambda = 0.71073$  Å) at 100 K.

### 2.5.3 Synthesis.

The synthesis of compounds 1, 2c-e, 3a-e, and 4 are described in the Supporting information.

### 2.5.4 Bioconjugation of BSA.

Bovine serum albumin was prepared at 0.15 mM in 25 mM bicarbonate buffer, pH 8.4. 100  $\mu$ L of BSA solution was placed in an Eppendorf tube. Quaternary labeling reagents (3b or 3c) were dissolved into DI H<sub>2</sub>O to a concentration of 15 mM immediately prior to addition to the reaction. 30  $\mu$ L of labeling reagent solution was added to the reaction mixture and gently vortexed. The Eppendorf is capped and placed in a 25 °C oven for 1 h at which point 5  $\mu$ L of NH<sub>4</sub>OH was added to end the reaction by reaction with the excess labeling reagent. 100  $\mu$ L of the reaction was purified via centrifugation using a Princeton Separations Pro-Spin CS-800 size exclusion spin column. 20  $\mu$ L of the purified solution was diluted with 480  $\mu$ L of 25 mM pH 8.4 bicarbonate solution and the absorbance or fluorescence is measured on a Molecular Devices Spectramax M5 plate reader using a 500  $\mu$ L quartz cuvette. Sample was then recovered and concentrated via lyophilization or centrifugal filtration units.

### 2.5.5 PCR amplification and In Vitro Transcription:

Primary PCR amplification of *Plasmodium falciparum* U3 small nucleolar RNA (snoRNA)<sup>24</sup> was done from genomic DNA of *Plasmodium falciparum* 3D7 strain using a

forward primer containing T7 RNA polymerase promoter sequence (underlined) at the 5' end and reverse primer as indicated in Table 1-2. The resultant PCR amplified DNA of PfU3 snoRNA was In Vitro Transcribed (IVT) and purified using the MEGAscript and MEGAclean kits (Life Technologies, USA) following the manufacturer's recommendation.

Table 2-2: List of primers used to amplify PfU3 snoRNA

Name	Sequence
Forward primer	5'ATATCCATGGGACATAATACGACTCACTATAGGGAG <u>AGTTGACGAATACTTAGTCAAGATCATTGC</u> 3'
Reverse Primer	5'GCATGTCGACAAGTTCAGCGGACGCAAGCTCAT TTA 3'
RT primer	5' cy5- AAG TTC AGC GGA CGC AAG CTC-3'

### 2.5.6 In Vitro SHAPE modification:

For SHAPE modification of PfU3snoRNA, 5 µg of *in vitro* Transcribed RNA was denatured by heating at 95 °C for 2 min followed by immediate incubation in ice for 3 min. The modification was carried out by incubation of the RNA in 3X SHAPE buffer consisting of 333 mM of NaCl, 333 mM of HEPES and 20 mM of MgCl<sub>2</sub> at 37 °C prior to the addition of desired concentration of 2-methylnicotinic acid imidazolid (NAI), 4, 3a and 3c for both 5 min and 15 min. Following incubation, the modified RNA was ethanol precipitated, dissolved in water and subjected to nanodrop analysis.

### 2.5.7 cDNA synthesis and RNA-SHAPE analysis by denaturing Polyacrylamide Gel Electrophoresis (PAGE):

1.6 µg of SHAPE modified RNA was denatured by heating at 95 °C for 2 min followed by immediate incubation in ice for 3 min. For Primer extension, 5' cy5 labeled primer for reverse transcription as indicated in Table 2 along with 10 mM dNTPs were

mixed with the SHAPE modified RNA and incubated at 65 °C for 5 min. Four dideoxy sequencing reactions using ddNTPs were also performed in parallel using the unmodified PfU3 snoRNA in the similar manner as mentioned above. Following incubation all the reactions were supplemented with 1X SuperScript II First-strand buffer (50mM Tris-HCl, pH 8.3, 75 mM KCl and 5 mM MgCl<sub>2</sub>), 10mM DTT and 1 µL of RiboLock RNase inhibitor and incubated at 42 °C for 2 min. After addition of SuperScript II Reverse Transcriptase enzyme, the whole reaction was incubated at 42 °C for 60 min, followed by inactivation at 70 °C for 15 min. The, resultant mixtures of cDNA were precipitated by adding isopropanol and then dissolved in water. cDNA generated from primer extension were mixed with 80% formaldehyde, heated at 80 °C for 3 min and then resolved by 12% denaturing PAGE gel. The gel was scanned using Typhoon phosphorimager and the band intensity for each lane was quantified using SAFA software.<sup>29</sup> Nucleotide position for each band was identified from dideoxy sequencing lane. Secondary structure of RNA was generated by Mfold and the SHAPE reactivities of accessible nucleotide were mapped on it.<sup>28</sup>

#### **2.5.8 Signal-to-background ratio calculation:**

Signal-to-background ratio (S/B) is the arithmetic mean of intensity of bands for each lane (except the dideoxy sequencing reactions control lanes) to the background. Mean Pixel intensity of randomly chosen three bands for each lane was calculated using ImageJ software and divided by the mean pixel intensity of background which is the low intensity region present adjacent to each band. Same position bands were used for calculation for each lane.

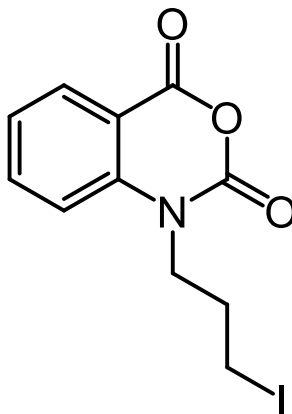


## 2.6 REFERENCES

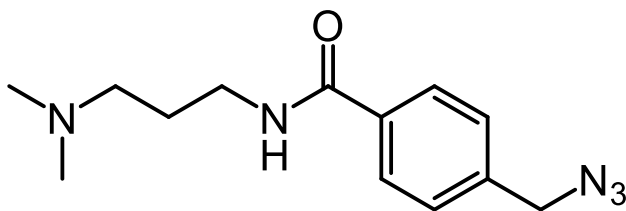
- (1) Levengood, M. R.; Zhang, X.; Hunter, J. H.; Emmerton, K. K.; Miyamoto, J. B.; Lewis, T. S.; Senter, P. D., (2017) Orthogonal Cysteine Protection Enables Homogeneous Multi-Drug Antibody–Drug Conjugates. *Angew. Chem. Int. Ed.* **56** (3), 733-737.
- (2) Diamantis, N.; Banerji, U., (2016) Antibody-drug conjugates—an emerging class of cancer treatment. *Brit. J. Cancer* **114**, 362.
- (3) Chudasama, V.; Maruani, A.; Caddick, S., (2016) Recent advances in the construction of antibody–drug conjugates. *Nat. Chem.* **8**, 114.
- (4) Dhar, S.; Gu, F. X.; Langer, R.; Farokhzad, O. C.; Lippard, S. J., (2008) Targeted delivery of cisplatin to prostate cancer cells by aptamer functionalized Pt(IV) prodrug-PLGA–PEG nanoparticles. *Proc. Natl. Acad. Sci.* **105** (45), 17356.
- (5) Gallina, M. E.; Zhou, Y.; Johnson, C. J.; Harris-Birtill, D.; Singh, M.; Zhao, H.; Ma, D.; Cass, T.; Elson, D. S., (2016) Aptamer-conjugated, fluorescent gold nanorods as potential cancer theradiagnostic agents. *Mater. Sci. Eng. C* **59**, 324-332.
- (6) Nevala, W. K.; Butterfield, J. T.; Sutor, S. L.; Knauer, D. J.; Markovic, S. N., (2017) Antibody-targeted paclitaxel loaded nanoparticles for the treatment of CD20+ B-cell lymphoma. *Sci. Rep.* **7**, 45682.
- (7) Shi, J.; Kantoff, P. W.; Wooster, R.; Farokhzad, O. C., (2016) Cancer nanomedicine: progress, challenges and opportunities. *Nat. Rev. Cancer* **17**, 20.
- (8) Xie, J.; Chen, K.; Lee, H.-Y.; Xu, C.; Hsu, A. R.; Peng, S.; Chen, X.; Sun, S., (2008) Ultrasmall c(RGDyK)-Coated Fe<sub>3</sub>O<sub>4</sub> Nanoparticles and Their Specific Targeting to Integrin  $\alpha\beta$ 3-Rich Tumor Cells. *J. Am. Chem. Soc.* **130** (24), 7542-7543.
- (9) Hermanson, G. T., (2013) *Bioconjugate techniques*. 3rd ed. ed.; Academic Press: London.
- (10) Hallam, T. J.; Wold, E.; Wahl, A.; Smider, V. V., (2015) Antibody Conjugates with Unnatural Amino Acids. *Mol. Pharm.* **12** (6), 1848-1862.
- (11) Greene, M. K.; Richards, D. A.; Nogueira, J. C. F.; Campbell, K.; Smyth, P.; Fernández, M.; Scott, C. J.; Chudasama, V., (2018) Forming next-generation antibody–nanoparticle conjugates through the oriented installation of non-engineered antibody fragments. *Chem. Sci.* **9** (1), 79-87.
- (12) McCombs, J. R.; Owen, S. C., (2015) Antibody Drug Conjugates: Design and Selection of Linker, Payload and Conjugation Chemistry. *AAPS J.* **17** (2), 339-351.
- (13) Arakawa, T.; Kita, Y.; Timasheff, S. N., (2007) Protein precipitation and denaturation by dimethyl sulfoxide. *Biophy. Chem.* **131** (1), 62-70.
- (14) Fessler, A.; Garmon, C.; Heavey, T.; Fowler, A.; Ogle, C., (2017) Water-soluble and UV traceable isatoic anhydride-based reagents for bioconjugation. *Org. Biomol. Chem.* **15** (45), 9599-9602.
- (15) Ehresmann, C.; Baudin, F.; Mougél, M.; Romby, P.; Ebel, J. P.; Ehresmann, B., (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15** (22), 9109-9128.
- (16) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M., (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protocols* **1**, 1610.
- (17) Weeks, K. M., (2010) Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Bio.* **20** (3), 295-304.

- (18) Ding, F.; Lavender, C. A.; Weeks, K. M.; Dokholyan, N. V., (2012) Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat. Methods* 9, 603.
- (19) Peattie, D. A.; Gilbert, W., (1980) Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. USA* 77 (8), 4679-4682.
- (20) Spitale, R. C.; Crisalli, P.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y., (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Bio.* 9 (1), 18-20.
- (21) Lee, J.; Vogt, C. E.; McBairty, M.; Al-Hashimi, H. M., (2013) Influence of Dimethylsulfoxide on RNA Structure and Ligand Binding. *Anal. Chem.* 85 (20), 9692-9698.
- (22) Watters, K. E.; Yu, A. M.; Strobel, E. J.; Settle, A. H.; Lucks, J. B., (2016) Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* 103, 34-48.
- (23) Mortimer, S. A.; Weeks, K. M., (2007) A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc.* 129 (14), 4144-4145.
- (24) Chakrabarti, K.; Pearson, M.; Grate, L.; Donohue, J. P.; Ares Jr, M.; Sterne-Weiler, T.; Deans, J., (2007) Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA* 13 (11), 1923-1939.
- (25) Hughes, J. M.; Ares, M., (1991) Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.* 10 (13), 4231-4239.
- (26) Mougin, A.; Grégoire, A.; Banroques, J.; Ségault, V.; Fournier, R.; Brulé, F.; Chevrier-Miller, M.; Branlant, C., (1996) Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA* 2 (11), 1079.
- (27) Granneman, S.; Pruijn, G. J. M.; Horstman, W.; Van Venrooij, W. J.; Lührmann, R.; Watkins, N. J., (2002) The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA in vitro. *J. Biol. Chem.* 277 (50), 48490-48500.
- (28) Zuker, M., (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids res.* 31 (13), 3406.
- (29) Das, R.; Laederach, A.; Pearlman, S. M.; Herschlag, D.; Altman, R. B., (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* 11 (3), 344.

## APPENDIX I: SUPPLEMENTARY INFORMATION.

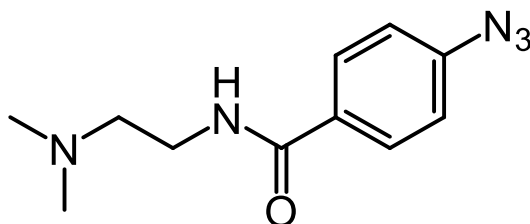


**Preparation of 1.** To a Schlenk flask under dry  $N_2$  was added 60% wt/wt NaH suspended in oil (1.5 g, 36 mmol). To this was added 15 ml of dry hexanes, the suspension briefly stirred then allowed to settle, and the hexanes removed via cannula. This process was repeated 3 x. To the freshly rinsed NaH was added anhydrous DMF (30 ml) at room temperature, resulting in a cloudy suspension. To this suspension was added isatoic anhydride (5.1 g, 30 mmol) in one portion. Following this addition, an additional 20 ml of anhydrous DMF was added and the resulting suspension was stirred for 60 min, upon which time 1-chloro-3-iodopropane was added (36 mmol) in one portion. The resulting suspension was stirred 12 h yielding a clear and colored solution containing the desired product. The resulting solution is concentrated under vacuum at 100 °C resulting in a darkly colored viscous residue. This residue is then dissolved into DCM (150 ml) and extracted 3 x with 100 ml of saturated  $NaHCO_3$  (aq). Finally, the organic layer is rinsed 1 x wash with 100 ml of brine solution. The organic layer is then collected and stirred with activated carbon (0.5 g) for 30 min. The resulting organic solution is then filtered through a plug of  $MgSO_4$  to both remove the activated carbon and dry the solution of residual water. The resulting solution is then concentrated under vacuum to remove the DCM resulting in a lightly colored solid that is then recrystallized from isopropanol. The final product (6.19 g, 62%) was isolated as an off white solid.  $C_{11}H_{10}N_2O_3I$ ,  $^1H$  NMR (500 MHz, acetone- $d_6$ ):  $\delta$  2.34 (m, 2H), 3.44 (t, 2H,  $J = 7.1$  Hz), 4.23 (t, 2H), 7.36 (t, 1H), 7.55 (d, 1H), 7.87 (t, 1H), 8.08 (d, 1H) ppm,  $^{13}C$  NMR (500 MHz, acetone- $d_6$ ):  $\delta$  2.4, 31.9, 45.9, 112.9, 115.2, 124.4, 130.8, 137.9, 142.6, 148.7, 159.6 ppm, EI MS: Calc. 330.97. Found: 331.0 ( $M^+$ , 11%), 132.1 (100%)  $m/z$ .



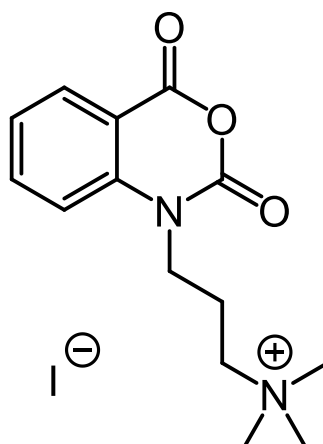
**2d.** To a dry round bottom flask equipped with a stir bar was added 4-(azidomethyl)benzoic acid<sup>2</sup> (355 mg, 2.0 mmol) and 1,1'-carbonyldiimidazole (360 mg, 2.2 mmol) followed by

the addition of 10 mL dry THF. This mixture was stirred until dissolution and an additional 1 h. To this crude solution was added ~1 mL H<sub>2</sub>O followed by N,N-dimethyl-1,3-propane diamine (0.40 mL, 3.2 mmol). The reaction mixture was stirred for 2 h, and the resulting solution was poured into a saturated NaHCO<sub>3</sub> solution. This solution was extracted 3x with 20 mL benzene. The organic layers were combined and extracted 3x with saturated NaHCO<sub>3</sub>. The organic layer was isolated and stirred over NaSO<sub>4</sub>. The solution is filtered and concentrated under reduced pressure. The product is isolated as a pale yellow oil, (418 mg, 80 %). C<sub>13</sub>H<sub>19</sub>N<sub>5</sub>O, <sup>1</sup>H NMR (500 MHz, acetone-d<sub>6</sub>): δ 1.75 (m, 2H), 2.20 (s, 6H), 2.38 (t, 2H), 3.44 (m, 2H), 4.52 (s, 2H), 7.47 (d, 2H), 7.89 (d, 2H), 8.25 (s, 1H) ppm, <sup>13</sup>C NMR (500 MHz, acetone-d<sub>6</sub>): δ 28.0, 40.4, 46.3, 55.1, 59.4, 128.8, 129.6, 136.5, 146.4, 167.0 ppm, ESI MS: M+ Calc. 262.17, found 262.19.

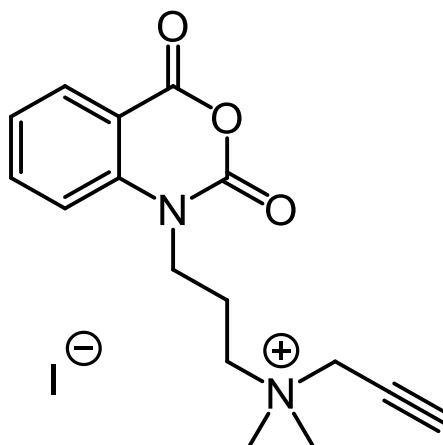


**2e.** To a dry round bottom flask equipped with a stir bar was added 4-azidobenzoic acid (330 mg, 2.0 mmol) and 1,1-carbonyldiimidazole (360 mg, 2.2 mol) followed by 10 mL dry THF. The reaction was stirred for 1 h at room temperature. To this crude solution was added ~1 mL H<sub>2</sub>O followed by N,N-dimethyl-1,2-ethanediamine (0.35 mL, 3.2 mmol). The reaction was stirred for 2 h, and the resulting solution was combined with 20 mL saturated NaHCO<sub>3</sub>. This solution is extracted 3 times with 20 mL benzene. The organic layers are combined and extracted 3x with saturated NaHCO<sub>3</sub>. The organic layer was collected and stirred over NaSO<sub>4</sub>. The solution is filtered and concentrated under reduced pressure. The product is isolated as a pale white solid, (384 mg, 82 %). C<sub>11</sub>H<sub>15</sub>N<sub>5</sub>O, <sup>1</sup>H NMR (500 MHz, benzene-d<sub>6</sub>): δ 1.89 (s, 6H), 2.04 (m, 2H), 3.40 (m, 2H), 6.42 (s, 1H), 6.56 (d, 2H), 7.65 (d, 2H) ppm, <sup>13</sup>C NMR (500 MHz, benzene-d<sub>6</sub>): δ 37.7, 45.4, 58.4, 119.4, 129.4, 132.2, 143.3, 165.8 ppm, ESI MS: M+ Calc. 234.13, found 234.17.

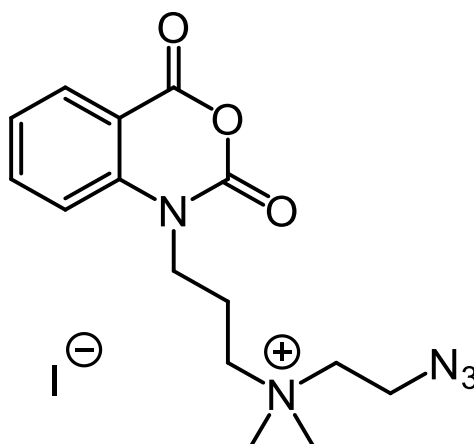
**General procedure 1.** To a 3 ml screw capped vial containing compound **1** (100 mg, 0.3 mmol) was added 1 ml of reagent grade acetone and the solution heated to completely dissolve the reagent then filtered through a 0.2 μm syringe filter. To this solution was added the corresponding nucleophile (0.3 mmol) and the solution was allowed to sit for 18 hrs at 40 °C. The solution was then stored at -20 °C overnight followed by filtration of the product. If no solid formed the Et<sub>2</sub>O was introduced to precipitate the quaternary ammonium salt.



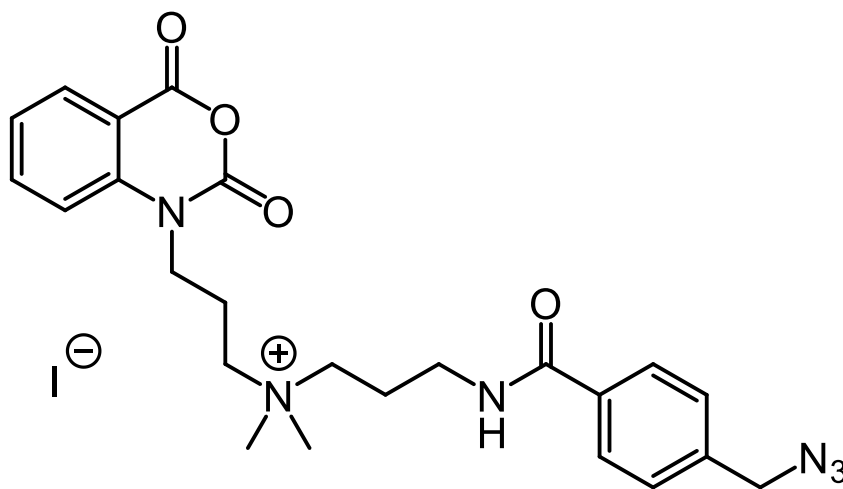
**3a.** General procedure 1 was followed as above using **2a** (trimethylamine) in excess as the nucleophile. The final product (80 mg, 68%) was isolated as a light yellow solid.  $C_{14}H_{19}N_2O_3I$ ,  $^1H$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  2.12 (m, 2H), 3.04 (s, 9H), 3.48 (m, 2H), 4.14 (t, 2H), 7.39 (t, 1H), 7.57 (d, 1H), 7.90 (t, 1H), 8.07 (d, 1H),  $^{13}C$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  20.9, 41.8, 52.5, 62.9, 112.0, 114.8, 124.3, 130.0, 137.6, 141.4, 148.3, 159.2. HRMS:  $M^+$  calc. 263.1396, found 263.1388  $m/z$ .



**3b.** General procedure 1 was followed as above using **2b** (N,N-dimethylprop-2-yn-1-amine) (36  $\mu$ L, 0.33 mmol) as the nucleophile. The final product (98 mg, 78%) was isolated as a light yellow solid.  $C_{16}H_{19}N_2O_3I$ ,  $^1H$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  2.14 (m, 2H), 3.07 (s, 6H), 3.56 (m, 2H), 4.08 (t, 1H), 4.13 (t, 2H), 4.37 (d, 2H), 7.38 (t, 1H), 7.57 (d, 1H), 7.89 (t, 1H), 8.05 (d, 1H) ppm,  $^{13}C$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  20.5, 41.3, 49.9, 53.9, 60.7, 72.2, 83.1, 111.8, 114.7, 123.8, 129.7, 137.3, 141.1, 147.8, 158.8 ppm. HRMS:  $M^+$  calc. 287.1396, found 287.1391  $m/z$ .

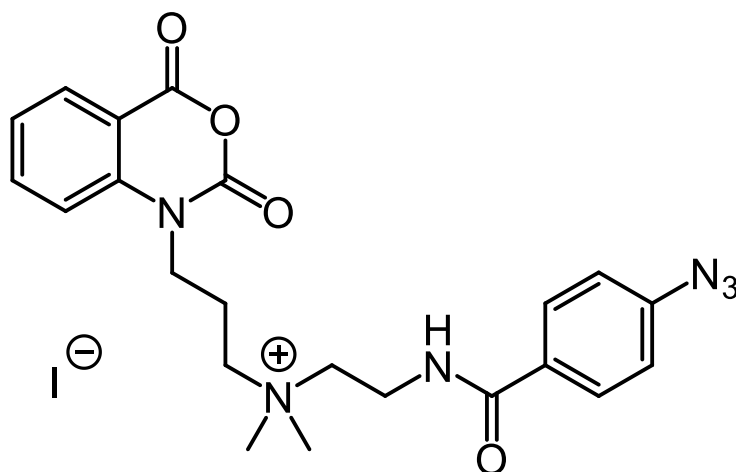


**3c.** General procedure 1 was followed with **2c**<sup>2</sup> (35 mg, 0.3 mmol) as the nucleophile. The final product (87 mg, 65 %) was isolated as a light brown solid. C<sub>15</sub>H<sub>20</sub>N<sub>5</sub>O<sub>3</sub>I, <sup>1</sup>H-NMR (500 MHz, dimethyl sulfoxide-d<sub>6</sub>): δ 2.02 (m, 2H), 3.05 (s, 6H), 3.49 (m, 4H), 3.93 (t, 2H), 4.12 (t, 2H), 7.37 (t, 1H), 7.56 (d, 1H), 7.88 (t, 1H), 8.05 (d, 1H) ppm, <sup>13</sup>C NMR (500 MHz, dimethyl sulfoxide-d<sub>6</sub>): δ 20.5, 41.5, 44.2, 50.9, 51.2, 61.6, 111.9, 114.9, 124.0, 129.39, 137.5, 141.3, 147.9, 159.0. HRMS: M<sup>+</sup> calc. 318.1566, found 318.1561

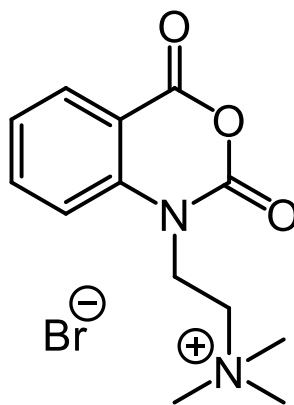


m/z.

**3d.** General procedure 1 was followed as above using **2d** (4-(azidomethyl)-N-(3-(dimethylamino)propyl)benzamide) (79 mg, 0.3 mmol) as the nucleophile. The final product (106 mg, 59%) was isolated as a light yellow solid. C<sub>24</sub>H<sub>29</sub>N<sub>6</sub>O<sub>4</sub>I (592.43), <sup>1</sup>H NMR (500 MHz, dimethyl sulfoxide-d<sub>6</sub>): δ 1.96 (m, 2H), 2.10 (m, 2H), 3.01 (s, 6H), 3.33 (m, 4H), 3.47 (m, 2H), 4.12 (t, 2H), 4.53 (s, 2H), 7.36 (m, 1H), 7.46 (d, 2H), 7.54 (d, 1H), 7.86 (m, 3H), 8.05 (d, 1H), 8.59 (t, 1H) ppm, <sup>13</sup>C NMR (500 MHz, dimethyl sulfoxide-d<sub>6</sub>): δ 20.3, 22.6, 36.3, 41.4, 50.0, 53.1, 60.4, 61.9, 111.8, 114.7, 123.8, 127.6, 128.2, 129.7, 134.0, 137.3, 138.8, 141.1, 147.8, 158.9, 166.1 ppm. HRMS: M<sup>+</sup> calc. 465.2251, found 465.2251 m/z.

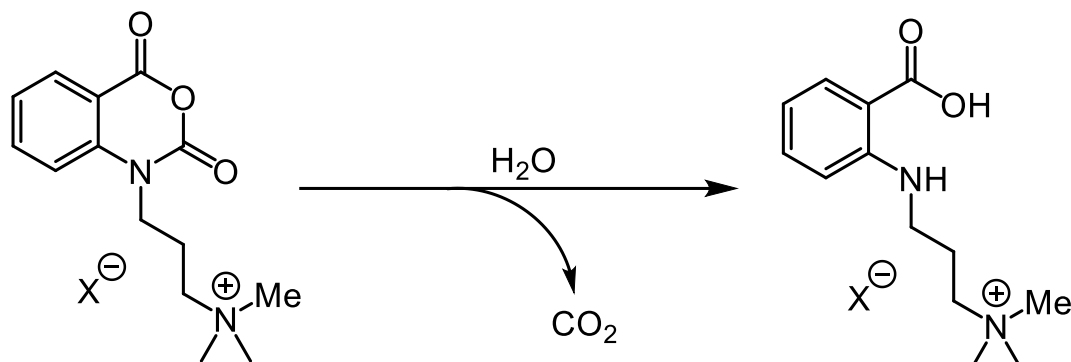


**3e.** General procedure 1 was followed as above using **2e** (4-azido-N-(2-(dimethylamino)ethyl)benzamide) (70 mg, 0.3 mmol) as the nucleophile. The final product (116 mg, 66%) was isolated as a light brown solid.  $C_{22}H_{25}N_6O_4I$ ,  $^1H$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  2.16 (m, 2H), 3.10 (s, 6H), 3.45 (t, 2H), 3.58 (m, 2H), 3.66 (q, 2H), 4.13 (t, 2H), 7.23 (d, 2H), 7.37 (m, 1H), 7.57 (d, 1H), 7.87 (m, 3H), 8.05 (d, 1H), 8.78 (t, 1H) ppm,  $^{13}C$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  20.4, 33.4, 41.4, 50.5, 60.8, 61.4, 111.8, 114.7, 119.1, 123.8, 129.1, 129.7, 130.0, 137.3, 141.1, 142.7, 147.8, 158.9, 165.7 ppm. HRMS:  $M^+$  calc. 437.1938, found 437.1938  $m/z$ .



**4.** To a 3 ml screw capped vial containing DMIA (100 mg, 0.4 mmol) was added 1 ml of dry reagent grade acetone. The solution heated to completely dissolve the reagent then filtered through a 0.2  $\mu m$  syringe filter. To this solution was added excess methyl bromide in THF ( $-78^\circ C$ ) and the solution was allowed to sit for 40 minutes at  $25^\circ C$ . The resulting precipitate was washed with acetone 3x and dried under reduced pressure. Product was isolated as a pure white solid.  $C_{13}H_{17}N_2O_3Br$  (329.19),  $^1H$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  3.23 (s, 9H), 3.62 (t, 2H,  $J = 7.8$  Hz), 4.49 (t, 2H,  $J = 7.8$  Hz), 7.40 (m, 1H), 7.53 (d, 1H,  $J = 8.5$  Hz), 7.91 (m, 1H), 8.06 (dd, 1H,  $J = 7.8, 1.6$  Hz) ppm,  $^{13}C$  NMR (500 MHz, dimethyl sulfoxide- $d_6$ ):  $\delta$  38.0, 52.7, 60.4, 112.1, 114.4, 124.0, 129.8, 137.3, 140.8, 147.6, 158.6 ppm, HRMS:  $M^+$  calc. 249.1239, found 249.1233  $m/z$ .

### Hydrolysis of quaternary derivatives.



Scheme S1: The hydrolysis of the anhydride (**3a**) resulting in the liberation of CO<sub>2</sub>. This ring-opening hydrolysis leads to a noticeable upfield shift in the <sup>1</sup>H NMR spectrum.

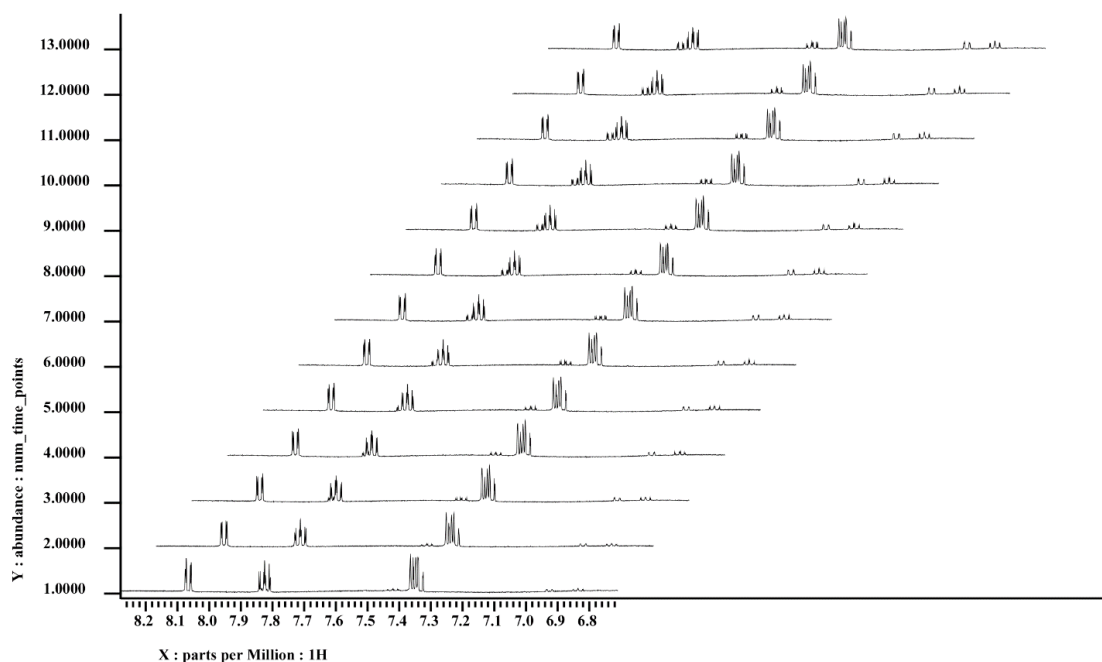


Figure I-S1: A kinetic plot of <sup>1</sup>H NMR spectra vs time of **3a** in D<sub>2</sub>O at 25 °C, zoomed on the aromatic region to show the slow disappearance of the peaks at 8.08, 7.83 and 7.35 corresponding to ring closed anhydride and appearance of peaks at 7.88, 7.44, 6.95 and 6.85 corresponding to the hydrolysis product. Scan 1 started 15 minutes after sample preparation, each scan is 1 hour apart, integration and extrapolation of the data resulted in a t<sub>1/2</sub> of 50 h.



