

Chemical Modifications for Spectral Tuning of Fluorogenic RNA Aptamers

A thesis submitted to the Nuffield Department of Medicine in fulfilment for the degree of

Doctor of Philosophy

Lynda Truong

Linacre College, Trinity 2019

University of Oxford

National Institutes of Health

Supervisors:

Prof. Tom Brown

Dr. Adrian Ferré-D'Amaré

Table of Contents

Abstract ..... **x**

Acknowledgements ..... **xii**

Abbreviations ..... **xiv**

CHAPTER 1. INTRODUCTION: THE DEVELOPMENT OF FLUOROGENIC RNA APTAMERS **1**

1.1 Fluorescent Biomolecules for Imaging 2

1.1.1 The Development of Green Fluorescent Protein 3

1.2 Fluorogenic RNA Aptamers 5

1.2.1 G-Quadruplexes: A Common Structural Motif in Fluorogenic RNA Aptamers 6

1.2.2 Spinach is an RNA Mimic of GFP 8

1.2.3 Corn and Chili Activate Red-shifted FP-related Fluorophores 13

1.2.4 High Fluorophore Affinity of Compact Mango Aptamers 16

1.2.5 Fluorogenic Aptamers without G-quadruplexes 20

1.3 Applications of Fluorescence Turn-on Aptamers 22

1.3.1 Imaging 22

1.3.2 Biosensors 23

1.3.3 Intermolecular interactions 23

1.4 Potential for Optimizing Fluorogenic RNA Aptamers 24

1.5 Research Objectives 27

1.5.1 Use of naturally occurring RNA modifications for spectral tuning 27

1.5.2 Studying nucleoside and fluorophore analogues in Spinach aptamers..... 28

1.5.3 Remote excitation in Mango aptamers using fluorescent dye labels..... 28

CHAPTER 2. SYNTHESIS OF tRNA MODIFICATIONS FOR FLUORESCENT TUNING..... 30

2. 1 INTRODUCTION..... 31

2.2 RESULTS AND DISCUSSION 36

2.2.1 Synthesis of nucleoside precursors..... 36

2.2.1.1 Synthesis of $PreQ_0$ nucleoside 36

2.2.1.2 Synthesis of Archaeosine tRNA nucleoside 39

2.2.1.3 Synthesis of Wyosine tRNA nucleoside 40

2.2.2 Synthesis of phosphoramidite forms of tRNA modifications 42

2.2.2.1 Phosphoramidites using TIPS protection..... 45

2.2.2.2 Phosphoramidites using TC protection..... 47

2.3 CONCLUSIONS AND FUTURE WORK..... 49

2.4 MATERIALS AND METHODS 50

2-Amino-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 50

4-Chloro-2-amino-7*H*-pyrrolo[2,3-d]pyrimidine. 51

4-Chloro-2-pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidine. 51

2-Pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 52

4-Chloro-2-pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidine. 53

4-Chloro-5-iodo-2-pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidine. 54

4-Chloro-5-iodo-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidine. 6 55

5-Iodo-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 56

5-Cyano-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 57

5-Cyano-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 58

2-Amino-5-cyano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one (PreQ0). 59

5-cyano-2-[(dibutylamino)methylidene]amino-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 60

5-cyano-2-[(dibutylamino)methylidene]amino-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 22 61

5-cyano-2-[(dibutylamino)methylidene]amino-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-2′-O-triisopropylsilyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 23 62

5-cyano-2-[(dibutylamino)methylidene]amino-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-2′-O-triisopropylsilyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one-3′-[2-cyanoethyl *N*,*N*-bis(1-methylethyl)phosphoramidite]. 63

5-cyano-2-[(dibutylamino)methylidene]amino-3′,5′-O-tetraisopropyldisiloxano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one..... 64

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-thiocarbonylimidazole-3′,5′-O-tetraisopropyldisiloxano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 65

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-3′,5′-O-(tetraisopropyldisiloxano)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 66

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 68

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one-3′-[2-cyanoethyl *N*,*N*-bis(1-methylethyl)phosphoramidite]. ... 69

2-Amino-5-amidino-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 70

2′,3′,5′-O-tri(tertbutyldimethylsilyl)-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 71

3′,5′-O-tetraisopropyldisiloxano-2-amino-1,9-dihydro-9-β-D-ribofuranosyl-6*H*-purin-6-one. 72

2′-O-acetyl-3′,5′-O-tetraisopropyldisiloxano-2-amino-1,9-dihydro-9-β-D-ribofuranosyl-6*H*-purin-6-one. 73

2′-O-acetyl-3′,5′-O-tetraisopropyldisiloxano-1-propargyl-2-amino-1,9-dihydro-9-β-D-ribofuranosyl-6*H*-purin-6-one. 74

2′-O-acetyl-3′,5′-O-tetraisopropyldisiloxano-3-β-D-ribofuranosyl-9-hydro-6-methyl-5*H*-oxoimidazo(1,2-a)purine. 75

2′-O-acetyl-3′,5′-O- tetraisopropyldisiloxano -3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 76

3′,5′-O-tetraisopropyldisiloxano-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 78

2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-3′,5′-O- tetraisopropyldisiloxano -3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 78

2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 80

2′-O-(1,1-dioxothiomorpholine-4-carbothioate)- 5′-O-(4,4′-dimethoxytrityl)-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine-3′-[2-cyanoethyl *N*,*N*-bis(1-methylethyl)phosphoramidite]. 81

CHAPTER 3. DESIGN AND MODIFICATION OF RNA SPINACH APTAMERS 83

3.1 INTRODUCTION 84

3.2 RESULTS AND DISCUSSION 86

3.2.1 Fluorescence effects of aptamer and fluorophore modifications 86

3.2.1.1 Spectral changes of Spinach aptamer when bound to different fluorophores 87

3.2.1.2 Structure of iSpinach bound to novel ligand 91

3.2.2 Tuning fluorescent wavelength with base triple mutations in Spinach 93

3.2.3 Split Baby Spinach 97

3.2.3.1 Designing split aptamer for solid phase synthetic accessibility 98

3.2.3.2 Fluorescence effects of modifying Split Baby Spinach with aminopurine and pyrrolocytosine 101

3.2.3.4 Fluorescence effects of modifying Split Baby Spinach with $PreQ_0$ and Wyosine 112

3.3 CONCLUSIONS AND FUTURE WORK 116

3.4 MATERIALS AND METHODS 119

3.4.1 *In vitro* RNA transcription 119

3.4.2 Fluorescent measurements 119

3.4.3 *i*Spinach/DFHBI-1T crystallization and data collection 120

CHAPTER 4. REMOTE EXCITATION OF RNA MANGO APTAMERS 122

4.1 INTRODUCTION 123

4.2 RESULTS AND DISCUSSION 124

4.2.1 RNA TERRA: a model G-quadruplex 124

4.2.1.1 Modifying G-quadruplex loop regions with aminopurine and pyrrolocytosine 124

4.2.1.2 Effects of fluorescent nucleotide modifications on spectral properties of fluorescent G-quadruplex binders 127

4.2.2 Fluorescence antenna effects with Mango-III aptamer 129

4.2.2.1 Modifying ligand binding region of Mango-III with aminopurine and pyrrolocytosine 129

4.2.2.2 Internal FRET system using fluorescein and rhodamine labels 131

4.3 CONCLUSIONS AND FUTURE WORK 139

4.3 MATERIALS AND METHODS 140

4.3.1 AP and PyC labelled oligonucleotides 140

4.3.2 Fluorescein and rhodamine labelled oligonucleotides 140

4.3.3 Fluorescent measurements 141

4.3.4 Calculations for apparent FRET efficiencies 141

CHAPTER 5. CONCLUSIONS AND FUTURE WORK 143

5.1 Conclusions and Future Work 144

REFERENCES i

APPENDIX xix

BS1 $PreQ_0$ HPLC trace and mass spectrum xx

BS2 $PreQ_0$ HPLC trace and mass spectrum xxii

BS1 Wyosine HPLC trace and mass spectrum xxiv

BS2 Wyosine HPLC trace and mass spectrum xxv

Table S1. Summary of Crystallographic Statistics xxvii

Table S2. RNA Sequences xxviii

NMR Compound 16 xxix

NMR Compound 17 xxxiii

NMR Compound 18 xxxvii

NMR Compound 19 xli

NMR Compound 20 xlv

NMR Compound 21 xlix

NMR Compound 22 liii

NMR Compound 23 lv

NMR Compound 24 lviii

NMR Compound 26a lxi

NMR Compound 27a lxv

NMR Compound 28a lxix

NMR Compound 29a lxxiii

NMR Compound 30a lxxvii

NMR Compound 28b lxxix

NMR Compound 29b lxxxiii

NMR Compound 30b lxxxvii

Abstract

Fluorogenic RNA aptamers such as Spinach, Corn, and Mango have provided a practical method for imaging RNAs. While much work has gone into developing different aptamer-ligand systems to expand the current toolbox of fluorogenic RNA aptamers, the potential of synthetic modifications has remained largely unexplored. To address this gap, I have investigated a range of canonical mutations, naturally occurring noncanonical modifications, and artificial nucleotide labels to study the effects of these synthetic modifications on the spectral properties of fluorogenic RNA aptamers. To access naturally occurring, noncanonical RNA modifications, I have synthesized novel phosphoramidites for tRNA modifications $PreQ_0$ (2-amino-5-cyano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one) and wyosine (3,4-dihydro-4,6-dimethyl-3-β-D-ribofuranosyl-9*H*-imidazo[1,2-a]purin-9-one) for incorporation into an RNA aptamer using solid-phase oligonucleotide synthesis. Guided by the crystal structure of the Spinach aptamer, I developed a Split Baby Spinach aptamer which is accessible via solid-phase oligonucleotide synthesis. I then explored the effects of modifications such as $PreQ_0$ and wyosine, as well as fluorescent nucleotide analogues 2-aminopurine and pyrrolocytosine, on the fluorogenic aptamer when bound to different fluorophore-ligands. While most mutations had a negligible effect on the fluorescent properties of the aptamer-fluorophore complex, one wyosine mutation near the ligand binding pocket exhibited enhanced fluorescence turn-on. Finally, guided by the crystal structure of *i*Mango-III, I have also individually introduced the fluorescent dyes, fluorescein and rhodamine, into five positions surrounding the *i*Mango-III ligand binding site. These modifications create an internal FRET system, providing a simple method for tuning the excitation or emissions wavelength of the RNA aptamer to those

of the introduced fluorophores. These synthetic modifications demonstrate the potential of noncanonical nucleotides for fluorescence modulation enhancement.

Acknowledgements

This research would not have been possible without the direct support of the Tom Brown lab at the University of Oxford and the Adrian Ferré-D'Amaré lab at the National Institutes of Health (NIH). Tom and Adrian have both provided endless guidance and unconditional support over the course of my degree, and for that I cannot thank them enough. It has been a rewarding experience to work with and learn from such inspiring mentors and colleagues on both sides of the Atlantic Ocean. Pawan Kumar and Ysobel (Yssy) Baker were amazing mentors and friends, patiently offering their expertise in my synthetic chemistry experiments. Robert Trachman, Christopher Jones, Michael Banco, and Natalia Demeshkina have all been incredible sources of knowledge as my project transitioned from chemistry to chemical and structural biology. I am forever grateful to everyone in both labs for the insightful scientific discussions, fun memories, and unwavering support.

This work was funded by the NIH Oxford-Cambridge Research program and the intramural program of the National Heart, Lung, Blood, Institute (NHLBI) at the National Institutes of Health (NIH). Grzegorz Piszczek and the Biophysics Core (NHLBI) provided expertise and equipment for fluorescence spectroscopy experiments. Duck-Yeon Lee and the Biochemistry Core (NHLBI) provided expertise and equipment for HPLC purification. The Oxford Chemistry Department and its NMR and mass spectrometry staff provided expertise and equipment for NMR and mass spectrometry experiments. Cameron Thorpe and Afaf El-Sagheer synthesized and purified my custom RNA oligonucleotides. My introduction was originally published in Protein Science and reused with the permission of Johnson Wiley & Sons.

On a personal note, I also want to take the time to thank my friends and family for their support. My husband, Matthew Dixon, in the ultimate display of love and support, left his life in the UK to join me in the US as I transitioned to my lab at the NIH. My parents, Henry Truong and Cindy Phan, taught me the value of learning and education, without ever having the chance to pursue higher education themselves. My little sister, Jennifer Truong, motivated me to aim high, so I can be someone she looks up to. My friends in Linacre College and in the OxCam program helped me grow as person, rather than just as a scientist. Ashley Tsai, my dynamic duo in lab and life, has been a constant source of positivity and emotional support throughout my four years. Without everyone named here and many more, this would not have been possible.

Abbreviations

A	adenosine
ACE	2′-bis(2-acetoxyethoxy)methyl
ADP	adenosine diphosphate
AMP	adenosine monophosphate
AP	2-aminopurine
BS	Baby Spinach
C	cytosine
DCM	dichloromethane
DEPC	diethyl pyrocarbonate
DFHBI	3,5-difluoro-4-hydroxybenzylidene imidazolinone
DFHBI-1T	(*Z*)-4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-1-(2,2,2-trifluoroethyl)-1*H*-imidazol-5(4*H*)-one
DFHBI-2T	(*Z*)-4-(3,5-difluoro-4-hydroxybenzylidene)-1-methyl-2-(trifluoromethyl)-1*H*-imidazol-5(4*H*)-one
DFHO	3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime
DIPEA	*N*,*N*-Diisopropylethylamine
DIR	dimethylindole red
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMHBI	3,5-dimethoxy-4-hydroxybenzylidene imidazolinone
DMSO	dimethyl sulfoxide
DMT	dimethoxytrityl

DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EBFP	enhanced blue fluorescent protein
ECFP	enhanced cyan fluorescent protein
EGFP	enhanced green fluorescent protein
FP	fluorescent protein
FRET	Förster resonance energy transfer
G	guanosine
GFP	green fluroescent protein
GMP	guanosine monophosphate
HBI	4-(*p*-hydroxybenzylidene)-5-imidazolinone
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high pressure liquid chromatography
IR	infrared
MeCN	acetonitrile
MeOH	methanol
MGA	malachite green aptamer
MGA	malachite green
MPD	2-methyl-2,4-pentanediol
NHS	*N*-hydroxysuccinimide
NMM	*N*-methylmorpholine
NMR	nuclear magnetic resonance
nt	nucleotide

OTB	oxale thiazole blue
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PyC	pyrrolocytosine
RFP	red fluorescent protein
RNA	ribonucleic acid
SAM	s-adenosylmethionine
SAXS	small angle x-ray scattering
SELEX	systematic evolution of ligands by exponential enrichment
T	thymidine
TBE	Tris/Borate/Ethylenediaminetetraacetic acid
TBMDS	tetrabutyldimethylsilylene
TC	2′-thiomorpholine-4-carbothioate
TERRA	telomeric Repeat-containing RNA
THF	tetrahydrofuran
THT	Thioflavin T
TIPDS	tetraisopropyldisilyl
TIPS	triisopropylsilylane
TLC	thin layer chromatography
TO	thiazole orange
TOM	triisopropylsilyloxymethyl
tRNA	transfer RNA
U	uridine
UV	ultraviolet

YFP	yellow fluorescent protein
YO3	oxale yellow

CHAPTER 1. INTRODUCTION: THE DEVELOPMENT OF FLUOROGENIC RNA APTAMERS[1]

[1] Originally published in Protein Science, adapted & reused with the permission of Johnson Wiley & Sons.

The discovery and subsequent application of green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* revolutionized the study of biology by providing visual access to the cellular environment.[1–3] Genetically encoded fluorescent protein tags allowed cellular processes to be monitored and analyzed in real time. Since their initial discovery, considerable effort has been dedicated to engineering the biophysical and biochemical properties of fluorescent proteins (FPs) for improved stability, enhanced brightness, and an extended fluorescent spectrum. This has led to the development of blue, cyan, yellow, and red fluorescent proteins in addition to super-folding FPs, far-red fluorescent protein probes, split FPs, and photoactivated FPs, among others.[4–9] More recently, these fluorescent protein tags have inspired a new branch of molecular imaging in the realm of RNA.

No RNAs with bright intrinsic fluorophores have been discovered. Instead, RNA mimics of FPs are molecules that bind to cognate small molecules, *i.e.*, aptamers.[10] These aptamers recognize compounds that are minimally fluorescent on their own, and increase their fluorescence by as much as two thousand times upon binding.[11,12] All such described fluorescence turn-on aptamers are the result of *in vitro* evolution and design experiments. In principle, the biophysical properties of fluorogenic RNA-fluorophore complexes can be tuned more broadly than those of FPs, since neither the chemical nature of the conditionally fluorescent small molecules nor the structure of the turn-on aptamer RNAs are constrained by evolutionary history.

1.1.1 The Development of Green Fluorescent Protein

GFP adopts the now classic 11-stranded β-barrel fold (Figure 1.1A), which provides the environment for spontaneous formation of 4-(*p*-hydroxybenzylidene)-5-imidazolinone (*p*-HBI, Figure 1.1B) from three amino acid residues (Ser65, Tyr66, and Gly67) within the protein core.[13,14] Protein residues surrounding the intrinsic *p*-HBI restrain the rotational and vibrational motion of its photoexcited state, reducing non-radiative decay and enhancing fluorescence (Figure 1.1C).[15,16] Consistent with this, denaturation of GFP destroys fluorescence, and free *p*-HBI in solution is non-fluorescent.[17,18] Wild-type GFP displays two distinct absorption bands that arise from the neutral (phenol) and anionic (phenolate) forms of *p*-HBI (395 nm 475 nm, respectively). Mutations stabilizing the phenolate form of *p*-HBI, such as those in enhanced GFP (EGFP), result in improved brightness.[6,19,20]

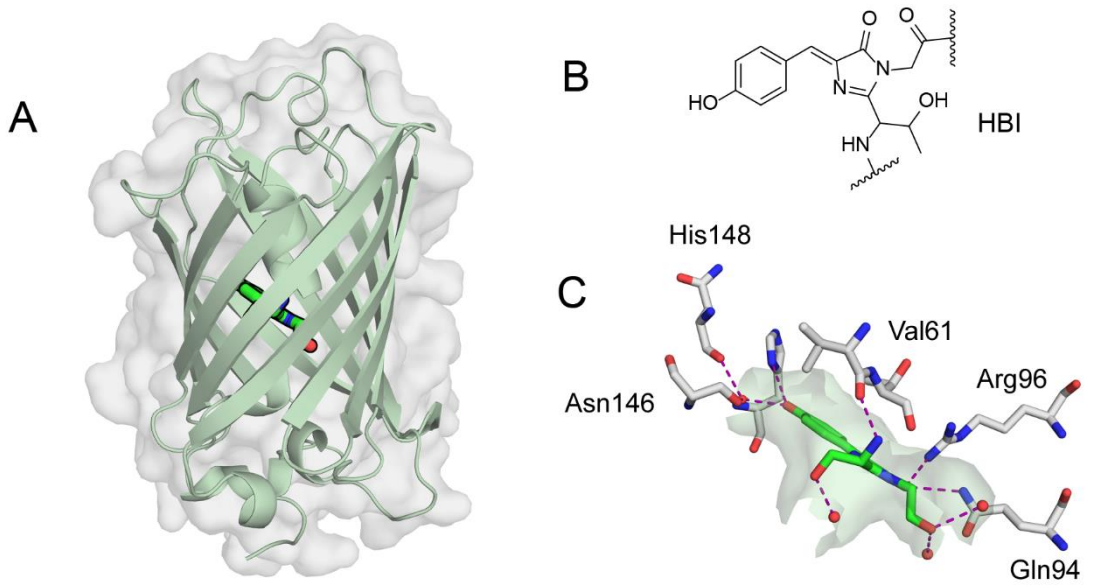


Figure 1.1. Structure of green fluorescent protein (GFP). (A) Characteristic 11 strand β-barrel of GFP containing buried fluorophore.[14] (B) Chemical structure of 4-(p-hydroxybenzylidene)-5-imidazolinone (HBI), the intrinsic chromophore in GFP. (C) Amino acid residues that form and stabilize HBI in GFP core. (PDB: 1GFL)

Formation of *p*-HBI is initiated by nucleophilic attack of the amide of Gly67 on the carbonyl of Ser65. This intramolecular cyclization is followed by dehydration and oxidation to yield the conjugated fluorophore.[21–23] Modification of the reactive amino acids, as well as their environment, gives rise to the diverse intrinsic fluorophores and fluorescent properties of related FPs. The blue-shifted EBFP and ECFP arise from mutation of Tyr66 to histidine or tryptophan, respectively.[23–25] Mutating an adjacent amino acid (Thr203Tyr) promotes π-π stacking between Tyr66 and Tyr203. This causes a 20 nm bathochromic shift, as seen in YFP.[26] The intrinsic fluorophores of red fluorescent proteins, isolated from corals, arise from a process similar to that of GFP. After intramolecular cyclization and an initial oxidation, a second oxidation step kinetically competes with a dehydration step (which would form the green

chromophore in GFP) in order to form the acylimine seen in the extended conjugation of the RFP chromophore.[27,28]

Considerable effort has been devoted to engineering improved folding, stability and oligomerization state of FPs for imaging applications. The improved variant EGFP retains its brightness at 37 °C and is commonly used for imaging and microscopy.[29] Many FPs form dimers or tetramers in solution.[30] Studies that employ FP fusion proteins can be disrupted by oligomerization. Monomeric variants of FPs have been developed by mutating residues at the oligomerization interface.[31] In some instances, these mutations abolish the FPs' intrinsic fluorescence and require further mutational analysis to recover fluorescence.[32] Psudeo-monomeric FPs have also been developed to address the challenges of oligomerization and aggregation, where two monomer FPs are connected using a flexible peptide linker. This intentional dimerization reduces the likelihood of further oligomerization during use.[33]

1.2 Fluorogenic RNA Aptamers

The invention of RNA *in vitro* selection (SELEX) led to the identification of numerous small-molecule binding aptamers.[10,34,35] The malachite green aptamer (MGA) was originally selected in order to generate a light-activated RNA-cleaving agent, as illumination of MG produces hydroxyl radicals.[36] Subsequently, it was found that upon binding, the triphenylmethane dye exhibited a 2300-fold increase in fluorescence.[37] Structure determination demonstrated that the dye is sandwiched between a base pair and base quadruple.[38] By restricting rotation around the central carbon bond connecting the phenyl rings, the aptamer reduces the fluorophore's ability to undergo non-radiative decay and thus enhances its fluorescence. The MGA provided proof of principle for

fluorescent turn-on aptamers, but its photoactivated RNA cleaving properties (and attendant light-induced cytotoxicity) limited imaging applications. Turn-on aptamers for derivatives of DIR and Hoechst were also identified.[39,40] Although these did not exhibit photoactivated RNA-cleavage, their application to live-cell imaging was limited, either because of modest fluorescence turn-on (DIR, 60-fold)[40] or non-specific nucleic-acid binding by the fluorophores (Hoechst derivative).[39]

1.2.1 G-Quadruplexes: A Common Structural Motif in Fluorogenic RNA Aptamers

Of the solved structures of fluorogenic RNA aptamers, many feature a G-quadruplex structural core. G-quadruplexes are non-canonical, helical structural motifs formed from two or more stacked g-tetrads.[41] Each tetrad consists of four guanines forming Hoogsten base pairs, arranged into a square planar layer (Figure 1.2A). The quadruplexes can be formed from one (unimolecular), two (bimolecular), or four (tetramolecular) strands of nucleic acid.[42] All fluorogenic RNA aptamers containing a G-quadruplex, to date, have contained unimolecular or bimolecular G-quadruplexes. Depending on the orientation of the strands forming their backbone, quadruplexes can also be described as parallel or antiparallel (Figure 1.1C). In quadruplexes, strands which are oriented in the same direction will adopt the same conformation. Therefore, parallel quadruplexes will be comprised of either all *anti* or all *syn* conformations (Figure 1.1B). However, changes in strand orientation will also cause a change in conformation.[43] G-quadruplexes can form lateral loops (joining two adjacent strands), diagonal loops (joining two diagonal strands), or propeller loops (joining the top of one strand to the bottom of a second adjacent strand or vice versa).[44] The G-quadruplexes found in fluorogenic RNA aptamers have had surprisingly complex connectivity,

displaying a mixture of parallel and anti-parallel strands and loops between 1 nt and 34 nt in length.[45,46]

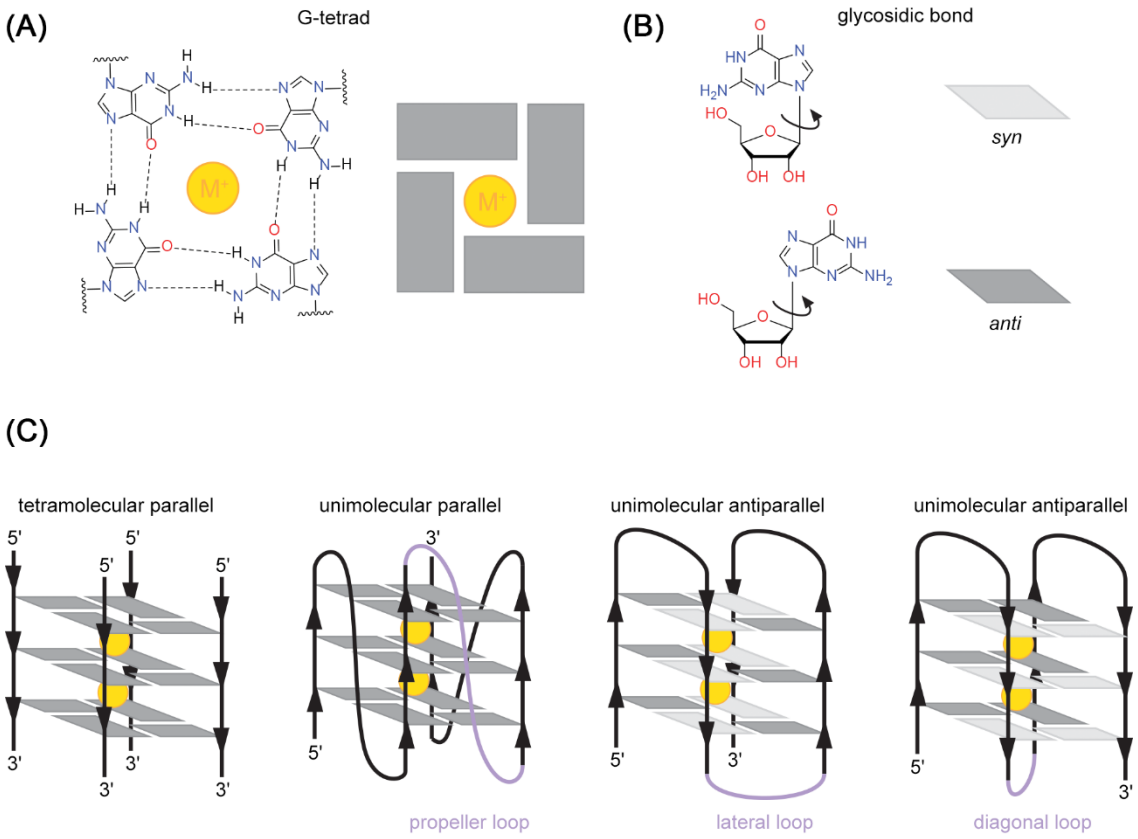


Figure 1.2. Structural overview of conventional G-quadruplexes. (A) The G-tetrad or G-quartet is comprised of four Hoogsten-bonded guanines on a plane. A cation, typically K+, is coordinated in the axial pore. (B) Each glyosidic bond in the G-quadruplex can adopt either anti or syn conformations. (C) G-quadruplexes are comprised of two or more tiers of G-tetrads, and can be oligomolecular or unimolecular. Strand orientations can be locally parallel or antiparallel, and loops connecting the stacked guanines can form propeller, lateral, or diagonal connections. Figure adapted from Hänsel-Hertch, et. al, *Nature Reviews Molecular Cellular Biology* (2017)[47]

1.2.2 Spinach is an RNA Mimic of GFP

In 2011, the Jaffrey group reported the isolation and characterization of a variety of turn-on aptamers selected for binding to *p*-HBI derivatives.[48] A 98 nt aptamer in complex with 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), was the

brightest, achieving a ~1000-fold turn-on and a quantum yield of 0.72, thus being ~50% as bright as EGFP. This RNA, termed Spinach, was selected to bind preferentially to DFHBI (Figure 1.3A,B), whose two fluorine substituents lower the pK_a of the fluorophore, which thus exists in the brighter phenolate form (analogous to the ionization state of *p*-HBI within EGFP). Spinach was used successfully as a live cell RNA-imaging tag by expressing ribosomal RNA-Spinach fusions in live cells that were soaked in the cell-permeable and non-cytotoxic DFHBI, demonstrating the potential of this aptamer-fluorophore complex as a partially genetically-encoded *in vivo* fluorescent tag.

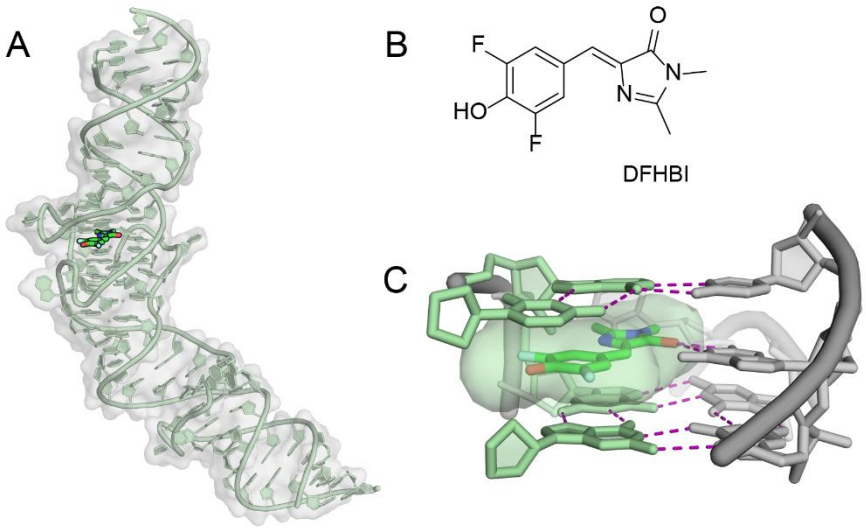


Figure 1.3. Structure of Spinach aptamer. (A) Structure of Spinach aptamer with bound DFHBI ligand, featuring a G-quadruplex fluorophore binding core flanked by two helices.[45] (B) Chemical structure of 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), fluorophore-ligand of Spinach. (C) The fluorophore pockets DFHBI-bound Spinach. Hydrogen bonds are depicted by purple dashed lines and stacking interactions in the fluorogenic RNA aptamers are highlighted in green. (PDB: 4TS0)

Crystallographic structure determination of Spinach-DFHBI complexes revealed an RNA comprised of a single coaxial helical stack organized around a G-quadruplex motif at its core.[45,46] The presence of a G-quadruplex in Spinach was unexpected because the aptamer lacks any of the canonical sequence signatures of G-quadruplexes (such as consecutive stretches of G-rich sequences) and computational secondary structure prediction had suggested only the presence of duplexes.[48] The bound DFHBI is sandwiched between the top tetrad (quartet) of the G-quadruplex and a base triple, and is further restrained by an unpaired guanine nucleobase that is coplanar with the two heterocycles of the fluorophore (Figure 1.3C). Mutagenesis demonstrated the importance of the G-quadruplex, base triple and unpaired guanine for fluorescence

turn-on, and NMR analysis confirmed that the G-quadruplex is present in solution and that it is stabilized by DFHBI binding.[45,46] A crystal structure of the fluorophore-free RNA indicated that the overall structure of the aptamer is pre-formed (as did SAXS analysis),[45,46] and that a nucleotide that is extruded from the helical stack in the DFHBI-bound state occupies the fluorophore binding pocket in the absence of the small molecule.

The presence of a G-quadruplex in the fluorophore binding site of a turn-on aptamer can be rationalized on the basis of three properties of this nucleic acid structural motif. First, by virtue of the cyclic symmetry of G-quartets, as well as the axial coordination of a cation, the exposed nucleobases of G-quadruplexes tend to be coplanar. This is in contrast to the ends of Watson-Crick duplexes or triplexes that often exhibit pronounced propeller twist or buckling. As maximal fluorescence activation requires that the two heterocycles of the bound DFHBI (and similar fluorophores) be restrained on one plane, coplanarity of the exposed nucleobases of G-quadruplexes could be advantageous for the binding site of turn-on aptamers. Second, the exposed planar faces of G-quadruplexes are larger than those of duplexes and triplexes, and this allows positioning of additional structural elements on the plane of the fluorophore that is stacking on the quadruplex. This is the case for the unpaired guanine in Spinach, which also hydrogen bonds to the bound DFHBI using a ribose from the RNA backbone. Third, axial cation coordination provides increased thermodynamic stability to G-quadruplexes over simple duplexes. This is especially the case under physiological conditions, because most G-quadruplexes (including Spinach) preferentially bind K^+, which is the major intracellular monovalent cation.

Despite its large fluorescence turn-on and overall brightness, Spinach has several shortcomings as a fluorescent tag for *in vivo* imaging of cellular RNAs. Spinach exhibits low thermostability and a propensity to misfold.[49] These two properties may arise from the complex, non-standard nature of its central G-quadruplex. The structure of Spinach coaxially stacks an A-form duplexes on each side of a G-quadruplex. This necessarily implies complex connectivity of the RNA main chain, since it must transition from an antiparallel duplex (A-form), to a 4-stranded structure (G-quadruplex), back to an antiparallel duplex. Indeed, the G-quadruplex of Spinach is among the most structurally complex described to date, comprised of two G-tetrads, one non-canonical quartet, several nonconsecutive guanines, a combination of parallel and antiparallel connectivity, and variable loop lengths, the longest of which is 34 nucleotides. In addition, Spinach-DFHBI exhibits poor photostability, losing fluorescence rapidly upon illumination (however, in the dark and in the presence of DFHBI, Spinach can rapidly recover fluorescence). This may arise from relatively weak binding to the fluorophore (K_d = 537 nM), and the fact that the fluorophore is exposed to bulk solvent. Lastly, as the autofluorescence of cells is in the blue-green range of the visible spectrum, fluorescent tags that are more red-shifted would be preferable.

Several experiments to improve Spinach properties through further SELEX as well as functional reselection have been reported. The melting temperature (T_m) of Spinach is 34 °C, limiting its application in live cells much as the limited thermal stability of wild-type GFP does. Mutational optimization by reducing bulges and mismatches, resulted in Spinach2.[50] Although Spinach2 has a T_m of 38 °C and was 2.8-fold brighter than Spinach at 37 °C, still only 37% of the improved aptamer was folded at this

temperature. To improve its *in vivo* performance, Spinach2 was fused to $tRNA^{Lys3}$, using it as a folding scaffold (tSpinach2). This increased the fraction of folded aptamer to 60% at 37 °C and overall brightness by 3 to 20-fold relative to Spinach. Broccoli is an improved variant isolated by further *in vitro* and *in vivo* selection,[51] and a selection experiment based solely on *in vivo* fluorescence also yielded a better variant.[52] A selection experiment employing microfluidic-assisted *in vitro* compartmentalization yielded another variant termed *i*Spinach that is optimized for *in vitro* applications.[53] Despite these efforts, the improved variants retain the fundamental structure of Spinach, with its complex fold and modest affinity for fluorophore. In order to achieve drastic improvements, several groups have turned to independent RNA scaffolds and different conditional fluorophores.

1.2.3 Corn and Chili Activate Red-shifted FP-related Fluorophores

Initial efforts to achieve red-shifted emission relative to Spinach led to the DFHBI derivatives DFHBI-1T and DFHBI-2T that contain a trifluoromethyl moiety. In conjunction with the Spinach or Broccoli aptamers these modified conditional fluorophores yielded 35 nm and 53 nm bathochromic shift in excitation and 4 nm and 22 nm bathochromic shift in emission wavelengths, respectively, which are better suited to existing filters designed for GFP fluorescence.[54] When the co-crystal structures of Spinach were determined, it was hypothesized that the poor photostability of Spinach-DFHBI may be a consequence of the relatively sparse hydrogen bonding network between the RNA and the fluorophore, and that if the fluorophore were to include additional hydrogen bonding "handles", photostability may be enhanced. To this end, 3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime (DFHO) was synthesized, which augments DFHBI by addition of an oxime moiety. *In vitro*

selection for aptamers that selectively bound DFHO yielded Corn, a ~30 nt aptamer that is red-shifted relative to Spinach and Broccoli, exciting at 505 nm at emitting at 545 nm, and which is ~1000 times more photostable than Spinach.[55] A parallel experiment in which variants of Spinach that recognize DFHO were selected yielded Orange Broccoli and Red Broccoli, which while being red-shifted relative to Broccoli-DFHBI, exhibit rapid photobleaching, indicating that the photostability of Corn-DFHO is a property of the RNA aptamer, rather than the fluorophore.

Crystallographic structure determination revealed that, unexpectedly, Corn is a homodimer that binds one molecule of DFHO at the inter-protomer interface (Figure 1.4A,B).[56] Subsequent biophysical analyses confirmed that the aptamer is a stable homodimer, and that the fluorescent species has a 2:1 RNA:DFHO stoichiometry. Each Corn protomer folds into a mixed quadruplex, composed of two G-quartets and two mixed sequence quartets. A G-quartet from each protomer flanks the bound DFHO, restraining it into a planar conformation (Figure 1.4C). Notably, the homodimer interface lacks any inter-protomer base pairing and is comprised only of three unpaired adenines from each RNA and the bound DFHO. One of these adenines makes a bidentate contact with the oxime moiety that characterizes DFHO, while another makes bifurcated hydrogen bonds bridging both heterocycles of DFHO. Because in the Corn-bound DFHO its two heterocycles are coplanar, the fluorophore has a plane of mirror symmetry that bisects the $(Corn)_2$-DFHO complex. However, since Corn is a chiral molecule, its dimer cannot have a plane of mirror symmetry. This paradox is resolved by quasisymmetry,[57] whereby the two RNA protomers are locally asymmetric at the DFHO binding interface, even though their overall structure is nearly indistinguishable.

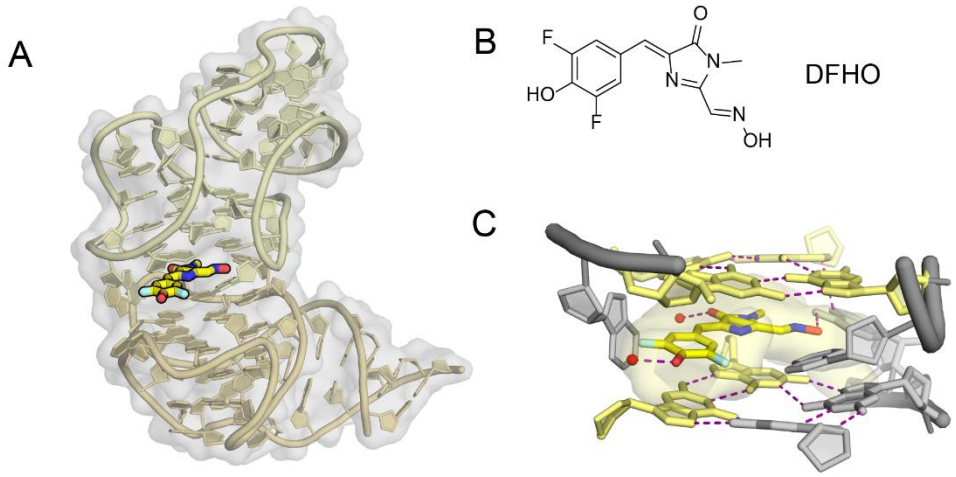


Figure 1.4. Structure of Corn aptamer. (A) Homodimer structure of Corn aptamer with DFHO ligand bound between the G-quadruplex faces of each Corn protomer.[56] (B) Chemical structure of 3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime (DFHO), fluorophore-ligand of Corn. (C) The fluorophore pockets DFHO-bound Corn. Hydrogen bonds are depicted by purple dashed lines and stacking interactions in the fluorogenic RNA aptamers are highlighted in yellow. (PDB: 5BJ0)

In an independent study, analogues of HBI featuring extended aromatic and cationic substituents on the imidazolinone heterocycle were synthesized. When complexed with the compact aptamer "13-2min" that was originally isolated by Jaffrey and co-workers in the same selection experiment that yielded Spinach, three of these new fluorophores (DMHBI-Imi, DMHBI$^+$, and DMHBO$^+$) fluoresced green, yellow, and red, leading the authors to term the complexes Chili.[58] In addition to providing access to a broad spectral range, Chili also has the largest Stokes shift reported for fluorogenic aptamers, ~130 nm for each of the DMHBI analogs. This is reminiscent of fluorescent proteins such as LSSmOrange and LSSmKate,[59,60] which were engineered for their sizeable Stokes shift. While no structure has been reported for Chili, initial NMR characterization of the 13-2min aptamer in complex with DFHBI indicated that this

aptamer also contains a G-quadruplex, that its G-quadruplex is different from that of Spinach, and that binding of fluorophore protects the G-quadruplex from proton exchange, suggesting that the fluorophore is in direct contact with it.[45]

1.2.4 High Fluorophore Affinity of Compact Mango Aptamers

In addition to spectral range, fluorescence turn-on, quantum yield and photostability, an important property of turn-on aptamer complexes of conditional fluorophores is the strength of the RNA-small molecule interaction. A high affinity allows more dilute fluorophore to be employed, and also makes it possible to investigate low copy-number RNAs in live cells. Small size and high affinity were explicit considerations in selecting RNAs that would bind and activate the thiazole orange (TO) derivative TO1-Biotin. TO has a large extinction coefficient, but in its unmodified form, it binds non-specifically to cellular nucleic acids yielding high background.[61] Addition of a substituent to the benzothiazole of TO suppresses non-specific binding, and incorporation of biotin allowed immobilization on streptavidin beads for *in vitro* selection. Initial selection experiments yielded the ~30 nt Mango-I aptamer (also 'RNA Mango') which bound its cognate fluorophore with ~3 nM K_d, and induced orange fluorescence with a quantum yield of 0.14.[62] Mango-I also binds the variant fluorophore TO3-Biotin, which has extended conjugation, albeit with decreased affinity. The Mango-I complex of TO3-Biotin is strongly red-shifted.

The crystal structure of Mango-I in complex with TO1-Biotin revealed a mostly parallel 3-tiered G-quadruplex flexibly linked to an A-form duplex (Figure 1.5A,B).[63] The fluorophore binds to one of the G-quadruplex faces, and is restrained by three flap-like unpaired nucleotides (Figure 1.5C). Unexpectedly, the structure revealed that all

elements of TO1-Biotin, including the fluorescent TO headgroup, the biotin, and the intervening PEG linker, are closely packed against the G-quadruplex, adopting a circular arrangement where the Biotin packs against the TO. A consequence of this arrangement is that the two heterocycles of TO (a benzothiazole and a methylquinoline) are bound by the aptamer making a 45° angle. As with the *p*-HBI family of fluorophores, TO is maximally fluorescent when its two heterocycles and the connecting methine carbon are coplanar. Thus, the modest quantum yield of Mango-I may be a result of this mode of fluorophore binding.

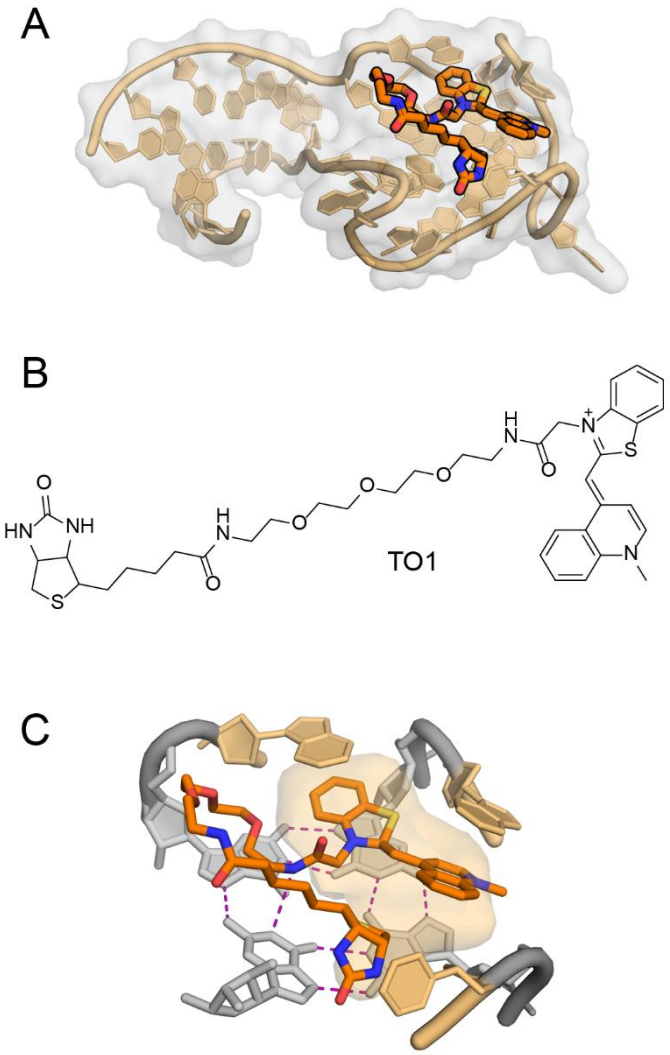


Figure 1.5. Structure of Mango-I aptamer. (A) Mango-I structure with TO1-biotin fluorophore bound in the G-quadruplex core, flanked by a short helix.[63] (B) Chemical structure of TO1-biotin, fluorophore-ligand of Mango-I aptamer. (C) The fluorophore pockets TO1-bound Mango-I. Hydrogen bonds are depicted by purple dashed lines and stacking interactions in the fluorogenic RNA aptamers are highlighted in orange. (PDB: 5V3F)

To improve the functional properties of Mango-I, the selection pool from which this aptamer was isolated was subjected to functional (*i.e.*, fluorescence-based) microfluidic reselection.[64] Three close variants of the parental aptamer were isolated: Mango-II,

Mango-III and Mango-IV. These all had improved brightness, and it was estimated that four individual aptamer molecules were sufficient for visualization in cells. Structure determination of the Mango-II aptamer in complex with TO1-Biotin and TO3-Biotin revealed a molecule with a simple overall architecture similar to that of Mango-I (a three-tiered G-quadruplex and an A-form duplex), but with a more open ligand binding pocket in which the TO headgroup of the fluorophore lies in a near-planar conformation, consistent with an improved quantum yield of 0.21 (for TO1-Biotin). The open pocket also accommodates the TO3 headgroup, explaining the similar affinity of both fluorophores, 1.1 and 1.3 nM). Biotin is not present in the ligand binding pocket, and while most of the PEG linker appeared to be excluded from the pocket, it was shown that its interactions with the RNA contribute to affinity and selectivity.[65]

Thermal melt analysis of the four related Mango aptamers revealed that only Mango-III exhibited a cooperative, biphasic melting profile, suggestive of tertiary structure. The co-crystal structure of the ~30 nt Mango-III aptamer in complex with TO1-Biotin revealed a remarkably complex structure for an RNA of this size, distinctly different from those of Mango-I and Mango-II.[66] Mango-III is also organized around a G-quadruplex (albeit of only two tiers), but incorporates two duplexes: one A-form analogous to those of Mango-I and Mango-II, and a second, non-canonical duplex formed between a propeller loop of the G-quadruplex and nucleotides outside the G-quadruplex. This connectivity, unprecedented for a G-quadruplex, is reminiscent of how pseudoknots arise from the connectivity of two duplexes.[67] The A-form duplex of Mango-III, rather than being flexibly linked to the G-quadruplex as in Mango-I and Mango-II, stacks coaxially on a base triple that in turn stacks on the G-quadruplex.

Rather than unpaired flaps, the ligand binding site of Mango-III is closed by a long-range Watson-Crick base pair (an unusual *trans* pair), and the TO headgroup of the fluorophore is constrained into a planar conformation between this tertiary pair and the G-quadruplex. Consistent with the near optimal conformation of the fluorophore, the quantum yield of the Mango-III-TO1-Biotin complex is 0.55. Mutation of the *trans*-Watson-Crick pair into a homopyrimidine pair further improved the quantum yield to 0.65. This mutant Mango-III in complex with TO1-Biotin is 50% brighter than EGFP.

1.2.5 Fluorogenic Aptamers without G-quadruplexes

Just as the 11-stranded β-barrel is not the only way to achieve protein fluorescence, some fluorescence turn-on RNAs are not organized around G-quadruplexes. A promiscuous aptamer, DIR2s, which turns on both dimethylindole red (DIR) and oxazole thiazole blue (OTB) allows imaging at both the blue and red edges of the visible spectrum.[68] Due to the high extinction coefficients of both conditional fluorophores and the sizable fluorescence turn-on, these are the brightest fluorogenic RNA aptamer-fluorophore complexes reported. Crystallographic structure determination demonstrated that the fluorophore binding site of DIR2s is not organized around a G-quadruplex (Figure 1.6A).[69] Instead, the OTB fluorophore is bound between a two-tiered stack of base triples and one unpaired adenosine (Figure 1.6B,C). Despite the potential for hydrogen bonding of OTB, its primary interaction with DIR2s is through relatively limited stacking, with hydrogen bonding limited to between the sulfonate group of OTB and a guanine in one of the base triples. The lack of extensive stacking or hydrogen bonding between the aptamer and OTB may explain the promiscuity of this turn-on aptamer for chemically similar ligands, including DIR and even TO1-biotin. However, as structures of DIR2s in complex with other ligands have

not yet been determined, the possibility remains that other fluorophores bind to a different sites, and interfere allosterically with binding to a second fluorophore.

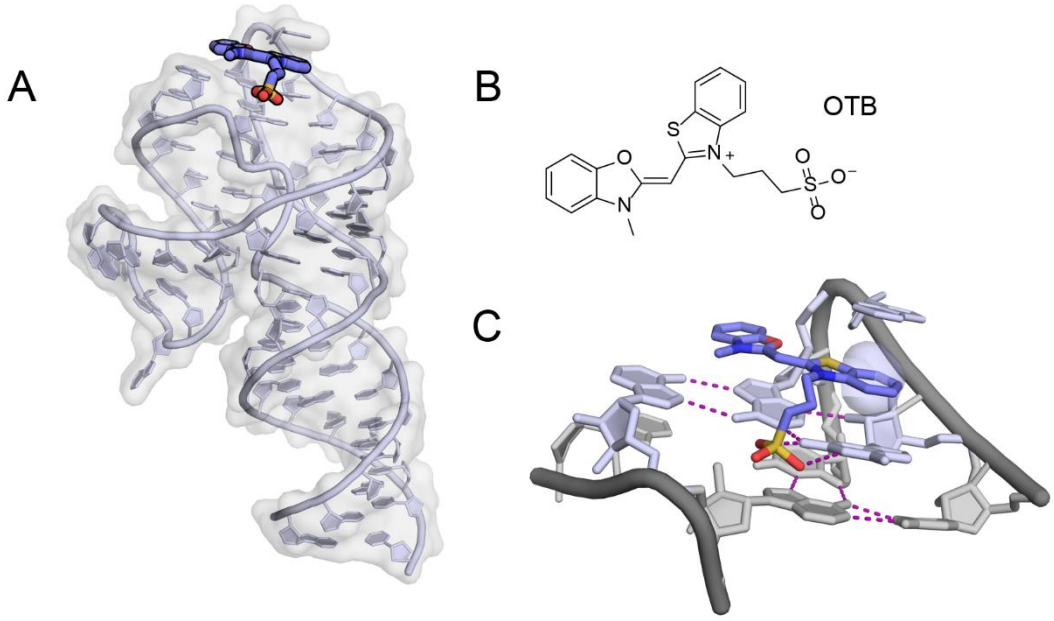


Figure 1.6. Structure of DIR2s aptamer. (A) DIR2s structure with bound OTB fluorophore between base triple and single unpaired nucleotide in the terminal region of aptamer.[69] (B) Chemical structure of oxazole thiazole blue (OTB), fluorophore-ligand of DIR2s aptamer. (C) The fluorophore pockets OTB-bound DIR2s. Hydrogen bonds are depicted by purple dashed lines and stacking interactions in the fluorogenic RNA aptamers are highlighted in light blue. (PDB: 6DB8)

Riboglow is the result not of *in vitro* selection from random RNA library, but of engineering of a bacterial riboswitch that evolved to bind cobalamin, primarily through shape complementarity in an interhelical junction.[70,71] Riboglow capitalizes on the fluorescence quenching properties of cobalamin. A series of fluorophores linked to cobalamin were synthesized that are poorly fluorescent in free form. Upon binding by the cobalamin riboswitch, the quencher is sequestered, thereby resulting in overall

fluorescence turn-on. Most recently, a new aptamer was developed to bind and induce the fluorescence of a new family of fluorophores, (4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenylacetonitrile (HBC), which spans the visible spectrum.[72] This new aptamer, Pepper, enhances the fluorescence of its ligand 3000-fold *in vitro* and notably retains visibly bright fluorescence *in vivo*. While the RNA has not yet been structurally characterized, biochemical studies indicating potassium-independent folding suggest the aptamer does not utilize a G-quadruplex binding motif.

1.3 Applications of Fluorescence Turn-on Aptamers

1.3.1 Imaging

Fusions of FPs to RNA-binding proteins, such as GFP-MS2 have long been employed to visualize RNAs (tagged with the cognate binding site for the RNA-binding protein) *in vivo*.[73] While successful, this approach suffers from low contrast since the fusion protein is intrinsically fluorescent, and thus any RNA-unbound molecules produce background. To overcome this, the binding site is multimerized, which in turn results in the fluorescent tag often becoming as large or larger than the RNA whose cellular behavior is ostensibly being studied. In addition, these multimeric tags are recombinogenic, and have also been shown to adversely affect post-transcriptional processing of the tagged RNAs.[74] Fluorescence turn-on aptamers overcome these limitations by virtue of their small size, and because their fluorophores are conditional, offer in principle very low background (limited by non-specific turn-on through binding to other cellular components). Fluorogenic RNA aptamers have been used to image both coding and noncoding RNA transcripts *in vivo* and *in vitro*.[75–77]

1.3.2 Biosensors

Beyond their use in directly imaging RNA transcripts, fluorogenic RNA aptamers have also been developed to work as biosensors for small molecule metabolites, Förster resonance energy transfer (FRET) sensors, and split fluorescence detectors for RNA-RNA interactions. Reported small molecule biosensors are designed such that the fluorescent RNA aptamer remains unstructured until a fused metabolite-sensing unit binds to its cognate ligand. Binding of the metabolite indirectly stabilizes the fluorophore binding site, triggering fluorescence. This approach has enabled monitoring *in vivo* of ADP, cyclic AMP-GMP, cyclic di-GMP, guanine, SAM, thiamine pyrophosphate, and theophylline.[78–82]

1.3.3 Intermolecular interactions

Many protein biosensors rely on FRET for readout. Recently, FRET between fluorescence turn-on aptamers has been demonstrated, opening the way to this approach. A new analogue of TO3-Biotin (oxazole yellow derivative, YO3-Biotin) was used to develop a FRET system in which Spinach-DFHBI and Mango-YO3-Biotin to report on conformational changes to an RNA scaffold within the cell.[83] The fluorescence of Spinach is also quenched when in close proximity to a protein, allowing the RNA-protein pair to be used to study RNA-protein interactions.[84] As the library of available fluorogenic RNA aptamers expands, it is conceivable that new RNA FRET pairs will be developed, providing a more extensive toolbox for studying intermolecular RNA interactions. Beyond FRET, split fluorogenic RNA aptamers have also been developed as useful tools for studying RNA-RNA or RNA-DNA interactions. Split Spinach and Broccoli Florets have both been developed to monitor RNA assembly *in vitro* and *in vivo*, respectively.[85,86] Furthermore, Split Spinach has been

developed to act as a sequence specific nucleic acid probe,[87] to report on ribozyme self-cleavage activity,[88] and to study RNA-RNA assembly *in vitro*.[89]

1.4 Potential for Optimizing Fluorogenic RNA Aptamers

At a global level, a striking difference between FPs and fluorescence turn-on aptamers is that while the former mostly are variants of one fold (the 11-stranded β-barrel), the latter adopt diverse, apparently unrelated 3D-architectures. The diversity of folds of fluorescence turn-on RNA aptamers opens the way to molecular engineering efforts aimed at optimizing a particular feature of interest to the researcher.

This is most obviously the case in the choice of fluorophore. Since the development of the first generation Spinach aptamer, RNA aptamers for several fluorophores have been developed to span the UV-visible spectrum (Figure 1.7). For use *in vivo*, efficient cellular uptake, lack of interaction with other cellular components (to provide low background fluorescence and molecular specificity), and low cytotoxicity (especially under illumination) are primary considerations. Some of the FP derived fluorophores such as DFHBI and DFHO meet all these criteria, although their extinction coefficients are only modest. Other fluorophores, such as those employed by Mango[62] and Riboglow[70] have higher extinction coefficients, but carry functional groups (biotin and cobalamin, respectively) that could interact in unexpected ways with cellular components.

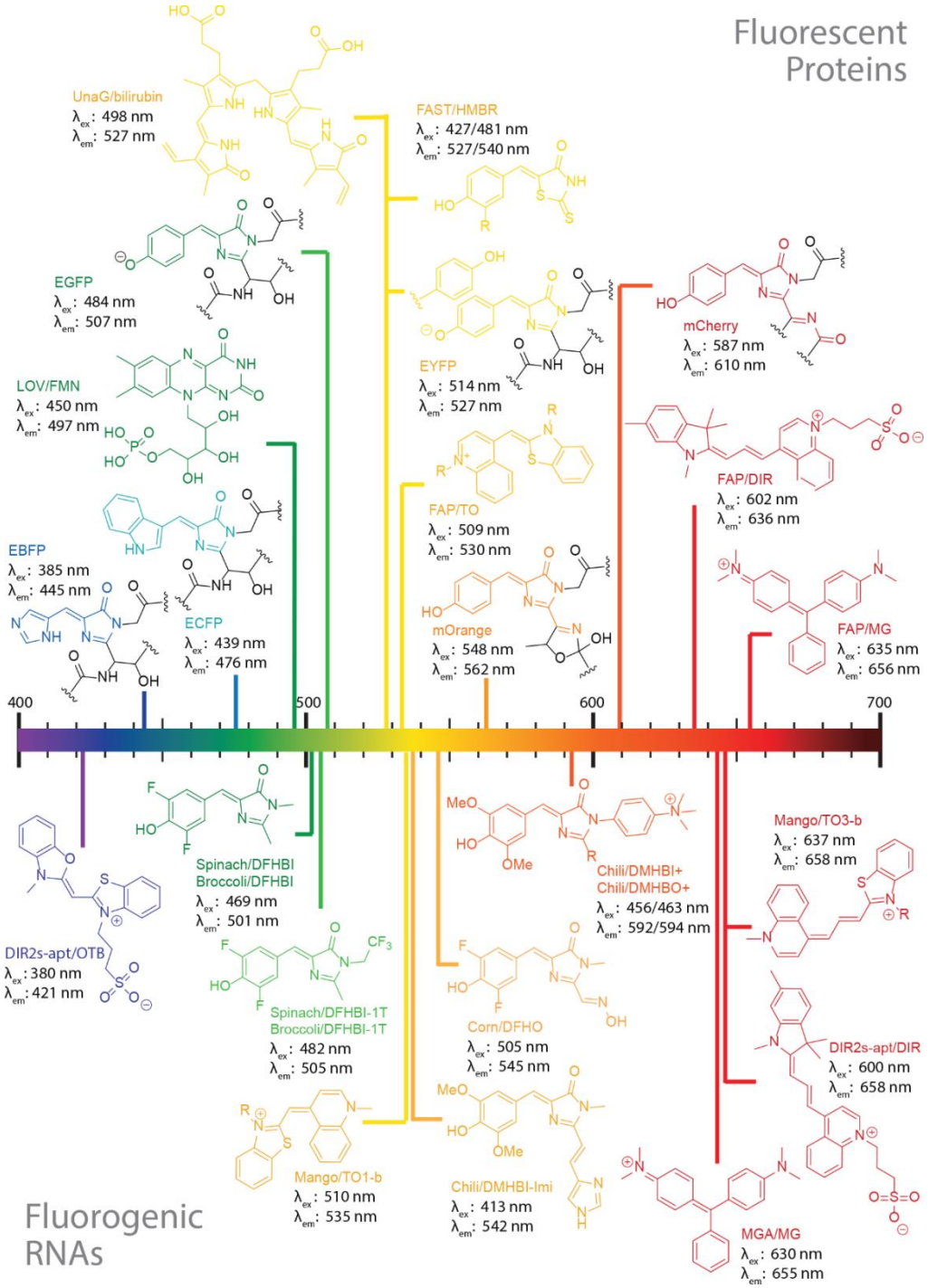


Figure 1.7. Comparison of spectral range for chromophores found in fluorescent proteins (top) and fluorogenic RNAs (bottom). Depicted colors reflect the visual appearance of chromophores in complex with their respective fluorescent protein/RNA, rather than their excitation or emissions wavelengths.

In the case of fluorescence turn-on RNA aptamers, the molecular environment of the fluorophore can be tuned to achieve the specific objectives of the researcher. Thus, while G-quadruplexes are widespread among the aptamers, DIR2s and the MGA show that there are other ways of assembling binding sites that induce fluorescence. Indeed, Spinach flanks its fluorophore with a G-quartet and a base triple on either side, and mutagenesis experiments demonstrate that the fluorophore and the base triple are closely coupled electronically, as changing the nature of the triple results in large absorption and emission wavelength changes.[45,46] G-quadruplexes have been reported to quench fluorescence,[90,91] but also have intrinsic fluorescence[92,93] so that in principle, they could function as antennas for intramolecular FRET. G-quadruplexes are stabilized by binding a cation in their axial pore, and experiments with Spinach revealed that changing the identity of the cation from the physiological K^+ to Ba^{2+} and Sr^{2+} resulted in spectral shifts.[45] How other RNA functional groups modulate fluorescence is just starting to be studied,[94] but indications are that both fluorophore-proximal and -distal groups can be important. Thus, in the case of Mango-III, mutation of the lone Watson-Crick base pair that constrains the fluorophore against the core G-quadruplex from A•U to U•U resulted in a ~18% increase in quantum yield. Functional reselection of Mango-III produced *i*Mango-III which achieves a comparable quantum yield but is blue-shifted by 4-7 nm. Mutational analysis indicates that this hypsochromic shift is not a result of changes in the immediate environment of the fluorophore. Rather, it is a function of nucleotides that are at least 10 Å away.[66] This is reminiscent of the effect of mutations away from the active sites of globular proteins on catalytic activity. The complex interplay of RNA structure and fluorescence

properties of RNA aptamers represent an experimental challenge, but also provides multiple avenues for fine-tuning the properties of these powerful molecular tools.

Structure-guided design efforts have met with moderate success, yielding more conservative changes in the biophysical and biochemical properties of the aptamers. Upon the structural characterization of Spinach, the aptamer was successfully miniaturized (from 98 nt to 51 nt) by removing nonessential regions of the RNA molecule while retaining full fluorescence turn-on.[45] Following the structural characterization of Mango-III, structure-guided randomization and functional (*i.e.*, fluorescence-based) reselection of nucleotides surrounding the ligand led to a new variant (*i*Mango-III) with 13% fluorescence enhancement and unexpected hypsochromic shift and *shorter* fluorescence lifetime.[66] The apparent disparity between the success of rational engineering of FPs, and the unexpected results of turn-on RNA aptamer development probably reflect the latter efforts being in an earlier phase of development, and also the fact that *in vitro* selection has provided access to a much wider sequence space for RNAs than that explored by the canonical FPs.

1.5 Research Objectives

1.5.1 Use of naturally occurring RNA modifications for spectral tuning

Spectral tuning of fluorogenic RNA aptamers through mutations to the RNA is limited by the availability of only four canonical nucleotides. In this research, I will expand the existing repertoire of available nucleotides by chemically synthesizing naturally occurring RNA modifications for incorporation into fluorogenic RNA aptamers. The modifications of interest are $PreQ_0$, archaeosine, and wyosine. Wyosine was chosen for study as it is the only naturally occurring nucleoside currently known to possess an

intrinsic fluorescence. $PreQ_0$ and archaeosine were chosen as the nitrile and amidine group were functionally interesting and both nucleosides could be accessed fromm similar synthetic schemes. While the synthesis of these nucleosides has previously been reported in the literature, their phosphoramidite forms, which are necessary for solid-phase RNA synthesis, have never been reported.

1.5.2 Studying nucleoside and fluorophore analogues in Spinach aptamers

Building upon previous research demonstrating shifts in absorbance and fluorescence spectra by mutating the base triple in Spinach, I will explore other combinations of base triple mutations. In addition to these canonical mutations and the study of naturally occurring RNA modifications, I will also investigate the effects of artificial nucleosides such as 2-aminopurine (AP) and pyrrolocytosine (PyC) on the fluorescence of Spinach aptamers when incorporated into sites surrounding the ligand binding pocket of the RNA. These studies will be further complemented by investigating the dual effects of concurrently modifying the RNA and the fluorophore ligand.

1.5.3 Remote excitation in Mango aptamers using fluorescent dye labels

As 2-aminopurine and pyrrolocytosine are weakly fluorescent, I will study their capacity to form a FRET pair with a TO1-bound Mango aptamer. Moving beyond nucleoside analogues, I will also investigate the use of fluorescent dyes as a FRET donor or acceptor when covalently linked to a Mango aptamer. Looking specifically at fluorescein and rhodamine labels paired with TO1-bound Mango-III and YO3-bound Mango-III, successful FRET between the fluorescent dyes and fluorogenic RNA aptamers would form the basis of an internal FRET antenna. This enables the

possibility of remote excitation of the aptamer, providing a straight-forward method for tuning its excitation or emission wavelength.

CHAPTER 2. SYNTHESIS OF tRNA MODIFICATIONS FOR FLUORESCENT TUNING

Fluorogenic RNA aptamers can be spectrally tuned by modifying the aptamer's cognate ligand or by introducing mutations to the aptamer. Mutations to the fluorogenic aptamer are limited by the availability of sites surrounding the ligand binding pocket that are not essential to folding and the existence of only four possible base substitutions. To address this, I have developed synthetic access to noncanonical, post-transcriptional tRNA modifications that can be site selectively incorporated into RNA using solid phase oligonucleotide synthesis.

While over one hundred tRNA modifications have been discovered in nature,[95–97] their study has been limited due to the inaccessibility of the modified RNAs. Traditional extraction of modified tRNAs from natural sources requires extensive purification and is infeasible in the worst cases. Chemical synthesis of modified tRNA using solid-phase RNA oligonucleotide synthesis is one route to access these otherwise difficult to obtain biomolecules.[98] However, the use of chemical synthesis to access these modified RNAs depends on the availability of their corresponding phosphoramidites. While some naturally occurring modifications such as formyl-C or psuedouridine are available as RNA phosphoramidite monomers, the large majority of these are only available in their deoxyribose form. Furthermore, uncommon naturally occurring modifications are often not commercially available at all, and many have never even been synthesized.

For these studies, I have chosen to synthesize the phosphoramidite forms of $PreQ_0$, archaeosine, and wyosine (Figure 2.1).[99–101] The synthesis of these phosphoramidites have not been reported in the literature, and thus have never been accessible through solid-phase oligonucleotide synthesis. These modifications are of interest for several

reasons. $PreQ_0$ contains a nitrile group, which produces a peak around 2250 cm^{-1} in infrared (IR) spectroscopy.[102] There are few other groups that produce a signal in this area of the spectrum, making the nitrile group and thus $PreQ_0$ a promising IR probe. Archaeosine is one of few naturally occurring nucleobases which carry a formal positive charge. This would allow me to study of the effects of electrostatic interactions on the chemical environment of the fluorophore-binding pocket of RNA aptamers, as charge differences have been shown to shift fluorescence signals in fluorogenic proteins and RNAs.[58] Finally, wyosine is the only naturally occurring RNA modification that is inherently fluorescent.[100] This will enable me to study the effects of extrinsic fluorescence on the fluorescent signal of the bound fluorophore-ligand to a fluorogenic RNA aptamer.

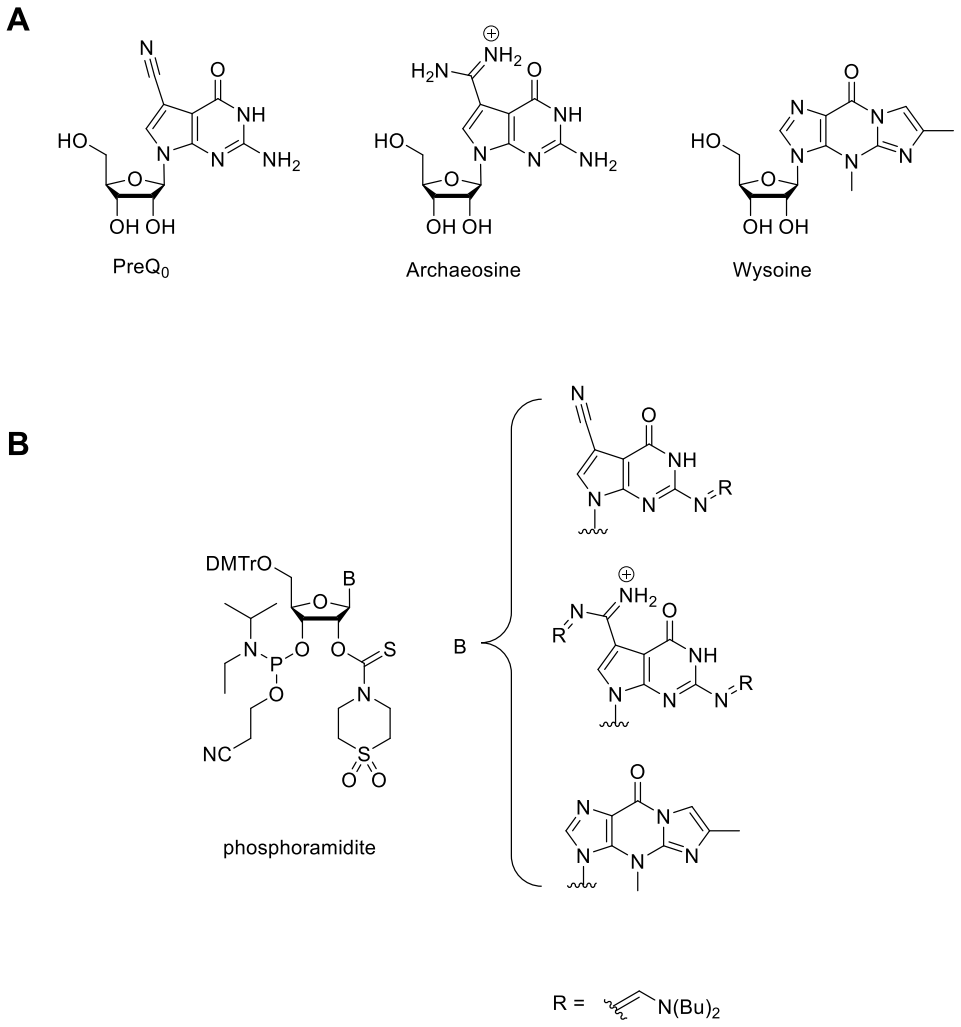


Figure 2.1. Target tRNA nucleosides (A) and corresponding phosphoramidite forms (B). The phosphoramidite forms of these nucleosides are necessary for site selective incorporation into RNA oligonucleotides using solid-phase synthesis.

The synthesis of the phosphoramidite forms of these naturally occurring, modified RNA nucleosides are then suitable for solid-phase RNA oligonucleotide synthesis (Figure 2.2). Unlike enzymatic transcription, solid-phase synthesis proceeds in the 3′ to 5′ direction. The first nucleotide is attached to a solid support at the 3′ hydroxyl, and the 5′-DMT group is removed via treatment with trichloroacetic acid. Next, the appropriate phosphoramidite for the desired sequence (A, G, C, or T) is introduced and activated

with tetrazole, resulting in coupling of the two monomers. The coupled product is then oxidized, and again the 5′-hydroxyl of the newly introduced nucleotide is deprotected. This cyclic process is repeated, and phosphoramidites are added singularly until the desired sequence is obtained. At this point, the oligonucleotide chain can be cleaved from solid support, and deprotected to yield an oligonucleotide strand. Due to the sequence selectivity of this process, modified phosphoramidites can be precisely incorporated in positions of interest.

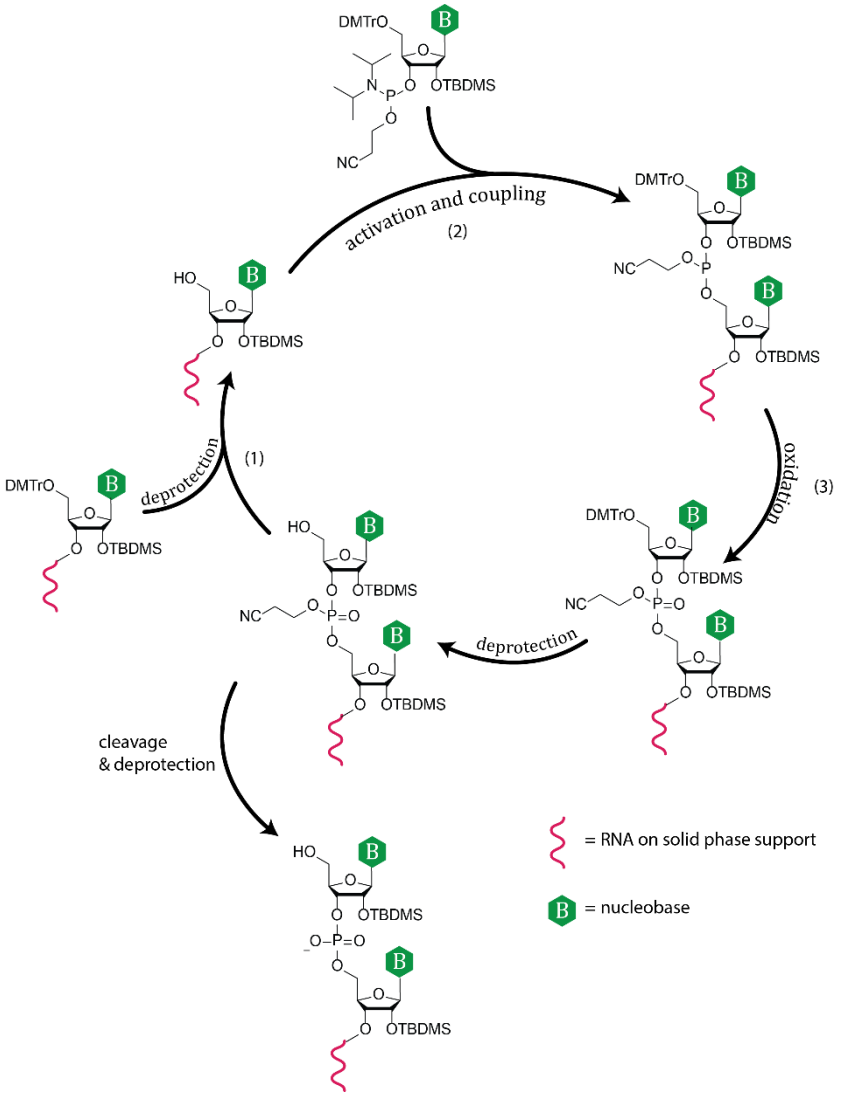


Figure 2.2. Solid-phase RNA synthesis cycle. Each cycle begins with (1) deprotection at the 5′ hydroxyl, proceeds to (2) activation and coupling with the next phosphoramidite monomer, then (3) oxidation of the phosphite. After an oligonucleotide chain is built, it can be cleaved and deprotected. Figure adapted from Hughes, et. al. *Methods in Enzymology* (2011)[103]

By developing the phosphoramidites for these modified RNA nucleosides, I will be able to study how modifications to the RNA aptamer, rather than the fluorophore ligand, can affect the spectral properties of fluorogenic RNA aptamers. Using naturally occurring modifications also maintains the possibility of translating this proof of principle study into an *in vivo* environment.

The nucleosides of $PreQ_0$, archaeosine, and wyosine have been previously synthesized, but their corresponding phosphoramidites have not been reported.[99,101] I will discuss the synthesis of the novel phosphoramidites for these tRNA nucleosides. As a number of the previously reported literature steps proved difficult or inefficient, I also explored alternative synthetic pathways and optimized the use of protecting groups for subsequent phosphoramidite chemistry.

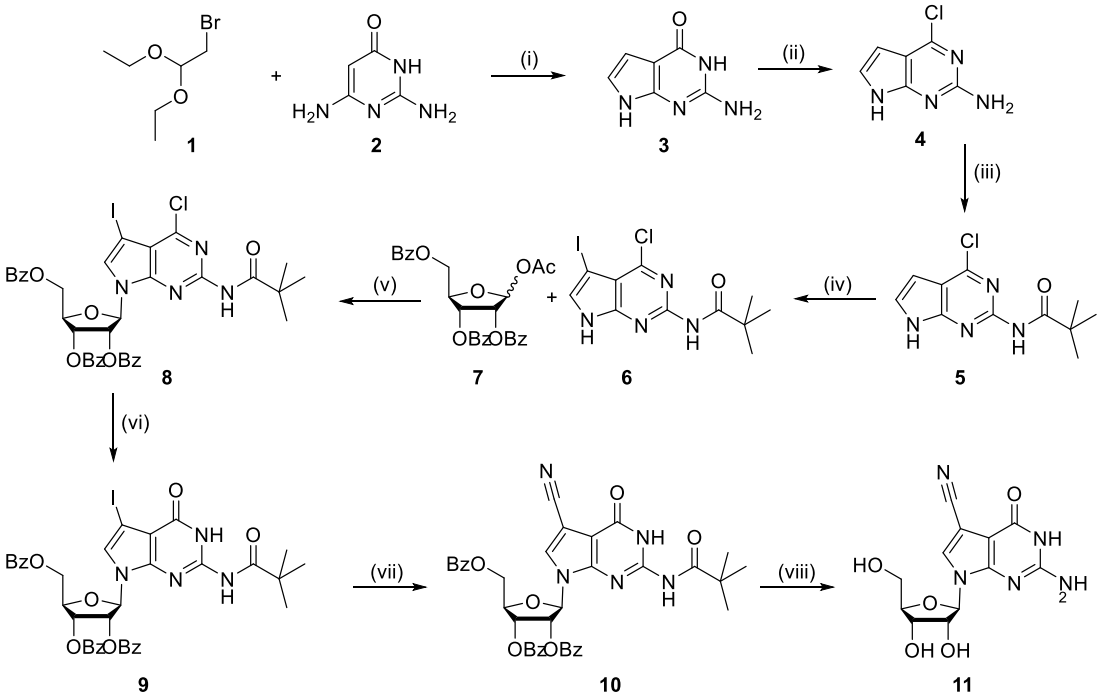
2.2.1 Synthesis of nucleoside precursors

Nucleosides are chemically accessible using one of two general schemes: nucleobases can be built around a C-1′ substituted sugar, or a nucleobase and sugar can be covalently joined by a glycosidic bond formation. In the case of C7 substituted guanosine analogues such as $PreQ_0$ and archaeosine, glycosidic bond formation offers the most direct path forward. However, the synthesis of wyosine is most efficiently accomplished by beginning with guanosine and extending the purine backbone to achieve wyosine's tricyclic structure.

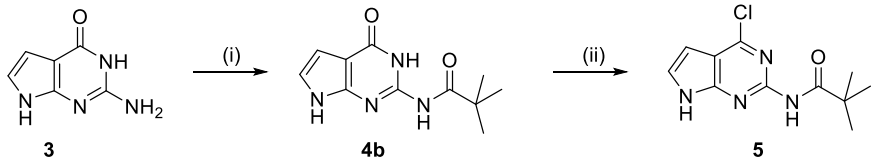
2.2.1.1 Synthesis of $PreQ_0$ nucleoside

The synthesis of the $PreQ_0$ nucleoside could be achieved over 8 steps (Scheme 1). **3** was successfully prepared from diamine substituted pyrimidinone **2** and bromoacetaldehyde diethyl acetal **1** as previously reported.[99] The subsequent chlorination of this guanine derivative **4** following literature procedures proved to be difficult to purify resulting in low yields in the subsequent pivaloyl protection step. This was due to the inherent insolubility of guanine and its analogues in most organic solvents. To circumvent this, I introduced a pivaloyl protecting group earlier in the

synthetic pathway to access **4b**, improving the solubility of the intermediates in organic solvents (Scheme 2). This improved the overall yield of compound **5** from 10% to 26%. Additionally, while literature methods have successfully recovered **5** from precipitation during work up, column chromatography was consistently found to give higher yields with improved purity. Subsequent iodination at the C7 position to give **6** proceeded smoothly with high yields. This newly introduced iodo-group provided a handle for future reactivity. Compound **6** was then suitable for glycosylation with **7** to generate a guanosine analogue **8** for further derivatization.



Scheme 1. Synthesis of $PreQ_0$ nucleoside. (i) (1) HCl, H_2O, 80 °C, 30 min (2) NaOAc, H_2O, 90 °C, 2 h, 81%; (ii) $POCl_3$, 120 °C, 2 h, 64%; (iii) pivaloyl chloride, pyridine, 120 °C, 2 h, 16%; (iv) *N*-iodosuccinimide, THF, rt, 1 h, 80%; (v) *N*-*O*-bismethylsilylacetamide, TMSOTf, MeCN, 50 °C, 24 h, 30%; (vi) CsOAc, DABCO, TEA, DMF, 24 h, 59%; (vii) CuCN, DMF, 130 °C, 5 h, 61%; (viii) NH_3, MeOH, 60 °C, 24 h, 91%.



Scheme 2. Alternative synthesis of 4-Chloro-2-pivaloylamino-7H-pyrrolo[2,3-d]pyrimidine. (i) (1) pyridine, 90 °C, 16 h (2) aq. NH_3, MeOH, rt, 30 min, 32%; (ii) $POCl_3$, benzyltriethyl ammonium chloride, *N*,*N*-dimethylaniline, MeCN, 120 °C, 1 h, 82%;

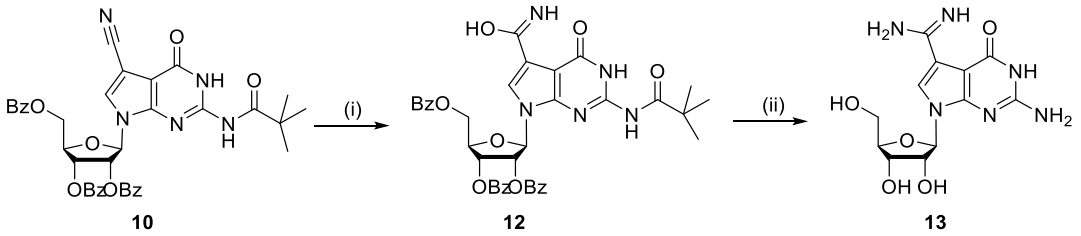
Glycosylation of **6** and **7** using the Vorbrüggen reaction[104,105] generated the 5-chloro-7-deaza-7-iodoguanosine **8**. Initial yields for the glycosylation were low, ranging from 15 – 20%. Despite using a large excess of **7** and extending reaction times to 72 h, the reaction never reached completion and starting material **6** was always observable via TLC. As trimethylsilyl trifluoromethanesulfonate (TMSOTf), an important catalyst for the formation of a key intermediate during the reaction, is water sensitive, I hypothesized the presence of trace amounts of water was inhibiting the glycosylation reaction. To address this, I carefully maintained anhydrous conditions under an argon atmosphere and added TMSOTf portion-wise. These additional conditions resulted in the complete consumption of the starting material, as evidence by TLC, and increased yields to 34%. The complete consumption of starting material and isolated yield of 34% indicates that **7** or **8** is likely decomposing possibly from the presence of residual triflic acid. It is also possible that the reaction was also hindered by the steric bulk of the benzoyl protecting groups on the ribofuranose, and that reaction times can be reduced and yields further improved by using 1,2,3,5-tetra-*O*-acetyl-β-D-ribofuranose which has less bulky acetyl protecting groups. In the interest of time, this reaction was not further optimized and the generated 5-chloro-7-deaza-7-iodoguanosine intermediate **8** was taken forward.

The substitution of the chloride with a carbonyl proceeded smoothly. Although literature reports indicate reaction times of 40 h,[99] the conversion of the chloro intermediate **8** to the ketone **9** was complete after 24 h. Yields for this reaction were consistently ~60%. Nitrile exchange was initially attempted by converting the iodo at the C7 position into a Grignard reagent to displace the tosylate group in *p*-toluenesulfonyl cyanide (TosCN). MeMgCl was used to deprotonate the nucleoside, and a Turbo-Grignard was used to convert the nucleoside into a Grignard reagent capable of attacking TosCN.[99] Using this method, several attempts to access **10** were unsuccessful. Instead, the iodo could be displaced with a nitrile group using copper cyanide in DMF under reflux. These reaction conditions were more straightforward, requiring only two reagents, and resulted in 61% yield after 24 h. In parallel, I investigated the use of microwave assisted iodo-nitrile exchange[106] on compound **9** to generate the cyano substituted guanosine derivative **10**, which uses copper cyanide and high temperatures to induce nucleophilic attack. The reaction was complete within 5 h and yields from the microwave assisted iodo-nitrile exchange using copper cyanide were consistent with those from reflux. Following the nitrile exchange, $PreQ_0$ was accessed from compound **10** by deprotection with ammonia. While literature methods for the complete deprotection of **10** were carried out in 28% NH_3 in H_2O, removal of the pivaloyl and benzoyl protecting groups was found to be more efficient in 4M NH_3 in MeOH and resulted in 91% yield.[99] This nucleoside was then suitable for conversion into an RNA phosphoramidite for solid-phase synthesis.

2.2.1.2 Synthesis of Archaeosine tRNA nucleoside

The synthesis of archaeosine could be easily achieved using pivaloyl protected $PreQ_0$ as a synthetic precursor (Scheme 3). The Pinner reaction was used to convert the nitrile

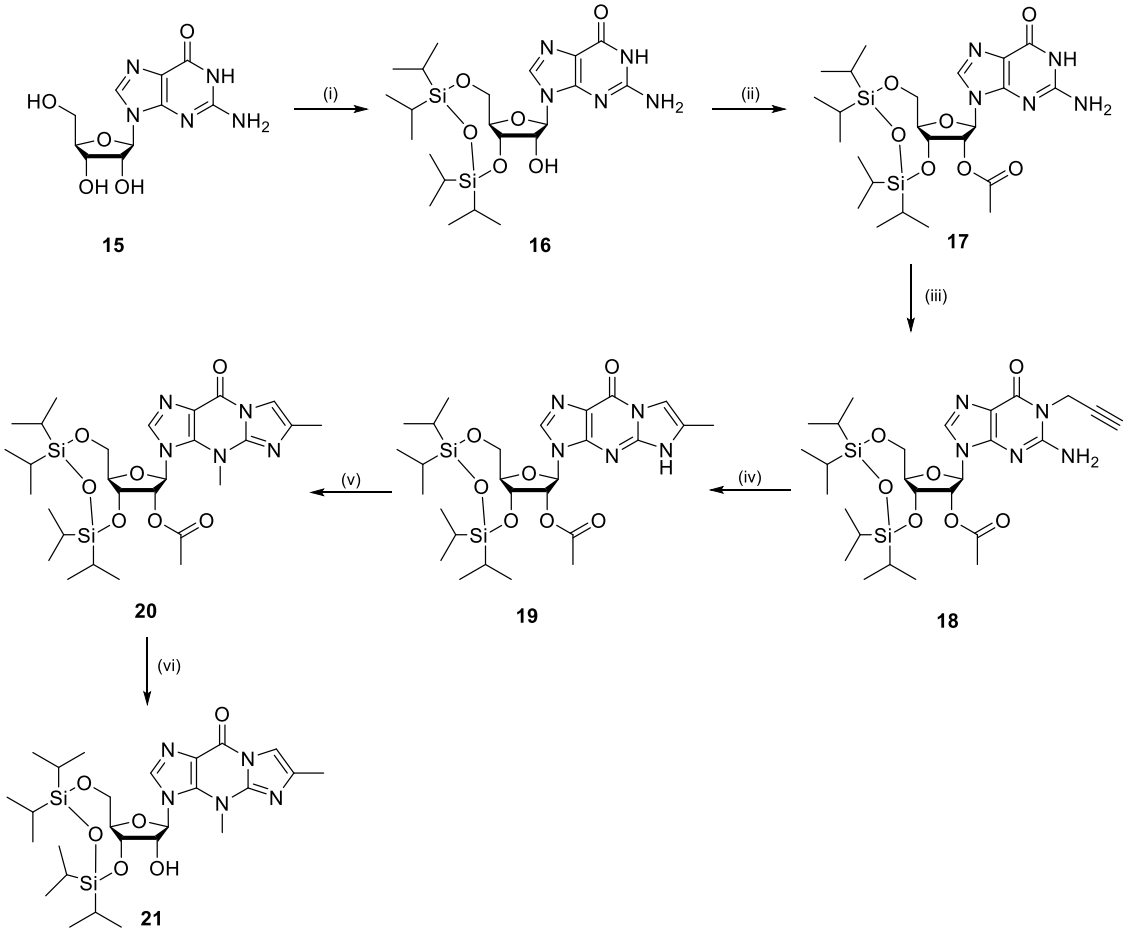
group on $PreQ_0$ into iminoether **12**.[99] This could then be converted into amidine **13** by treatment with 4M NH_3 in MeOH. While archaeosine was unable to be purified by standard silica gel-based column chromatography, it recrystallized from water in high purity resulting in 84% overall yield. This easy conversion from $PreQ_0$ to archaeosine provides convenient access to two naturally occurring guanosine analogues.



Scheme 3. Synthesis of Archaeosine nuceloside. (i) (1) HCl, MeOH, 0 °C, 5 min; (ii) NH_3, MeOH, 16 h, 84%.

2.2.1.3 Synthesis of Wyosine tRNA nucleoside

Unlike $PreQ_0$ and archaeosine which require formation of a new glycosidic bond, wyosine is most easily accessed by building upon guanosine **15** (Scheme 4). Before modifying the nucleobase moiety of guanosine, it was important to protect the ribose hydroxyls from reactivity. While literature methods use the acetyl protecting group,[101] I chose to use a combination of silyl and acetyl protecting groups for greater control over 2′ and 3′ hydroxyl differentiation. I optimized this decision specifically for TC phosphoramidite chemistry, which is discussed later (Section 2.2.2.2). Selective protection of the 5′ and 3′ hydroxyls with TIPDS gave **16** in 80% yield. The 3′ hydroxyl could then be protected with an acetyl group, giving **17** in 74% yield. The immediate use of hydrophobic protecting groups also increases the solubility of these guanosine derivatives making future reactions and purifications more straightforward.



Scheme 4. Synthesis of Wyosine nucleoside. (i) $TIPDSiCl_2$, pyridine, DMF, rt, 16 h, 80%; (ii) acetic anhydride, anh. MeCN, DMAP, TEA, 0 °C for 1 h, rt for 3 h, 74%; (iii) (1) NaH, anh. DMF, 0 °C, 1 h, (2) propargyl bromide, anh. DMF, rt, 3 h, 68%; (iv) Zn dust, 1,2-dibromoethane, MeCN, 95 °C, 3 h, 44%; (v) diiodomethane, diethyl zinc, dimethoxyethane, anh. diethyl ether, 45 min, 53%; (vi) NH_3, MeOH, rt, 2 h, 86%.

To extend the carbon backbone of guanosine, I performed an alkylation of the N1 position using sodium hydride and propargyl bromide to give **18** in 68% yield. As the use of propargyl bromide alone would alkylate the exocyclic amine, sodium hydride is needed to first deprotonate the N1 position. This deprotonation activates N1, preferentially forming **18**. The extended carbon backbone provided by the propargyl group can then be cyclized using ZnBr catalyzed hydroamination giving **19** in 44%

yield. This provides the core tricyclic structure of wyosine. ZnOTf was briefly tested as an alternative catalyst for hydroamination to improve yields, but no product was observed via TLC and *in situ* generated ZnBr was used from that point forward. Additionally, access to **19** was attempted following NaH deprotonation using chloroacetone. While a similar reaction has been reported using bromoacetone,[107,108] the same reaction with chloroacetone resulted in the degradation of starting material without an observable production of the target compound. Following generation of **19**, methylation of N3 was achieved by $Zn(Et)_2$ and iodomethane, affording the conjugated and fluorescent protected wyosine nucleoside **20** in 53% yield. The acetyl group can then be selectively removed in ammonia to give **21** in 86% yield. **21** is then suitable for phosphoramidite conversion using TC chemistry.[109]

2.2.2 Synthesis of phosphoramidite forms of tRNA modifications

Monomers for solid-phase oligonucleotide synthesis must contain a variety of selective protecting groups to adhere to the necessary reaction protocol for solid-phase synthesis (Section 2.1). This normally includes a trityl protecting group on the 5′ hydroxyl, the most common being dimethoxy trityl, which can be selectively removed using mild acidic conditions. Exocyclic primary amines are nucleophilic and must also be protected. Common protection schemes include isobutryl, benzoyl, and dimethylformamidyl groups. The 3′ hydroxyl should carry the phosphoramidite moiety that is later oxidized to form the phosphate backbone of the oligonucleotide.

When working with RNA oligonucleotides, the 2′ hydroxyl adds an extra layer of complexity as it must be protected to avoid 2′ to 3′ migration during oligonucleotide synthesis.[110–112] Many protecting groups have been developed for the 2′ position,

though the most commonly used is the tetrabutyldimethylsilyl (TBDMS) group.[113] Others include triisopropylsilyl (TIPS), triisopropylsilyloxymethyl (TOM), 2′-thiomorpholine-4-carbothioate (TC), and 2′-bis(2-acetoxyethoxy)methyl (ACE) (Figure 2.3).[109,113–115] Synthesis of RNA phosphoramidite monomers are difficult due to the necessity of protecting the 2′ hydroxyl and phosphitylating the 3′ hydroxyl, which are chemically similar.

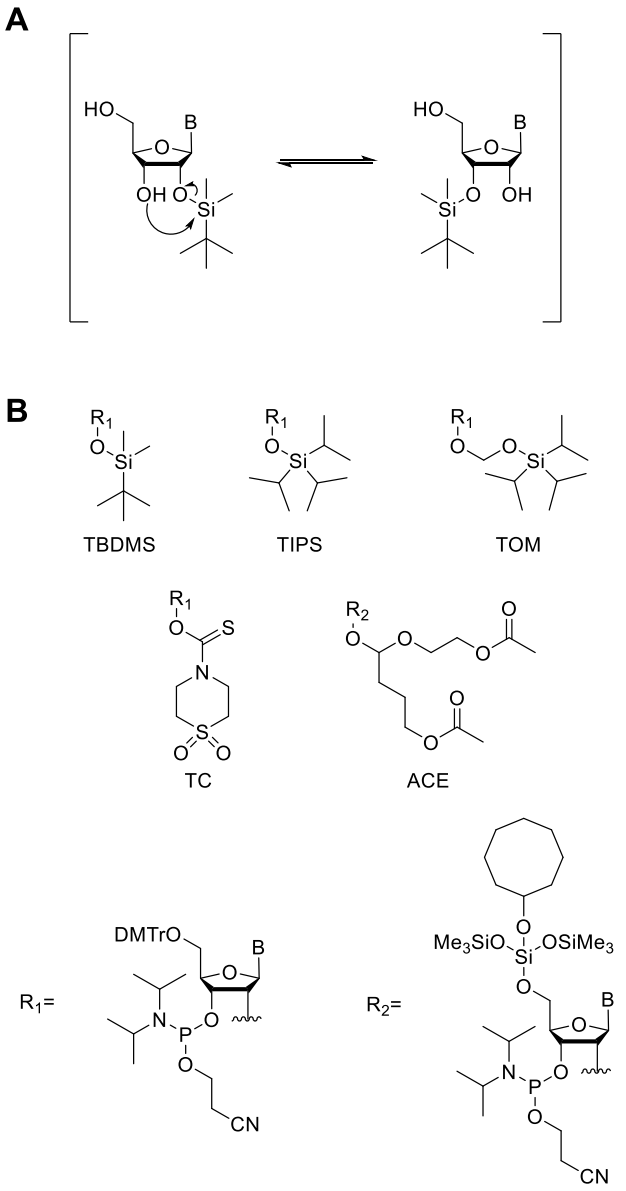
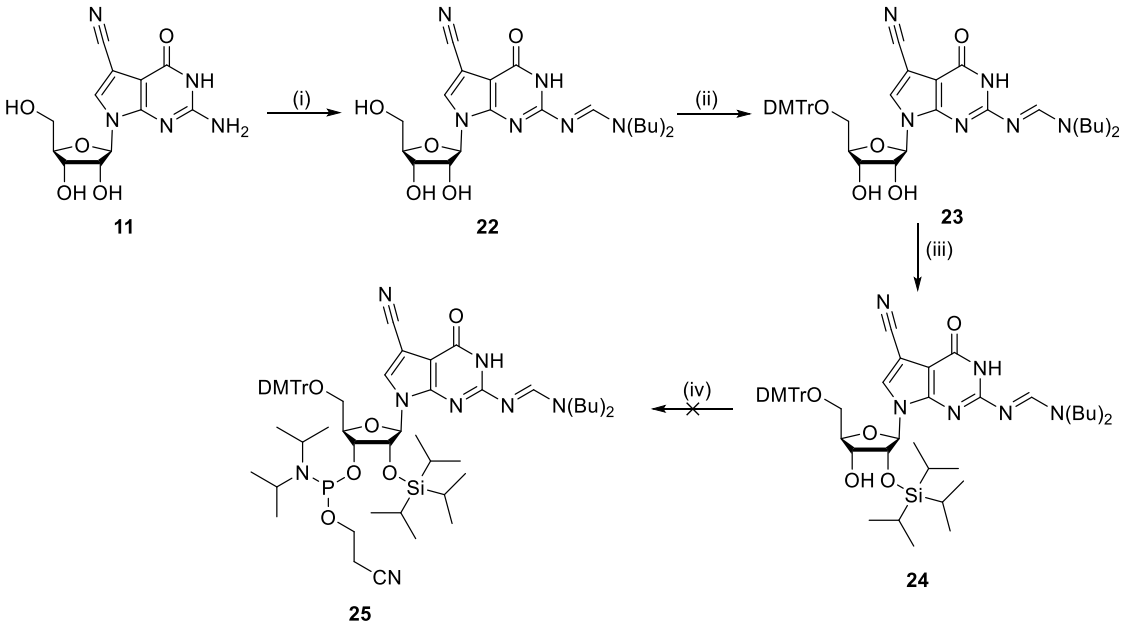


Figure 2.3. RNA monomers must be distinguishable at their 2′ and 3′ hydroxyl positions. (A) When protecting the 2′ hydroxyl, migration between the 2′ and 3′ positions is commonly observed under slightly basic or acidic conditions. (B) Several protecting groups have been developed for the 2′ hydroxyl position for RNA solid-phase synthesis.

It is a common technique to non-selectively protect both the 2′ or 3′ hydroxyl groups with TBDMS, isolate the 2′ protected nucleoside, and re-isomerize the 3′ protected nucleoside and purify again to improve yields. Some nucleosides are less amenable to

this methodology, as the isolation of the 2′-protected isomer from the mixture of isomers can be very difficult depending on how closely the isomers migrate during chromatography. The literature also reports that the split of 2′ and 3′ protected isomers is not precisely a 50/50 split, with the TBDMS protecting group favouring the 3′ position on guanosine.[116] Given the expected difficulty of the protection of the 2′ position and potential subsequent isomerization, I sought to use protecting groups developed to hinder or prevent migration. I thus chose to attempt phosphoramidite synthesis using either the TIPS or the TC protecting group.

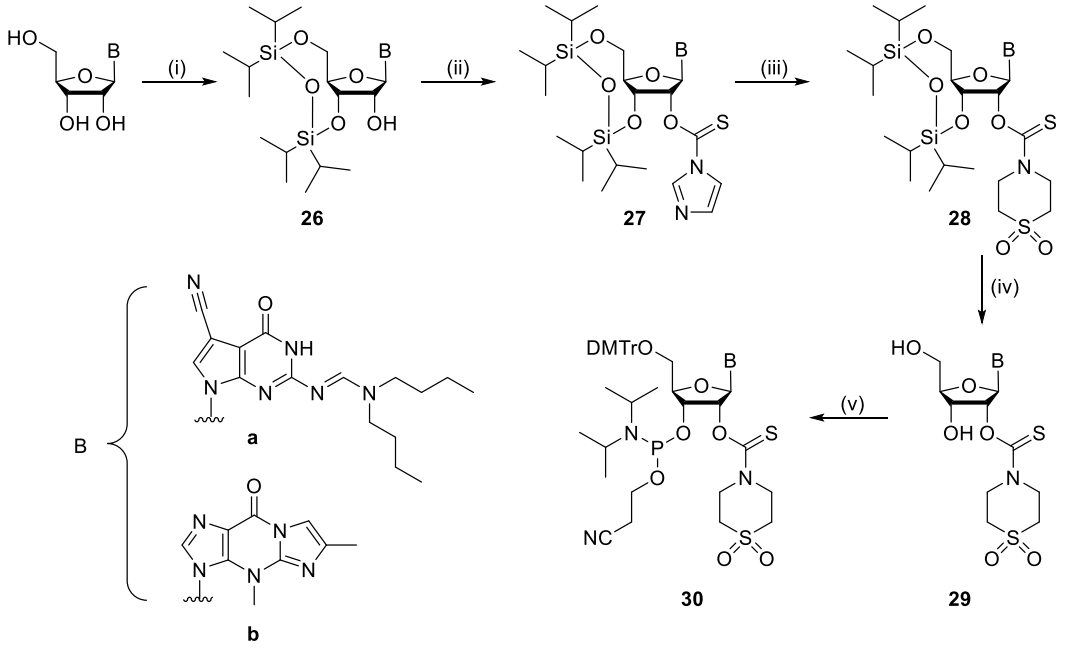
2.2.2.1 Phosphoramidites using TIPS protection



Scheme 5. TIPS protecting group. (i) *N*,*N*-dibutylformamide dimethyl acetal, anh. DMF, rt, 24 h, 70%; (ii) dimethoxytrityl chloride, anh. pyridine, rt, 16 h, 60%; (iii) TIPS-Cl, $AgNO_3$, anh. pyridine, anh. THF, 24 h, 22%; (iv) 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite, DIPEA, anh. DCM, rt, 24 h.

To first protect the exocyclic amine in $PreQ_0$ **11**, I used *N*,*N*-dibutylformamide dimethyl acetal to introduce a dibutylformamidyl group giving **22** in 70% yield. Dimethylformamidyl groups are commonly used as they are selective for primary amines. The long carbon chains on the dibutylformamidyl group has the additional benefit of increasing the overall hydrophobicity of the compound and the protecting group was intentionally selected to avoid potential solubility issues commonly associated with nucleosides. The 5′ position was protected using DMT-Cl to give **23** in 60% yield. The 2′ position was protected using $(i\text{-}Pr)_3SiCl$ in the presence of silver nitrate[116] in order to push the reaction favourably toward 2′ protection over 3′ protection and the 2′ isomer **24** was isolated in 22% yield. With the appropriately protected nucleoside **24** in hand, the phosphitylation was attempted using large excesses of the chloro-phosphitylating reagent with an extended 24 h reaction time. No reaction was observable via TLC nor ^{31}P NMR, and **25** was not accessible using this method. This was most likely a result of the steric hindrance provided by the large isopropyl groups of the TIPS group, which also serves to prevent 2′ to 3′ migration. The TIPS protecting group was not used further.

2.2.2.2 Phosphoramidites using TC protection



Scheme 6. TC protecting group. (i) TIPS-Cl, pyridine, 2 h, 78%; (ii), 1,1′-thiocarbonyldiimidazole, DMAP, MeCN, rt, 16 h, 95%; (iii) thiomorpholine 1,1-dioxide, MeCN, 90 °C, 3 h, 92%; (iv) HF/pyridine, anh. THF, rt, 4 h, 62-75%; (v) 1) DMT-Cl, NMM, anh. DCM, rt, 1 h, (2) 2-cyanoethyl N,N-diisopropylchlorophosphoramidite, anh. DCM, rt, 2 h, 37-64%.

2'-Thiomorpholine-4-carbothioate protecting groups were recently developed to address the low yields and sequence length limitations of solid-phase RNA synthesis relative to DNA.[109] Using 1,2-diamines such as ethylenediamine, both the 2′ thionocarbamate and nucleobase protecting groups can be removed in an efficient one step deprotection. Furthermore, the thionocarbamate group is reportedly stable against 2′ and 3′ isomerization. To selectively protect the 2′ hydroxyl, I used the TIPDS protecting group previously mentioned (section 2.2.1.3) to first block the 5′ and 3′ positions. The first silyl group will preferentially protect the 5′ hydroxyl due to steric hindrance of the

2′ and 3′ hydroxyls. The second will preferentially protect the 3′ hydroxyl due to the stability of 8-membered rings relative to 9-membered rings.

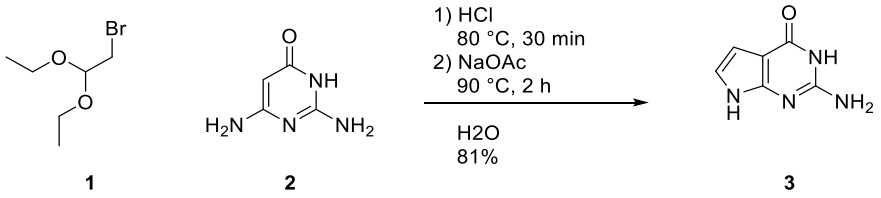
Dibutylformamidyl-protected $PreQ_0$ (**22**) was selectively protected at the 5′ and 3′ hydroxyl positions using 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane to give **26a** in 78% yield. The 2′ hydroxyl was then activated with thiocarbonyldiimidazole to give **27a** in 95% yield. The imidazole group could then be displaced with thiomorpholine dioxide to give **28a** in 92% yield. Following selective protection of the 2′ position with the thionocarbamate group, the TIPDS group on the 5′ and 3′ could be selectively removed using 70% hydrogen fluoride in pyridine to give the pure 2′ thionocarbamate isomer **29a** in 62% yield. Finally, **29a** could be protected at the 5′ position with DMT-Cl and phosphitylated at the 3′ position to produce the $PreQ_0$ phosphoramidite **30a** in 37% yield over two steps. This compound is then suitable for use in solid-phase RNA oligonucleotide synthesis.

The TC phosphoramidite synthesis for wyosine was approached similarly. As wyosine does not have any exocyclic amines, it was only necessary to protect the 5′ and 2′ hydroxyls. The 2′ hydroxyl on **26b** could be similarly activated with thiocarbonylimidazole and displaced with thiomorpholine to give **28b** in 94% yield over two steps. Upon removal of the TIPDS group using 70% hydrogen fluoride, 2′ to 3′ isomerization of the thionocarbamate group was observed via NMR (compound **29b** in 75% yield). The isomers were not separable by column chromatography, and in the interest of time, the mixture of 2′ and 3′ isomers were used as is. The 5′ position was protected with DMT-Cl and the 3′ or 2′ hydroxyl was phosphitylated to give the isomers **30b** in 64% yield. These phosphoramidites were thus suitable for solid-phase RNA oligonucleotide synthesis.

Protection of the primary amine on archaeosine (**14**) was attempted with *N*,*N*-dibutylformamide dimethyl acetal, but no reaction was observed via TLC. Due to the possible resonance between the two nitrogens of the amidine group, protecting this group proved to be non-trivial. The phosphoramidite form of archaeosine was thus not pursued further in the interest of time.

2.3 CONCLUSIONS AND FUTURE WORK

Following some optimization of the literature protocols, the tRNA nucleosides $PreQ_0$, archaeosine, and wyosine were successfully synthesized over 8, 2, and 6 steps respectively. Archaeosine was not pursued in its phosphoramidite form following failure to protect the amidine group. To access phosphoramidite forms of the nucleosides, the TC protecting group was found to perform better than TIPS protecting group. While 2′ and 3′ isomerization of wyosine was observed, $PreQ_0$ was selectively converted into its 3′ phosphoramidite form and both wyosine and $PreQ_0$ were successfully phosphitylated. These compounds are thus suitable for solid-phase oligonucleotide synthesis, providing site selective access to two post-transcriptional tRNA modifications for the first time. These modified nucleotides are of interest for studying the spectral properties of fluorogenic RNA aptamers, and were incorporated into a Spinach variant aptamer for further studies in Chapter 3.



1 2 3

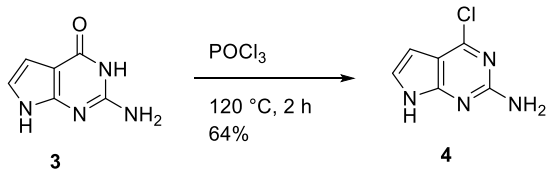
2-Amino-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. Concentrated HCl (4 mL, 0.26 mol, 2.5 equiv.) was added to a suspension of bromoacetaldehyde diethyl acetal **1** (28.3 mL, 0.188 mol, 1 equiv.) in water (93 mL). The suspension was stirred at 90 °C for 30 min, resulting in a clear solution. The solution was then cooled to room temperature, and NaOAc•$3H_2O$ (29.8 g, 0.363 mol, 2 equiv.) was added to create a buffered solution of pH 5. To the resulting solution was added a suspension of 2,4-diamino-6-hydroxypyrimidine **2** (26.4 g, 0.21 mol, 1.1 equiv.) and NaOAc•$3H_2O$ (15.3 g, 0.187 mol, 1 equiv.) in water (198 mL). The resulting mixture was stirred at 80 °C. After 16 h, the solution was removed from heat and stirred at room temperature to form a pale white precipitate. The reaction was then stirred on ice for an additional 90 min. The resulting precipitate was filtered, washed with cold water, and dried under high vacuum overnight to give a pale pink powder (81% yield).

R_f: 0.26 (15% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.97 (s, 1H, NH), 10.24 (s, 1H, NH), 6.61 (dd, J = 3.4, 2.2 Hz, 1H, NHCHC*H*), 6.18 (dd, J = 3.4, 2.2 Hz, 1H, C*H*NH), 6.05 (s, 2H, NH_2).

^{13}C NMR (101 MHz, DMSO-d_6) δ 159.3, 152.7, 151.6, 117.1, 102.0, 100.3.

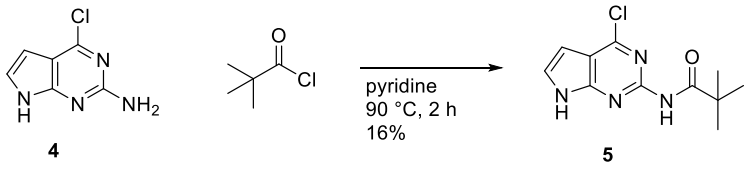
MS: calculated for $[C_6H_7N_4O]^+$: 151.0, Found: 151.0.



4-Chloro-2-amino-7*H*-pyrrolo[2,3-d]pyrimidine. 3 (5.00 g, 33.3 mmol, 1 equiv.) was suspended in $POCl_3$ (31 mL, 330 mmol, 10 equiv.) and heated under reflux for 2 h. The reaction was cooled to room temperature, and the majority of the remaining $POCl_3$ was removed under reduced pressure. Ice water was slowly added to the resulting residue to quench the last remnants of $POCl_3$. The mixture was filtered, and the filtrated was slowly adjusted to pH 2 using conc. ammonia. A yellow precipitate formed upon addition of conc. ammonia, and this solid was collected by vacuum filtration and washed with cold water. 3.57 g of **4** was isolated (64% yield). This compound was taken forward without further purification.

An NMR of this compound was not obtained due to poor solubility in deuterated solvents.

MS: calculated for $[C_6H_6ClN_4O]^+$: 169.0, Found: 169.1.



4-Chloro-2-pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidine. (2a) (2.57 g, 15.4 mmol, 1 equiv.) was co-evaporated with anhydrous pyridine (30 mL x3). Once dry, **4** was suspended in anhydrous pyridine (36 mL) and pivaloyl chloride (6.5 mL, 53 mmol, 3.4

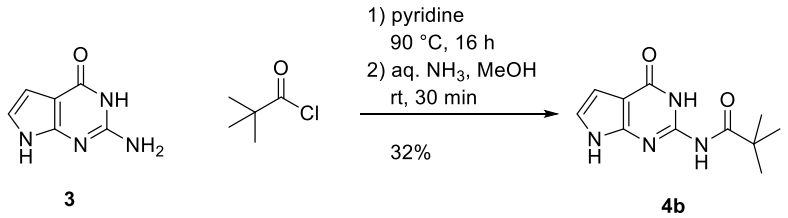
equiv.) was added. The reaction was heated under reflux for 2 h. The reaction was quenched with MeOH (100 mL) and the remaining pyridine and MeOH was removed under reduced pressure. The resulting residue was dissolved in dichloromethane and washed with 0.1 N HCl, dried over $MgSO_4$, and purified using flash chromatography (59:1 CH_2Cl_2/MeOH) to afford 0.636 g of a white solid (16% yield).

R_f: 0.47 (5% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 12.4 (s, 1H, NH), 10.1 (s, 1H, NHC(O)), 7.6 (d, J = 3.6 Hz, 1H, NHC*H*CH), 6.5 (d, J = 3.6 Hz, 1H, C*H*NH), 1.2 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 176.1, 152.5, 151.7, 150.6, 128.5, 114.7, 99.5, 27.5 (3C).

MS: calculated for $[C_{11}H_{14}ClN_4O]^+$: 253.0, Found: 253.0.



3 **4b**

2-Pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 3 (10.0 g, 66.6 mmol, 1 equiv.) was co-evaporated with anhydrous pyridine, then suspended in anhydrous pyridine (80 mL). Pivaloyl chloride (39.2 mL, 0.218 mol, 3.2 equiv.) was added and the reaction was stirred at 90 °C for 16 h. The reaction was quenched with MeOH. The solvent was then removed under reduced pressure, and the resulting residue was dissolved in conc. aqueous ammonia (17 mL, 254 mmol, 4.5 equiv.) and MeOH (160 mL). The reaction

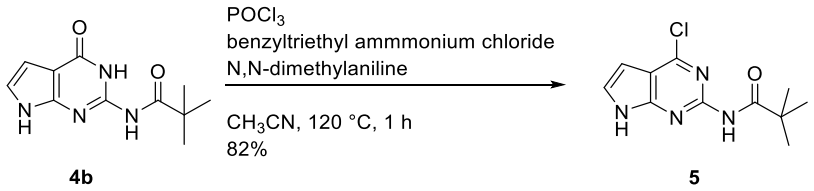
was stirred at room temperature for 30 min. The resulting solid was collected via vacuum filtration. The filtrate was concentrated under reduced pressure, affording a precipitate which was also collected via vacuum filtration and washed with water (5.03 g, 21.5 mmol, 32% yield).

R_f: 0.14 (3% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.9 (s, 1H, NH), 11.7 (s, 1H, NH), 10.9 (s, 1H, NHC(O)), 7.0 (dd, *J* = 3.4, 2.3 Hz, 1H, NHCHC*H*), 6.5 (dd, *J* = 3.4, 2.0 Hz, 1H, C*H*NH), 1.3 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, DMSO-d_6) δ 181.3, 157.5, 148.2, 147.0, 120.2, 104.3, 102.7, 26.8 (3C).

MS: calculated for $[C_{11}H_{15}ClN_4O_2]^+$: 235.1, Found: 235.0.



4-Chloro-2-pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidine. 4b (1.02 g, 4.35 mmol, 1 equiv.) was co-evaporated in anhydrous CH_3CN (10 mL x3). Once dry, (4b) was suspended in 10 mL of anhydrous CH_3CN. *N*,*N*-dimethylaniline (1.1 mL, 8.7 mmol, 2 equiv.), benzyltriethylammonium chloride (1.96 g, 8.60 mmol, 2 equiv.), and $POCl_3$ (2.4 mL, 25.8 mmol, 6 equiv.) were added to the reaction. The reaction was heated to reflux for 1 h, and CH_3CN was removed under reduced pressure. The remaining residue

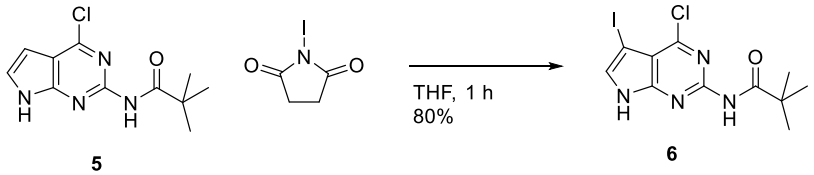
was poured onto water, and resulting precipitate was filtered (0.894 g, 3.55 mmol, 81% yield).

R_f: 0.40 (1:1, Hexane:EtOAc)

^{1}H NMR (400 MHz, DMSO-d_6) δ 12.4 (s, 1H, NH), 10.1 (s, 1H, NHC(O)), 7.6 (d, J = 3.6 Hz, 1H, NHCHC*H*), 6.5 (d, J = 3.6 Hz, 1H, C*H*NH), 1.2 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 176.1, 152.5, 151.7, 150.6, 128.5, 114.7, 99.5, 27.5 (3C).

MS: calculated for $[C_{11}H_{14}ClN_4O]^+$: 253.1, Found: 253.0.



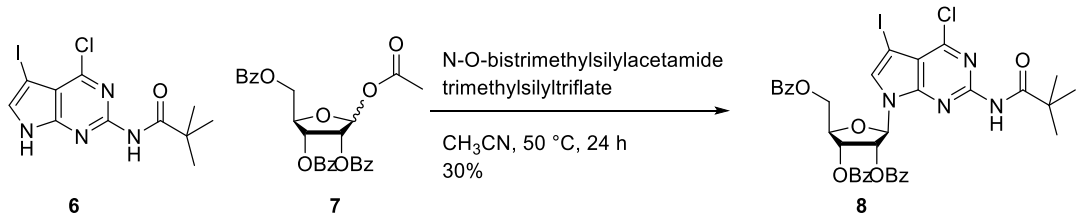
4-Chloro-5-iodo-2-pivaloylamino-7*H*-pyrrolo[2,3-d]pyrimidine. 5 (0.445 g, 1.76 mmol, 1 equiv.) was dissolved in anhydrous THF (3 mL). N-iodosuccinimide (0.436 g, 1.94 mmol, 1.1 equiv.) was added, and the reaction stirred for 1 h in the dark at room temperature. The solution was diluted with CH_2Cl_2, washed with water and a sodium thiosulfate solution, and dried over $MgSO_4$. The resulting solution was concentrated under reduced pressure and purified via flash chromatography (98:2 CH_2Cl_2/MeOH) to afford a white solid (80% yield).

R_f: 0.24 (2% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 12.7 (d, J = 2.4 Hz, 1H, NH), 10.1 (s, 1H, NHC(O)), 7.8 (d, J = 2.4 Hz, 1H, C*H*NH), 1.2 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, DMSO-d_6) δ 176.3, 153.0, 151.9, 151.1, 133.2, 112.8, 100.0, 27.3 (3C).

MS: calculated for $[C_{11}H_{13}ClIN_4O]^+$: 379.0, Found: 379.0.



4-Chloro-5-iodo-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidine. 6 (0.110 g, 0.290 mmol, 1 equiv.) was co-evaporated in anhydrous CH_3CN, then suspended in anhydrous CH_3CN (2 mL) and *N*-*O*-bistrimethylsilylacetamide (87 μL, 0.349 mmol, 1.2 equiv.) was added. The reaction was stirred at room temperature for 5 min. TMSOTf (73 μL, 0.400 mmol, 1.4 equiv.) was added. 1-*O*-acetyl-2,3,5-tribenzyl-D-ribofuranose (2.92 g, 0.580 mmol, 2 equiv.) was added in three portions while stirring at 50 °C. The reaction was cooled to room temperature and purified using flash chromatography (CH_2Cl_2) to afford a yellow foam (1.12 g, 1.39 mmol, 30% yield).

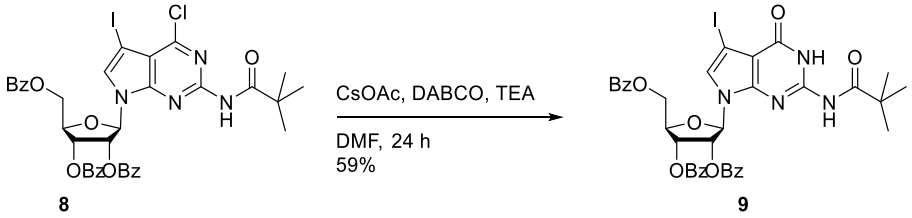
R_f: 0.64 (1% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.4 (s, 1H, NH), 8.1 (s, 1H, NC*H*), 7.9 (dddd, J = 9.6, 6.8, 2.9, 1.4 Hz, 6H, HAr), 7.7 – 7.6 (m, 3H, HAr), 7.5 – 7.4 (m, 6H, HAr), 6.5 (d,

J = 3.9 Hz, 1H, C_1H), 6.5 (app. t, *J* = 6.0 Hz, 1H, C_3H), 6.4 (dd, *J* = 6.0, 3.9 Hz, 1H, C_2H), 4.9 – 4.7 (m, 2H, C_5H_2), 4.7 (dd, *J* = 11.7, 5.7 Hz, 1H, C_4H), 1.2 (d, *J* = 0.9 Hz, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, DMSO-d_6) δ 176.3, 165.8, 165.0, 152.3, 152.0, 151.8, 134.4 (2C), 134.2, 133.9, 129.9 (3C), 129.8 (2C), 129.7 (2C), 129.6 (3C), 129.2 (2C), 129.1 (2C), 128.8, 113.7, 88.1, 79.6, 74.4, 71.7, 64.2, 60.2, 54.6, 39.9, 27.1 (3C).

MS: calculated for $[C_{37}H_{32}Cl_1N_4O_8]^-$: 823.1, Found: 823.0.



8 **9**

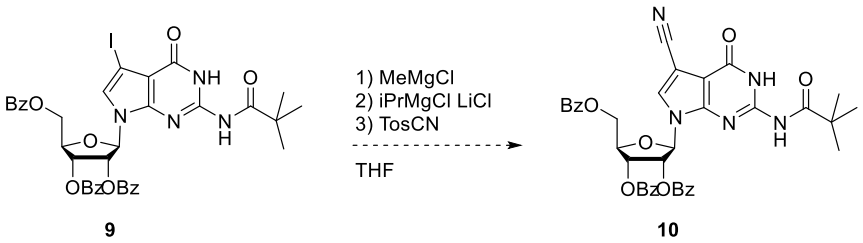
5-Iodo-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 8 (0.102 g, 0.124 mmol, 1 equiv.) was dissolved in anhydrous DMF (70 μL). CsOAc (96.1 mg, 0.627 mmol, 5 equiv.), DABCO (25.7 mg, 0.212 mmol, 1.7 equiv), and TEA (50 μL, 0.36 mmol, 3 equiv.) was added, and the solution was stirred for at room temperature for 24 h. The reaction was quenched with water. The solution was extracted using CH_2Cl_2, washed with brine, and dried over $MgSO_4$. The resulting residue was purified via flash chromatography (99:1 CH_2Cl_2/MeOH) to afford a pale yellow solid (58.3 mg, 0.0725 mmol, 59% yield).

R_f: 0.62 (0.5% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 12.0 (s, 1H, NH), 10.8 (s, 1H, NHC(O)), 8.1 – 7.3 (m, 16H, NC*H*, HAr), 6.5 (d, *J* = 5.8 Hz, 1H, C_1H), 6.2 (app. t, *J* = 5.8 Hz, 1H, C_2H), 6.0 (dd, *J* = 5.8, 5.6 Hz, 1H, C_3H), 4.9 – 4.7 (m, 2H, C_5H_2), 4.7 (dd, *J* = 12.0, 5.6 Hz, 1H, C_4H), 1.2 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 180.1, 166.1, 165.2, 156.7, 146.4, 146.2, 133.8 (2C), 133.6, 129.8, 129.2 (3C), 128.9 (2C), 128.8 (2C), 128.7 (3C), 128.6 (2C), 128.5 (2C), 127.1, 106.6, 89.4, 78.5, 77.2, 74.5, 70.6, 61.4, 56.1, 40.2, 26.9 (3C).

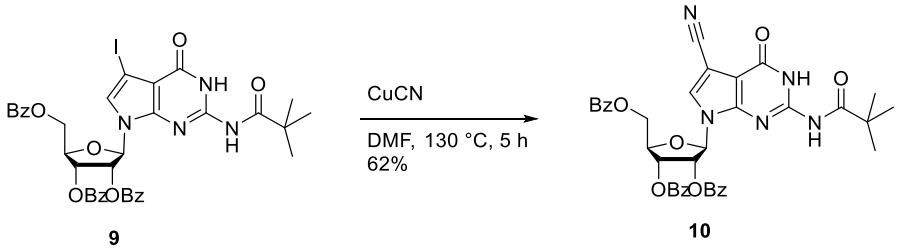
MS: calculated for $[C_{37}H_{33}IN_4O_9]^+$: 805.1, Found: 805.2.



9 **10**

5-Cyano-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 9 (58 mg, 0.072 mmol, 1 equiv.) was dissolved in anhydrous toluene (580 µL) and cooled to -78 °C in an acetone/dry ice cooling bath. MeMgCl (3 M in THF, 48 µL, 0.144 mmol, 2 equiv.) was added dropwise and the reaction stirred for 1 h. iPrMgCl•LiCl (1.3 M in THF, 143 µL, 0.186 mmol, 2.6 equiv.) was taken up in 410 µL of anhydrous toluene and added over the course of 4 h using a syringe pump. TosCN (82.1 mg, 0.453 mmol, 6 equiv.) was added, and the solution was allowed to warm to room temperature over 16 h. The reaction was diluted with water and extracted with EtOAc. The organic layer was concentrated under reduced

pressure and purified via flash chromatography (4:1 hexane/EtOAc). The recovered, major product was **9** (33.2 mg, 0.0413 mmol, 57% recovery).



9 **10**

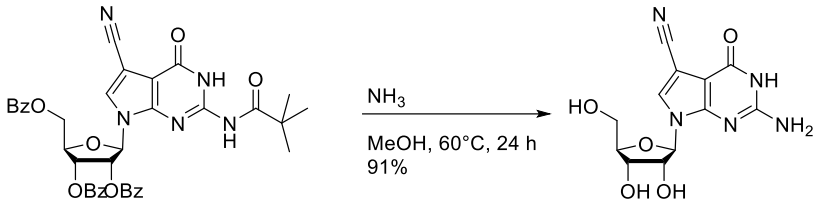
5-Cyano-2-pivaloylamino-7-[(2,3,5-tri-O-benzoyl)-β-D-ribofuranosyl]-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 9 (0.650 g, 0.790 mmol, 1 equiv.) was dissolved in 8.5 mL of anhydrous DMF, CuCN (1.08 g, 12.1 mmol, 15 equiv.) was added, and the reaction was refluxed. After 5 h, the reaction was cooled and DMF was removed under reduced pressure. The residue was purified via flash chromatography (3:2 Hex/EtOAc) to give a yellow foam (0.343 g, 0.0542 mmol, 61% yield).

R_f: 0.25 (3:2, Hexane:EtOAc)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 11.7 (s, 1H, NH), 8.8 (s, 1H, NHC(O)), 8.0 – 7.2 (m, peak obscured by residual $CHCl_3$, NC*H*, HAr), 6.8 (dd, *J* = 8.3, 4.9 Hz, 1H, C_3H), 6.4 (dd, *J* = 4.9, 1.7 Hz, 1H, C_2H), 6.0 (d, *J* = 1.7 Hz, 1H, C_1H), 4.7 (dt, *J* = 8.3, 3.1 Hz, 1H, C_4H), 4.6 (d, *J* = 3.1 Hz, 2H, C_5H_2), 1.3 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 180.4, 166.0, 165.2, 155.4, 148.0, 146.5, 134.0 (2C), 133.9, 133.8, 130.4, 129.8 (3C), 129.8 (2C), 129.2 (2C), 128.8 (3C), 128.7 (2C), 128.6 (2C), 128.5, 113.5, 105.3, 90.0, 79.0, 74.3, 70.5, 61.3, 60.4, 40.3, 26.8 (3C).

MS: calculated for $[C_{38}H_{33}N_5O_9]^-$: 702.2, Found: 702.2.



10 11

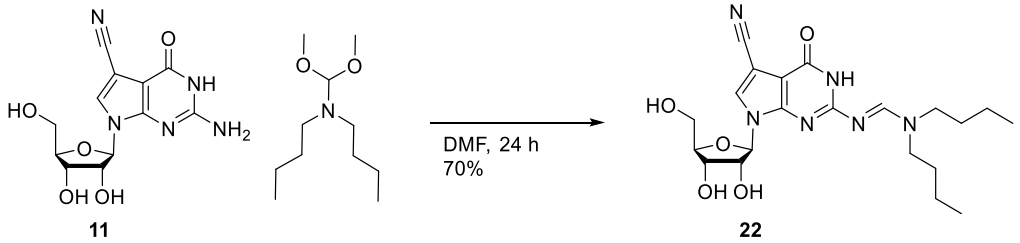
2-Amino-5-cyano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one (PreQ0). **10** (0.207 g, 0.294 mmol) was suspended in 4 M NH_3/MeOH (4 mL) and stirred at 60 °C for 24 h. MeOH was removed under reduced pressure and the resulting residue was purified via flash chromatography (10% MeOH/CH_2Cl_2) to give a white foam (0.223, 0.726 mmol, 91%).

R_f: 0.22 (10% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.9 (s, 1H, NH), 7.9 (s, 1H, NC*H*), 6.6 (s, 2H, NH_2), 5.9 (d, *J* = 5.9 Hz, 1H, C_1H), 5.4 (d, *J* = 5.9 Hz, 1H, $O_2'H$), 5.1 (d, *J* = 4.4 Hz, 1H, $O_3'H$), 5.1 (t, *J* = 5.4 Hz, 1H, $O_5'H$), 4.3 (q, *J* = 5.9 Hz, 1H, C_2H), 4.1 (d, *J* = 4.4 Hz, 1H, C_3H), 3.9 (q, *J* = 3.9 Hz, 1H, C_4H), 3.6 (dd, *J* = 5.4, 3.9 Hz, 2H, C_5H_2).

^{13}C NMR (101 MHz, DMSO-d_6) δ 157.8, 154.4, 151.9, 128.0, 115.8, 99.1, 86.9, 86.8, 85.6, 74.5, 70.7, 61.7.

MS: calculated for $[C_{12}H_{13}N_5O_5]^-$: 306.1, Found: 306.1.



11 22

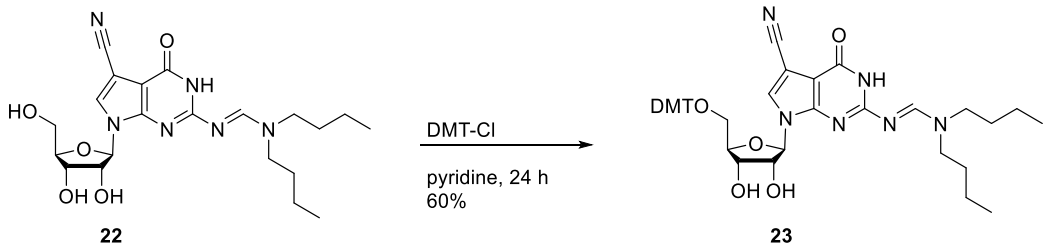
5-cyano-2-[(dibutylamino)methylidene]amino-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 11 (0.220 g, 0.715 mmol, 1 equiv.) was dissolved in anhydrous DMF (2 mL). Di-n-butylformamide dimethyl acetal (0.335 g, 1.65 mmol, 2 equiv.) was added and the reaction was allowed to stir at room temperature for 24 h. DMF was removed under reduced pressure, and the residue was dissolved in CH_2Cl_2, washed with sat. sodium bicarb solution, and dried over Na_2SO_4. CH_2Cl_2 was removed under reduced pressure and the remaining residue was purified via column chromatography (1% to 7% MeOH/CH_2Cl_2) to afford a white foam (0.223 g, 0.499 mmol, 70% yield).