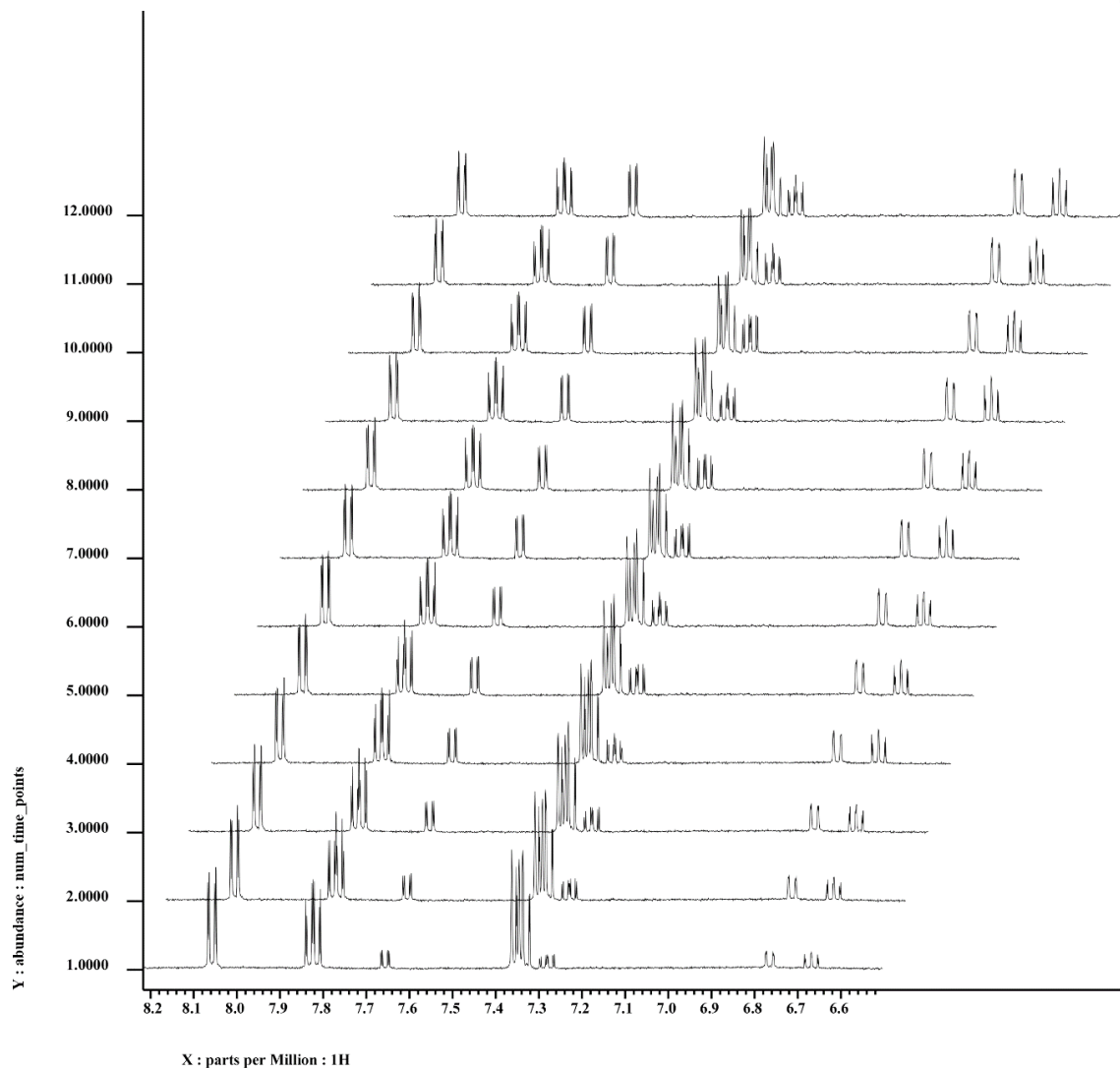


Figure I-S2: Kinetic  $^1\text{H}$ -NMR of **3a** in 25 mM bicarbonate pH 8.4  $\text{D}_2\text{O}$ , at 25  $^\circ\text{C}$  zoomed in on the aromatic hydrogens to show disappearance of peak at 8.08 corresponding to ring closed anhydride and appearance of the corresponding peak at 7.66 of the hydrolysis product. Scan 1 started 15 minutes after sample preparation, each scan is 15 minutes apart, integration of the data resulted in a  $t_{1/2}$  of 237 minutes.

Table S1: Half-life data for bicarbonate buffered  $\text{D}_2\text{O}$  at 25  $^\circ\text{C}$ .

Buffer Concentration (mM)	Half-life (minutes)	Half-life (hours)
0	3000	50
10	484	8.07
25	237	3.95
50	90.4	1.51
100	42.9	0.715

### Degree of labeling (DOL).

DOL was determined using Equation 1 below:

$$DOL = \frac{A_{330}\epsilon_{prot}}{(A_{280} - A_{330}C_{280})\epsilon_{reagent}}$$

Equation S1: Degree of labeling formula

Table S2: Degree of labeling variables

Abbreviation	Definition
$A_{330}$	Absorbance of the sample at 330 nm
$\epsilon_{prot}$	Molar extinction coefficient (in $M^{-1} cm^{-1}$ ) of pure protein at 280 nm
$A_{280}$	Absorbance of the sample at 280 nm
$C_{280}$	Correction factor of reagents contribution to absorbance at 280 nm
$\epsilon_{reagent}$	Molar extinction coefficient (in $M^{-1} cm^{-1}$ ) of the reagent at 330 nm

Table S3: Example data for bioconjugate for determination of DOL

Abbreviation	Value
$A_{330}$	0.1707
$\epsilon_{prot}$	43,824 $M^{-1} cm^{-1}$
$A_{280}$	0.7215
$C_{280}$	0.3184
$\epsilon_{reagent}$	3750 $M^{-1} cm^{-1}$

$$DOL = \frac{0.1707 * 43824 M^{-1}cm^{-1}}{(0.7215 - 0.1707 * 0.3184) * 3750 M^{-1}cm^{-1}} = 3.0$$

Equation S2: Example calculation for bioconjugate DOL

Table S4: Normalization Data for 50mM SHAPE analysis with 15-minute labeling, NAI is delivered in DMSO (S/B ratio of 1.92) whereas the isatoic anhydride reagents were delivered in water (S/B ratio of 1.28). Showing enhanced signal to solvent background of isatoic anhydride reagents due to reduced signal of solvent lane comparing water and DMSO.

Item	Normalized data	Normalizing lane
NAI	1.8	DMSO
4	2.94	Water
3a	2.89	Water
3c	2.73	Water

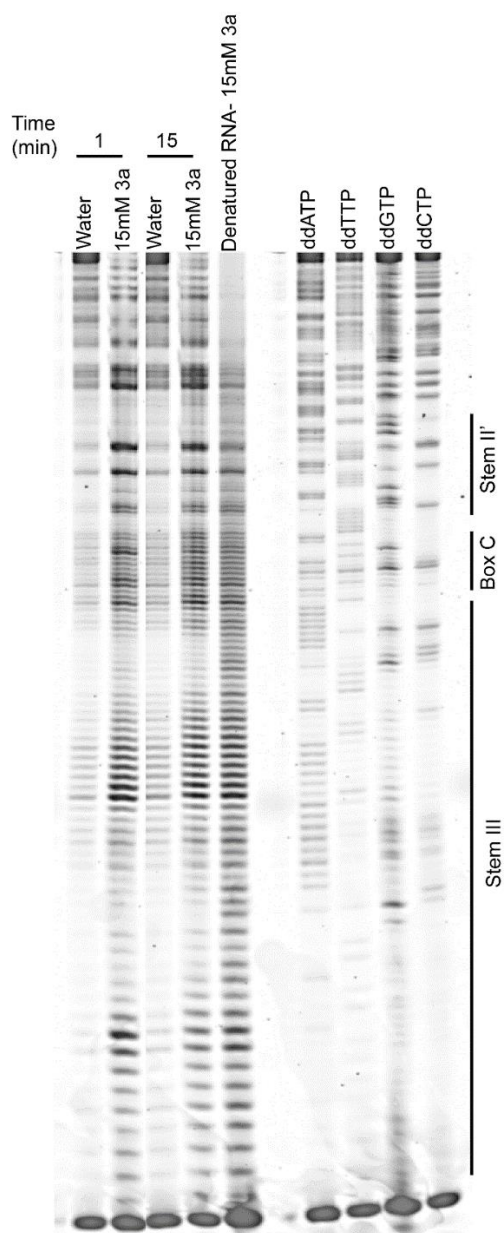


Figure I-S4: SHAPE- RT optimization of the water-soluble SHAPE electrophile **3a**. Cy5 labeled denaturing gel comparing modification of synthetic pfU3 snoRNA using 15mM of SHAPE electrophile **3a** at 1 min and 15 min respectively. Modification of the RNA at 1 min shows quick reactivity of the compound **3a**.

## CHAPTER 3: INNATELY WATER SOLUBLE ISATOIC ANHYDRIDES WITH MODULATED REACTIVITIES FOR RNA SHAPE ANALYSIS

Adapted with permission from *Bioconjugate Chem.* 2020, 31, 3, 884–888. Copyright (2020) American Chemical Society.

### 3.1 ABSTRACT:

1-methyl-7-nitroisatoic anhydride (1M7) and 2-methylnicotinic acid imidazolidine (NAI) are two of the most commonly applied RNA-SHAPE electrophiles; 1M7 due to its high reactivity and NAI for its solubility and cell permeability. While the addition of a nitro group yields desirable activation of the reagent, it also leads to poorer water solubility. This limited solubility has motivated the development of water-soluble reagents. We present alternative, isatoic anhydride-based reagents possessing variable reactivities that are simultaneously water soluble. Solubility is gained by using a quaternary ammonium while modulation of the reactivity is obtained by functionalization of the aryl ring. The syntheses of the reagents are discussed, and the electrophiles are demonstrated to be suitable for use for an *in vitro* RNA SHAPE experiment when directly compared to 1M7.

### 3.2 BODY:

The field of RNA selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) and additional mutational profiling (SHAPE-MAP) have provided novel insight into the structural dynamics of RNA both *in vitro* and *in vivo*.<sup>1-6</sup> This has been done through the use of numerous acylating reagents, where all use an activated carbonyl for the reaction with the 2'-hydroxyl of the ribose sugars on flexible RNA nucleotides.<sup>2, 7</sup> The current SHAPE electrophiles can be classified as water-soluble or reactive, with no single reagent

simultaneously containing both attractive features in appreciable amounts.<sup>7</sup> Despite the differences in reactivities, all SHAPE electrophiles are cross reactive with common intercellular nucleophiles including; alcohols, thiols, amines, and water.

The highly reactive 1M7 has played an important role in the development and maturation of both RNA SHAPE and SHAPE-MaP experiments (Figure 3-1).<sup>6, 8, 9</sup> This reagent possesses a strongly electron withdrawing nitro group para to the electrophilic carbonyl of the anhydride ring, Figure 1. The electron withdrawing effects of the nitro functionality on benzoic acid have been previously quantified using the Hammett equation.<sup>8, 10</sup> Hammett constants ( $\sigma$ ) for nitro substitution in the para ( $\sigma_p$ ) and meta ( $\sigma_m$ ) positions of benzoic acid are  $\sigma_p = 0.78$  and  $\sigma_m = 0.71$  respectively.<sup>10</sup> In the case of 1M7 this greatly increases electrophilicity thereby accelerating the reaction of 1M7 with nucleophiles including the 2'-hydroxyl of RNA. Accordingly, the reported half-life of 1M7 is 14 s in RNA SHAPE buffer (pH 8.0, 37 °C) used for the probing *in vitro* RNA.<sup>4</sup> The increased reactivity of 1M7 has been deemed advantageous for SHAPE experiments due to the “self-quenching” hydrolysis of the reagent in water which eliminates the need for disruptive quenching reagents.

2-methylnicotinic acid imidazolide (NAI) has been valuable for *in vivo* probing, attributed to its attenuated reactivity ( $t_{1/2} = 33$  min) and water solubility.<sup>4</sup> One downside to the attenuated reactivity of NAI is the requirement for an increase in either assay time or reagent concentration for robust RNA modification.<sup>4, 7</sup> In spite of this, NAI is frequently applied to SHAPE assays due to its advantageous water solubility which has been shown to improve probing output.

We have previously reported on a new class of isatoic anhydride-based reagents which incorporate a quaternary ammonium salt for increased water solubility.<sup>11, 12</sup> These reagents

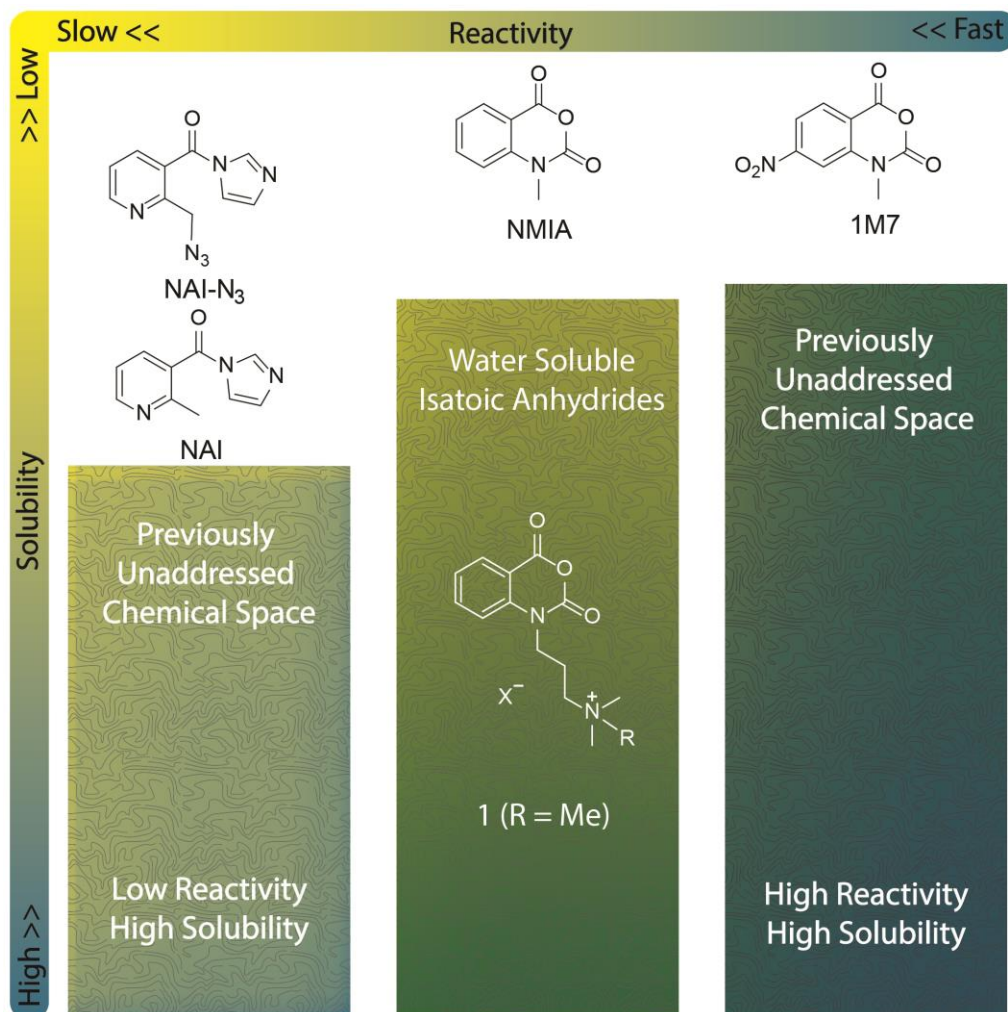


Figure 3-1: Graphical representation of the current coverage of chemical space for RNA SHAPE reagents

are highly adaptable and can be readily modified to include a number of chemical handles. Several reagents from this platform have been used as SHAPE reagents to probe the secondary structure of *Plasmodium falciparum* U3 small nucleolar RNA (U3 snoRNA).<sup>12</sup> These reagents demonstrated comparable reactivity to N-methylisatoic anhydride (NMIA) based on hydrolysis half-life, which is more reactive than NAI. The increased reactivity

and low relative solubility of NMIA has been previously suggested to be the cause of poor SHAPE outcomes.<sup>4, 13</sup> Our previous work detailing the use of our platform in the SHAPE analysis of U3 snoRNA clearly showed an improved signal to background noise compared to NAI in an *in vitro* RNA SHAPE experiment, suggesting the problem with NMIA may be the poor water solubility rather than its reactivity. The water solubility of **1** (Figure 3-1) could be tuned from 25-130 mM based on counter anion ( $I^- < Br^- < Cl^-$ ).

Recently, interest in reagents for RNA SHAPE which are both highly reactive and water soluble was highlighted,<sup>7</sup> and the importance of water solubility and chemical probe design was reiterated.<sup>14</sup> We sought to address this gap in the chemical space (Figure 1) using our previously published strategies.<sup>11, 12</sup> Additionally, we aimed to expand our efforts to include the development of isatoic anhydride reagents exhibiting attenuated reactivity, more similar to that of NAI. Initial efforts were aimed at directly preparing a water-soluble analogue of 1M7 (**4a**, Figure 3-2) and a second, slower reagent (**4b**, Figure 3-2) using a sterically hindering methyl group much like that of NAI. Individual steps in preparation of **4a** proved difficult resulting in substantial amounts of anhydride hydrolysis, resulting in low yields and difficult isolation. The water solubility of **4a** was disappointing, found to be < 1 mg/mL (< 2.6 mM) as the bromide salt. This experimentally determined solubility is below that of the calculated solubility of unmodified 1M7,  $\log S = -2.4$  (4 mM) bringing into question the usefulness and accuracy of relying on calculated water solubilities. The poor water solubility of **4a** limited interest in this derivative as previously prepared isatoic anhydride reagents possess solubility roughly 20X greater as the bromide salts. Compound **4b** was prepared and isolated in modest yield and found to have a water solubility of roughly 20 mg/mL (56 mM). The half-life of **4b** in 25 mM bicarbonate (pH 8.4, 25 °C) in

D<sub>2</sub>O, was determined using time-resolved <sup>1</sup>H-NMR to monitor hydrolysis over 24 h by monitoring the changes in the resonances from 7.8 - 6.9 ppm. The half-life was found to be 16 h under these conditions, about four times as long as **1**. Compound **4b** was found to have a half-life of 16 minutes in deuterated 1X SHAPE buffer (pH 8.0, 37 °C) using the same experimental setup. Having reasonable solubility and reactivity, **4b** was therefore chosen for use in additional experimentation. The poor solubility of **4a** raised interest in determining the true solubility of 1M7, to provide insight if an alternative to using the nitro group for activation of the electrophile should be sought. To test the accuracy of the logS solubility calculations we chose to prepare and isolate 1M7 as previously described.<sup>15</sup> The determined water solubility of 1M7 was found to be ≈0.1 mM at pH 7.0, roughly 40 times lower than the 4 mM calculated logS solubility which has been reported.<sup>2</sup> This experiment is detailed in the supporting information with spectra provided, Figure II-S1.

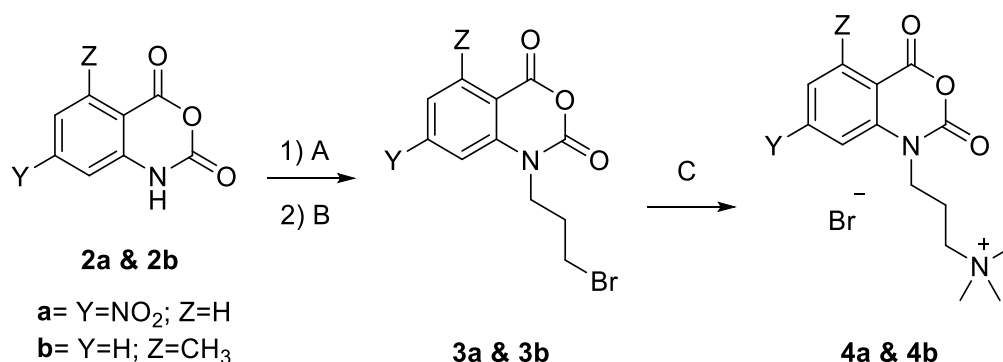


Figure 3-2: Synthesis of **4a** & **4b** A) NaH, DMF B) Br(CH<sub>2</sub>)<sub>3</sub>Cl C) Me<sub>3</sub>N, DMF

Due to the unexpectedly poor characteristics of compound **4a** and insolubility stemming from the nitro-functionalized core,<sup>16</sup> we sought an alternative route for the preparation of a water-soluble, and highly reactive isatoic anhydride derivative. Investigating the Hammett values, we noted the value for the Me<sub>3</sub>N<sup>+</sup>- group meta to the electrophilic



carbonyl has a sigma value of  $\sigma_m = 0.88$ .<sup>10</sup> This is more withdrawing than that of a nitro group in the para position  $\sigma_p = 0.78$ , as in 1M7. The use of a  $\text{Me}_3\text{N}^+$  appended to the aryl ring at the 6 position therefore should have a dual benefit of activating the carbonyl by withdrawing electron density from the aromatic ring while concomitantly solubilizing the reagent.<sup>8, 10, 17</sup> The synthesis of this new compound (**6**), Figure 3-3, is a simple procedure starting from the commercially available 5-fluoro-2-nitrobenzoic acid and following previously published procedure to prepare the 5-dimethylaminoisatoic anhydride.<sup>18</sup> We chose this starting material to utilize nucleophilic aromatic substitution for addition of a dimethyl amino group on the ring. After ring closure with triphosgene, the dimethylamino group ( $\sigma_m = -0.211$ ) deactivates the anhydride due to electron donation.<sup>10</sup> This feature minimizes the hydrolysis of the anhydride during additional synthetic steps. Compound **5** is readily isolated from side products and starting materials by extraction under basic conditions. Upon quaternization of **5** in acetone with methyl bromide, compound **6** with a bromide counter ion is isolated as a light brown precipitate in modest yields.

Upon preparation of **6**, we experimentally determined aqueous solubility to be similar to other previously prepared reagents as the bromide salt, roughly 20 mg/mL ( $\approx 70$  mM). This allows for preparation of DMSO free stock solutions at concentrations suitable for use in traditional SHAPE experiments as demonstrated with previous reagents. We chose to pursue **6** as a possible alternative for 1M7. The hydrolysis half-life of **6** was determined as previously described for compound **4b**. The half-life of **6** was determined to be 55 minutes in 25 mM bicarbonate (8.4, 25 °C) in  $\text{D}_2\text{O}$ . This is considerably shorter than what was previously demonstrated for reagent **1** which demonstrated a half-life of roughly 4 hours under identical conditions.<sup>12</sup> The half-life in deuterated SHAPE buffer was not obtained

for **6** due to the rapid nature of the hydrolysis under these conditions, exceeding the experimental capability when using our standard NMR techniques. Despite the reactive functionality of **6** and hygroscopic nature of the quaternary ammonium salts, the reagents were able to be dissolved in water and lyophilized with little to no degradation. Lyophilized samples stored under an inert atmosphere are stable at 4 °C for longer term storage.

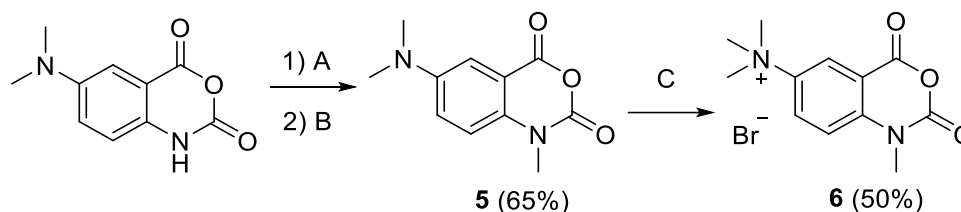


Figure 3-3: Synthesis of **6**, A) NaH, DMF B) MeI (1 eq) C) MeBr, acetone

To demonstrate the ability of compounds **6** and **4b** as an alternative to conventional SHAPE reagents, we probed the secondary structure of *in vitro* transcribed *Plasmodium falciparum* U3 small nucleolar RNA (snoRNA) (Figure II-S3). For these RNA-SHAPE experiments, 1M7 was used as a probing control. The reagents, **6** and **4b** provided similar structural output for PfU3snoRNA (Fig II-S2) as previously reported using our water soluble electrophile **1**.<sup>12</sup> Using **4b**, **6** and 1M7 we further probed the secondary structure of *in vitro* transcribed *Trypanosoma brucei* Telomerase RNA (TbTR) catalytic core. Within eukaryotic cells, telomerase RNA (TR) associates with Telomerase Reverse Transcriptase (TERT) protein to form an active Telomerase ribonucleoprotein (RNP) complex.<sup>19</sup> Telomerase replenish the linear ends of eukaryotic chromosomes called telomeres which are essential to avoid cell death and apoptosis, resulting in unlimited cellular proliferation.<sup>20</sup> For maintaining telomeres, TERT uses the template region of TR and adds complementary nucleotides to telomeres *de novo*.<sup>21</sup> We and others have previously

identified distinct features of TR in *Trypanosoma brucei*, (TbTR)<sup>22-25</sup> which are essential for telomerase holoenzyme assembly and catalytic activity. TbTR in *Trypanosoma brucei*, a group of parasitic protozoa, is ~900 nt long and contains a ‘template’ domain flanked by template boundary element (TBE) at 5′ and template distal helix at 3′ end.<sup>22, 24</sup> The template region of TbTR contains the following sequence: CCCUAACCC which is used by TERT to add TTAGGG repeats to the chromosome (telomere) end.<sup>22, 23</sup> The template boundary element is in the form of stem loop helix and act as a boundary, preventing TERT from adding nucleotides beyond the template region which would otherwise have deleterious consequences.<sup>24</sup> The template distal helix can involve in long range tertiary interactions,<sup>25</sup> thus forming the triple helix structure termed a ‘pseudoknot’, a common feature in all known TR.<sup>24</sup> Together, the template, template boundary element and pseudoknot form the catalytic core which is minimally required by telomerase for its catalytic activity.<sup>24</sup>

Probing of the synthetic *Trypanosoma brucei* telomerase RNA (TbTR) catalytic core was performed with DMSO soluble 1M7 and water soluble **4b** and **6**. These each revealed distinct modifications of unpaired nucleotides compared to controls in the denaturing gel. Our RNA-SHAPE modifications are consistent with previously published SHAPE assay data for TbTR.<sup>25</sup> Furthermore, a ~2-fold enhancement in the signal to noise ratio was observed for compound **6** as compared to 1M7 (See Appendix II for information), suggesting the higher reactivity of **6** in modifying RNA nucleotides as compared to 1M7 (Figure 3-4B and 3-4C). Background signal was substantially lower for both **4b** and **6** compared to the water and DMSO controls alone, suggesting that the water solubility could make it straightforward to distinguish the signal originating from the SHAPE modifications from that of the background noise of DMSO or water alone. Interestingly, in spite of the

expected attenuated reactivity, **4b** provided robust labeling, similar to 1M7, but weaker when compared to **6**. These results highlight the importance of both water solubility and reagent reactivity for SHAPE probing.

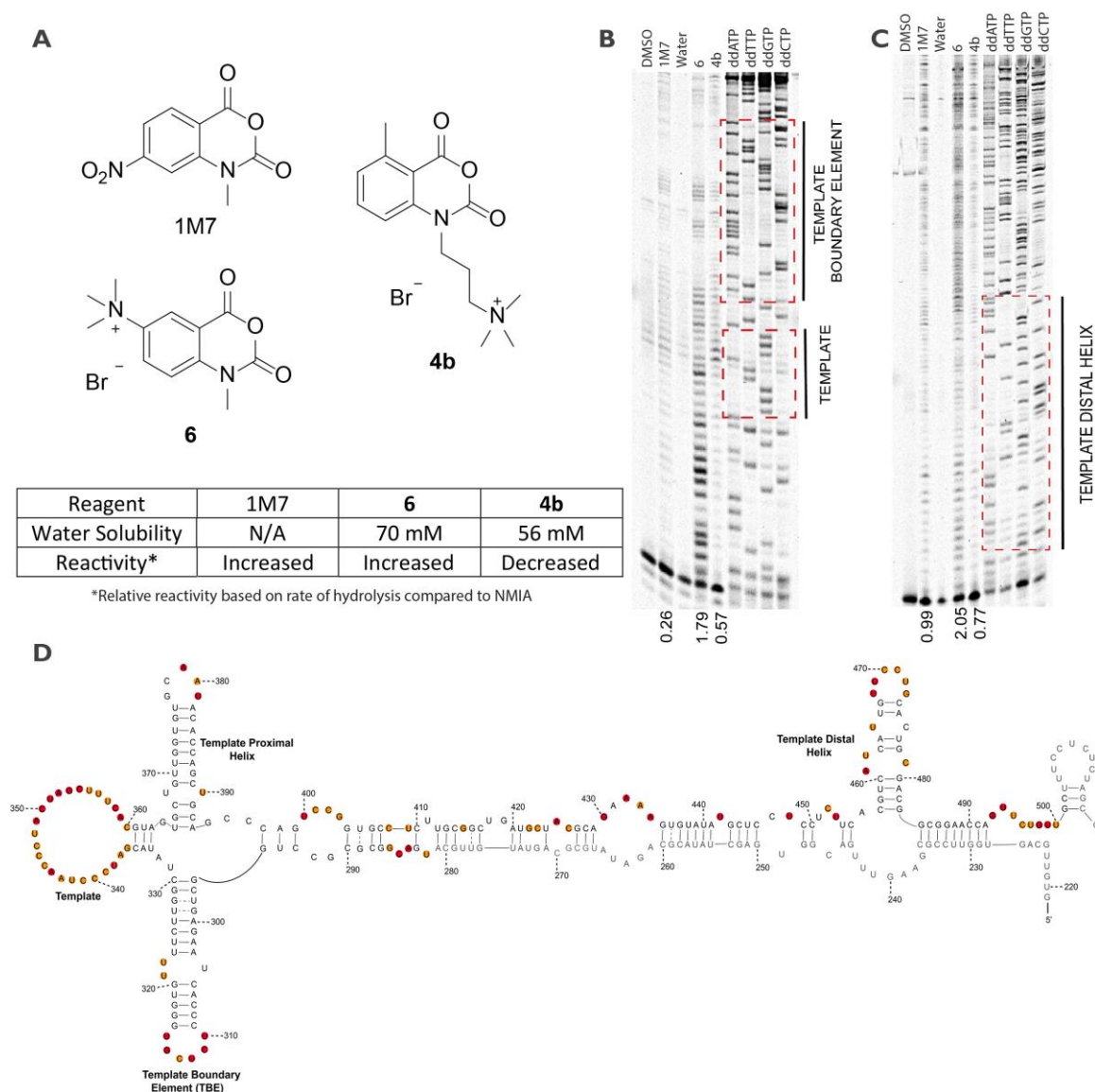


Figure 3-4: *In vitro* SHAPE-RT modification of synthetic TbTR catalytic core, (A) Structure and reagent comparison. Cy5 labeled denaturing gel indicates modification in the unpaired nucleotide of (B) Template and template boundary element (C) Template distal helix. (D) Secondary structure of the TbTR catalytic core based on *in vitro* SHAPE modification of RNA using compound **6**. Unpaired RNA nucleotides are shown in red and orange color based on their reactivities, where high reactivity is signified by red color and low reactivity by orange with the SHAPE reagent. Nucleotides denoted by grey color were not modified in this study.

Based on the modifications obtained from our denaturing gel experiments, we derived the secondary structure of TbTR catalytic core (Figure 3-4D) using RNAstructure.<sup>26</sup> Nucleotides comprising the template domain and those immediately present at 3' end of template (nt336- nt360) were highly modified and hence indicates that these nucleotides are single-stranded or not involved in any base pairing interactions within the TbTR catalytic core model *in vitro*. The template boundary element is in the form of a stem loop helix with the base being stabilized by base pairing of 330C-296G and culminating in hexa-nucleotide apical loop (nt310-nt315) with the latter showing intense nucleotide modifications indicative of their single stranded nature. Additionally, TBE also contains an internal UU (nt321-nt322) bulge with moderate modifications as apparent in the denaturing gel (Figure 3-4B). Instead of a RNA triple helix found at the 3' of TR template region of other eukaryotes, our TbTR model depicts two stem-loop helices named herewith as a template proximal and template distal helix. Like TBE, the template proximal helix also contains an UU internal bulge and a penta-nucleotide loop at its apical end (nt377-nt381) with partial modification in nucleotides. The template distal helix, however, has a two nucleotide internal bulge and ends in a hexa-nucleotide (nt467-nt473) apical loop with high reactivity towards SHAPE reagents. Our TbTR secondary structure also identified two canonical tetraloops of UNCG (nt400-nt403) and GNRA (nt430-nt433) types positioned in between the template proximal and distal helix. Although, nucleotides comprising these tetraloops were found to be reactive towards SHAPE reagents, previous *in vivo* studies suggested that these tetraloops are essential for maintaining RNA structure either by extensive base-pairing interactions within the RNA or by interacting with proteins.<sup>27, 28</sup> Overall, the TbTR catalytic core secondary structure model obtained from 6 modifications

in this RNA-SHAPE experiment strongly aligns with the TbTR models previously reported.<sup>22, 25</sup>

These experiments highlight new water soluble isatoic anhydride SHAPE reagents; Compound **6** possessing high reactivity and solubility and **4b** with attenuated reactivity and high solubility. These reagents stock solutions can be directly prepared in water due to their innate solubility. This feature seems to lower the background signal as compared to the DMSO control which shows a considerable degree of signal overlap with the structural probing. Derivative **6** utilizes a water-solubilizing quaternary ammonium appended to the aryl ring which simultaneously imparts its highly reactive character. The new reagent, **6**, with its attractive synthetic features is demonstrated as a direct alternative for 1M7 in use an *in vitro* RNA SHAPE analysis. Optimization of labeling concentrations to further improve results are underway. The future testing of the new quaternary ammonium reagents for applications in SHAPE-MaP experiments will prove to be an important step to determine their capacity to serve as new SHAPE electrophiles.

### 3.3 REFERENCES

1. Watters, K. E., Yu, A. M., Strobel, E. J., Settle, A. H., and Lucks, J. B. (2016) Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq), *Methods* 103, 34-48.
2. Lee, B., Flynn, R., Kadina, A., Guo, J., Kool, E., and Chang, H. (2017) Comparison of SHAPE reagents for mapping RNA structures inside living cells, *RNA* 23, 169-174.
3. Smola, M. J., and Weeks, K. M. (2018) In-cell RNA structure probing with SHAPE-MaP, *Nature Protoc.* 6, 1181-1195.
4. Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013) RNA SHAPE analysis in living cells, *Nature Chem. Biol.* 9, 18-20.
5. Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution, *Nature Protoc.* 1, 1610.
6. Martin, S., Blankenship, C., Rausch, J. W., and Sztuba-Solinska, J. (2019) Using SHAPE-MaP to probe small molecule-RNA interactions, *Methods* 167, 105-116.