R_f: 0.09 (5% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 9.2 (br. s, 1H, NH), 8.5 (s, 1H, NC*H*N), 7.4 (s, 1H, NC*H*C), 5.8 (d, *J* = 6.7 Hz, 1H, C_1H), 5.5 (br. s, 1H, O_2'H), 5.0 (dd, *J* = 6.7, 5.3 Hz, 1H, C_2H), 4.4 (dd, *J* = 5.3, 2.1 Hz, 1H, C_3H), 4.3 (br. s, 1H, O_5'H), 4.2 (q, *J* = 2.1 Hz, 1H, C_4H), 3.9 (br. s, 1H, O_3'H), 3.9 – 3.7 (m, 2H, C_5H_2), 3.5 – 3.2 (m, 4H, N(CH_2)$_2$), 1.8 – 1.5 (m, 5H, N(CH_2CH_2)$_2$), 1.4 – 1.2 (m, 4H, N(CH_2CH_2CH_2)$_2$), 1.0 – 0.8 (m, 6H, (CH_3)$_2$).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 158.47, 157.07, 149.94, 130.55, 114.39, 103.18, 91.42, 86.91, 86.20, 77.25, 73.28, 71.88, 62.61, 52.58, 45.85, 30.84, 28.96, 20.17, 19.79, 13.86, 13.77.

HR-MS: calculated for $[C_{21}H_{30}N_6O_5]^-$: 447.2278, Found: 447.2351.



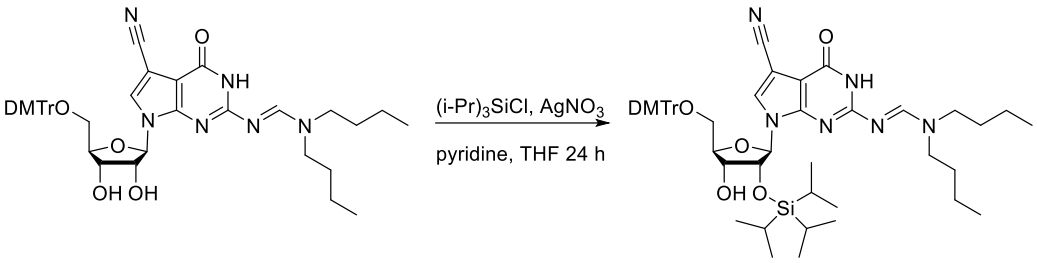
5-cyano-2-[(dibutylamino)methylidene]amino-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. **22** (0.200 g, 0.578 mmol, 1 equiv.) was co-evaporated with anhydrous pyridine (1 mL x2) and dissolved in 2 mL of anhydrous pyridine. Dimethoxytrityl chloride (0.274, 0.811 mmol, 1.4 equiv.) was added in portions over 16 h. Pyridine was removed under reduced pressure and the resulting residue was purified via column chromatography (0.25% pyridine/CH_2Cl_2 to 3% MeOH/CH_2Cl_2) to afford a white foam (0.260 g, 0.347 mmol, 60% yield).

R_f: 0.30 (5% MeOH/CH_2Cl_2)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.8 (br. s, 1H, NH), 8.4 (s, 1H, NC*H*N), 7.2 – 6.9 (m, peak partially obscured by $CHCl_3$, NC*H*C, HAr), 6.6 (d, *J* = 8.2 Hz, 4H, HAr), 5.9 (d, *J* = 5.9 Hz, 1H, C_1H), 5.4 (br. s, 1H, $O_{2'}H$), 4.4 (t, *J* = 5.6 Hz, 1H, C_2H), 4.1 (dd, *J* = 5.1, 2.6 Hz, 1H, C_3H), 4.1 – 4.0 (m, 1H, C_4H), 3.5 (s, 6H, $(OCH_3)_2$), 3.3 – 3.2 (m, 2H, C_5H_2), 3.1 (dt, *J* = 14.5, 6.5 Hz, 4H, $N(CH_2)_2$), 1.7 (s, 1H, $O_{3'}H$), 1.4 (h, *J* = 8.9, 8.2 Hz, 4H, $N(CH_2CH_2)_2$), 1.1 (h, *J* = 7.5 Hz, 4H, $N(CH_2CH_2CH_2)_2$), 0.8 – 0.7 (m, 6H, $(CH_3)_2$).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 158.6, 158.4, 157.7, 157.0, 150.1, 144.4, 136.2, 135.5, 135.5, 130.1, 130.1, 128.1, 127.9, 127.4, 127.0, 123.8, 114.4, 113.3, 102.8, 88.8, 87.9, 86.7, 84.7, 75.3, 72.1, 63.8, 55.3, 52.5, 45.9, 30.9, 29.0, 20.2, 19.8, 13.8, 13.7.

HR-MS: calculated for $[C_{42}H_{48}N_6O_7]^-$: 747.3584, Found: 747.3511.



23 **24**

5-cyano-2-[(dibutylamino)methylidene]amino-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-2′-O-triisopropylsilyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. **23** (0.250 g, 0.333 mmol, 1 equiv.) and $AgNO_3$ were combined in dry pyridine (6 mL) and stirred for 5 min. $(i\text{-}Pr)_3SiCl$ (107 μL, 0.500 mmol, 1.5 equiv.) in THF (6 mL) was added and the reaction stirred for 24 h. Column purification (10% to 50% Hexane/EtOAc) gave a white foam (67.2 mg, 0.0742 mmol, 22% yield).

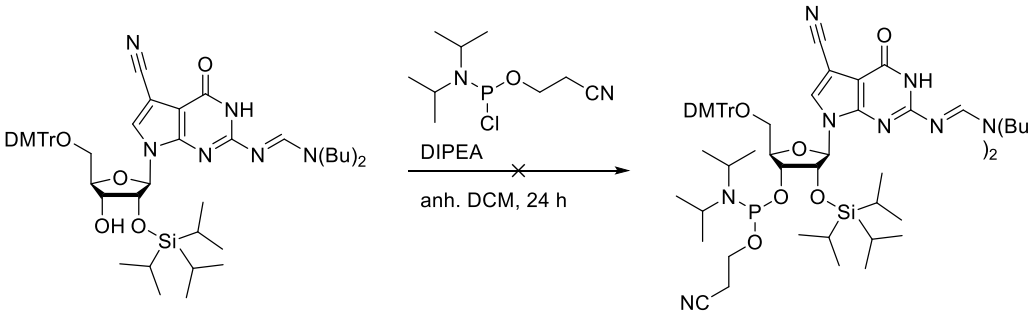
R_f: 0.27 (5:2, Hexane:EtOAc)

^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.8 (s, 1H, NH), 8.6 (s, 1H, NC*H*N), 7.5 – 7.2 (m, peak obscured by $CHCl_3$, NC*H*C, HAr), 6.9 – 6.7 (m, 4H, HAr), 6.2 (d, *J* = 6.2 Hz, 1H, C_1H), 4.7 (dd, *J* = 6.2, 5.0 Hz, 1H, C_2H), 4.4 (dt, *J* = 5.0, 2.6 Hz, 1H, C_3H), 4.2 (app. q, *J* = 2.6 Hz, 1H, C_4H), 3.8 (s, 6H, $(OCH_3)_2$), 3.7 – 3.4 (m, 4H, $N(CH_2)_2$), 3.4 – 3.3 (m, 2H, C_5H_2), 3.0 (d, *J* = 2.6 Hz, 1H, O_3H), 1.7 – 1.5 (m, 4H, $N(CH_2CH_2)_2$), 1.4 –

1.3 (m, 4H, $N(CH_2CH_2CH_2)_2$), 1.1 – 0.8 (m, 27H, $(CH_3)_2$, $Si((CH)CH_3)_3$, $Si((CH)CH_3)_3$).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 157.7, 157.2, 156.4, 156.0, 149.8, 148.8, 143.2, 134.3, 134.1, 129.1, 129.0, 127.1, 127.0, 126.2, 125.6, 122.7, 113.3, 112.3, 112.2, 101.9, 88.5, 86.1, 86.0, 83.0, 76.2, 71.4, 62.8, 54.3, 51.3, 44.7, 30.0, 27.9, 19.1, 18.7, 16.7, 16.4, 12.8, 12.7, 11.0.

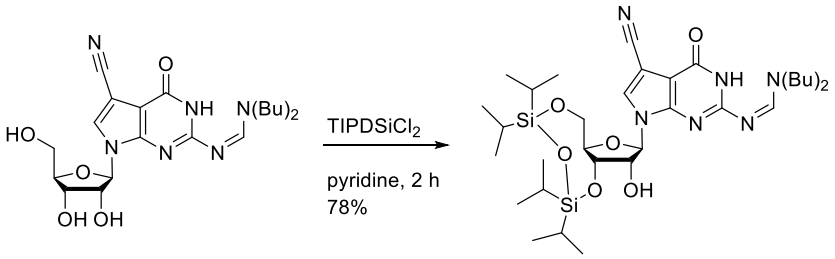
HR-MS: calculated for $[C_{51}H_{68}N_6O_7Si]^-$: 903.4919, Found: 903.4851.



24 **25**

5-cyano-2-[(dibutylamino)methylidene]amino-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-2′-O-triisopropylsilyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one-3′-[2-cyanoethyl *N*,*N*-bis(1-methylethyl)phosphoramidite]. **24** was co-evaporated with anh. DCM and dried under high vacuum overnight. **24** (65 mg, 71.8 µmol, 1 equiv.) was dissolved in 2 mL of anh. dry DCM, and *N*,*N*-diisopropylethylamine (70 µL, 400 µmol, 6 equiv.) and 2-cyanoethoxy-*N*,*N*-diisopropylaminochlorophosphine (50 µL, 225 µmol, 3.5 equiv.) were added to the reaction. The reaction was stirred at room temperature overnight. No change in TLC was observed. The reaction mixture was diluted with DCM, washed with brine, and dried over Na_2SO_4. DCM was removed

under reduced pressure and the remaining residue was purified via column chromatography (0% to 40% EtOAc/Hex). No peaks for the product were observed via ^{31}P NMR, and only peaks for the starting material were observed by ^{1}H NMR.



22 26a

5-cyano-2-[(dibutylamino)methylidene]amino-3′,5′-O-tetraisopropyldisiloxano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. (1.93 g, 4.32 mmol, 1 equiv.) was co-evaporated with anh. pyridine and dissolved in anh. pyridine. 1,3-Dichloro-1,1,3,3-tetraisopropyldisiloxane (2.1 mL, 6.7 mmol, 1.6 equiv.) was added dropwise and stirred at room temperature for 2 h. The reaction was quenched with MeOH, diluted with DCM, washed with saturated sodium bicarbonate, and dried over Na_2SO_4. The solvent was removed under reduced pressure, and purified via column chromatography to afford a white solid (2.31 g, 3.35 mmol, 78% yield).

^{1}H NMR (500 MHz, Chloroform-d) δ 8.95 (s, 1H, N*H*), 8.65 (s, 1H, NC*H*N), 7.53 (s, 1H, NC*H*C), 6.20 (s, 1H, $C_{1'}H$), 4.53 (dd, J = 8.5, 5.0 Hz, 1H, $C_{3'}H$), 4.22 (dd, J = 13.0, 2.7 Hz, 1H, $C_{5'}H_2$), 4.18 (d, J = 5.0 Hz, 1H, $C_{2'}H$), 4.14 (dt, J = 8.5, 2.7 Hz, 1H, $C_{4'}H$), 4.07 (dd, J = 13.0, 2.7 Hz, 1H, C_5H_2), 3.51 (t, J = 7.5 Hz, 2H, N(CH_2)$_2$), 3.39 (t, J = 7.5 Hz, 2H, N(CH_2)$_2$), 2.99 (s, 1H, $O_{2'}H$), 1.69 – 1.58 (m, 4H, N(CH_2CH_2)$_2$), 1.38 (hd, J =

7.4, 3.8 Hz, 4H, $N(CH_2CH_2C\mathit{H}_2)_2$), 1.20 – 1.01 (m, 27H, $((CH_3)_4(CH)_2SiO)_2$), 0.99 (dt, J = 8.8, 7.4 Hz, 6H, (CH3)2).

^{13}C NMR (126 MHz, CDCl3) δ 158.3, 157.6, 157.2, 149.4, 126.2 (N*C*HN), 114.5, 103.3, 88.6, 88.1 ($C_{1'}$), 81.7 ($C_{4'}$), 75.9 ($C_{2'}$), 69.6 ($C_{3'}$), 60.6 ($C_{5'}$), 52.3 ($N(\mathit{C}H_2)_2$), 45.8 ($N(\mathit{C}H_2)_2$), 30.9 ($N(CH_2\mathit{C}H_2)_2$), 29.0 ($N(CH_2\mathit{C}H_2)_2$), 20.2 ($N(CH_2CH_2\mathit{C}H_2)_2$), 19.7 ($N(CH_2CH_2\mathit{C}H_2)_2$), 17.5, 17.4, 17.3, 17.3, 17.1, 17.0, 16.9, 16.9, 13.8, 13.7, 13.4, 13.0, 12.9, 12.6.

HR-MS: calculated for $[C_{33}H_{57}N_6O_6Si_2]^+$: 689.3800, Found: 689.3873.



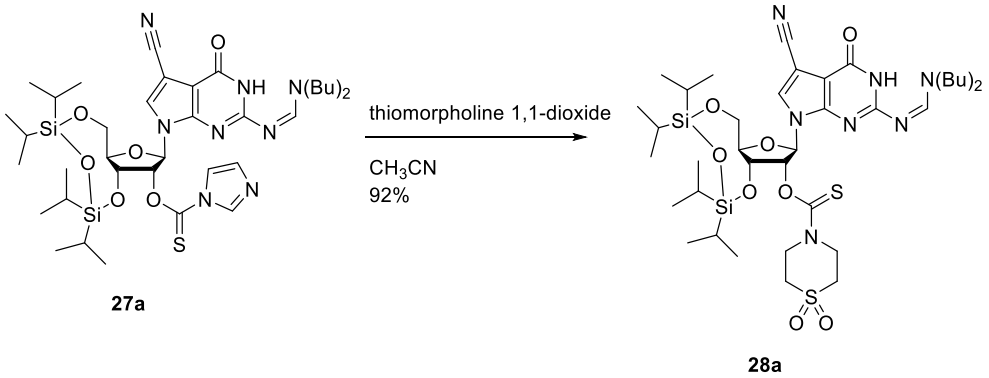
26a **27a**

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-thiocarbonylimidazole-3′,5′-O-tetraisopropyldisiloxano-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one.

26a (0.250 g, 0.363 mmol, 1 equiv.) was dissolved in anh. CH_3CN (2 mL). 1,1′-Thiocarbonyldiimidazole (77.6 mg, 0.435 mmol, 1.2 equiv.) and DMAP (29.0 mg, 0.237 mmol, 0.65 equiv.) were added, and the reaction was stirred at room temperature for 16 h. CH_3CN was removed under reduced pressure and the resulting residue was purified via column chromatography (0% to 40% EtOAc/Hex) to afford an off-white solid (0.275 g, 0.344 mmol, 95% yield).

^{1}H NMR (500 MHz, Chloroform-d) δ 8.94 (broad s, 1H, N*H*), 8.56 (s, 1H, NC*H*N(Bu)$_2$), 8.42 (t, J = 1.1 Hz, 1H, SCNC*H*N), 7.70 (t, J = 1.5 Hz, 1H, SCNC*H*CHN), 7.69 (s, 1H, C8H), 7.14 (dd, J = 1.6, 0.8 Hz, 1H, SCNCHC*H*N), 6.38 (d, J = 4.4 Hz, 1H, $C_{2'}$H), 6.23 (s, 1H, $C_{1'}$H), 4.71 (dd, J = 9.2, 4.5 Hz, 1H, $C_{3'}$H), 4.34 (dd, J = 13.4, 1.7 Hz, 1H, $C_{5'}H_1$), 4.20 (dt, J = 9.1, 2.1 Hz, 1H, $C_{4'}$H), 4.10 (dd, J = 13.4, 2.8 Hz, 1H, $C_{5'}H_2$), 3.54 (dt, J = 13.3, 7.5 Hz, 1H, N(C*H*$_2$)$_2$), 3.45 (dt, J = 13.3, 7.6 Hz, 1H, N(C*H*$_2$)$_2$), 3.31 – 3.26 (m, 2H), 1.65 – 1.56 (m, 4H, N(CH$_2$C*H*$_2$)$_2$), 1.41 – 1.25 (m, 5H, N(CH$_2$CH$_2$C*H*$_2$)$_2$), 1.18 – 1.10 (m, 13H), 1.06 (ddd, J = 13.8, 6.9, 2.2 Hz, 2H), 1.01 – 0.91 (m, 19H).

^{13}C NMR (126 MHz, CDCl3) δ 182.97, 158.11, 157.76, 157.42, 148.90, 136.97, 131.42, 125.96, 117.94, 114.29, 103.68, 88.84, 86.86, 84.08, 82.82, 68.29, 59.73, 52.39, 45.82, 31.01, 28.97, 20.12, 19.77, 17.45, 17.31, 17.28, 17.16, 17.00, 16.95, 16.82, 16.72, 13.80, 13.64, 13.43, 12.93, 12.92, 12.50.



27a **28a**

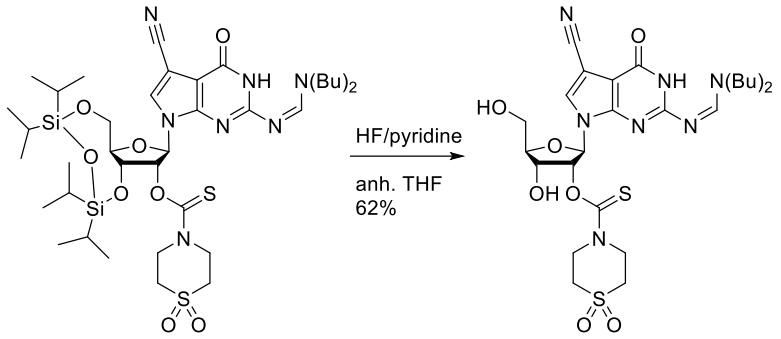
5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-3′,5′-O-(tetraisopropyldisiloxano)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 27a (0.264 g, 3.30 mmol, 1 equiv.) was suspended in

CH_3CN (2 mL) and gently heated until a solution was obtained. The reaction was then heated under reflux for 3 h and cooled to room temperature. CH_3CN was removed under reduced pressure, and the resulting residue was purified via column chromatography (0% to 2% MeOH/DCM) to afford an off-white solid (0.263 g, 3.04 mmol, 92% yield).

^{1}H NMR 1H NMR (500 MHz, Chloroform-d) δ 8.72 (s, 1H, NC*H*N(Bu)$_2$), 7.70 (s, 1H, C8H), 6.48 (d, J = 4.3 Hz, 1H, $C_{2'}H$), 6.10 (s, 1H, $C_{1'}H$), 5.05 (d, J = 14.3 Hz, 1H, $C_{2'}H$), 4.70 (d, J = 14.3 Hz, 1H), 4.60 (dd, J = 8.7, 4.4 Hz, 1H, $C_{1'}H$), 4.31 (dd, J = 13.7, 1.8 Hz, 1H, $C_{5'}H_1$), 4.17 (t, J = 11.5 Hz, 1H, $C_{4'}H$), 4.06 (dd, J = 13.7, 1.8 Hz, 1H, $C_{5'}H_2$), 4.05 – 4.03 (m, 1H), 3.99 (t, J = 11.5 Hz, 1H), 3.55 (dt, J = 15.0, 7.3 Hz, 1H), 3.44 (dt, J = 15.0, 7.3 Hz, 1H), 3.36 (t, J = 7.6 Hz, 2H, N(C*H*$_2$)$_2$), 3.26 – 3. 17 (m, 2H), 3.17 – 3.04 (m, 2H), 1.64 (dt, J = 15.0, 7.3 Hz, 4H), 1.37 (dt, J = 15.3, 7.3 Hz, 4H), 1.16 – 0.95 (m, 34H).

^{13}C NMR (126 MHz, CDCl3) δ 187.55, 158.23, 158.12, 148.80, 126.32, 114.42, 103.73, 88.40, 87.56, 83.10, 83.00, 77.28, 77.03, 68.09, 59.71, 52.35, 51.71, 51.31, 48.54, 45.84, 44.13, 31.03, 29.72, 28.99, 20.14, 19.83, 17.47, 17.28, 17.11, 16.98, 16.90, 16.86, 16.75, 13.82, 13.75, 13.41, 12.93, 12.18.

HR-MS: calculated for $[C_{38}H_{64}N_7O_8S_2Si_2]^+$: 866.3718, Found: 866.3782.



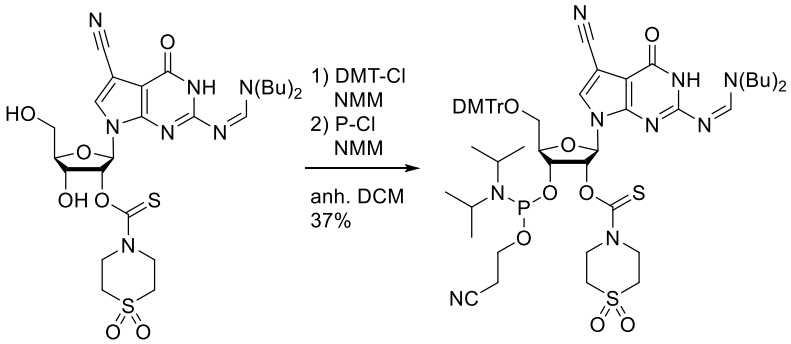
28a 29a

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 28a (0.529 g, 0.611 mmol, 1 equiv) was dissolved in anhydrous THF (3 mL) and transferred to a plastic falcon tube. Anh. pyridine (0.30 mL, 3.7 mmol, 6 equiv.) was added and the reaction was cooled to 0 °C in an ice bath. HF/pyridine (70% HF/ 30% pyridine) was added dropwise with stirring and the reaction was stirred at room temperature for 4 h. The reaction was directly purified via column chromatography (0% to 10% MeOH/DCM) to afford a white foam (0.238 g, 0.382 mmol, 62% yield).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.08 (broad s, 1H, N*H*), 8.62 (s, 1H, NC*H*N(Bu)$_2$), 7.65 (s, 1H, C8H), 6.40 (t, *J* = 5.3 Hz, 1H, C$_{2'}$H), 6.18 (d, *J* = 5.3 Hz, 1H, C$_{1'}$H), 4.92 – 4.87 (m, 1H, C$_{3'}$H), 4.81 (d, *J* = 14.2 Hz, 1H), 4.55 (d, *J* = 14.2 Hz, 1H), 4.26 – 4.17 (m, 2H), 4.13 (broad s, 1H, C$_{3'}$OH), 4.09 – 4.01 (m, 1H), 3.99 (broad s, 1H, C$_{5'}$OH), 3.95 (d, *J* = 11.8 Hz, 1H, C$_{5'}$H$_1$), 3.84 (d, *J* = 11.8 Hz, 1H, C$_{5'}$H$_2$), 3.54 – 3.44 (m, 2H), 3.43 – 3.38 (m, 2H), 3.38 – 3.31 (m, 1H), 3.30 – 3.21 (m, 1H), 3.02 – 3.14 (m, 2H, N(C*H*$_2$)$_2$), 1.64 (dp, *J* = 22.8, 7.5 Hz, 4H, N(C*H*$_2$)$_2$), 1.37 (hept, *J* = 7.5 Hz, 4H, N(CH$_2$C*H*$_2$)$_2$), 0.99 (t, *J* = 7.5 Hz, 3H, (CH$_3$)$_2$), 0.96 (t, *J* = 7.5 Hz, 3H, (CH$_3$)$_2$).

^{13}C NMR (126 MHz, $CDCl_3$) δ 187.28, 158.55, 158.29, 157.54, 149.99, 130.13, 114.88, 103.75, 88.21, 87.78, 86.06, 81.81, 69.27, 61.81, 52.46, 51.56, 51.04, 48.56, 45.96, 44.05, 30.92, 28.99, 20.15, 19.79, 13.83, 13.78.

HR-MS: calculated for $[C_{26}H_{38}N_7O_7S_2]^+$: 624.2196, Found: 624.2266.



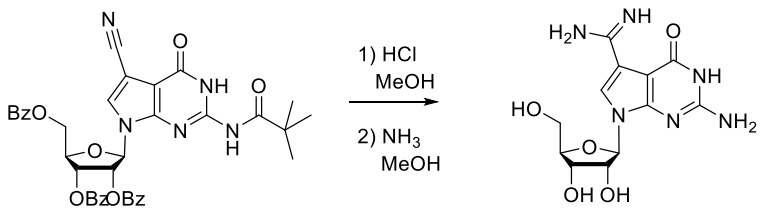
29a **30a**

5-cyano-2-[(dibutylamino)methylidene]amino-2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-5′-O-(4,4′-dimethoxytrityl)-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one-3′-[2-cyanoethyl *N*,*N*-bis(1-methylethyl)phosphoramidite]. 29a (0.300 g, 0.481 mmol, 1 equiv.) was dissolved in anh. DCM (5 mL) with 4,4′-dimethoxytritylchloride (0.196 g, 0.577 mmol, 1.2 equiv.). 4-Methylmorpholine (63 µL, 0.577 mmol, 1.2 equiv.) was added dropwise, and the reaction was stirred at room temperature for 1 h. 2-Cyanoethoxy-N,N-diisopropylaminochlorophosphine (110 µL, 0.49 mmol, 1 equiv.) and 4-methylmorpholine (55 µL, 0.500 mmol, 1 equiv.) were added and the reaction was stirred at room temperature for 2 h. The solution was diluted with DCM, washed with brine, and dried over Na_2SO_4. DCM was removed under reduced pressure and the product was purified via column chromatography (0% to 50% acetone/hexane) to afford a white foam (0.207 g, 0.184 mmol, 38%).

^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 9.27 (s, 1H), 8.62 (d, *J* = 8.6 Hz, 1H), 7.58 (d, *J* = 2.3 Hz, 1H), 7.50 – 7.41 (m, 2H), 7.38 – 7.18 (m, 8H, DMT), 6.89 – 6.79 (m, 5H, DMT), 6.33 (dd, *J* = 19.7, 4.9 Hz, 1H), 6.22 (dt, *J* = 20.0, 5.1 Hz, 1H), 4.98 (dt, *J* = 10.2, 5.1 Hz, 1H), 4.89 (dt, *J* = 9.6, 4.7 Hz, 1H), 4.77 (t, *J* = 15.2 Hz, 1H), 4.48 (s, 1H), 4.21 – 3.87 (m, 3H), 3.75 (d, *J* = 5.1 Hz, 6H, DMT OCH_3), 3.35 (dd, *J* = 8.8, 5.1 Hz, 3H), 3.17 – 2.88 (m, 5H), 2.63 – 2.57 (m, 1H), 1.65 – 1.50 (m, 4H), 1.38 – 1.23 (m, 4H), 1.21 (dd, *J* = 6.8, 5.3 Hz, 4H), 1.11 (dd, *J* = 11.4, 6.8 Hz, 9H, CH_3), 1.03 (d, *J* = 6.8 Hz, 3H, CH_3), 0.98 – 0.85 (m, 6H).

^{31}P NMR (126 MHz, CD_3CN) δ 150.47, 149.47, 13.74.

HR-MS: calculated for $[C_{56}H_{72}N_9O_{10}PS_2]^+$: 1126.4581, Found: 1126.4652.



10 13

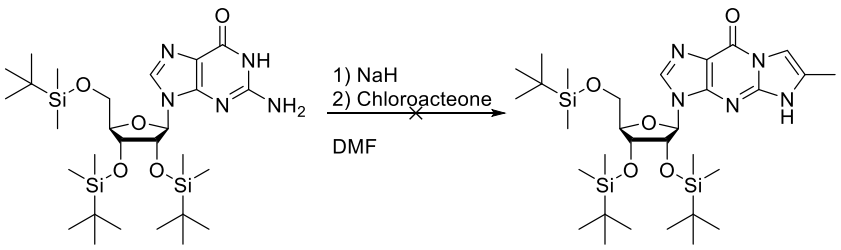
2-Amino-5-amidino-7-β-D-ribofuranosyl-7*H*-pyrrolo[2,3-d]pyrimidin-4-one. 10 (1.94 g, 2.76 mmol, 1 equiv.) was dissolved in absolute EtOH (30 mL) and cooled to 0 °C in an ice bath. Acetyl chloride (10 mL) was slowly added dropwise with stirring and the reaction was stirred at room temperature for 3 h. The reaction was flushed under argon for 2 h, and an outlet was fitted to the flask and bubbled through a solution of sodium bicarbonate to neutralize HCl gas. Methanolic ammonia (30 mL) was added to the flask and the reaction was stirred overnight. The solvent was removed under

reduced pressure and the crude product was recrystallized in water to give archaeosine (0.754 g, 2.33 mmol, 84% yield).

^{1}H NMR (500 MHz, DMSO-d_6) δ 9.99 (broad s, 2H), 8.60 (broad s, 2H), 8.32 (s, 1H, C8H), 6.89 (broad s, 2H), 5.94 (d, J = 6.1 Hz, 1H, $C_{1'}H$), 5.48 (d, J = 4.1 Hz, 1H, $C_{2'}OH$), 5.25 (s, 1H, $C_{3'}OH$), 4.99 (t, J = 5.2 Hz, 1H, $C_{5'}OH$), 4.28 – 4.19 (m, 1H, $C_{2'}H$), 4.10 – 4.06 (m, 1H, $C_{3'}H$), 3.89 (dd, J = 10.6, 4.5 Hz, 1H, $C_{4'}H$), 3.56 (ddt, J = 32.7, 10.6, 5.2 Hz, 2H, $C_{5'}H_2$).

^{13}C NMR (126 MHz, DMSO) δ 160.98, 159.19, 154.53, 154.10, 125.38, 107.81, 96.92, 86.52, 85.73, 74.50, 70.89, 62.15.

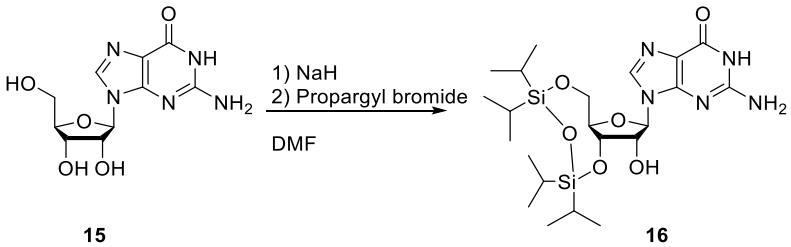
HR-MS: calculated for $[C_{12}H_{17}N_6O_5]^+$: 325.1182, Found: 325.1252.



17a **19a**

2′,3′,5′-O-tri(tertbutyldimethylsilyl)-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 17a (1 g, 1.59 mmol, 1 equiv.) was dissolved in anh. DMF (15 mL) and cooled to 0 °C in an ice bath. NaH (60% dispersion in mineral oil) (77 mg, 1.92 mol, 1.2 equiv.) was added portion-wise over 5 min, and the reaction was stirred at room temperature for 1 h. Chloroacetone (154 μL) was added, and the reaction was stirred at room temperature and monitored by the hour. Over time,

TLC analysis indicated the consumption of starting material but no obvious formation of a singular product.



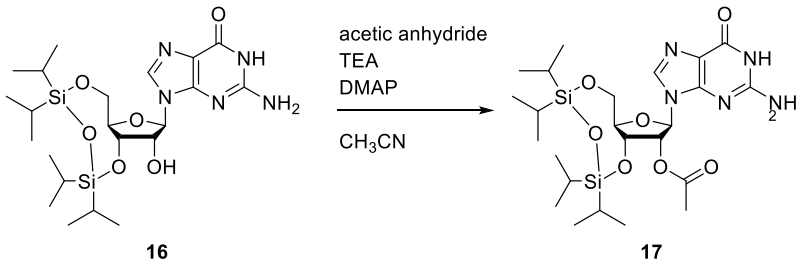
15 16

3′,5′-O-tetraisopropyldisiloxano-2-amino-1,9-dihydro-9-β-D-ribofuranosyl-6*H*-purin-6-one. Guanosine **15** (2.00 g, 7.06 mmol, 1 equiv.) was suspended in anh. pyridine/DMF (20 mL/4 mL). $TIPDSiCl_2$ (2.5 mL, 7.81 mmol, 1.1 equiv.) was added dropwise and the reaction stirred overnight at room temperature. MeOH (5 mL) was added to quench the reaction, and the solvents were removed under reduced pressure. The crude product was recrystallized in MeOH/Et2O to afford a white solid (2.96 g, 5.63 mmol, 80% yield).

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.71 (s, 1H, N*H*), 7.80 (s, 1H, C8H), 6.54 (broad s, 2H, NH_2), 5.69 (d, *J* = 1.5 Hz, 1H, $C_{1'}$H), 5.62 (broad s, 1H, $C_{2'}$OH), 4.35 (dd, *J* = 8.1, 5.0 Hz, 1H, $C_{2'}$H), 4.27 (dd, *J* = 5.0, 1.5 Hz, 1H, $C_{3'}$H), 4.09 (dd, *J* = 12.8, 2.9 Hz, 1H, $C_{5'}H_1$), 4.00 (dt, *J* = 8.1, 2.9 Hz, 1H, $C_{4'}$H), 3.93 (dd, *J* = 12.8, 2.9 Hz, 1H, $C_{5'}H_2$), 1.14 – 0.97 (m, 28H, $((CH_3)_4(CH)_2SiO)_2$).

^{13}C NMR (101 MHz, DMSO) δ 157.00, 154.36, 150.96, 134.55, 116.86, 88.33, 81.49, 74.31, 70.10, 61.19, 17.83, 17.68, 17.66, 17.63, 17.44, 17.37, 17.33, 17.27, 13.24, 12.92, 12.79, 12.49.

HR-MS: calculated for $[C_{22}H_{40}N_5O_6Si_2]^+$: 526.2439, Found: 526.2510.



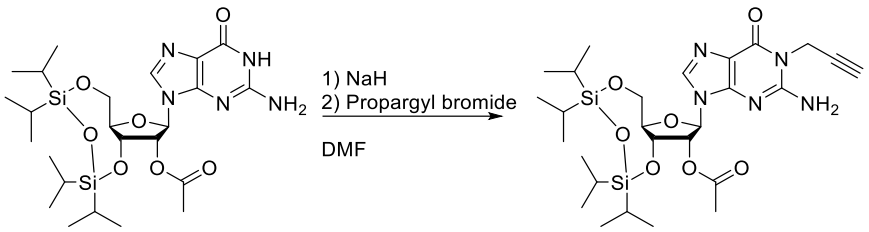
16 17

2′-O-acetyl-3′,5′-O-tetraisopropyldisiloxano-2-amino-1,9-dihydro-9-β-D-ribofuranosyl-6*H*-purin-6-one. 16 (2.95, 5.61 mmol, 1 equiv.) was suspended in anh. CH_3CN with DMAP (68.5 mg, 0.561 mmol, 0.1 equiv.) and TEA (2.0 mL, 14.3 mmol, 2.5 equiv.). The reaction was cooled to 0 °C in an ice bath, and acetic anhydride (0.65 mL, 6.9 mL, 1.2 equiv.) was added dropwise. The reaction was stirred on ice for 1 h, then at room temperature for 3 h. MeOH (5 mL) was added to quench the reaction, and the solvents were removed under reduced pressure. The resulting residue was dissolved in DCM, then washed with brine and dried over Na_2SO_4. DCM was removed under reduced pressure and the product was purified by column chromatography (0% to 10% MeOH/DCM) to afford an off-white foam (2.35 g, 4.14 mmol, 74% yield).

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.72 (s, 1H, N*H*), 7.88 (s, 1H, C8H), 6.44 (broad s, 2H, NH_2), 5.89 (d, J = 2.8 Hz, 1H, $C_{1'}H$), 5.60 (dd, J = 6.0, 2.8 Hz, 1H, $C_{2'}H$), 4.66 (dd, J = 7.3, 6.0 Hz, 1H, $C_{3'}H$), 4.05 (dd, J = 12.4, 3.3 Hz, 1H, $C_{5'}H_1$), 3.96 (dd, J = 12.4, 3.3 Hz, 1H, $C_{4'}H$), 3.92 (dt, J = 7.3, 3.3 Hz, 1H, $C_{5'}H_2$), 2.10 (s, 3H, $COCH_3$), 1.09 – 0.97 (m, 29H, $((CH_3)_4(CH)_2SiO)_2$).

^{13}C NMR (101 MHz, DMSO) δ 169.43, 157.08, 154.36, 151.10, 135.43, 117.20, 85.18, 82.42, 75.11, 69.98, 61.86, 20.87, 17.75, 17.65, 17.60 (2), 17.29, 17.24 (2), 17.19, 13.14, 12.91, 12.68, 12.62.

HR-MS: calculated for $[C_{24}H_{42}N_5O_7Si_2]^+$: 568.2545, Found: 568.2615.



17 18

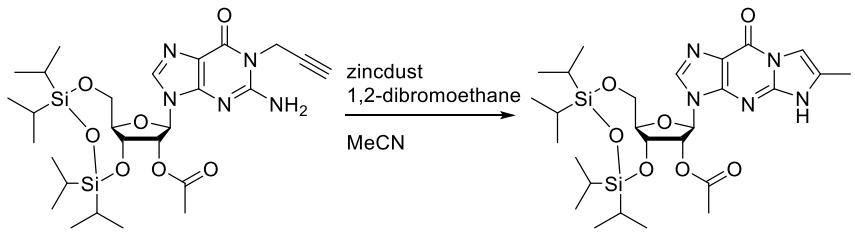
2′-O-acetyl-3′,5′-O-tetraisopropyldisiloxano-1-propargyl-2-amino-1,9-dihydro-9-β-D-ribofuranosyl-6*H*-purin-6-one. 17 (1.60 g, 2.82 mmol, 1 equiv.) was dissolved in anh. DMF and cooled to 0 °C in an ice bath. NaH (60% dispersion in mineral oil) (0.201 g, 5.02 mmol, 1.8 equiv.) was added portion-wise over 5 min, and the reaction was stirred at room temperature for 1 h. Propargyl bromide (0.52 mL, 6.7 mmol, 2.4 equiv.) was added and the reaction was stirred at room temperature for 3 h. DMF was removed under reduced pressure, and the residue was re-dissolved in DCM, washed with brine, and dried over Na_2SO_4. DCM was removed under reduced pressure, and the product was purified via column chromatography (0% to 7% MeOH/DCM) to afford a pale yellow foam (0.744 g, 1.23 mmol, 44% yield).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.73 (s, 1H, C8H), 5.85 (d, *J* = 1.2 Hz, 1H, $C_{1'}$H), 5.68 (dd, *J* = 5.2, 1.2 Hz, 1H, $C_{2'}$H), 5.19 (broad s, 2H, NH_2), 4.96 (d, *J* = 2.6 Hz, 1H, NC*H*$_2$CCH), 4.94 (d, *J* = 2.6 Hz, 1H, NC*H*$_2$CCH), 4.72 (dd, *J* = 8.6, 5.2 Hz,

1H, $C_{3'}H$), 4.18 – 4.10 (m, 1H, $C_{5'}H_1$), 4.05 – 4.02 (m, 1H, $C_{5'}H_2$), 4.01 – 3.98 (m, 1H, $C_{4'}H$), 2.38 (t, *J* = 2.6 Hz, 1H, $NCH_2CC\mathit{H}$), 2.16 (s, 3H, $COCH_3$), 1.14 – 0.99 (m, 29H, $((CH_3)_4(CH)_2SiO)_2$).

^{13}C NMR (101 MHz, $CDCl_3$) δ 169.27, 155.70, 152.91, 147.77, 135.96, 118.34, 86.73, 81.97, 76.78, 75.31, 74.13, 69.02, 60.56, 30.52, 20.72, 17.44, 17.38, 17.34, 17.29, 16.97, 16.93 (2), 16.89, 13.41, 12.98, 12.92, 12.72.

HR-MS: calculated for $[C_{27}H_{44}N_5O_7Si_2]^+$: 606.2701, Found: 606.2771.



18 **19**

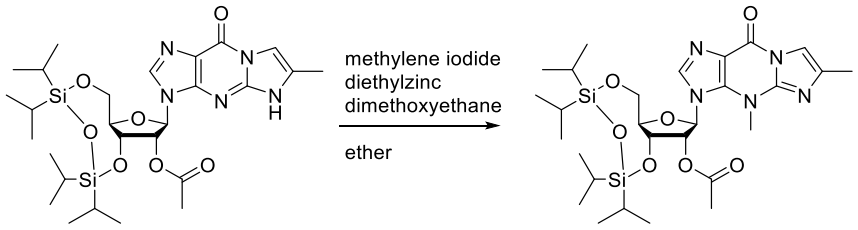
2′-O-acetyl-3′,5′-O-tetraisopropyldisiloxano-3-β-D-ribofuranosyl-9-hydro-6-methyl-5*H*-oxoimidazo(1,2-a)purine. Zinc dust (1.02 g, 15.6 mmol, 6 equiv.) was suspended in anh. CH_3CN (40 mL) and heated to reflux. Dibromoethane (0.65 mL, 7.51 mmol, 3 equiv.) was added and the reaction was stirred under reflux for 2 h. Additional dibromoethane (0.65 mL, 7.51 mmol, 3 equiv.) was added and the reaction was stirred under reflux for 1 h. **18** (1.54 g, 2.54 mmol, 1 equiv) was dissolved in anh. DCM and transferred to the reaction flask. The reaction was stirred under reflux for 24 h. CH_3CN was removed under reduced pressure, and the residue was diluted with DCM. This was washed with an ammonium chloride solution, dried over Na_2SO_4, and DCM was removed under reduced pressure. The product was purified via column

chromatography (0% to 2% MeOH/EtOAc) to afford a pale yellow foam (0.672 g, 1.11 mmol, 44% yield).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.90 (broad s, 1H, N*H*), 7.82 (s, 1H, C8*H*), 7.31 (d, *J* = 1.1 Hz, 1H, NC*H*CCH$_3$), 5.90 (s, 1H, C$_{1'}$H), 5.63 (d, *J* = 5.4 Hz, 1H, C$_{2'}$H), 4.66 (dd, *J* = 8.5, 5.4 Hz, 1H, C$_{3'}$H), 4.14 – 4.07 (m, 1H, C$_{4'}$H), 4.01 – 3.93 (m, 2H, C$_{5'}$H$_2$), 2.27 (d, *J* = 1.1 Hz, 3H, NCHCC*H*$_3$), 2.10 (s, 3H, COCH$_3$), 1.07 – 0.93 (m, 34H, ((CH$_3$)$_4$(CH)$_2$SiO)$_2$).

^{13}C NMR (126 MHz, CDCl$_3$) δ 169.41, 151.81, 148.74, 145.81, 135.99, 124.94, 117.17, 104.48, 86.86, 82.04, 75.40, 68.84, 60.48, 20.72, 17.46, 17.37, 17.33, 17.30, 16.98, 16.92, 16.91 (2), 13.40, 12.99, 12.90, 12.71, 11.09.

HR-MS: calculated for [C$_{27}$H$_{44}$N$_5$O$_7$Si$_2$]$^+$: 606.2701, Found: 606.2771.



19 **20**

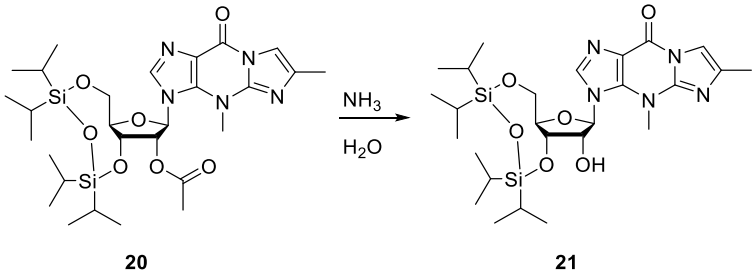
2′-O-acetyl-3′,5′-O- tetraisopropyldisiloxano -3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. Diiodomethane (2.3 mL, 29 mmol, 18 equiv.) was combined with anh. diethyl ether (20 mL), and diethylzinc (1.0 M in hexanes) (14.4 mL, 14.4 mmol, 9.1 equiv.) was added dropwise so that the temperature of the reaction never exceeded 30 °C. The reaction was stirred at room temperature for 40 min, then cooled to 0 °C in an ice bath. Anh. dimethoxyethane (20 mL) was added

dropwise, and a solution of **19** (0.960 g, 1.58 mmol, 1 equiv.) was added quickly to the reaction. After stirring for 5 min, the reaction was quenched with a solution of ammonium bicarbonate and filtered over a pad of celite. The filtrate was diluted with DCM, and washed with a solution of sodium thiosulfate, then with brine. The organic layer was dried over Na_2SO_4 and the solvents were removed under reduced pressure. The product was purified via column chromatography (0% to 5% MeOH/DCM) to afford a white foam (0.517 g, 0.834 mmol, 53% yield).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.82 (s, 1H, C8H), 7.36 (d, *J* = 1.1 Hz, 1H, NC*H*CCH$_3$), 6.13 (s, 1H, C$_{1'}$H), 5.59 (d, *J* = 5.1 Hz, 1H, C$_{2'}$H), 4.57 (dd, *J* = 8.8, 5.1 Hz, 1H, C$_{3'}$H), 4.14 (dd, *J* = 13.3, 2.3 Hz, 1H, C$_{5'}$H$_1$), 4.05 (dt, *J* = 8.8, 2.3 Hz, 1H, C$_{4'}$H), 3.99 (s, 3H, NCH$_3$), 3.96 (dd, *J* = 13.3, 2.3, 1H, C$_{5'}$H$_2$), 2.25 (d, *J* = 1.1 Hz, 3H, NCHCC*H*$_3$), 2.14 (s, 3H, COCH$_3$), 1.07 – 0.92 (m, 28H, ((CH$_3$)$_4$(CH)$_2$SiO)$_2$).

^{13}C NMR (101 MHz, $CDCl_3$) δ 169.10, 152.13, 142.40, 138.48, 138.04, 133.88, 106.82, 87.98, 82.86, 75.82, 68.25, 59.78, 34.02, 20.51, 17.42, 17.32, 17.26 (2), 16.94, 16.86, 16.79 (2), 14.31, 13.42, 12.95, 12.83, 12.68.

HR-MS: calculated for $[C_{28}H_{46}N_5O_7Si_2]^+$: 620.2858, Found: 620.2922.



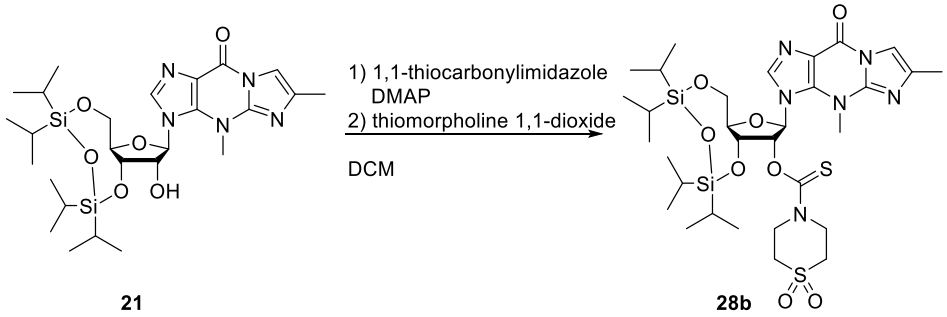
20 **21**

3′,5′-O-tetraisopropyldisiloxano-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 20 (0.516 g, 0.832 mmol) was dissolved in methanolic ammonia (4.0 M) and stirred for 2 h at room temperature. Methanol was removed under reduced pressure, and the product was purified via column chromatography to afford a white solid (0.412 g, 0.713 mmol, 86% yield).

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.03 (s, 1H, C8H), 7.40 (d, J = 1.1 Hz, 1H, NC*H*CCH$_3$), 6.29 (s, 1H, C$_{1'}$H), 6.00 (broad s, 1H, C$_{2'}$OH), 4.58 (d, J = 4.4 Hz, 1H, C$_{2'}$H), 4.33 (dd, J = 9.2, 4.4 Hz, 1H, C$_{3'}$H), 4.25 (d, J = 13.3 Hz, 1H, C$_{5'}$H$_1$), 4.15 (dt, J = 9.2, 2.2 Hz, 1H, C$_{4'}$H), 4.07 (s, 3H, NCH$_3$), 3.98 (dd, J = 13.3, 2.2 Hz, 1H, C$_{5'}$H$_2$), 2.24 (d, J = 1.1 Hz, 3H, NCHCC*H*$_3$), 1.15 – 0.89 (m, 28H, ((CH$_3$)$_4$(CH)$_2$SiO)$_2$).

^{13}C NMR (126 MHz, DMSO) δ 152.08, 142.85, 139.77, 137.80, 134.64, 116.61, 106.17, 91.30, 81.51, 74.97, 68.81, 60.27, 34.47, 17.85, 17.70, 17.66, 17.62, 17.42, 17.35, 17.31, 17.23, 14.62, 13.25, 12.85, 12.80, 12.43.

HR-MS: calculated for [$C_{26}H_{44}N_5O_6Si_2$]$^+$: 578.2752, Found: 578.2822.



21 28b

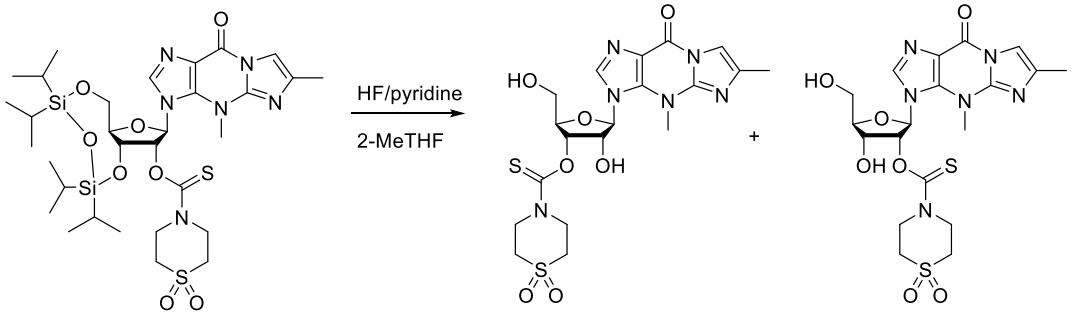
2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-3′,5′-O- tetraisopropyldisiloxano -3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 21

(0.405 g, 0.701 mmol, 1 equiv.) was dissolved in anh. DCM with 1,1′-thiocarbonyldiimidazole (0.175 g, 0.981 mmol, 1.4 equiv.) and DMAP (8.6 mg, 70 μmol, 0.1 equiv.). The reaction was stirred at room temperature for 2 h, and thiomorpohline 1,1-dioxide (0.114 g, 0.841 mmol, 1.2 equiv.) was added. The reaction was fitted with a reflux condenser and heated under reflux for 16 h. DCM was removed under reduced pressure and the product was purified via column chromatography to afford a white solid (0.496 g, 0.657 mmol, 94% yield).

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.11 (s, 1H, C8H), 7.40 (d, J = 1.1 Hz, 1H, NC*H*CCH$_3$), 6.64 (s, 1H, C$_{1'}$H), 6.47 (d, J = 5.3 Hz, 1H, C$_{2'}$H), 4.66 (dd, J = 8.9, 5.3 Hz, 1H, C$_{3'}$H), 4.52 – 4.36 (m, 2H, NC*H*$_2$CH$_2$S), 4.36 – 4.25 (m, 2H, NC*H*$_2$CH$_2$S), 4.23 (d, J = 2.7 Hz, 1H, C$_{5'}$H$_1$), 4.19 (dt, J = 8.9, 2.7 Hz, 1H, C$_{4'}$H), 4.03 (s, 3H, NCH$_3$), 4.02 – 3.96 (m, 1H, C$_{5'}$H$_2$), 3.49 – 3.40 (m, 2H, NCH$_2$C*H*$_2$S), 3.18 – 3.04 (m, 2H, NCH$_2$C*H*$_2$S), 2.22 (d, J = 1.1 Hz, 3H, NCHCC*H*$_3$), 1.10 – 0.95 (m, 28H, ((CH$_3$)$_4$(CH)$_2$SiO)$_2$).

^{13}C NMR (101 MHz, DMSO) δ 187.00, 151.99, 142.77, 139.88, 137.88, 135.27, 116.43, 106.36, 88.05, 82.50, 82.35, 68.94, 60.33, 51.64, 51.16, 48.87, 44.31, 34.29, 17.75, 17.66, 17.58 (2), 17.30 (3), 17.16, 14.58, 13.13, 12.83, 12.67, 12.24.

HR-MS: calculated for [$C_{31}H_{51}N_6O_8S_2Si_2$]$^+$: 755.2670, Found: 755.2748.



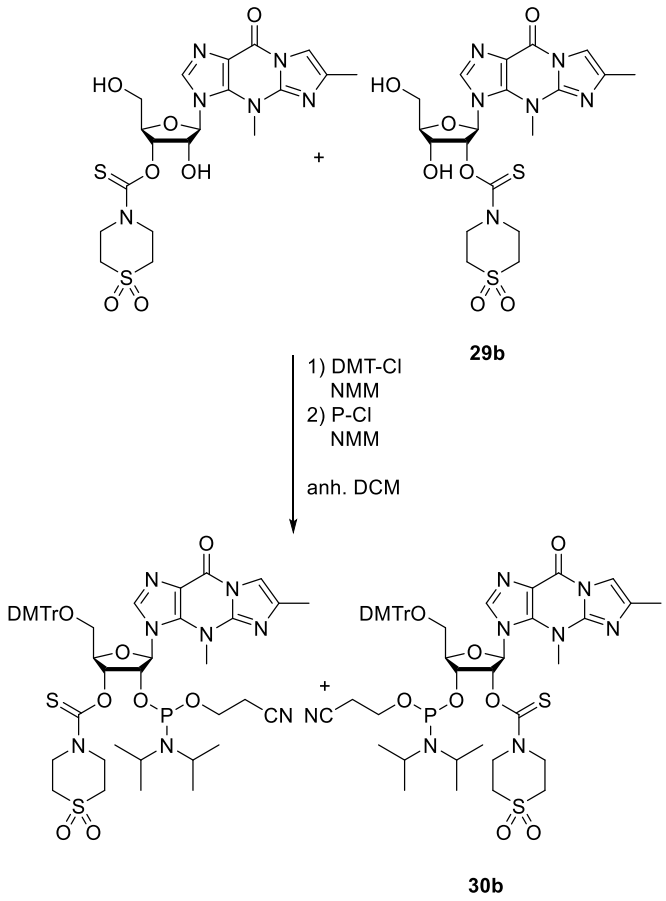
28b **29b**

2′-O-(1,1-dioxothiomorpholine-4-carbothioate)-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine. 28b (0.350 g, 0.464 mmol, 1 equiv) was suspended in anhydrous 2-MeTHF (7 mL) in a plastic falcon tube. Anh. pyridine (22.4 µL, 0.278 mmol, 0.6 equiv.) was added and the reaction was cooled to 0 °C in an ice bath. HF/pyridine (70% HF/ 30% pyridine) (70.8 µL, 2.72 mmol, 5.9 equiv.) was added dropwise with stirring and the reaction was stirred at room temperature for 2 h. The reaction was diluted in 10% MeOH/DCM, washed with water, and dried over Na_2SO_4. The solvents were removed under reduced pressure and the product was purified via column chromatography (0% to 10% MeOH/DCM) to afford a white foam (0.178g, 0.347 mmol, 75% yield). NMR analysis suggests a mixture of 2′ and 3′ isomers.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.24 (s, 1H, C8H), 7.41 (dq, *J* = 5.1, 1.3 Hz, 1H, NC*H*CCH$_3$), 6.25 (d, *J* = 6.2 Hz, 1H, C$_{1'}$H), 6.22 (d, *J* = 5.5 Hz, 1H, C$_{3'}$OH), 5.79 – 5.74 (m, 1H, C$_{3'}$H), 5.34 (t, *J* = 5.3 Hz, 1H, C$_{5'}$OH), 4.81 (q, *J* = 5.7 Hz, 1H, C$_{2'}$H), 4.46 – 4.42 (m, 2H, NC*H*$_2$CH$_2$S), 4.41 – 4.39 (m, 1H, C$_{4'}$H), 4.24 (q, *J* = 5.8 Hz, 2H, NC*H*$_2$CH$_2$S), 4.12 (s, 3H, NCH$_3$), 3.76 – 3.71 (m, 2H, C$_{5'}$H$_2$), 3.45 – 3.37 (m, 2H, NCH$_2$C*H*$_2$S), 3.27 – 3.20 (m, 2H, NCH$_2$C*H*$_2$S), 2.24 (d, *J* = 1.3 Hz, 3H, NCHCC*H*$_3$).

^{13}C NMR (126 MHz, DMSO) δ 187.13, 151.95, 150.08, 142.89, 140.61, 137.88, 136.60, 135.66, 124.38, 116.08, 106.34, 88.28, 83.67, 79.85, 73.64, 60.71, 55.39, 51.28, 50.98, 48.57, 44.06, 34.34, 14.62, 14.60.

HR-MS: calculated for $[C_{19}H_{25}N_6O_7S_2]^+$: 513.1148, Found: 513.1216.



2′-O-(1,1-dioxothiomorpholine-4-carbothioate)- 5′-O-(4,4′-dimethoxytrityl)-3-β-D-ribofuranosyl-4,9-dihydro-4,6-dimethyl-9*H*-oxoimidazo(1,2-a)purine-3′-[2-cyanoethyl *N*,*N*-bis(1-methylethyl)phosphoramidite]. 29b (0.173 g, 0.338 mmol, 1 equiv.) was suspended in anh. DCM (3 mL) with 4,4′-dimethoxytritylchloride (0.160 g, 0.472 mmol, 1.4 equiv.). 4-Methylmorpholine (52 μL, 4.73 mmol, 1.4 equiv) was

added and the reaction was stirred at room temperature for 1 h. 2-cyanoethoxy-N,N-diisopropylaminochlorophosphine (135 μL, 0.605 mmol, 1.8 equiv.) and 4-methylmorpholine (67 μL, 0.606 mmol, 1.8 equiv.) were added and the reaction was stirred at room temperature for 2 h. The solution was diluted with DCM, washed with brine, and dried over Na_2SO_4. DCM was removed under reduced pressure and the product was purified via column chromatography (0% to 70% acetone/hexane) to afford a white foam (0.220 g, 0.216 mmol, 64% yield).

^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 7.94 (s, 1H), 7.44 – 7.40 (m, 1H), 7.35 (dddd, *J* = 7.7, 6.0, 3.0, 1.4 Hz, 6H), 7.32 – 7.19 (m, 13H), 6.87 – 6.75 (m, 7H), 6.63 – 6.51 (m, 2H), 6.48 (d, *J* = 4.5 Hz, 1H), 5.03 – 4.92 (m, 2H), 4.81 (t, *J* = 18.9 Hz, 2H), 4.38 – 4.28 (m, 1H), 4.20 (s, 1H), 4.19 (s, 3H), 4.16 (s, 2H), 4.07 (td, *J* = 11.0, 9.9, 4.5 Hz, 1H), 3.97 – 3.83 (m, 2H), 3.79 – 3.72 (m, 12H), 3.69 – 3.54 (m, 5H), 3.54 – 3.42 (m, 2H), 3.34 (ddd, *J* = 17.6, 11.1, 3.1 Hz, 1H), 3.29 – 3.00 (m, 9H), 2.73 – 2.66 (m, 2H), 2.54 – 2.43 (m, 2H), 2.29 (t, *J* = 1.3 Hz, 6H), 1.20 – 1.11 (m, 16H), 1.07 (d, *J* = 6.8 Hz, 6H).

^{31}P NMR (126 MHz, CD_3CN) δ 151.86, 150.04, 149.84.

HR-MS: calculated for $[C_{49}H_{60}N_8O_{10}PS_2]^+$: 1015.3533, Found:1015.3602.

CHAPTER 3. DESIGN AND MODIFICATION OF RNA SPINACH APTAMERS

Fluorescence is a valuable tool for biological studies and enables the visualization of cellular processes both *in vivo* and *in vitro*. Fluorescence occurs when a molecule absorbs a photon to transition from the ground state (S_0) to the excited state (S_1) and subsequently emits that energy in the form of luminescence rather than nonradiative decay (Figure 3.1). Fluorescence is usually observed at a longer wavelength, or lower energy, than the corresponding absorption of energy due to vibrational relaxation in the excited state, which occurs on a faster timescale (10^{-12} – 10-^{10} s) than fluorescence (10^{-10} – 10^{-7} s). This change in energy is the foundation of the Stokes shift. Consequently, the Stokes shift of a fluorescent system is affected by changes in the chemical environment of the molecule that result in different vibrational relaxation rates or excited state energy levels.

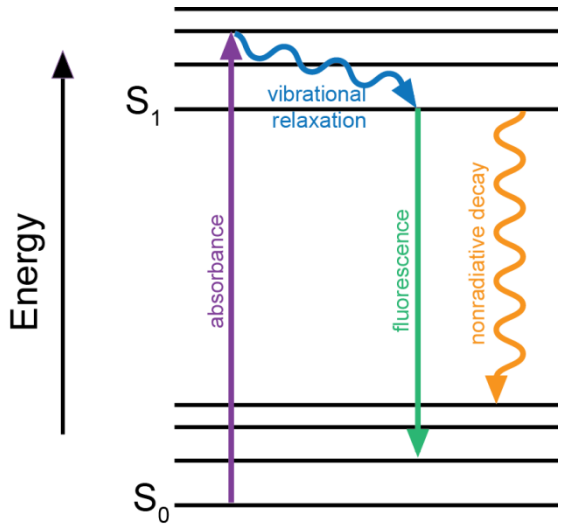


Figure 3.1. Jablonski diagram illustrating processes for absorption (purple), vibrational relaxation (blue), fluorescence (green), and nonradiative decay (orange).

Spinach is often considered the first RNA mimic of green fluorescent protein (GFP) due to the aptamer's ability to bind and induce the fluorescence of DFHBI, an analogue of the HBI chromophore found in GFP.[48] Much like the development of fluorescent proteins, which began from the isolation and characterization of GFP and expanded into a large tool box of brighter fluorescent proteins spanning the UV-visible spectrum, Spinach and other fluorogenic aptamers are being developed for improved spectral properties.[5] Since the original development of Spinach, there have been a number of approaches to expand the toolbox of fluorogenic RNA aptamers, from stabilizing and reselecting variants of Spinach to developing aptamers for novel fluorophores.[53,54,117]

While the selection of new aptamers based on novel fluorophore ligands provides a powerful method for accessing new fluorescent wavelengths, the process can be time and labor intensive. An alternative method for accessing an expanded toolbox of fluorogenic RNA aptamers follows the historic methodology of fluorescent protein development: using structure guided mutations for spectral tuning.[118,119] The solved crystal structure of the Spinach aptamer has illustrated DFHBI ligand-binding between a three-tiered quadruplex and base triple.[45] The structure thus provides a guideline for nondisruptive mutations. To investigate the feasibility of introducing mutation-induced fluorescence shifts, I have studied Spinach and variants in a multi-pronged approach: 1) studying the effects of different classes of fluorophore ligands bound to a mutated Spinach aptamer, 2) exploring mutations of the base triple which caps Spinach, and 3) introducing noncanonical base mutations to the Spinach aptamer. These studies indicate that base triple mutations in Spinach effectively shift the absorbance and fluorescence spectra of the aptamer only when bound to DFHBI analogues.

3.2.1 Fluorescence effects of aptamer and fluorophore modifications

When developing fluorogenic aptamers, there are generally two options available to access a new excitation or emission wavelength: 1) modify the fluorophore or 2) modify the RNA. There exist several examples of modifying fluorophores, ranging from changing a single heteroatom to modifying positions and types of functional groups on the heterocycle.[48,54,83] However, in most reported studies, changing the RNA typically involves selecting for an entirely new aptamer. There are very few examples of achieving significant changes in excitation or emission wavelengths with simple base mutations to an existing aptamer. Known mutations that induce spectral shifts include: 1) mutating a base triple in *i*Mango-III to shift fluorescence 7 nm,[66] 2) mutating a nucleotide flanking the ligand binding pocket in Broccoli to shift fluorescence 20 nm,[94] and 3) mutating the base triple stacks on top of bound DFHBI in Spinach to induce a 12 nm hypsochromic shift.[45] Of these, the Spinach base triple mutation is of particular interest due to not only the 12 nm shift in the fluorescence spectrum but also the 45 nm blue shift in the absorbance spectrum.

The fluorogenic RNA aptamer Spinach binds its cognate fluorophore, DFHBI, through stacking interactions between a G-quadruplex core and a U-U-A base triple. Prior investigations of this ligand binding pocket demonstrate surprisingly large shifts in excitation and emission wavelengths upon mutating the base triple. The A64U, U32A mutations (changing the base triple from U-U-A to A-U-U) induce a moderate hypsochromic shift in the fluorescence emission maximum from 502 nm to 490 nm and a dramatic hypsochromic shift in the absorbance maximum from 468 nm to 430 nm.[45]

Taking advantage of this reported mutation-induced fluorescence shift, I investigated the effects of this base triple modification on other classes of fluorophore-ligands.

3.2.1.1 Spectral changes of Spinach aptamer when bound to different fluorophores

While fluorogenic RNA aptamers are selected to be specific for a single fluorophore ligand, Spinach exhibits promiscuity in ligand binding. In addition to binding DFHBI (Figure 3.1A), Spinach has also been reported to non-selectively bind and induce the fluorescence of TO1 (Figure 3.1D), the cognate fluorophore of the Mango aptamers, and ThT (Figure 3.1C), a small molecule fluorescent probe for amyloids and G-quadruplexes.[62,120–122] While no studies have explicitly determined the affinity of Spinach for DFHO, the cognate fluorophore for the fluorogenic Corn aptamer, the chemical similarity of DFHBI and DFHO suggests that Spinach may also nonspecifically bind DFHO (Figure 3.1B).[55,56]

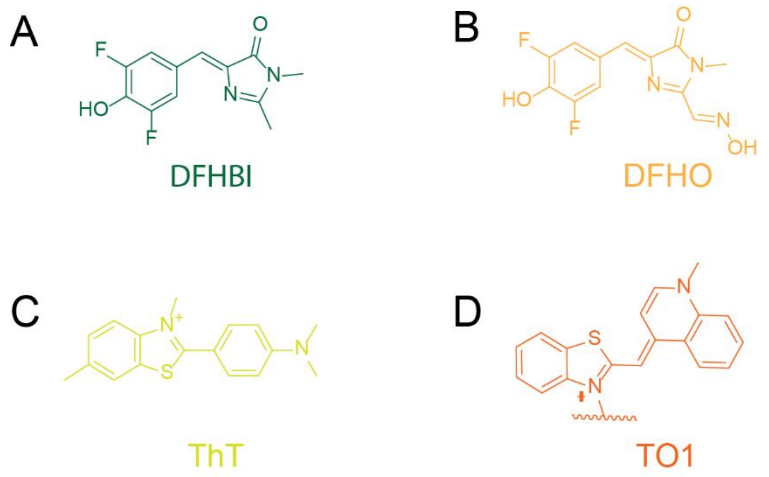


Figure 3.1. Fluorophore ligands complexed with Baby Spinach A-U-U base triple mutation. (A) DFHBI is the cognate ligand for Spinach aptamers. (B) DFHO is the cognate ligand for Corn and Red-Broccoli aptamers. (C) ThT is a small molecule fluorescent probe for G-quadruplexes which has been shown to bind Spinach. (D) TO1 is the cognate ligand for Mango aptamers, and has been shown to bind Spinach.

The most dramatic shifts in absorbance and fluorescence spectra from base mutations have been observed in Spinach and Broccoli, a reselected variant of Spinach.[45,51,94] It remains unclear whether this is a result of a unique feature of the family of DFHBI analogues, or if it is possible to observe similar shifts for other fluorophore binding aptamers. To investigate this, I took advantage of Spinach's promiscuity for other fluorophore-ligands. Using the A-U-U base triple mutation as a scaffold, I investigated the fluorescence of TO1, ThT, and DFHO when bound to Baby Spinach AUU, a miniaturization of Spinach.

While DFHBI-bound Baby Spinach AUU exhibited a large 48 nm hypsochromic shift in its absorbance maxima, no such shifts were observed for TO1-bound or ThT-bound Baby Spinach AUU (Figure 3.2A,C,D). Regardless of the RNA, TO1 bound to either aptamer had an excitation wavelength maximum at 520 nm. Similarly, ThT bound to

either aptamer had a maximum at 450 nm. Fluorescence spectra for both ligands also saw little to no shift as TO1-bound RNA had a consistent fluorescence maximum at 535 nm and ThT-bound RNA had a consistent maximum at 480 nm. Interestingly, Baby Spinach AUU strongly induced the fluorescence of TO1, ThT, and DFHBI, but did not induce the fluorescence of DFHO to the same extent. Still, the 13-fold fluorescence enhancement of DFHO-bound Baby Spinach and the 4-fold enhancement of DFHO-bound Baby Spinach AUU was enough to observe a 14 nm blue shift in fluorescence maximum. While DFHO-bound Corn is normally excited at 510 nm, the absorbance maximum for DFHO-bound Baby Spinach was observed to be 495 nm. Furthermore, the excitation maximum for DFHO-bound Baby Spinach AUU was observed at 471 nm, which is a 24 nm shift from DFHO-bound Baby Spinach and a 39 nm shift from DFHO-bound Corn (Figure 3.2B).

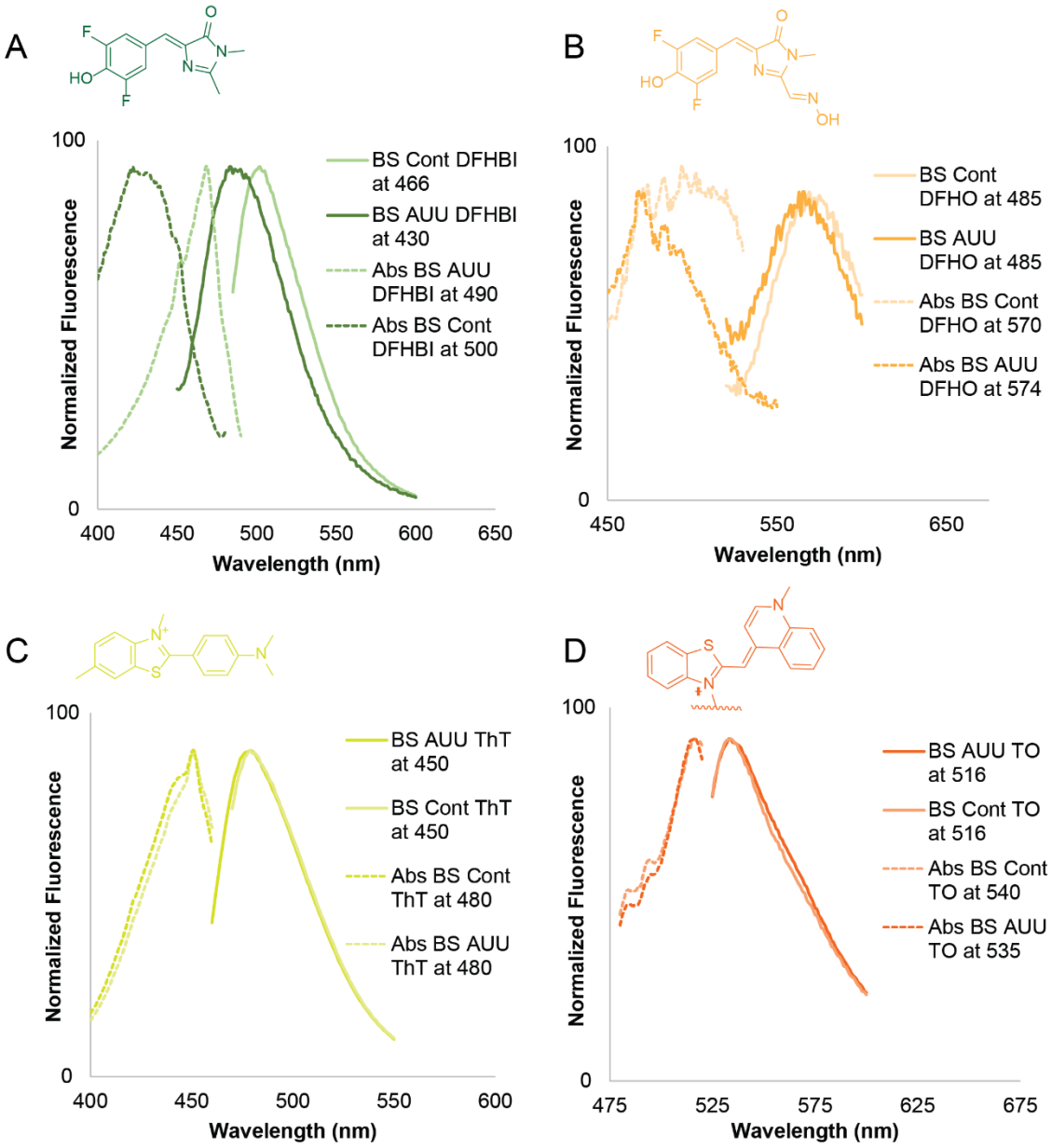


Figure 3.2. Absorbance and fluorescence spectra for Baby Spinach A-U-U base triple mutation when bound to different fluorophore ligands. (A) Baby Spinach AUU bound to DFHBI. (B) Baby Spinach AUU bound to DFHO. (C) Baby Spinach AUU bound to ThT. (D) Baby Spinach AUU bound to TO1.

The absorbance and fluorescence shifts for Baby Spinach and Baby Spinach AUU bound to DFHO suggests the phenomenon of the base-triple induced spectral shifts are specific to DFHBI and analogous fluorophore ligands. Furthermore, the absorbance

spectra of unmodified Baby Spinach and Baby Spinach AUU in the presence of DFHO indicate the presence of a minor UV band, which could account for shifts in the fluorescence spectra when excited.

3.2.1.2 Structure of iSpinach bound to novel ligand

To better understand the structural basis of the spectral changes to Spinach aptamers when bound to DFHBI analogues, I solved the crystal structure of the *i*Spinach variant in complex with DFHBI-1T. While several fluorophore ligands have been created to optimize the brightness of Spinach and its variants, DFHBI-1T is commonly used due to its 1.8x enhanced brightness relative to DFHBI.[54] *i*Spinach, a re-selected variant of Spinach, was chosen for crystallography studies due to its thermostability relative to Spinach, which tends to misfold.[53] Prior crystallographic studies of DFHBI-bound *i*Spinach also displayed a nearly identical structural resemblance to DFHBI-bound Spinach, indicating a conserved ligand binding pocket despite distant mutations for aptamer stabilization.[123]

The structure of DFHBI-1T bound *i*Spinach (Figure 3.3) was solved at 2.01 Å (Supplementary Table S1) and is in good agreement with those structures of Spinach and variants. DFHBI-1T sits on top of a three-tiered quadruplex and is capped by a base triple. Diffraction data were collected to 2.01 A, providing clear electron density for the trifluoromethyl group that distinguishes DFHBI-1T from DFHBI. This functional group packs tightly against the flanking A50 nucleotide, an interaction that could help lock the fluorophore into place in the ligand binding pocket (Figure 3.3D). This restrained positioning might then reduce the possibility of non-radiative decay and result in the ligand's characteristic fluorescence enhancement relative to DFHBI.

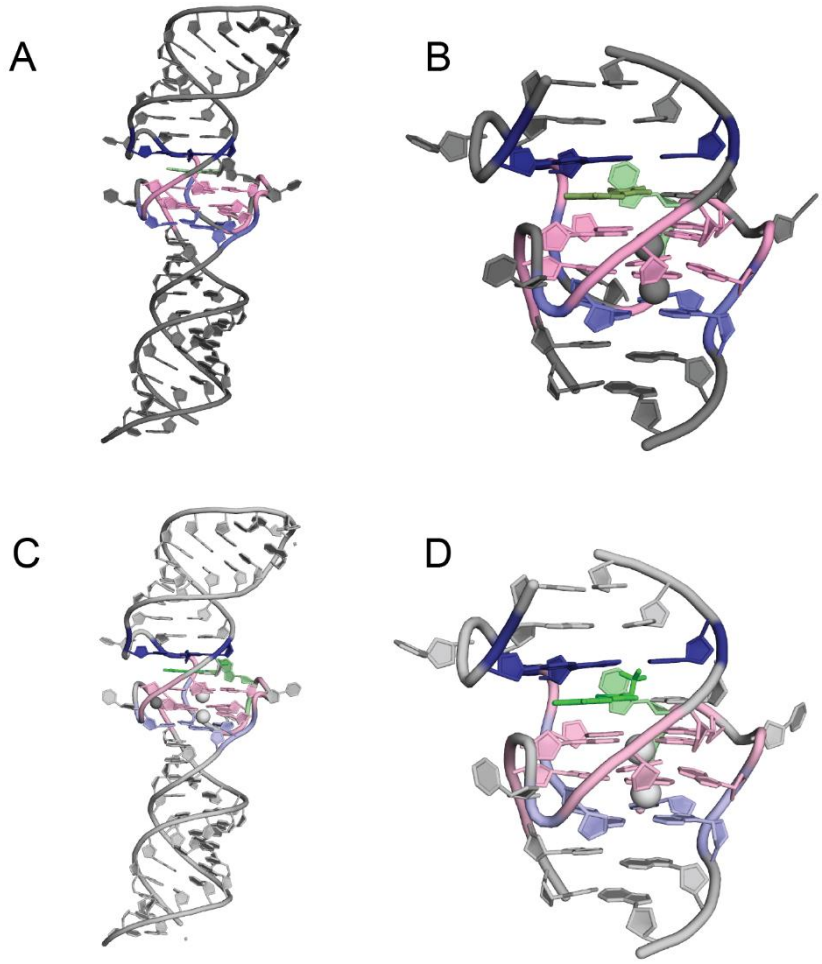


Figure 3.3. Structures for *i*Spinach-DFHBI complex (A) and ligand binding pocket (B) and structures for novel *i*Spinach-DFHBI-1T complex (C) and ligand binding pocket (D). Base triple is highlighted in indigo, G-quadruplex tiers are highlighted in pink, and bottom-most mixed base-quadruple is highlighted in lilac. The flanking A50 nucleotide which packs against the trifluoromethyl group in DFHBI-1T is highlighted in green. (PDB: 5OB3)

While structural studies of base triple mutants of *i*Spinach were also attempted, these mutated RNA constructs did not provide crystals in any initial screening conditions. It is possible that the base triple mutations to the *i*Spinach aptamer were too destabilizing to form well ordered crystals. In the interest of time, additional constructs were not

screened. The solved structure of *i*Spinach bound to a novel DFHBI analogue demonstrates the conserved fold of Spinach variants when bound to similar ligands. When bound to DFHBI-1T, the tight packing of the trifluoromethyl group provides a structural basis for the fluorophore's enhanced brightness upon binding the RNA aptamer.

3.2.2 Tuning fluorescent wavelength with base triple mutations in Spinach

To further study the effects of mutations of the Spinach binding pocket on the fluorescent wavelengths of the fluorophore-aptamer complex, I investigated the effects of various combinations of base triples on Baby Spinach, a miniaturization of the Spinach aptamer.[45] Due to the complex fold of the ligand-binding core of Spinach and variants, mutations must be carefully placed to preserve the aptamer's overall structure and thus its ability to bind DFHBI. Keeping this in mind, I preserved the A64-U32 base pair interaction and cross-checked the existence of known base triples with the RNA base triple database when designing mutant base triples (Table 3.1).[124]

Table 3.1. Sequences of Baby Spinach base triple mutations

CONSTRUCT	SEQUENCE
UUA	GGGACUACGGUGAGGGUCGGG**U**CCAGUAGGUACGCCU ACUG**U**UG**A**GUAGAGUGUGGGCUCCGUAGUCCC
AUU	GGGACUACGGUGAGGGUCGGG**A**CCAGUAGGUACGCCU ACUG**U**UG**U**GUAGAGUGUGGGCUCCGUAGUCCC
GUC	GGGACUACGGUGAGGGUCGGG**G**CCAGUAGGUACGCCU ACUG**U**UG**C**GUAGAGUGUGGGCUCCGUAGUCCC
AAU	GGGACUACGGUGAGGGUCGGG**A**CCAGUAGGUACGCCU ACUG**A**UG**U**GUAGAGUGUGGGCUCCGUAGUCCC
ACU	GGGACUACGGUGAGGGUCGGG**A**CCAGUAGGUACGCCU ACUG**C**UG**U**GUAGAGUGUGGGCUCCGUAGUCCC
AGU	GGGACUACGGUGAGGGUCGGG**A**CCAGUAGGUACGCCU ACUG**G**UG**U**GUAGAGUGUGGGCUCCGUAGUCCC
UAA	GGGACUACGGUGAGGGUCGGG**U**CCAGUAGGUACGCCU ACUG**A**UG**A**GUAGAGUGUGGGCUCCGUAGUCCC
CGG	GGGACUACGGUGAGGGUCGGG**C**CCAGUAGGUACGCCU ACUG**G**UG**G**GUAGAGUGUGGGCUCCGUAGUCCC
GAC	GGGACUACGGUGAGGGUCGGG**G**CCAGUAGGUACGCCU ACUG**A**UG**C**GUAGAGUGUGGGCUCCGUAGUCCC
GCC	GGGACUACGGUGAGGGUCGGG**G**CCAGUAGGUACGCCU ACUG**C**UG**C**GUAGAGUGUGGGCUCCGUAGUCCC
CUG	GGGACUACGGUGAGGGUCGGG**C**CCAGUAGGUACGCCU ACUG**U**UG**G**GUAGAGUGUGGGCUCCGUAGUCCC
UCA	GGGACUACGGUGAGGGUCGGG**U**CCAGUAGGUACGCCU ACUG**C**UG**A**GUAGAGUGUGGGCUCCGUAGUCCC
GGC	GGGACUACGGUGAGGGUCGGG**G**CCAGUAGGUACGCCU ACUG**G**UG**C**GUAGAGUGUGGGCUCCGUAGUCCC

Even taking these precautions, of the 13 base triple mutations studied, seven constructs exhibited no discernible fluorescent signal above background suggesting unfolded or misfolded RNA. Of the remaining six base triple mutations (Figure 3.4), all exhibited large reductions in fluorescence signal compared to native Baby Spinach. Surprisingly, four of the six constructs consistently exhibited 48 nm and 12 nm hypsochromic shifts in excitation and emission maxima (Figure 3.5). The remaining two constructs had emission maxima at 502 nm, consistent with unmodified Baby Spinach. The consistency of the excitation and emission spectra of Baby Spinach with modified base triples suggests altering the electronic environment of the ligand binding pocket might excite a minor band in DFHBI's UV spectrum.

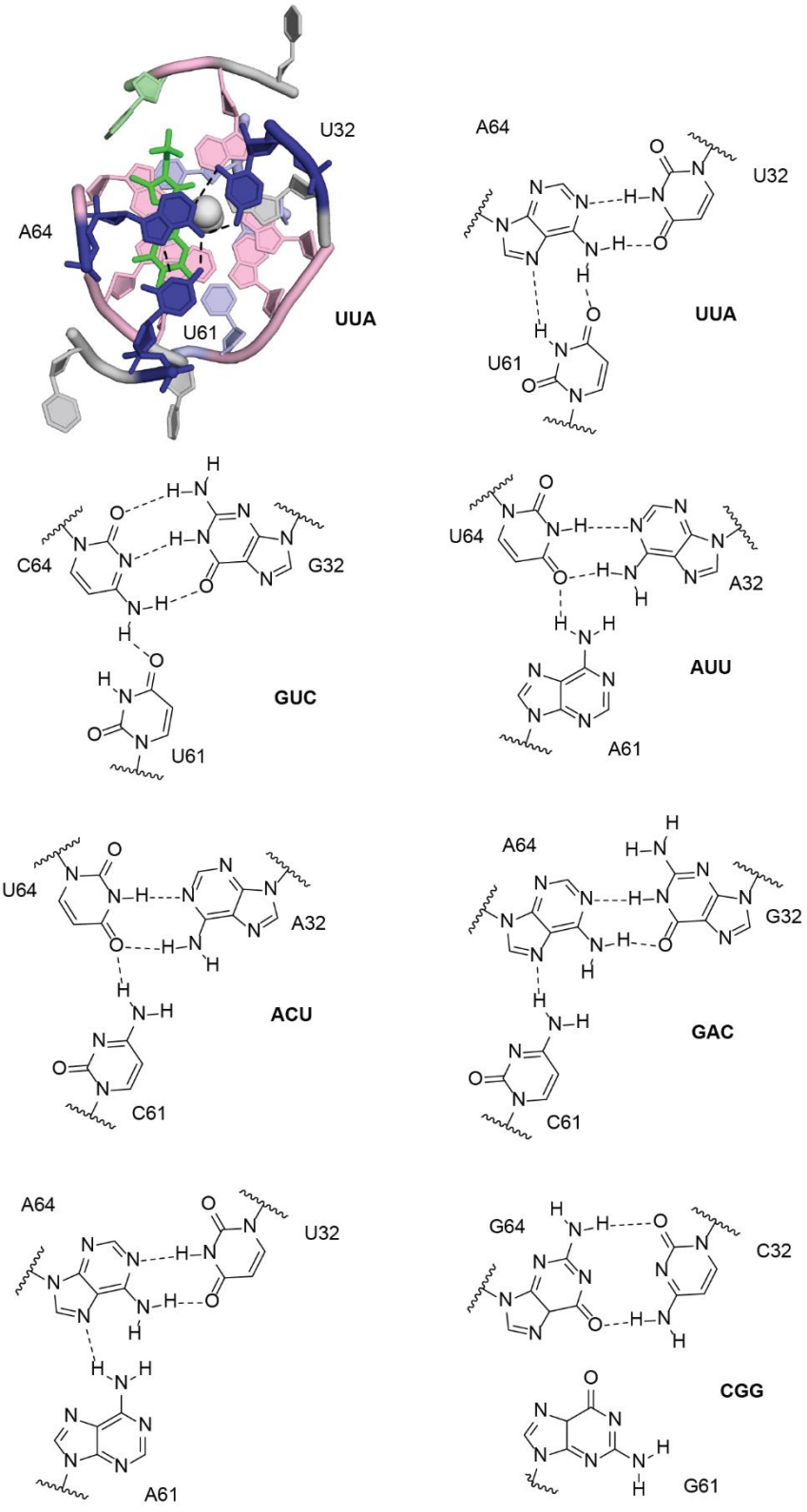


Figure 3.4. Base triple mutations that exhibited fluorescence turn on with predicted hydrogen bonding. (PDB: 5OB3)

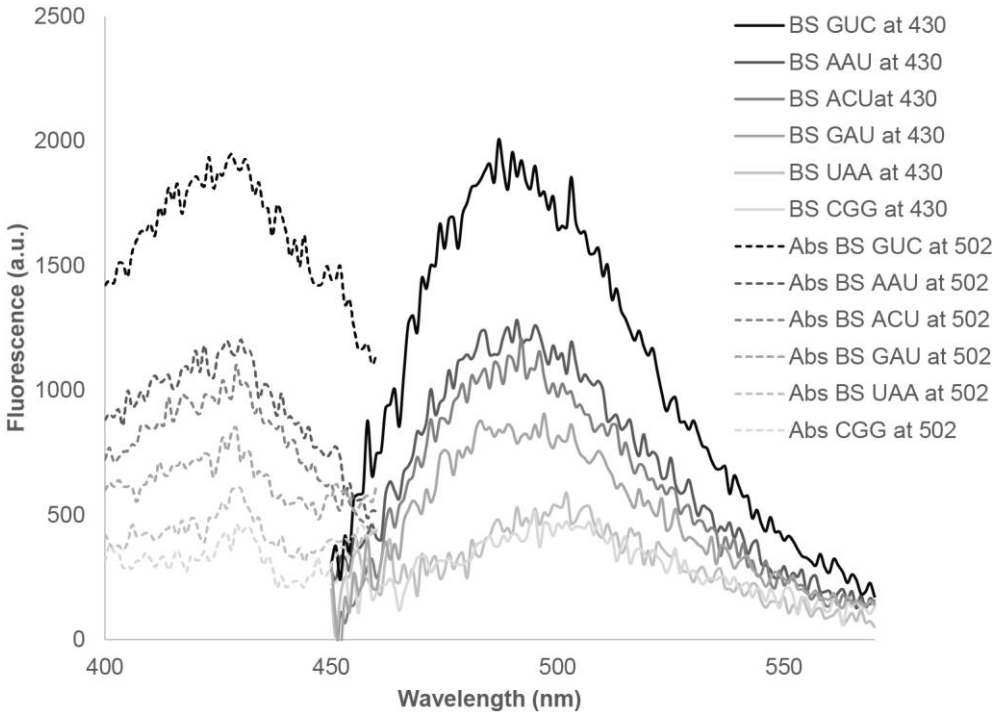


Figure 3.5. Absorbance and fluorescence spectra for Baby Spinach base triple mutations which exhibited florescence turn on.

3.2.3 Split Baby Spinach

After probing the base triple mutations in Baby Spinach, I next sought to investigate the use of modified nucleoside analogues to shift the fluorescence spectrum of the fluorophore-bound RNA aptamer. The introduction of modified nucleosides to the aptamer requires the use of solid-phase RNA synthesis, but conventional coupling yields limit the overall length of chemically synthesized RNA constructs to approximately 50 nt in length.[125] The full-length Spinach aptamer, which is 94 nt, is far beyond the capabilities of standard solid-phase synthesis. Even the miniaturization of the aptamer, Baby Spinach, reaches the upper limits of chemical synthesis at 52 nt. In order to test the possibility to tuning the fluorescent spectrum of Baby Spinach with

modified nucleosides, it was first necessary to further reduce the length of the RNA sequence with minimal impact on the aptamer's overall folding and brightness.

3.2.3.1 Designing split aptamer for solid phase synthetic accessibility

Based on the structure of Spinach, and extrapolating to Baby Spinach, the ligand binding pocket of the RNA aptamer is comprised of a three-tiered G-quadruplex topped by a base triple (Figure 3.6A). This core is flanked on either side by a duplex stem, and as demonstrated by one solved crystal structure of Spinach, the P3 stem of the aptamer can be truncated and split without significant effects on folding or fluorescence. As Spinach could be split at the P3 stem while maintaining both structure and fluorescence,[45] it was probable that Baby Spinach could be similarly split. This division of the aptamer into two separate strands of RNA which could then be refolded in the presence of DFHBI offers a straightforward method for accessing shorter RNA sequences that are chemically accessible (Figure 3.6B).

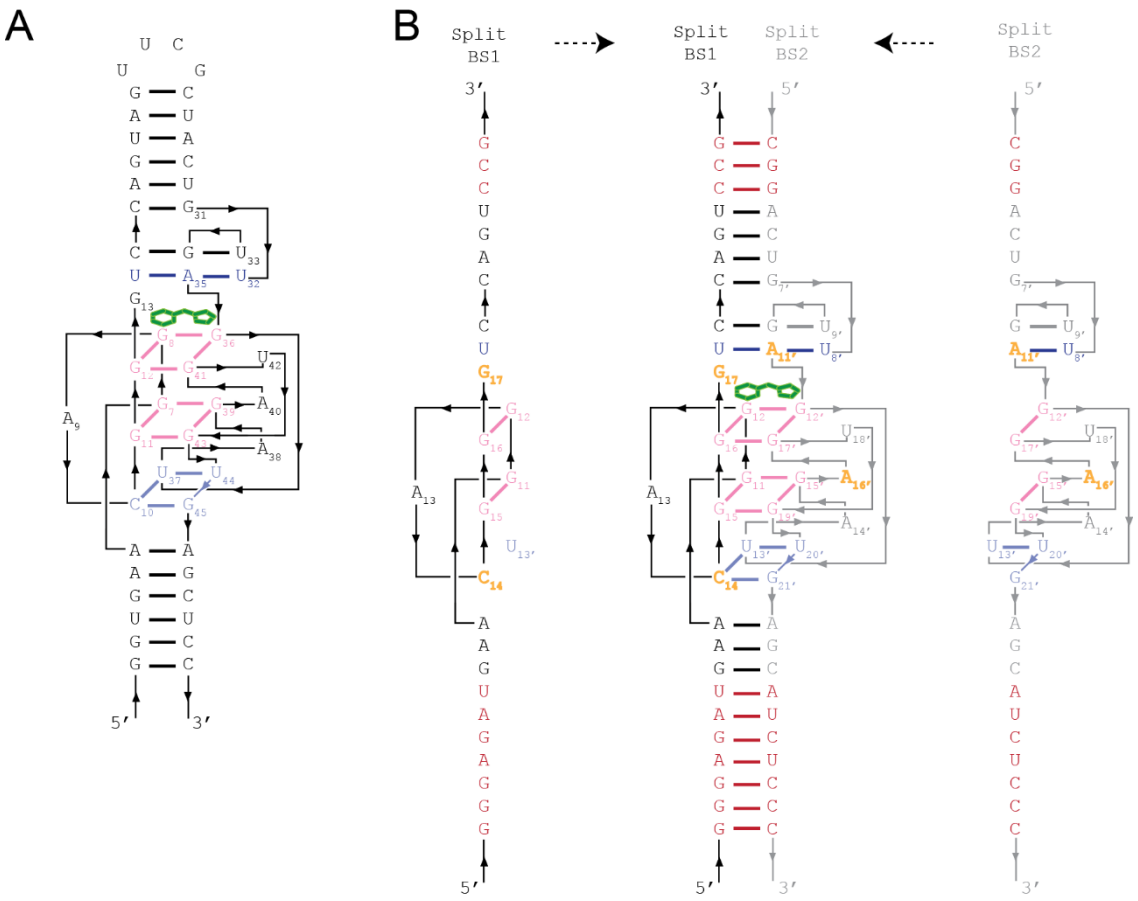


Figure 3.6. Secondary structures for unmodified Baby Spinach (A) and Split Baby Spinach complex (B). DFHBI ligand is shown in green, base triple is highlighted in indigo, G-quadruplex tiers are highlighted in pink, and bottom most mixed quadruplex is highlighted in lilac. Chosen sites of modification are highlighted in yellow. Red bases represent extended P1 and P3 stems for stability in Split Baby Spinach.

The initial split of Baby Spinach to yield 5′- GGTGGAAGGACGGGUCCAGU (BS1) and 5′- ACTGUUGAGUAGAGUGUGAGCUCC (BS2) based on the RNA's predicted secondary structure was unsuccessful. The split aptamer's fluorescence signal was heavily impacted, most likely as a result of inadequate folding. To address this, I extended the P1 and P3 duplex stems of the aptamer to help trigger and stabilize the overall structure. In extending the duplex stems each by 3 base pairs (from 5 bp to 8 bp

for P1 and from 3 bp to 6 bp for P3), the extended Split Baby Spinach aptamer produced not just comparable, but higher fluorescence signals than that of Baby Spinach (Figure 3.7).

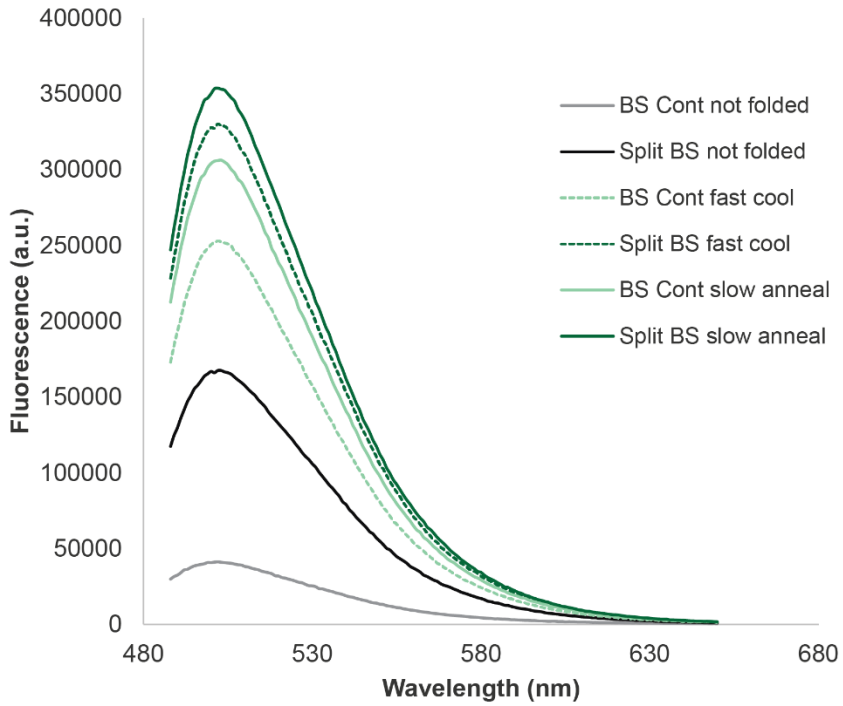


Figure 3.7. Fluorescence spectra for unmodified Baby Spinach and Split Baby Spinach using different folding protocols.

Additionally, while literature protocols for folding Spinach consisted of several steps involving heating, snap cooling, re-heating, and slow cooling, Split Baby Spinach was especially amenable to the simple slow anneal typical of duplex nucleic acids. Under the heat, snap cool, re-heat, and slow cool method of folding, Split Baby Spinach had a 31% increase in fluorescence signal compared to Baby Spinach. With a slow anneal, Split Baby Spinach demonstrated a 16% increase in fluorescence signal compared to Baby Spinach. When comparing the two folding methods, the slow anneal produced the brightest aptamer-fluorophore complex with a 6% increase in signal for Split Baby

Spinach. Notably, even when the split aptamer is not subjected to any folding protocols, 47% of the aptamer still seems to fold and fluoresce just in the presence of buffer and ligand. These measurements confirm Split Baby Spinach's suitability for fluorescence studies, and the shortened length of each RNA strand (26 nt and 31 nt) make it significantly more amenable to solid-phase RNA synthesis.

3.2.3.2 Fluorescence effects of modifying Split Baby Spinach with aminopurine and pyrrolocytosine

Previous studies have demonstrated that small mutations to a fluorogenic RNA aptamer can alter the chemical environment of a bound fluorophore-ligand and induce changes in the fluorescent spectrum. With the synthetic accessibility provided by Split Baby Spinach, I thus set out to determine if these moderate spectral changes could be further shifted by incorporating mutations beyond the four canonical RNA bases. 2-Aminopurine (AP) and pyrrolocytosine (PyC) are two weakly fluorescent purine analogues (Figure 3.8A) that are commonly used to report on nucleic acid structure and folding.[126] Using these analogues, I modified several positions surrounding the ligand binding pocket of Split Baby Spinach with either AP or PyC and probed the fluorescence spectrum of the aptamer-fluorophore complex (Figure 3.8B,C). I also tested these modifications in the presence of three fluorophore ligands: DFHBI, DFHBI-1T, and DFHO.

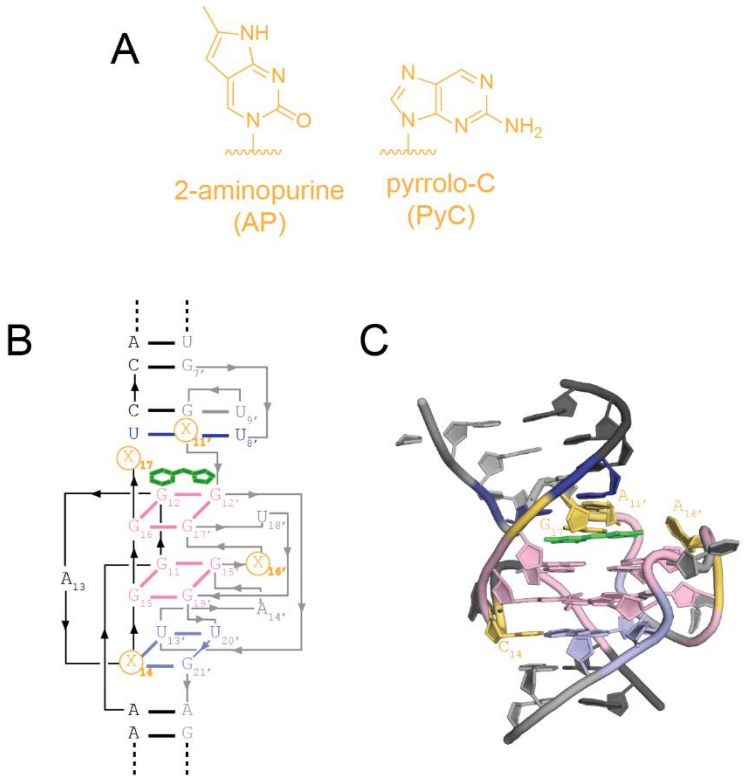


Figure 3.8. 2-Aminopurine and pyrrolocytosine purine nucleoside analogues (A) which were used to modify various positions in Split Baby Spinach. The predicted secondary structure (B) and tertiary structure (C) of Split Baby Spinach which positions for modification highlighted in yellow. (PDB: 4TS0)

AP and/or PyC were introduced into four positions surrounding the ligand binding pocket in Split Baby Spinach: BS1 C14, BS1 G17, BS2 A11, and BS2 A16 (Table 3.2). While all four modifications impacted the fluorogenic aptamer's overall fluorescence signal to some extent, all but BS2 A11 (Split BS 4) still exhibited significant fluorescence turn-on (Figure 3.9A). The BS2 A11AP mutation, despite a relatively conservative change, destroyed the aptamer's fluorescence signal. As BS2 A11 participates in the base triple that stacks on top of the bound fluorophore and G-quadruplex, modifying the position of the exocyclic amine on the purine likely impacted the hydrogen bonding within the base triple and subsequently destabilized

RNA folding. Conversely BS1 C14 (Split BS 8), which participates in a base quadruple that stacks below the G-quadruplex, can be mutated to PyC and still demonstrate fluorescence turn on (Figure 3.9A). Surprisingly the minor modification of BS1 G17 to AP (Split BS 6), effectively removing a carbonyl group, produced a drastic decrease in fluorescence signal. BS1 G17 is a singular guanine that stacks between the G-quadruplex and base triple, in the same plane as the bound ligand. As the overall shape and size of AP relative to guanosine is comparable, it's unlikely that the mutation has overcrowded the ligand binding pocket to reduce the ligand binding affinity. More likely, the change in the Watson-Crick base pairing edge of the nucleoside impacted hydrogen bonding between BS1 G17 and the bound ligand thus impacting the binding affinity of the fluorophore ligand. As the binding pocket of Spinach is predicted to be preformed in the absence of ligand, it is unclear how this mutation affects the RNA fold. Of all mutations studied, BS1 C14 and BS2 A16 are the best tolerated while BS1 G17 seems reasonably tolerated and BS2 A11 is not tolerated at all.

Table 3.2. Modified sequences of Split Baby Spinach with reference numbers used in following figures and text.

REF. NUMBER	MODIFIED CONSTRUCTS	SEQUENCE NAME
1	GGGAGATGAAGGACGGGTCCAGTCCG	Split BS 1 WT
	CGGACTGTTGAGTAGAGTGTGAGCATCTCCC	Split BS 2 WT
2	GGGAGATGAAGGACGGGACCAGTCCG	Split BS 1 AUU
	CGGACTGTTGTGTAGAGTGTGAGCATCTCCC	Split BS 2 AUU
3	GGGAGATGAAGGACGGGGCCAGTCCG	Split BS 1 GUC
	CGGACTGTTGCGTAGAGTGTGAGCATCTCCC	Split BS 2 GUC
4	GGGAGATGAAGGACGGGTCCAGTCCG	Split BS 1 WT
	CGGACTGTTG**A**GTAGAGTGTGAGCATCTCCC	Split BS 2 WT A11AP
5	GGGAGATGAAGGACGGG**A**CCAGTCCG	Split BS 1 AUU U18AP
	CGGACTGTTGTGTAGAGTGTGAGCATCTCCC	Split BS 2 AUU
6	GGGAGATGAAGGACGG**A**TCCAGTCCG	Split BS 1 WT G17AP
	CGGACTGTTGAGTAGAGTGTGAGCATCTCCC	Split BS 2 WT
7	GGGAGATGAAGGACGGGTCCAGTCCG	Split BS 1 WT
	CGGACTGTTGAGTAG**A**GTGTGAGCATCTCCC	Split BS 2 WT A16AP
8	GGGAGATGAAGGA**C**GGGTCCAGTCCG	Split BS 1 WT C14PC
	CGGACTGTTGAGTAGAGTGTGAGCATCTCCC	Split BS 2 WT
9	GGGAGATGAAGGACGGGGCCAGTCCG	Split BS 1 GTC
	CGGACTGTTG**C**GTAGAGTGTGAGCATCTCCC	Split BS 2 GUC C11PC
10	GGGAGATGAAGGACGGGTCCAGTCCG	Split BS 1 WT
	CGGACTGTTGAGTAG**C**GTGTGAGCATCTCCC	Split BS 2 WT A16PC

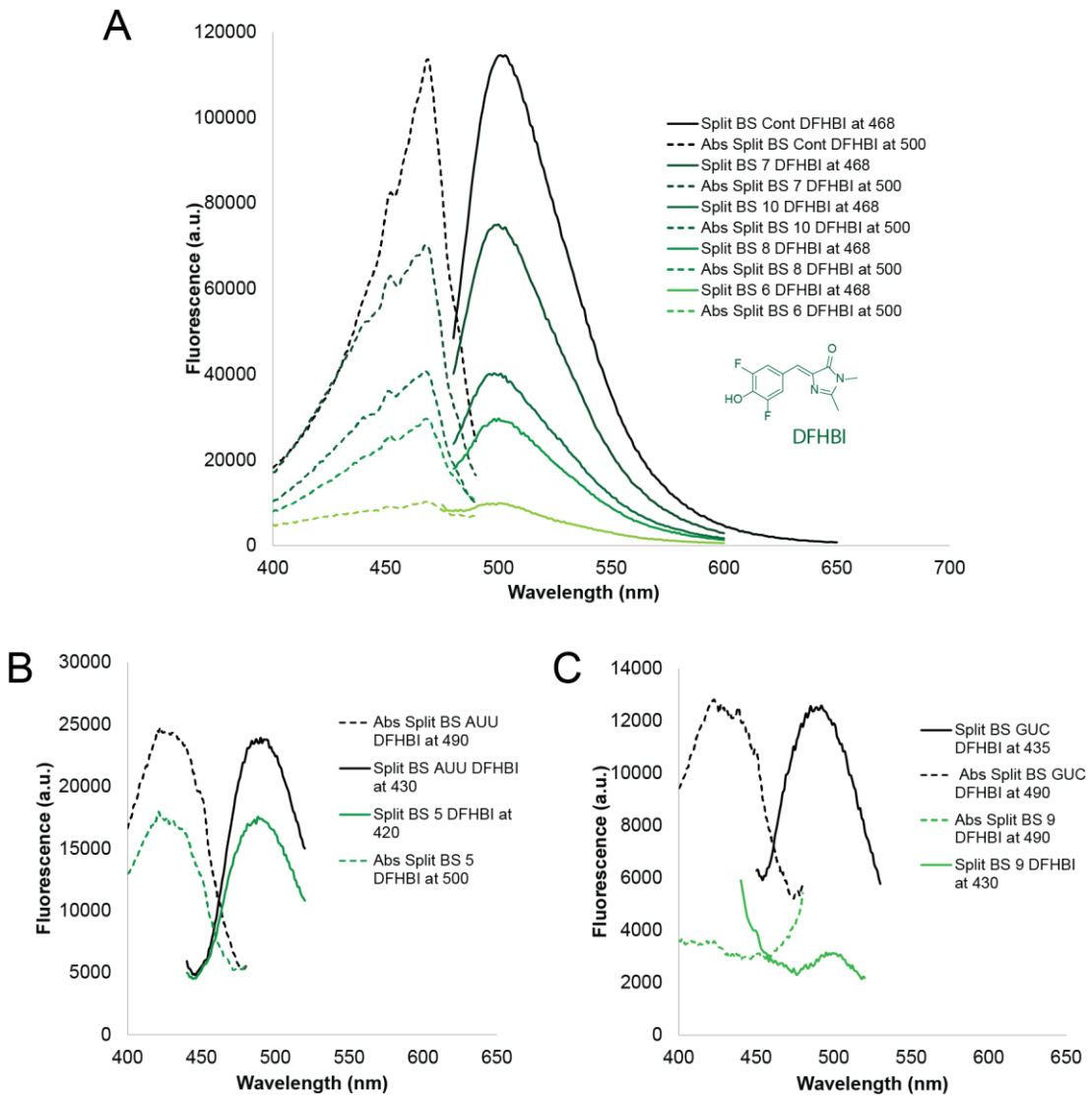


Figure 3.9. Absorbance and fluorescence spectra for modified constructs of Split Baby Spinach in complex with DFHBI. (A) Spectra for unmutated Split BS UUA base triple with AP and PyC modifications. (B) Spectra for mutated Split BS AUU base triple with AP modification. (C) Spectra for mutated Split BS GUC base triple with PyC modification.

In studying these modified RNA constructs with DFHBI, no shifts in the absorbance or fluorescence spectra were observable. All modified constructs had an absorbance maximum at 468 nm and a fluorescence maximum at 502 nm. Modified constructs bound to DFHBI-1T similarly displayed consistent spectra, with an absorbance

maximum at 470 and a fluorescence maximum at 504 nm (Figure 3.10). Additionally, modifications were tested with base triples BS AUU, where adenosine was mutated to AP, and BS GTC, where cytosine was mutated to PyC. Interestingly, BS (AP)UU bound to DFHBI did not see a large reduction in fluorescence signal whereas BS (AP)UU bound to DFHBI-1T demonstrated a large decrease (Figure 3.9B, 3.10B). Additionally, BS (AP)UU in the presence of DFHBI did not exhibit a shift in the absorbance or fluorescence spectra while the reduced fluorescence spectrum of BS (AP)UU in the presence of DFHBI-1T exhibited a slight shift in its fluorescence spectrum peak from 495 nm, which is typical of the BS AUU base triple, to 501 nm, which is closer to the fluorescence peak seen for the Baby Spinach with no modifications at all (Figure 9B, 10B). BS GT(PyC) in the presence of DFHBI displayed a larger reduction in fluorescence signal than BS GT(PyC) in the presence of DFHBI-1T (Figure 3.9C, 3.10C). However, in both cases, the fluorescence spectra of the modified constructs also displayed a modest shift in the emissions peak from 490 nm to 502 nm (in the presence of DFHBI) and from 498 nm to 504 nm (in the presence of DFHBIT-1T). Thus, the base triple mutations BS AUU and BS GTC seem to impart a hypsochromic shift in the fluorescence spectra compared to Baby Spinach, while the modified base triples BS (AP)UU and BS GT(PyC) introduce a bathochromic shift, negating the overall effect of the base triple mutations.

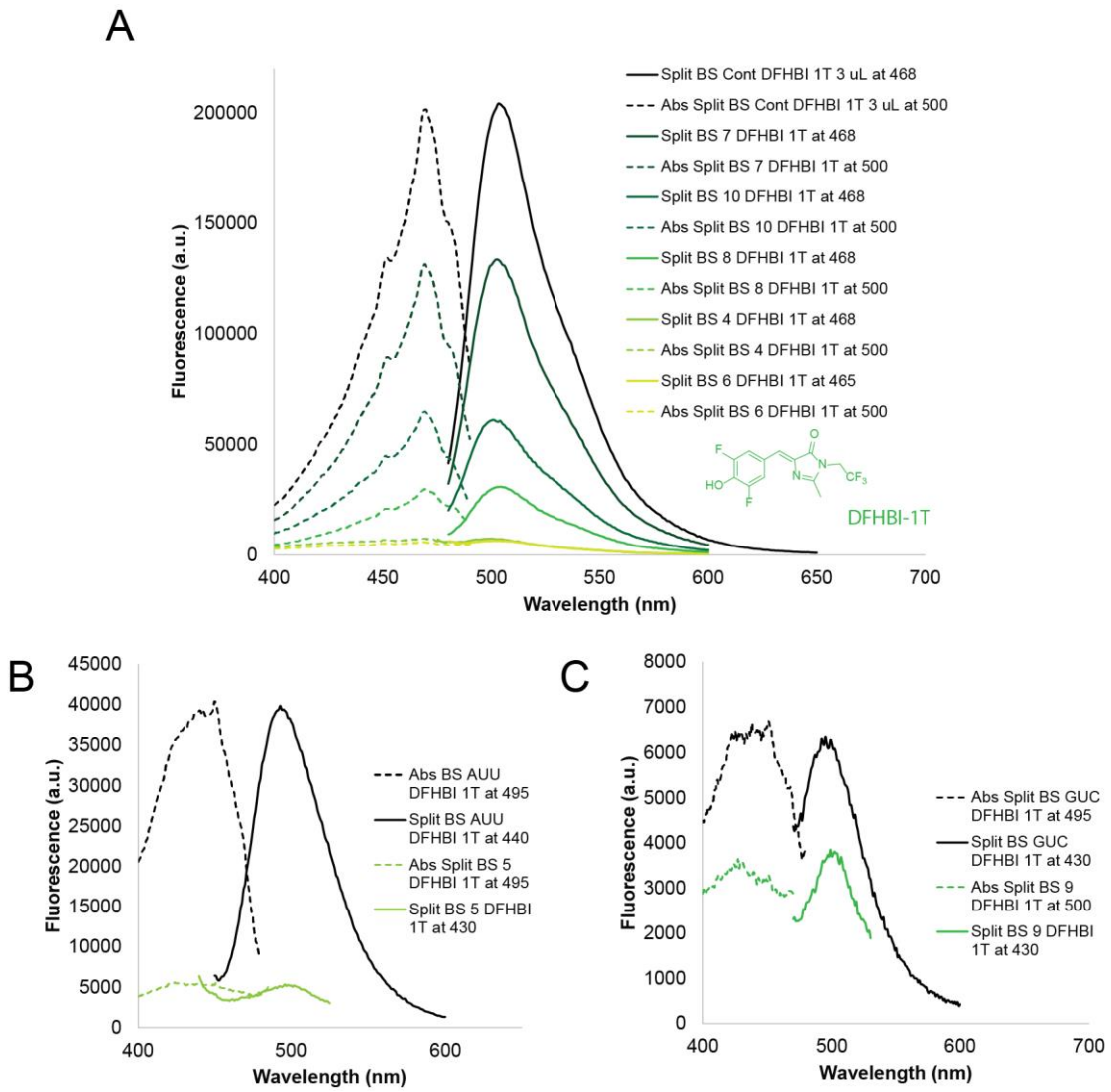


Figure 3.10. Absorbance and fluorescence spectra for modified constructs of Split Baby Spinach in complex with DFHBI-1T. (A) Spectra for unmutated Split BS UUA base triple with AP and PyC modifications. (B) Spectra for mutated Split BS AUU base triple with AP modification. (C) Spectra for mutated Split BS GUC base triple with PyC modification.

Modest shifts in fluorescence spectrum of AP or PyC modified Split Baby Spinach were also observed in the presence of DFHO (Figure 3.11A). While the unmodified construct had a fluorescence peak at 574 nm, BS2 A16(AP) (Split BS 7) and BS2 A16(PyC) (Split BS 10) exhibited peaks at 570 nm and 560 nm, respectively. Interestingly, the absorbance spectra for both modified complexes in the presence of

DFHO also displayed multiple peaks, and the fluorescence maximum shifted depending on the chosen excitation wavelength. The strength of these peaks in the absorbance spectra varied depending on the modification, and minor peaks became more apparent for modifications such as BS1 G17(AP) (Split BS 6). This indicates that shifts in the fluorescence spectra of Baby Spinach can be achieved by exciting some of the minor bands, which can also be manipulated by mutating the aptamer. Unlike BS1 C14(PyC) (Split BS 8) in the presence of DFHBI or DFHBI-1T, the BS1 C14 mutation is not well tolerated.

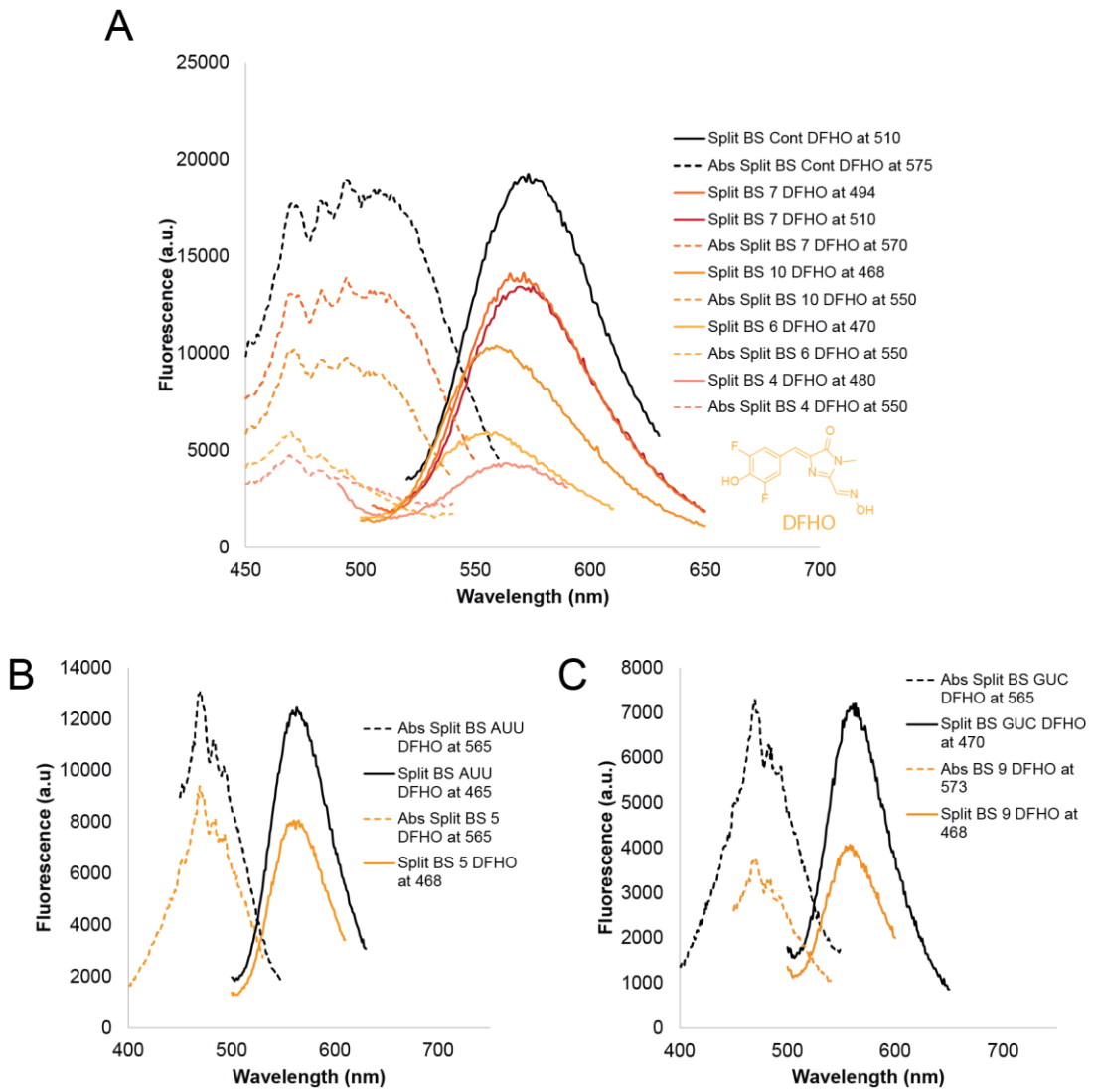


Figure 3.11. Absorbance and fluorescence spectra for modified constructs of Split Baby Spinach in complex with DFHO. (A) Spectra for unmutated Split BS UUA base triple with AP and PyC modifications. (B) Spectra for mutated Split BS AUU base triple with AP modification. (C) Spectra for mutated Split BS GUC base triple with PyC modification.

To investigate how far the system could be pushed, I examined the fluorescence spectra of complexes containing two base modifications (Table 3.3). In all cases, large reductions in fluorescence signal were observed (Figure 3.12), indicating that multiple modifications in the chosen positions are not well tolerated.

Table 3.3. Sequences of Split Baby Spinach containing two modifications (one of Split BS1 and one on Split BS strands) with reference numbers used in following figures and text.

REF. NUMBER	MODIFIED CONSTRUCTS	SEQUENCE NAME
11	GGGAGATGAAGGACGG**A**TCCAGTCCG	Split BS 1 WT G17AP
	CGGACTGTTG**A**GTAGAGTGTGAGCATCTCCC	Split BS 2 WT G11AP
12	GGGAGATGAAGGACGG**A**TCCAGTCCG	Split BS 1 WT G17AP
	CGGACTGTTGAGTAG**A**GTGTGAGCATCTCCC	Split BS 2 WT A16AP
13	GGGAGATGAAGGACGG**A**TCCAGTCCG	Split BS 1 WT G17AP
	CGGACTGTTGAGTAG**C**GTGTGAGCATCTCCC	Split BS 2 WT A16PC
14	GGGAGATGAAGGA**C**GGGTCCAGTCCG	Split BS 1 WT C14PC
	CGGACTGTTG**A**GTAGAGTGTGAGCATCTCCC	Split BS 2 WT G11AP
15	GGGAGATGAAGGA**C**GGGTCCAGTCCG	Split BS 1 WT C14PC
	CGGACTGTTGAGTAG**A**GTGTGAGCATCTCCC	Split BS 2 WT A16AP
16	GGGAGATGAAGGA**C**GGGTCCAGTCCG	Split BS 1 WT C14PC
	CGGACTGTTGAGTAG**C**GTGTGAGCATCTCCC	Split BS 2 WT A16PC

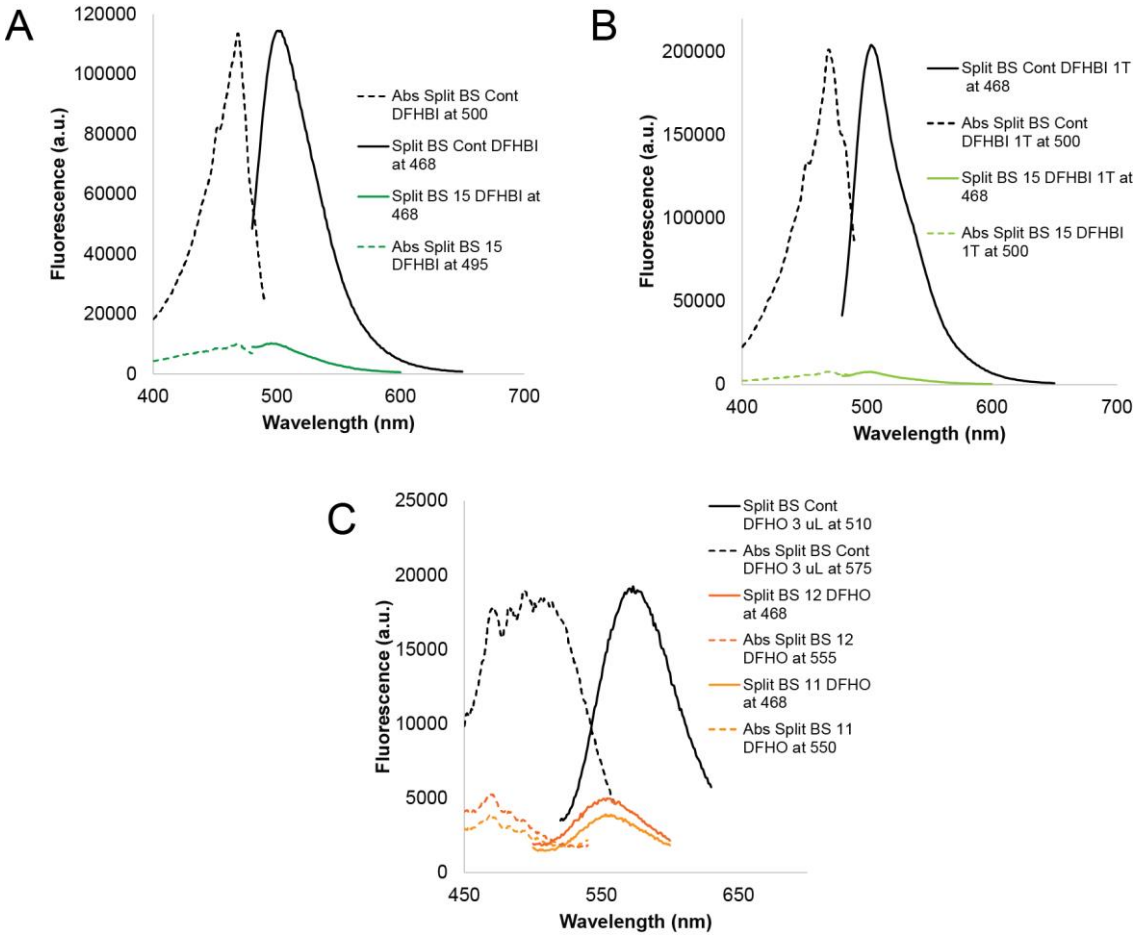


Figure 3.12. Absorbance and fluorescence spectra for constructs of Split Baby Spinach containing two modifications. (A) Spectra of modified construct in complex with DFHBI. (B) Spectra of modified construct in complex with DFHBI-1T. (C) Spectra of modified construct in complex with DFHO.

Aside from reducing overall signal intensity, AP and PyC modifications had little effect on the absorbance and fluorescence spectra of Split Baby Spinach when bound to DFHBI or DFHBI-1T. However, when the aptamer is bound to DFHO, shifts in wavelength are observed. Still, these shifts in fluorescence spectra seem to be more closely related to excitation wavelengths than specifically to modifications on the aptamer. The absorbance spectra of DFHO-bound Split Baby Spinach demonstrate a number of peaks in the absorbance spectra, and each peak seems to result in a different

fluorescence maximum. Thus, this spectral tuning through minor band excitation is an inherent feature of the DFHO fluorophore although it is possible that modifications to the aptamer may help enhance some minor UV bands.

3.2.3.4 Fluorescence effects of modifying Split Baby Spinach with $PreQ_0$ and Wyosine

Following preliminary studies of AP and PyC in Split Baby Spinach, the aptamer was modified using the synthesized post-transcriptional RNA modifications $PreQ_0$ and wyosine (Chapter 2, Section 2.2.2.2) in positions BS1 G17 and BS2 A16 (Figure 3.13A). These two positions were least distruptive when probing the aptamer using AP and PyC modifications. In all cases, shifts in absorbance or fluorescence maxima were minimal or not apparent (Figure 3.13B,C,D). BS2 A16($PreQ_0$) and BS2 A16(Wyosine) exhibited modest (4-5 nm) hypsochromic shifts in their fluorescence maxima when bound to DFHBI, while BS2 A16(PreQ0) exhibited a 8 nm bathochromic shift in its fluorescence maximum. All other constructs displayed fluorescence maxima that were consistent with the unmodified control RNA aptamer when bound to the respective ligands.

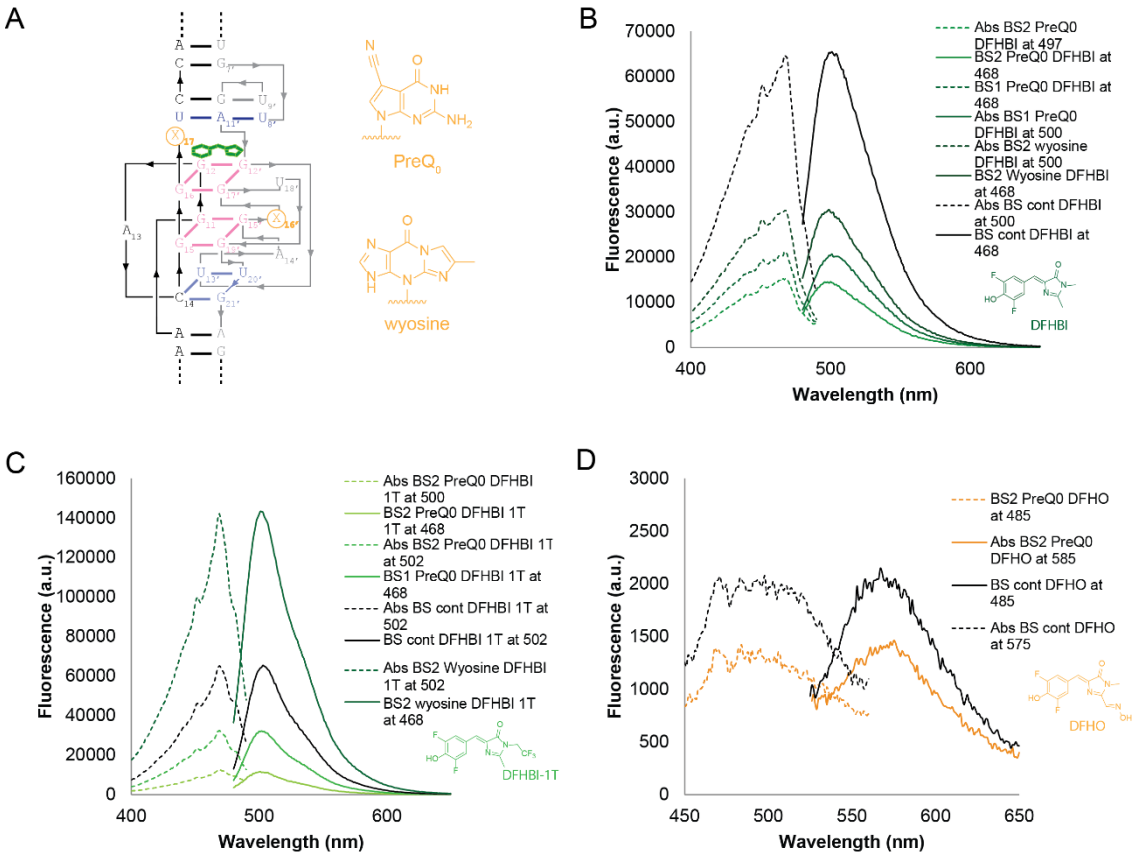


Figure 3.13. Modified positions of Split Baby Spinach with $PreQ_0$ and wyosine and corresponding fluorescence spectra. (A) Positions modified in Split Baby Spinach with $PreQ_0$ or wyosine (yellow). (B) Absorbance and fluorescence spectra for modified Split Baby Spinach with $PreQ_0$ or wyosine when bound to DFHBI. (C) Absorbance and fluorescence spectra for modified Split Baby Spinach with $PreQ_0$ or wyosine when bound to DFHBI 1T. (D) Absorbance and fluorescence spectra for modified Split Baby Spinach with $PreQ_0$ or wyosine when bound to DHFO.