7. Busan, S., Weidmann, C. A., Sengupta, A., and Weeks, K. M. (2019) Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies, *Biochemistry US* 58, 2655-2664.
8. Mortimer, S. A., and Weeks, K. M. (2007) A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry, *J. Am. Chem. Soc.* 129, 4144-4145.
9. Incarnato, D., Morandi, E., Simon, L. M., and Oliviero, S. (2018) RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications, *Nucleic Acids Res.* 46, e97-e97.
10. Charton, M. (1981) Electrical Effect Substituent Constants for Correlation Analysis, *Prog. Phys. Org. Chem.*, pp 119-251.
11. Fessler, A., Garmon, C., Heavey, T., Fowler, A., and Ogle, C. (2017) Water-soluble and UV traceable isatoic anhydride-based reagents for bioconjugation, *Org. Biomol. Chem.* 15, 9599-9602.
12. Fessler, A. B., Dey, A., Garmon, C. B., Finis, D. S., Saleh, N.-A., Fowler, A. J., Jones, D. S., Chakrabarti, K., and Ogle, C. A. (2018) Water-Soluble Isatoic Anhydrides: A Platform for RNA-SHAPE Analysis and Protein Bioconjugation, *Bioconjugate Chem.* 29, 3196-3202.
13. Habibian, M., Velema, W. A., Kietrys, A. M., Onishi, Y., and Kool, E. T. (2019) Polyacetate and Polycarbonate RNA: Acylating Reagents and Properties, *Organic Lett.* 21 (14), 5413-5416.
14. Velema, W. A., and Kool, E. T. (2020) The chemistry and applications of RNA 2'-OH acylation, *Nature Rev. Chem.* 4 (1), 22-37.
15. D'Souza, A. M., Rivinoja, D. J., Mulder, R. J., White, J. M., Meyer, A. G., Hyland, C. J. T., and Ryan, J. H. (2018) Competitive 1,3-Dipolar Cycloaddition Reactions of an Azomethine Ylide with Aromatic and Carbonyl Groups of Nitro-Substituted Isatoic Anhydrides, *Aust. J. Chem.* 71, 690-696.
16. Sagawa, N., and Shikata, T. (2014) Are all polar molecules hydrophilic? Hydration numbers of nitro compounds and nitriles in aqueous solution, *Phys. Chem. Chem. Phys.* 16, 13262-13270.
17. Hansch, C., Leo, A., and Taft, R. W. (1991) A survey of Hammett substituent constants and resonance and field parameters, *Chem. Rev.* 91, 165-195.
18. Aranyi, P., Balazs, L., Balogh, M., Bata, I., Batori, S., Nagy-Lajos, T., Timari, G., Boer, K., Finance, O., Kapui, Z., et al., (2005) Aminoquinoline Derivatives and their use as Adenosine A3 Ligands, WO, P., Ed.
19. Greider, C. W., and Blackburn, E. H. (1989) A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis, *Nature* 337, 331-337.
20. Blackburn, E. H. (1991) Structure and function of telomeres, *Nature* 350, 569-573.
21. Greider, C. W., and Blackburn, E. H. (1985) Identification of a specific telomere terminal transferase activity in *tetrahymena* extracts, *Cell* 43, 405-413.
22. Sandhu, R., Sanford, S., Basu, S., Park, M., Pandya, U. M., Li, B., and Chakrabarti, K. (2013) A trans-spliced telomerase RNA dictates telomere synthesis in *Trypanosoma brucei*, *Cell Res.* 23, 537-551.

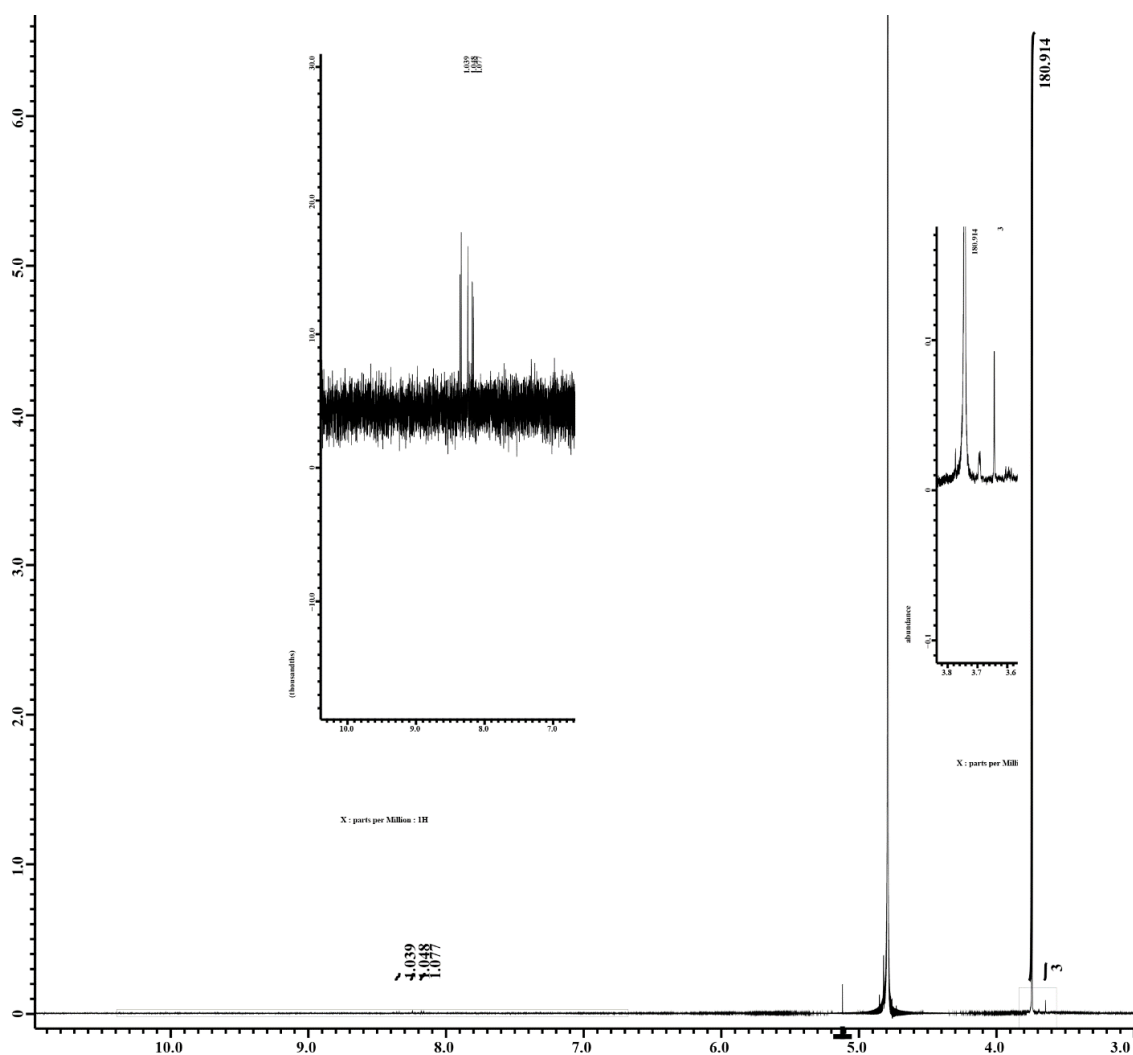
23. Gupta, S. K., Kolet, L., Doniger, T., Biswas, V. K., Unger, R., Tzfati, Y., and Michaeli, S. (2013) The *Trypanosoma brucei* telomerase RNA (TER) homologue binds core proteins of the C/D snoRNA family, *FEBS Lett.* 587, 1399-1404.
24. Dey, A., and Chakrabarti, K. (2018) Current Perspectives of Telomerase Structure and Function in Eukaryotes with Emerging Views on Telomerase in Human Parasites, *Int. J. Mol. Sci.* 19, 333.
25. Podlevsky, J. D., Li, Y., and Chen, J. J. L. (2016) The functional requirement of two structural domains within telomerase RNA emerged early in eukaryotes, *Nucleic Acids Res.* 44, 9891-9901.
26. Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis, *BMC Bioinformatics* 11, 129.
27. Molinaro, M., and Tinoco, I., Jr. (1995) Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications, *Nucleic Acids Res.* 23, 3056-3063.
28. Jaeger, L., Michel, F., and Westhof, E. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions, *J. Mol. Biol.* 236, 1271-1276.



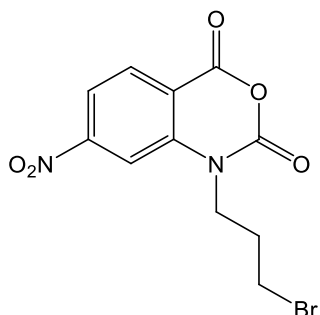
## APPENDIX II: SUPPLEMENTARY INFORMATION

### Determination of 1M7 water solubility.

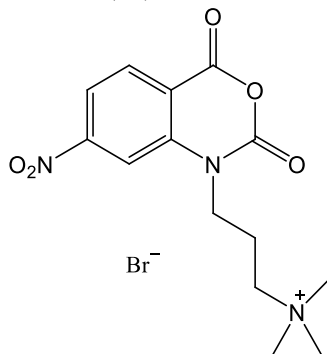
A 2 mM standard solution of 1,4-dioxane in D<sub>2</sub>O was prepared by dilution of 0.17  $\mu$ L 1,4-dioxane to 1.0 mL final volume with D<sub>2</sub>O. This solution was equilibrated to 25 °C. 1 mL was added to an Eppendorf tube containing ~15 mg 1M7. The tube was vortexed vigorously for 30 s and the resulting suspension was filtered through a 0.2  $\mu$ M syringe filter into an NMR tube and run on a 500 MHz JOEL EC500 NMR for 16 scans with 20 s relaxation. The integration values of 1M7 was set by the N-methyl group of the isatoic anhydride at 3.77 ppm. 1,4-dioxane peak at 3.64 ppm (8H) integrated ~22.5 times high. The concentration of 1M7 was therefore determined to be  $\approx$  0.1 mM.



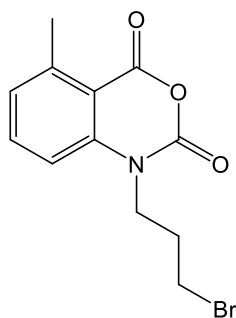
**Figure II-S1:** NMR determination of 1M7 water solubility 25° C in D<sub>2</sub>O, picture in picture images shown to demonstrate clean 1M7 and relative integrations of aromatic protons.



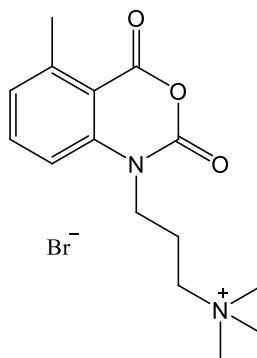
**3a.** A dry Schlenk flask equipped with a stir bar and flushed with N<sub>2</sub> was charged with NaH (0.06 g, 2.3 mmol) and to this, dry DMF was added (10 mL). 6-nitroisatoic anhydride (300 mg, 1.4 mmol) was added to the solution slowly and then subsequently rinsed down the glass with dry DMF. To the reaction mixture, 1-bromo-3-chloropropane (1.0 eq, 1.4 mmol) was added in one portion. The reaction stirred at RT overnight. The solution was then concentration under reduced pressure, the crude residue was dissolved with DCM and washed 6x with ice cold DI H<sub>2</sub>O. The organic layer was dried overnight with Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under reduced pressure to afford a bright orange powder. (160 mg, 34%) C<sub>11</sub>H<sub>9</sub>BrN<sub>2</sub>O<sub>5</sub> (329.11), <sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>): δ 8.30 (d, 1H), 8.21 (t, 1H), 8.09 (d, 1H), 4.23 (t, 2H), 3.83 (t, 2H), 2.1 (t, 2H) ppm. <sup>13</sup>C NMR (300 MHz, DMSO-d<sub>6</sub>): δ 157.9, 152.5, 147.3, 141.7, 131.3, 117.6, 117.1, 109.6, 41.5, 41.3, 26.5 ppm. MS (EI): Calc.: 327.97 (M); Found: 328.0 (M) m/z.



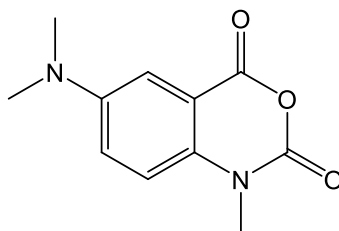
**4a.** To a 1.5 mL vial flushed with N<sub>2</sub> was added **3a** (20 mg, 61 μmol) in 250 μL dry DMF. To this was added 2.5 eq 2M trimethylamine in THF (54 μL, 0.152 mmol). The vial is sealed and placed at RT vial remained at RT overnight. Ethyl ether was added to triturate product from DMF. Vial is washed 5x with ethyl ether then the residual solvent is removed under reduced pressure. Product was isolated as bright yellow solid. (2 mg, 8%). C<sub>14</sub>H<sub>18</sub>BrN<sub>3</sub>O<sub>5</sub> (387.04), <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>): δ 8.30 (d, 1H), 8.16 (s, 1H), 8.09 (d, 1H), 4.24 (t, 2H), 3.40 (m, 2H), 3.06 (s, 9H) 2.14 (t, 2H) ppm. <sup>13</sup>C NMR (500 MHz, DMSO-d<sub>6</sub>): δ 157.9, 152.8, 147.5, 141.9, 131.6, 117.9, 116.9, 109.8, 62.6, 52.3, 41.8, 20.5 ppm. HRMS (ESI): Calc.: 308.1246 (M<sup>+</sup>); Found: 308.1226 (M<sup>+</sup>) m/z.



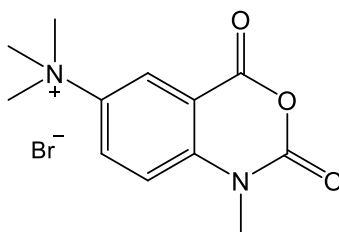
**3b.** A dry Schlenk flask equipped with a stir bar and flushed with N<sub>2</sub> was charged with NaH (0.06 g, 2.3 mmol) and to this, dry DMF was added (10 mL). 6-methylisatoic anhydride (300 mg, 1.7 mmol) was added to the solution slowly and then subsequently rinsed down the glass with dry DMF (10 mL). To the reaction mixture, 1-bromo-3-chloropropane (1.0 eq, 1.7 mmol) was added in one portion. The reaction stirred at RT overnight. The solution was then concentration under reduced pressure, the crude residue was dissolved with DCM and washed 6x with ice cold DI H<sub>2</sub>O. The organic layer was dried overnight with Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under reduced pressure, isolated as a light brown solid (1.1 g, 51%). C<sub>12</sub>H<sub>12</sub>BrNO<sub>3</sub> (298.14), <sup>1</sup>H NMR (500 MHz, DMSO): δ 7.69 (t, 1H), 7.35 (d, 1H), 7.18 (d, 1H), 4.06 (t, 2H), 3.37 (t, 2H), 2.64 (s, 3H), 2.15 (p, 1H). <sup>13</sup>C NMR (300 MHz, DMSO): δ 157.9, 141.8, 143.2, 142.4, 135.9, 126.4, 112.5, 110.4, 42.2, 30.5, 22.3, 3.73 ppm. MS (EI): Calc. 297.0; Found: 297.1 (M) m/z



**4b.** To a 1.5 mL vial flushed with N<sub>2</sub> was added **3b** (168 mg, 56 μmol) in 1000 μL dry acetone. To this was added roughly 2.5 eq 2M trimethylamine in THF (54 μL, 0.152 mmol). The vial is sealed and placed at RT vial remained at RT overnight. Vial is washed 5x with ethyl ether then the residual solvent is removed under reduced pressure Final product isolated as a white powder (119 mg, 59%). C<sub>15</sub>H<sub>21</sub>BrN<sub>2</sub>O<sub>3</sub> (356.07), <sup>1</sup>H NMR (500 MHz, DMSO): δ 7.72 (t, 1H), 7.42 (d, 1H), 7.22 (d, 1H), 4.11 (t, 2H), 3.49 (m, 2H), 3.04 (s, 9H), 2.66 (s, 3H), 2.09 (m, 9H) ppm. <sup>13</sup>C NMR (500 MHz, DMSO): δ 157.9, 147.9, 143.4, 142.2, 136.1, 126.6, 112.7, 110.31, 62.6, 52.2, 41.7, 22.3, 20.6, ppm. HRMS (ESI): Calc: 277.1552 (M<sup>+</sup>); Found: 277.1544 (M<sup>+</sup>) m/z



**5.** A dry Schlenk flask equipped with a stir bar and flushed with N<sub>2</sub> was charged with NaH (0.06 g, 2.3 mmol) and to this, dry DMF was added (10 mL). 5-dimethylaminoisatoic anhydride (300 mg, 1.5 mmol) was added to the solution slowly and then subsequently rinsed down the glass with dry DMF (2-3 mL). To the reaction mixture, CH<sub>3</sub>I (1.0 eq, 1.5 mmol) was added in one portion. The reaction stirred at RT overnight. The solution was then concentration under reduced pressure, the crude residue was dissolved with DCM and washed 6x with an ice cold solution of saturated sodium bicarbonate. The organic layer was dried overnight with Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated under reduced pressure to afford a faint green solid. (207 mg, 65%). C<sub>11</sub>H<sub>12</sub>N<sub>2</sub>O<sub>3</sub>, <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>): δ 7.31 (d, 2H), 7.14 (t, 1H), 3.42 (s, 3H), 2.94 (s, 6H) ppm. <sup>13</sup>C NMR (500 MHz, DMSO-d<sub>6</sub>): δ 159.9, 148.1, 147.3, 147.3, 133.3, 122.7, 116.4, 110.1, 40.9, 32.1 ppm. HRMS (EI): Calc. 221.0926; Found: 221.0917 (MH<sup>+</sup>) m/z.



**6.** To a 1.5 mL vial flushed with N<sub>2</sub> was added **5** (100 mg, 0.45 mmol) in 750 μL dry acetone through a 0.2 μm filter. To this was added excess MeBr condensed in 250 μL cold acetone. The vial is sealed and placed at RT vial remained at RT overnight. Vial is washed 5x with ethyl ether then the residual solvent is removed under reduced pressure. (65 mg, 50%). C<sub>12</sub>H<sub>15</sub>BrN<sub>2</sub>O<sub>3</sub> (315.17), <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>): δ 8.44 (m, 2H), 7.69 (s, 1H), 3.66 (s, 9H), 3.51 (s, 3H) ppm. <sup>13</sup>C NMR (500 MHz, DMSO-d<sub>6</sub>): δ 158.1, 147.4, 142.9, 141.9, 129.2, 121.3, 116.7, 112.2, 56.5, 32.1 ppm. HRCMS (ESI): Calc. 235.1083; Found: 235.1071 (M<sup>+</sup>) m/z

#### PCR amplification and *in vitro* Transcription:

Primary PCR amplification of *Plasmodium falciparum* U3 small nucleolar RNA (snoRNA) was done from genomic DNA of *Plasmodium falciparum* 3D7 strain as described earlier.<sup>1</sup> The resultant PCR amplified DNA of PfU3 snoRNA was *in vitro* Transcribed (IVT) and purified using the MEGAscript and MEGAclear kits (Life Technologies, USA) following the manufacturer's recommendation.

### ***in vitro* SHAPE modification:**

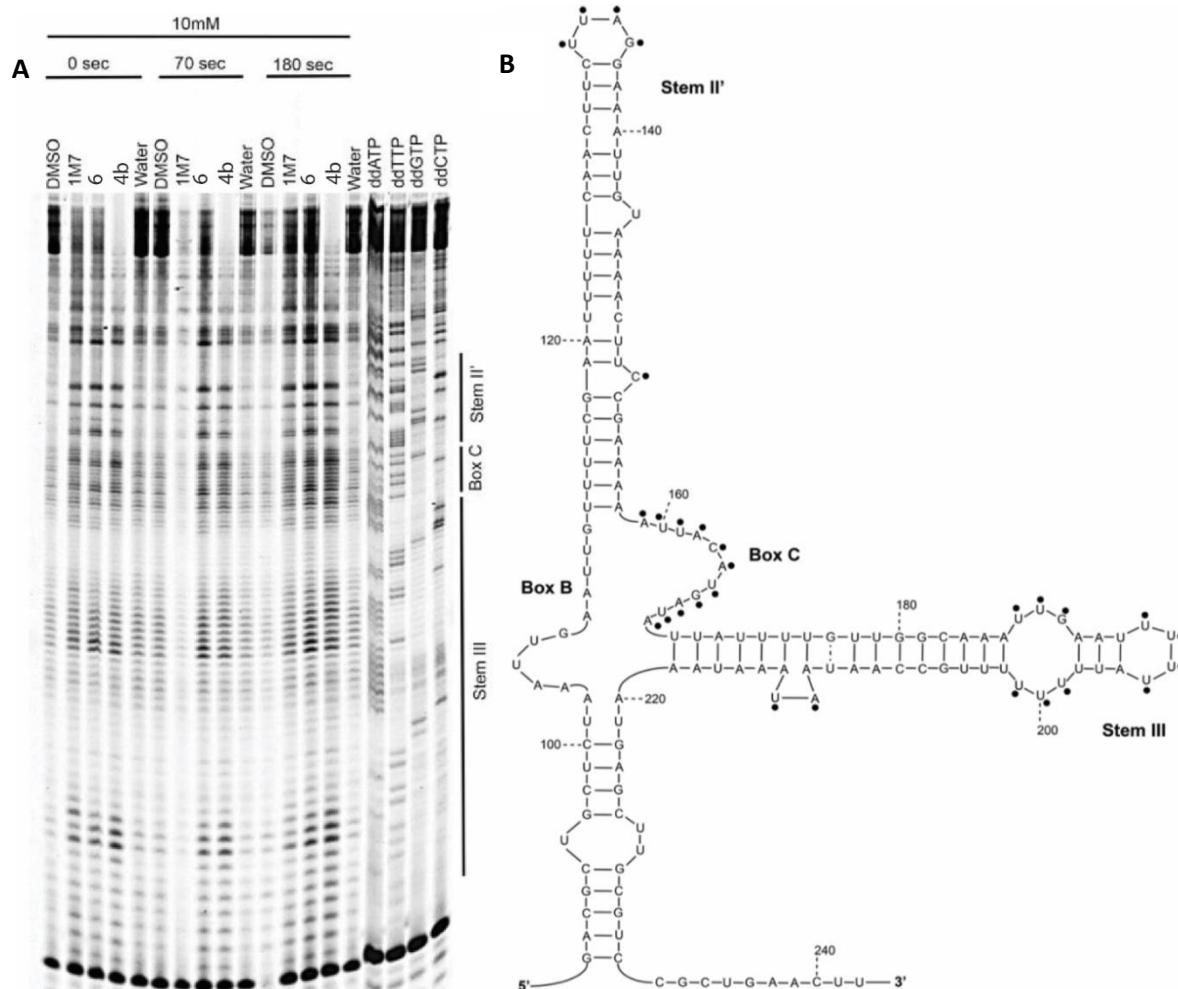
For SHAPE modification of PfU3snoRNA, 5 µg of *in vitro* transcribed RNA was denatured by heating at 95 °C for 2 min followed by immediate incubation on ice for 3 min. The modification was carried out by incubation of the RNA in 3X SHAPE buffer consisting of 333 mM of NaCl, 333 mM of HEPES and 20 mM of MgCl<sub>2</sub> at 37 °C prior to the addition of 10mM of 1M7, **4b** and **6** for 70 sec. Following incubation, the modified RNA was ethanol precipitated, dissolved in water and subjected to nanodrop analysis.

### **cDNA synthesis and RNA-SHAPE analysis by denaturing Polyacrylamide Gel Electrophoresis (PAGE):**

1.6 µg of SHAPE modified RNA was denatured by heating at 95 °C for 2 min followed by immediate incubation on ice for 3 min. For Primer extension, 5' cy5 labeled primer (Table S1) for reverse transcription along with 10 mM dNTPs were mixed with the SHAPE modified RNA and incubated at 65 °C for 5 min. Four dideoxy sequencing reactions using ddNTPs were also performed in parallel using the unmodified PfU3 snoRNA in the similar manner as mentioned above. Following incubation, all the reactions were supplemented with 1X SuperScript II First-strand buffer (50mM Tris-HCl, pH 8.3, 75 mM KCl and 5 mM MgCl<sub>2</sub>), 10mM DTT and 1 µL of RiboLock RNase inhibitor and incubated at 42 °C for 2 min. After addition of SuperScript II Reverse Transcriptase enzyme, the whole reaction was incubated at 42 °C for 60 min, followed by inactivation at 70 °C for 15 min. The resultant mixtures of cDNA were precipitated by adding isopropanol and then dissolved in water. cDNA generated from primer extension were mixed with 80% formaldehyde, heated at 80 °C for 3 min and then resolved by 12% denaturing PAGE gel. The gel was scanned using Typhoon phosphorimager (Figure S2). Nucleotide position for each band was identified from dideoxy sequencing lane (Figure S2). Secondary structure of RNA was generated by Mfold<sup>2</sup> and the SHAPE reactivities of accessible nucleotide were mapped on it.

Table S1: Primer used for PfU3 snoRNA cDNA synthesis

Name	Sequence
PfU3 snoRNA RT	5' cy5- AAG TTC AGC GGA CGC AAG CTC-3'



**Figure II-S2:** (A) cy5 labeled denaturing gel comparing modifications in the synthetic pfU3 snoRNA using 1M7 and 4b and 6 at 10 mM concentration (B) Partial secondary structure of PfU3 snoRNA. Nucleotides modified by the SHAPE electrophiles are indicated by black circles

#### ***in vitro* Transcription of TbTR catalytic core:**

The catalytic core of *T. brucei* telomerase RNA (TbTR) comprising of the Template boundary element (TBE), Template and Template proximal and distal helix regions were synthesized as a G-block oligos from IDT. T7 RNA promoter sequence was added to the 5' end of TBE during the synthesis. The resultant oligos was *in vitro* transcribed (IVT) using the MEGascript and purified MEGAclean kit (Life Technologies, USA) following the manufacturer's recommendation. The resultant RNA was quantified by nanodrop analysis.

#### ***in vitro* SHAPE modification:**

For SHAPE modification of TbTR catalytic core, 5 µg of *in vitro* transcribed RNA was denatured by heating at 95 °C for 2 min followed by immediate incubation on ice for 3 min. The modification was carried out by incubating the RNA in 3X SHAPE buffer consisting of 333 mM of NaCl, 333 mM of HEPES and 20 mM of MgCl<sub>2</sub> at 37 °C prior to the addition of 10 mM of 1M7, **6** and **4b** for 70 sec. As a negative control equivalent amount of DMSO/water was added to the RNA. Following incubation, the modified RNA was ethanol precipitated, dissolved in water and subjected to nanodrop analysis prior to cDNA synthesis.

### **cDNA synthesis and RNA-SHAPE analysis by denaturing Polyacrylamide Gel**

#### **Electrophoresis (PAGE):**

1.6 µg of SHAPE modified RNA was denatured by heating at 95 °C for 2 min followed by immediate incubation on ice for 3 min. For Primer extension, 5' cy5 labeled primer (Table S2) for reverse transcription along with 10mM dNTPs were mixed with the SHAPE modified RNA and incubated at 65 °C for 5 min. Four dideoxy (dd) sequencing reactions were also performed using the unmodified TbTR catalytic core in the similar manner as mentioned above. Additionally, to the individual sequencing reactions were also added ddATP, ddTTP, ddGTP and ddCTP. Following incubation all the reactions were supplemented with 1X SuperScript II First-strand buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl and 5 mM MgCl<sub>2</sub>), 10 mM DTT and 1 µl of RiboLock RNase inhibitor and incubated at 42 °C for 2 min. After addition of SuperScript II Reverse Transcriptase enzyme, the whole reaction was incubated at 42 °C for 60 min, followed by inactivation at 70 °C for 15 min. The, resultant mixtures of cDNA were precipitated by adding isopropanol and then dissolved in water. cDNA generated from primer extension were mixed with 80% formaldehyde, heated at 80 °C for 3 min and then resolved by 12% denaturing PAGE gel. The gel was scanned using Typhoon phosphorimager. Nucleotide position for each band was identified from dideoxy sequencing lane. Secondary structure of TbTR catalytic core was generated by RNAstructure software<sup>3</sup> and the SHAPE reactivities of accessible nucleotide were mapped on it.

Table II-S2: Primer used for TbTR Catalytic core cDNA synthesis

Name	Sequence
TbTR-Template RT	5' cy5- TAT TGC ACA CCA ACA GCA CTA -3'
TbTR-Template Proximal helix RT	5' cy5- CAA TGA TGA CGT GAA GAG GAG GA -3'
TbTR-Template Distal helix RT	5' cy5- TAA ACA AAC CGG CTA GAG AGG -3'

#### **Signal-to-background ratio calculation:**

Signal-to-background ratio (S/B) is the arithmetic mean of intensity of bands for sample lanes (except the dideoxy sequencing reactions and control lanes) to the background. Mean Pixel intensity of randomly chosen three bands for sample lane was calculated using

ImageJ software and divided by the mean pixel intensity of background which is the low intensity region present adjacent to each band. Same position bands were used for calculation for each lane.

### **High Resolution Mass Spectrometry:**

High resolution MS data were analyzed on an Orbitrap XL in positive mode at a resolution of 60000, spray voltage of 5 kV, sheath 12, Aux 10, Sweep 0, Capillary Temp 275 °C, Capillary Voltage 32.00 V, Tube Lens of 85.00 V Compounds were dissolved in 1 mL ACN, then 8  $\mu$ L was diluted in 1:1 H<sub>2</sub>O:ACN with 0.1% Formic acid. 2  $\mu$ L of this solution was injected using an Accela 1250 LC system, at a solvent flowrate of 200  $\mu$ L/min of 1:1 H<sub>2</sub>O:ACN with 0.1% Formic Acid. X-ray data were collected on an Agilent/Oxford Diffraction Gemini A Ultra diffractometer utilizing Mo radiation ( $\lambda = 0.71073$  Å) at 100 K.



## CHAPTER 4: A NOVEL SHAPE REAGENT ENABLES THE ANALYSIS OF RNA STRUCTURE IN LIVING CELLS WITH UNPRECEDENTED ACCURACY

Adapted with permission from Nucleic Acids Research, 2021;. gkaa1255. Copyright (2021) Oxford Academic.

### 4.1 ABSTRACT:

Due to the mounting evidence that RNA structure plays a critical role in regulating almost any physiological as well as pathological process, being able to accurately define the folding of RNA molecules within living cells has become a crucial need. We introduce here 2-aminopyridine-3-carboxylic acid imidazolide (2A3), as a general probe for the interrogation of RNA structures in vivo. 2A3 shows moderate improvements with respect to the state-of-the-art selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) reagent NAI on naked RNA under in vitro conditions, but it significantly outperforms NAI when probing RNA structure in vivo, particularly in bacteria, underlining its increased ability to permeate biological membranes. When used as a restraint to drive RNA structure prediction, data derived by SHAPE-MaP with 2A3 yields more accurate predictions than NAI-derived data. Due to its extreme efficiency and accuracy, we can anticipate that 2A3 will rapidly take over conventional SHAPE reagents for probing RNA structures both in vitro and in vivo.

### 4.2 INTRODUCTION

RNA structure is a jack-of-all-trades. Besides being crucial to the function of structural non-coding RNAs (ncRNAs), RNA structural elements are now widely recognized as key players in almost any essential biological process, ranging from transcriptional and post-transcriptional control of gene expression, to catalysis and sensing of environmental stimuli such as metabolite and temperature changes (1).

In light of its importance, being able to define the structure of an RNA molecule is a key need towards understanding its mechanism of action. It is however still impossible for most RNAs to define their secondary structures starting solely from their primary sequence. Main reason for this is the huge size of the RNA folding space, so that a  $n$  nucleotides-long RNA can theoretically fold in up to  $1.8^n$  distinct conformations (2). A widely used approach to define the putative structure of an RNA based on its sequence relies on predicting its minimum free energy (MFE) structure, the theoretically most stable structure, based on a set of experimentally-determined thermodynamic parameters known as nearest neighbor model (or Turner rules) (3). Unfortunately, the accuracy of MFE predictions is very limited, with only  $\sim 66\%$  of predicted base pairs actually occurring in experimentally-validated structures (4). This is likely to be a consequence of the fact that, beyond thermodynamics, a multitude of cellular factors play a key role in regulating RNA folding in the context of the living cell, such as RNA post-transcriptional modifications, protein binding, crowding and many others (5–7). The accuracy of thermodynamics-driven RNA structure determination algorithms can be however improved by incorporating experimental constraints from RNA footprinting experiments. These experiments are based on the use of specific chemicals or nucleases to probe the conformation of RNA residues. When informed with these experimentally-derived constraints, the accuracy of RNA structure prediction algorithms greatly improves (8–12).

Only few chemical probes can readily permeate cell membranes and can thus be used *in vivo*, while the majority of them are only suitable for *in vitro* and *ex vivo* RNA structure probing experiments. Two classes of compounds are mainly used to date. On

the one hand, nucleobase-specific probes, such as dimethyl sulfate (DMS) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), can be respectively used to probe adenine (A) and cytosine (C), or guanine (G) and uracil (U) bases (13,14). On the other hand, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) reagents can be used to probe the local flexibility of the RNA backbone in a nucleobase-independent fashion (15), as they can form adducts with the 2'-OH of the ribose moiety. Among SHAPE reagents, 3 have been extensively characterized: 1-methyl-7-nitroisatoic anhydride (1M7), 2-methyl-3-furoic acid imidazolide (FAI) and 2-methylnicotinic acid imidazolide (NAI) (16,17), with the latter being the most effective at probing RNA in living cells (18,19).

Typical readout of RNA probing experiments is the detection of reverse transcription (RT) drop-off events induced by the probing reagent-modified RNA bases (20). More recently, it has been shown that, by either using specific RT enzymes or RT conditions, it is possible to favor the RT read-through on modified RNA bases, ultimately resulting into the recording of multiple modification sites as mutations within the same cDNA molecule. These techniques have been therefore dubbed mutational profiling approaches (MaP) (21–24) and are rapidly taking over traditional RT drop-off methods. While the use of SHAPE compounds over nucleobase-specific probes has in theory a considerable advantage, as they enable the unbiased probing of all four RNA bases, reagents such as DMS still exhibit a higher signal-to-noise ratio in MaP experiments. To this end, we here sought to optimize *in vivo* RNA probing by synthesizing six new SHAPE reagents and by assaying their ability to probe *in vivo* Gram-positives, Gram-negatives and mammalian cells. One of these compounds, 2-aminopyridine-3-carboxylic

acid imidazolide (2A3), showed both increased reactivity with RNA and higher permeability to biological membranes, resulting in a significantly higher signal-to-noise ratio when compared to NAI. Importantly, when used to perform experimentally-informed prediction of in vivo RNA structures, 2A3 produced markedly more accurate predictions than NAI.