In most instances, modification with $PreQ_0$ or wyosine reduced fluorescence intensity compared to the unmodified aptamer. BS1 G17(Wyosine) exhibited no fluorescence enhancement in the presence of DFHBI, DFHBI 1T, or DFHO. This indicates that the BS1 G17(Wyosine) mutation was detrimental to either aptamer folding or ligand

binding. In structural characterizations of Spinach and variants, the analogous BS1 G17 nucleotide shares the ligand binding pocket of the aptamer with the bound fluorophore, stacking in a co-planar conformation (Figure 3.14). The bulky tricyclic nucleobase of wyosine could prevent ligand binding through sterics, resulting in the diminished fluorescence turn-on.

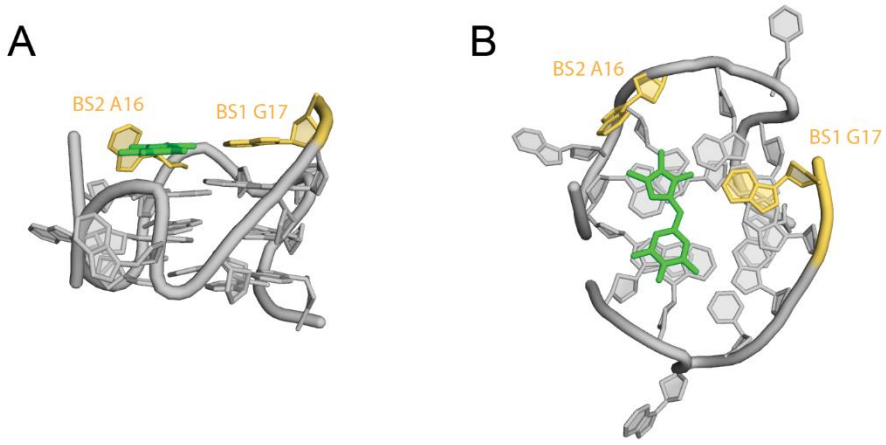


Figure 3.15. Structure of Spinach binding pocket with analogous mutation positions. (A) Side view of the ligand binding pocket of Spinach. Mutated positions are highlighted in yellow, while the ligand is shown in green. (B) Top view of the ligand binding pocket of Spinach. Nucleotide BS1 G17 packs co-planarly with the fluorophore ligand. (PDB: 4TS0)

Notably, the modification of BS2 A16(Wyosine) when bound to DFHBI 1T exhibited greater than two-fold fluorescence enhancement relative to the unmodified aptamer. Where the tricyclic structure of wyosine was detrimental in the position BS1 G17 of the ligand binding pocket, its bulk could help restrain the fluorophore in position BS2 A16 (Figure 3.15). This would prevent nonradiative decay, leading to stronger fluorescence enhancement. However, due to the isomeric mixture of the original wyosine phosphoramidite and thus the resulting 2′ 3′ isomers of the modified RNA, it is

important ensure the enhanced fluorescence signal is a result of the wyosine modification and not the 2′ isomer. To achieve this, the wyosine phosphoramidite would need to be synthesized again, taking care to avoid isomerization. Ultimately, the strong fluorescence enhancement observed with the BS2 A16(Wyosine) modification demonstrates the potential of noncanonical modifications for fluorogenic RNA aptamers.

Table 3.4. Summary of fluorescence changes for modified Baby Spinach

	RNA	Fluorophore	Relative change in fluorescence intensity	Excitation maximum	Emission maximum
Base Triples	GTC	DFHBI	-	430	487
	AAT	DFHBI	-36%	430	491
	ACT	DFHBI	-40%	430	492
	GAC	DFHBI	-58%	430	490
	TAA	DFHBI	-71%	430	502
	CGG	DFHBI	-78%	430	503
Modifications	Split BS	DFHBI	-	468	502
	RNA 7	DFHBI	-34%	468	499
	RNA 10	DFHBI	-65%	468	499
	RNA 8	DFHBI	-74%	468	502
	RNA 6	DFHBI	-92%	468	499
	Split BS AUU	DFHBI	-	430	490
	RNA 5	DFHBI	-27%	430	490
	Split BS GUC	DFHBI	-	435	489
	RNA 9	DFHBI	-75%	430	500
	Split BS	DFHBI-1T	-	468	503
	RNA 7	DFHBI-1T	-35%	468	502
	RNA 10	DFHBI-1T	-70%	468	500
	RNA 8	DFHBI-1T	-85%	468	503
	RNA 4	DFHBI-1T	-96%	468	503
	RNA 6	DFHBI-1T	-96%	468	500
	Split BS AUU	DFHBI-1T	-	440	493
	RNA 5	DFHBI-1T	-87%	430	500
	Split BS GUC	DFHBI-1T	-	440	495
	RNA 9	DFHBI-1T	-39%	430	499
	Split BS	DFHO	-	510	573
	RNA 7	DFHO	-26%	494	573

	RNA 10	DFHO	-30%	510	570
	RNA 6	DFHO	-36%	468	559
	RNA 4	DFHO	-46%	468	560
	Split BS AUU	DFHO	-	468	564
	RNA 5	DFHO	-35%	465	561
	Split BS GUC	DFHO	-	470	565
	RNA 9	DFHO	-44%	468	560

While changes in the fluorescence properties of modified Split Baby Spinach were observed with DFHBI, DFHBI-1T, and DFHO (Table 3.4), studies for binding affinities for these modified aptamers were not explored. Differences in binding affinities of modified aptamers for a fluorophore ligand may also affect the observed fluorescence properties.

3.3 CONCLUSIONS AND FUTURE WORK

Large changes to the absorbance and fluorescence spectra of Spinach family aptamers were best accomplished by changing the base triple that stacks on top of the ligand binding pocket. While the A-U-U base triple mutation had been previously reported to induce large shifts in absorbance and fluorescence maxima, all combinations of base triple mutations which did not completely abolish fluorescence surprisingly exhibited the same shifts in their absorbance and fluorescence spectra. Furthermore, these shifts are only observed with DFHBI and analogous ligands such as DFHO. Fluorophores such as ThT and TO1, which have been shown to bind the Spinach aptamer, displayed consistent fluorescence spectra despite the base triple mutations. Crystallographic studies of *i*Spinach with DFHBI-1T, an analogue of DFHBI with greater fluorescence enhancement, indicated a conserved structural core consisting of a base triple, two-tiered G-quadruplex, and mixed base quadruple despite the modified ligand. The

fluorescence enhancement of DFHBI-1T may stem from tighter packing of the molecule's trifluoromethyl group against the adjacent A50 nucleotide.

Baby Spinach was successfully divided into a split aptamer that was brighter than the unmodified whole Baby Spinach aptamer. This split aptamer, consisting of a 26 nt and 31 nt strand, was accessible using RNA solid-phase synthesis, and 2-aminopurine and pyrrolocytosine were placed in varying positions surrounding the aptamer's ligand binding pocket. When bound to DFHBI or DFHBI-1T, changes to the aptamer-fluorophore complex's absorbance or fluorescence spectra were not observed relative to the unmodified control sequences. Modifications to Split Baby Spinach bound to DFHO did result in shifts in the fluorescence spectra maxima. This is consistent with previous studies of Red and Orange Broccoli.[94] However, the observed changes in the maxima of the fluorescence spectra appeared to result from excitation at different wavelengths rather than the modifications to the aptamer. This indicates that the fluorescence of the aptamer when bound to DFHO can be tuned by exciting a minor band in the absorbance spectrum.

Finally, the use of post-transcriptional RNA modifications $PreQ_0$ and wyosine in select positions of Split Baby Spinach bound to DFHBI, DFHBI 1T, or DFHO produced absorbance and fluorescence maxima that were generally consistent with the unmodified aptamer. While most positions studied with these modifications displayed reductions in overall fluorescence intensity, Split Baby Spinach BS1 G17(Wyosine) exhibited a greater than two-fold fluorescence increase when bound to DFHBI 1T relative to the unmodified aptamer. Further research is necessary to confirm the fluorescence enhancement is a result of the wyosine modification and not the 3′ isomer present in solution. However, this signal increase could demonstrate the potential of

noncanonical nucleosides to improve the spectral properties of fluorogenic RNA aptamers. Additional IR spectroscopy experiments can be performed using the nitrile group in $PreQ_0$ to study how the ligand and the nucleotides in these positions might interact.

3.4.1 *In vitro* RNA transcription

RNA constructs for triple base mutation studies and *i*Spinach structural characterization were prepared using *in vitro* transcription from PCR templates. Plasmids encoding the templates were purchased from Eurofin Genomics and PCR amplified in a 1.2 mL PCR mixture containing 0.2 µg/µL plasmid, 0.2 u/µL of DreamTaq polymerase, 2 µM each of forward and reverse primers, 0.2 mM dNTPs, and the DreamTaq buffer at the recommended working concentration. The PCR mixture was thermocycled with an initial denaturing step at 95 °C for 5 min; followed by 20 cycles of 95 °C for 30 s, 75 °C for 45 s (with a reduction of 1.5 °C per cycle), and 72 °C for 60 s; followed by an additional 20 cycles of 95 °C for 30 s, 48 °C for 45, and 72 °C. These PCR products were taken forward without further purification and *in vitro* transcribed in 5 mL using 300 mM Tris-HCl pH 8.1, 20 mM spermidine, 250 mM $MgCl_2$, 0.1% Triton X-100, 5 mM NTPs, 5 mM DTT, and 1 u/mL of inorganic pyrophosphatase. After incubation at 37 °C for 16 h, the resulting RNA was purified on a 12% polyacrylamide, 1x TBE, 8 M urea gels, electroeluted over 16 h, and concentrated using Amicon centrifugal concentrators (Millipore). For the *i*Spinach construct, the gel-purified RNA was electroluted over 16 h and ethanol precipitated in the presence of 300 mM sodium acetate (pH 5.5) and three volumes of cold ethanol for 2 h at -80 °C. The solution was centrifuged at 12,000 g for 45 min, the supernatant poured off, and pellet washed three times with 70% ethanol. The RNA pellet was resuspended in DEPC treated water until further use.

3.4.2 Fluorescent measurements

To fold Baby Spinach, 1 μM RNA in 100 mM Tris-HCl pH 7.4 and 100 mM KCl was heated to 95 °C for 5 min and snap cooled on ice for 2 min. $MgCl_2$ was then added to a final concentration of 10 mM and the appropriate fluorophore ligand was added in a 1:5 ratio of RNA:ligand. The solution was heated to 65 °C for 5 min and slowly cooled to room temperature over 1 h. Fluorescence measurements were recorded using a Varian Cary Eclipse spectrophotometer at 468 nm excitation/500 nm emission for DFHBI and DFHBI-1T, 468 nm excitation/550 nm emission nm for DFHO, 510 nm excitation/535 nm emission nm for TO1, 450 nm excitation/480 nm emission nm for ThT unless otherwise stated. Monochromator slits were set to 4 nm and integration time was 1 s. Measurements were taken on samples with a volume of 120 μL.

3.4.3 *i*Spinach/DFHBI-1T crystallization and data collection

A solution of purified, in vitro transcribed RNA was heated to 85 °C and cooled to 25 °C using a thermocycler. A solution of 25 mM DFHBI-1T in DMSO was added in a 1:1.2 ratio and Tris-HCl, KCl, and MgCl2 were added to final concentrations of 210 mM RNA, 252 mM DFHBI-1T, 40 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM MgCl2. This solution was stored at -20 °C until use.

Using sitting-drop vapor-diffusion for crystallization, the prepared *i*Spinach/DFHBI-1T solution was combined 1:1 with a reservoir solution of 40 mM NaCl2, 32% MPD, 40 mM Na cacodylate pH 7.0, and 12 mM spermine. Crystals grew to maximum dimensions of 200 x 500 μm within 1 week with leaf-like plate morphology. Crystals were cryoprotected using a solution of 36% MPD, 10% glycerol, 40 mM NaCl2, 40 mM Na cacodylate pH 7.0, and 12 mM spermine and frozen in liquid nitrogen. X-ray diffraction data were collected at beamline 5.0.1 of the Advanced Light Source (ALS),

phased using molecular replacement with a previously solved *i*Spinach crystal structure and PHASER (PHENIX), and manually refined using COOT.

CHAPTER 4. REMOTE EXCITATION OF RNA MANGO APTAMERS

4.1 INTRODUCTION

The Mango family of fluorogenic RNA aptamers are high affinity aptamers that bind and induce the fluorescence of a thiazole orange derivative.[62,64] The fluorescence spectra for the Mango aptamers are red shifted, relative to Spinach and other DFHBI-based aptamers, and like many fluorogenic RNA aptamers, Mango aptamers bind its cognate ligand using a G-quadruplex structural core.[63,65] To date, the Mango aptamers are some of the best fluorogenic RNA aptamers due to their strong binding affinities, brightness, dramatic fluorescence enhancement, and small sizes.[127]

Forster resonance energy transfer (FRET) is a common technique used to detect intramolecular conformational changes and intermolecular protein-protein interactions.[128–130] Recently, a FRET system using fluorogenic aptamers Spinach and Mango was also developed, demonstrating the applicability of genetically encodable FRET systems to fluorogenic RNA aptamers.[83] In these systems, a donor fluorophore in an excited state can transfer its excitation energy to a nearby acceptor, reporting on changes in distance between the donor fluorophore and acceptor. Taking advantage of the fluorescence of the fluorophore-Mango aptamer complex, I have developed an internal FRET antenna using covalently attached fluorescent dyes. Although there exist examples of fluorescently labelled RNA for conformational and single molecule studies,[64,131–133] this is the first instance of fluorescently labelling a fluorogenic RNA aptamer for remote excitation of the labelled dye or the bound fluorophore. The versatility of the internal FRET antenna provides a straightforward method for adjusting the working excitation or emission wavelength of these fluorogenic RNA aptamers.

To demonstrate the adaptability of fluorescent labels for remote excitation of a bound fluorophore ligand in a fluorogenic RNA aptamer complex, I have explored the use of fluorescent nucleotide analogues and fluorescent dyes. Furthermore, I have studied these modifications in a model G-quadruplex and in the fluorogenic RNA aptamers, Mango-III and *i*Mango-III. In all cases, a fluorescence of the FRET acceptor was observed upon excitation of the donor.

4.2.1 RNA TERRA: a model G-quadruplex

4.2.1.1 Modifying G-quadruplex loop regions with aminopurine and pyrrolocytosine

2-aminopurine (AP) and pyrrolocytosine (PyC) are fluorescent nucleotide analogues of guanosine, which are commonly used to study conformational dynamics in nucleic acids.[126] Thiazole orange is a fluorescent dye that binds nucleic acids, and TO1 is a thiazole orange derivative developed for the Mango class of fluorogenic RNA aptamers.[62,134] Due to the overlapping excitation and emission spectra of AP and PyC with TO1 (Figure 4.1), these fluorophores have the capacity to function as a Förster resonance energy transfer (FRET) pair. The FRET potential of these fluorescent nucleotide analogues provides a route to easily tune the spectral properties of fluorogenic RNA aptamers by allowing excitation of the donor to observe fluorescence from the acceptor.

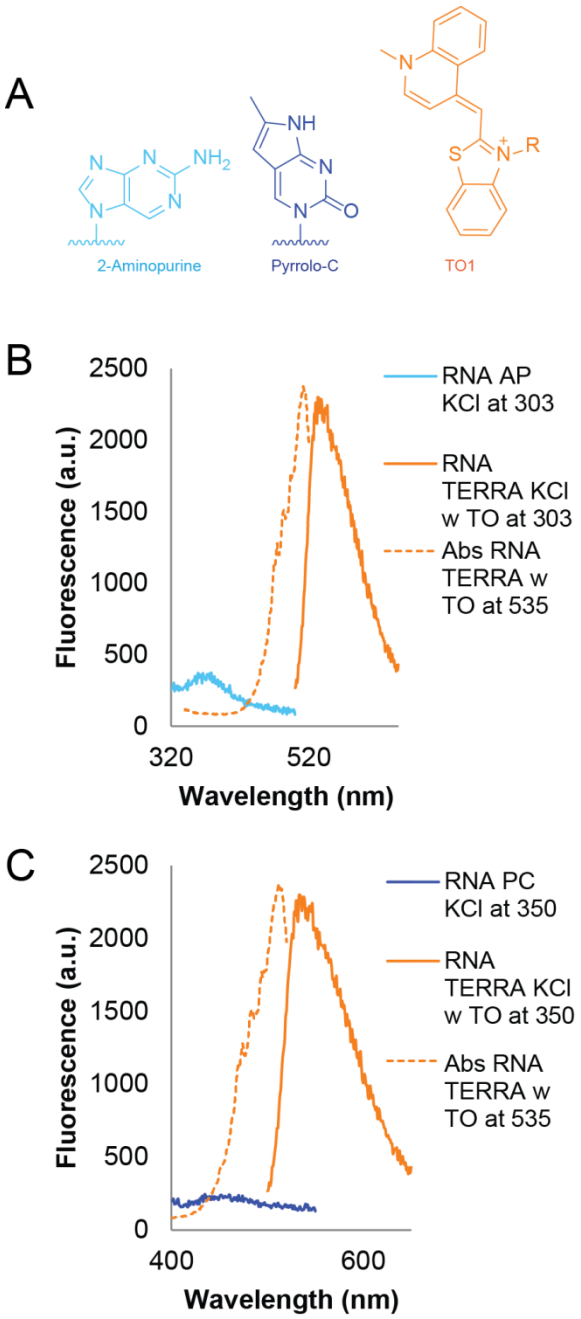


Figure 4.1. Spectral overlap of fluorescent nucleoside analogues and a thiazole orange derivative. (A) Chemical structures of pyrrolocytosine (PyC) and 2-aminopurine (AP), two fluorescent analogues of guanine; and a thiazole orange derivative (TO1), a fluorogenic G-quadruplex binding ligand. (B) Spectral overlap of excitation and emission spectra for 2-aminopurine and TO1. (C) Spectral overlap of excitation and emission spectra for pyrrolocytosine and TO1.

Like many of the fluorogenic RNA aptamers whose structures have been solved, the Mango aptamers use a G-quadruplex motif to bind and induce the fluorescence of its cognate ligand.[63,65,66] To study the effects of AP and PyC in a model G-quadruplex system, I introduced the fluorescent analogues into telomeric repeat-containing RNA (TERRA), a well-studied tetrameric RNA G-quadruplex (Figure 4.2).[135,136] Using the structure of RNA TERRA as a guide, I substituted position A6 in RNA TERRA with either AP or PyC. Since RNA TERRA folds into a tetrameric G-quadruplex, each tetramolecular G-quadruplex contains 4 modified nucleotides. A6 is positioned in the loop regions of the G-quadruplex and forms a base stacking interaction with its corresponding nucleotide on the adjacent oligonucleotide strand, providing a stable site for modification without directly interfering with the G-quadruplex fold.

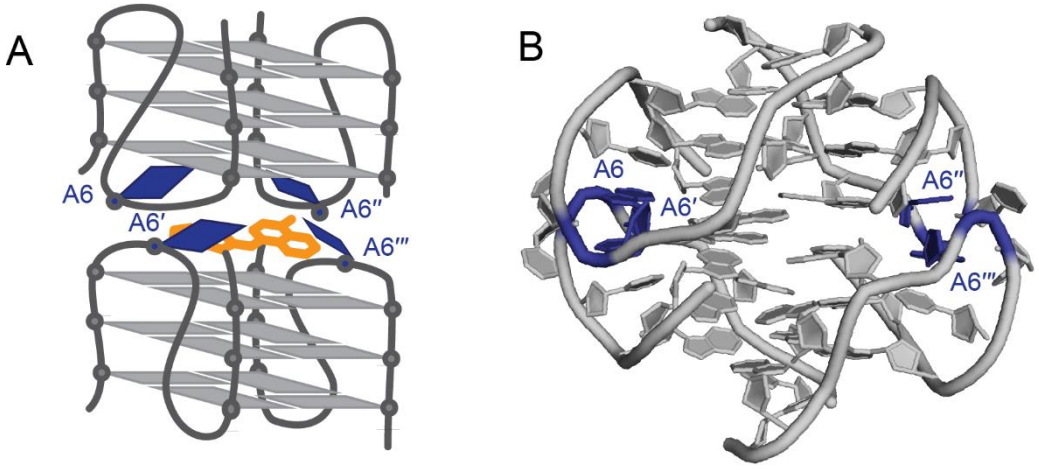


Figure 4.2. Sites of modification in RNA TERRA. (A) Cartoon depiction of RNA TERRA G-quadruplex with modified nucleosides in loop regions. Guanines involved in g-quadruplex are depicted in gray, and modified nucleoside sites are depicted in blue. Bound TO1 ligand is depicted in yellow. (B) Structure of RNA TERRA with modified sites highlighted in blue. (PDB: 2M18).

Although G-quadruplexes are particularly stable nucleic acid structural motifs, regions for modifications must be carefully selected to avoid destabilizing the overall RNA fold while still remaining in close enough proximity to the binding site to permit FRET transfer. The loop regions of the G-quadruplex are thus ideal positions for modifications.

4.2.1.2 Effects of fluorescent nucleotide modifications on spectral properties of fluorescent G-quadruplex binders

In the absence of TO1, AP modified RNA TERRA can be excited at 303 nm to produce an emission maximum at 320 nm (Figure 4.3A). Similarly, in the absence of TO1 PyC modified RNA TERRA can be excited at 350 nm to give an emission maximum at 450 nm (Figure 4.3B). Once TO1 is added to the system and binds to the G-quadruplex, AP

and PyC modified RNA TERRA excited at 303 nm or 350 nm, respectively, will both produce an emission maximum at 535 nm (Figure 4.3A,B). This shift in the emission spectra indicates an efficient energy transfer between a fluorescent modification in the loop regions of a model G-quadruplex to a bound fluorescent ligand.

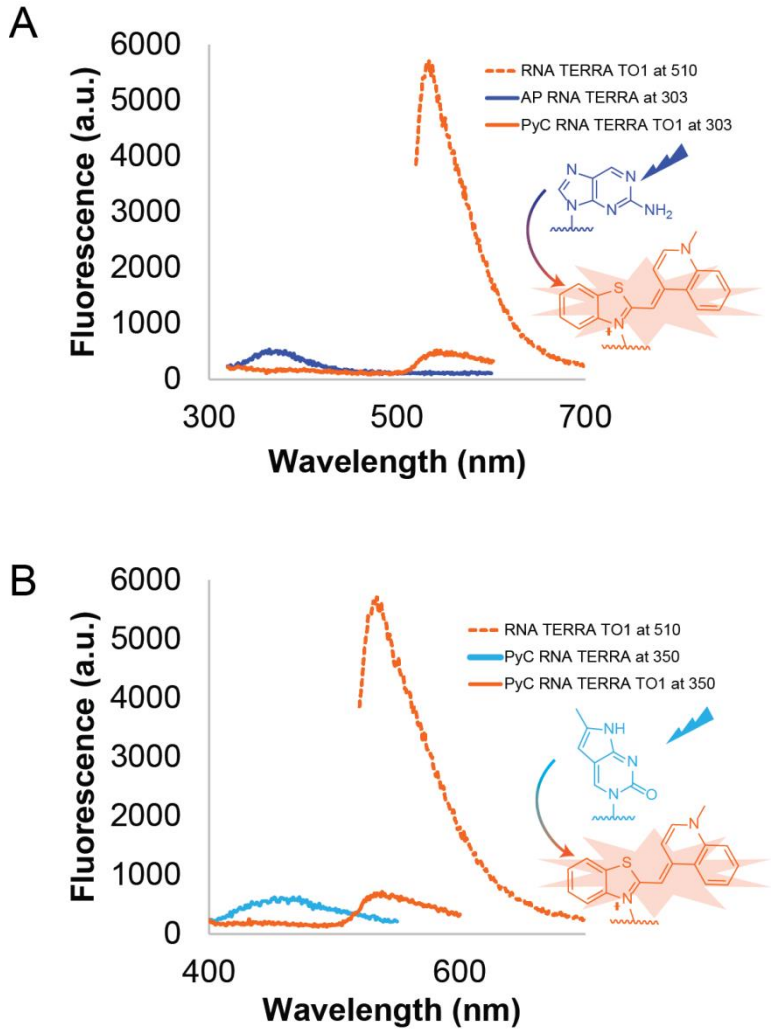


Figure 4.3. Fluorescence spectra of TO1-bound RNA TERRA modified with AP or PyC. (A) Fluorescence spectra of AP modified RNA TERRA with and without TO1. (D) Fluorescence spectra of PyC modified RNA TERRA with and without TO1.

Previous studies of planar aromatic heterocycles have been shown to bind to G-quadruplexes between G-tetrad faces, so it is expected that TO1 would preferentially

bind RNA TERRA at the tetramolecular G-quadruplex face. However, as it was not explicitly determined that the G-quadruplex is the site of binding, it was important to apply this proof-of-principle FRET system from a G-quadruplex model system to a fluorogenic aptamer with a characterized binding pocket.

4.2.2 Fluorescence antenna effects with Mango-III aptamer

4.2.2.1 Modifying ligand binding region of Mango-III with aminopurine and pyrrolocytosine

After confirming the viability of the FRET antenna using RNA TERRA as a model G-quadruplex, I next introduced AP and PyC into Mango-III. Mango-III, like other RNA Mango aptamers, was selected to bind TO1 with high affinity. Of the family of Mango aptamers, TO1-bound Mango-III has the brightest fluorescence.[64] Using its structure as a guide, I modified positions A15 or U17 with AP or PyC (Figure 4.4). Nucleotide A15 is part of a base pair that stacks on top of the bound ligand, forming a lid for the G-quadruplex binding pocket. Previous studies demonstrate an A15U mutation affects the aptamer's spectral properties with an increase in fluorescence.[66] Position U17's proximity to the ligand binding pocket, but indirect involvement in its formation, also makes it an ideal position for modification.

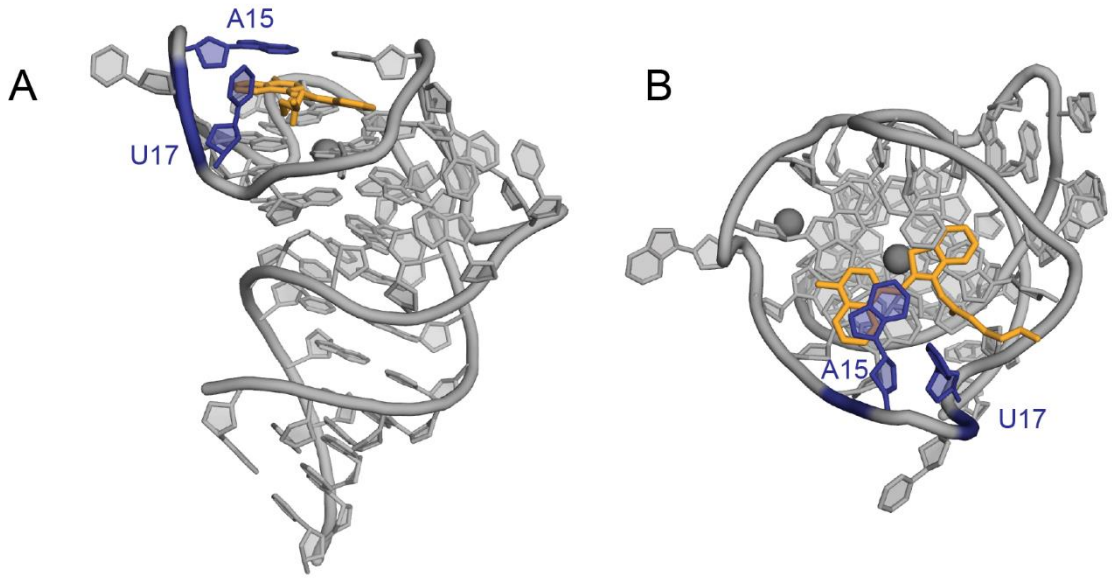


Figure 4.4. Sites of modification in TO1-bound Mango-III aptamer. (A) Structure of TO-1 bound Mango-III with modified sites highlighted in blue. (B) Top view of structure of TO1-bound Mango-III with modified sites highlighted in blue. (PDB: 6E8S)

TO1-bound Mango-III has an excitation maximum at 535 nm when excited at 510 nm. Upon substituting *i*Mango-III A15 or U17 with AP, in the absence of ligand, exciting AP at 303 nm results in a fluorescent signal at 370 nm typical of AP. In the presence of ligand, excitation of AP at 303 nm resulted in a shifted emission maximum of 535 nm, characteristic of TO1-bound Mango-III (Figure 4.5A,C). Similarly, excitation of PyC with Mango-III A15(PyC) or U17(PyC) in the absence of ligand produces a fluorescence emission maximum at 450 nm when excited at 350 nm. This fluorescence signal is characteristic of PyC. Upon binding TO1, PyC modified Mango-III has a shifted emission spectrum with a maximum at 535 nm (Figure 4.5B,D). This indicates the energy transfer of PyC to TO1 upon excitation of PyC.

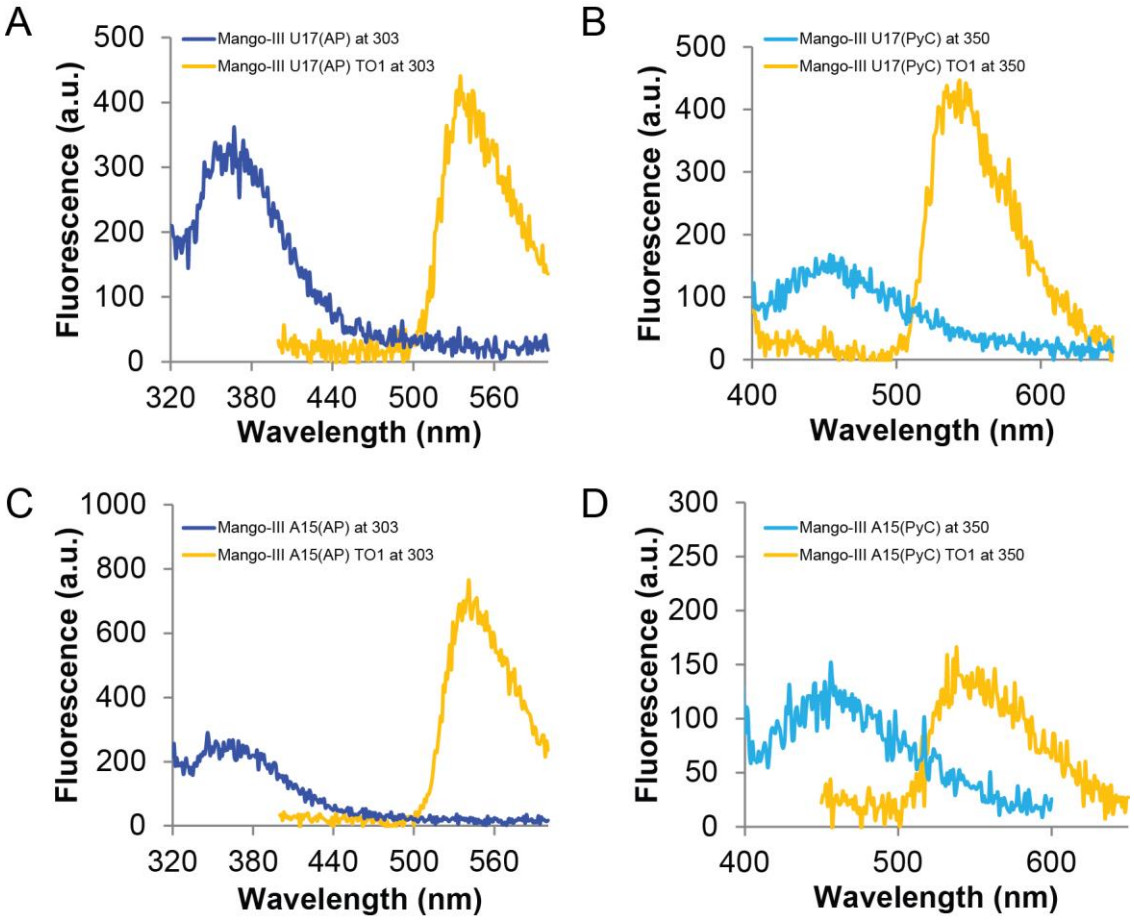


Figure 4.5. Fluorescence spectra of TO1-bound Mango-III modified with AP or PyC. (A) Fluorescence spectra of Mango-III U17(AP) with and without TO1. (B) Fluorescence spectra of Mango-III U17(PyC) with and without TO1. (E) Fluorescence spectra of Mango-III U15(AP) with and without TO1. (F) Fluorescence spectra of Mango-III U15(PyC) with and without TO1.

Introducing simple modifications of fluorescent nucleoside analogues to sites adjacent to the ligand binding pocket of fluorogenic Mango-III provided a reliable method for remote excitation of the bound fluorophore ligand.

4.2.2.2 Internal FRET system using fluorescein and rhodamine labels

While AP and PyC are conventional fluorescent nucleotide reporters used in studying nucleic acid structures, they are not as bright as conventional fluorogenic aptamers. The

nucleotide analogues are most useful for their structural similarity to canonical nucleotides, therefore reducing the potential of perturbing folding. To better utilize the strong fluorescence output of fluorogenic aptamers, I investigated fluorescent dyes as modifications to the RNA aptamer *i*Mango-III, a functionally reselected variant of Mango-III.[66] *i*Mango-III can bind both TO1 or YO3 to induce fluorescence (Figure 4.6 A,C). Upon binding TO1, the fluorophore-aptamer complex can be excited at 510 nm for a fluorescence emission maximum at 535 nm. Upon binding YO3, the fluorophore-aptamer complex can be excited at 580 nm for a fluorescence emission maximum at 620 nm. The spectral overlap of TO1 with fluorescein and YO3 with rhodamine (Figure 4.6B,D) indicates the potential for these ligand-dye sets to act as FRET pairs.

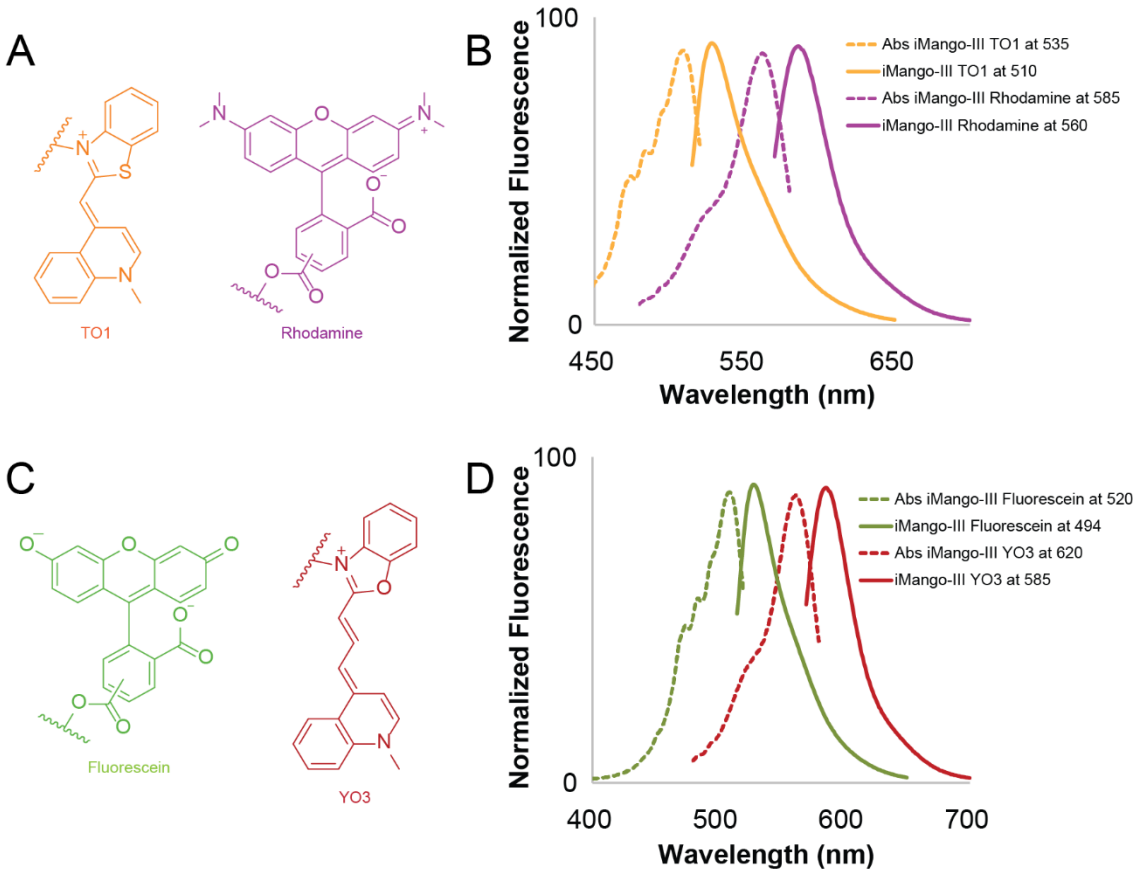


Figure 4.6. Spectral overlap of fluorescent modifications (rhodamine and fluorescein) with fluorogenic ligands (TO1 and YO3). (A) Chemical structures of TO1 and rhodamine. (B) Spectral overlap of excitation and emission spectra for TO1 (FRET donor) and rhodamine (FRET acceptor). (C) Chemical structures of fluorescein and YO3. (D) Spectral overlap of excitation and emission spectra for fluorescein (FRET donor) and YO3 (FRET acceptor).

By incorporating an amino-modified nucleotide with a 6 carbon linker, many readily available *N*-hydroxysuccinimide (NHS) dyes could be covalently attached to the RNA aptamer (Figure 4.7A). Using this method, I studied *i*Mango-III covalently bonded to fluorescein or rhodamine. I modified six positions surrounding the binding pocket of *i*Mango-III (Figure 4.7B,C) to determine optimal locations to minimize disruptions to folding while maintaining proximity to the bound ligand.

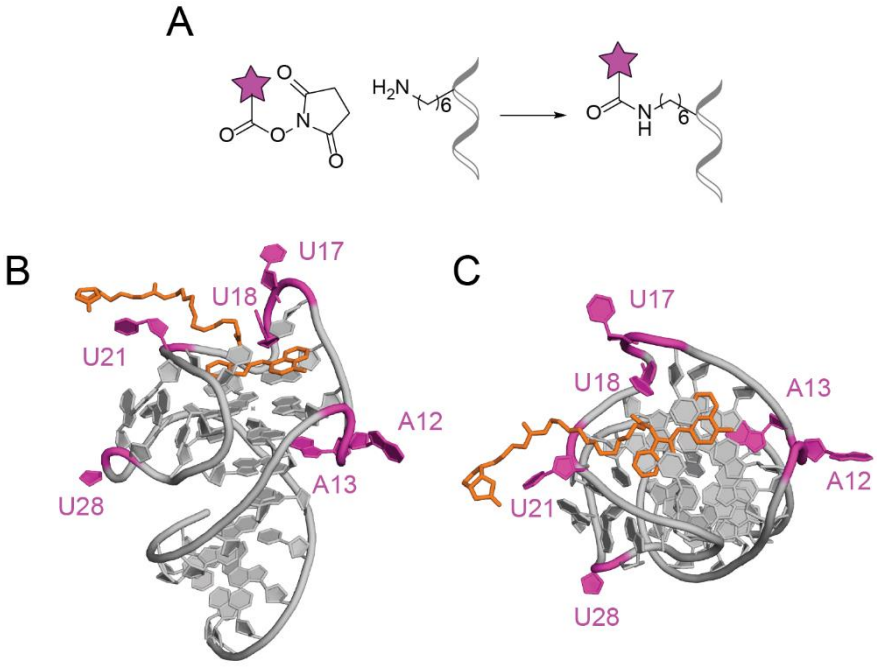


Figure 4.7. Sites of modification to *i*Mango-III using *N*-hydroxysuccinimide (NHS) chemistry. (A) NHS labeling technique used to introduce fluorescent modifications (rhodamine or fluorescein) onto RNA aptamer. (B) Structure of *i*Mango-III with modified sites highlighted in pink and TO1 ligand in orange. (C) Top view of structure of *i*Mango-III with modified sites highlighted in pink and TO1 ligand in orange. (PDB: 6E8U)

When complexed with TO1, rhodamine labelled *i*Mango-III could be excited at 510 nm to induce a fluorescence signal with an emission maximum at 580 nm (Figure 4.8). This is a 50 nm shift when compared to unmodified *i*Mango-III, which has an emission maximum at 535 nm. Rhodamine labelled *i*Mango-III also has high apparent FRET efficiencies (Table 4.1), though the overall fluorescence intensity varies with position labelled. For all positions, the calculated FRET efficiencies for the TO1/rhodamine pair range from 91-98%. Fluorescence intensities for the rhodamine labelled *i*Mango-III from highest to lowest are: 28, 12, 21, 17, 13, 18. Modifications at positions 13 and 18 resulted in 50% less fluorescence signal compared to the brightest modified aptamer,

R28. This suggests introducing a rhodamine label at these positions interferes with the folding of the aptamer thereby reducing its binding affinity for TO1 and decreasing fluorescence output.

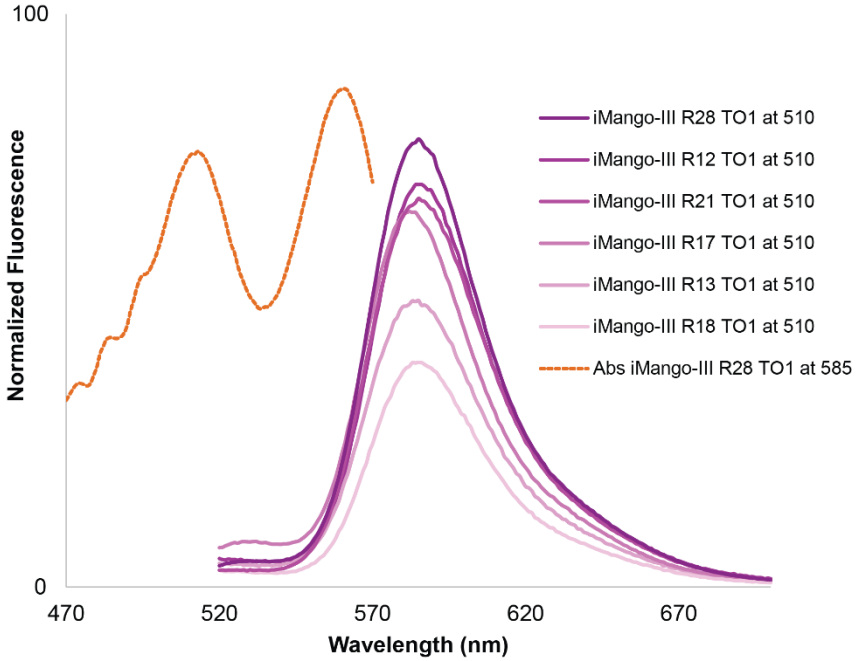


Figure 4.8. Fluorescent spectra for *i*Mango-III modified with rhodamine. Each pink spectra represents a different position of rhodamine labelled *i*Mango-III with the darkest shades indicating strongest fluorescence intensity. The absorbance spectrum for a representative rhodamine modified *i*Mango-III aptamer is shown in orange.

Table 4.1. Apparent FRET Efficiencies for rhodamine and fluorescein labeled *i*Mango-III.

Modified Position	Apparent FRET Efficiency	
	TO1/ Rhodamine	Fluorescein/ YO3
12	95%	44%
13	95%	21%
17	91%	44%
18	97%	44%
21	98%	46%
28	96%	35%

YO3-bound *i*Mango-III can be excited at 580 nm for a fluorescence signal with an emission maximum at 620 nm. When labelled with fluorescein, the aptamer-fluorophore complex can be excited at 494 nm resulting in an emissions spectrum with two maxima at 520 nm and 620 nm (Figure 4.9). The emission maxima observed at 520 nm corresponds to the fluorescence signal of fluorescein, while the maxima observed at 620 nm corresponds to the YO3-bound *i*Mango-III aptamer. The presence of the emission maxima at 620 nm suggests some internal FRET transfer between the fluorescein label and the YO3 ligand. The additional observed fluorescence signal of fluorescein at 520 nm indicates incomplete FRET transfer, and the FRET efficiencies for the fluorescein/YO3 pair are lower than those of the TO1/rhodamine pair with modifications in equivalent positions (Table 4.1). Similar to rhodamine modified positions 13 and 18, the fluorescent intensities at 620 nm for fluorescein labels in these

positions were significantly reduced compared to the brightest modified aptamers. This is consistent with the hypothesis that dramatic mutations at these positions disrupts the overall aptamer fold. Unlike in the rhodamine labelled aptamer, where the position 28 resulted in the highest fluorescence intensity, the modified positions resulting in highest fluorescence intensities at 620 nm were 21 and 12 with moderate FRET efficiencies of 46% and 44% respectively.

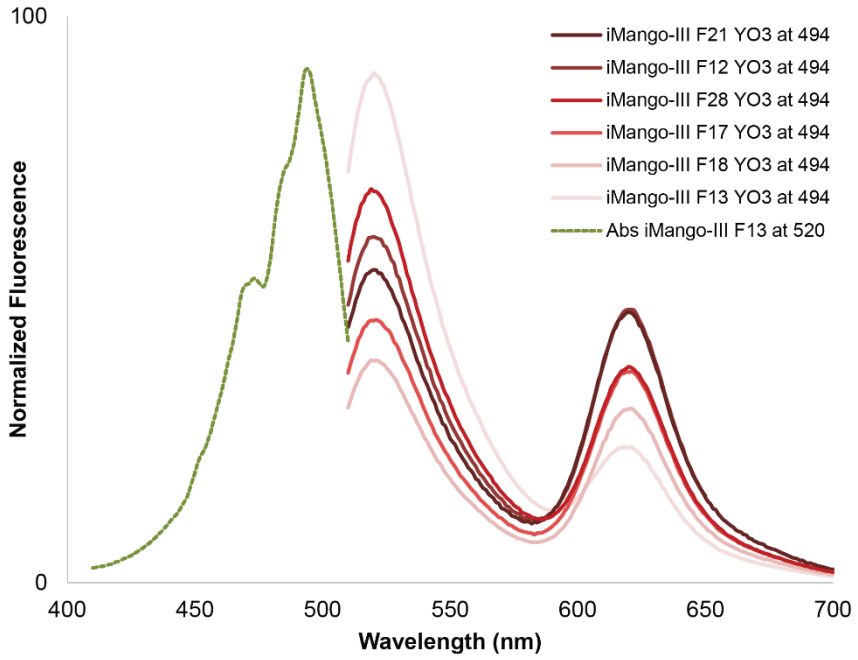


Figure 4.9. Fluorescent spectra for *i*Mango-III modified with fluorescein. Each red spectra represents a different position of fluorescein labelled *i*Mango-III with the darkest shades indicating strongest fluorescence intensity at 620 nm. The absorbance spectrum for a representative fluorescein modified *i*Mango-III aptamer is shown in green.

The effect of introducing fluorescent dyes to aptamers could reduce the binding affinities of *i*Mango-III for TO1 or YO3. Covalently linking planar, aromatic heterocycles to the RNA increases the dye's effective molarity which may cause it to outcompete the aptamer's cognate ligands. The observed FRET between the dye label

and fluorophore ligand in both the rhodamine and fluorescein modified systems suggests there is a population of the labelled fluorogenic aptamer which folds and binds its fluorophore ligand. However, reduced affinity may contribute to the observation of two fluorescence maxima in fluorescein labelled *i*Mango-III. I attempted to experimentally determine the binding affinities of the labelled aptamers using fluorescence titration but was unable to observe a reliable fluorescent signal using the nanomolar concentrations necessary for the nanomolar binding affinities of the Mango aptamers.

Overall, the observed FRET between TO1/rhodamine and fluorescein/YO3 demonstrates the system's potential. Using NHS chemistry to covalently attach dyes to amino-modified aptamers provides a straightforward method for tuning the spectral properties of existing fluorogenic RNA aptamers. The attached dye can function as either a FRET donor or FRET acceptor, affording the possibility of changing the fluorogenic aptamer's excitation or emissions wavelengths.

Table 4.2 Summary of fluorescence properties for modified Mango aptamers.

RNA	Fluorophore	Relative change in Fluorescence Intensity	Excitation maximum (nm)	Emission maximum (nm)
Mango-III U17(AP)	-	-	303	367
Mango-III U17(AP)	TO1	22%	303	535
Mango-III U17(PyC)	-	-	350	456
Mango-III U17(PyC)	TO1	187%	350	538
Mango-III A15(AP)	-	-	303	365
Mango-III A15(AP)	TO1	137%	303	541
Mango-III A15(PyC)	-	-	350	456
Mango-III A15(PyC)	TO1	10%	350	538
iMango III	TO1	-	510	530
iMango III, 12-Rhodamine	TO1	-	510	586
iMango III, 13-Rhodamine	TO1	-10%	510	585

iMango III, 17-Rhodamine	TO1	-36%	510	582
iMango III, 18-Rhodamine	TO1	-16%	510	585
iMango III, 21-Rhodamine	TO1	-50%	510	585
iMango III, 28-Rhodamine	TO1	-	510	585
iMango III	YO3	-	510	520
iMango III, 12-Fluorescein	YO3	-1%	494	521, 620
iMango III, 13-Fluorescein	YO3	-50%	494	520, 620
iMango III, 17-Fluorescein	YO3	-22%	494	522, 621
iMango III, 18-Fluorescein	YO3	-36%	494	520, 621
iMango III, 21-Fluorescein	YO3	-	494	520, 622
iMango III, 28-Fluorescein	YO3	-20%	494	522, 620

4.3 CONCLUSIONS AND FUTURE WORK

The fluorescent nucleoside analogues 2-aminopurine and pyrrolocytosine can be incorporated into sites surrounding the G-quadruplex face of RNA TERRA, a model G-quadruplex, and the ligand binding pocket of Mango-III, a fluorogenic RNA aptamer. These modifications then allow the remote excitation of a bound fluorophore, TO1, through internal FRET from the fluorescent nucleoside analogue to the bound fluorophore. Rhodamine and fluorescein, two common fluorescent dyes, were also covalently attached to amino modified *i*Mango-III in six positions surrounding the aptamer's ligand binding pocket using NHS chemistry. FRET was observed between the fluorescein label and YO3-bound *i*Mango-III, providing an effective stokes shift of 126 nm. FRET was also observed between TO1-bound *i*Mango-III and the rhodamine label, demonstrating high apparent FRET efficiencies. Of the positions studied, two modified positions (A13 and U18) displayed reduced fluorescence signal, suggesting the labelled aptamer misfolded. Additional studies to determine the effects of these modifications on the aptamer's binding affinity can provide further guidance on optimal positions for modification. The demonstration of remote excitation between 2-

aminopurine, pyrrolocytosine and TO1 in the model G-quadruplex and in the Mango-III aptamer in addition to the observed FRET between fluorescein/YO3 and TO1/rhodamine establishes the versatility of these internal FRET antennas for modifying the working excitation and emission wavelengths of fluorogenic RNA aptamers.

4.3 MATERIALS AND METHODS

4.3.1 AP and PyC labelled oligonucleotides

RNA TERRA and *i*Mango-III control oligonucleotide sequences and those containing 2-aminopurine and pyrrolocytosine were purchased from Dharmacon with HPLC-purification. RNAs were stored at 80 µM in DEPC-treated water until use. To fold RNA TERRA, the oligonucleotide was brought up to 1 µM in 20 mM potassium cacodylate pH 7.0, 150 mM KCl, heated at 95 °C for 5 min and slowly cooled to room temperature over 2 h. The folded RNA was run on a Superdex 75 increase size exclusion column to confirm homogeneity and analytical ultracentrifugation was performed to confirm a tetrameric fold. To fold *i*Mango-III, the oligonucleotide was brought up to 1 µM in 20 mM HEPES-KOH pH 7.0, 150 mM KCl, heated at 95 °C for 5 min, and slowly cooled to room temperature over 1 h. RNA TERRA and *i*Mango-III sequences used in this study are listed in (Supplementary Table S2).

4.3.2 Fluorescein and rhodamine labelled oligonucleotides

Amino-modified *i*Mango-III was purchased from Dharmacon and stored at 10 µg/µL in DEPC-treated water at -20 °C until use. NHS-fluorescein or NHS-rhodamine (Thermofisher) was dissolved in DMSO at 60 µg/µL, and freshly prepared for each use. NHS labelling was performed with 6x excess of the NHS-dye in 100 mM Na

tetraborate pH 8.5 and incubated at room temperature with shaking for 16 h. The reaction mixture was passed over a G-25 microspin column (GE-Healthcare) to remove excess dye and purified by HPLC (Buffer A: 100% TEA pH 8.0, Buffer B: 50% TEA pH 8.0, 50% MeCN; 0% to 30% Buffer B over 1 h). TEA and MeCN were removed using a table top vacuum centrifuge, and the HPLC purified product was run on a 15% urea PAGE gel and imaged without staining to confirm successful labelling. The labelled RNA was stored in DEPC-treated water at -20 °C until use.

4.3.3 Fluorescent measurements

Absorbance and fluorescence measurements were recorded using a Varian Cary Eclipse spectrophotometer at 303 nM (AP), 350 nm (PyC), 494 nm (fluorescein) 510 nm (TO1), 585 nm (rhodamine), 620 nm (YO3). In all cases, RNA was folded in the absence of ligand and incubated with ligand for 5 min before fluorescence measurements. Monochromator slits were set to 4 nm and integration time was 1 s. Measurements were taken on samples with a volume of 120 μL. Fluorescence signals were normalized using the absorbance signal of the labeled dye to ensure consistent concentrations across samples.

4.3.4 Calculations for apparent FRET efficiencies

Ratiometric FRET intensities were calculated using the donor/acceptor intensity ratios for TO1/rhodamine and fluorescein/YO3. This was given by

$$E_{rel} = \frac{I_A}{I_D + I_A}$$

where I_A is the fluorescence intensity of the acceptor and I_D is the intensity of the donor following excitation of the donor.

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

The development of fluorogenic RNA aptamers enables the imaging of native RNA molecules without the need for traditional protein-based methods. Aptamers such as Spinach, Corn, Mango, and DIR2s provide a basic toolbox of RNA-based imaging probes that function at different fluorescent wavelengths. While the first-generation versions of these fluorogenic RNA aptamers demonstrate incredible promise, further optimization to improve folding and stability, enhance brightness, and broaden the spectrum of available wavelengths would facilitate the general and widespread use of these new tools. Previous efforts to provide access to additional wavelengths have focused primarily on changing the fluorophore-ligand, which often demands the selection of an entirely new aptamer. This approach is time and labor intensive, and the development of a more versatile method for accessing additional wavelengths would improve the utility and convenience of these fluorogenic RNA aptamers. As part of this effort, I have explored the potential of incorporating chemical modifications to the RNA aptamer for spectral tuning. In my research, I have investigated the use of canonical mutations, naturally occurring modifications, artificial nucleotide analogues, and fluorescent dye labels in tuning the fluorescent properties of these RNA aptamers.

The use of naturally occurring modifications in RNA is currently limited by the inaccessibility of such modifications. To address this, I have synthesized the phosphoramidite forms of two naturally occurring tRNA modifications: $PreQ_0$ and wyosine. While the nucleosides for these tRNA modifications have previously been synthesized, their corresponding phosphoramidites have never been reported in current literature. These phosphoramidites enable the site selective incorporation of $PreQ_0$ and wyosine into a strand of RNA using solid-phase oligonucleotide synthesis. Basing my

synthesis on previously established methods, I optimized the synthetic scheme of these nucleosides for phosphoramidite synthesis. For $PreQ_0$, I reorganized the synthetic steps and utilized a dibutylformamidyl protecting group to improve solubility of the nucleoside. For wyosine, I developed a new protecting group strategy using a combination of TIPDS and acetyl groups for selective access to the nucleosides 2′ hydroxyl.

To approach the phosphoramidite conversion, I explored two methods to address the 2′ and 3′ isomerization issue often faced in RNA phosphoramidite synthesis. First, I investigated the TIPS route, using the bulky isopropyls on the TIPS protecting group to prevent 2′ 3′ isomerization. While the steric hinderance provided by the group did seem to prevent isomerization, the bulk also prevented phosphitylation at the 3′ position – the final synthetic step necessary for phosphoramidite synthesis. This led me instead to explore the use of thionocarbamate chemistry for phosphoramidite generation. Using TC chemistry, I was able to access the phosphoramidite form of $PreQ_0$ and wyosine. While there was no isomerization observed with the $PreQ_0$ phosphoramidite, some isomerization was observed with the wyosine phosphoramidite. Due to time constraints, both phosphoramidites were used as is. In future, to address isomerization of the wyosine phosphoramidite, an alternative protection scheme such as TOM may be investigated. Finally, while the nucleoside form of archaeosine was also generated, the corresponding phosphoramidite was not pursued due to the difficulty of protecting the exocyclic amidine group. Archaeosine remains an interesting modification for its formal positive charge, and further exploration of the phosphoramidite synthesis can be conducted by investigating the use of other protecting groups suitable for the amidine.

In addition to exploring the use of naturally occurring post-transcriptional tRNA modifications, I investigated the use of canonical mutations and artificial nucleotide analogues in a Spinach aptamer. Previous studies have indicated that the absorbance and fluorescence wavelengths of the Spinach aptamer could be blue shifted by a mutation to the base triple atop the aptamer's ligand binding core. In testing this mutation in complex with several fluorophore ligands (DFHBI, DFHO, ThT, and TO1), all of which will non-selectively bind to Spinach, I discovered that only DFHBI and analogues exhibit a spectral shift. To further probe the potential of the base triple mutations, I investigated several base triple mutations in Baby Spinach, a miniaturization of the Spinach aptamer. Surprisingly all base triple mutations, which still exhibited some degree of fluorescence turn on, consistently blue shifted the absorbance maxima 48 nm and the fluorescence maxima 12 nm. This suggests that mutating the base triple produces a hypsochromic shift by enhancing a minor UV band in DFHBI. While preliminary attempts to crystallize a Spinach aptamer variant with base triple mutations were unsuccessful, structural characterization of these mutants would provide additional insight on the interplay of the fluorophore-ligand and the flanking base triple.

To study the effects of artificial nucleotide analogues on Spinach, I first developed a Split Baby Spinach aptamer which would be amenable to solid-phase oligonucleotide synthesis. I then modified several positions surrounding the ligand-binding pocket of the aptamer with 2-aminopurine or pyrrolocytosine and studied the effects on fluorescence when complexed with DFHBI, DFHBI-1T, and DFHO. While the modifications alone did not produce a noticeable shift in absorbance or fluorescence, modified Split Baby Spinach in complex with DFHO exhibited multiple absorbance

bands that could be separately excited to produce fluorescence spectra with maxima at various wavelengths. This was not apparent when Split Baby Spinach was bound to DFHBI or DFHBI-1T. The use of $PreQ_0$ and wyosine modifications in positions surrounding the ligand binding pocket generally had negligible effects on the aptamer's spectral properties, with the exception of the modified BS2 A16(Wyosine) aptamer which exhibited greater than two-fold fluorescence enhancement when bound to DFHBI 1T. Additional study would utilize the $PreQ_0$ modified aptamer to probe the binding pocket of Split Baby Spinach, as the nitrile functional group on $PreQ_0$ can be used as an IR probe.

Finally, I used fluorescent labels on Mango-III aptamers to determine the feasibility of developing an internal FRET antenna for remote excitation of the fluorogenic RNA aptamer. Using RNA TERRA as a model G-quadruplex, I demonstrated the 2-aminopurine and pyrrolocytosine labels could be used to remotely excite a bound TO1 ligand. Translating this system into Mango-III, I similarly demonstrated this versatile technique could be used to remotely excite TO1-bound Mango-III using 2-aminopurine or pyrollocytosine. Taking this internal FRET antenna one step further, I labelled *i*Mango-III in positions surrounding the ligand binding pocket with fluorescent dyes. Using fluorescein labelled *i*Mango-III bound to YO3, I demonstrated remote excitation of fluorescein at 494 nm for a fluorescence maximum at 620 nm, a 126 nm Stokes shift. I also showed the fluorophore could function as a FRET donor, and the label as a FRET acceptor. Using rhodamine labelled *i*Mango-III bound to TO1, I was able to excite the fluorogenic aptamer at 510 nm to observe a fluorescence signal from the rhodamine label at 585 nm. This system thus provides a straightforward method for modifying the

working excitation or emission wavelength of a fluorogenic RNA aptamer by introducing an internal FRET antenna.

Through this research, I have explored the use of chemical modifications for spectrally tuning fluorogenic RNA aptamers. Using canonical mutations and the chose naturally occurring modifications generally had little effect on shifting the absorbance and fluorescence spectra of the modified aptamers, and the effects that were observed stemmed primarily from properties of the fluorophore-ligand. Other modifications may be explored with greater success, although changes in the resulting fluorescence remains unpredictable. For more precise tuning, fluorescent dye labels can be covalently attached to the fluorogenic RNA aptamer to create an internal FRET antenna for remote excitation of the fluorophore or fluorescent label. This provides a more versatile approach for tuning the aptamer, though the system is currently better suited for *in vitro* rather than *in vivo* imaging.

REFERENCES

1. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, purification and properties of Aequorin, a bioluminescent protein from the luminous Hydromedusan,Aequorea. *J. Cell. Comp. Physiol.* **59,** 223–239 (1962).

2. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263,** 802–5 (1994).

3. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111,** 229–233 (1992).

4. Day, R. N. & Davidson, M. W. (Michael W. *The fluorescent protein revolution*.

5. Rodriguez, E. A. *et al.* The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends Biochem. Sci.* **42,** 111–129 (2017).

6. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67,** 509–544 (1998).

7. Day, R. N. & Davidson, M. W. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.* **38,** 2887 (2009).

8. Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins as a toolkit for *in vivo* imaging. *Trends Biotechnol.* **23,** 605–613 (2005).

9. Cubitt, A. B. *et al.* Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20,** 448–455 (1995).

10. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822 (1990).

11. Ouellet, J. RNA fluorescence with light-up aptamers. *Front. Chem.* **4,** 29 (2016).

12. Bouhedda, F. *et al.* Light-up RNA aptamers and their cognate fluorogens: from their development to their applications. *Int. J. Mol. Sci.* **19,** 44 (2017).

13. Ormö, M. *et al.* Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273,** 1392–5 (1996).

14. Yang, F., Moss, L. G. & Phillips, G. N. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14,** 1246–1251 (1996).

15. Zimmer, M. Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem. Rev.* **102,** 759–782 (2002).

16. Meech, S. R. Excited state reactions in fluorescent proteins. *Chem. Soc. Rev.* **38,** 2922 (2009).

17. Ward, W. W. & Bokman, S. H. Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry* **21,** 4535–4540 (1982).

18. Niwa, H. *et al.* Chemical nature of the light emitter of the Aequorea green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13617–22 (1996).

19. Zhang, G., Gurtu, V. & Kain, S. R. An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem. Biophys. Res. Commun.* **227,** 707–711 (1996).

20. Arpino, J. A. J., Rizkallah, P. J. & Jones, D. D. Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations

for Glu222. *PLoS One* **7,** e47132 (2012).

21. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G. & Ward, W. W. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. *Biochemistry* **32,** 1212–1218 (1993).

22. Reid, B. G. & Flynn, G. C. Chromophore formation in green fluorescent protein. *Biochemistry* **36,** 6786–6791 (1997).

23. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 12501–4 (1994).

24. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6,** 178–182 (1996).

25. Mena, M. A., Treynor, T. P., Mayo, S. L. & Daugherty, P. S. Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat. Biotechnol.* **24,** 1569–1571 (2006).

26. Wachter, R. M., Elsliger, M.-A., Kallio, K., Hanson, G. T. & Remington, S. J. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6,** 1267–1277 (1998).

27. Strack, R. L., Strongin, D. E., Mets, L., Glick, B. S. & Keenan, R. J. Chromophore formation in DsRed occurs by a branched pathway. *J. Am. Chem. Soc.* **132,** 8496–8505 (2010).

28. Bravaya, K. B., Subach, O. M., Korovina, N., Verkhusha, V. V. & Krylov, A. I. Insight into the common mechanism of the chromophore formation in the red fluorescent proteins: the elusive blue intermediate revealed. *J. Am. Chem. Soc.* **134,** 2807–2814 (2012).

29. Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14,** 315–319 (1996).

30. Nienhaus, G. U. & Wiedenmann, J. Structure, dynamics and optical properties of fluorescent proteins: perspectives for marker development. *ChemPhysChem* **10,** 1369–1379 (2009).

31. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7877–82 (2002).

32. Merzlyak, E. M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* **4,** 555–557 (2007).

33. Fradkov, A. F. *et al.* Far-red fluorescent tag for protein labelling. *Biochem. J.* **368,** 17–21 (2002).

34. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–10 (1990).

35. Robertson, D. L. & Joyce, G. F. Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344,** 467–468 (1990).

36. Grate, D. & Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6131–6 (1999).

37. Babendure, J. R., Adams, S. R. & Tsien, R. Y. Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.* **125,** 14716–14717 (2003).

38. Baugh, C., Grate, D. & Wilson, C. 2.8 Å crystal structure of the malachite green aptamer. *J. Mol. Biol.* **301,** 117–128 (2000).

39. Sando, S., Narita, A. & Aoyama, Y. Light-up hoechst–DNA aptamer pair: generation of an aptamer-selective fluorophore from a conventional DNA-staining dye. *ChemBioChem* **8,** 1795–1803 (2007).

40. Tudor P. Constantin, † *et al.* Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules. (2008). doi:10.1021/OL702920E

41. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334,** 364–366 (1988).

42. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342,** 825–829 (1989).

43. Wang, Y., Jin, R., Gaffney, B., Jones, R. A. & Breslauer, K. J. Characterization by ^{1}H NMR of glycosidic conformations in the tetramolecular complex formed by d(GGTTTTTGG). *Nucleic Acids Res.* **19,** 4619–4622 (1991).

44. Smirnov, I. & Shafer, R. H. Effect of Loop Sequence and Size on DNA Aptamer Stability †. *Biochemistry* **39,** 1462–1468 (2000).

45. Warner, K. D. *et al.* Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **21,** 658–663 (2014).

46. Huang, H. *et al.* A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **10,** 686–691 (2014).

47. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18,** 279–284 (2017).

48. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333,** 642–6 (2011).

49. Han, K. Y., Leslie, B. J., Fei, J., Zhang, J. & Ha, T. Understanding the photophysics of the Spinach–DFHBI RNA aptamer–fluorogen complex to improve live-vell RNA imaging. *J. Am. Chem. Soc.* **135,** 19033–19038 (2013).

50. Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA. *Nat. Methods* **10,** 1219–1224 (2013).

51. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136,** 16299–16308 (2014).

52. Zou, J. *et al.* Selection of intracellularly functional RNA mimics of green fluorescent protein using fluorescence-activated cell sorting. *J. Mol. Evol.* **81,** 172–178 (2015).

53. Autour, A., Westhof, E. & Ryckelynck, M. iSpinach: a fluorogenic RNA aptamer optimized for *in vitro* applications. *Nucleic Acids Res.* **44,** 2491–2500 (2016).

54. Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* **136,** 1198–1201 (2014).

55. Song, W. *et al.* Imaging RNA polymerase III transcription using a photostable RNA–fluorophore complex. *Nat. Chem. Biol.* **13,** 1187–1194 (2017).

56. Warner, K. D. *et al.* A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nat. Chem. Biol.* **13,** 1195–1201 (2017).

57. Jones, C. P. & Ferré-D'Amaré, A. R. RNA quaternary structure and global symmetry. *Trends Biochem. Sci.* **40,** 211–220 (2015).

58. Steinmetzger, C., Palanisamy, N., Gore, K. R. & Höbartner, C. A multicolor large stokes shift fluorogen-activating RNA aptamer with cationic chromophores. *Chem. – A Eur. J.* **25,** 1931–1935 (2019).

59. Shcherbakova, D. M., Hink, M. A., Joosen, L., Gadella, T. W. J. & Verkhusha, V. V. An orange fluorescent protein with a large stokes shift for single-excitation multicolor FCCS and FRET imaging. *J. Am. Chem. Soc.* **134,** 7913–7923 (2012).

60. Piatkevich, K. D. *et al.* Monomeric red fluorescent proteins with a large Stokes shift. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 5369–74 (2010).

61. Nygren, J., Svanvik, N. & Kubista, M. The interactions between the fluorescent

dye thiazole orange and DNA. *Biopolymers* **46,** 39–51 (1998).

62. Dolgosheina, E. V. *et al.* RNA Mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.* **9,** 2412–2420 (2014).

63. Trachman, R. J. *et al.* Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat. Chem. Biol.* **13,** 807–813 (2017).

64. Autour, A. *et al.* Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat. Commun.* **9,** 656 (2018).

65. Trachman, R. J. *et al.* Crystal structures of the Mango-II RNA aptamer reveal heterogeneous fluorophore binding and guide engineering of variants with improved selectivity and brightness. *Biochemistry* **57,** 3544–3548 (2018).

66. Trachman, R. J. I. *et al.* Structure and functional reselection of the Mango-III fluorogenic RNA aptamer. *Nat. Chem. Biol.* 472–479 (2019).

67. Pleij, C. W., Rietveld, K. & Bosch, L. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13,** 1717–31 (1985).

68. Tan, X. *et al.* Fluoromodules consisting of a promiscuous RNA aptamer and red or blue fluorogenic cyanine dyes: selection, characterization, and bioimaging. *J. Am. Chem. Soc.* **139,** 9001–9009 (2017).

69. Shelke, S. A. *et al.* Structural basis for activation of fluorogenic dyes by an RNA aptamer lacking a G-quadruplex motif. *Nat. Commun.* **9,** 4542 (2018).

70. Braselmann, E. *et al.* A multicolor riboswitch-based platform for imaging of

RNA in live mammalian cells. *Nat. Chem. Biol.* **14,** 964–971 (2018).

71. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441,** 1167–1171 (2006).

72. Chen, X., Zhang, D. & Su, N. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nat. Biotechnol.* doi:10.1038/s41587-019-0249-1

73. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2,** 437–445 (1998).

74. Garcia, J. F. & Parker, R. MS2 coat proteins bound to yeast mRNAs block 5' to 3' degradation and trap mRNA decay products: implications for the localization of mRNAs by MS2-MCP system. *RNA* **21,** 1393–1395 (2015).

75. Guet, D. *et al.* Combining Spinach-tagged RNA and gene localization to image gene expression in live yeast. *Nat. Commun.* **6,** 8882 (2015).

76. Höfer, K., Langejürgen, L. V. & Jäschke, A. Universal aptamer-based real-time monitoring of enzymatic RNA synthesis. *J. Am. Chem. Soc.* **135,** 13692–13694 (2013).

77. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene position more strongly influences vell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.* **3,** 363–371 (2014).

78. Paige, J. S., Nguyen-Duc, T., Song, W. & Jaffrey, S. R. Fluorescence imaging of

cellular metabolites with RNA. *Science (80-.).* **335,** 1194 (2012).

79. You, M., Litke, J. L. & Jaffrey, S. R. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E2756-65 (2015).

80. Kellenberger, C. A. *et al.* A minimalist biosensor: Quantitation of cyclic di-GMP using the conformational change of a riboswitch aptamer. *RNA Biol.* **12,** 1189–1197 (2015).

81. Kellenberger, C. A. *et al.* GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc. Natl. Acad. Sci.* **112,** 5383–5388 (2015).

82. Sekella, P. T., David, R. & Walter, N. G. A biosensor for theophylline based on fluorescence detection of ligand-induced hammerhead ribozyme cleavage. *RNA* **8,** S1355838202028066 (2002).

83. Jepsen, M. D. E. *et al.* Development of a genetically encodable FRET system using fluorescent RNA aptamers. *Nat. Commun.* **9,** 18 (2018).

84. Roszyk, L., Kollenda, S. & Hennig, S. Using a specific RNA–protein interaction to quench the fluorescent RNA Spinach. *ACS Chem. Biol.* **12,** 2958–2964 (2017).

85. Rogers, T. A., Andrews, G. E., Jaeger, L. & Grabow, W. W. Fluorescent monitoring of RNA assembly and processing using the split-Spinach aptamer. *ACS Synth. Biol.* **4,** 162–166 (2015).

86. Chandler, M. *et al.* Broccoli fluorets: split aptamers as a user-friendly fluorescent toolkit for dynamic RNA nanotechnology. *Molecules* **23,** 3178 (2018).

87. Kikuchi, N. & Kolpashchikov, D. M. Split spinach aptamer for highly selective recognition of DNA and RNA at ambient temperatures. *ChemBioChem* **17,** 1589–1592 (2016).

88. Ausländer, S., Fuchs, D., Hürlemann, S., Ausländer, D. & Fussenegger, M. Engineering a ribozyme cleavage-induced split fluorescent aptamer complementation assay. *Nucleic Acids Res.* **44,** e94–e94 (2016).

89. Alam, K. K., Tawiah, K. D., Lichte, M. F., Porciani, D. & Burke, D. H. A fluorescent split aptamer for visualizing RNA–RNA assembly *in vivo*. *ACS Synth. Biol.* **6,** 1710–1721 (2017).

90. Seidel, C. A. M., Schulz, A. & Sauer, M. H. M. Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *J. Phys. Chem.* **100,** 5541–5553 (1996).

91. Yang, X. *et al.* G-quadruplex fluorescence quenching ability: a simple and efficient strategy to design a single-labeled DNA probe. *Anal. Methods* **4,** 895 (2012).

92. Dao, N. T., Haselsberger, R., Michel-Beyerle, M.-E. & Phan, A. T. Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Lett.* **585,** 3969–3977 (2011).

93. Sherlock, M. E. *et al.* Steady-state and time-resolved studies into the origin of the intrinsic fluorescence of G-quadruplexes. *J. Phys. Chem. B* **120,** 5146–5158 (2016).

94. Filonov, G. S., Song, W. & Jaffrey, S. R. Spectral tuning by a single nucleotide controls the fluorescence properties of a fluorogenic aptamer. *Biochemistry* **17,** 44 (2019).

95. Cantara, W. A. *et al.* The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39,** D195–D201 (2011).

96. Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip. Rev. RNA* **4,** 35–48 (2013).

97. Chow, C. S., Lamichhane, T. N. & Mahto, S. K. Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chemical Biology* **2,** 610–619 (2007).

98. Caruthers, M. H. Chemical synthesis of DNA and DNA analogs. *Acc. Chem. Res.* **24,** 278–284 (1991).

99. Brückl, T., Thoma, I., Wagner, A. J., Knochel, P. & Carell, T. Efficient Synthesis of Deazaguanosine-Derived tRNA Nucleosides PreQ0, PreQ1, and Archaeosine Using the Turbo-Grignard Method. *European J. Org. Chem.* **2010,** 6517–6519 (2010).

100. Nakatsuka, S., Ohgi, T. & Goto, T. Synthesis of wyosine (nucleoside Yt), a strongly fluorescent nucleoside found in Torulopsis utilis tRNAPhe, and 3-

methylguanosine. *Tetrahedron Lett.* **19,** 2579–2582 (1978).

101. Hienzsch, A., Deiml, C., Reiter, V. & Carell, T. Total Synthesis of the Hypermodified RNA Bases Wybutosine and Hydroxywybutosine and Their Quantification Together with Other Modified RNA Bases in Plant Materials. *Chem. - A Eur. J.* **19,** 4244–4248 (2013).

102. Lindquist, B. A., Furse, K. E. & Corcelli, S. A. Nitrile groups as vibrational probes of biomolecular structure and dynamics: an overview. *Phys. Chem. Chem. Phys.* **11,** 8119 (2009).

103. Hughes, R. A., Miklos, A. E. & Ellington, A. D. in *Methods in Enzymology* **498,** 277–309 (Academic Press Inc., 2011).

104. Vorbrüggen, H. & Höfle, G. Nucleoside Syntheses, XXIII1) On the Mechanism of Nucleoside Synthesis. *Chem. Ber.* **114,** 1256–1268 (1981).

105. Vorbrüggen, H., Lagoja, I. M. & Herdewijn, P. in *Current Protocols in Nucleic Acid Chemistry* **27,** 1.13.1-1.13.16 (John Wiley & Sons, Inc., 2006).

106. Ming, X. & Seela, F. Efficient Synthesis of the tRNA Nucleoside PreQ0, 7-Cyano-7-deazaguanosine, via Microwave-Assisted Iodo→Carbonitrile Exchange. *Chem. Biodivers.* **7,** 2616–2621 (2010).

107. Kasai, H. *et al.* Structure of wye (Yt base) and wyosine (Yt) from Torulopsis utilis phenylalanine transfer ribonucleic acid. *Biochemistry* **15,** 898–904 (1976).

108. Jahnz-Wechmann, Z., Framski, G. R., Januszczyk, P. A. & Boryski, J. Base-Modified Nucleosides: Etheno Derivatives. *Front. Chem.* **4,** 19 (2016).

109. Dellinger, D. J. *et al.* Streamlined Process for the Chemical Synthesis of RNA Using 2′- *O* -Thionocarbamate-Protected Nucleoside Phosphoramidites in the Solid Phase. *J. Am. Chem. Soc.* **133,** 11540–11556 (2011).

110. Somoza, Á. Protecting groups for RNA synthesis: an increasing need for selective preparative methods. *Chem. Soc. Rev.* **37,** 2668 (2008).

111. Komatsu, Y. & Ohtsuka, E. Chemical RNA Synthesis (Including RNA with Unusual Constituents). *Compr. Nat. Prod. Chem.* 81–96 (1999). doi:10.1016/B978-0-08-091283-7.00140-5

112. Marshall, W. S. & Kaiser, R. J. Recent advances in the high-speed solid phase synthesis of RNA. *Curr. Opin. Chem. Biol.* **8,** 222–229 (2004).

113. Usman, N., Ogilvie, K. K., Jiang, M. Y. & Cedergren, R. J. The automated chemical synthesis of long oligoribuncleotides using 2’-O-silylated ribonucleoside 3’-O-phosphoramidites on a controlled-pore glass support: synthesis of a 43-nucleotide sequence similar to the 3’-half molecule of an Escherichia coli formylmethionine tRNA. *J. Am. Chem. Soc.* **109,** 7845–7854 (1987).

114. Pitsch, S., Weiss, P. A., Jenny, L., Stutz, A. & Wu, X. Reliable Chemical Synthesis of Oligoribonucleotides (RNA) with 2′-O-[(Triisopropylsilyl)oxy]methyl(2′-O-tom)-Protected Phosphoramidites. *Helv. Chim. Acta* **84,** 3773–3795 (2001).

115. Scaringe, S. A., Kitchen, D., Kaiser, R. J. & Marshall, W. S. in *Current Protocols in Nucleic Acid Chemistry* **16,** 2.10.1-2.10.16 (John Wiley & Sons,

Inc., 2004).

116. Hakimelahi, G. H., Proba, Z. A. & Ogilvie, K. K. New catalysts and procedures for the dimethoxytritylation and selective silylation of ribonucleosides. *Can. J. Chem.* **60,** 1106–1113 (1982).

117. Strack, R. L. & Jaffrey, S. R. Live-cell imaging of mammalian RNAs with Spinach2. *Methods Enzymol.* **550,** 129–146 (2015).

118. Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3,** 751 (2012).

119. Ng, H.-L. & Lin, M. Z. Structure-guided wavelength tuning in far-red fluorescent proteins. *Curr. Opin. Struct. Biol.* **39,** 124–133 (2016).

120. Jeng, S. C. Y., Chan, H. H. Y., Booy, E. P., McKenna, S. A. & Unrau, P. J. Fluorophore ligand binding and complex stabilization of the RNA Mango and RNA Spinach aptamers. *RNA* **22,** 1884–1892 (2016).

121. Zhang, X. F., Xu, H. M., Han, L., Li, N. B. & Luo, H. Q. A Thioflavin T-induced G-Quadruplex Fluorescent Biosensor for Target DNA Detection. *Anal. Sci.* **34,** 149–153 (2018).

122. Sjekloća, L. & Ferré-D'Amaré, A. R. Binding between G Quadruplexes at the Homodimer Interface of the Corn RNA Aptamer Strongly Activates Thioflavin T Fluorescence. *Cell Chem. Biol.* **26,** 1159–1168.e4 (2019).

123. Fernandez-Millan, P., Autour, A., Ennifar, E., Westhof, E. & Ryckelynck, M. Crystal structure and fluorescence properties of the iSpinach aptamer in complex

with DFHBI. *RNA* **23,** 1788–1795 (2017).

124. Abu Almakarem, A. S., Petrov, A. I., Stombaugh, J., Zirbel, C. L. & Leontis, N. B. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.* **40,** 1407–23 (2012).

125. Reese, C. B. Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Org. Biomol. Chem.* **3,** 3851 (2005).

126. Wilhelmsson, L. M. Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.* **43,** 159–183 (2010).

127. Trachman, R. J. & Ferré-D'Amaré, A. R. Tracking RNA with light: selection, structure, and design of fluorescence turn-on RNA aptamers. *Q. Rev. Biophys.* **52,** e8 (2019).

128. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 719–26 (1967).

129. Bajar, B., Wang, E., Zhang, S., Lin, M. & Chu, J. A Guide to Fluorescent Protein FRET Pairs. *Sensors* **16,** 1488 (2016).

130. Algar, W. R., Hildebrandt, N., Vogel, S. S. & Medintz, I. L. FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. *Nat. Methods* **16,** 815–829 (2019).

131. Haller, A., Rieder, U., Aigner, M., Blanchard, S. C. & Micura, R. Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.* **7,** 393–400 (2011).

132. Ha, T. *et al.* Ligand-induced conformational changes observed in single RNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9077–82 (1999).

133. Juskowiak, B. Nucleic acid-based fluorescent probes and their analytical potential. *Anal. Bioanal. Chem.* **399,** 3157–3176 (2011).

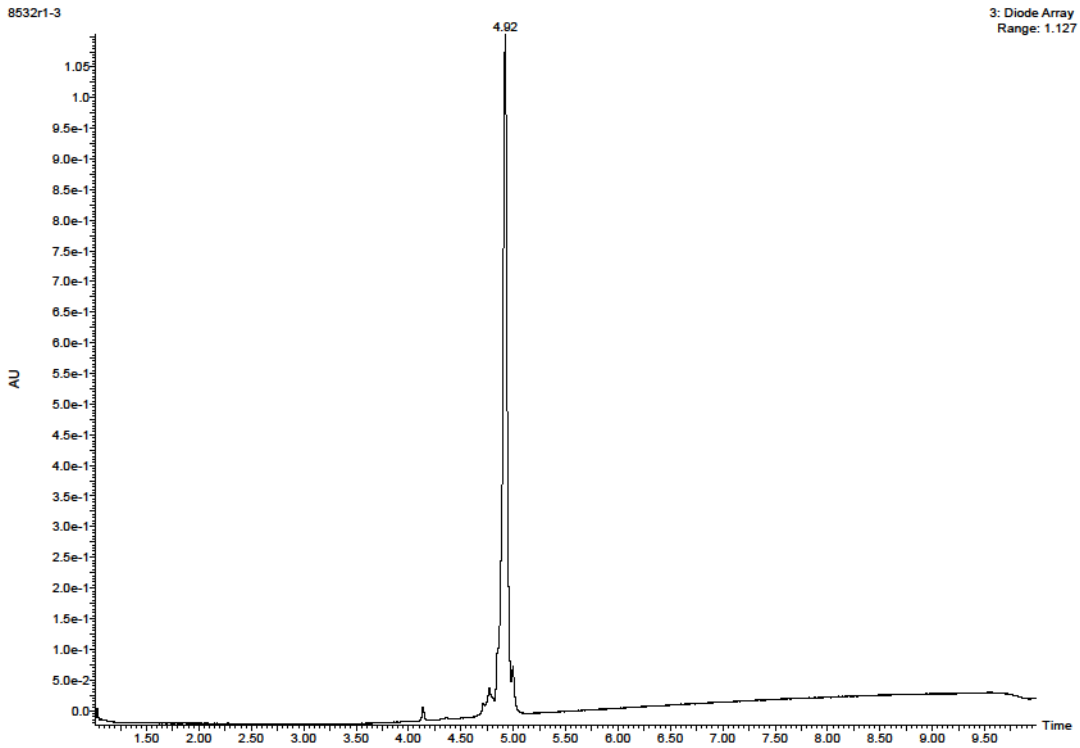
134. Lee, L. G., Chen, C.-H. & Chiu, L. A. Thiazole orange: A new dye for reticulocyte analysis. *Cytometry* **7,** 508–517 (1986).

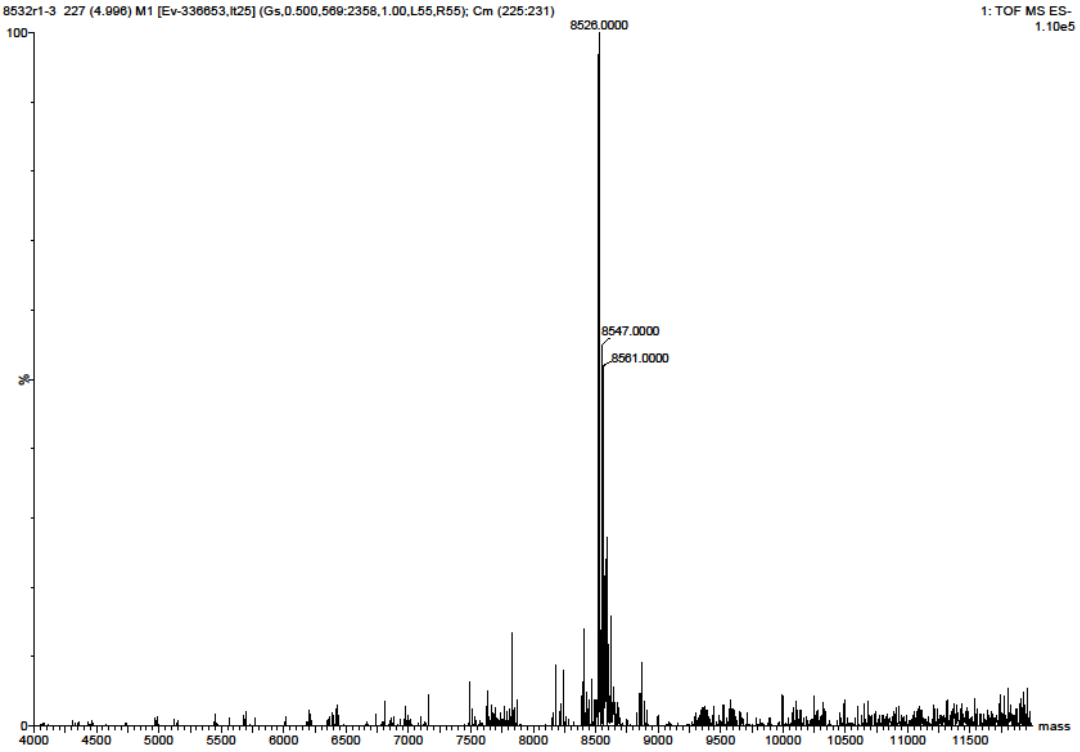
135. Cusanelli, E. & Chartrand, P. Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front. Genet.* **6,** 143 (2015).

136. Martadinata, H. & Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K^+ Solution. *Biochemistry* **52,** 2176–2183 (2013).

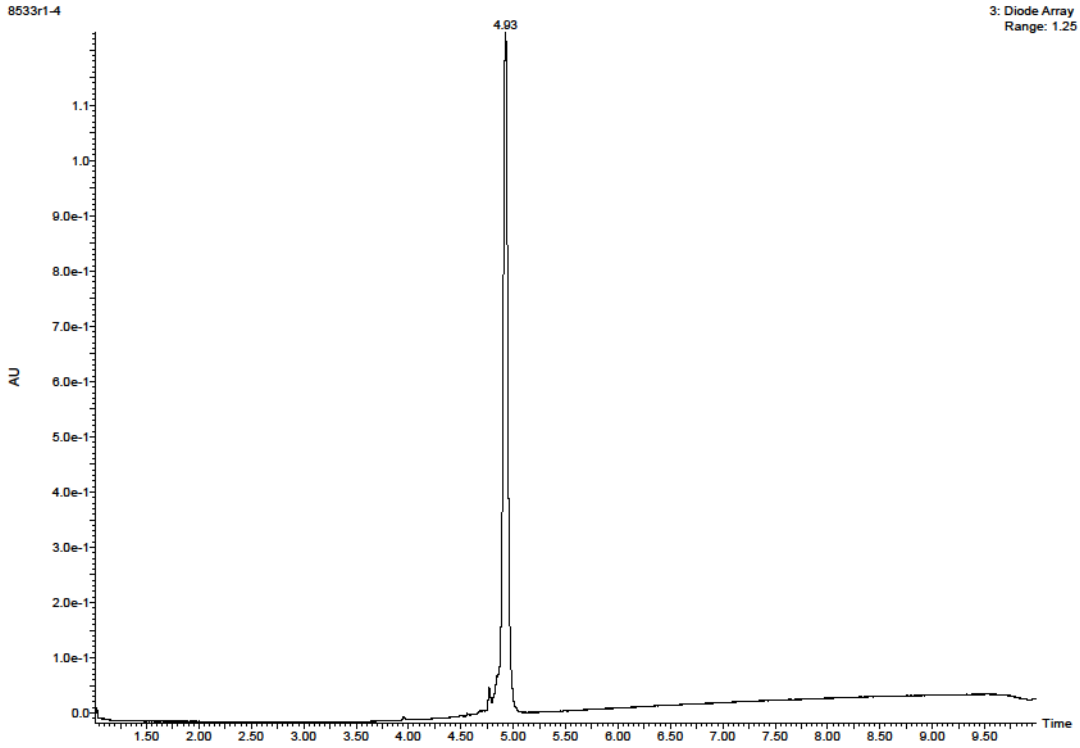
APPENDIX

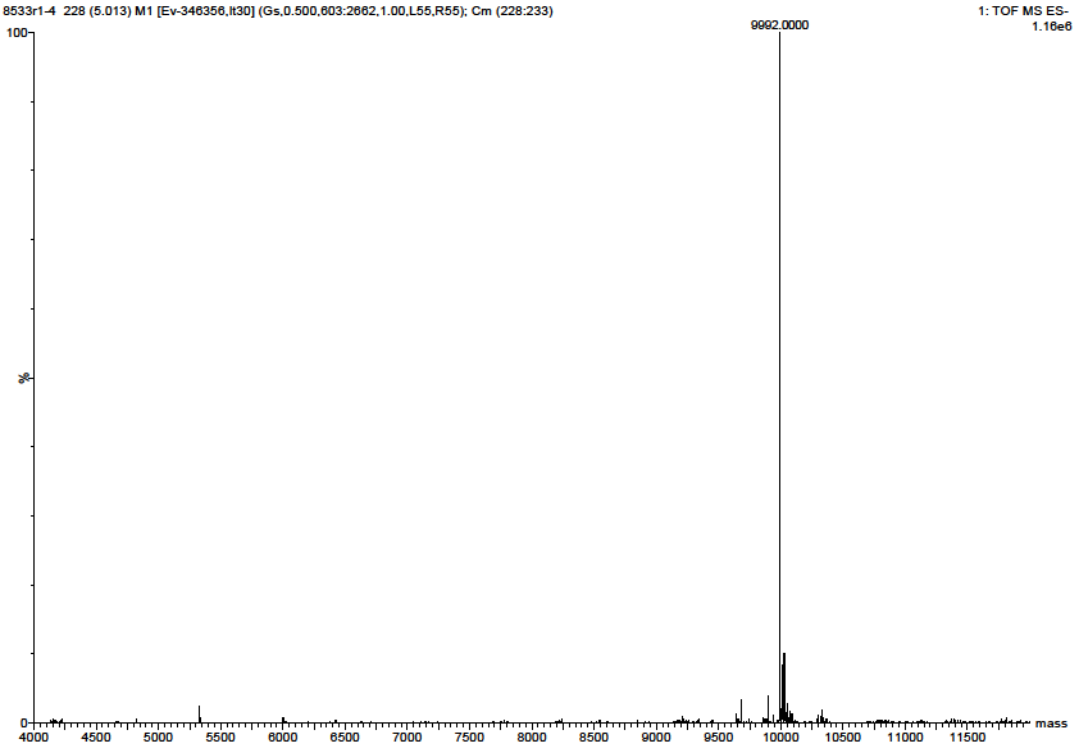
BS1 $PreQ_0$ HPLC trace and mass spectrum

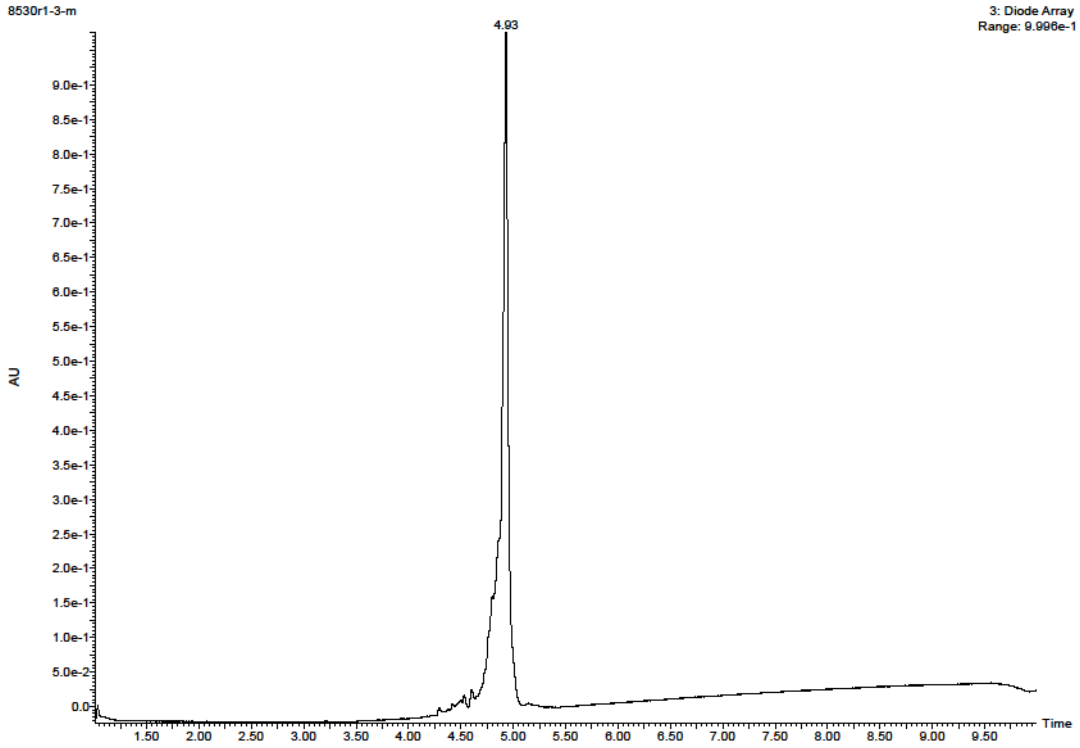


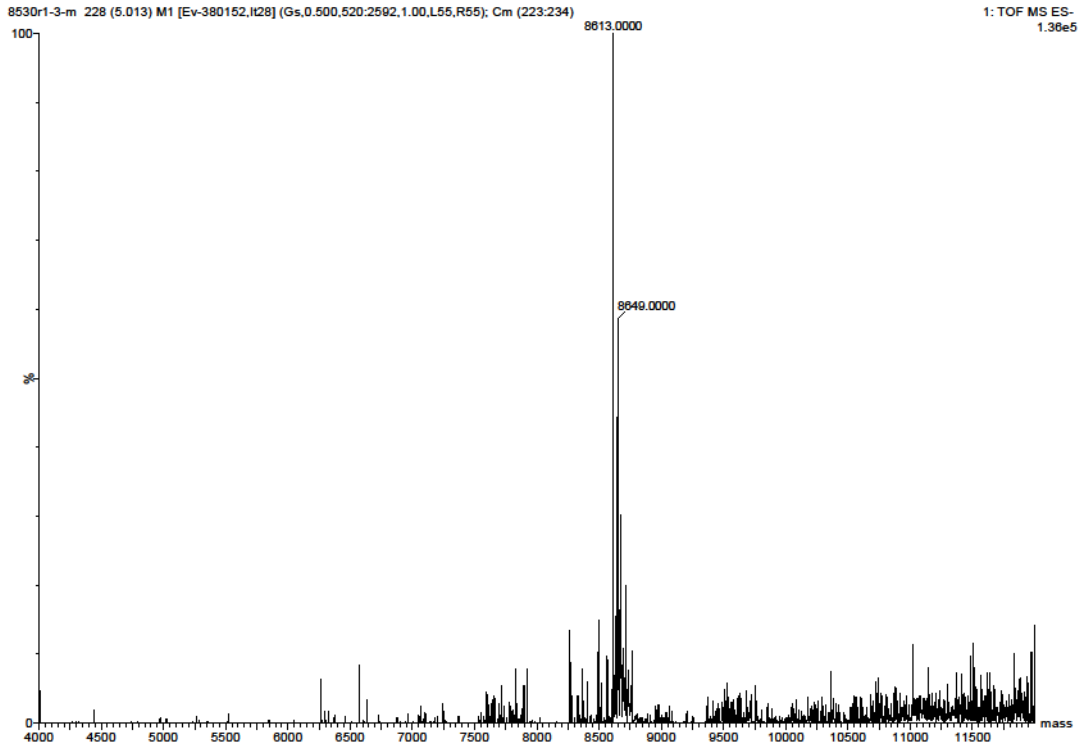

8532r1-3 227 (4.996) M1 [Ev-336653,It25] (Gs,0.500,569:2358,1.00,L55,R55); Cm (225:231)
1: TOF MS ES-
1.10e5
100
%
0
8526.0000
8547.0000
8561.0000
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500
11000
11500
mass

BS2 $PreQ_0$ HPLC trace and mass spectrum

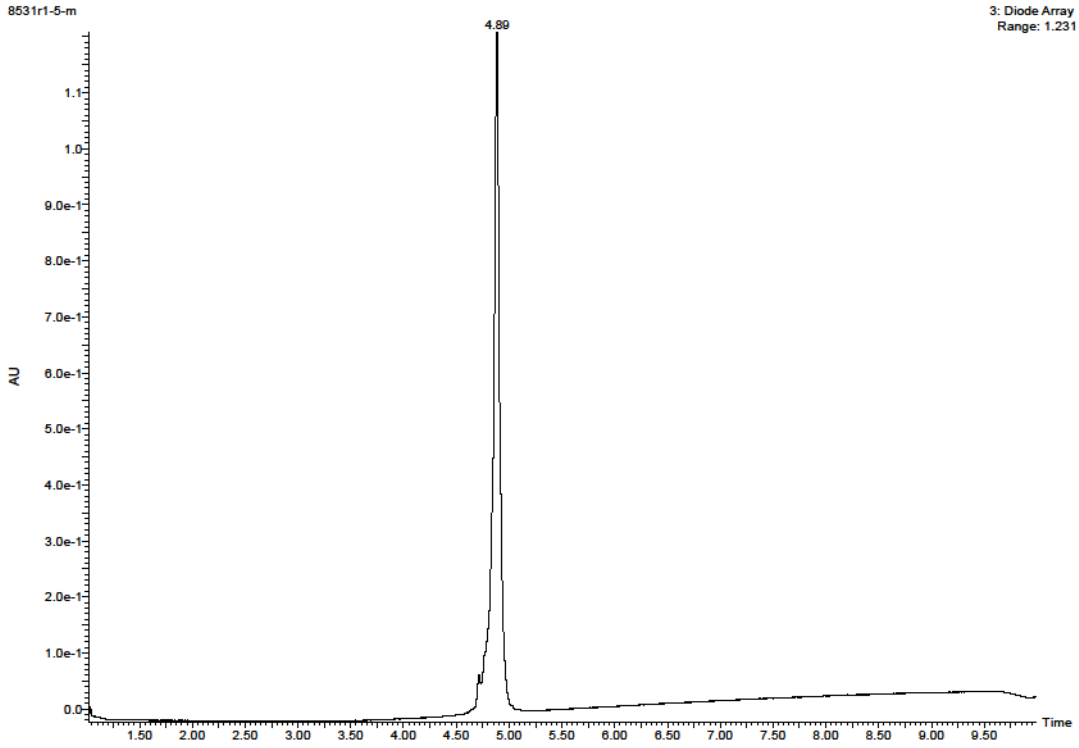



8533r1-4 228 (5.013) M1 [Ev-346356,It30] (Gs,0.500,603:2662,1.00,L55,R55); Cm (228:233)
1: TOF MS ES-
1.16e6
9992.0000
100
%
0
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500
11000
11500
mass


8530r1-3-m
3: Diode Array
Range: 9.996e-1
4.93
AU
Time


8530r1-3-m 228 (5.013) M1 [Ev-380152,It28] (Gs,0.500,520:2592,1.00,L55,R55); Cm (223:234)
1: TOF MS ES-
1.38e5
8613.0000
8649.0000
%
mass

BS2 Wyosine HPLC trace and mass spectrum



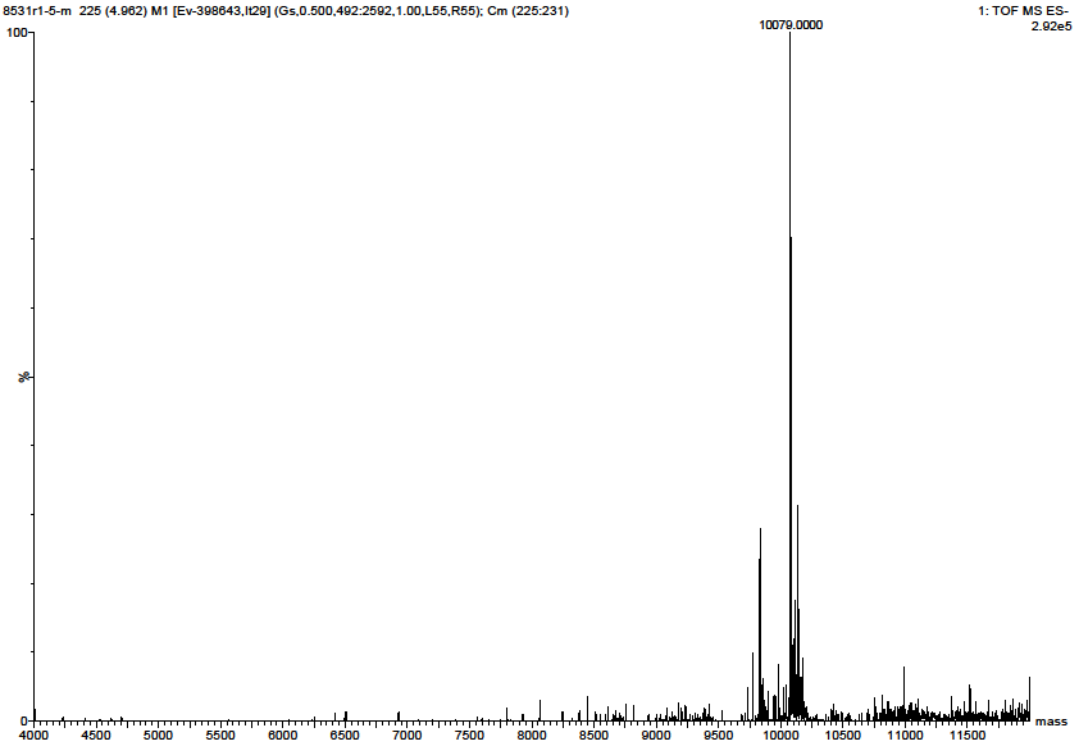

8531r1-5-m 225 (4.962) M1 [Ev-398643,It29] (Gs,0.500,492:2592,1.00,L55,R55); Cm (225:231)
1: TOF MS ES-
2.92e5
10079.0000
100
%
0
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500
11000
11500
mass

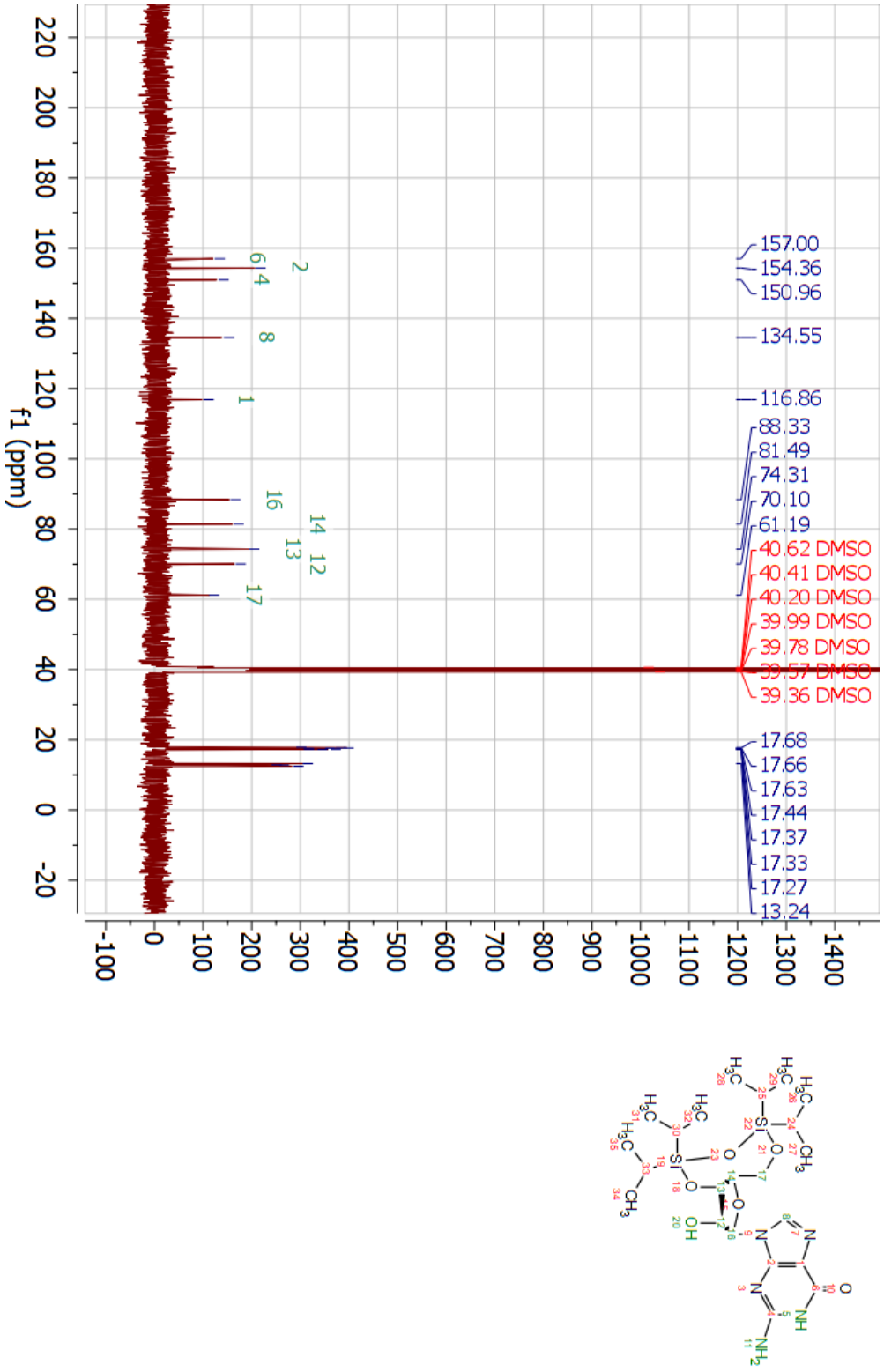
Table S1. Summary of Crystallographic Statistics

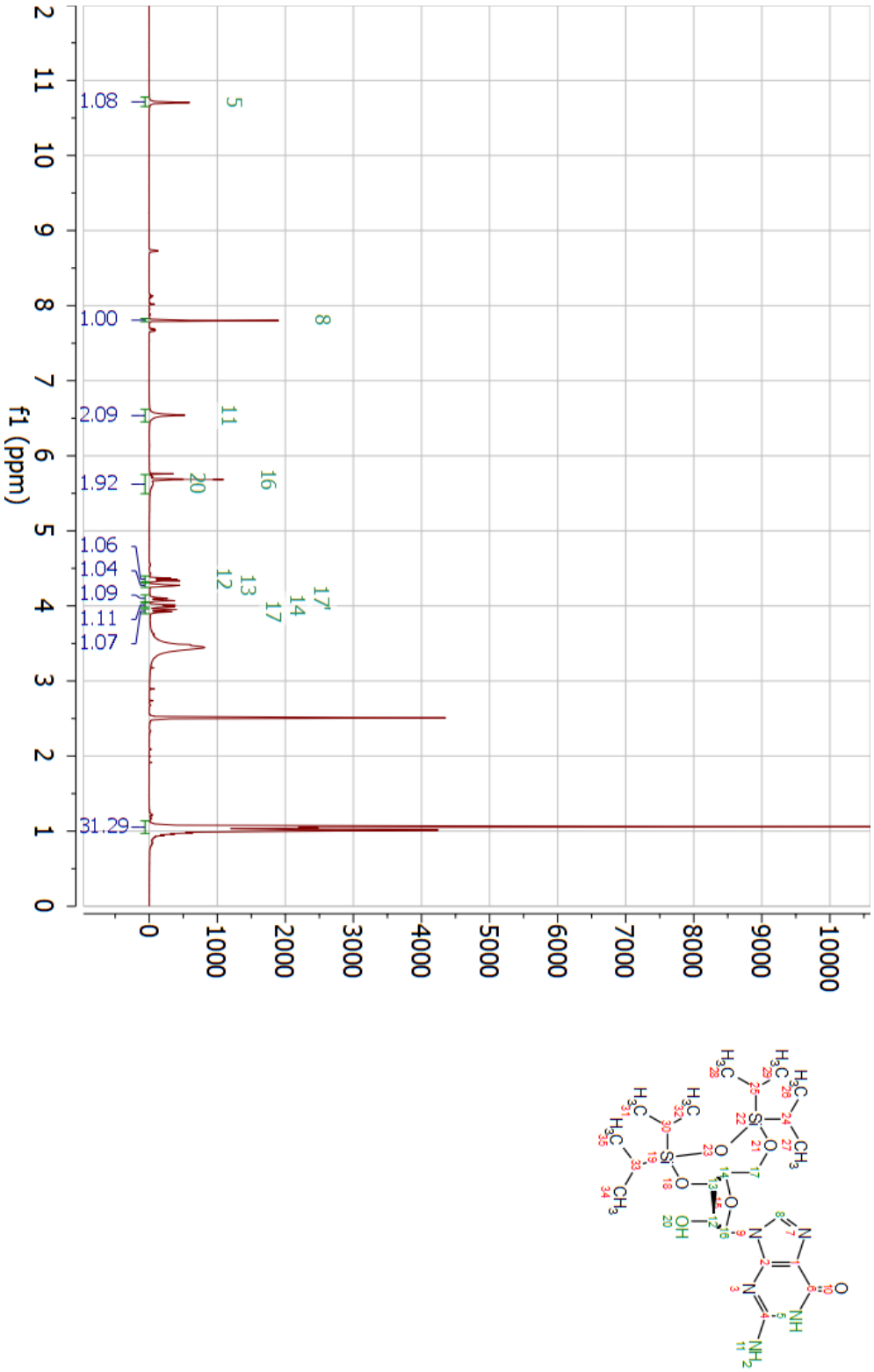
Data Collection	
Resolution (Å)	75.26 – 2.01 (2.07 – 2.01)[a]
Space group	C121
Unit Cell Dimensions	150.56, 48.69, 30.61, 90.00, 91.36, 90.00
Multiplicity	5.3 (5)
Completeness (%)	98.7 (92.1)
Mean (I) / σ (I)	13.5 (1.1)
R_{merge}	0.054
$CC_{1/2}$	0.999 (0.627)
Refinement	
No. Reflections	11686 (1167)
R_{work} / R_{free}	0.1959 / 0.2324

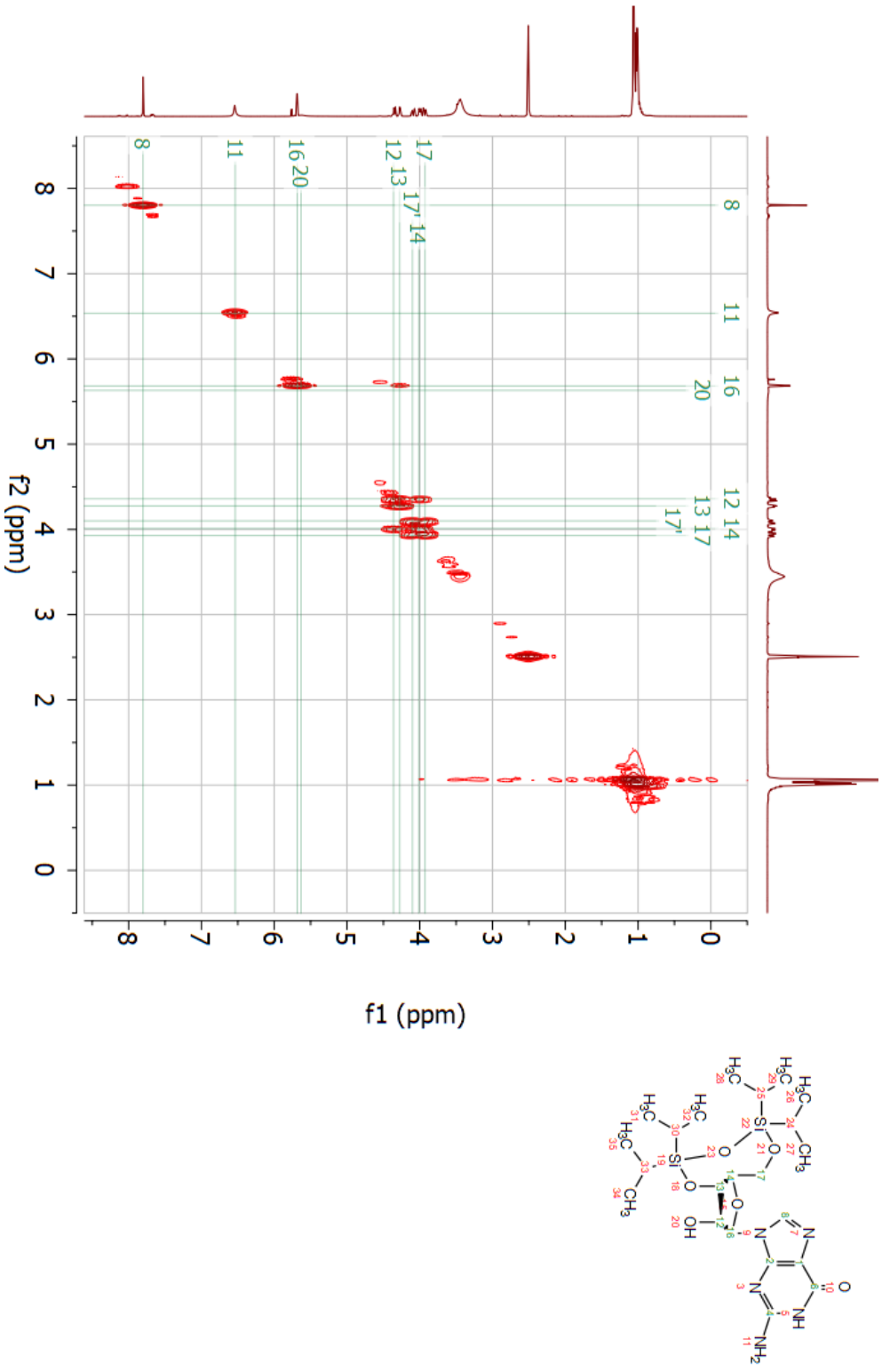
[a]Values in parentheses are for highest resolution shell.

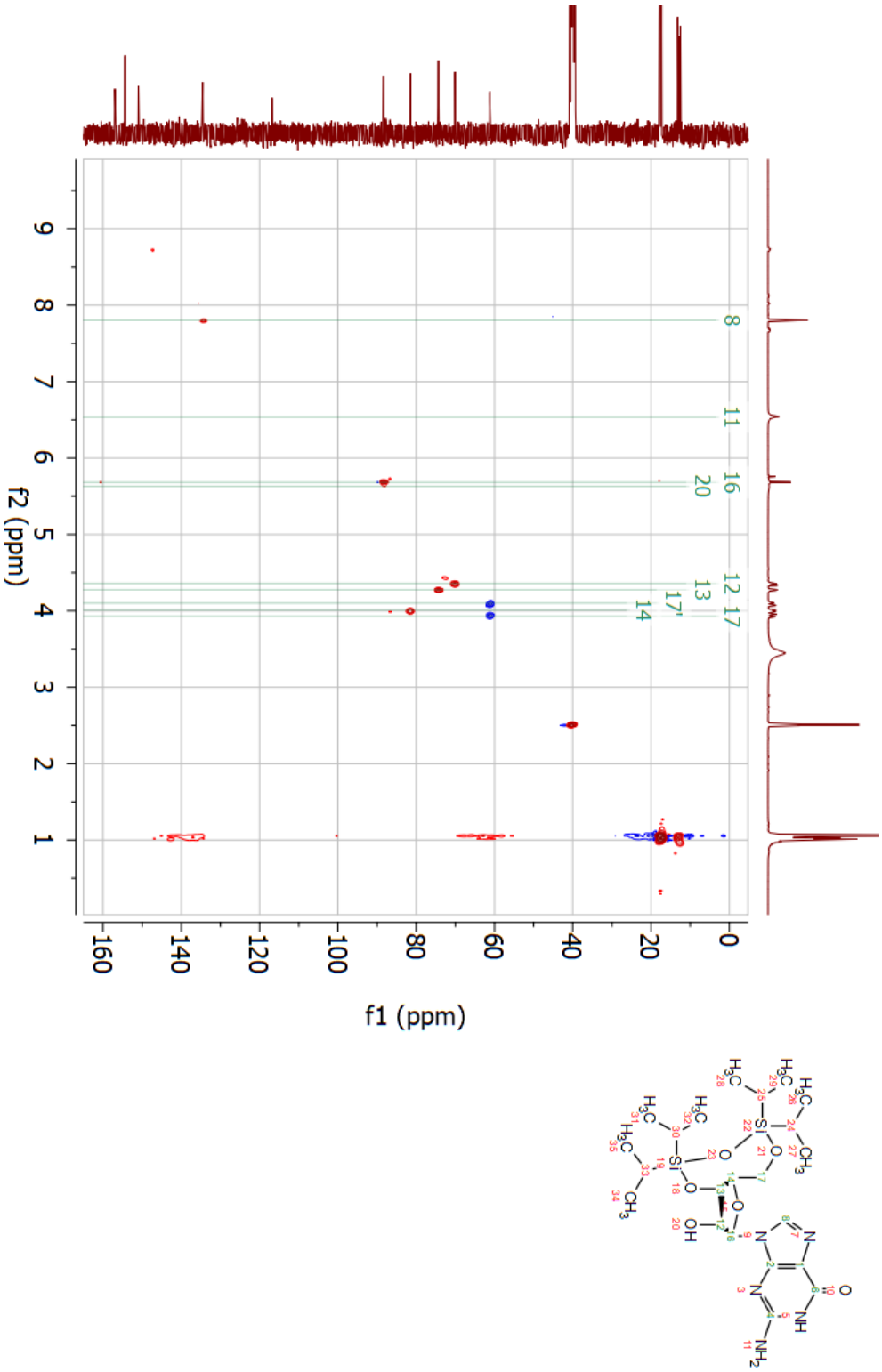
Table S2. RNA Sequences

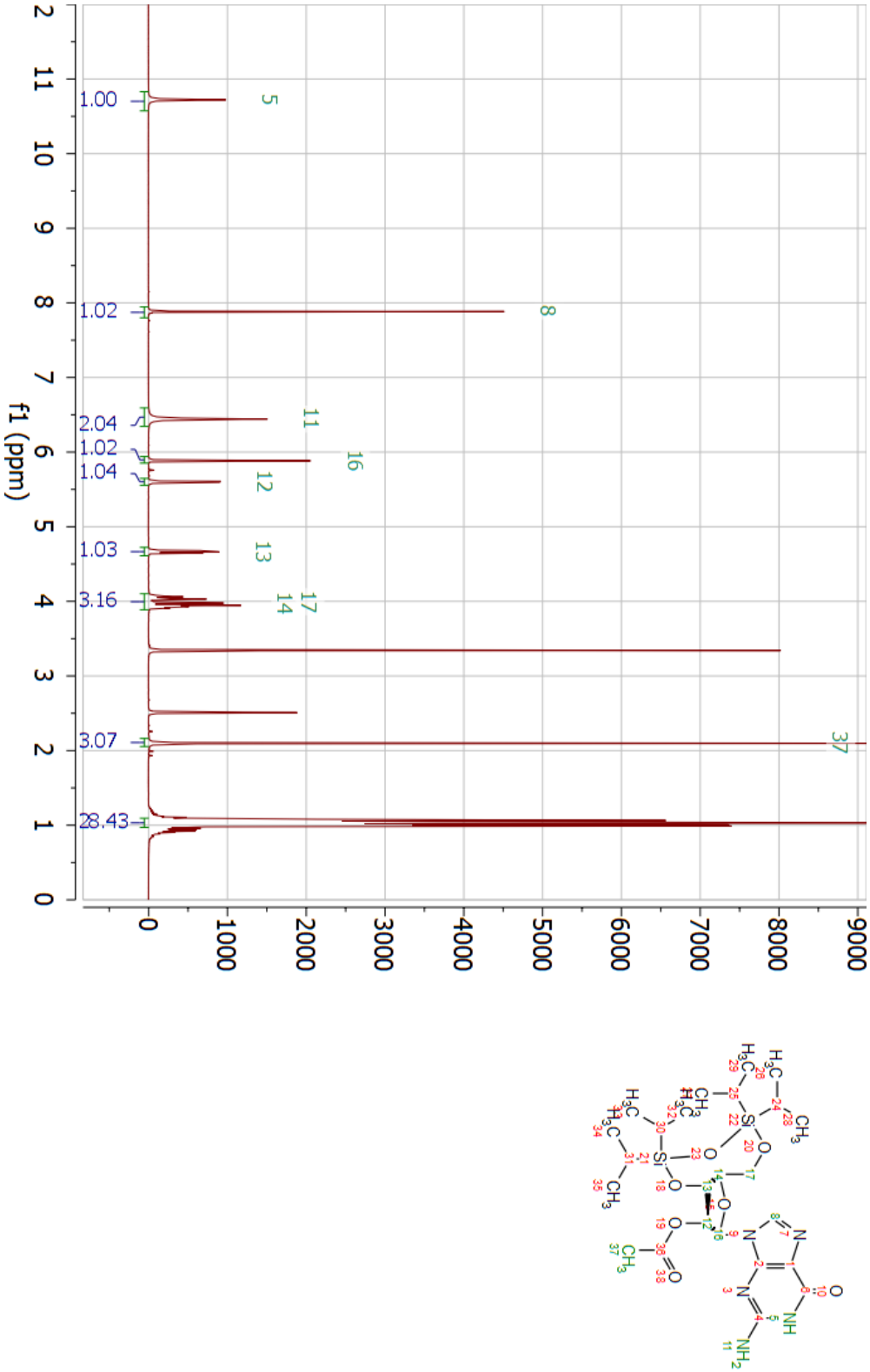
NAME	SEQUENCE
RNA TERRA	UAGGGUUAGGGU
RNA TERRA AP	U(AP)GGGUUAGGGU
RNA TERRA PC	U(PyC)GGGUUAGGGU
MANGO-III	GCUACGAAGGAAGGAUUGGUAUGUGGUAUAUUCGUAGC
MANGO-III U17AP	GCUACGAAGGAAGGAU(AP)GGUAUGUGGUAUAUUCGUAGC
MANGO-III U17PC	GCUACGAAGGAAGGAU(PyC)GGUAUGUGGUAUAUUCGUAGC
MANGO-III A15PC	GCUACGAAGGAAGG(PyC)UUGGUAGGUGGUAUAUUCGUAGC
MANGO-III A15AP	GCUACGAAGGAAGG(AP)UUGGUAUG
IMANGO-III	GUGUACGAAGGUAGGUUUGGUAUGGGGUAGUUGUCGUACAC
IMANGO-III U12(AMINO U)	GUGUACGAAGG(AminoU)AGGUUUGGUAUGGGGUAGUUGUCGUACAC
IMANGO-III U17(AMINO U)	GUGUACGAAGGAAGGU(AminoU)UGGUAUGGGGUAGUUGUCGUACAC
IMANGO-III U18(AMINO U)	GUGUACGAAGGAAGGUU(AminoU)GGUAUGGGGUAGUUGUCGUACAC
IMANGO-III U21(AMINO U)	GUGUACGAAGGAAGGUUUGG(AminoU)AUGGGGUAGUUGUCGUACAC
IMANGO-III U28(AMINO U)	GUGUACGAAGGAAGGUUUGGUAUGGGG(AminoU)AGUUGUCGUACAC

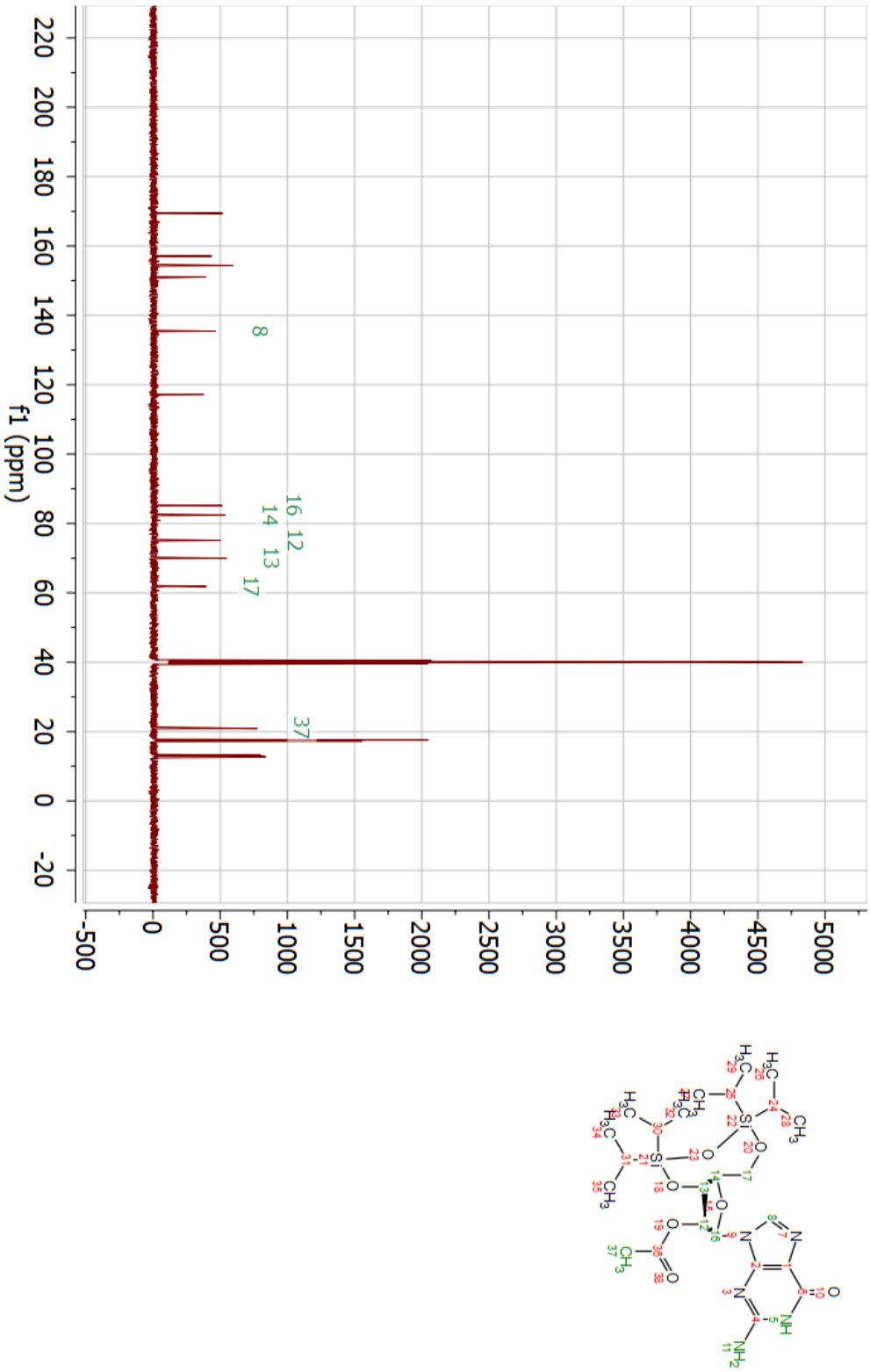

f1 (ppm)
157.00
154.36
150.96
134.55
116.86
88.33
81.49
74.31
70.10
61.19
40.62 DMSO
40.41 DMSO
40.20 DMSO
39.99 DMSO
39.78 DMSO
39.57 DMSO
39.36 DMSO
17.68
17.66
17.63
17.44
17.37
17.33
17.27
13.24

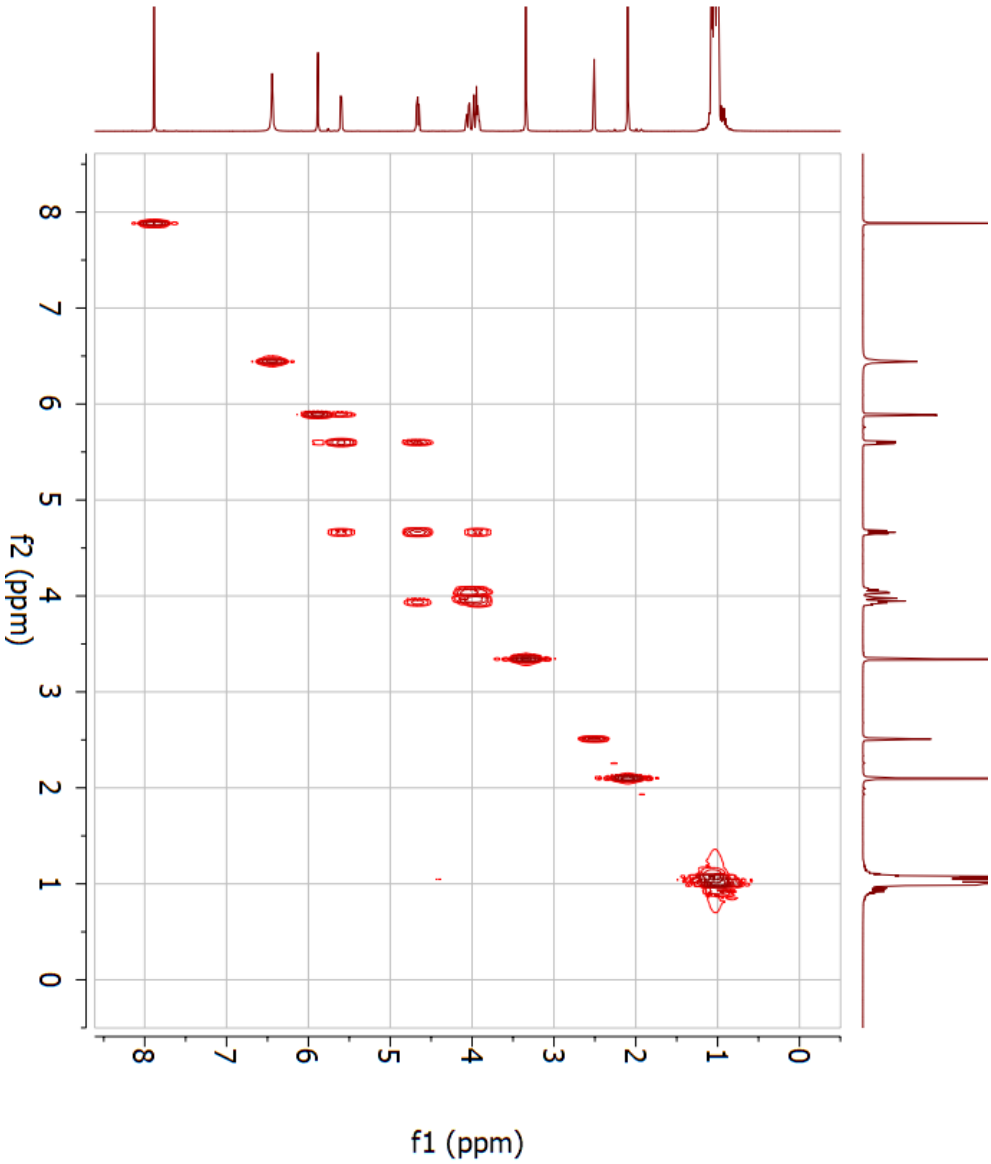

f1 (ppm)
1.08
1.00
2.09
1.92
1.06
1.04
1.09
1.11
1.07
31.29
5
8
11
16
20
12
13
17
14
17'
NH
NH2
OH
O
Si
CH3