## **4.3 MATERIALS AND METHODS**

### **4.3.1 Synthesis of SHAPE reagents**

For the design of novel SHAPE reagents, we first compiled a list of commercially-available carboxylic acids. We then selected those containing nitrogen groups or for which the putative adduct-forming group had a pKa close to (or slightly above) 7. The pKa values in water were obtained (where possible) using the DataWarrior tool (25). The underlying rationale was to slow-down the hydrolysis in water and to favor reaction with RNA. For the synthesis of SHAPE reagents, 2-methylpyridine-3-carboxylic acid (cat. 325228), isoquinoline-6-carboxylic acid (cat. APO455831358–250MG), indoline-5-carboxylic acid (cat. CDS019637), 1-methylimidazole-4-carboxylic acid (cat. 679720), 6-aminopyridine-3-carboxylic acid (cat. 216879), benzotriazole-5-carboxylic acid (cat. 304239), nicotinic acid (cat. N4126), 2-aminopyridine-3-carboxylic acid (cat. A68300) and 1,1'-carbonyldiimidazole (CDI, cat. 21860) were purchased from Sigma Aldrich. For the synthesis of NAI, I6, I5, 1M4, B5, 6A3, NIC and 2A3, respectively 137.14, 173.17, 163.17, 126.11, 163.13, 138.12, 123.11 and 138.12 mg were resuspended in 500  $\mu$ l DMSO anhydrous (Sigma Aldrich, cat. 276855). ~1.3 g of CDI were then resuspended in 4 ml DMSO anhydrous and 500  $\mu$ l of this solution were added to each 500  $\mu$ l of carboxylic acids in DMSO while constantly stirring, over a period of 5 min. Reaction mixtures were then incubated at room temperature with constant stirring for 1 h.

Residual unreacted material was then removed by briefly centrifuging the solutions for 5 min at  $17\,000 \times g$ , after which the cleared solution was transferred to a clean tube, aliquoted into 50  $\mu$ l aliquots and stored at  $-80^{\circ}\text{C}$ .

#### **4.3.2 Chemical characterization**

NMR spectra were recorded on a JEOL-ECA 500 MHz NMR spectrometer or JEOL-ECX 300 MHz NMR in DMSO- $d_6$  or D $_2$ O. Spectra were referenced to the respective solvent peaks for  $^1\text{H}$  and  $^{13}\text{C}$ . Routine ESI MS data were collected on a ThermoFisher MSQ Plus single quadrupole mass spectrometer. High resolution MS data were analyzed on a ThermoFisher Orbitrap XL in positive mode. Chemical characterization is detailed in the Supplementary Information.

#### **4.3.3 Half-life determination**

SHAPE electrophiles were dissolved in 700  $\mu$ l PBS buffer  $1\times$  pH 7.4 prepared with D $_2$ O, and any undissolved solids were centrifuged out prior to transfer to the NMR tube. As needed, some samples were dissolved in DMSO  $d_6$  (10% final concentration) and added to D $_2$ O to achieve quantifiable concentrations. Samples were incubated at  $25^{\circ}\text{C}$  or  $37^{\circ}\text{C}$  and kinetic  $^1\text{H}$  NMR scanning was performed. The disappearance of product resonances and reappearance of starting materials (correlating to hydrolysis) was monitored as previously described (26). Half-life data is provided in Supplementary Table S1.

#### **4.3.4 Extraction of native deproteinized *E. coli* rRNAs**

Deproteinized *Escherichia coli* RNA was prepared essentially as previously described (23), with minor changes. Briefly, a single colony of *E. coli* K-12 DH10B was picked and inoculated in LB medium without antibiotics, then grown to exponential phase ( $\text{OD}_{600} \sim 0.5$ ). 2 ml aliquots were collected by centrifugation at  $1000 \times g$  ( $4^{\circ}\text{C}$ )

for 5 min. Cell pellets were resuspended in 1 ml of Resuspension Buffer [15 mM Tris–HCl pH 8.0; 450 mM sucrose; 8 mM EDTA], and lysozyme was added to a final concentration of 100 µg/ml. After incubation at 22°C for 5 min and on ice for 10 min, protoplasts were collected by centrifugation at  $5000 \times g$  (4°C) for 5 min. Pellets were resuspended in 120 µl Protoplast Lysis Buffer [50 mM HEPES pH 8.0; 200 mM NaCl; 5 mM MgCl<sub>2</sub>; 1.5% SDS], supplemented with 0.2 µg/µl Proteinase K. Samples were incubated for 5 min at 22°C and for 5 min on ice. Sample was then extracted 2 times with phenol:chloroform:isoamyl alcohol (25:24:1, pre-equilibrated three times with RNA folding buffer [50 mM HEPES pH 8.0; 200 mM NaCl; 5 mM MgCl<sub>2</sub>]), and once with chloroform. 20 U SUPERase•In™ RNase Inhibitor (ThermoFisher Scientific, cat. AM2696) were then added and RNA was equilibrated at 37°C for 20 min prior to probing.

#### **4.3.5 Ex vivo probing of E. coli rRNAs**

180 µl of deproteinized E. coli rRNAs in RNA folding buffer, pre-equilibrated at 37°C for 20 min, were mixed with 20 µl of each SHAPE compound (assuming complete conversion, stocks should be ~1 M, resulting in a final concentration of 100 mM in the probing reaction). RNA was then allowed to react at 37°C for 15 min, with moderate shaking, after which 200 µl of 1 M DTT were added, to quench the reaction. Samples were then vortexed briefly and 1 ml of ice-cold QIAzol Lysis Reagent (Qiagen, cat. 79306) was added to each sample, followed by extensive vortexing.

#### **4.3.6 In vivo probing of E. coli cells**

A single colony of E. coli K-12 DH10B was picked and inoculated in LB medium without antibiotics, then grown to exponential phase (OD<sub>600</sub> ~ 0.5). 2 ml aliquots were collected by centrifugation at  $1000 \times g$  for 5 min. Cell pellets were resuspended in 450 µl

1× PBS (pH 7.4) and 50 µl of each SHAPE compound were added, followed by brief vortexing. Cells were allowed to react at 37°C for 20 min, with moderate shaking, after which 500 µl of 1 M DTT were added, to quench the reaction. Cells were pelleted by centrifugation at  $17\,000 \times g$  for 2 min (4°C) and supernatant discarded. Cell pellets were then resuspended in 100 µl of Buffer A [10 mM Tris pH 8.0; 100 mM NaCl], supplemented with 5 U SUPERase•In™ RNase Inhibitor, by vigorously vortexing for 30 s. 25 µl of Buffer B [50 mM EDTA; 120 mM Tris pH 8.0], supplemented with 100 µg/ml final lysozyme, were then added, followed by 30 s vortexing. Samples were incubated 1 min at room temperature, after which 125 µl of Buffer C [0.5% Tween-20; 0.4% sodium deoxycholate; 2 M NaCl; 10 mM EDTA] were added to lyse protoplasts. Samples were then incubated at room temperature for 5 min, after which 100 µl of lysate were transferred to 1 ml ice-cold QIAzol and thoroughly mixed by vortexing.

#### **4.3.7 In vivo probing of *B. subtilis* cells**

A single colony of *Bacillus subtilis* 168 was picked and inoculated in LB medium without antibiotics, then grown to exponential phase ( $OD_{600} \sim 0.5$ ). 2 ml aliquots were collected by centrifugation at  $1000 \times g$  for 5 min. Cell pellets were resuspended in 450 µl 1× PBS (pH 7.4) and 50 µl of each SHAPE compound were added, followed by brief vortexing. Cells were allowed to react at 37°C for 20 min, with moderate shaking, after which 500 µl of 1 M DTT were added, to quench the reaction. Cells were pelleted by centrifugation at  $17\,000 \times g$  for 2 min (4°C) and supernatant discarded. Cell pellets were then resuspended in 200 µl TE Buffer [10 mM Tris pH 8.0; 1 mM EDTA pH 8.0], supplemented with 1.1% final SDS. 0.25 g glass beads were then added and samples were beaten in a Mini-Beadbeater-24 (Glen Mills) for 1 min. Samples were then incubated on ice for 1 min, after which beating was repeated once. 1 ml ice-cold QIAzol

was then added, followed by vigorous vortexing. Samples were then incubated at 70°C for 20 min, to allow complete cell lysis.

#### **4.3.8 In vivo probing of HEK293 cells**

HEK293 cells were grown in DMEM (4.5 g/l D-glucose), supplemented with 10% heat-inactivated FBS, 0.1 mM NEAA, 1 mM sodium pyruvate, 25 U/ml penicillin and 25 µg/ml streptomycin, at 37°C (5% CO<sub>2</sub>), to ~75% confluence. Cells were then covered with a thin layer of trypsin–EDTA (0.25%) solution, incubated at room temperature for 1 min, then dissociated by pipetting up and down with complete medium. Cells were centrifuged at 180 × g for 1 min and medium discarded. Cells were then resuspended in 450 µl 1× PBS (pH 7.4) and 50 µl of each SHAPE compound were added, followed by gentle mixing by tapping the tube. Cells were allowed to react at 37°C for 15 min, with moderate shaking, after which 500 µl of 1 M DTT were added, to quench the reaction. Cells were pelleted by centrifugation at 10 000 × g for 1 min (4°C) and supernatant discarded. Cells pellets were lysed by direct addition of 1 ml ice-cold QIAzol, followed by a brief incubation at 56°C for 5 min to completely dissolve cell aggregates.

#### **4.3.9 Total RNA extraction**

After samples have been collected in 1 ml QIAzol, 200 µl of chloroform was added, followed by vigorous vortexing for 15 s. Samples were then incubated at room temperature for 2 min, after which they were centrifuged at 12 500 × g for 15 min (4°C). After centrifugation, the upper aqueous phase was transferred to a clean 2 ml tube, and mixed with 2 volumes of 100% ethanol by vigorous vortexing. The entire volume was then transferred to an RNA Clean & Concentrator™-5 column (Zymo Research, cat. R1015) and RNA was purified as per manufacturer instructions. RNA integrity was



ensured by gel electrophoresis on a denaturing 2% agarose gel. Besides samples treated with I6, all other samples appeared to be perfectly intact. Before proceeding to library preparation, traces of genomic DNA were removed by treatment with 2 U TURBO™ DNase (ThermoFisher Scientific, cat. AM2239) at 37°C for 20 min.

#### **4.3.10 SHAPE-MaP library preparation**

Total RNA was first fragmented to a median size of 150 nt by incubation at 94°C for 8 min in RNA Fragmentation Buffer [65 mM Tris–HCl pH 8.0; 95 mM KCl; 4 mM MgCl<sub>2</sub>], then purified with NucleoMag NGS Clean-up and Size Select beads (Macherey Nagel, cat. 744970), supplemented with 10 U SUPERase•In™ RNase Inhibitor, and eluted in 2 µl (for SSII and HIV RT) or 2.5 µl (for TGIRT-III) NF H<sub>2</sub>O. Eluted RNA was supplemented with 0.5 µl 20 µM random hexamers and either 0.25 µl (for SSII) or 0.5 µl (for HIV RT and TGIRT-III) dNTPs (10 mM each), then incubated at 70°C for 5 min and immediately transferred to ice for 1 min. Reverse transcription reactions were conducted in a final volume of 5 µl. For HIV RT, reaction was supplemented with 1 µl 5× RT Buffer [250 mM Tris–HCl pH 8.3; 375 mM KCl; 40 mM MgCl<sub>2</sub>], 0.5 µl DTT 0.1 M, 5 U SUPERase•In™ RNase Inhibitor and 5 U HIV RT (Worthington Biochemical Corporation, cat. HIVRT). For TGIRT-III, reaction was supplemented with 1 µl 5× RT Buffer [250 mM Tris–HCl pH 8.3; 375 mM KCl; 15 mM MgCl<sub>2</sub>], 0.25 µl DTT 0.1 M, 5 U SUPERase•In™ RNase Inhibitor and 50 U TGIRT-III RT (InGex, cat. TGIRT-50). For SSII, reaction was supplemented with 1 µl 5× RT Buffer [250 mM Tris–HCl pH 8.3; 375 mM KCl], 0.5 µl DTT 0.1 M, 0.25 µl MnCl<sub>2</sub> 120 mM, 5 U SUPERase•In™ RNase Inhibitor and 50 U SuperScript II RT (ThermoFisher Scientific, cat. 18064014). Reactions were incubated at 25°C for 10 min to allow partial primer extension, followed by 2 h at 37°C (for HIV RT), 2 h at 57°C (for TGIRT-III) or 3 h at 42°C (for SSII). SSII

and HIV RT were heat-inactivated by incubating at 75°C for 20 min. As TGIRT-III tightly binds the cDNA-RNA complex, 0.5 µg of Proteinase K were added and reaction was incubated at 37°C for 20 min, after which 0.5 µl of a 1:2 dilution of protease inhibitor cocktail (Sigma Aldrich, cat. P9599) in water was added to stop the reaction. For SSII, the buffer exchange step was omitted. Instead, 6 mM final EDTA was added to chelate Mn<sup>2+</sup> ions, followed by 5 min incubation at room temperature and addition of 6 mM final MgCl<sub>2</sub>. Reverse transcription reactions were then used as input for the NEBNext® Ultra II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs, cat. E6111L). Second strand synthesis was performed by incubating 1 h at 16°C, as per manufacturer instructions. DsDNA was purified using NucleoMag NGS Clean-up and Size Select beads, and used as input for the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, cat. E7645L), following manufacturer instructions.

#### **4.3.11 Analysis of SHAPE-MaP data**

All the relevant data analysis steps have been conducted using RNA Framework v2.6.9 (27). Reads pre-processing and mapping was performed using the rf-map module (parameters: -ctn -cmn 0 -cqo -cq5 20 -b2 -mp ‘-very-sensitive-local’). Reads were trimmed of terminal Ns and low-quality bases (Phred < 20). Reads with internal Ns were discarded. Mapping was performed using the ‘very-sensitive-local’ preset of Bowtie2 (28). The mutational signal was derived using the rf-count module (parameters: -m -rd), enabling the right re-alignment of deletions. Grid search of optimal slope/intercept pairs was performed using the rf-jackknife module (parameters: -rp ‘-md 600 -nlp’ -x) and ViennaRNA package v2.4.11 (29), disallowing isolated base-pairs, setting the maximum base-pairing distance to 600 nt and enabling the relaxed structure comparison mode (8) (a

basepair  $i/j$  is considered to be correctly predicted if any of the following pairs exist in the reference structure:  $i/j$ ;  $i - 1/j$ ;  $i + 1/j$ ;  $i/j - 1$ ;  $i/j + 1$ ).

#### 4.3.12 Reference structure of rRNAs

16S and 23S rRNA reference structures for *E. coli* and *B. subtilis* were obtained from the Comparative RNA Web (30). 18S and 28S rRNA reference structures for *H. sapiens* were obtained from RNACentral (31).

#### 4.3.13 Generation of Receiver Operating Characteristic (ROC) curves

For the generation of ROC curves, the mutation frequency threshold was linearly varied from 0 to 1 with a step of 0.001, calculating at each increment the number of unpaired (true positive) and base-paired (true negative) residues (according to the accepted reference rRNA structures), whose mutation frequency exceeded that of the threshold. Terminal base-pairs (beginning/end of helices and isolated base-pairs) were excluded. For in vivo samples, only solvent-accessible bases were considered, as previously described (32). For solvent accessible surface area (SASA) calculation, ribosome structures were obtained from the Protein Data Bank, using accessions 5IT8 (*E. coli*), 6HA1 (*B. subtilis*) and 4UG0 (*H. sapiens*) in CIF format and converted to PDB format using PyMOL (v2.3.5). SASA per residue was then calculated using POPS (v3.1.2) and a probe radius of 3 Å. Accessible residues were defined as those residues with a SASA  $> 2$  Å. Area under the curve (AUC) was calculated using the MESS package (v0.5.6) for R.

#### 4.3.14 Signal-to-noise estimation

From reference rRNA structures in dot-bracket notation, we extracted all hairpin-loops with a stem length  $\geq 3$  and a loop size  $\geq 3$ , using the regular expression  $'([\backslash 3\backslash ., [\backslash 3\backslash ., [\{3\}])'$ . For each hairpin-loop, we then calculated the median reactivity on

the first and last two bases of the loop and the median reactivity on the three bases of the stem preceding and following the loop. The signal-to-noise value was then calculated as the ratio between the median reactivity across all the loops and the median reactivity across all the stems.

## 4.4 RESULTS

### 4.4.1 Comparison of RT enzymes for SHAPE-MaP

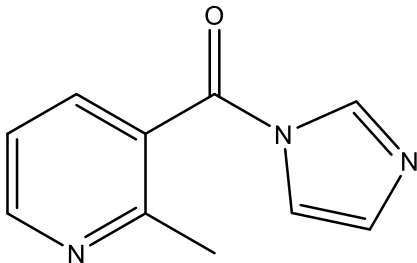
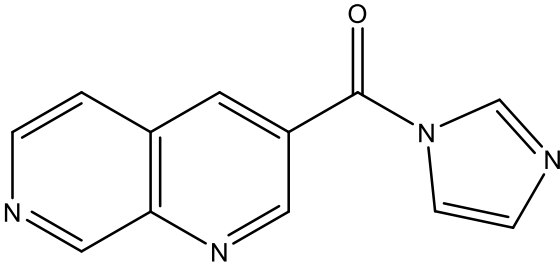
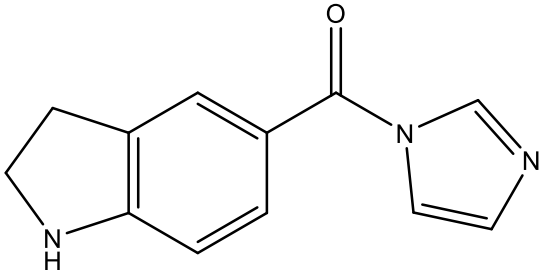
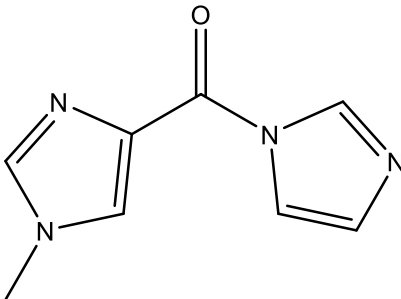
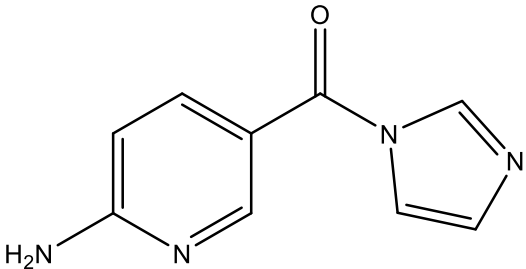
In a first attempt to optimize the signal-to-background ratio of SHAPE mutational profiling approaches (SHAPE-MaP), we sought to find the best reverse transcription conditions. To this end, we extracted deproteinized total RNA from *E. coli* cells, under conditions previously shown to preserve the native RNA folding (8,23,33) and probed it in solution with NAI (or neat DMSO as a control). We then performed reverse transcription using three different reverse transcriptases: SuperScript II (SSII) in  $Mn^{2+}$ -containing buffer, thermostable group II intron RT (TGIRT-III) and HIV RT (22,24,34). These enzymes have been previously shown to be able to generate mutational signatures in cDNA when reading through SHAPE adducts (SSII), DMS-induced alkylations (SSII and TGIRT-III), or N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT) adducts (HIV RT). Analysis of the distribution of per-base mutation frequencies across 16S and 23S rRNAs showed that, while TGIRT-III has lower background mutation frequencies compared to SSII (median DMSO control:  $0.75 \times 10^{-3}$  for TGIRT-III versus  $2.19 \times 10^{-3}$  for SSII) as previously reported (22), SSII shows a higher and more significant increase in the mutation rates measured in the NAI treated sample with respect to the DMSO control (ratio median NAI versus DMSO: 2.95 for SSII, 1.37 for TGIRT-III and 1.16 for HIV RT; Supplementary Figure S1A). Accordingly, receiver operating characteristic (ROC) curves built with respect to

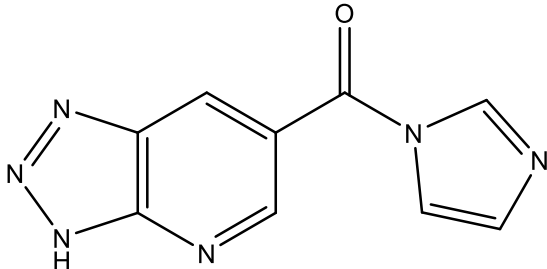
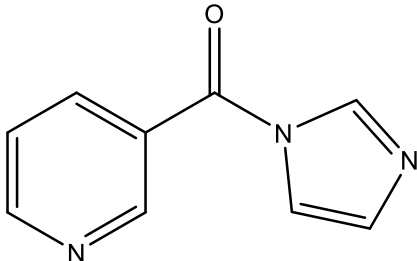
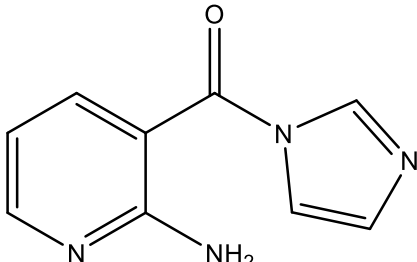
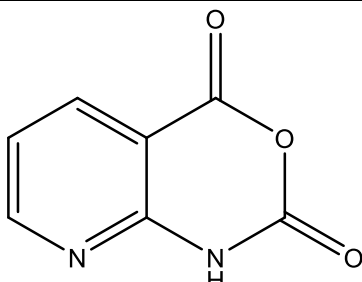
unpaired versus paired residues from phylogenetically-inferred *E. coli* rRNA secondary structure models (30) and inspection of SHAPE-MaP signals confirmed the higher ability of SSII to translate SHAPE adducts on structurally-flexible bases into a mutational signature (Supplementary Figure S1B, C). Altogether, these results identify SSII in Mn<sup>2+</sup>-containing buffer as the optimal RT enzyme for SHAPE-MaP experiments.

#### **4.4.2 Synthesis of six novel candidate SHAPE reagents**

While designing new candidate SHAPE compounds, we sought to meet the following requirements: (i) permeable to biological membranes (suitable for in vivo probing experiments); (ii) high signal over background (deriving from a good tradeoff between slow hydrolysis in water, reactivity with RNA and membrane permeability); (iii) easy and safe to synthesize. The two best-characterized in vivo SHAPE reagents to date are NAI and FAI. These compounds can be easily synthesized via the carbonyldiimidazole (CDI)-mediated activation of carboxylic acids, resulting in the corresponding carboxylic acid imidazolide, an activated carbonyl, that can acylate nucleophiles present in the system, especially 2' hydroxyl groups of RNA. We decided to adopt the same strategy, trying to optimize the acylating group. To this end, we selected seven commercially-available carboxylic acids, preferring those containing nitrogen groups or for which the putative adduct-forming group had a pK<sub>a</sub> in water close to (or slightly above) 7 (Table 1; see Materials and Methods).

Table 4-1. SHAPE reagents used in this study. Chemical structure, full name and acronym of the SHAPE reagents tested in this study. NAI, used as a standard, is reported for comparison.

Compound	Full name	Acronym
	2-Methylpyridine-3-carboxylic acid imidazolid	NAI
	Isoquinoline-6-carboxylic acid imidazolid	I6
	Indoline-5-carboxylic acid imidazolid	I5
	1-Methylimidazole-4-carboxylic acid imidazolid	1M4
	6-Aminopyridine-3-carboxylic acid imidazolid	6A3

	Benzotriazole-5-carboxylic acid imidazolid	B5
	Nicotinic acid imidazolid	NIC
	2-Aminopyridine-3-carboxylic acid imidazolid	2A3
	3-Azaisatoic anhydride (byproduct of 2A3 synthesis)	3AIA

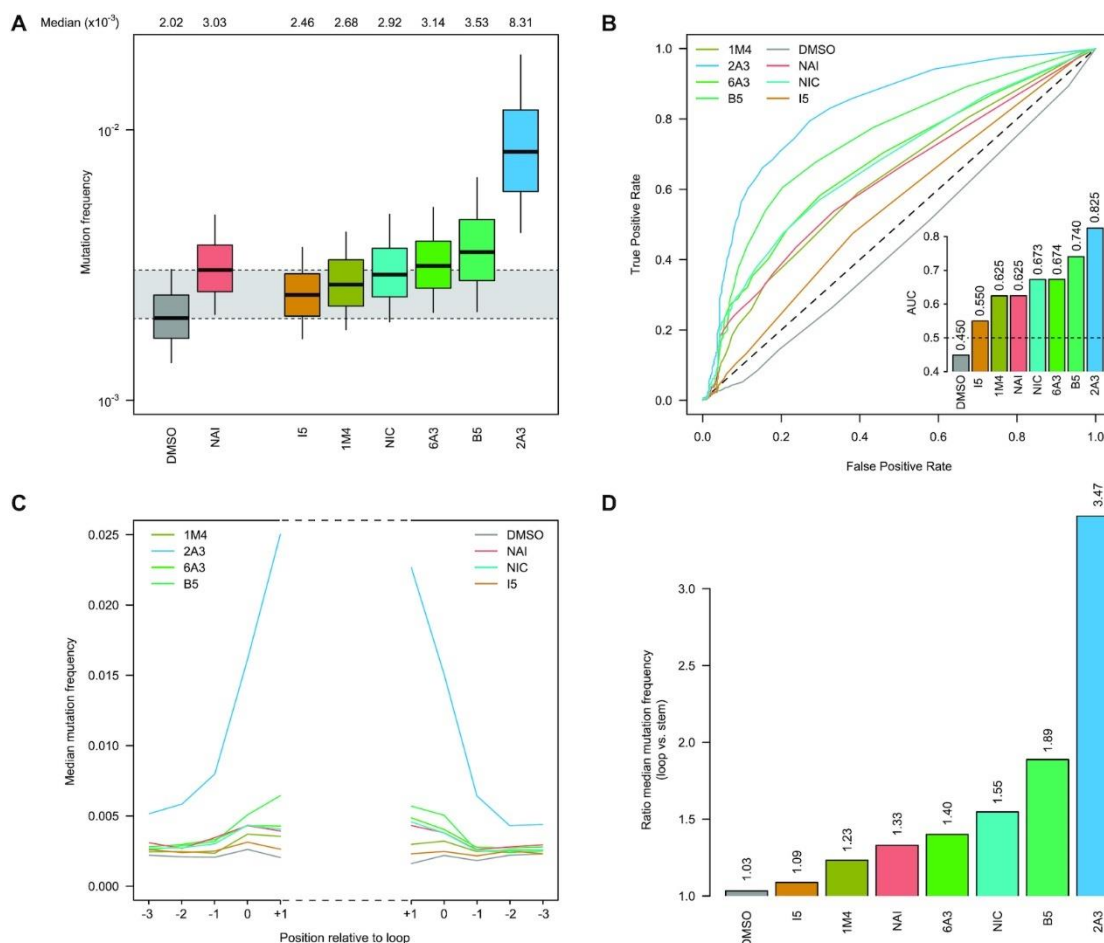
We then evaluated all reagents for their ability to react with structurally-flexible RNA residues using *E. coli* ex vivo deproteinized total RNA and NAI probing as the gold standard. We further discarded isoquinoline-6-carboxylic acid imidazolid (I6) as it was too reactive, hence resulting in very fast RNA degradation (data not shown). When looking at overall distributions of mutation rates, indoline-5-carboxylic acid imidazolid (I5) and 1-methylimidazole-4-carboxylic acid imidazolid (1M4) showed a lower reactivity towards RNA compared to NAI (median:  $3.07 \times 10^{-3}$  for I5,  $4.50 \times 10^{-3}$  for 1M4). Oppositely, the other four compounds, nicotinic acid imidazolid (NIC),

benzotriazole-5-carboxylic acid imidazolide (B5), 6-aminopyridine-3-carboxylic acid imidazolide (6A3) and 2-aminopyridine-3-carboxylic acid imidazolide (2A3), all showed increased reactivity compared to NAI (median:  $8.90 \times 10^{-3}$  for NIC,  $9.90 \times 10^{-3}$  for B5,  $11.9 \times 10^{-3}$  for 6A3 and  $15.2 \times 10^{-3}$  for 2A3), with 2A3 having ~2.4X higher mutation frequencies than NAI (Figure 1A). ROC curves built with respect to unpaired versus paired residues from *E. coli* rRNAs revealed nearly comparable accuracies for 6A3, B5, NIC and 2A3, with NIC and 2A3 showing a slightly higher area under the curve (AUC), independently of DMSO background signal subtraction (Figure 1B and Supplementary Figure S2A). This intrinsically implies that these two reagents have a slightly higher signal-to-noise ratio with respect to NAI, as demonstrated by the higher mutation frequencies observed on single-stranded RNA regions (Figure 4-1C). To systematically quantify the actual signal-to-noise ratio of the different reagents, we extracted all the individual hairpin-loops from the known structures of *E. coli* 16S and 23S rRNAs, and calculated the median mutation frequency across the first and last two bases of the loop, versus the three stem bases preceding or following the loop (Supplementary Figure S2B). Notably, the ratio between the median loop reactivity and the median stem reactivity for 2A3 (4.01), B5 (4.01) and NIC (4.19) was higher than that of NAI (3.59), confirming that these reagents might represent better choices for in vivo RNA probing.





At first, we looked again into the distribution of mutation frequencies for the different reagents. Unexpectedly, while the observed distributions for HEK293 cells approximately mirrored those observed on *E. coli* ex vivo-modified RNAs (Supplementary Figure S3A), we observed that, for both *E. coli* and *B. subtilis*, mutation frequency distributions for most reagents, including NAI, were extremely close to those of the control DMSO sample (Figure 4-2A and Supplementary Figure S3B). On the other hand, 2A3 showed significantly higher mutation frequencies compared to NAI (ratio median 2A3 versus NAI: 2.74 in *E. coli*, 1.61 in *B. subtilis* and 1.77 in HEK293), independently of the probed cell type, partially followed by B5 in *B. subtilis* and NIC in HEK293 cells.



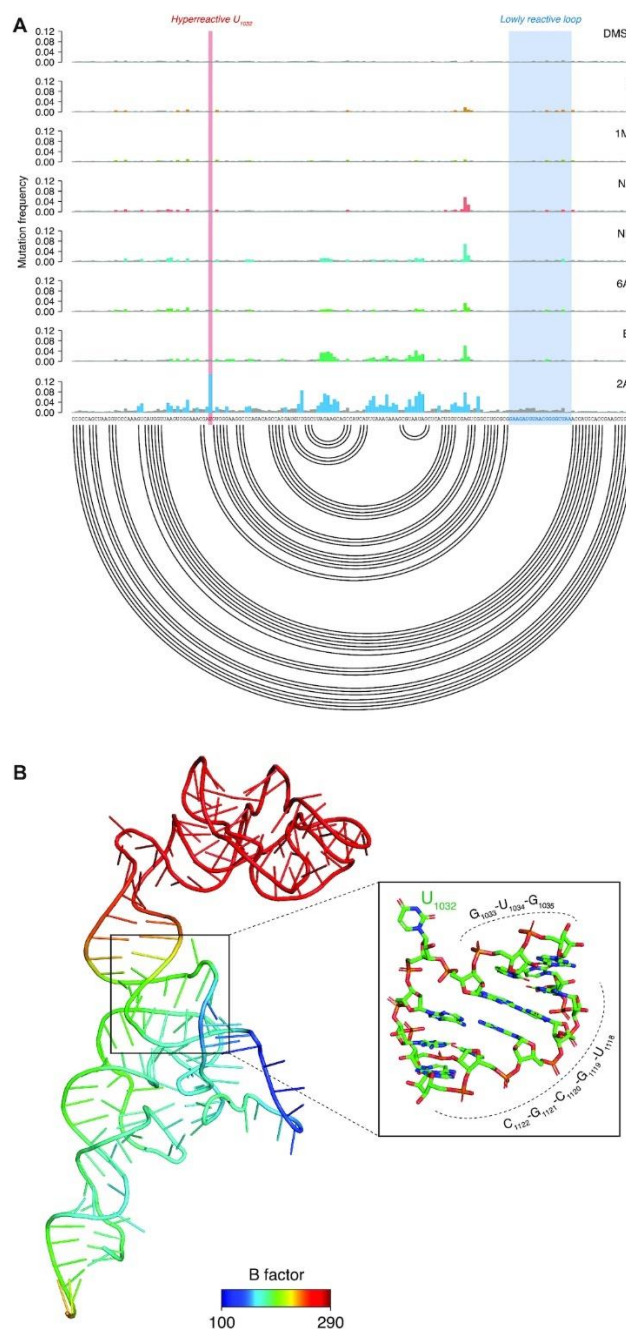
**Figure 4-2.** Comparison of SHAPE reagents under in vivo conditions. (A) Boxplot of SHAPE–MaP mutation frequencies for *E. coli* 16S and 23S rRNAs probed in vivo. Box plots span the interquartile range (from Q1 to Q3). The grey area spans from the median in the DMSO sample (control) to the median in the NAI sample (reference). (B) ROC curves for all tested SHAPE reagents, calculated on solvent-exposed residues in the crystal structure of the *E. coli* ribosome (PDB: 5IT8), with respect to the accepted phylogenetically-inferred 16S and 23S rRNA structures from the Comparative RNA Web (30). The inset reports the AUC for each compound. (C) Median in vivo SHAPE–MaP mutation frequencies across all stem–loops in the accepted *E. coli* 16S and 23S rRNA structures. Bases are numbered relatively to the loop. Positions –3 to –1 correspond to stem bases, while positions 0 and +1 correspond to loop bases. (D) Ratio between the median loop mutation frequency and the median stem mutation frequency, calculated on all stem–loops from C.

Oppositely to what we observed for the ex vivo-probed RNAs, ROC curves built with respect to unpaired versus paired residues from *E. coli*, *B. subtilis* and *H. sapiens* rRNAs (calculated on residues that are solvent-exposed in the respective ribosome crystal structures; see Materials and Methods) revealed a marked difference in

sensitivity/specificity of NAI compared to 2A3, with 2A3 showing up to 17–20% increased discrimination of unpaired versus paired residues, independently of DMSO background signal subtraction (AUC *E. coli*: 0.825/0.874 for 2A3  $-/+$  DMSO subtraction, 0.625/0.708 for NAI  $-/+$  DMSO subtraction; Figure 4-2B and Supplementary Figure S4–5). Accordingly, quantification of the SHAPE-MaP signal on loops versus stem regions confirmed the strikingly higher signal-to-noise ratio of 2A3 (*E. coli*: 3.47; *B. subtilis*: 2.36; HEK293: 3.01) with respect to NAI (*E. coli*: 1.33; *B. subtilis*: 1.56; HEK293: 2.08), independently of the examined cell type (Figure 4-2C, D and Supplementary Figure S6). Base reactivity distributions for 2A3-modified RNAs on unpaired, internally-paired and terminally-paired bases, under all the analyzed conditions, appeared to be in line with those previously reported for other SHAPE reagents (35) (Supplementary Figure S7).