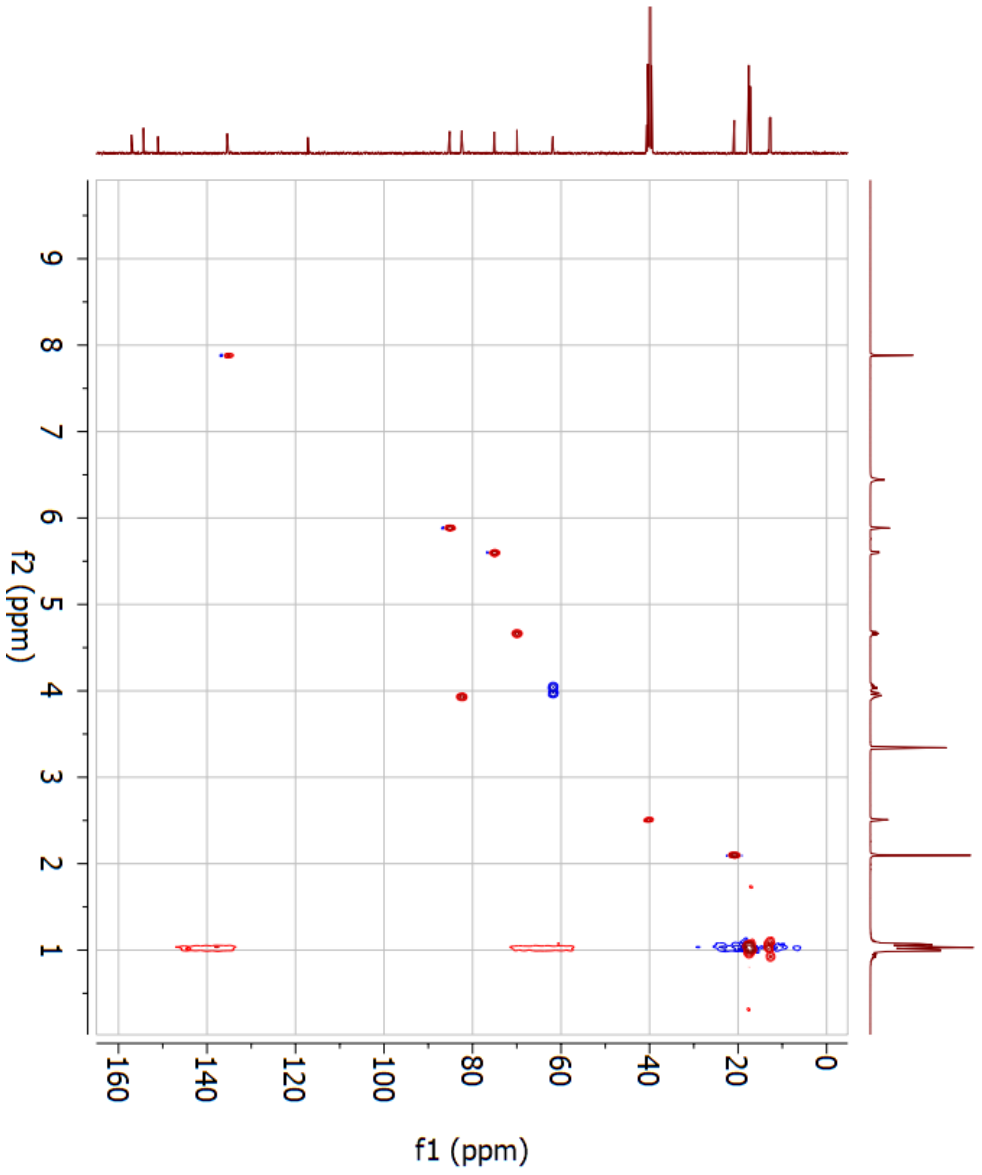

f2 (ppm)
f1 (ppm)
8
11
16 20
12 13 17' 14
17
8
11
16
20
12 14
13 17
17'

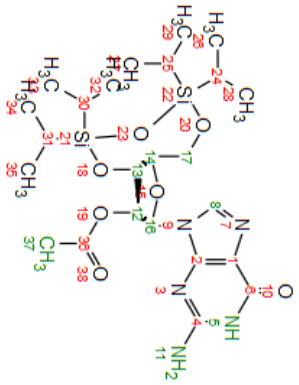

f2 (ppm)
f1 (ppm)
8
11
16
20
12
13
17
17'
14

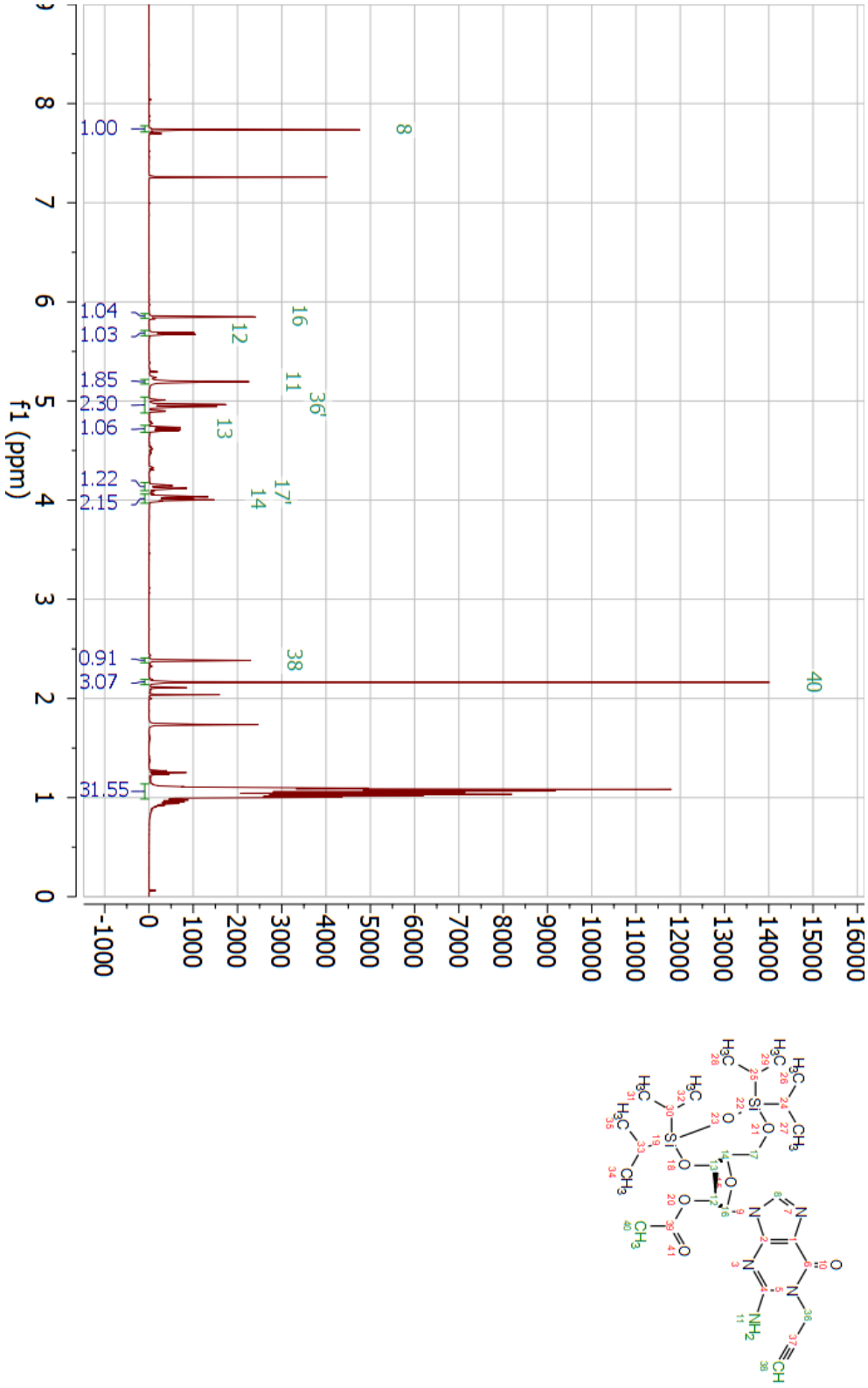

f1 (ppm)
1.00
1.02
2.04
1.02
1.04
1.03
3.16
3.07
28.43
5
8
11
16
12
13
17
14
37

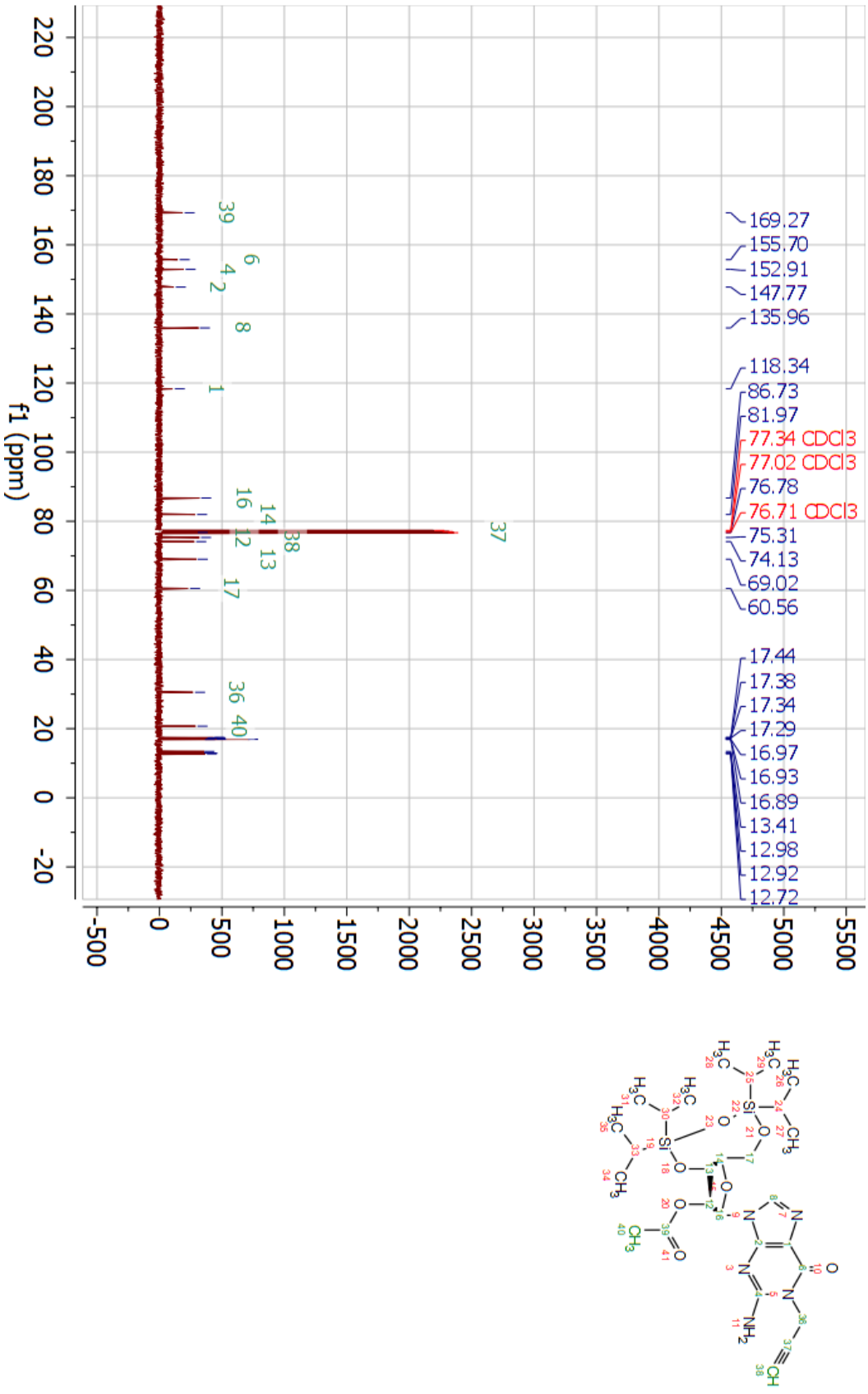

f1 (ppm)
220
200
180
160
140
120
100
80
60
40
20
0
-20
-500
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
8
16
12
14
13
17
37

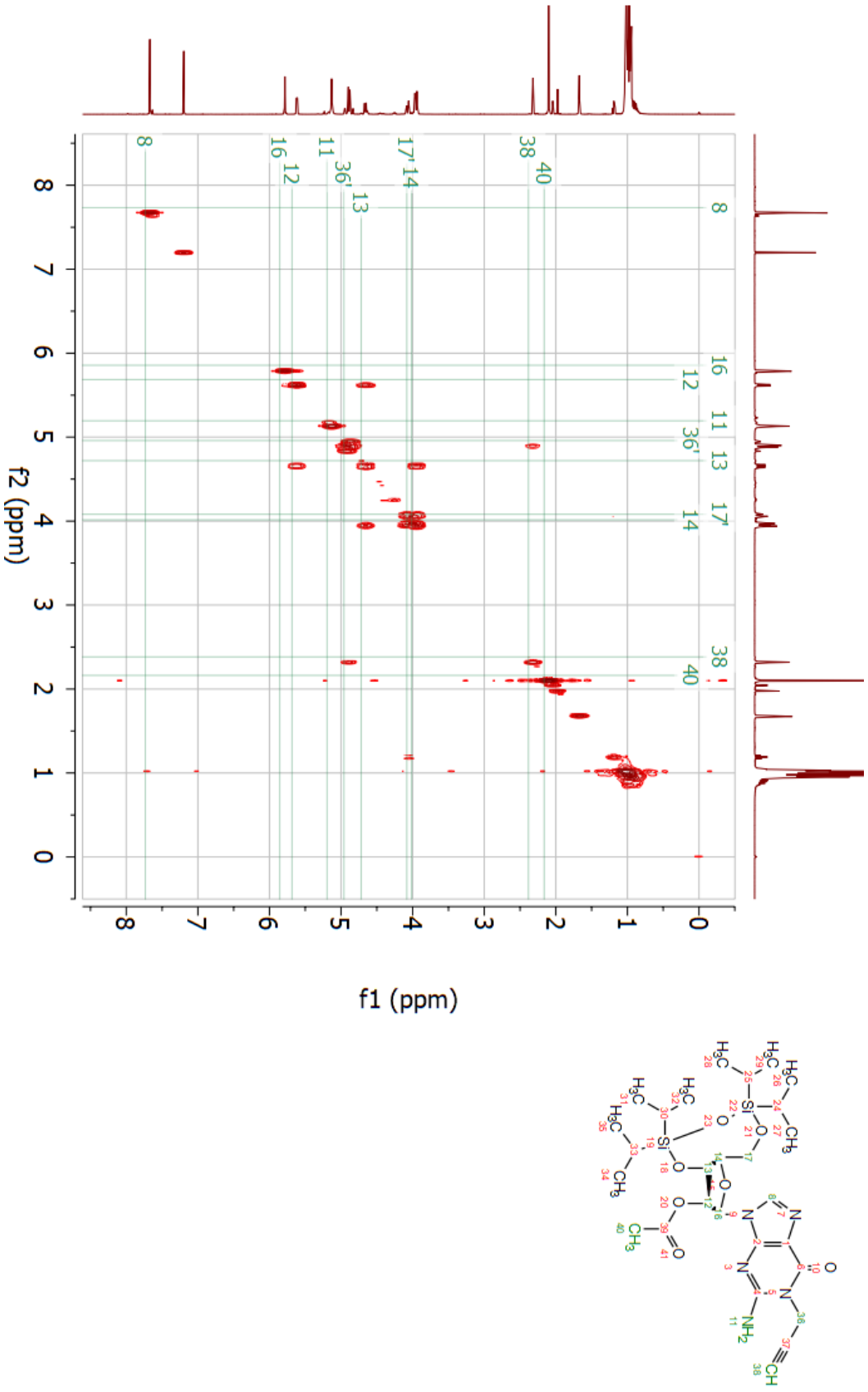

f2 (ppm)
f1 (ppm)
0
1
2
3
4
5
6
7
8

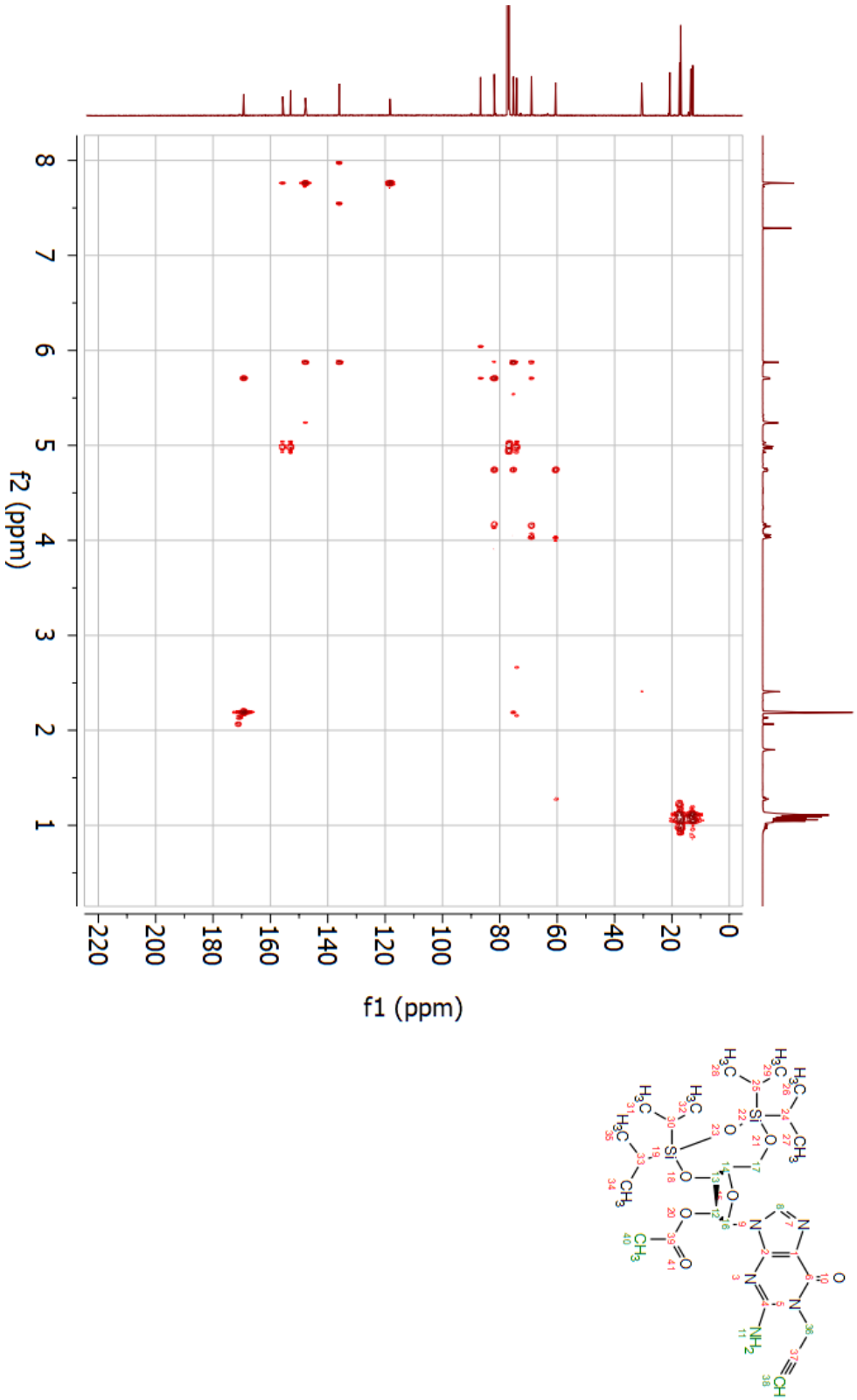

f2 (ppm)
f1 (ppm)

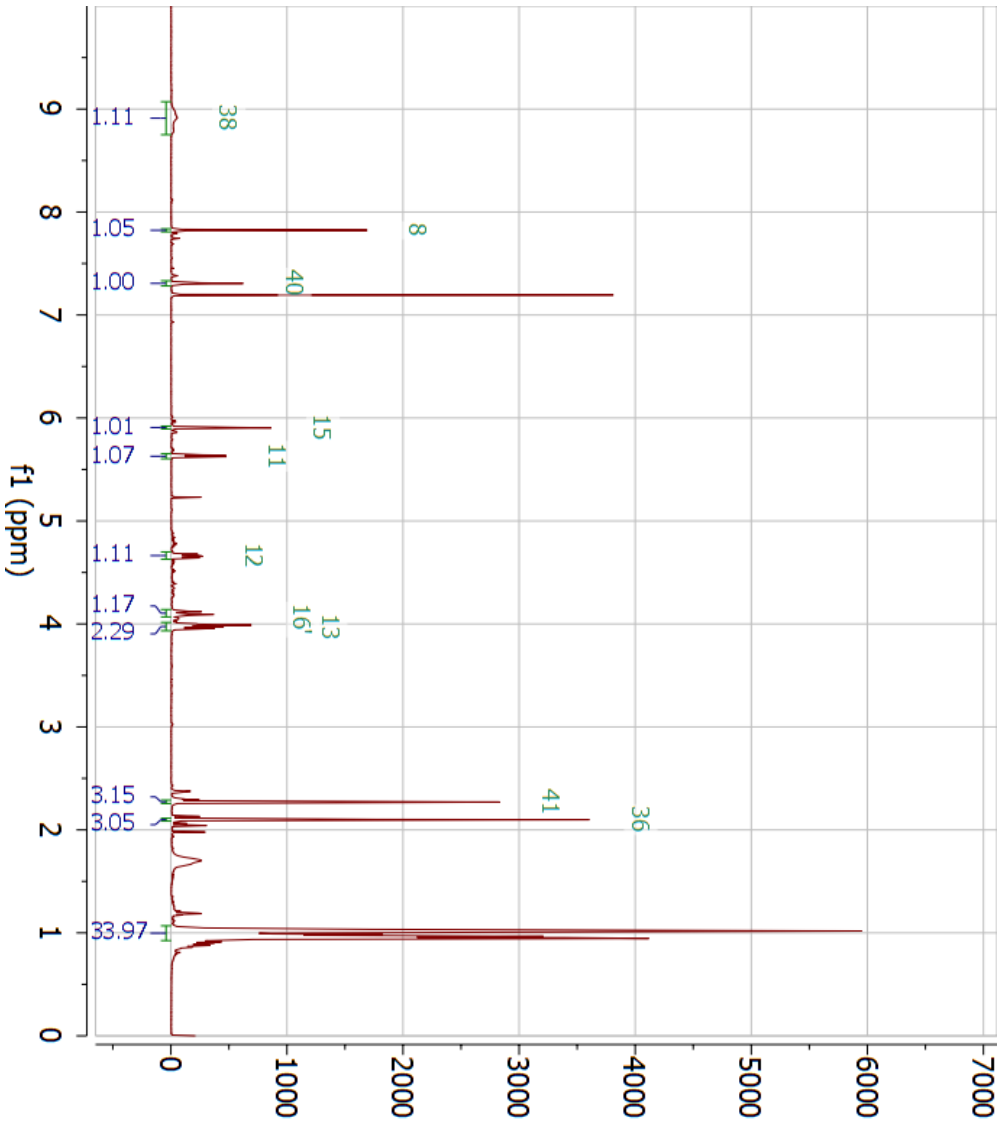


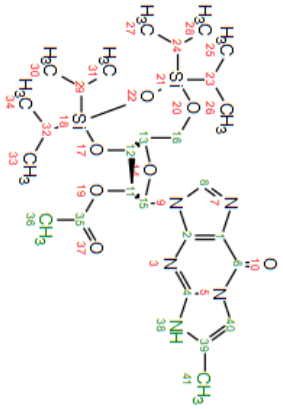

f1 (ppm)
1.00
1.04
1.03
1.85
2.30
1.06
1.22
2.15
0.91
3.07
31.55
8
16
12
11
36'
13
17'
14
38
40

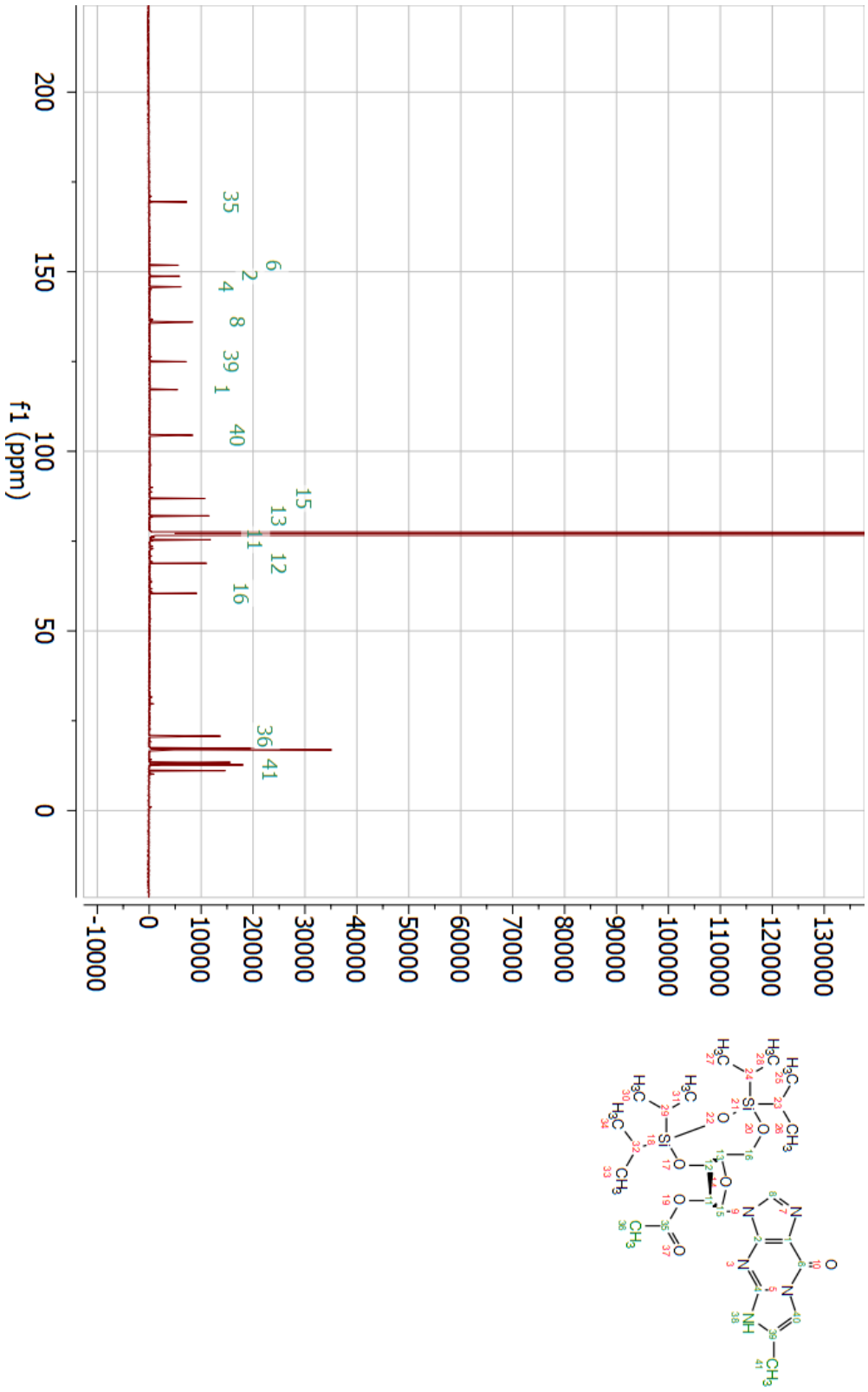

169.27
155.70
152.91
147.77
135.96
118.34
86.73
81.97
77.34 CDCl3
77.02 CDCl3
76.78
76.71 CDCl3
75.31
74.13
69.02
60.56
17.44
17.38
17.34
17.29
16.97
16.93
16.89
13.41
12.98
12.92
12.72
f1 (ppm)

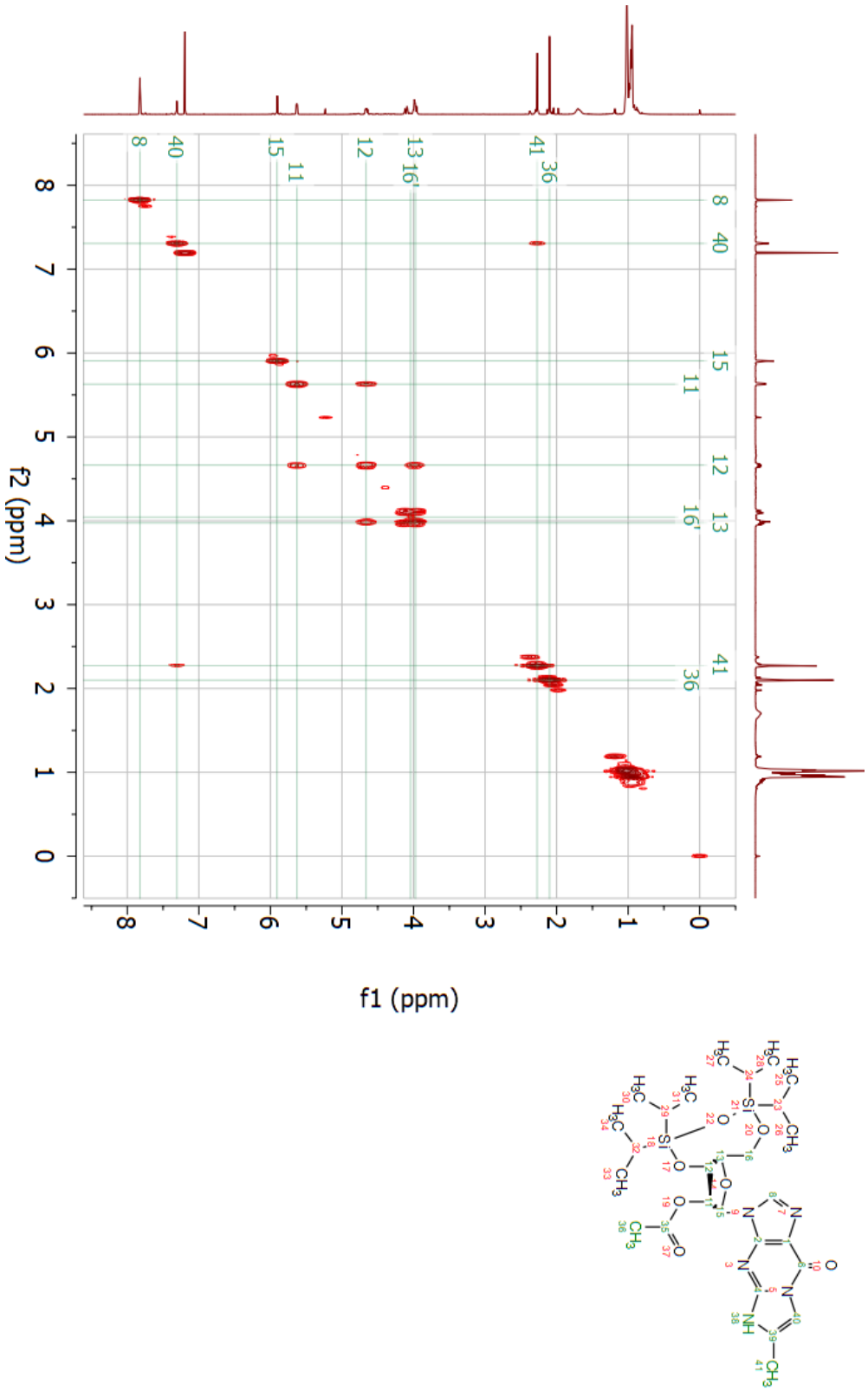

f2 (ppm)
f1 (ppm)
8
16
12
11
36'
13
17'
14
38
40

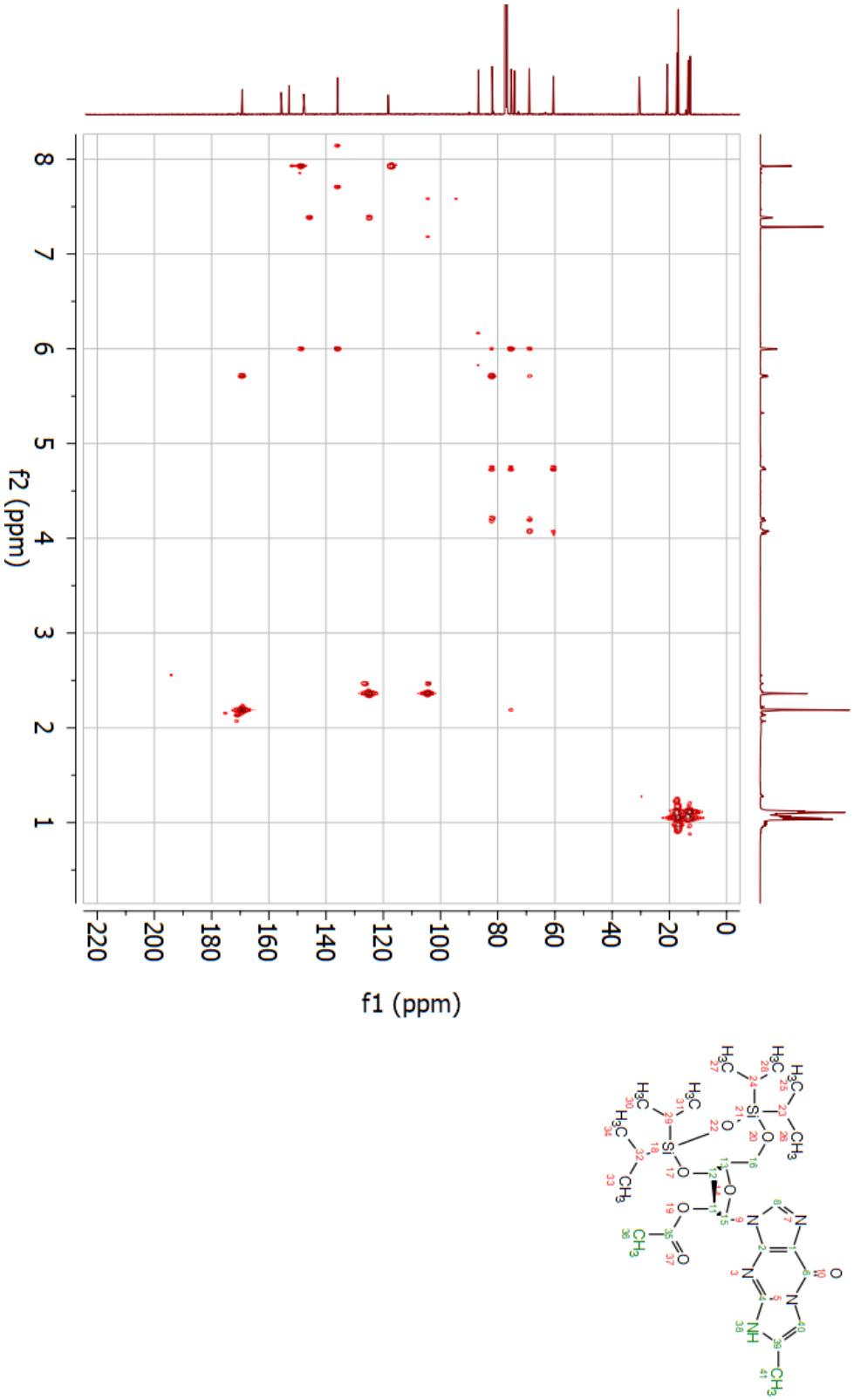

f2 (ppm)
f1 (ppm)

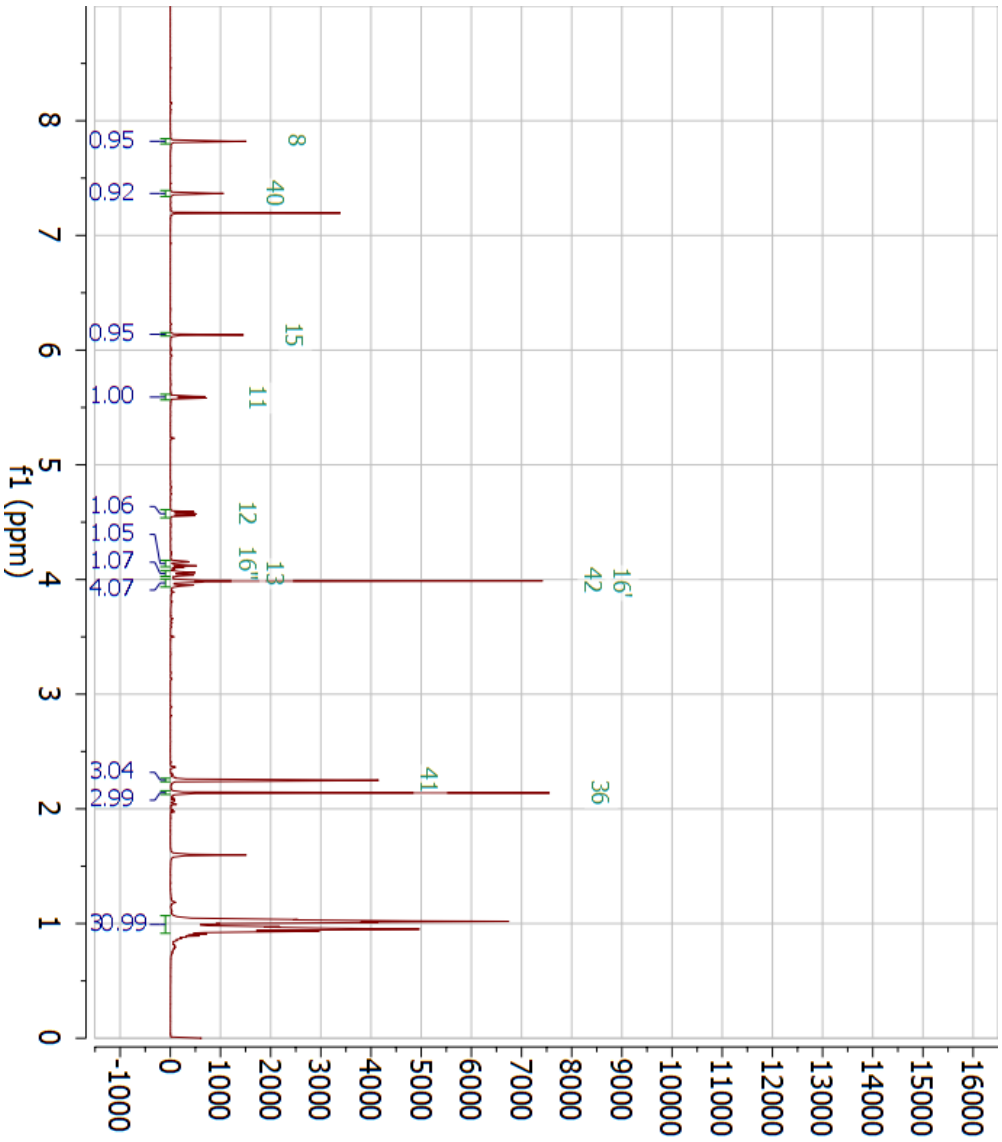

f1 (ppm)
0
1
2
3
4
5
6
7
8
9
0
1000
2000
3000
4000
5000
6000
7000
1.11
1.05
1.00
1.01
1.07
1.11
1.17
2.29
3.15
3.05
33.97
38
8
40
15
11
12
13
16'
41
36

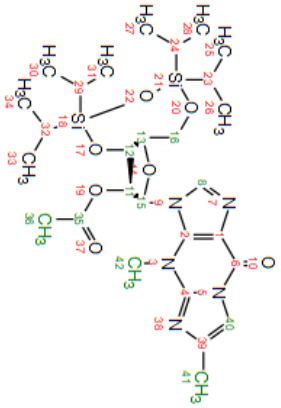


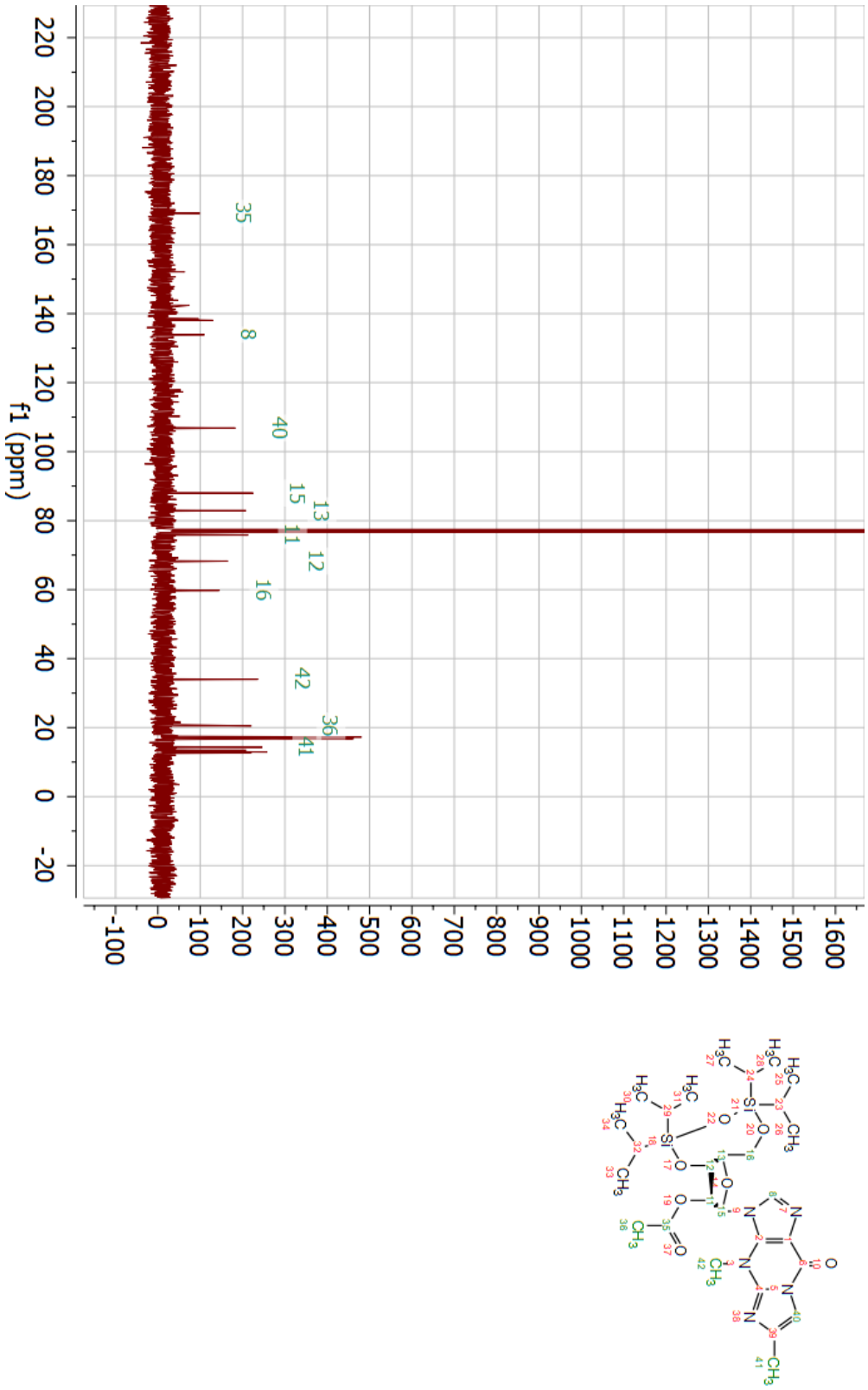

f1 (ppm)
200
150
100
50
0
-10000
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
110000
120000
130000
35
6
2
4
8
39
1
40
15
13
11
12
16
36
41

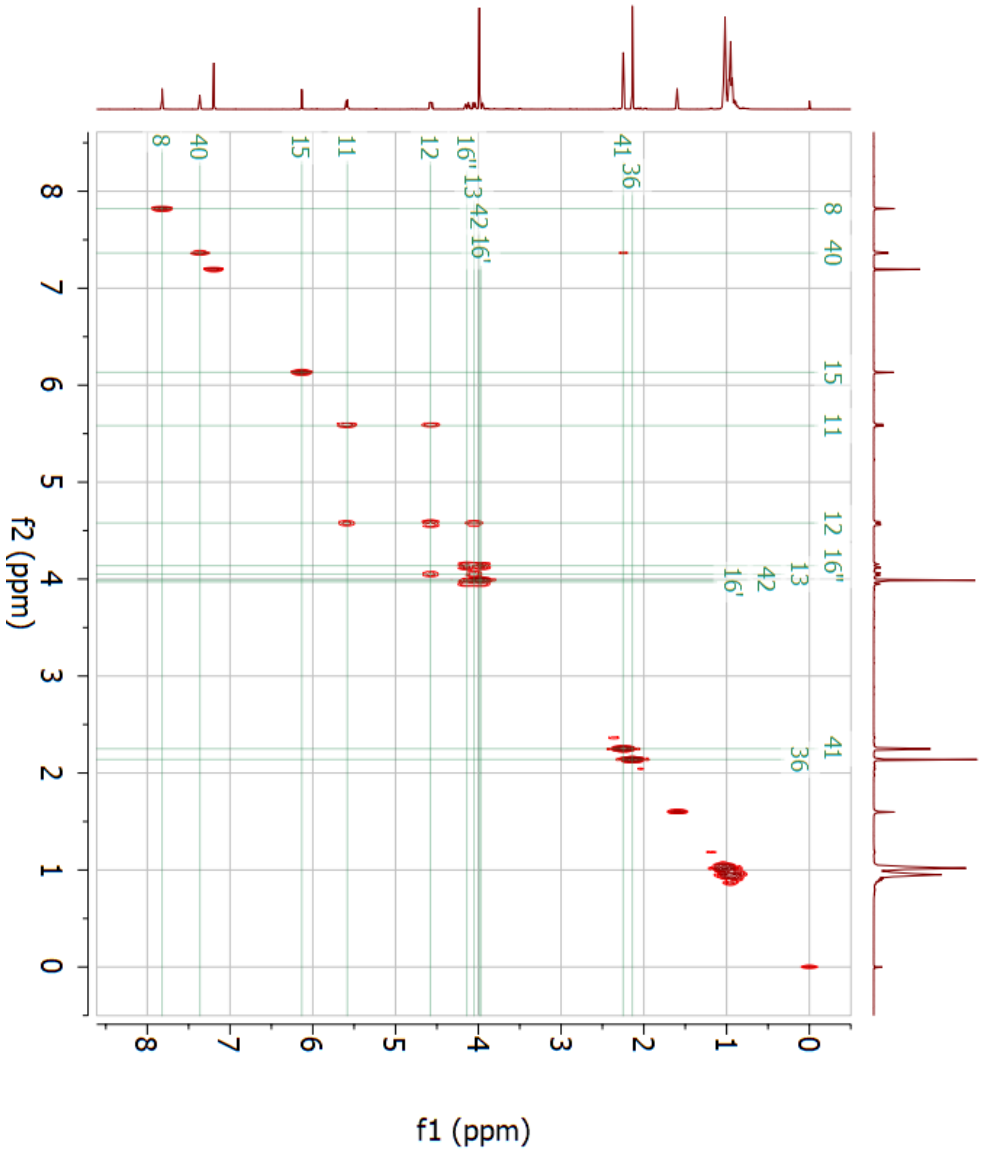

f2 (ppm)
f1 (ppm)
8
40
15
11
12
13
16'
41
36

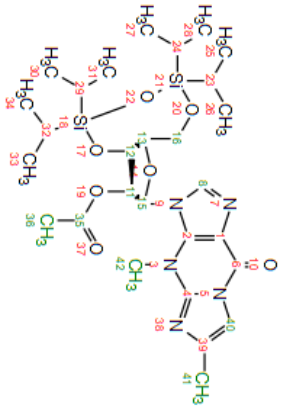

f2 (ppm)
f1 (ppm)

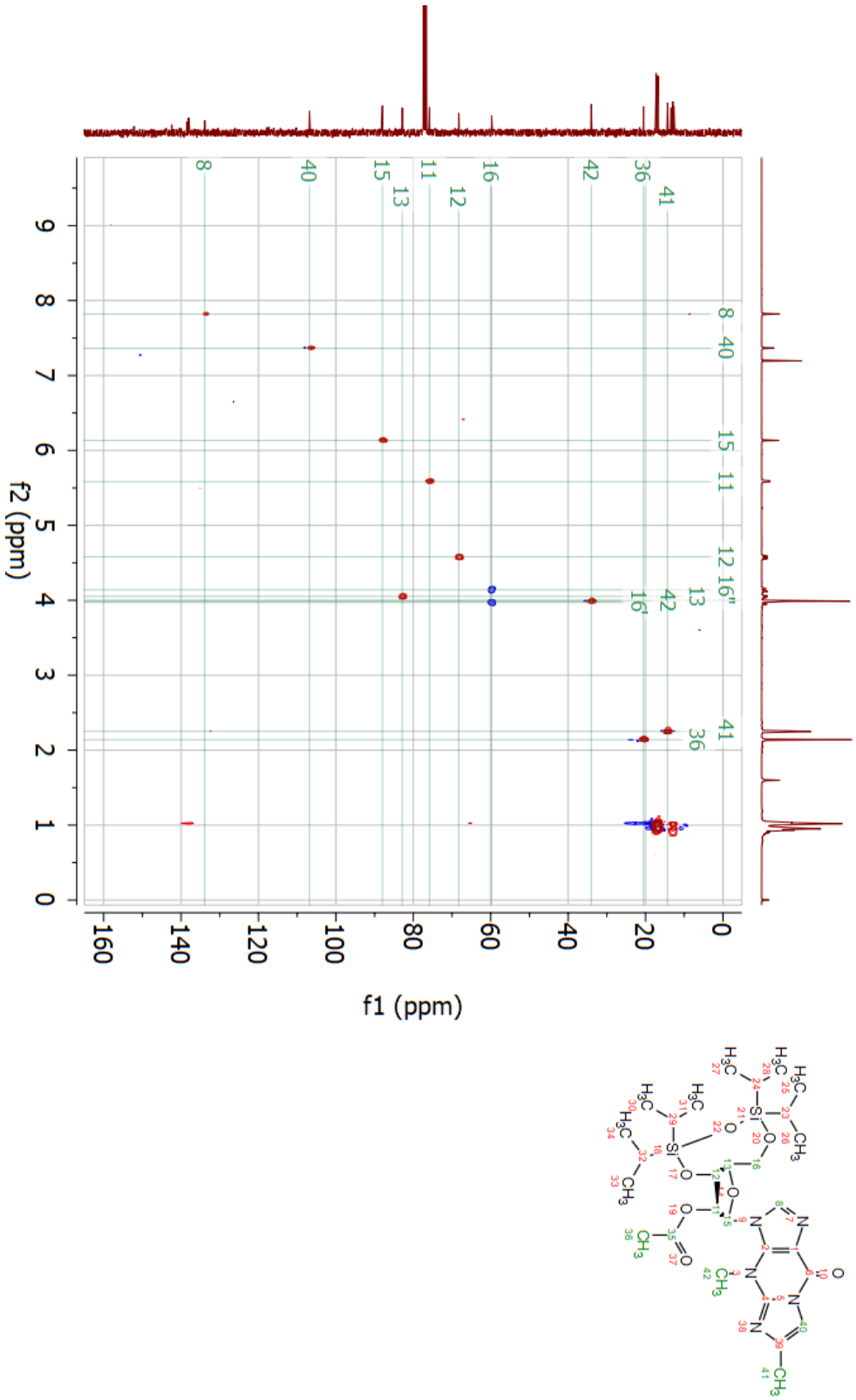

f1 (ppm)
0.95
0.92
0.95
1.00
1.06
1.05
1.07
4.07
3.04
2.99
30.99
8
40
15
11
12
16''
13
16'
42
41
36

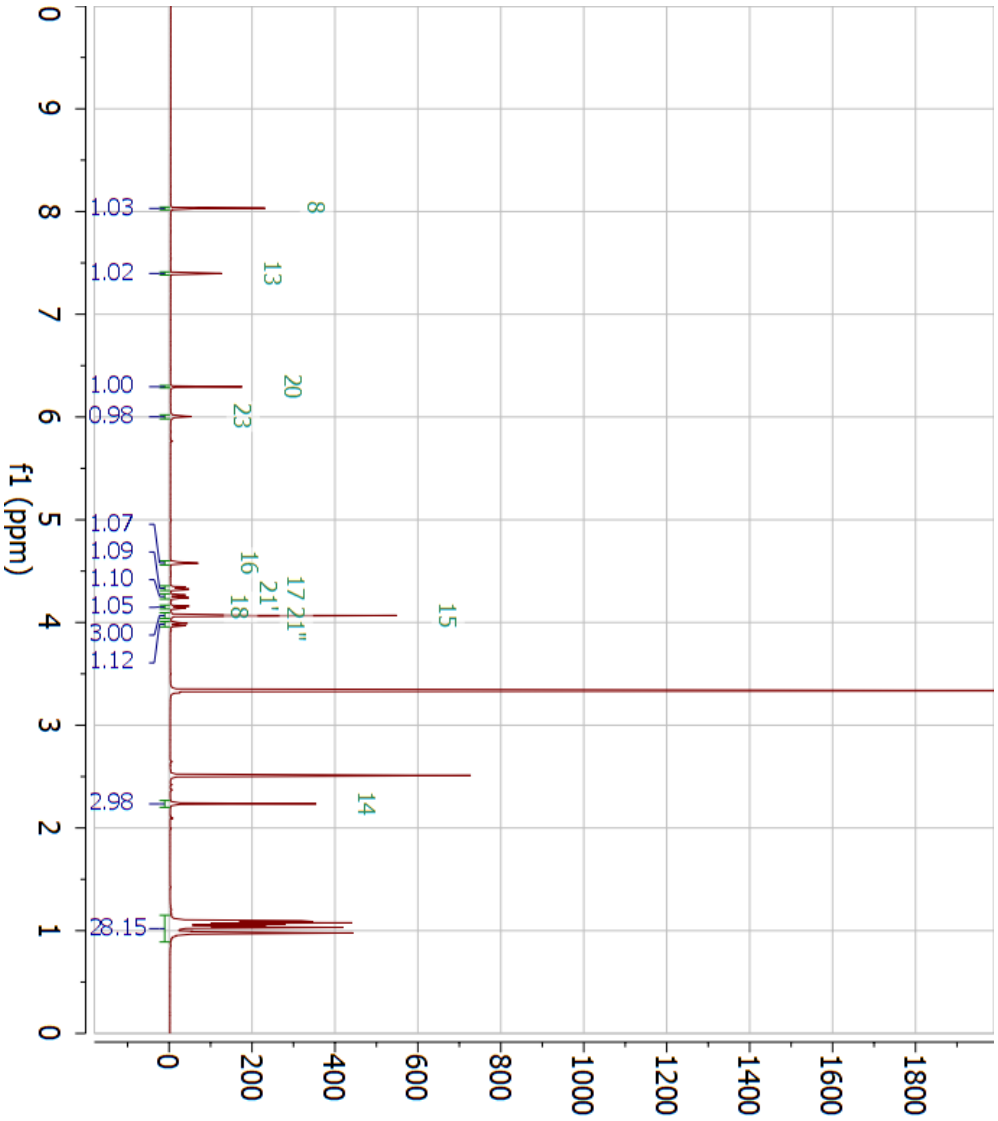


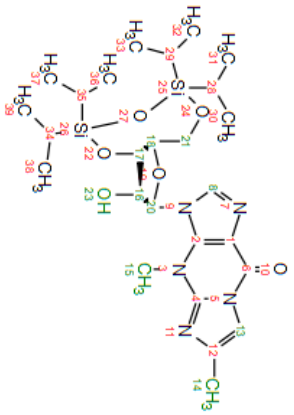

f1 (ppm)
220 200 180 160 140 120 100 80 60 40 20 0 -20
-100 0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600
35
8
40
15
13
11
12
16
42
36
41

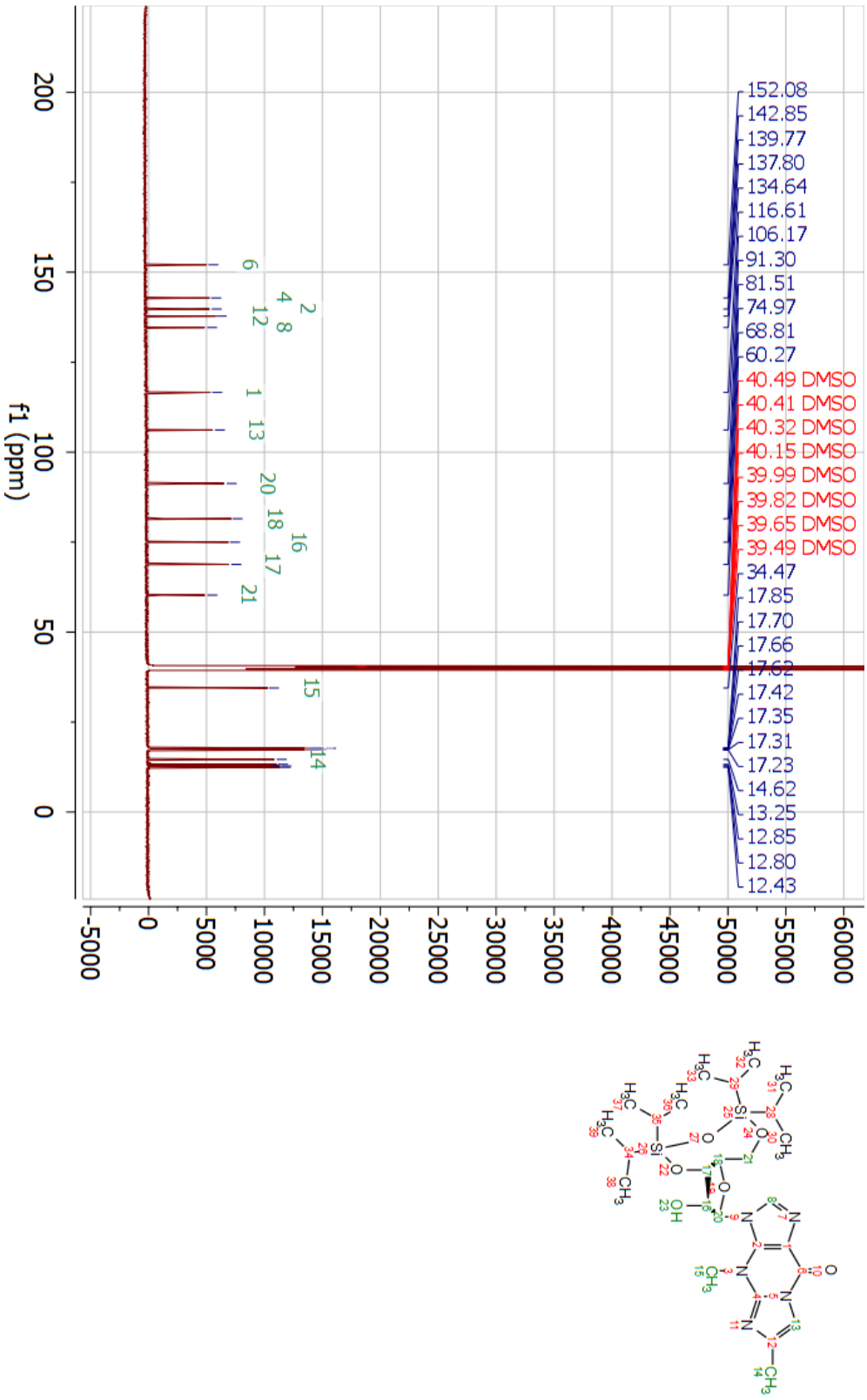

f2 (ppm)
f1 (ppm)
8
40
15
11
12
16''
13
42
16'
41
36

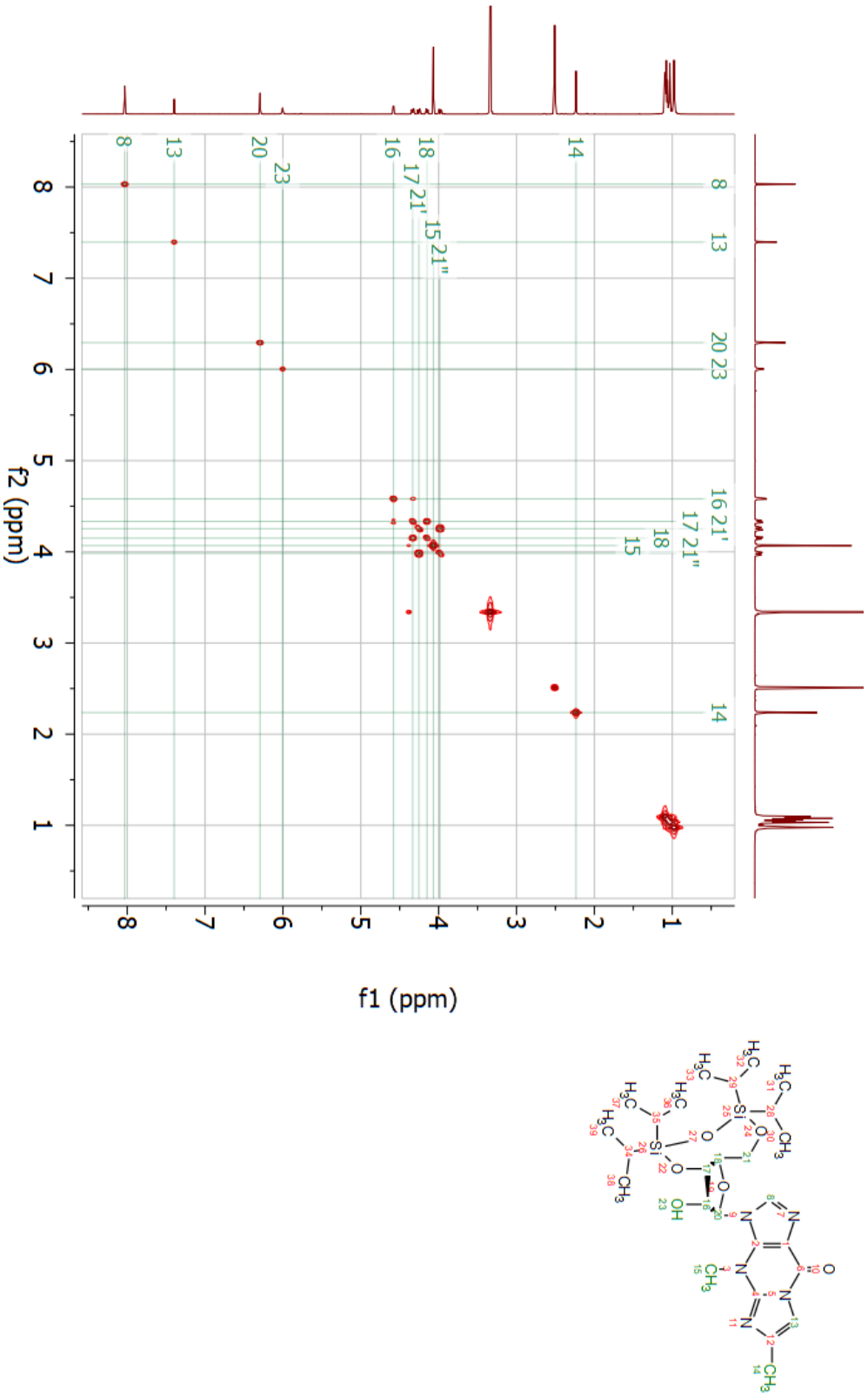


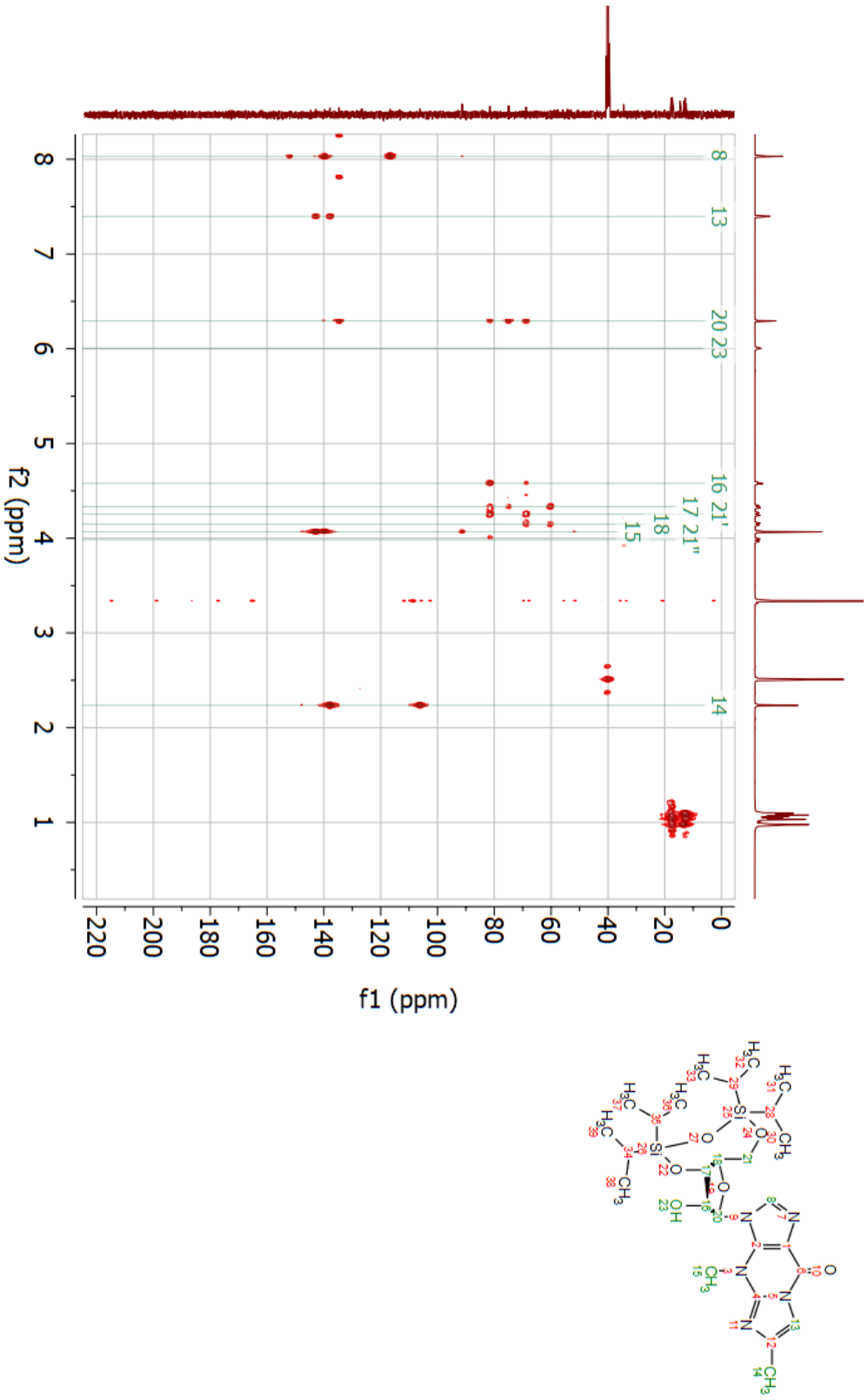

f2 (ppm)
f1 (ppm)
8
40
15
11
12
16"
13
42
16'
36
41

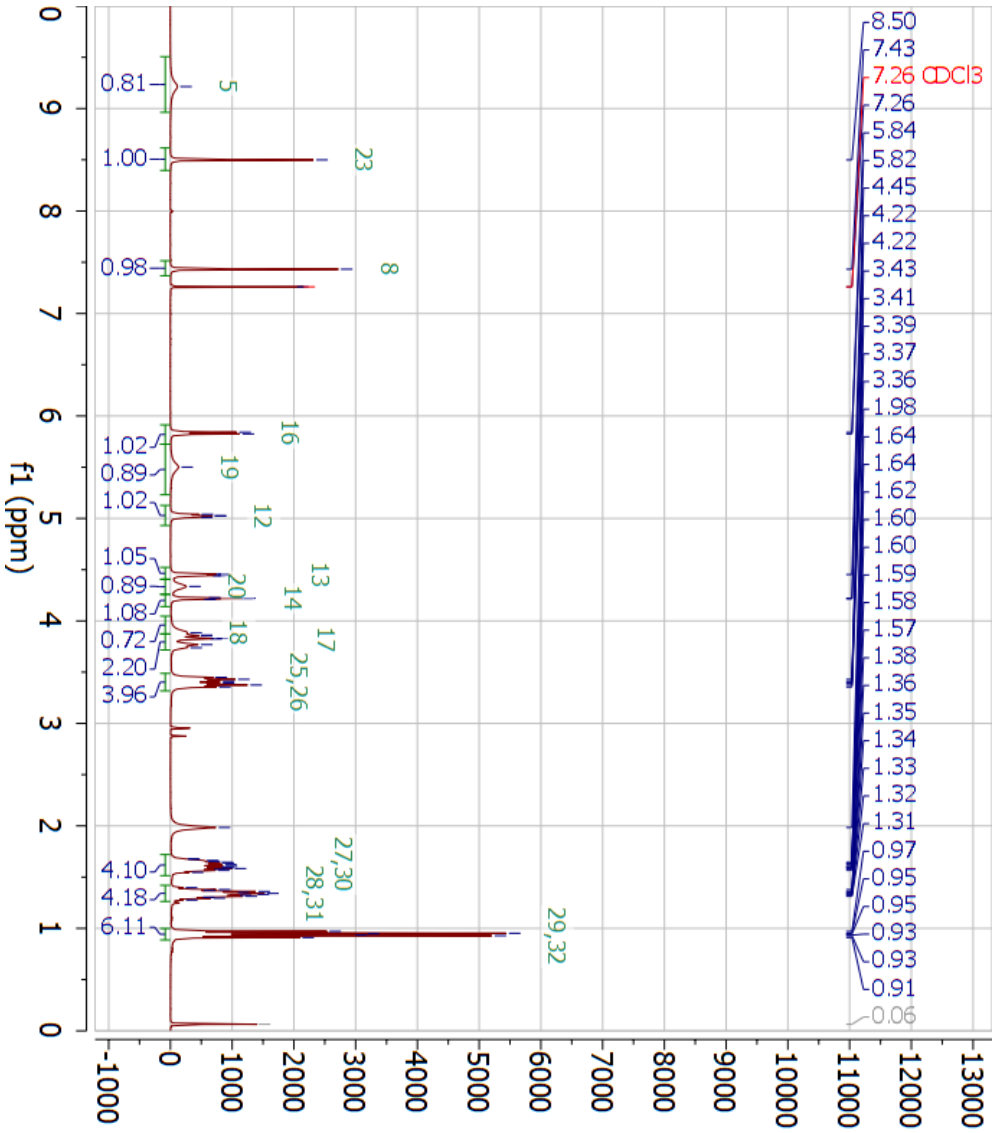

f1 (ppm)
1.03
1.02
1.00
0.98
1.07
1.09
1.10
1.05
3.00
1.12
2.98
28.15
8
13
20
23
16
18
21'
17
21"
15
14