Deeper inspection of SHAPE-MaP signals revealed that, regions of bacterial ribosomes that were almost completely blind to NAI and most of the other reagents, were instead partially accessible to B5 and readily probed with high efficiency by 2A3, with mutational signatures in strong agreement with the known rRNA structure (Figure 4-3A and Supplementary Figure S8). In these regions, not only 2A3 effectively probed most RNA loops, but also it better captured local nucleotide dynamics. As an example, base U1032 of *E. coli* 23S rRNA is hyperreactive to 2A3, but almost completely unreactive to any other tested reagent, including NAI (Figure 4-3A). Coherently, analysis of this base in the context of the crystal structure of the fully assembled *E. coli* ribosome (36) reveals its extreme flexibility, as it is twisted, completely flipping out of the parent helix (Figure 4-3B). Oppositely, a large loop of 19 contiguous single-stranded residues between

positions 1124 and 1181 showed very low reactivities. Notably, this loop has a very low B factor in the crystal, suggesting that these residues are weakly dynamic (lowly flexible) in the context of the ribosome. In agreement with this observation, these residues became highly reactive towards 2A3 upon deproteinization (Supplementary Figure S9). Similarly, the hyperreactivity of U1032 was abrogated upon rRNA deproteinization. In HEK293, although NAI readily accessed most regions of the ribosome, showing very similar reactivity patterns to the other tested reagents, we observed that probing with 2A3 (as well as with NIC) resulted in a stronger and neater SHAPE-MaP signal (Supplementary Figure S10).

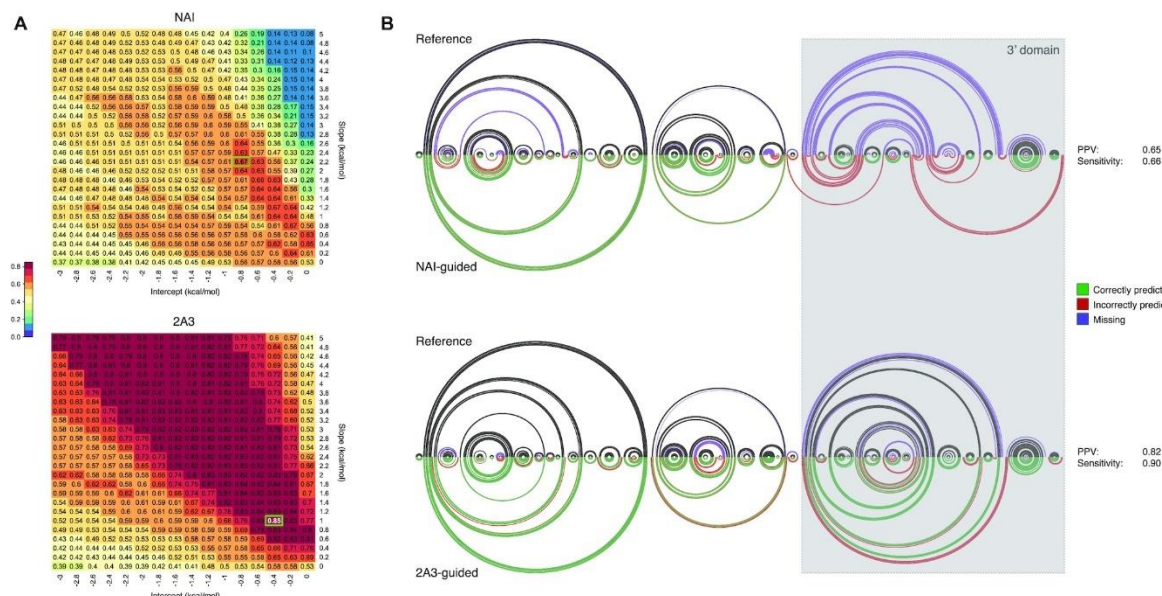


**Figure 4-3.** 2A3 successfully queries regions of the ribosome that are blind to other reagents. (A) Sample of SHAPE–MaP mutation frequencies for all tested compounds across a region spanning nucleotides 990 to 1162 of *E. coli* 23S rRNA. Colored bases are those whose mutation frequencies exceed by 2-fold the median mutation frequency in the analyzed region. The accepted structure is reported as an arc plot. The hyperreactive residue U1032 and the unreactive stretch of 19 nucleotides are respectively marked in red and blue. (B) Three-dimensional model of the domain depicted in A, colored by B factor (PDB: 5IT8). The inset zooms on the helical region containing the 2A3-hyperreactive residue U1032.

#### 4.4.4 2A3 markedly improves the accuracy of RNA structure prediction

To definitively prove the better suitability of 2A3 for in vivo RNA folding studies, we used SHAPE-inferred reactivities to perform experimentally-driven RNA structure modeling. SHAPE reactivities can be incorporated into thermodynamics-based prediction algorithms by converting them into pseudo-free energy contributions (8). This conversion relies on the use of two empirically-determined parameters, the slope ( $m$ ) and the intercept ( $b$ ). As these parameters depend on the experimental setup and on the employed reagent, we first used the ViennaRNA Package 2.0 (29) to perform a grid search of the optimal  $m/b$  pairs for NAI and 2A3 for *E. coli* rRNAs, in order to avoid skewing prediction results because of the use of the same  $m/b$  pair for both reagents (Figure 4-4A). For each  $m/b$  pair, we calculated the geometric mean of the prediction's sensitivity (the fraction of known base pairs correctly predicted) and positive predictive value (PPV; the fraction of predicted pairs that are present in the known reference structure) for both 16S and 23S rRNAs. Notably, grid search already revealed that the accuracy of 2A3-driven predictions was largely independent of the employed slope/intercept pair and that, even the most accurate NAI-driven prediction ( $m$ : 2.2;  $b$ : -0.8; geometric mean of PPV/sensitivity: 0.67) was ~10% less accurate than most 2A3-driven predictions, independently of the used  $m/b$  pair and ~18% less accurate than the top-scoring 2A3-driven prediction ( $m$ : 1.0;  $b$ : -0.4; geometric mean of PPV/sensitivity: 0.85). As an example, although incorporation of NAI data slightly increased the accuracy with respect to the unconstrained prediction (PPV: 0.47 for unconstrained, 0.65 for NAI-constrained; sensitivity: 0.54 for unconstrained, 0.66 for NAI-constrained), the NAI-guided prediction resulted in the misfolding of the 3' domain of the 16S rRNA, while the 2A3-guided prediction correctly recapitulated most of the known base-pairs from the

accepted reference structure (PPV: 0.82; sensitivity: 0.90; Figure 4-4B). We observed the same trend on other cellular RNAs (Supplementary Figures S11 and S12). A similar improvement was observed also when modeling the structure of *B. subtilis* rRNAs (Supplementary Figure S13). Use of RNAstructure (37) rather than ViennaRNA produced similar results (Supplementary Figure S14), hence demonstrating that our conclusions are software-independent.



**Figure 4-4.** 2A3 outperforms NAI at experimentally-driven RNA structure modeling. (A) Grid search (jackknifing) of optimal slope/intercept value pairs for *E. coli* 16S and 23S rRNAs in vivo probing data for NAI and 2A3. Values represent the geometric mean of sensitivity and PPV for the secondary structures predicted using each slope/intercept value pair. The chosen value pair is boxed in green. (B) Arc plot comparison of *E. coli* 16S rRNA reference structure (top), and structure inferred using either NAI-derived or 2A3-derived restraints (bottom). Black/green arcs correspond to correctly predicted base-pairs, violet arcs to non-predicted base-pairs and red arcs to mispredicted base-pairs. PPV and sensitivity for each prediction are indicated.

#### 4.4.5 Characterization of 2A3

Analytical preparations of 2A3 in DMSO unexpectedly revealed that a secondary product was occurring. This reaction proceeded slowly at room temperature after the rapid initial formation of the desired imidazolide, 2A3. The resulting product was a



cyclization reaction leading to the formation of 3-azaisatoic anhydride (3AIA). This unwanted side product increased over time, eventually reaching a maximum of ~90% as compared to 2A3 in solution during an overnight reaction in a sealed vessel. The formation of 2A3 proceeds rapidly as CDI is added to the 2-aminonicotinic acid. The conversion of 2A3 to 3AIA then proceeds at room temperature in the crude reaction mixture. The unwanted side reaction proceeds slowly and can be minimized with vigorous stirring under a constant flow of inert gas.

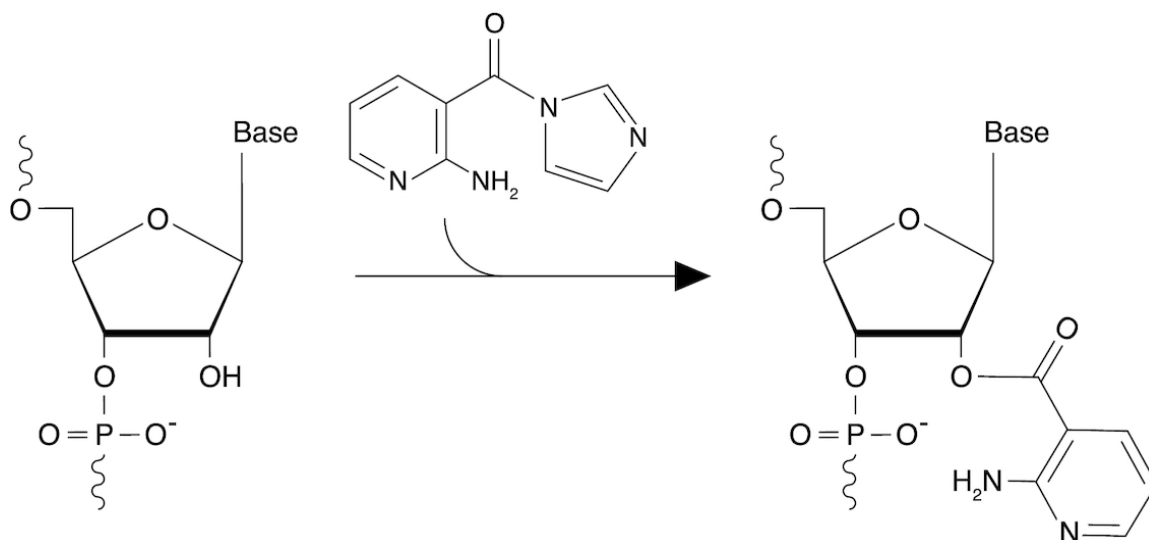
To rule out the possibility that 3AIA rather than 2A3 was acting as an *in vivo* SHAPE probe, we obtained commercially-available pure 3AIA (Synchem, Germany) and used it to probe living *E. coli* cells. Notably, 3AIA showed modification frequencies indistinguishable from those of the DMSO control *in vivo* (Supplementary Figure S15), hence confirming that 2A3 was indeed responsible for the previously observed RNA modification.

To further rule out the possibility that the higher probing efficiencies observed for 2A3 with respect to NAI were due to cell lysis, we incubated HEK293, *E. coli* and *B. subtilis* cells with 100 mM 2A3 at 37°C for 30 min. Visual inspection of the HEK293 cells by bright-field microscopy did not reveal any detectable morphological change, nor cell lysis (Supplementary Figure S16A). Similarly, the measured optical density for both *E. coli* and *B. subtilis* did not decrease over time, indicating that even after probing for 30 min, cells were still intact (Supplementary Figure S16B, C).

## 4.5 DISCUSSION

We have introduced here 2-aminopyridine-3-carboxylic acid imidazolide (2A3; Figure 4-5), a novel powerful and accurate SHAPE reagent for the interrogation of RNA

structures, both in vitro and in vivo. Briefly, after synthesizing six new candidate SHAPE reagents, we evaluated their ability to modify RNA in cell-free conditions. Although all compounds were able to react with RNA to different extents, four of them (B5, NIC, 6A3 and 2A3) showed partial improvements with respect to NAI. Surprisingly, when tested in vivo, 2A3 significantly outperformed NAI and all the other tested reagents under all the analyzed conditions, with the most striking differences being observed in bacterial cells. Indeed, for both *E. coli* and *B. subtilis*, several regions of the ribosome resulted to be blind to probing by NAI and most of the other tested reagents, while they were readily probed by 2A3. In this context, 2A3 better captured local nucleotide dynamics, as confirmed by in depth analysis of available ribosome 3D structures. This reveals that, besides showing a higher ability to react with RNA, 2A3 has also a greater capacity to permeate biological membranes. It is conceivable that the peptidoglycan of the bacterial cell wall might represent the main limitation when it comes to in vivo SHAPE probing. Indeed, although 2A3 is able to probe both gram-positive and gram-negative bacteria, we observed higher efficiencies in gram-negatives, suggesting that the thicker peptidoglycan of gram-positives limits the permeability to the SHAPE compound. The presence of an amine rather than a methyl group at position C2, that differentiates 2A3 from NAI, might explain this increased permeability, in agreement with a recent study showing that small molecules containing amine groups are more likely to permeate and accumulate into bacteria (38).



**Figure 4-5:** Mechanism of 2A3 reaction with RNA. Reaction of 2A3 with the 2'-OH of structurally-flexible RNA residues, resulting in the formation of a 2'-O-adduct.

Furthermore, we showed that the higher signal-to-noise ratio achieved by 2A3 results in a significantly higher RNA structure prediction accuracy when performing 2A3-informed RNA folding. It is worth noticing that 2A3, as opposed to NAI, outperforms probing with dimethyl sulfate (DMS) as well. Indeed, probing with DMS usually shows a higher signal-to-noise ratio with respect to NAI, hence resulting in more accurate experimentally-informed RNA structure predictions (see for example Figure 4-3B in (27) and Figure 4-4A in this work; geometric mean of PPV/sensitivity on *E. coli* 16S and 23S rRNAs: 0.71 for DMS, 0.67 for NAI) even if providing information on roughly only a half of RNA bases.

Compounds NIC, 2A3 and 6A3 were designed around the nicotinic acid core of NAI, where 2A3 and 6A3 were chosen to be electronically similar but sterically different. The electronic effects predict similar reactivities for 2A3 and 6A3, but the amino group ortho to the activated carbonyl in 2A3 is expected to decrease the reactivity due to steric considerations much the way the ortho methyl in NAI decreases the reactivity of NAI relative to NIC. Surprisingly, 2A3 was found to be more reactive than 6A3 as shown by

their hydrolysis half-lives (Table S1). It may be that 2A3 falls in a ‘sweet’ spot for reactivity compared to the other tested electrophiles both in vitro and in vivo.

In summary, 2A3 represents a concrete and significant advance in our ability to interrogate RNA structures in living cells. We can anticipate that, being able to readily probe even regions of ribonucleoprotein complexes that are normally hidden to other compounds, 2A3 will rapidly overtake conventional SHAPE reagents, revealing new and previously unobserved features of the in vivo RNA structurome.

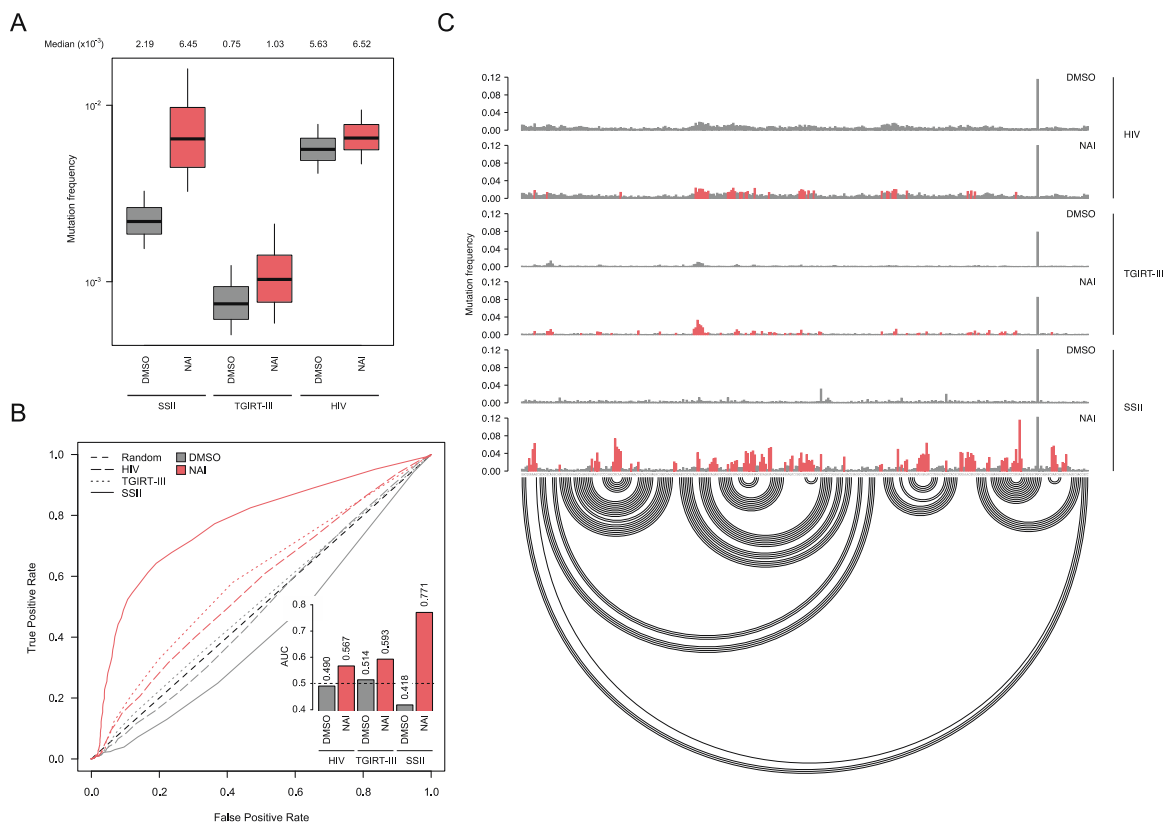
#### 4.6 REFERENCES:

1. Mortimer, S. A., Kidwell, M. A. and Doudna, J. A. (2014) Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.*, 15, 469–479.
2. Zuker, M. and Sankoff, D. (1984) RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46, 591–621.
3. Xia, T., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C. and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37, 14719–14735.
4. Mathews, D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, 10, 1178–1190.
5. Leamy, K. A., Assmann, S. M., Mathews, D. H. and Bevilacqua, P. C. (2016) Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biophys.*, 49, e10.
6. Helm, M. (2006) Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.*, 34, 721–733.
7. Weeks, K.M. (1997) Protein-facilitated RNA folding. *Curr. Opin. Struct. Biol.*, 7, 336–342.
8. Deigan, K. E., Li, T. W., Mathews, D. H. and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 97–102.
9. Incarnato, D., Neri, F., Anselmi, F. and Oliviero, S. (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.*, 15, 491.
10. Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H. and Weeks, K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *PNAS*, 110, 5498–5503.
11. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 7287–7292.

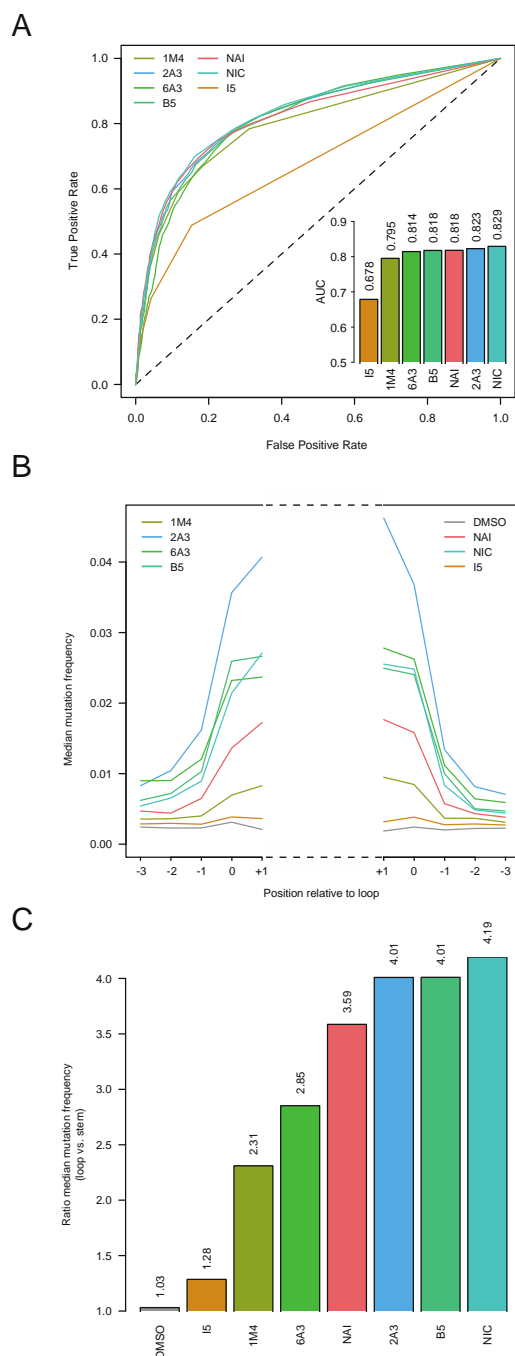
12. Cordero, P., Kladwang, W., VanLang, C. C. and Das, R. (2012) Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry*, 51, 7037–7039.
13. Wells, S. E., Hughes, J. M., Igel, A. H. and Ares, M. (2000) Use of dimethyl sulfate to probe RNA structure in vivo. *Methods Enzymol.*, 318, 479–493.
14. Mitchell, D., Renda, A. J., Douds, C. A., Babitzke, P., Assmann, S. M. and Bevilacqua, P. C. (2018) In vivo RNA structural probing of uracil and guanine base pairing by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). *RNA*, 25, 147–157.
15. McGinnis, J. L., Dunkle, J. A., Cate, J. H. D. and Weeks, K. M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.*, 134, 6617–6624.
16. Mortimer, S. A. and Weeks, K. M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, 129, 4144–4145.
17. Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T. and Chang, H. Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, 9, 18–20.
18. Lee, B., Flynn, R. A., Kadina, A., Guo, J. K., Kool, E. T. and Chang, H. Y. (2016) Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA*, 23, 169–174.
19. Busan, S., Weidmann, C. A., Sengupta, A. and Weeks, K. M. (2019) Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry-us*, 58, 2655–2664.
20. Incarnato, D. and Oliviero, S. (2017) The RNA epistructurome: uncovering RNA function by studying structure and Post-Transcriptional modifications. *Trends Biotechnol.*, 35, 318–333.
21. Homan, P. J., Favorov, O. V., Lavender, C. A., Kursun, O., Ge, X., Busan, S., Dokholyan, N. V. and Weeks, K. M. (2014) Single-molecule correlated chemical probing of RNA. *PNAS*, 111, 13858–13863.
22. Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S. and Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*, 14, 75–82.
23. Simon, L. M., Morandi, E., Luganini, A., Gribaudo, G., Martinez-Sobrido, L., Turner, D. H., Oliviero, S. and Incarnato, D. (2019) In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs. *Nucleic Acids Res.*, 47, 7003–7017.
24. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. and Weeks, K. M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, 11, 959–965.
25. Sander, T., Freyss, J., Korff, M. and von and Rufener, C. (2015) DataWarrior: An Open-Source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.*, 55, 460–473.
26. Fessler, A., Garmon, C., Heavey, T., Fowler, A. and Ogle, C. (2017) Water-soluble and UV traceable isatoic anhydride-based reagents for bioconjugation. *Org. Biomol. Chem.*, 15, 9599–9602.
27. Incarnato, D., Morandi, E., Simon, L. M. and Oliviero, S. (2018) RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Res.*, 46, e97.

28. Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, 9, 357–359.
29. Lorenz, R., Bernhart, S. H., Siederdisen, C. H. Z., Tafer, H., Flamm, C., Stadler, P. F. and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorith. Mol. Biol.: AMB*, 6, 26.
30. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D’Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M. et al. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3, 2.
31. Consortium, T. R. N. A., Sweeney, B. A., Petrov, A. I., Burkov, B., Finn, R. D., Bateman, A., Szymanski, M., Karlowski, W. M., Gorodkin, J., Seemann, S. E. et al. (2018) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.*, 47, D221–D229.
32. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505, 701–705.
33. Giannetti, C. A., Busan, S., Weidmann, C. A. and Weeks, K. M. (2019) SHAPE probing reveals human rRNAs are largely unfolded in solution. *Biochemistry-US*, 58, 3377–3385.
34. Zhou, K. I., Clark, W. C., Pan, D. W., Eckwahl, M. J., Dai, Q. and Pan, T. (2018) Pseudouridines have context-dependent mutation and stop rates in high-throughput sequencing. *Rna Biol*, 15, 892–900.
35. Eddy, S. R. (2014) Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. *Biophysics*, 43, 433–456.
36. Cocozaki, A. I., Altman, R. B., Huang, J., Buurman, E. T., Kazmirski, S. L., Doig, P., Prince, D. B., Blanchard, S. C., Cate, J. H. D. and Ferguson, A. D. (2016) Resistance mutations generate divergent antibiotic susceptibility profiles against translation inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, 113, 8188–8193.
37. Reuter, J. S. and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11, 129.
38. Richter, M. F., Drown, B. S., Riley, A. P., Garcia, A., Shirai, T., Svec, R. L. and Hergenrother, P. J. (2017) Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature*, 545, 299–304.

## 4.7 APPENDIX III: SUPPLEMENTARY INFORMATION

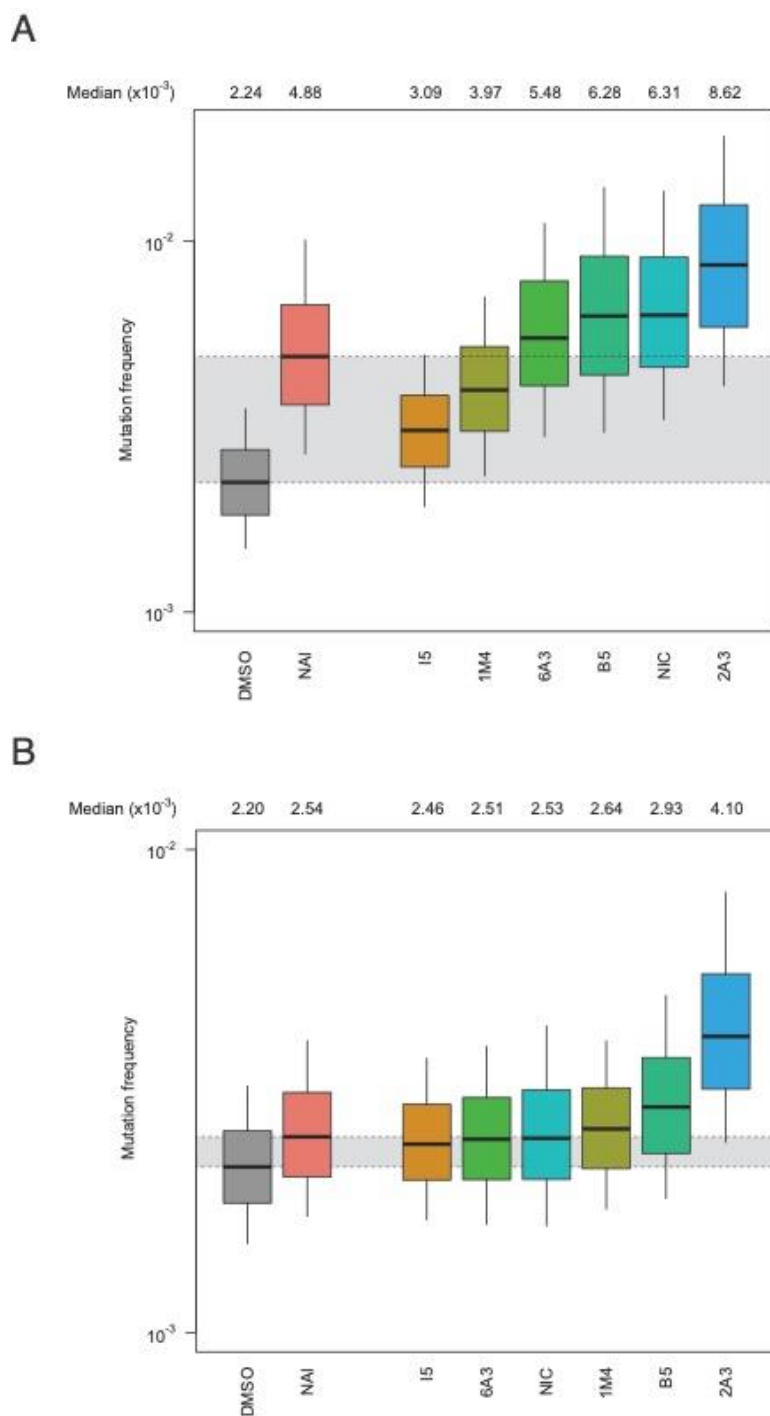


**Figure S1.** (A) Boxplot of SHAPE-MaP mutation frequencies for *E. coli* 16S and 23S rRNAs probed *ex vivo* after deproteinization. Box plots span the interquartile range (from Q1 to Q3). (B) ROC curve for all tested reverse transcriptases, calculated with respect to the accepted phylogenetically inferred 16S and 23S rRNA structures from CRW. The inset reports the area under the curve (AUC) for each condition. (C) Sample of SHAPE-MaP mutation frequencies for all tested conditions across a region spanning nucleotides 567 to 884 of *E. coli* 16S rRNA. Colored bases are those whose mutation frequencies exceed by 2-folds the median mutation frequency in the analyzed region. The accepted structure is reported as an arc plot.

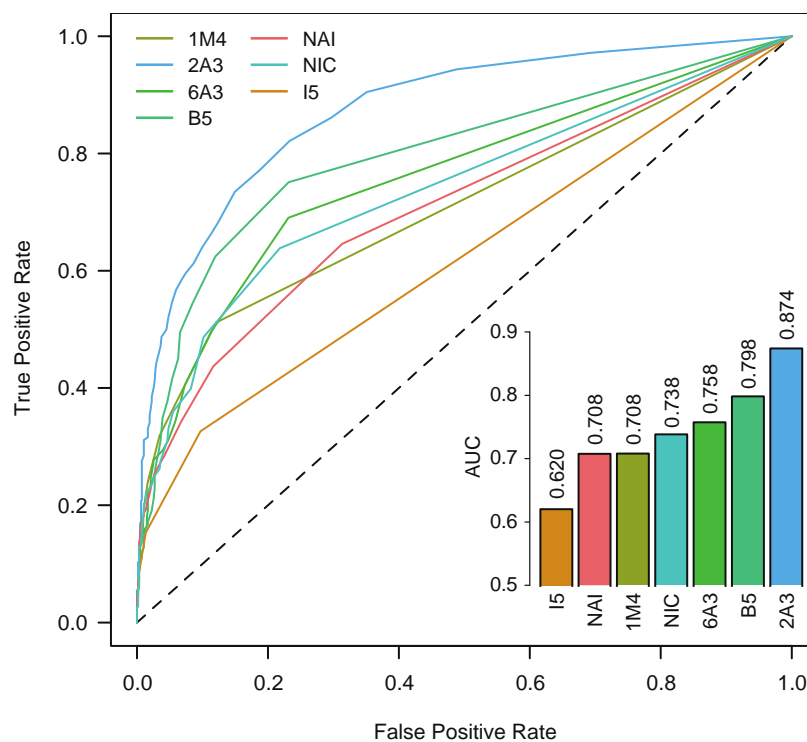


**Figure S2.** (A) ROC curve for all tested SHAPE reagents, calculated with respect to the accepted phylogenetically inferred *E. coli* 16S and 23S rRNA structures from CRW. The inset reports the AUC for each compound. (B) Median SHAPE-MaP mutation frequencies across all stem-loops in the accepted 16S and 23S rRNA structures. Bases are numbered relative to the loop. Positions -3 to -1 correspond to stem bases, while positions 0 and +1 correspond to loop bases. (C) Ratio between the median loop mutation frequency and the median stem mutation frequency, calculated on all stem-loops from B.

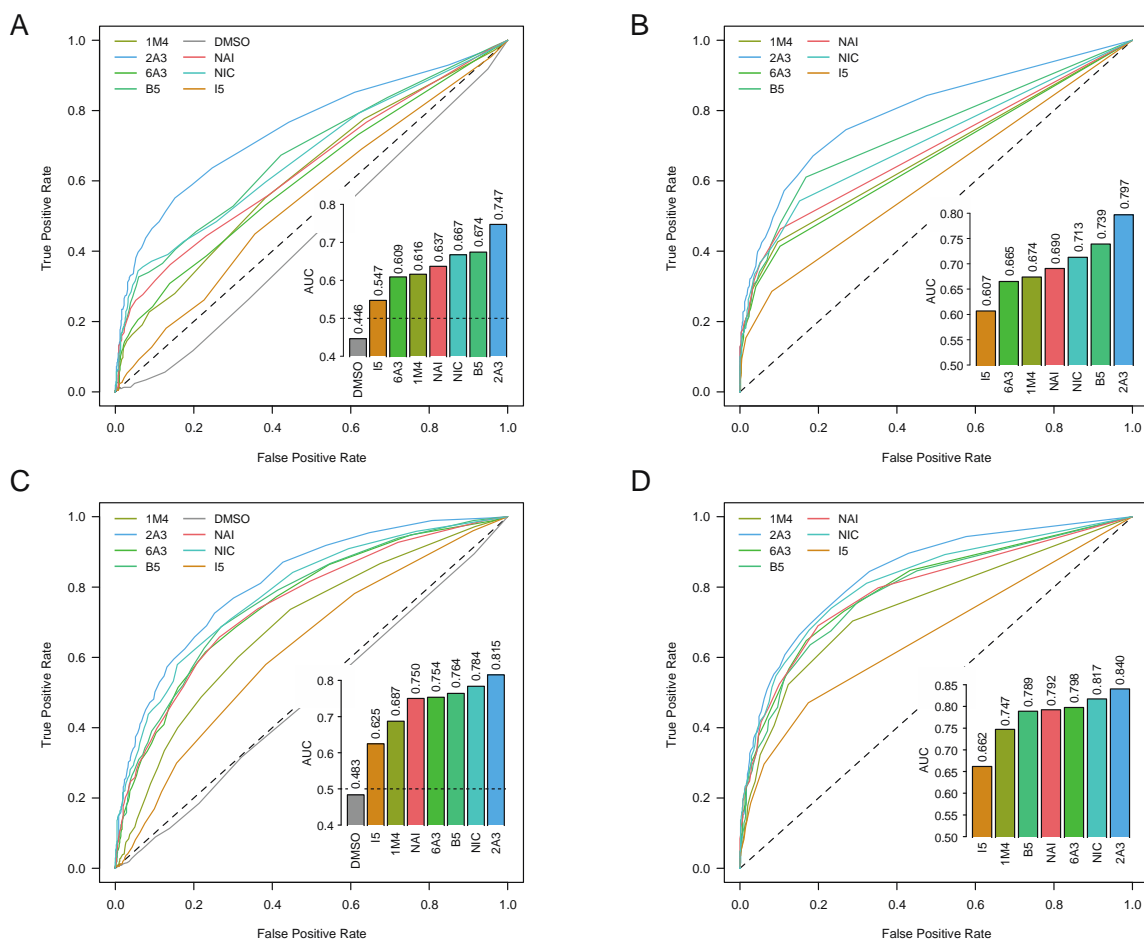




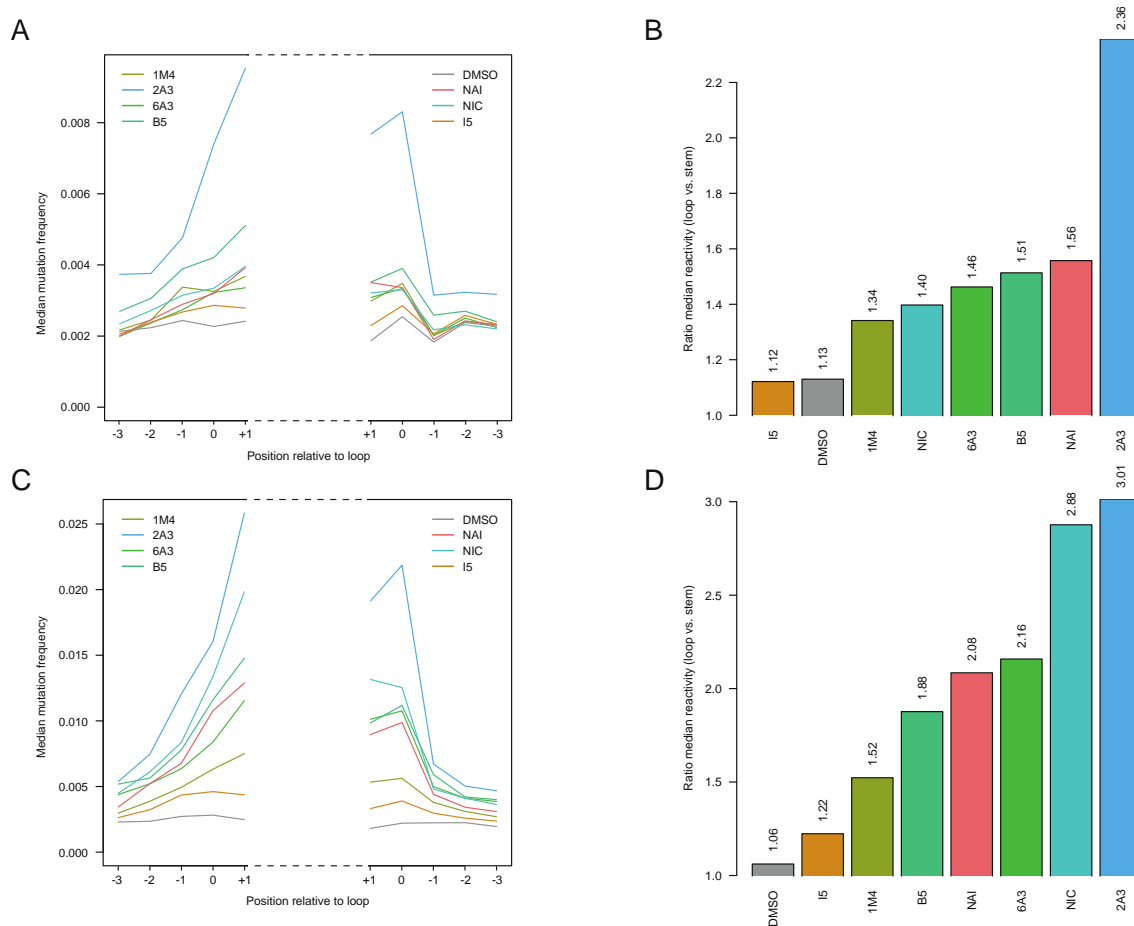
**Figure S3.** (A) Boxplot of SHAPE–MaP mutation frequencies for *H. sapiens* 18S and 28S rRNAs probed *in vivo*. Box plots span the interquartile range (from Q1 to Q3). The grey area spans from the median in the DMSO sample (control) to the median in the NAI sample (reference). (B) Boxplot of SHAPE–MaP mutation frequencies for *B. subtilis* 16S and 23S rRNAs probed *in vivo*. Box plots span the interquartile range (from Q1 to Q3). The grey area spans from the median in the DMSO sample (control) to the median in the NAI sample (reference).



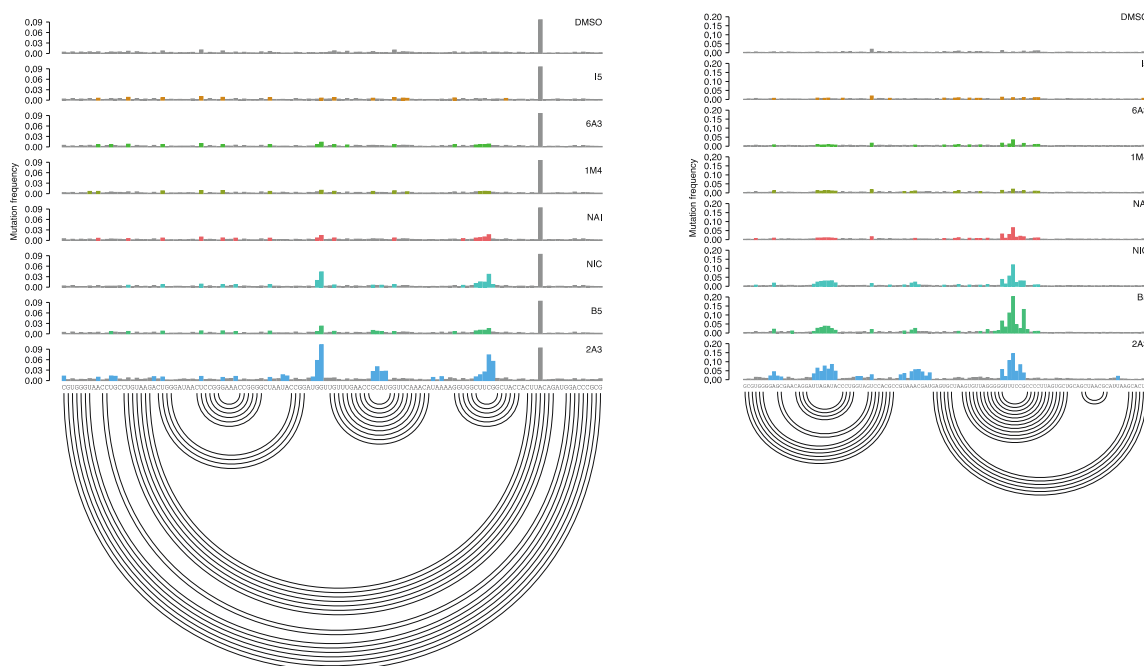
**Figure S4.** ROC curve for *in vivo* probing with all tested SHAPE reagents after subtracting DMSO mutation frequencies, calculated with respect to the accepted phylogenetically inferred *E. coli* 16S and 23S rRNA structures from CRW. The inset reports the AUC for each compound.



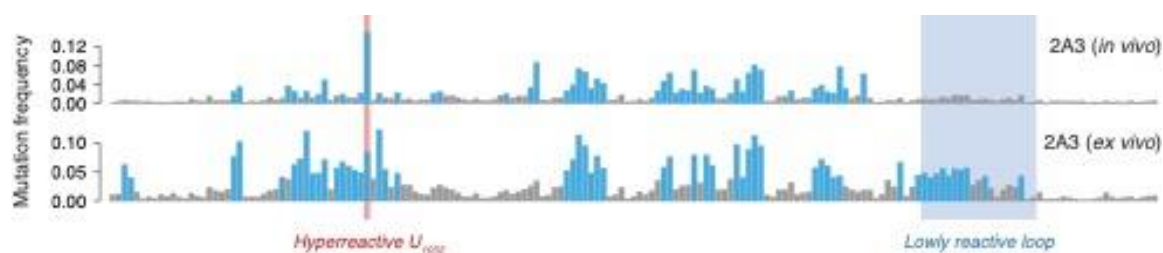
**Figure S5.** (A) and (B). ROC curve for *in vivo* probing with all tested SHAPE reagents before and after subtracting DMSO mutation frequencies, calculated on solvent-exposed residues in the crystal structure of the *B. subtilis* ribosome (PDB: 6HA1), with respect to the accepted phylogenetically inferred *B. subtilis* 16S and 23S rRNA structures from CRW. (C) and (D). ROC curve for *in vivo* probing with all tested SHAPE reagents before and after subtracting DMSO mutation frequencies, calculated on solvent-exposed residues in the crystal structure of the *H. sapiens* ribosome (PDB: 4UG0), with respect to the accepted phylogenetically inferred *H. sapiens* 18S and 28S rRNA structures from RNACentral. The insets report the AUC for each compound.



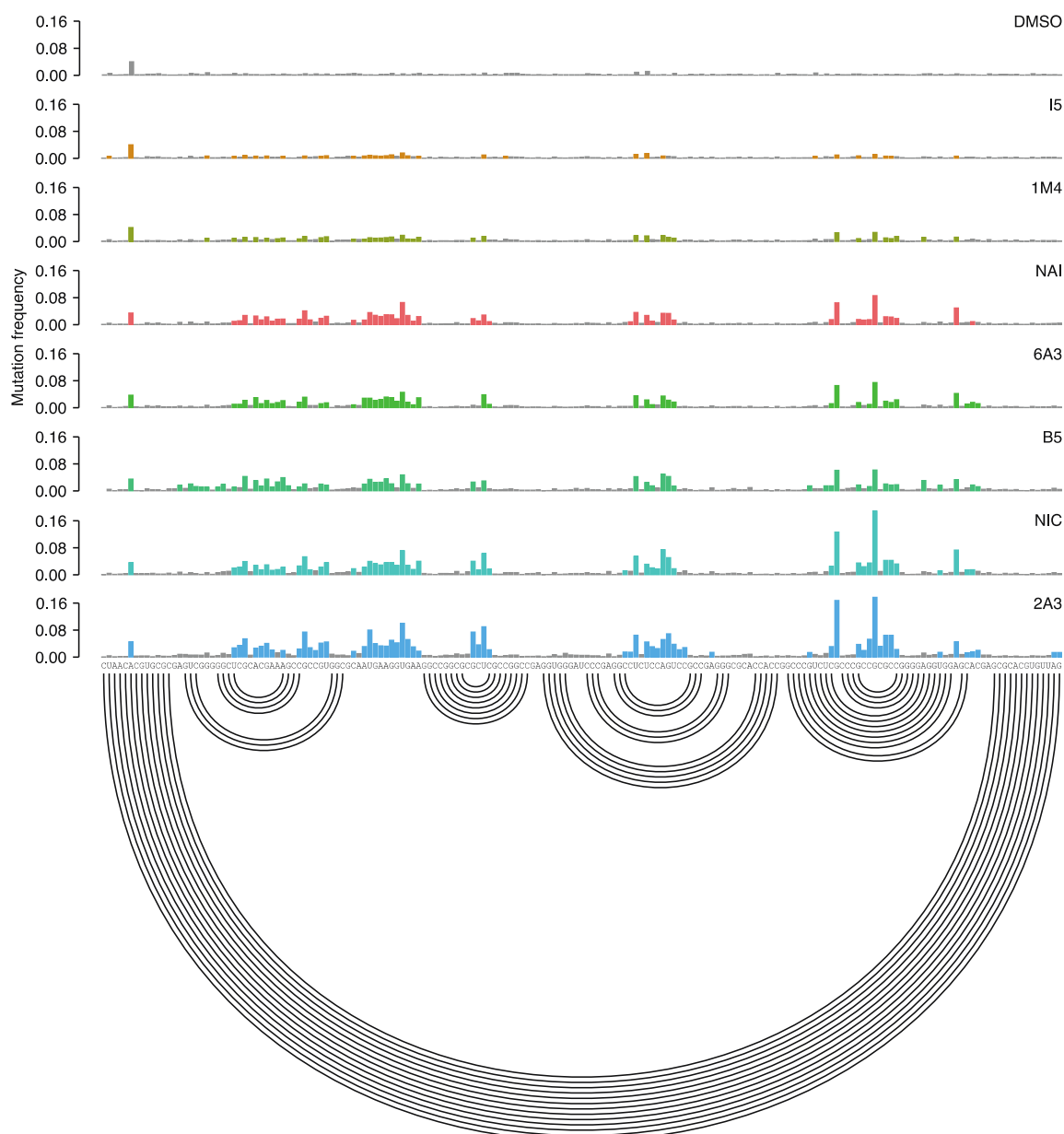
**Figure S6.** (A) Median *in vivo* SHAPE-MaP mutation frequencies across all stem-loops in the accepted *B. subtilis* 16S and 23S rRNA structures. Bases are numbered relative to the loop. Positions -3 to -1 correspond to stem bases, while positions 0 and +1 correspond to loop bases. (B) Ratio between the median loop mutation frequency and the median stem mutation frequency, calculated on all stem-loops from A. (C) Median *in vivo* SHAPE-MaP mutation frequencies across all stem-loops in the accepted *H. sapiens* 18S and 28S rRNA structures. Bases are numbered relative to the loop. Positions -3 to -1 correspond to stem bases, while positions 0 and +1 correspond to loop bases. (D) Ratio between the median loop mutation frequency and the median stem mutation frequency, calculated on all stem-loops from C.



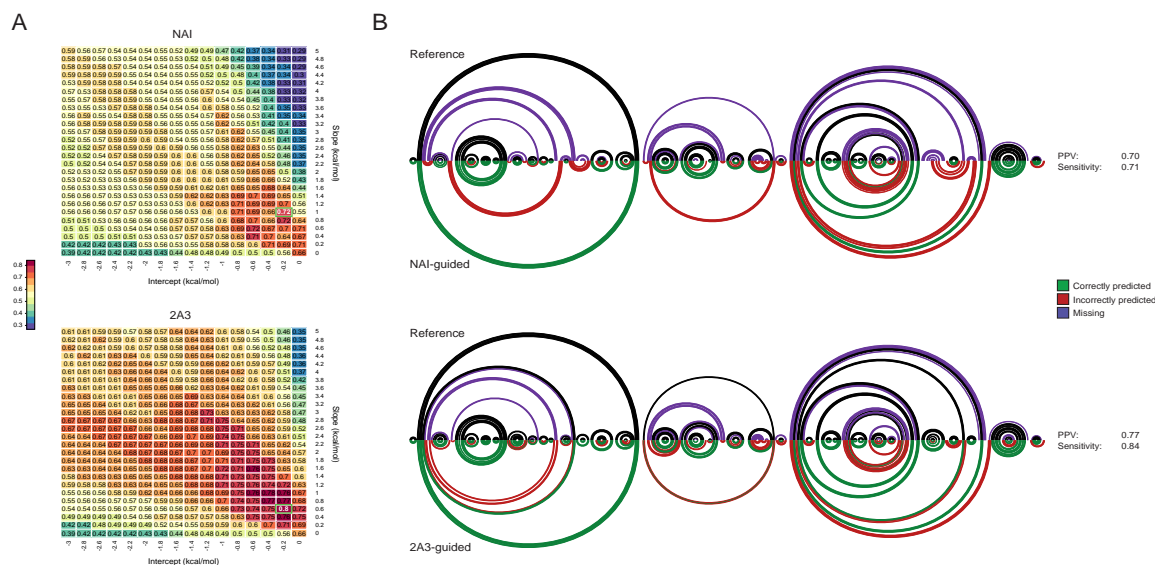
**Figure S7.** Sample of *in vivo* SHAPE–MaP mutation frequencies for all tested compounds across two regions, respectively spanning nucleotides 120 to 245 and nucleotides 777 to 888 of *B. subtilis* 16S rRNA. Colored bases are those whose mutation frequencies exceed by 2-folds the median mutation frequency in the analyzed region. The accepted structure is reported as an arc plot.



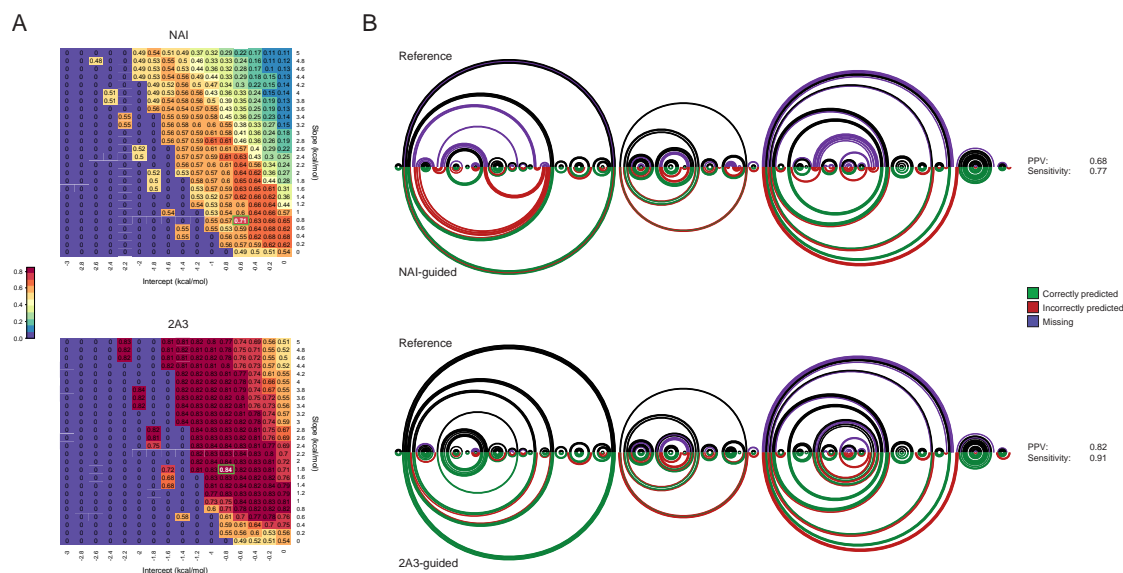
**Figure S8.** SHAPE–MaP mutation frequencies for 2A3 *ex vivo* and *in vivo* across a region spanning nucleotides 990 to 1162 of *E. coli* 23S rRNA (relative to Figure 3A). Colored bases are those whose mutation frequencies exceed by 2-folds the median mutation frequency in the analyzed region. The hyperreactive residue U<sub>1032</sub> and the unreactive stretch of 19 nucleotides are respectively marked in red and blue.



**Figure S9.** Sample of *in vivo* SHAPE–MaP mutation frequencies for all tested compounds across a region spanning nucleotides 1340 to 1516 of *H. sapiens* 28S rRNA. Colored bases are those whose mutation frequencies exceed by 2-folds the median mutation frequency in the analyzed region. The accepted structure is reported as an arc plot.

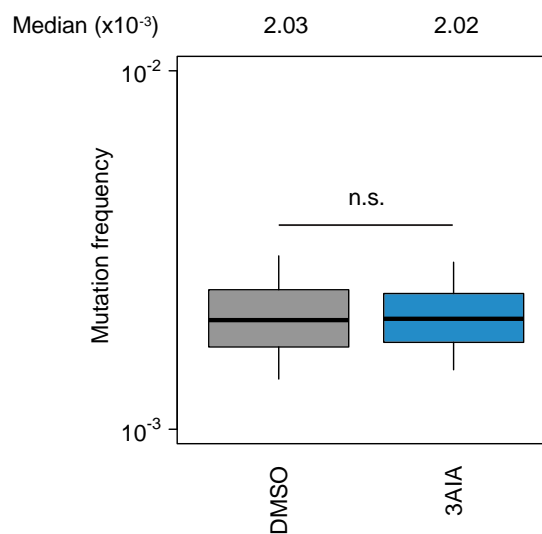


**Figure S10.** (A) Grid search (jackknifing) of optimal slope/intercept value pairs for *B. subtilis* 16S and 23S rRNAs *in vivo* probing data for NAI and 2A3, performed using ViennaRNA 2.0. Values represent the geometric mean of sensitivity and PPV for the secondary structures predicted using each slope/intercept value pair. The chosen value pair is boxed in green. (B) Arc plot comparison of *B. subtilis* 16S rRNA reference structure (top), and structure inferred using either NAI-derived or 2A3-derived restraints (bottom). Black/green arcs correspond to correctly predicted base-pairs, violet arcs to non-predicted base-pairs and red arcs to mispredicted base-pairs. PPV and sensitivity for each prediction are indicated.



**Figure S11.** (A) Grid search (jackknifing) of optimal slope/intercept value pairs for *E. coli* 16S and 23S rRNAs *in vivo* probing data for NAI and 2A3, performed using RNAstructure. Values represent the geometric mean of sensitivity and PPV for the secondary structures predicted using each slope/intercept value pair. Pairs with geometric mean equal to 0, produced an empty structure (no base-pairs). The chosen value pair is boxed in green. (B) Arc plot comparison of *E. coli* 16S rRNA reference structure (top), and structure inferred using either NAI-derived or 2A3-derived restraints (bottom). Black/green arcs correspond to correctly predicted base-pairs, violet arcs to non-predicted base-pairs and red arcs to mispredicted base-pairs. PPV and sensitivity for each prediction are indicated.

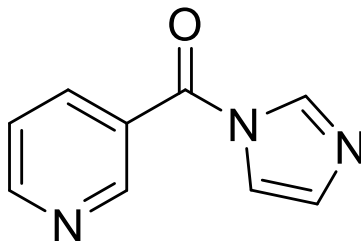




**Figure S12.** Boxplot of SHAPE–MaP mutation frequencies for *E. coli* 16S and 23S rRNAs probed *in vivo*. Box plots span the interquartile range (from Q1 to Q3). 3AIA is unable to probe RNA *in vivo* as shown by absence of any significant difference with respect to the DMSO control ( $P = 0.2$ , Wilcoxon rank sum test).

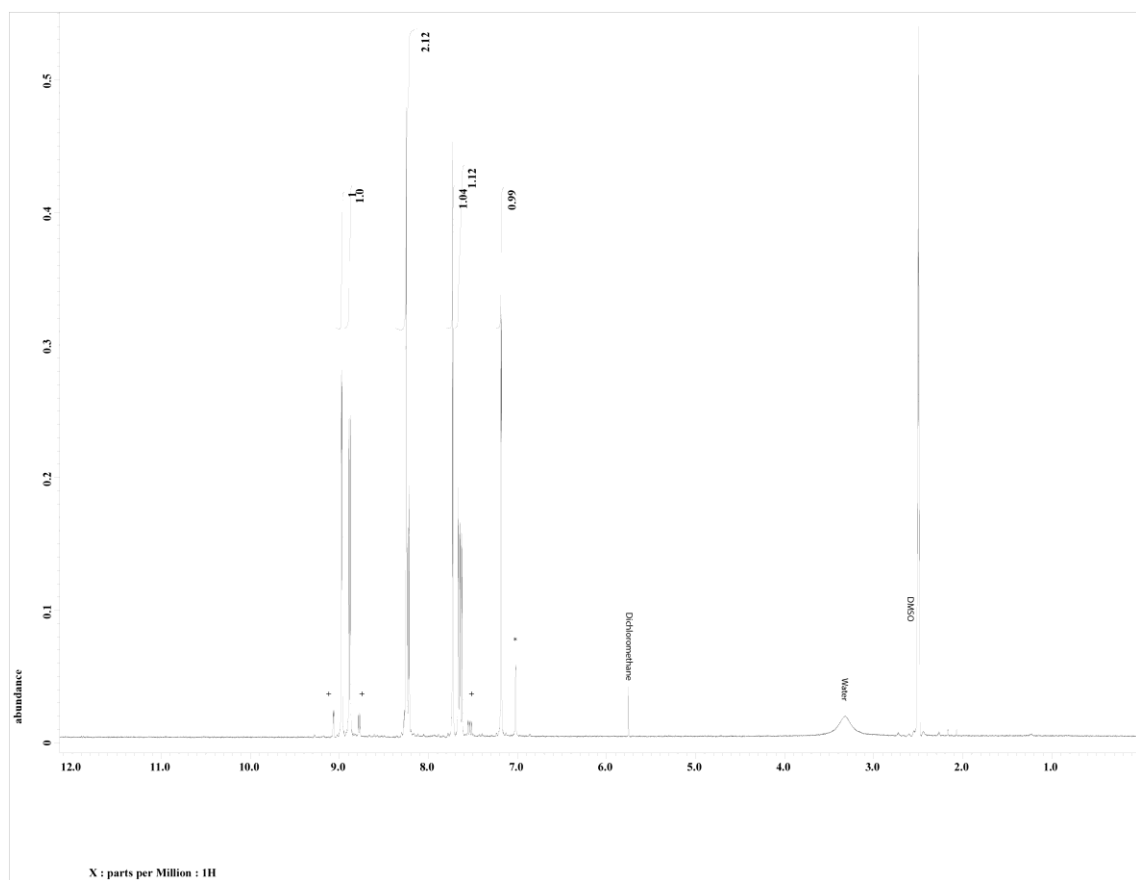
Electrophile	25 °C	37 °C
NIC	30 minutes	8 minutes
2A3	260 minutes	60 minutes
6A3	510 minutes	300 minutes
B5	70 minutes	30 minutes
1M4	900 minutes	290 minutes
3AIA	44 minutes	14 minutes
NAI	300 minutes	120 minutes

**Table S1.** Half-life data for SHAPE electrophiles in pH 7.4 PBS 1X at 25°C and 37°C

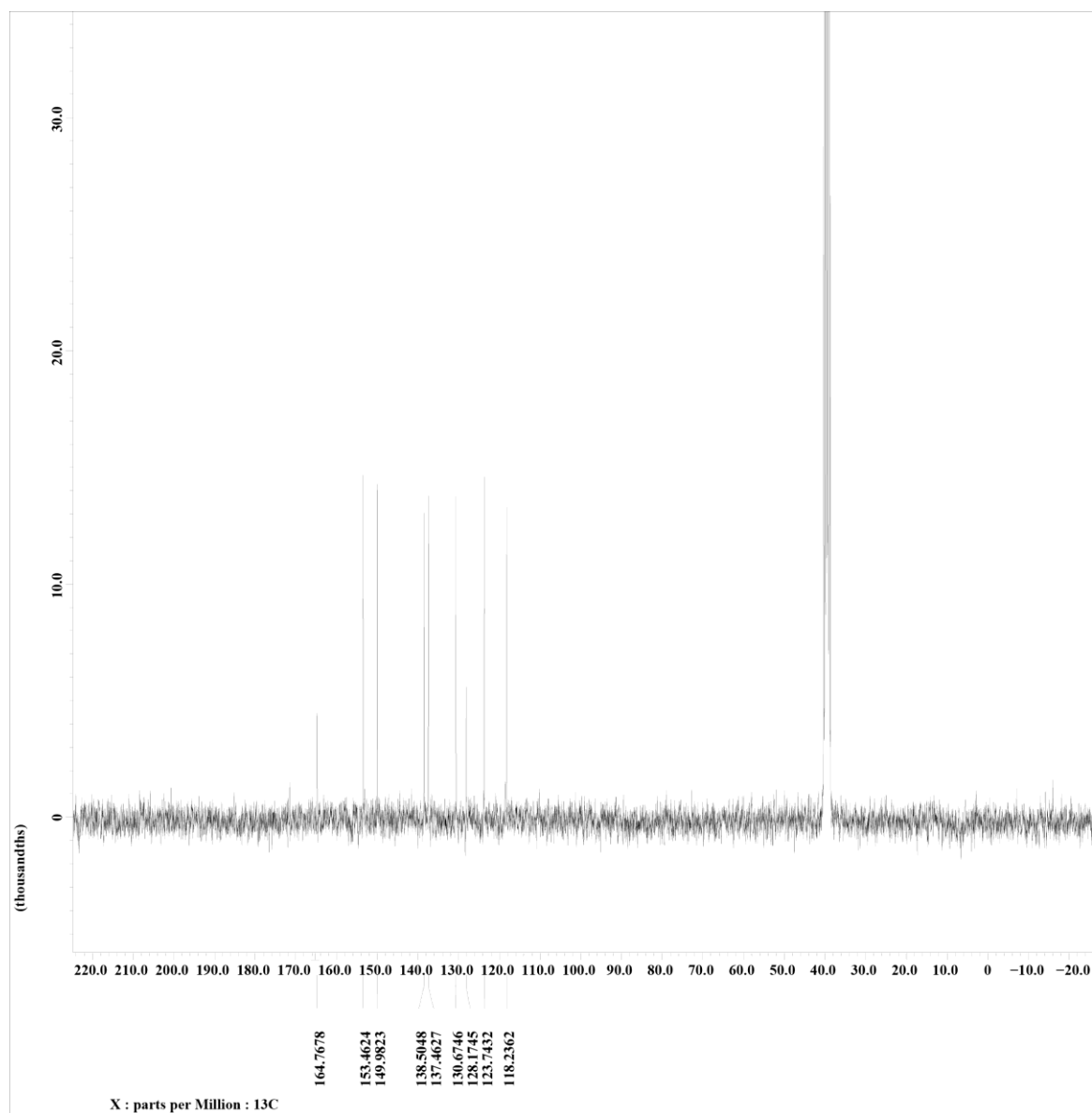
**Nicotinic acid imidazolid (NIC)**

NIC for SHAPE experiment was obtained following the general procedure described in the Materials and Methods section. For the analytical sample, a portion of the crude mixture was diluted into dichloromethane and extracted 3X with saturated sodium bicarbonate solution. The organic layer was dried over  $\text{MgSO}_4$  and concentrated under reduced pressure.

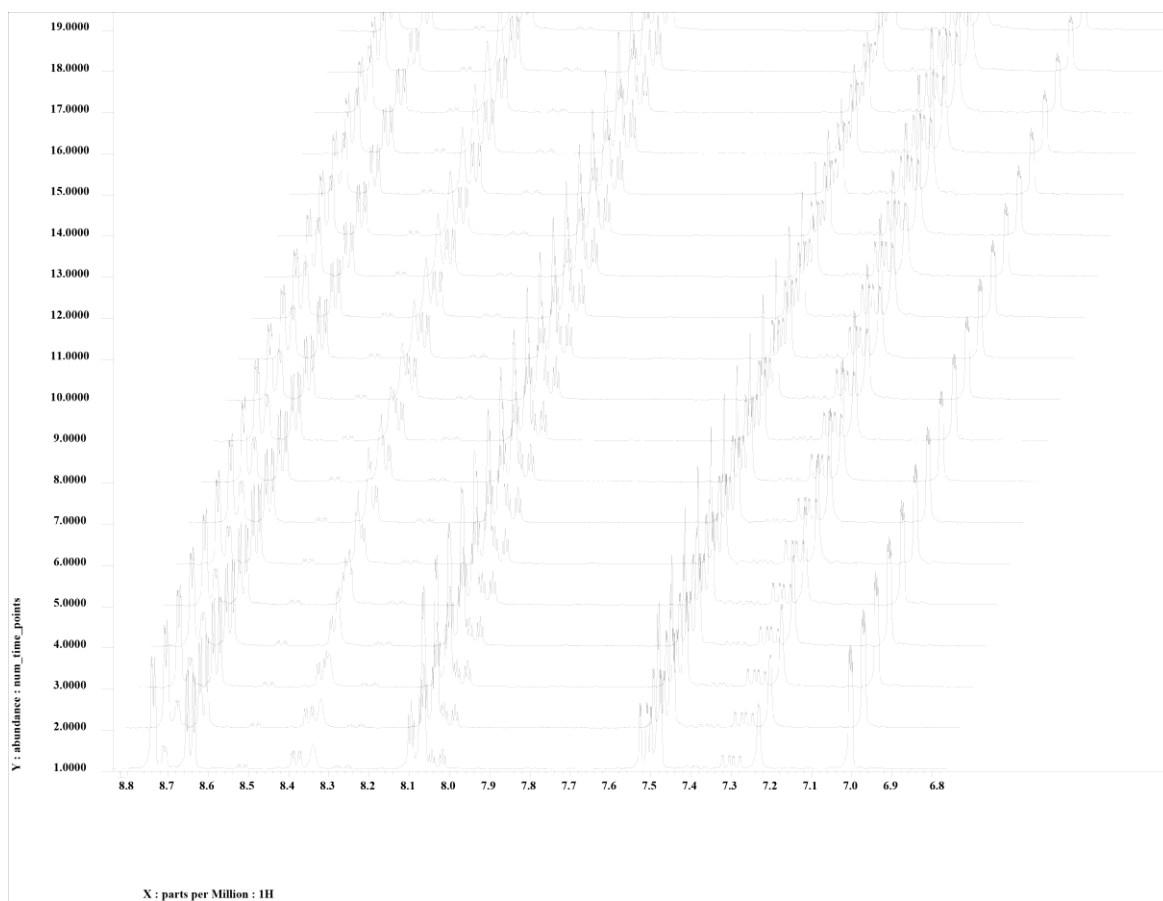
$\text{C}_9\text{H}_7\text{N}_3\text{O}$ ,  $^1\text{H}$  NMR (300 MHz,  $\text{DMSO-d}_6$ ):  $\delta$  7.20 (s, 1H), 7.65 (dd, 1H), 7.73 (s, 1H), 8.22 (dt, 1H), 8.25 (s, 1H), 8.90 (d, 1H), 8.98 (s, 1H) ppm,  $^{13}\text{C}$  NMR (75 MHz,  $\text{DMSO-d}_6$ ):  $\delta$  118.24, 123.74, 128.17, 130.67, 137.46, 138.51, 149.98, 153.46, 164.77 ppm, HRMS: Calc: 173.0589 m/z, Found: Compound hydrolyzes into starting materials.



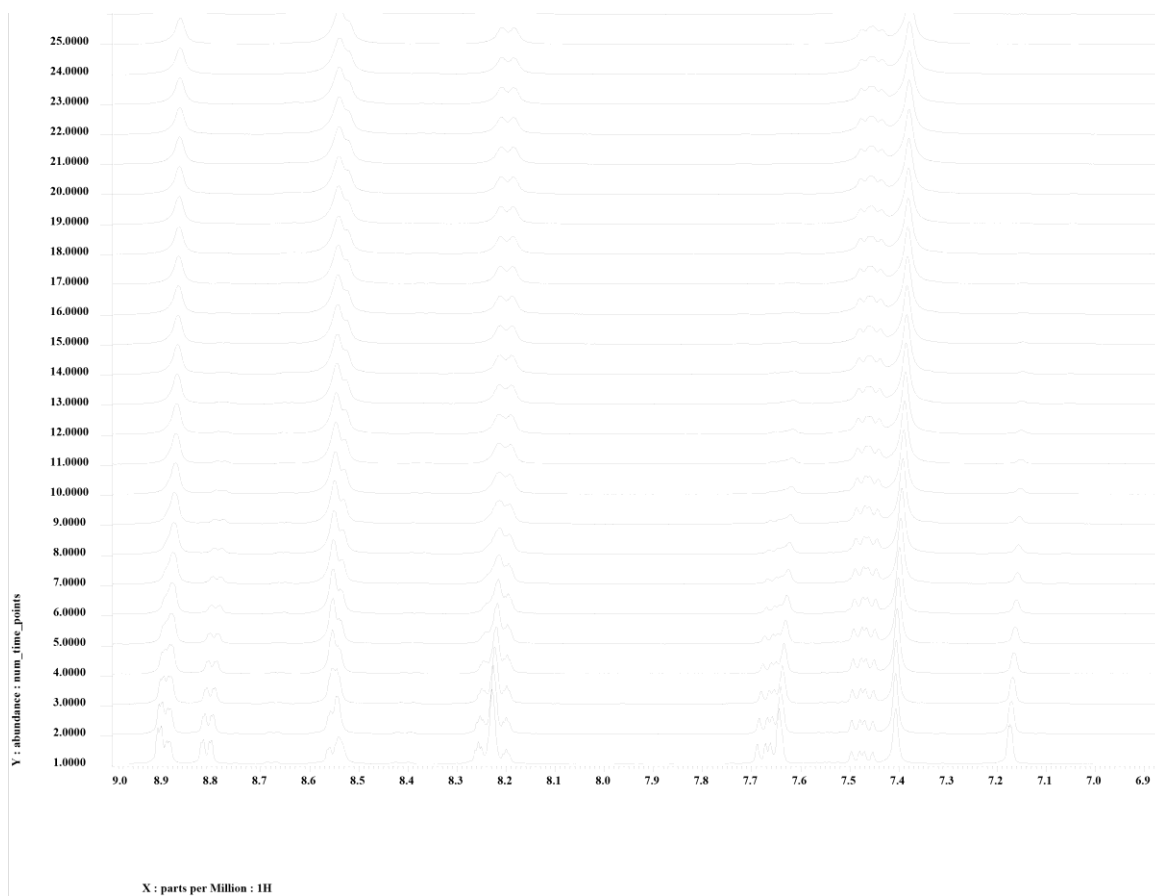
**Figure S13.**  $^1\text{H}$  NMR of NIC in DMSO- $\text{d}_6$ . + signifies hydrolyzed product protons and \* shows imidazole protons.



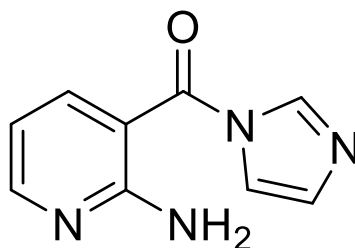
**Figure S14.**  $^{13}\text{C}$ -NMR of NIC in DMSO- $\text{d}_6$ .



**Figure S15.** Hydrolysis experiment of NIC at 25°C. Scan 1 occurred at 5 minutes, each subsequent scan was performed in 2 minutes intervals. Monitoring the disappearance of product peaks at 8.81 and 7.48 ppm.

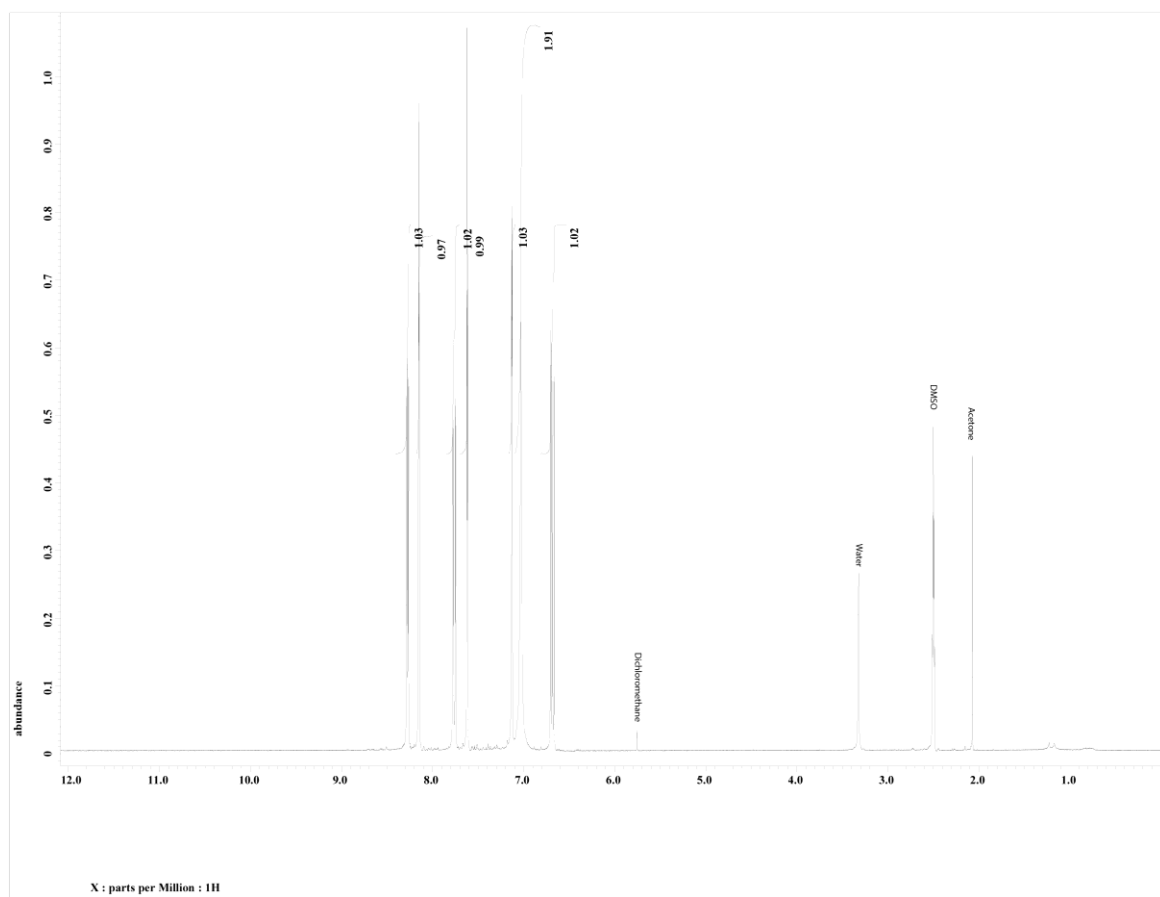


**Figure S16.** Hydrolysis experiment of NIC at 37°C. Scan 1 occurred at 5 minutes, each subsequent scan was performed in 2 minutes intervals. Monitoring the disappearance of product peaks at 8.81 and 7.48 ppm.

**2-aminopyridine-3-carboxylic acid imidazolide (2A3)**

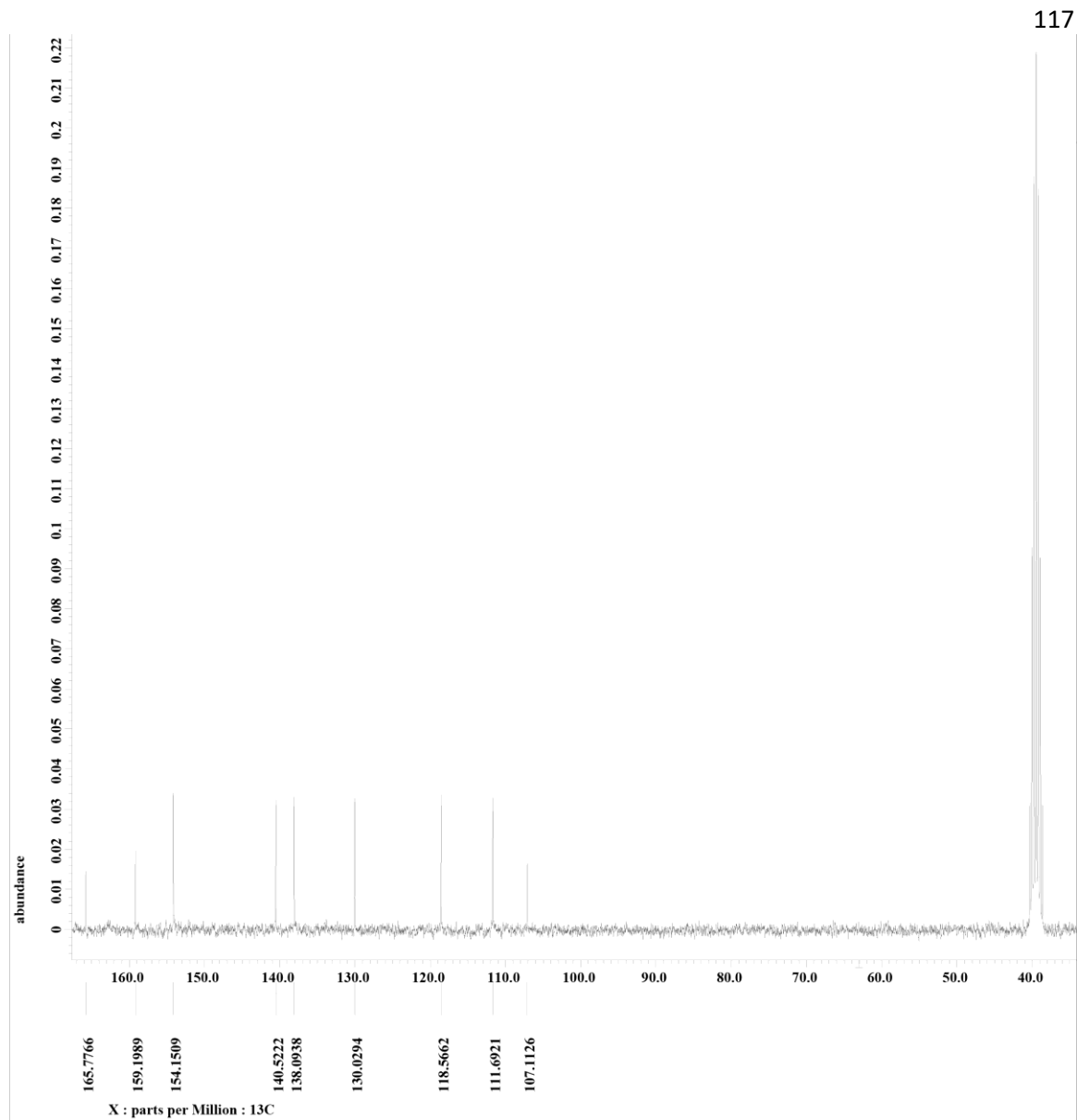
2A3 was obtained following the general procedure described in the Materials and Methods section but using MeCN in place of DMSO. For the analytical sample, a portion of the crude mixture was diluted into dichloromethane and extracted 3X with saturated sodium bicarbonate solution. The organic layer was dried over MgSO<sub>4</sub> and concentrated under reduced pressure.

C<sub>9</sub>H<sub>8</sub>N<sub>4</sub>O, <sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>): δ 6.69 (t, 1H), 7.04 (s, 2H), 7.15 (s, 1H), 7.65 (s, 1H), 7.76 (d, 1H), 8.18 (s, 1H), 8.28 (dd, 1H) ppm, <sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>): δ 107.11, 111.69, 118.57, 130.03, 138.09, 140.52, 154.15, 159.20, 165.78 ppm, HRMS Calc: 188.0698 m/z, Found: 189.0771 (M+H) m/z.

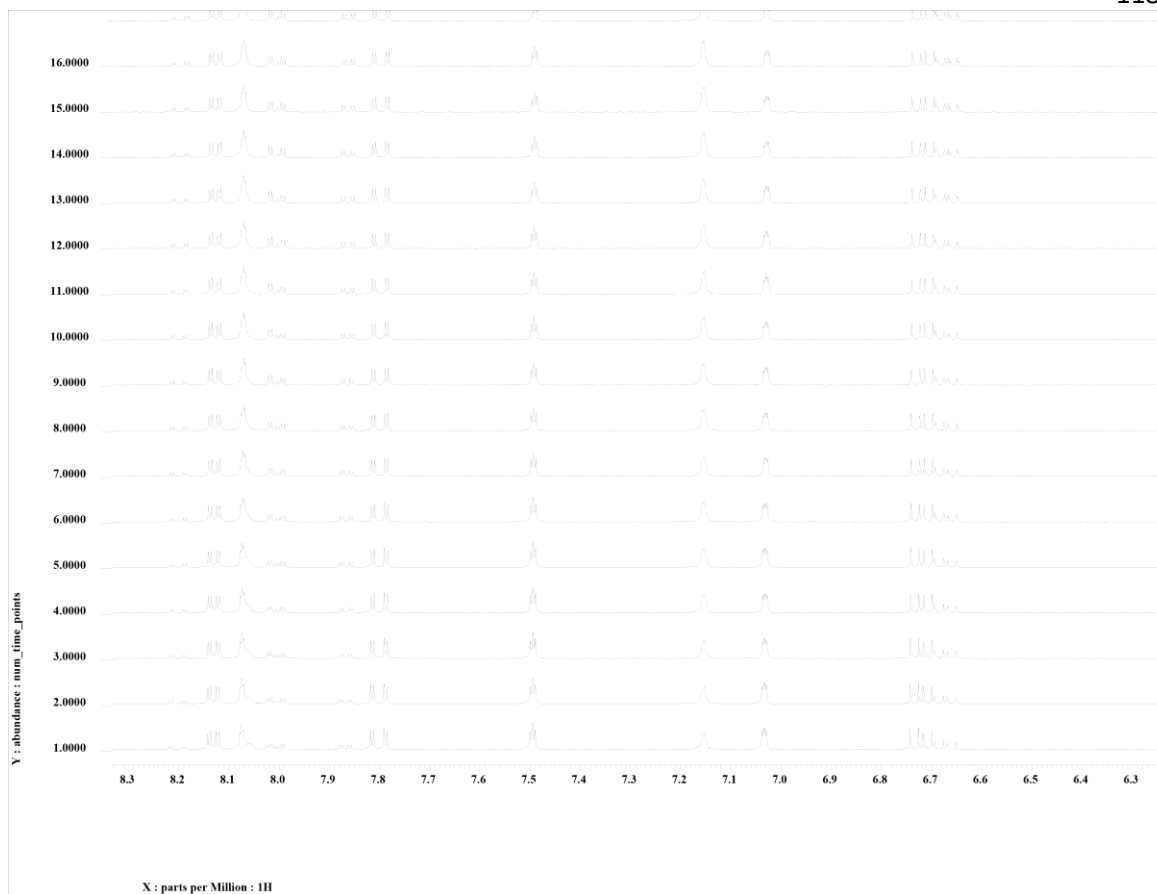


**Figure S17.**  $^1\text{H}$  NMR of 2A3 in DMSO- $d_6$ .

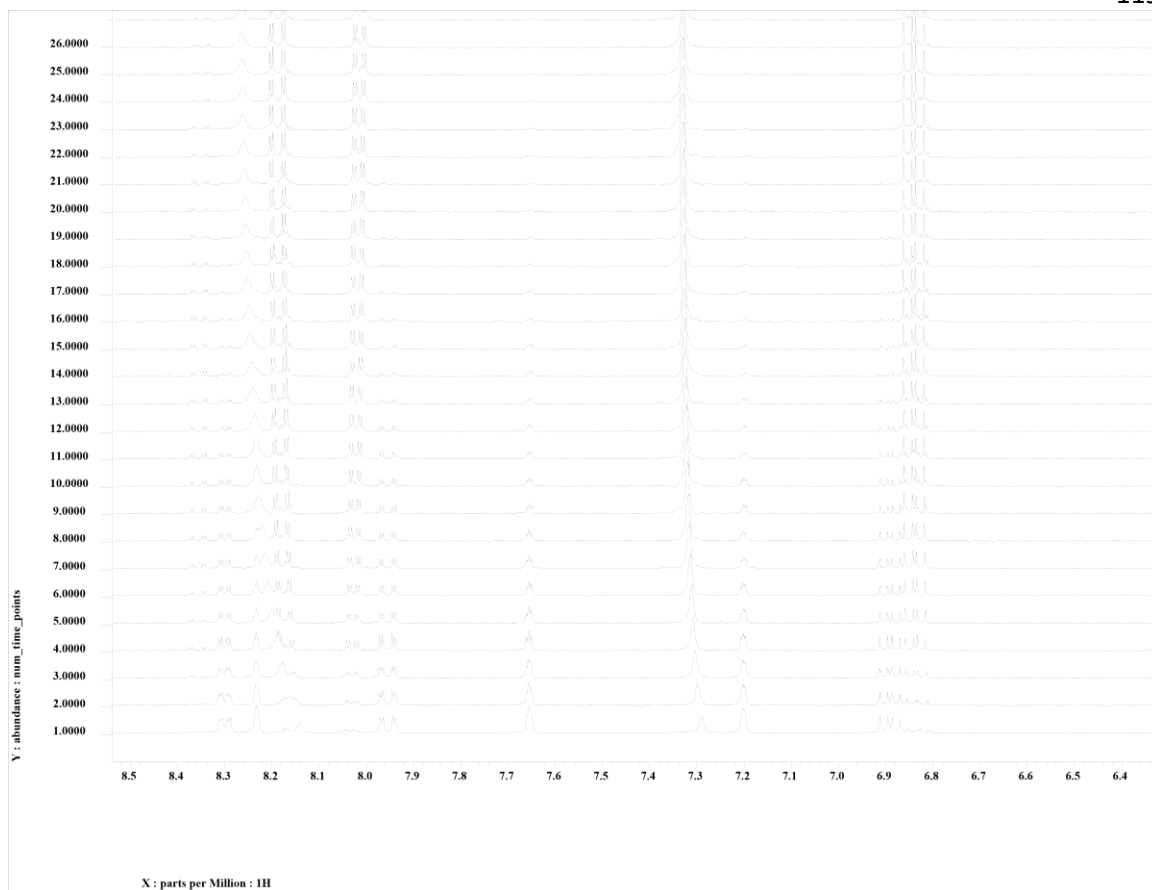




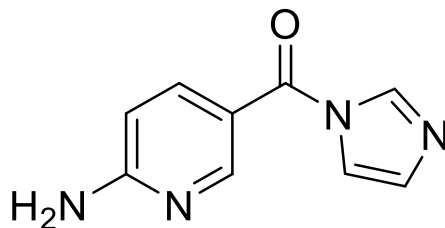
**Figure S18.**  $^{13}\text{C}$ -NMR of 2A3 in DMSO- $d_6$ .



**Figure S19.** Hydrolysis experiment of 2A3 at 25°C. Showing 85 minutes to 185 minutes, where each scan was performed in 5 minutes intervals. Monitoring the disappearance of product peaks at 8.81 and 7.48 ppm.

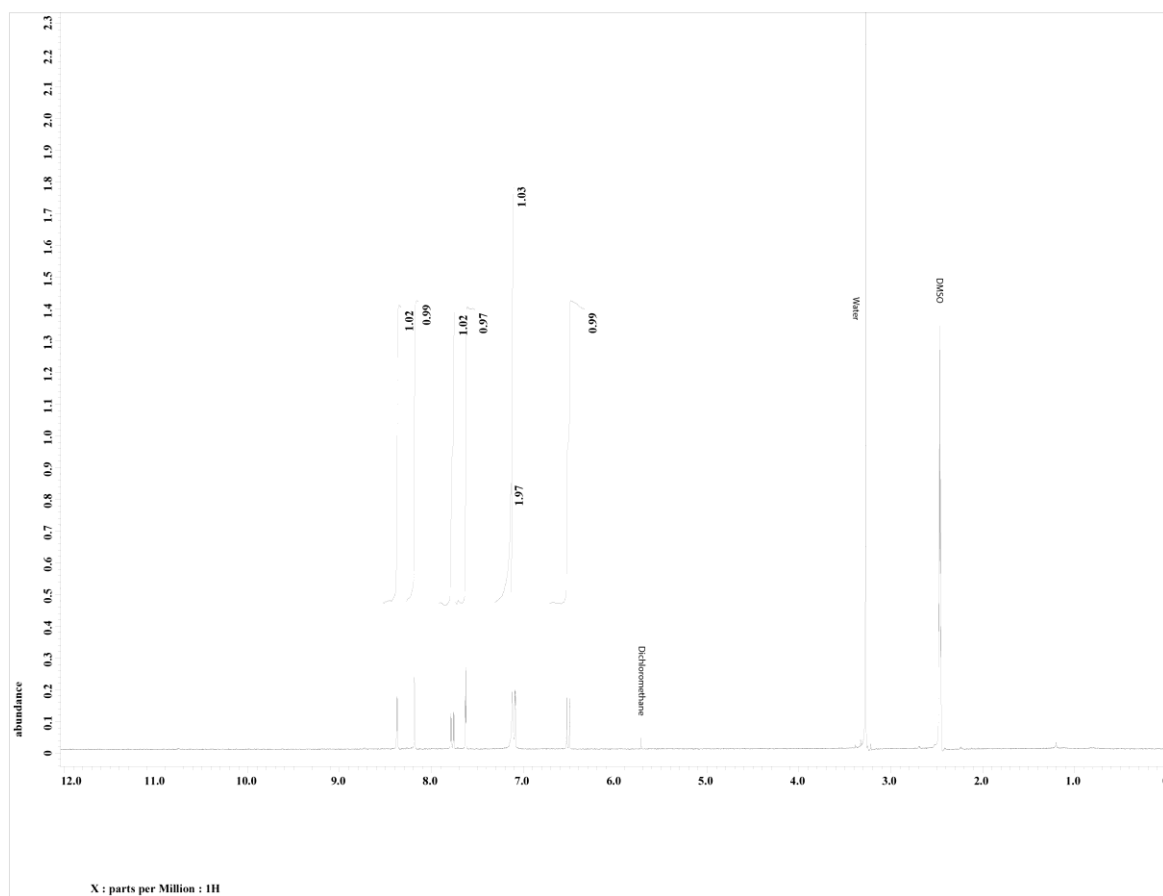


**Figure S20.** Hydrolysis experiment of 2A3 at 37°C. Showing 85 minutes to 185 minutes, where each scan was performed in 5 minutes intervals. Monitoring the disappearance of product peaks at 8.81 and 7.48 ppm.

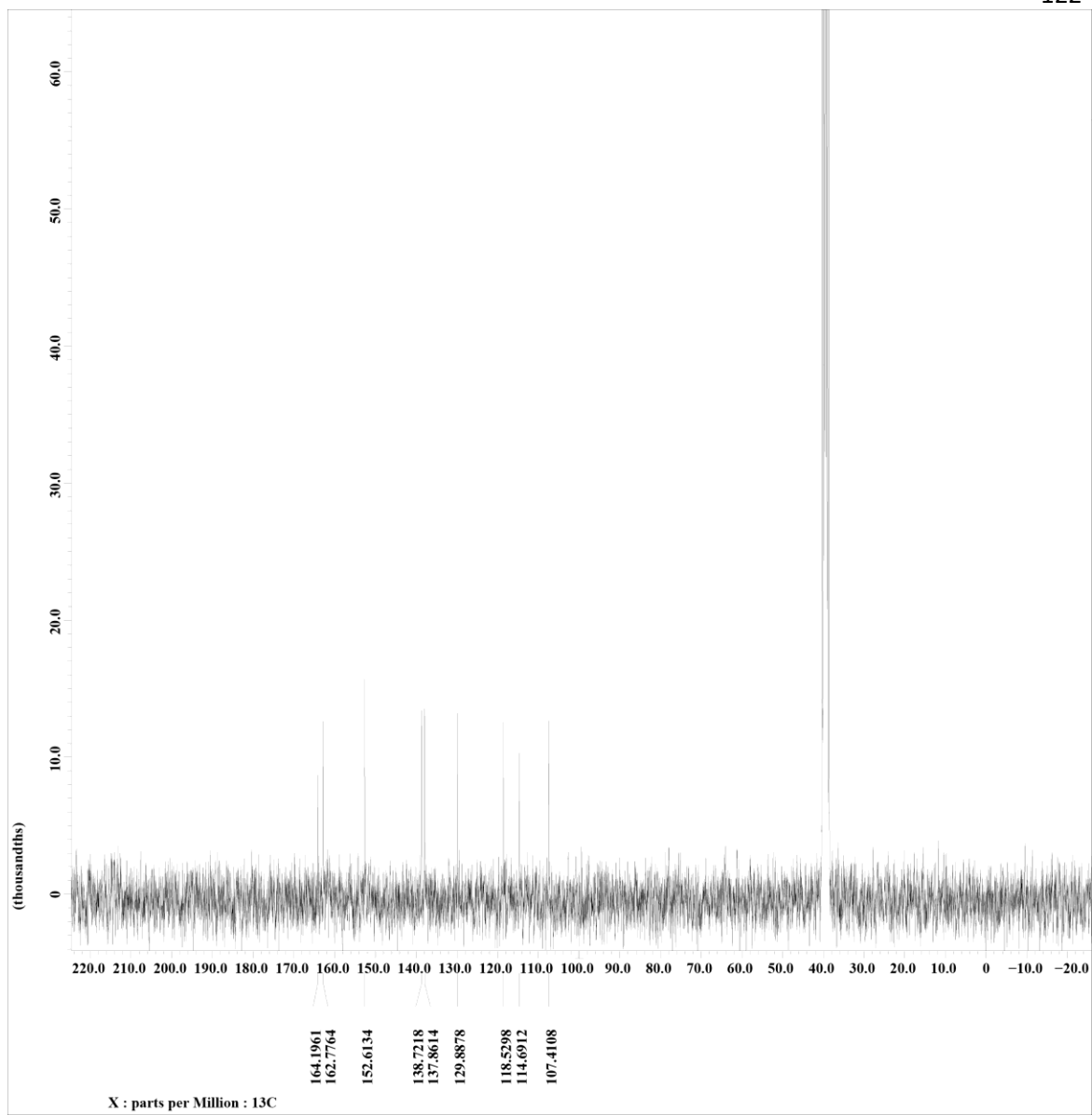
**6-aminopyridine-3-carboxylic acid imidazolid (6A3)**

6A3 was obtained following the general procedure described in the Materials and Methods section. For the analytical sample, a portion of the crude mixture was diluted into dichloromethane and extracted 3X with saturated sodium bicarbonate solution. The organic layer was dried over  $\text{MgSO}_4$  and concentrated under reduced pressure.

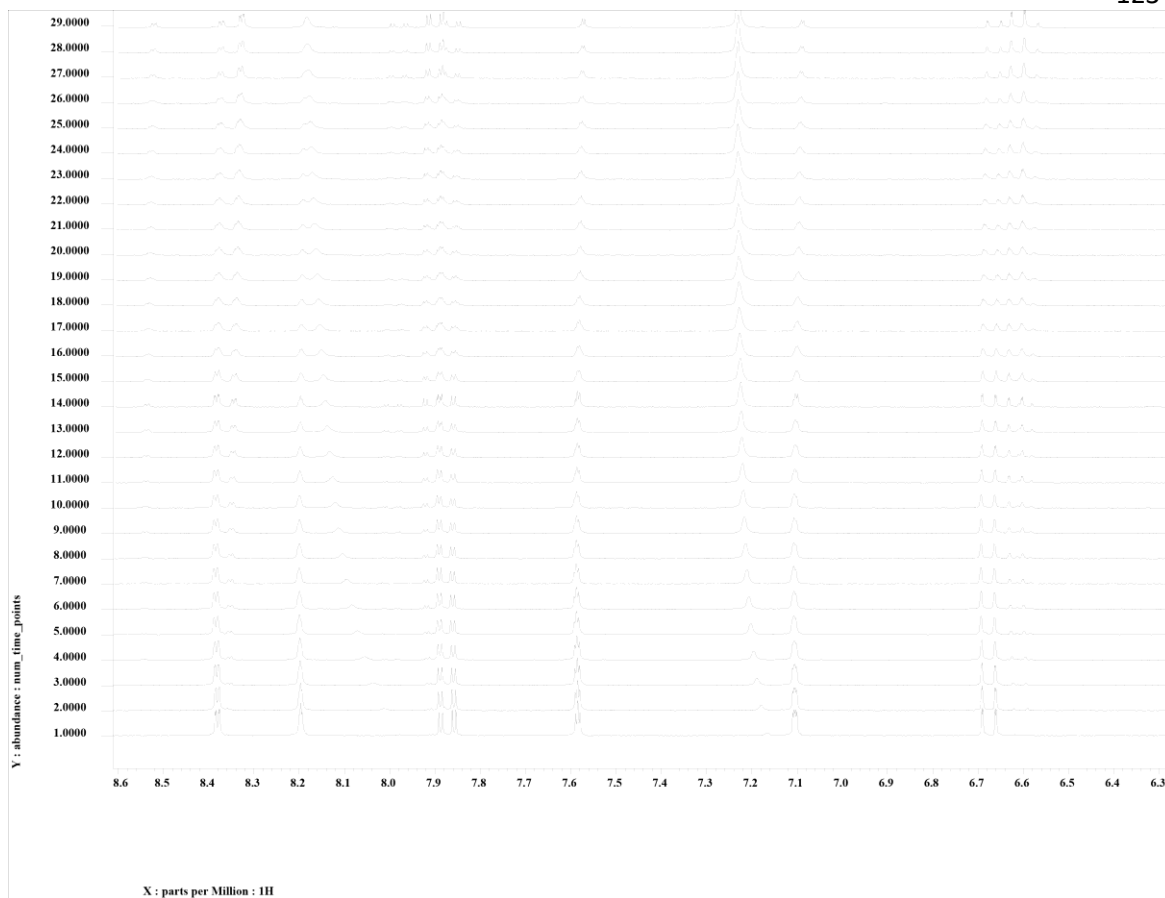
$\text{C}_9\text{H}_8\text{N}_4\text{O}$ ,  $^1\text{H}$  NMR (300 MHz, DMSO- $d_6$ ):  $\delta$  6.55 (d, 1H), 7.12 (s, 1H), 7.16 (s, 2H), 7.78 (s, 1H), 7.82 (dd, 1H), 8.24 (s, 1H), 8.42 (d, 1H) ppm,  $^{13}\text{C}$  NMR (75 MHz, DMSO- $d_6$ ):  $\delta$  107.43, 114.11, 118.55, 129.90, 137.95, 138.73, 152.63, 162.92, 164.21 ppm, HRMS Calc: 188.0698 m/z Found: 189.0771 (M+H) m/z.



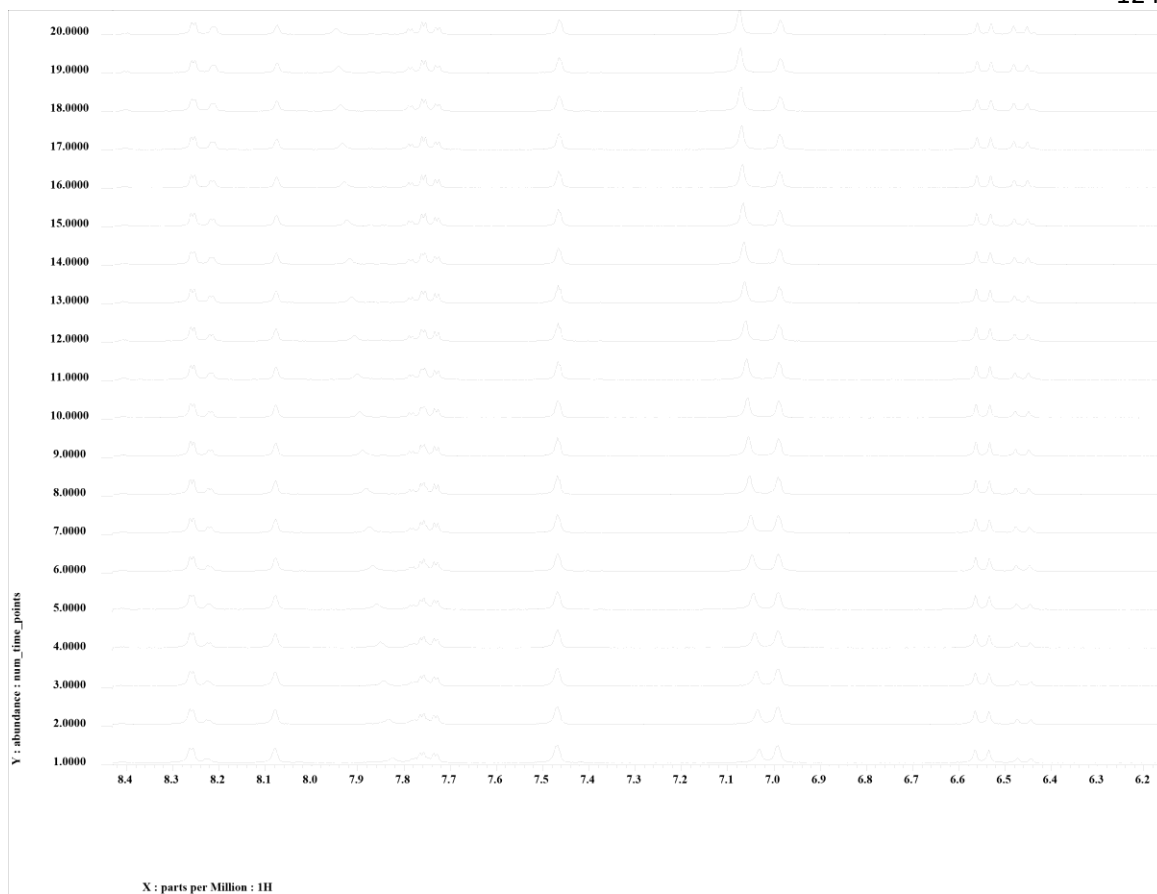
**Figure S21.**  $^1\text{H}$  NMR of 6A3 in DMSO- $\text{d}_6$



**Figure S22.**  $^{13}\text{C}$  NMR of 6A3 in DMSO-d6

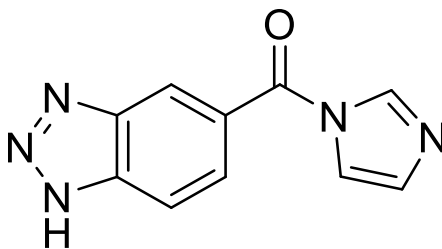


**Figure S23.** Hydrolysis experiment of 6A3 at 25°C. Showing 5 minutes to 905 minutes, where each scan was performed in 30 minutes intervals. Monitoring the disappearance of product peaks at 8.39 ppm and appearance at 8.36 ppm.



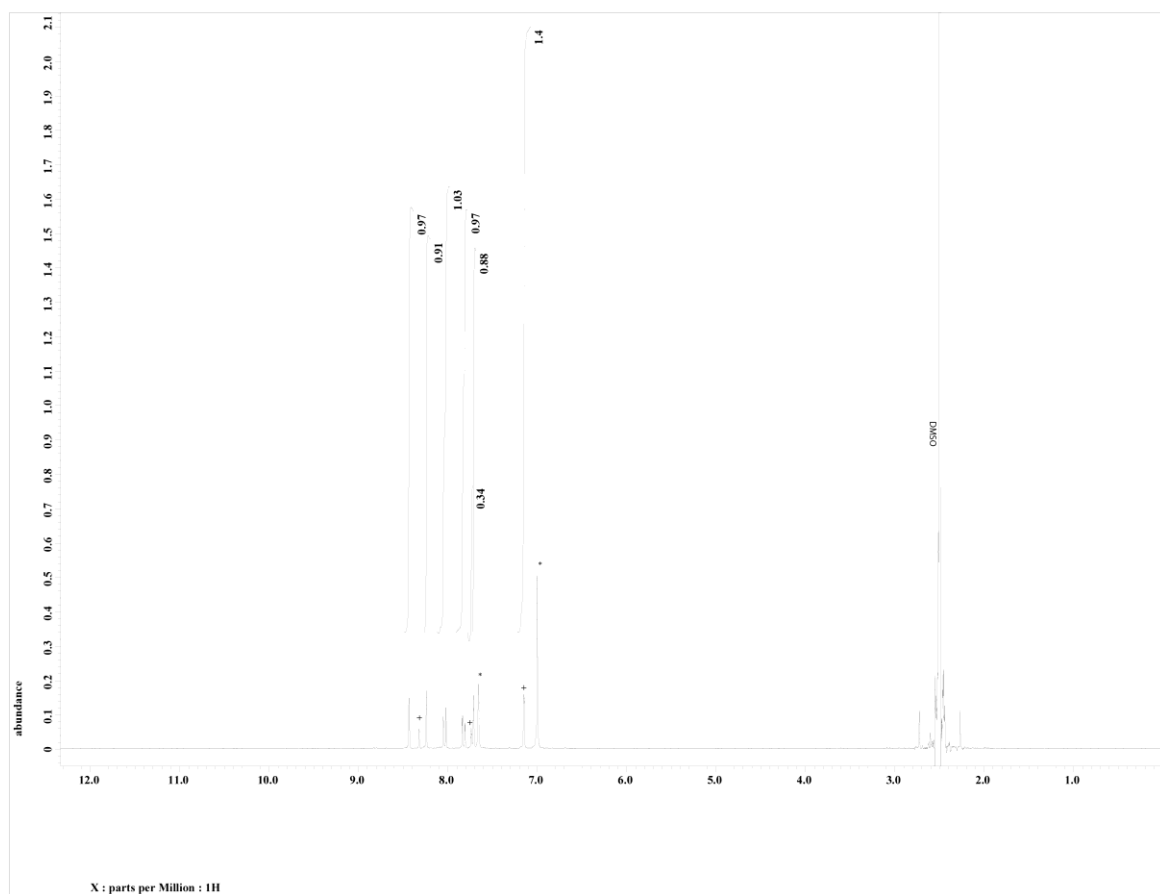
**Figure S24.** Hydrolysis experiment of 6A3 at 37°C. Showing 5 minutes to 905 minutes, where each scan was performed in 30 minutes intervals. Monitoring the disappearance of product peaks at 8.39 ppm and appearance at 8.36 ppm.



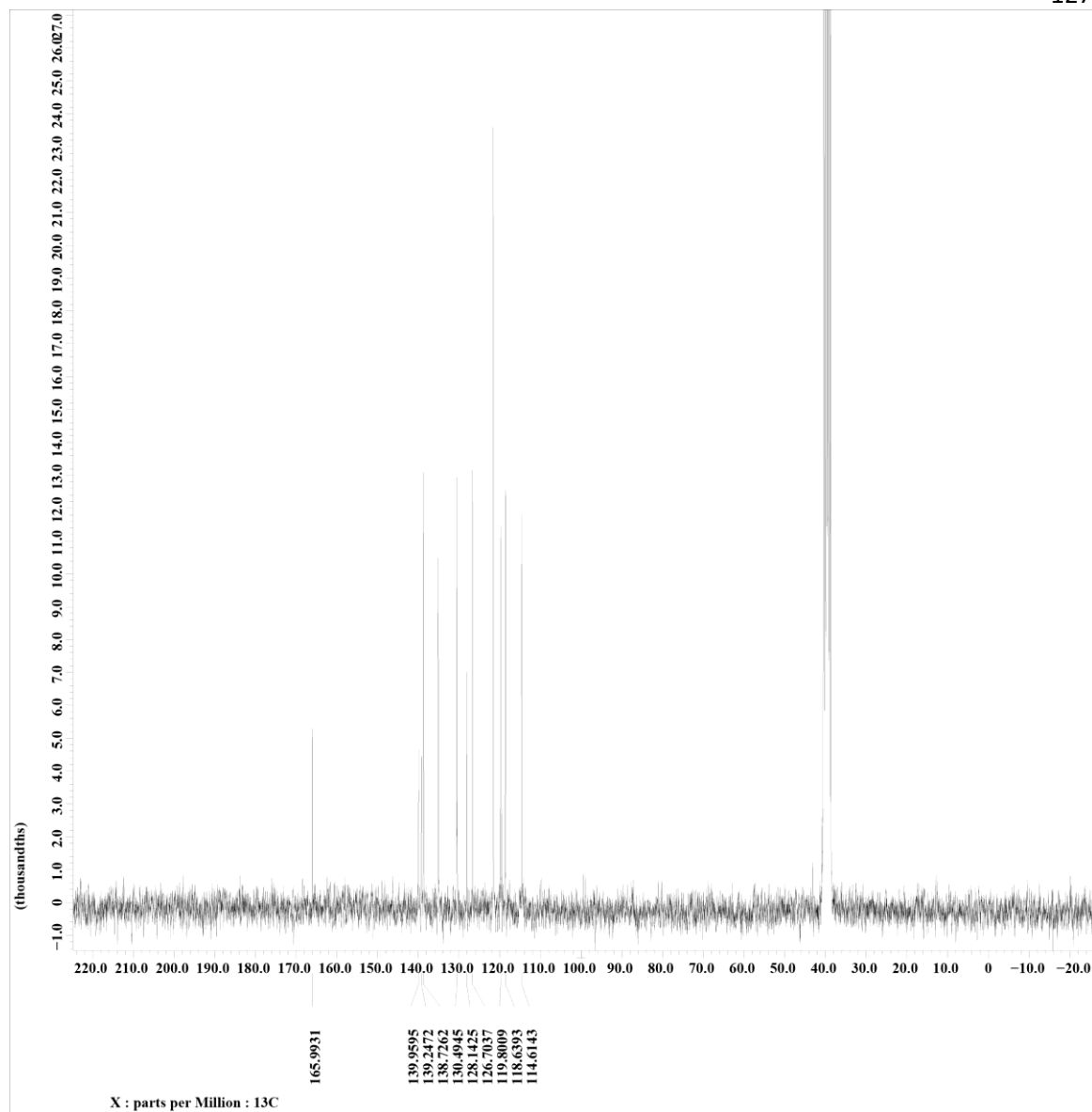
**Benzotriazole-5-carboxylic acid imidazolide (B5)**

B5 was obtained following the general procedure described in the Materials and Methods section. Analytical samples were taken directly from the reaction mixture, compound does not survive extraction.

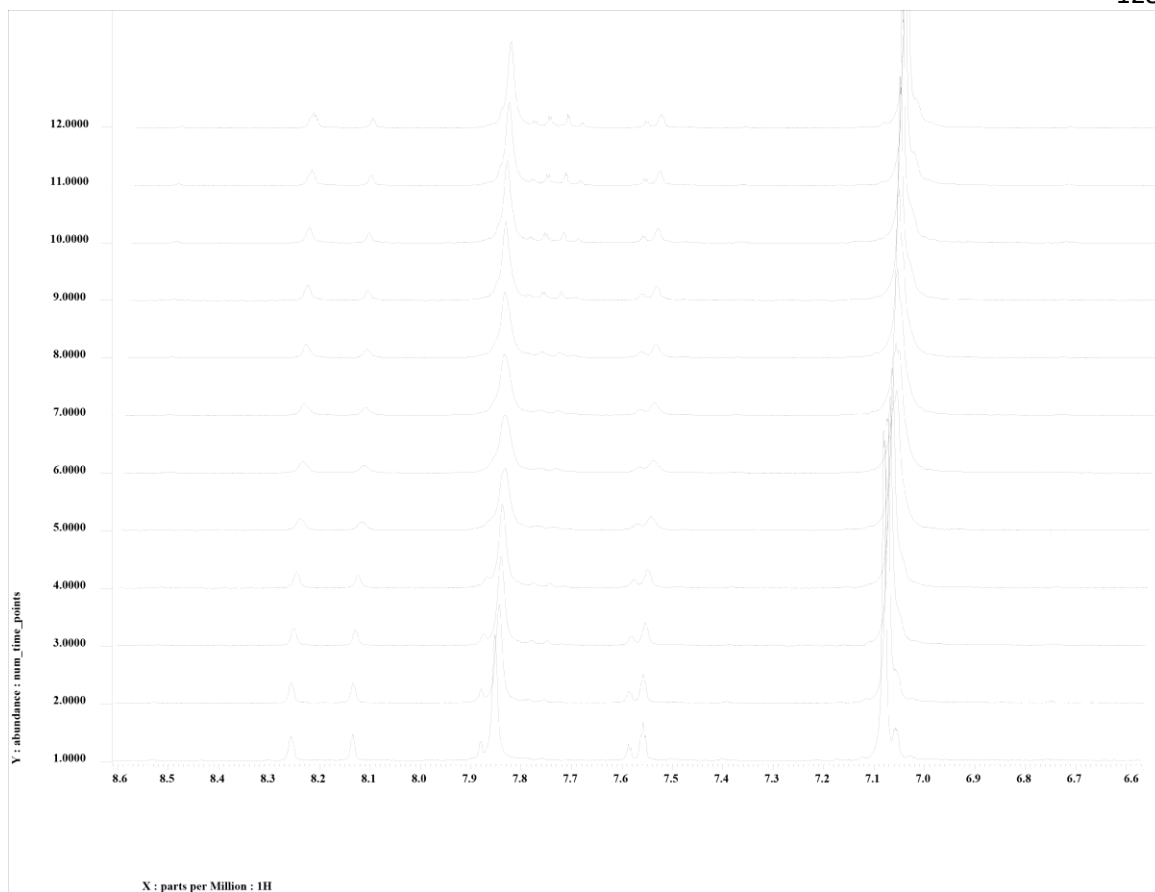
C<sub>10</sub>H<sub>7</sub>N<sub>5</sub>O, <sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>): δ 7.18 (s, 1H), 7.72 (s, 1H), 7.84 (d, 2H), 8.05 (d, 1H), 8.26 (s, 1H), 8.45 (s, 1H) ppm, <sup>13</sup>C NMR (75 MHz, DMSO-d<sub>6</sub>): δ 114.61, 118.64, 119.80, 126.70, 128.14, 130.49, 138.72, 139.05, 139.90, 165.99 ppm, HRMS: Calc. 213.0651 m/z, Found: Compound hydrolyzes to starting materials.



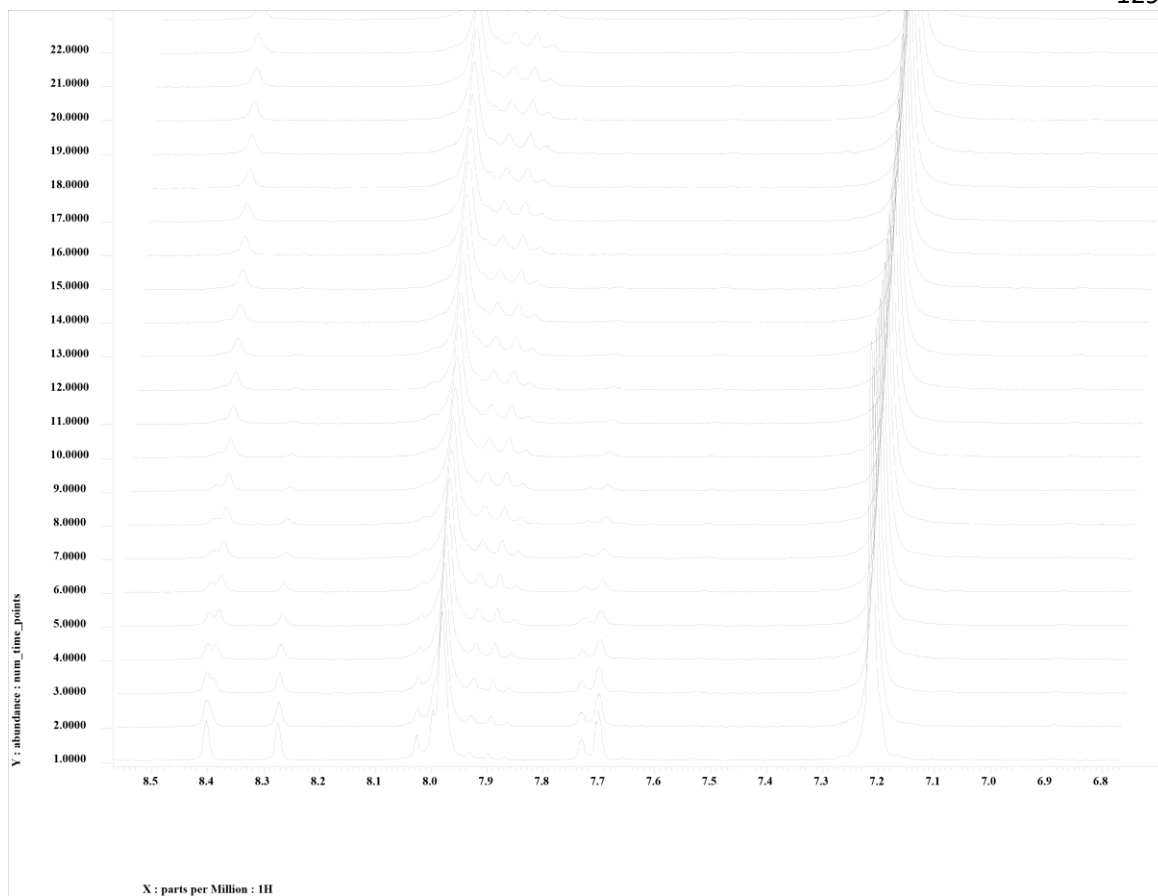
**Figure S25.**  $^1\text{H}$  NMR of B5 in DMSO. Integration of three protons from excess CDI shown with +, where one falls under product peak leading to the high integration. Two signals from imidazole shown with \*.



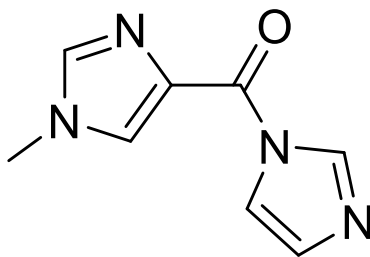
**Figure S26.**  $^{13}\text{C}$ -NMR of B5 in DMSO- $d_6$ . Liberated imidazole carbons shown with \*.



**Figure S27.** Hydrolysis experiment of B5 at 25°C. Showing from 5 minutes to 125 minutes, where each scan was performed in 10 minutes intervals. Monitoring the disappearance of product peaks at 8.28 ppm.

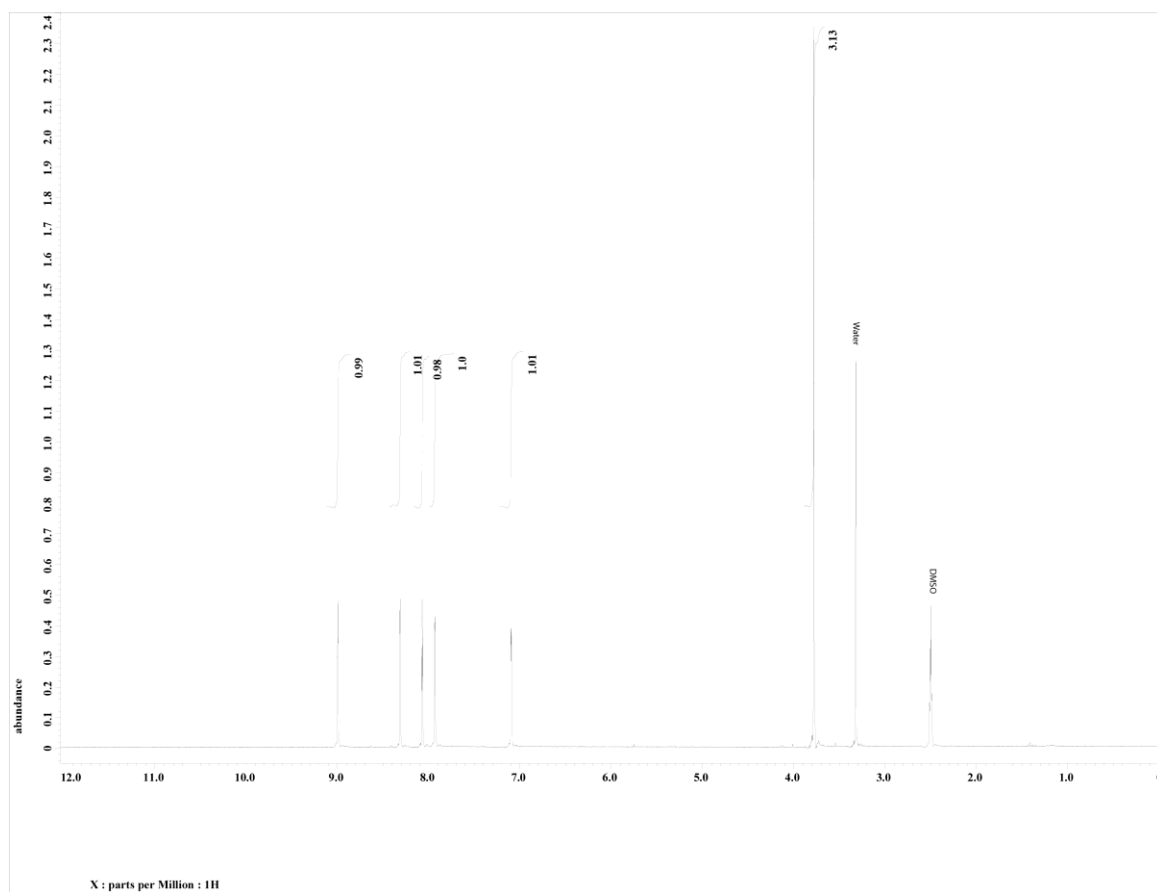


**Figure S28.** Hydrolysis experiment of B5 at 37°C. Showing from 5 minutes to 125 minutes, where each scan was performed in 10 minutes intervals. Monitoring the disappearance of product peaks at 8.28 ppm.

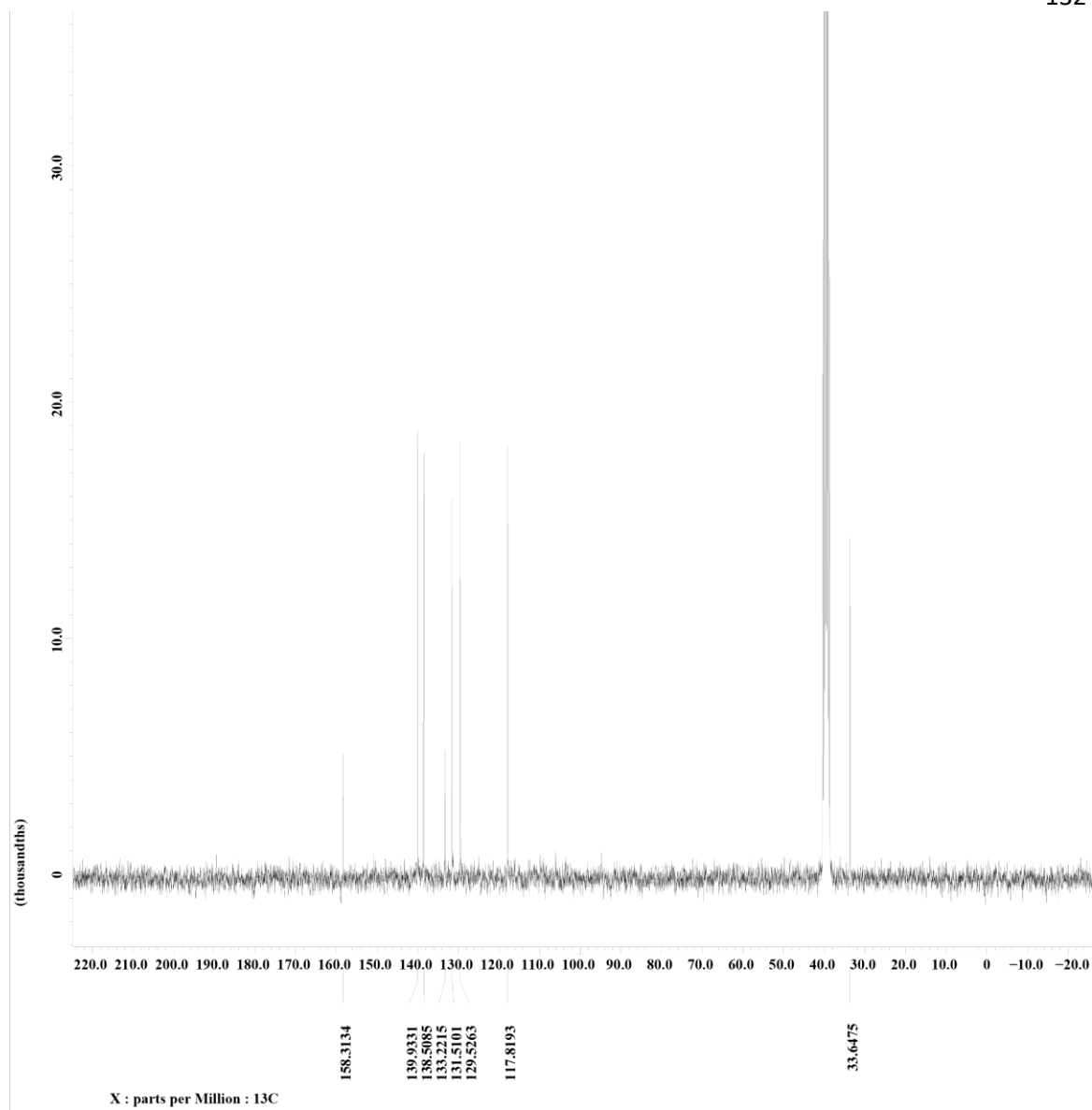
**1-methylimidazole-4-carboxylic acid imidazolide (1M4)**

1M4 was obtained following general procedure 1 above. For the analytical sample, a portion of the crude mixture was diluted into dichloromethane and extracted 3X with saturated sodium bicarbonate solution. The organic layer was dried over  $\text{MgSO}_4$  and concentrated under reduced pressure.

$\text{C}_8\text{H}_9\text{N}_4\text{O}$ ,  $^1\text{H}$  NMR (300 MHz,  $\text{DMSO-d}_6$ ):  $\delta$  3.78 (s, 3H), 7.10 (s, 1H), 7.94 (s, 1H), 8.08 (s, 1H), 8.32 (s, 1H), 9.00 (s, 1H) ppm,  $^{13}\text{C}$  NMR (75 MHz,  $\text{DMSO-d}_6$ ):  $\delta$  33.65, 117.82, 129.53, 131.51, 133.22, 138.51, 138.22, 138.51, 139.93, 158.31 ppm, HRMS Calc: 177.0776 m/z Found: 177.0771 (M+H) m/z.

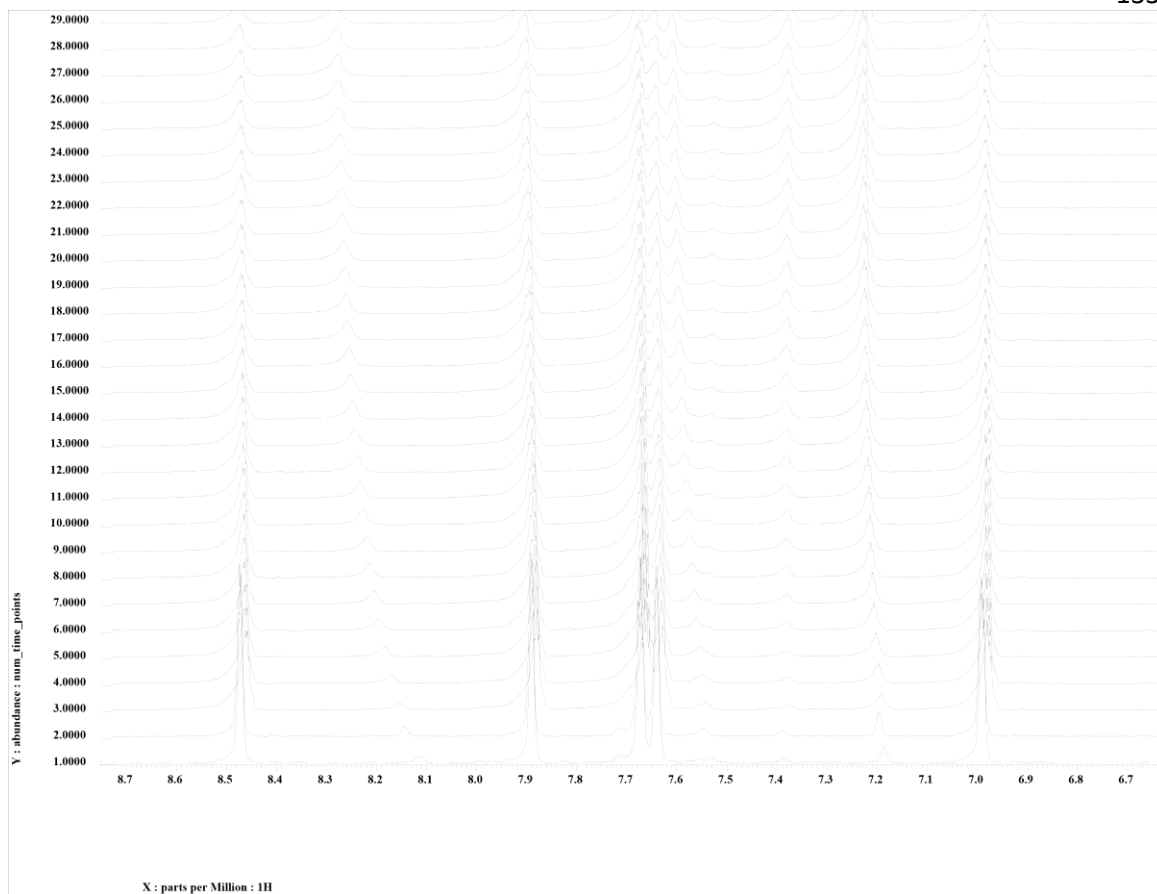


**Figure S29.** <sup>1</sup>H NMR of 1M4 in DMSO-d<sub>6</sub>

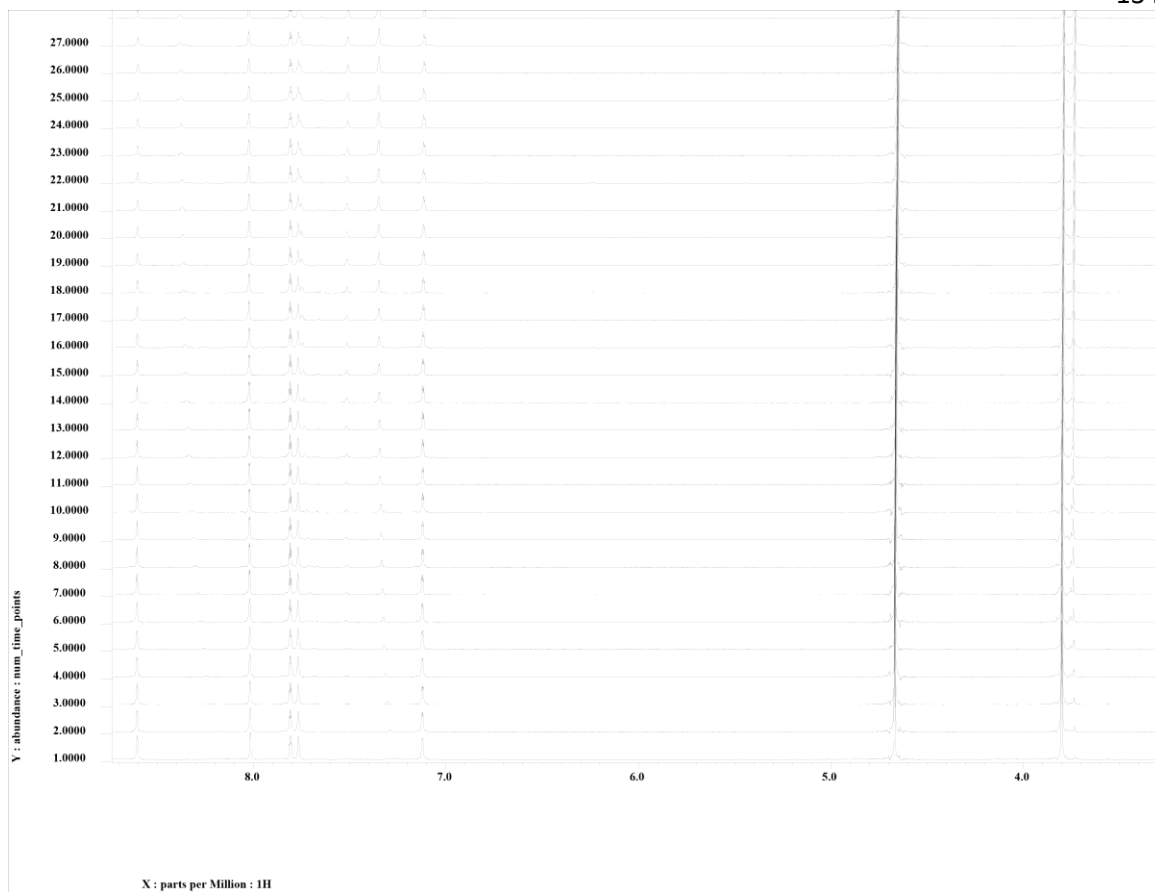


**Figure S30.**  $^{13}\text{C}$  NMR of 1M4 in DMSO- $\text{d}_6$ .

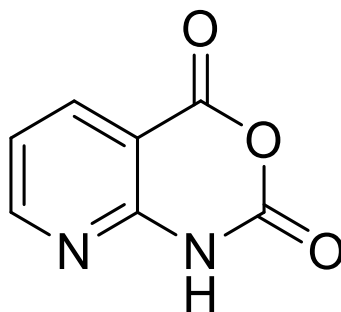




**Figure S31.** Hydrolysis experiment of 1M4 at 25°C. Showing from 5 minutes to 905 minutes, where each scan was performed in 30 minutes intervals. Monitoring the disappearance of product peaks at 3.81 ppm and appearance of 3.75 ppm.



**Figure S32.** Hydrolysis experiment of 1M4 at 37°C. Showing from 5 minutes to 905 minutes, where each scan was performed in 30 minutes intervals. Monitoring the disappearance of product peaks at 3.81 ppm and appearance of 3.75 ppm.

**3-azaisatoic anhydride (3AIA)**

$C_7H_4N_2O_3$ ,  $^1H$  NMR (300 MHz, DMSO- $d_6$ ):  $\delta$  7.29 (t, 1H), 8.31 (dd, 1H), 8.65 (d, 1H), 11.70 (s, 1H) ppm,  $^{13}C$  NMR (75 MHz, DMSO- $d_6$ ):  $\delta$  106.71, 119.74, 138.25, 147.12, 153.22, 155.93, 159.63 ppm, HRMS: Calc: 164.0222 m/z, Found: Compound hydrolyzes. Spectra was compared to commercially available materials.

## CHAPTER 5: CONCLUSIONS

The experiments and compounds described in this work represent novel constructs from which an improved output for RNA SHAPE in vitro and in vivo was demonstrated. Through tuning of reagent reactivity, improved water solubility and use of structural design, significant improvement to signal to noise was able to be achieved. The field of protein bioconjugation and RNA SHAPE can benefit from attention to details for bioconjugation reagents. The design of reagents with the solvation environment considered offers inherent benefits to the use of chemical probes, bioconjugation strategies, and numerous other roles in chemical biology.

In chapter two, the development of a water-soluble multifunctional platform is demonstrated and the preparation and its modulation are described. The application of the reagents within the field of RNA probing through SHAPE chemistry were demonstrated on the small subunit of the ribosomal RNA of the malaria pathogen *Plasmodium falciparum* in vitro. This was directly compared to NAI under its traditional probing conditions and provided a significant increase in the probing output, alluding to an increased reactivity without loss of specificity. These water-soluble reagents possessed an identical reactive site to that of the original SHAPE reagent NMIA which has since been phased out due to low signal. The increased signal using a water-soluble derivative directly demonstrated the importance of solubility, largely reiterating simple concentration effects which were largely ignored in the development of early probing strategies.

In chapter three, the use of well characterized electronic effects on the reactive site of the water-soluble reagents was demonstrated through the use of chemical activation or deactivation. This aimed to increase the footprint of the reactivity of these reagents to expand into the areas of ideal reactivity and solubility which had been recently imagined by Weeks et al. A strategy which simultaneously activated and solubilized the probe was demonstrated using a quaternary ammonium directly bound to the probe. An alternative strategy which electronically deactivated and sterically blocked the reactive site was also demonstrated. The direct comparison of these reagents to the highly reactive 1M7 *in vitro* clearly indicated advantages to high reactivity, but even the poorly reactive probe provided increased signal compared to 1M7. This result demonstrated the concentration of the reagent to be more important factor than simply its reactivity.

In chapter four, the development of the new nicotinic acid imidazolidine reagent 2A3 demonstrated the use of a unique intramolecular activation strategy which resulted in a superior probe *in vivo*. This strategy aligns with the goals of the work done to the isatoic anhydride probes in chapter two and chapter three. For *in vitro* experimentation we demonstrated more reactive reagents as largely superior, moving to *in vivo* experimentation the highly reactive constructs did not offer tangible benefits. Alternatively, 2A3 relies on an activation strategy that provides the reagent time to accumulate and results in an increased probing output. This result agrees with the large volume of validation which has been previously demonstrated with NAI for *in vivo* RNA probing.

The development of new and novel bioconjugation strategies and reagents that provide benefit for the final constructs' functionality or increases the throughput of preparation is important. Methods to improve functionality or use of strategies to improve conjugation efficacy and stability and prepare more precise conjugates are important for the future of medicinal biologics. Continuing research into chemical probing and conjugation will offer improvements which may improve therapeutic techniques, research output and provide important fundamental research.