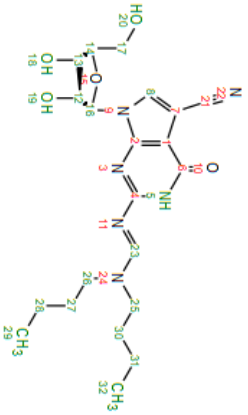


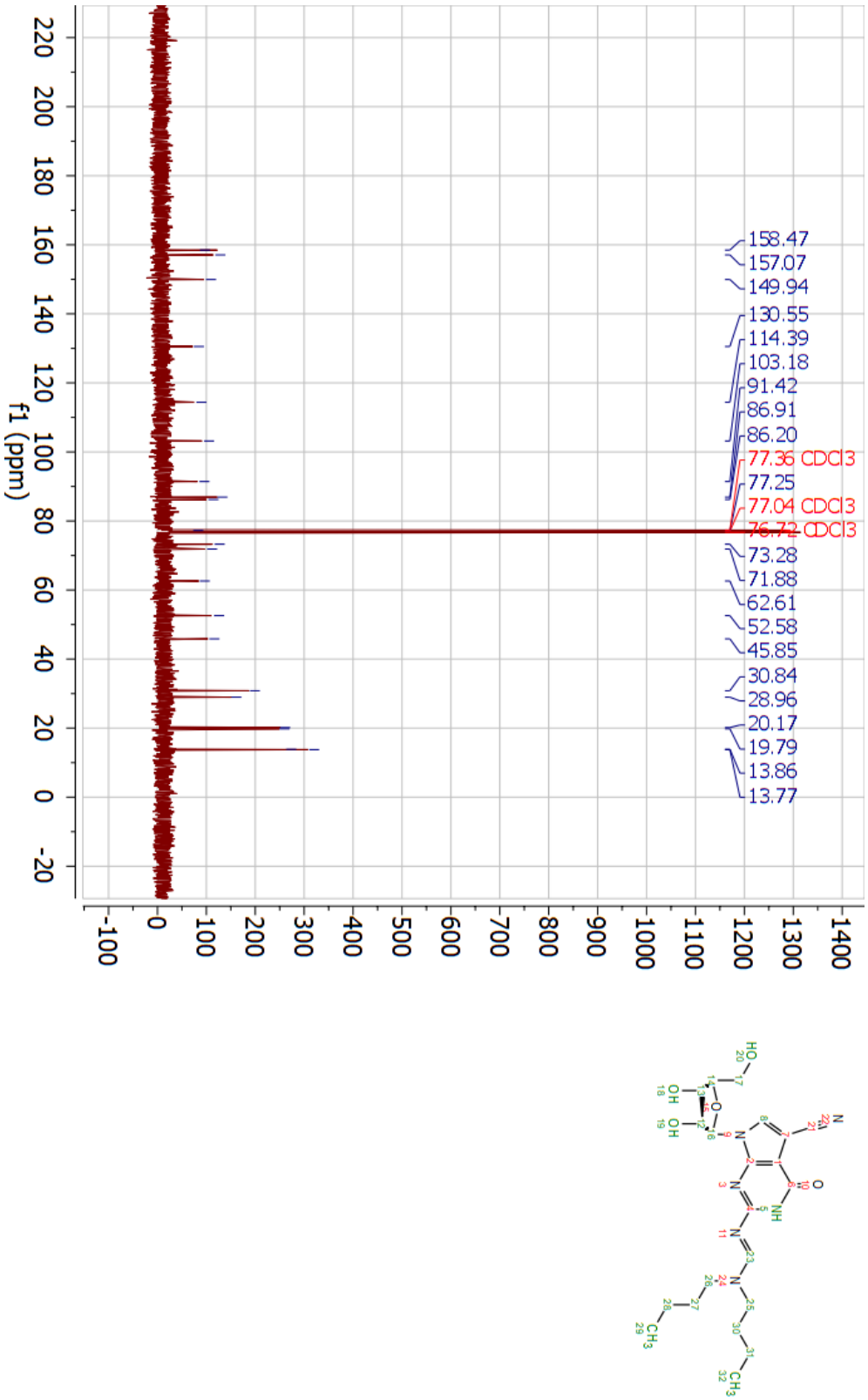

f1 (ppm)
200
150
100
50
0
-5000
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000
60000
152.08
142.85
139.77
137.80
134.64
116.61
106.17
91.30
81.51
74.97
68.81
60.27
40.49 DMSO
40.41 DMSO
40.32 DMSO
40.15 DMSO
39.99 DMSO
39.82 DMSO
39.65 DMSO
39.49 DMSO
34.47
17.85
17.70
17.66
17.62
17.42
17.35
17.31
17.23
14.62
13.25
12.85
12.80
12.43
6
2
4
8
12
1
13
20
18
16
17
21
15
14

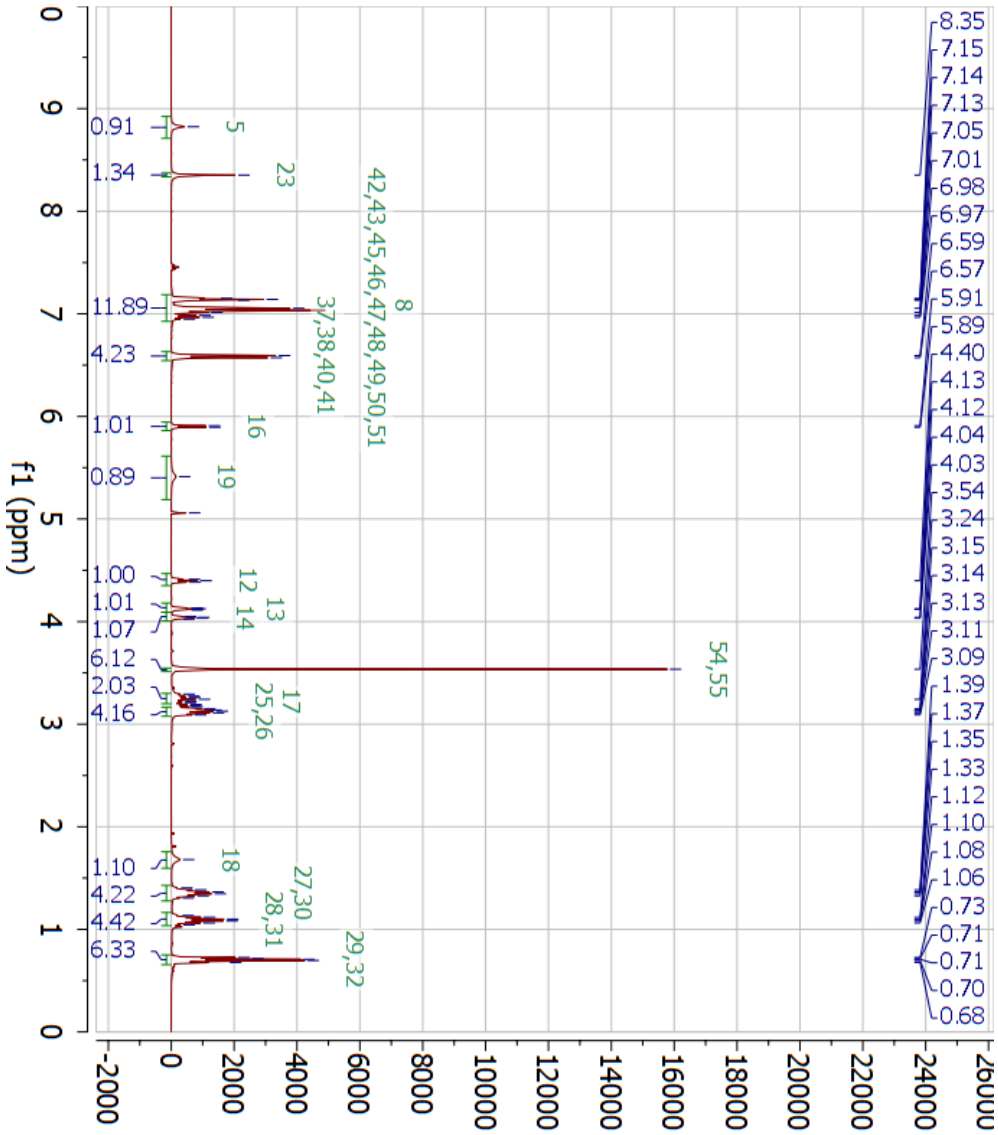

f2 (ppm)
f1 (ppm)
8
13
20 23
16 17 21' 15 21" 18
14

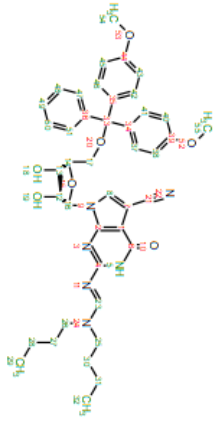

8
13
20 23
16 21'
17 21"
18
15'
14
f2 (ppm)
f1 (ppm)

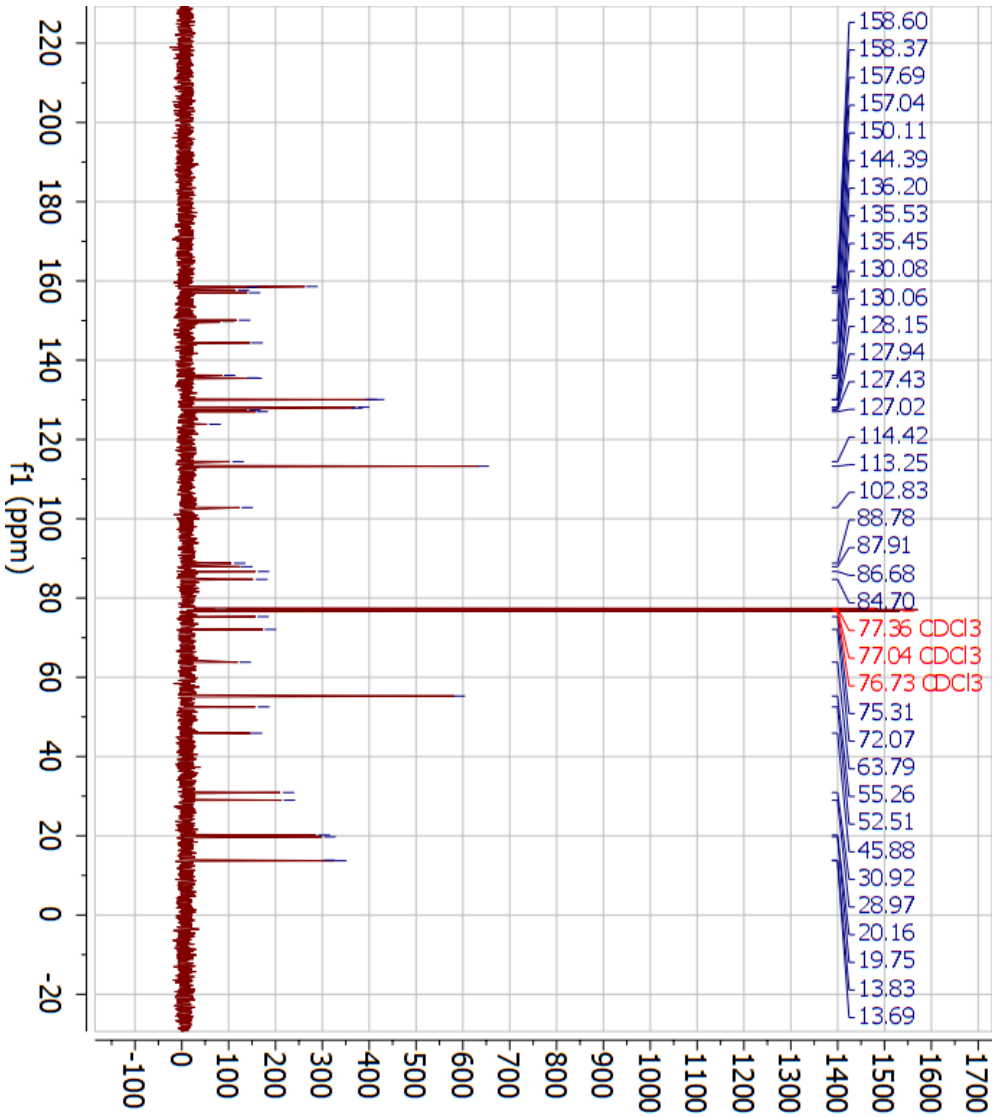

f1 (ppm)
0.81
1.00
0.98
1.02
0.89
1.02
1.05
0.89
1.08
0.72
2.20
3.96
4.10
4.18
6.11
5
23
8
16
19
12
13
14
20
18
17
25,26
27,30
28,31
29,32
8.50
7.43
7.26 CDCl3
7.26
5.84
5.82
4.45
4.22
4.22
3.43
3.41
3.39
3.37
3.36
1.98
1.64
1.64
1.62
1.60
1.60
1.59
1.58
1.57
1.38
1.36
1.35
1.34
1.33
1.32
1.31
0.97
0.95
0.95
0.93
0.93
0.91
0.06

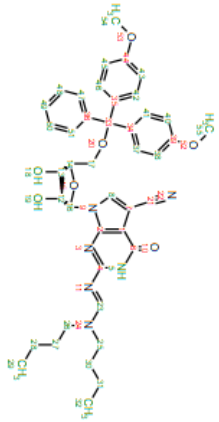


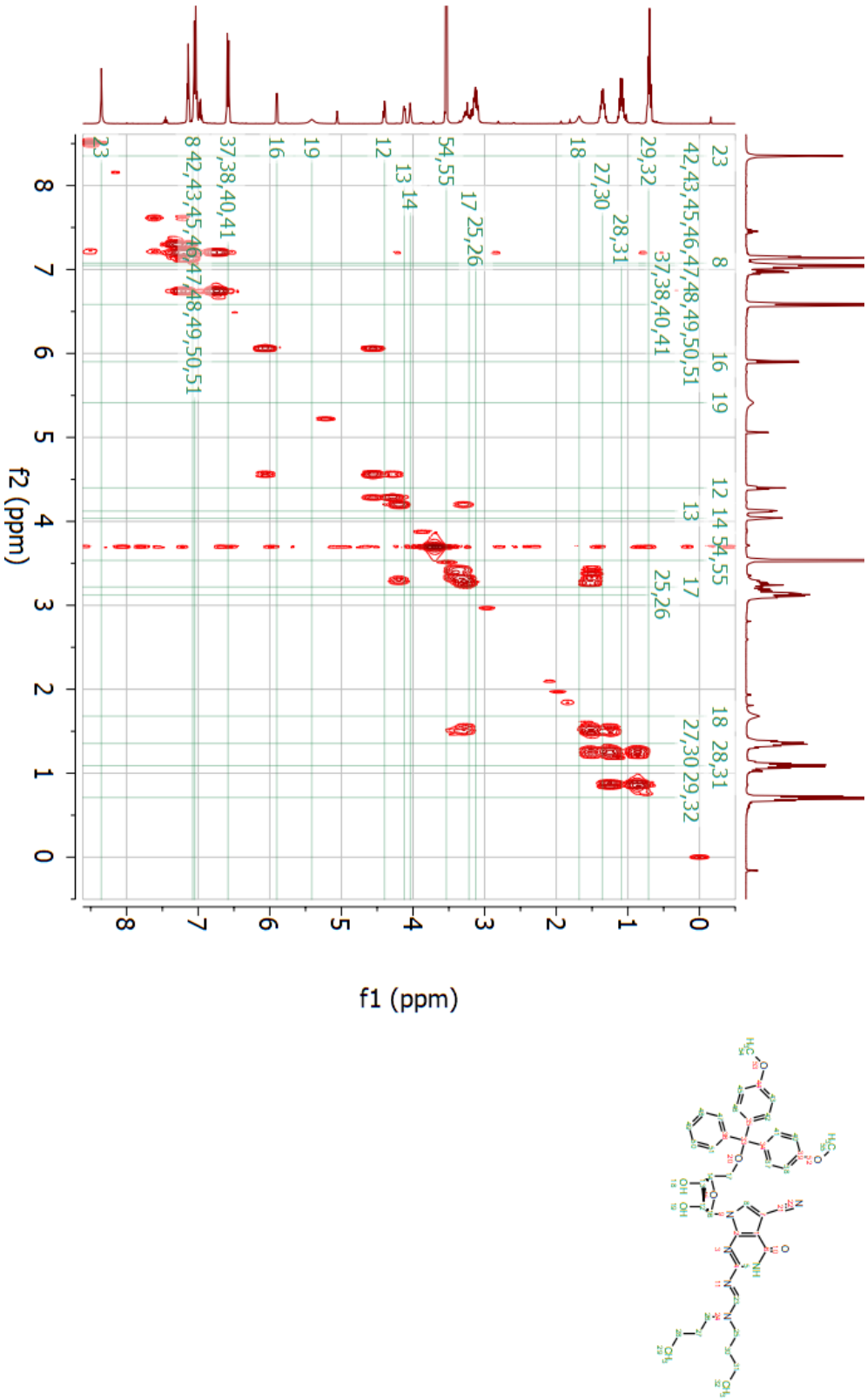

158.47
157.07
149.94
130.55
114.39
103.18
91.42
86.91
86.20
77.36 CDCl3
77.25
77.04 CDCl3
76.72 CDCl3
73.28
71.88
62.61
52.58
45.85
30.84
28.96
20.17
19.79
13.86
13.77
f1 (ppm)

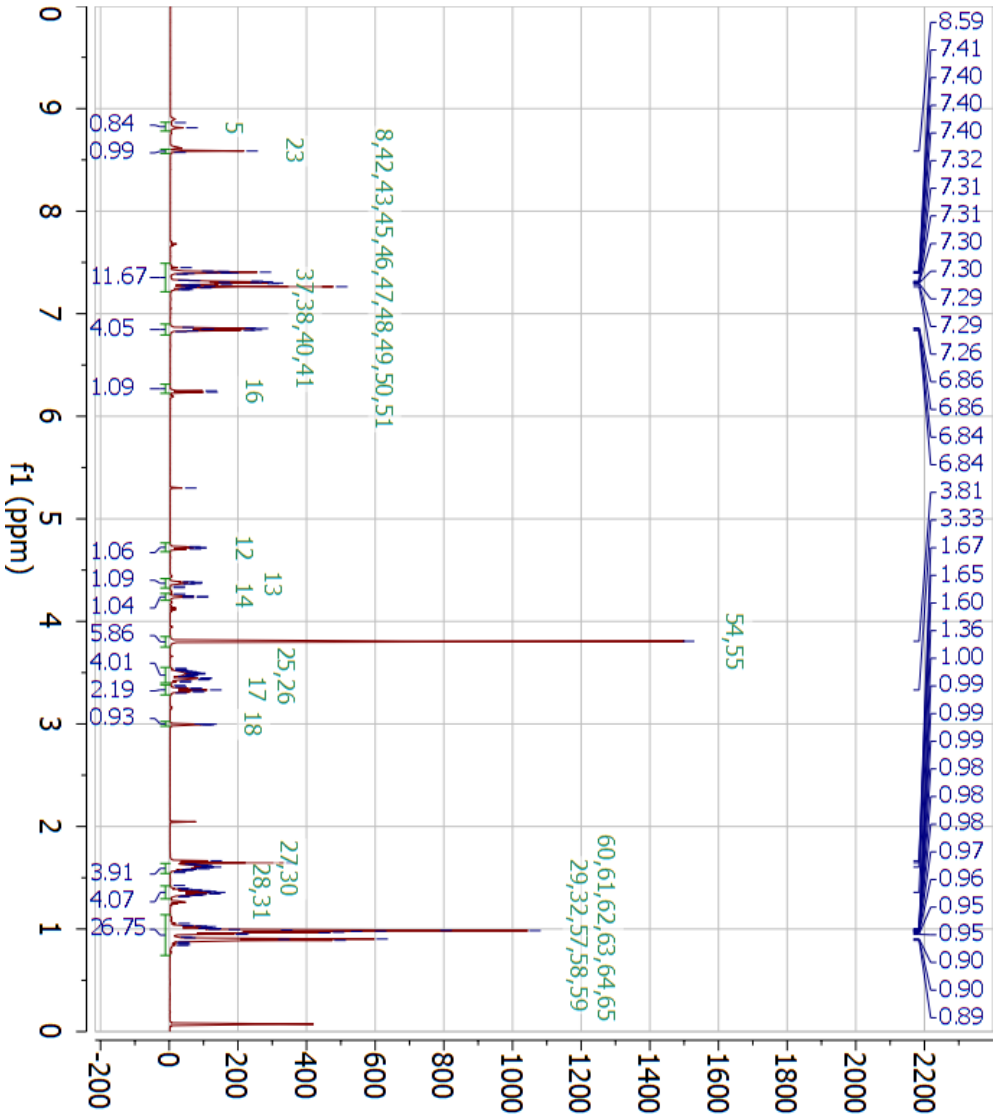

f1 (ppm)
0.91
1.34
11.89
4.23
1.01
0.89
1.00
1.01
1.07
6.12
2.03
4.16
1.10
4.22
4.42
6.33
5
23
8
42,43,45,46,47,48,49,50,51
37,38,40,41
16
19
12
14
13
54,55
17
25,26
18
27,30
28,31
29,32
8.35
7.15
7.14
7.13
7.05
7.01
6.98
6.97
6.59
6.57
5.91
5.89
4.40
4.13
4.12
4.04
4.03
3.54
3.24
3.15
3.14
3.13
3.11
3.09
1.39
1.37
1.35
1.33
1.12
1.10
1.08
1.06
0.73
0.71
0.71
0.70
0.68

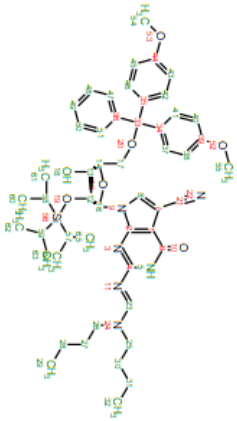


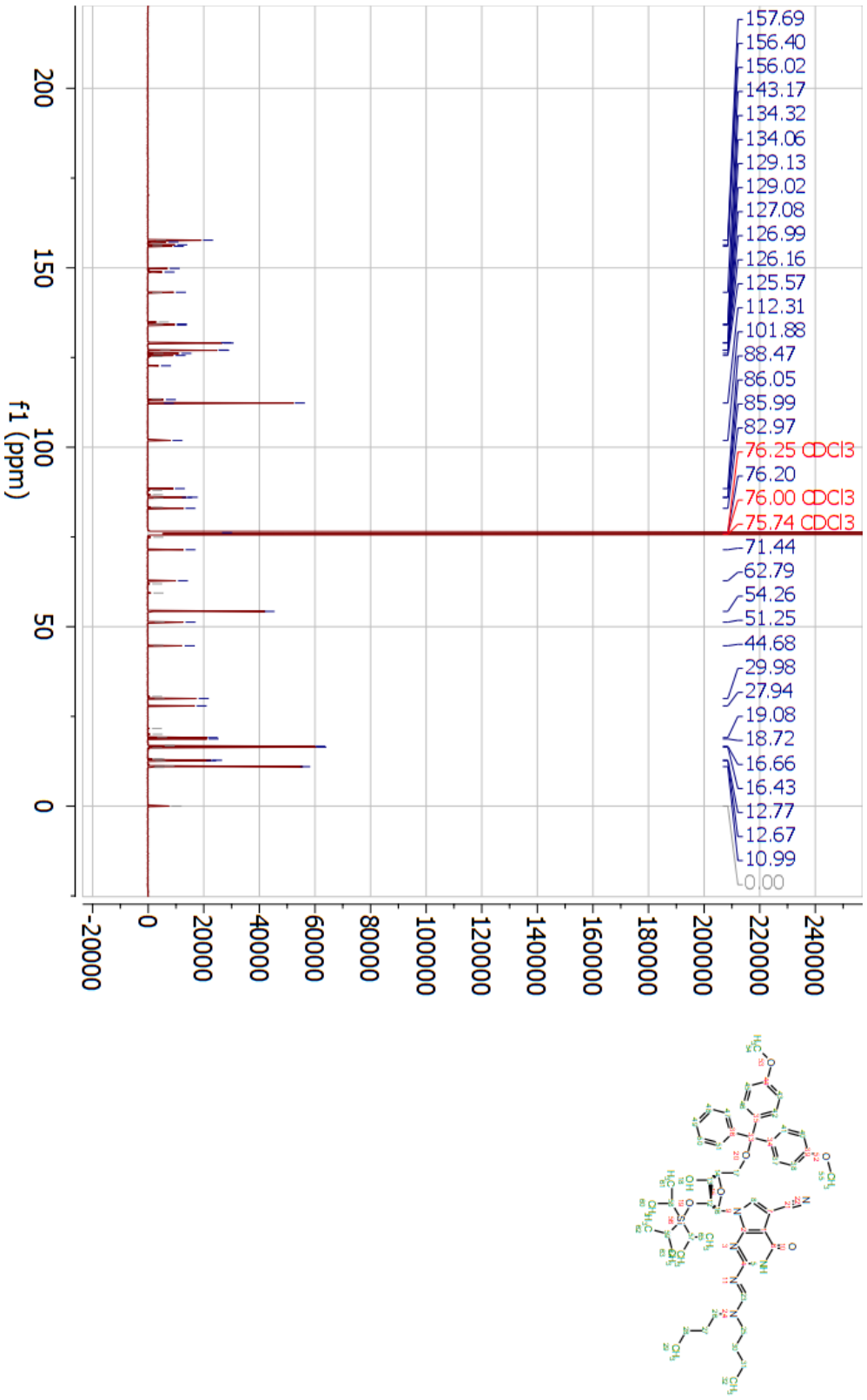

158.60
158.37
157.69
157.04
150.11
144.39
136.20
135.53
135.45
130.08
130.06
128.15
127.94
127.43
127.02
114.42
113.25
102.83
88.78
87.91
86.68
84.70
77.36 CDCl3
77.04 CDCl3
76.73 CDCl3
75.31
72.07
63.79
55.26
52.51
45.88
30.92
28.97
20.16
19.75
13.83
13.69
f1 (ppm)

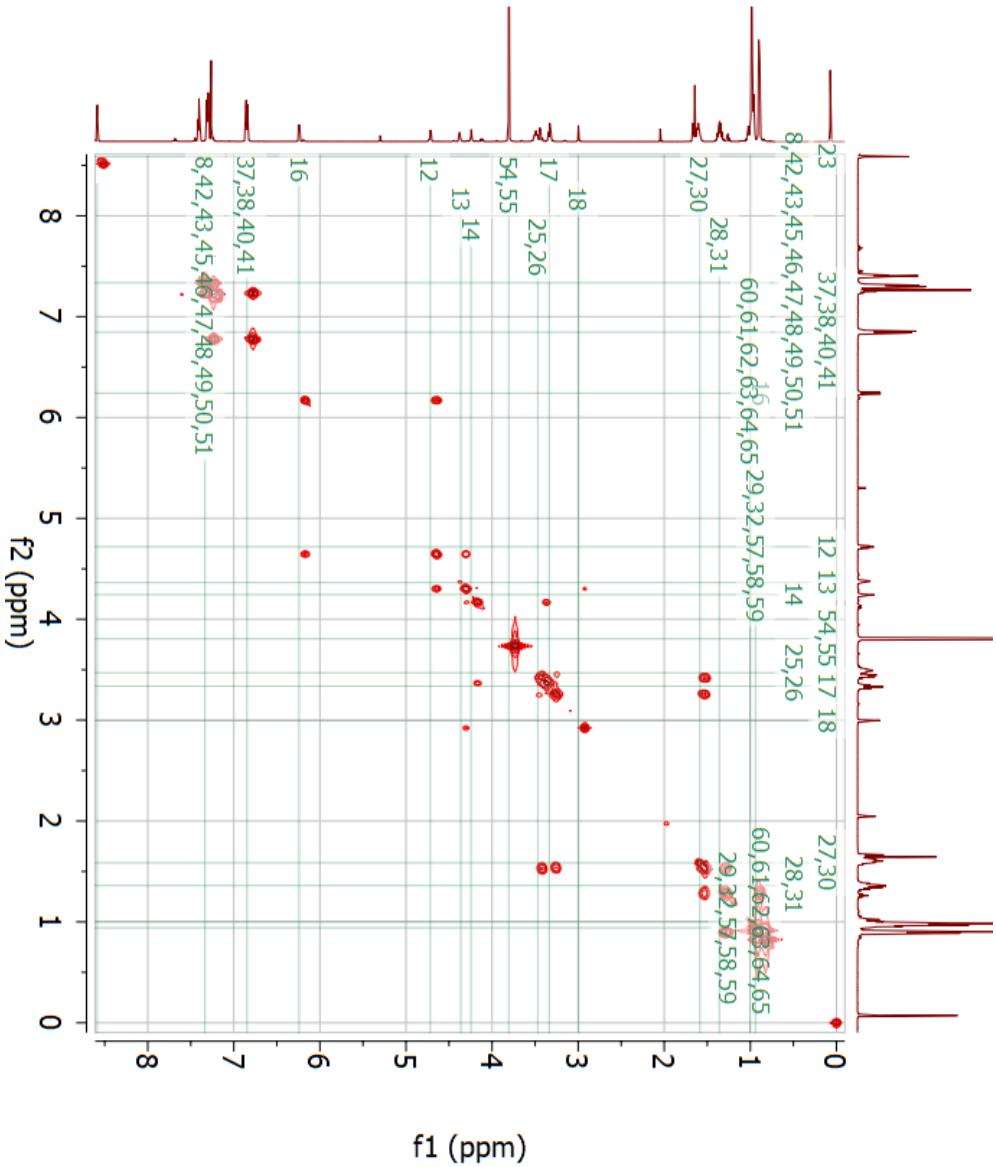


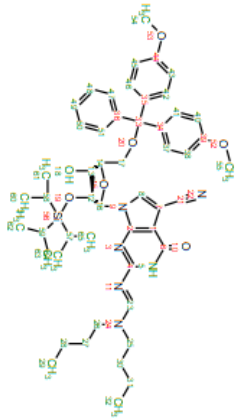

f2 (ppm)
f1 (ppm)
23
8
16 19
12 14 54,55
13
17
25,26
18 28,31
27,30 29,32
42,43,45,46,47,48,49,50,51
37,38,40,41

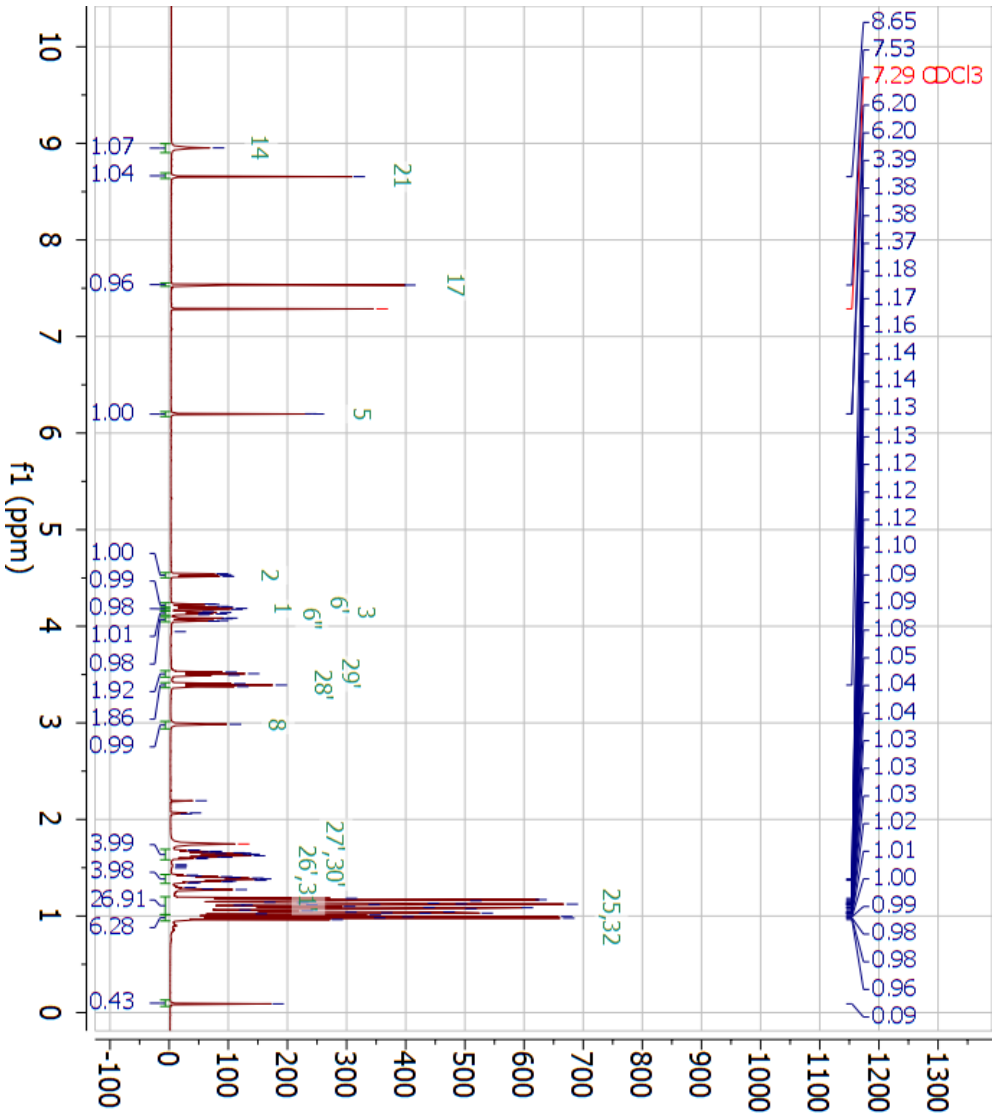

f1 (ppm)
0.84
0.99
11.67
4.05
1.09
1.06
1.09
1.04
5.86
4.01
2.19
0.93
3.91
4.07
26.75
5
23
8,42,43,45,46,47,48,49,50,51
37,38,40,41
16
12
13
14
54,55
25,26
17
18
27,30
28,31
60,61,62,63,64,65
29,32,57,58,59
8.59
7.41
7.40
7.40
7.40
7.32
7.31
7.31
7.30
7.30
7.29
7.29
7.26
6.86
6.86
6.84
6.84
3.81
3.33
1.67
1.65
1.60
1.36
1.00
0.99
0.99
0.99
0.98
0.98
0.98
0.97
0.96
0.95
0.95
0.90
0.90
0.89

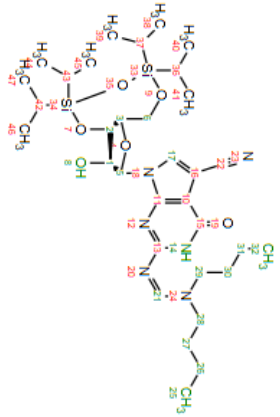


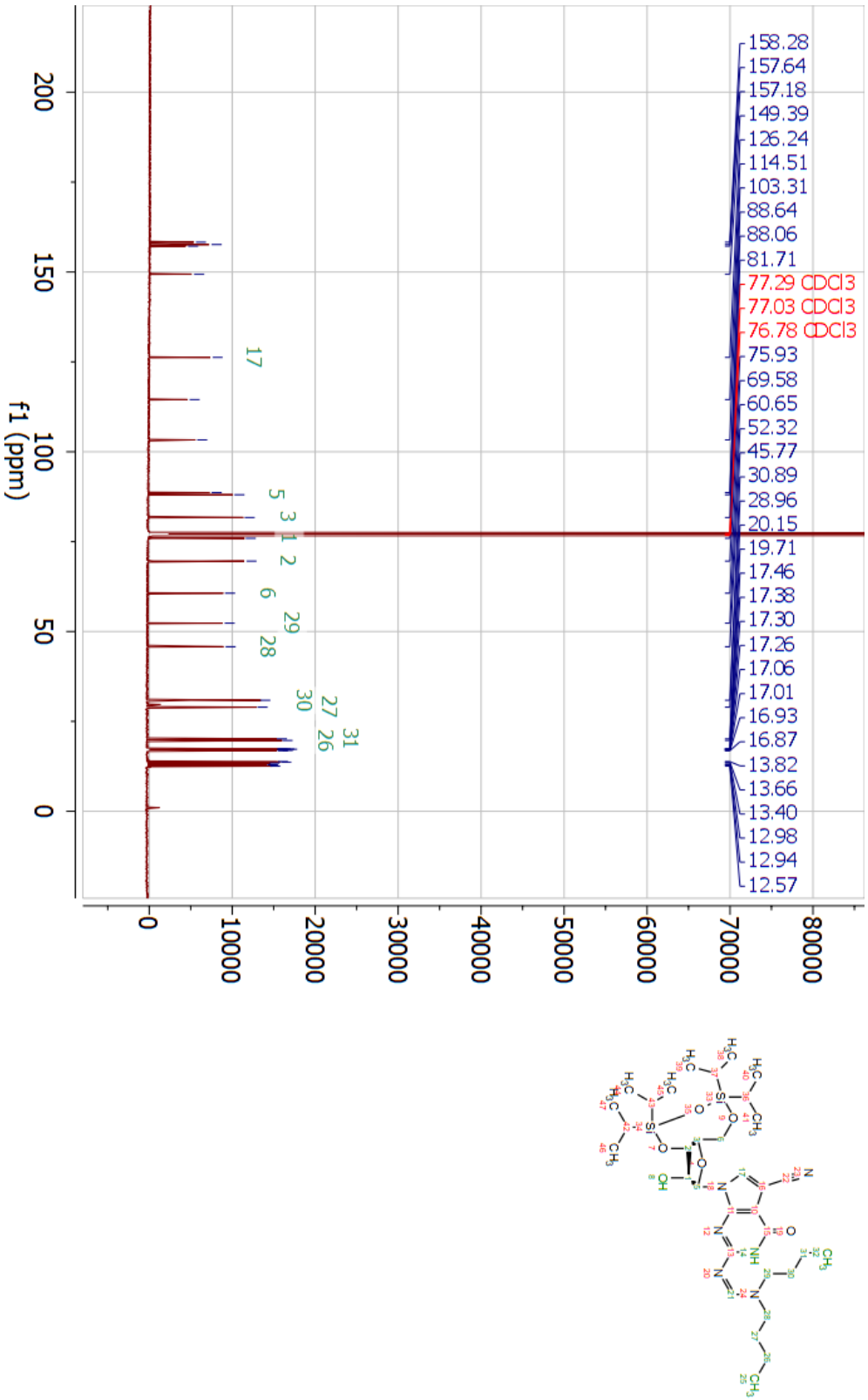

157.69
156.40
156.02
143.17
134.32
134.06
129.13
129.02
127.08
126.99
126.16
125.57
112.31
101.88
88.47
86.05
85.99
82.97
76.25 CDCl3
76.20
76.00 CDCl3
75.74 CDCl3
71.44
62.79
54.26
51.25
44.68
29.98
27.94
19.08
18.72
16.66
16.43
12.77
12.67
10.99
0.00
f1 (ppm)
200
150
100
50
0
-20000
0
20000
40000
60000
80000
100000
120000
140000
160000
180000
200000
220000
240000

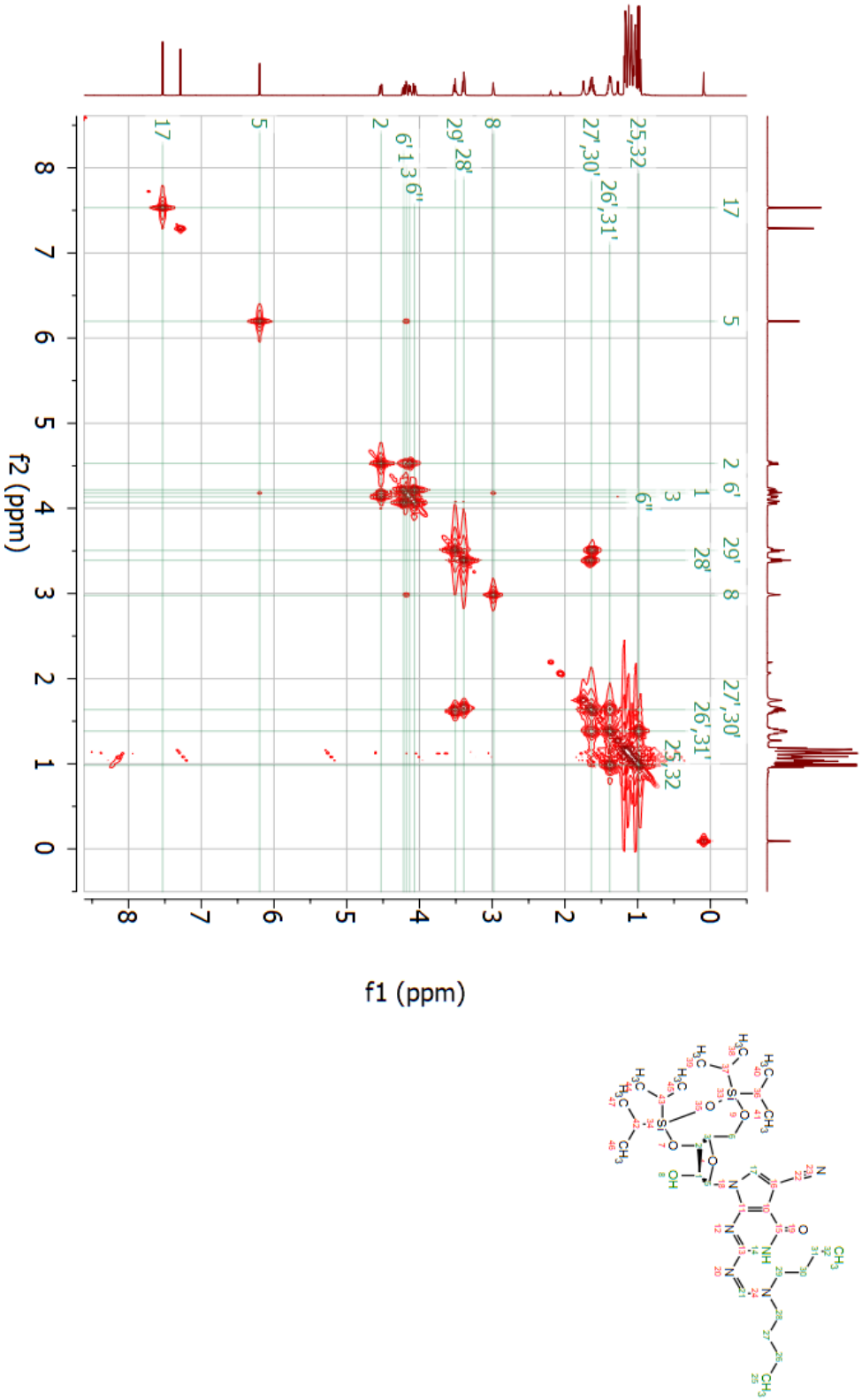

f2 (ppm)
f1 (ppm)
23
8,42,43,45,46,47,48,49,50,51
37,38,40,41
12 13 54,55 14 25,26 17 18
27,30
28,31
60,61,62,63,64,65 29,32,57,58,59
16
37,38,40,41
8,42,43,45,46,47,48,49,50,51
12
13 14
54,55
17 25,26
18
27,30
28,31
60,61,62,63,64,65
29,32,57,58,59

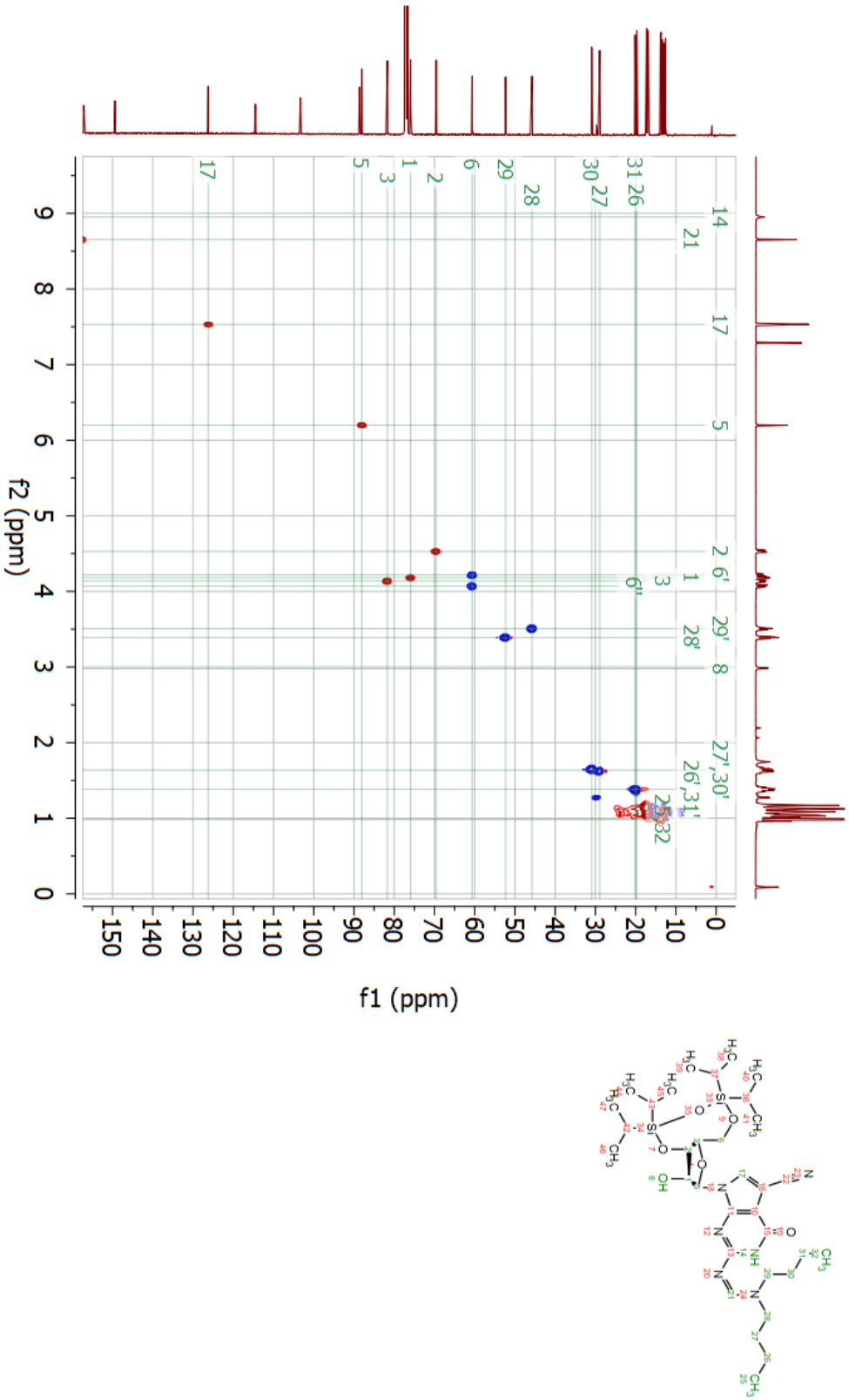


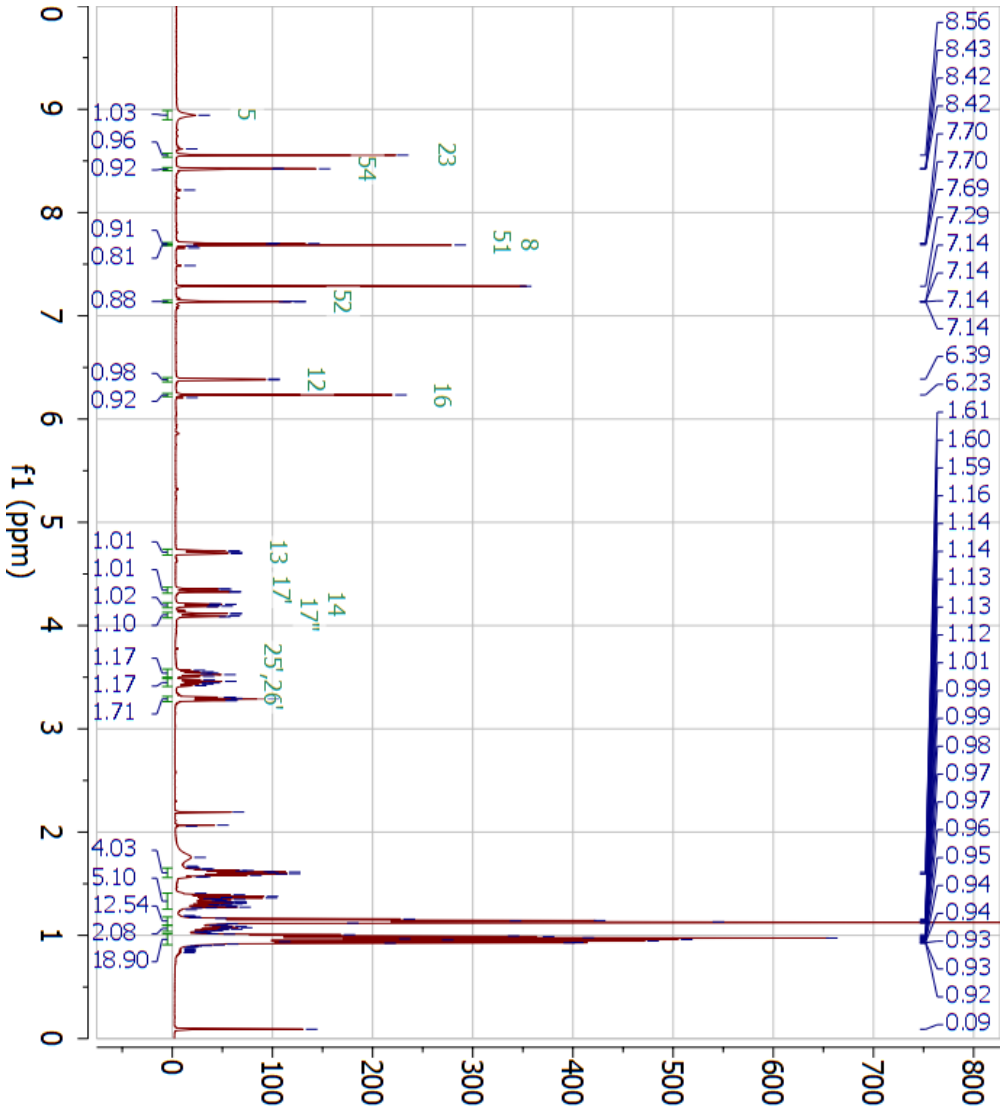

f1 (ppm)
1.07
1.04
0.96
1.00
1.00
0.99
0.98
1.01
0.98
1.92
1.86
0.99
3.99
3.98
26.91
6.28
0.43
14
21
17
5
2
1
3
6'
6''
29'
28'
8
27',30'
26',31'
25,32
8.65
7.53
7.29 CDCl3
6.20
6.20
3.39
1.38
1.38
1.37
1.18
1.17
1.16
1.14
1.14
1.13
1.13
1.12
1.12
1.12
1.10
1.09
1.09
1.08
1.05
1.04
1.04
1.03
1.03
1.03
1.02
1.01
1.00
0.99
0.98
0.98
0.96
0.09

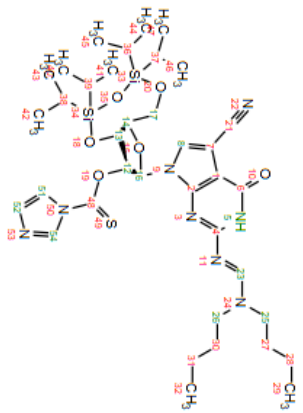


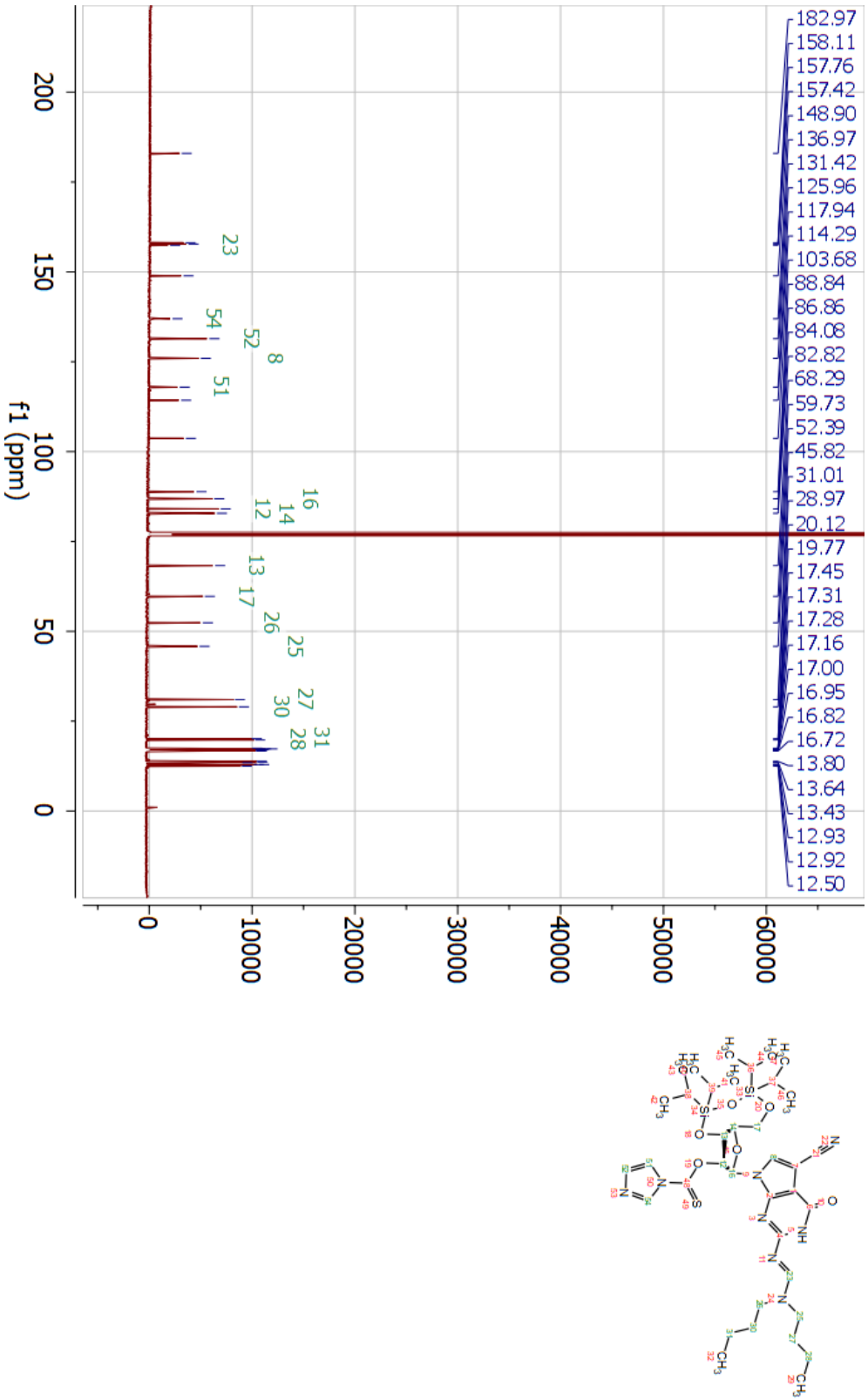

158.28
157.64
157.18
149.39
126.24
114.51
103.31
88.64
88.06
81.71
77.29 CDCl3
77.03 CDCl3
76.78 CDCl3
75.93
69.58
60.65
52.32
45.77
30.89
28.96
20.15
19.71
17.46
17.38
17.30
17.26
17.06
17.01
16.93
16.87
13.82
13.66
13.40
12.98
12.94
12.57
17
5
3
1
2
6
29
28
30
27
26
31
f1 (ppm)
200
150
100
50
0
0
10000
20000
30000
40000
50000
60000
70000
80000

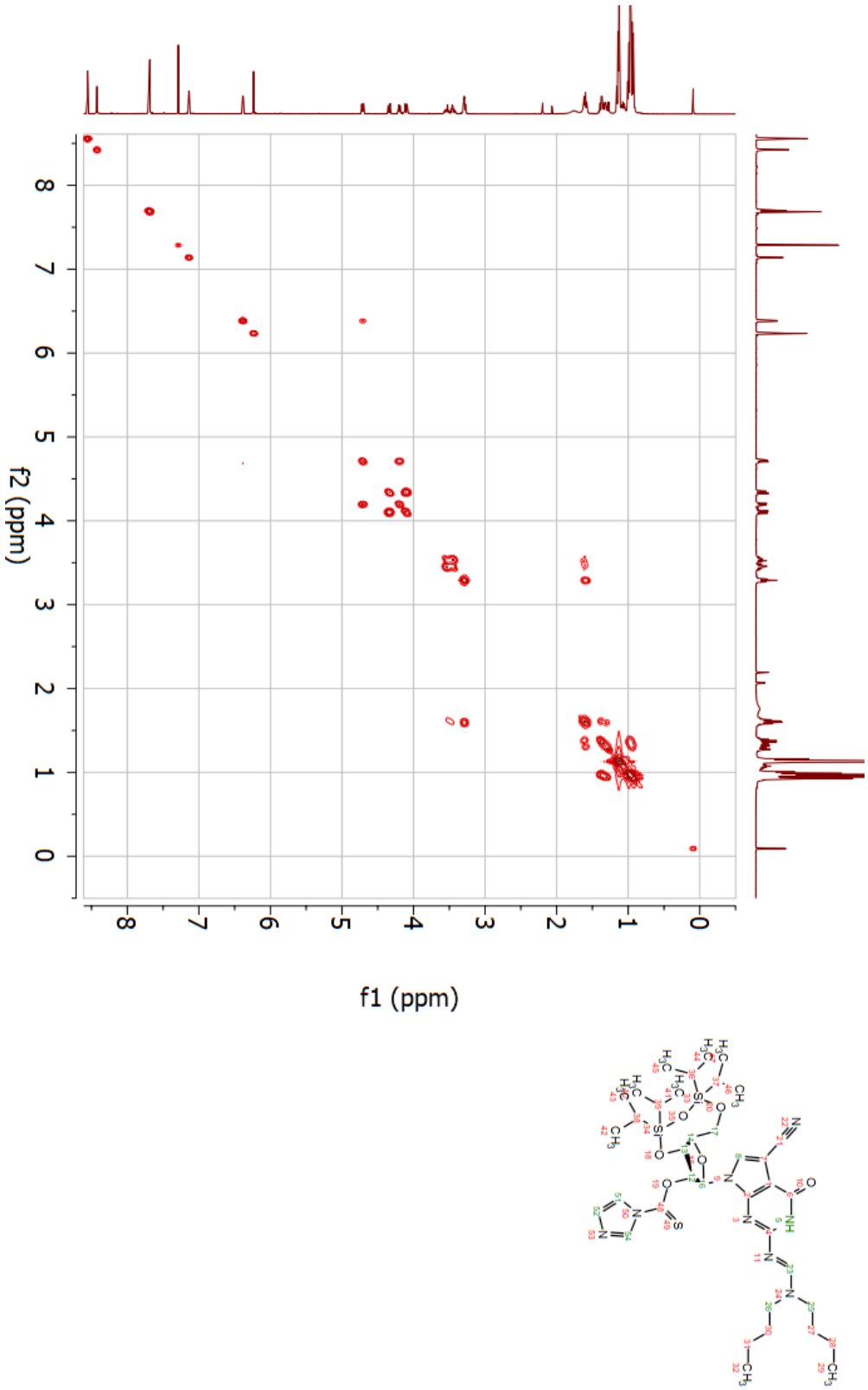

f2 (ppm)
f1 (ppm)
17
5
2
6'
1
3
6''
29'
28'
8
27',30'
26',31'
25,32

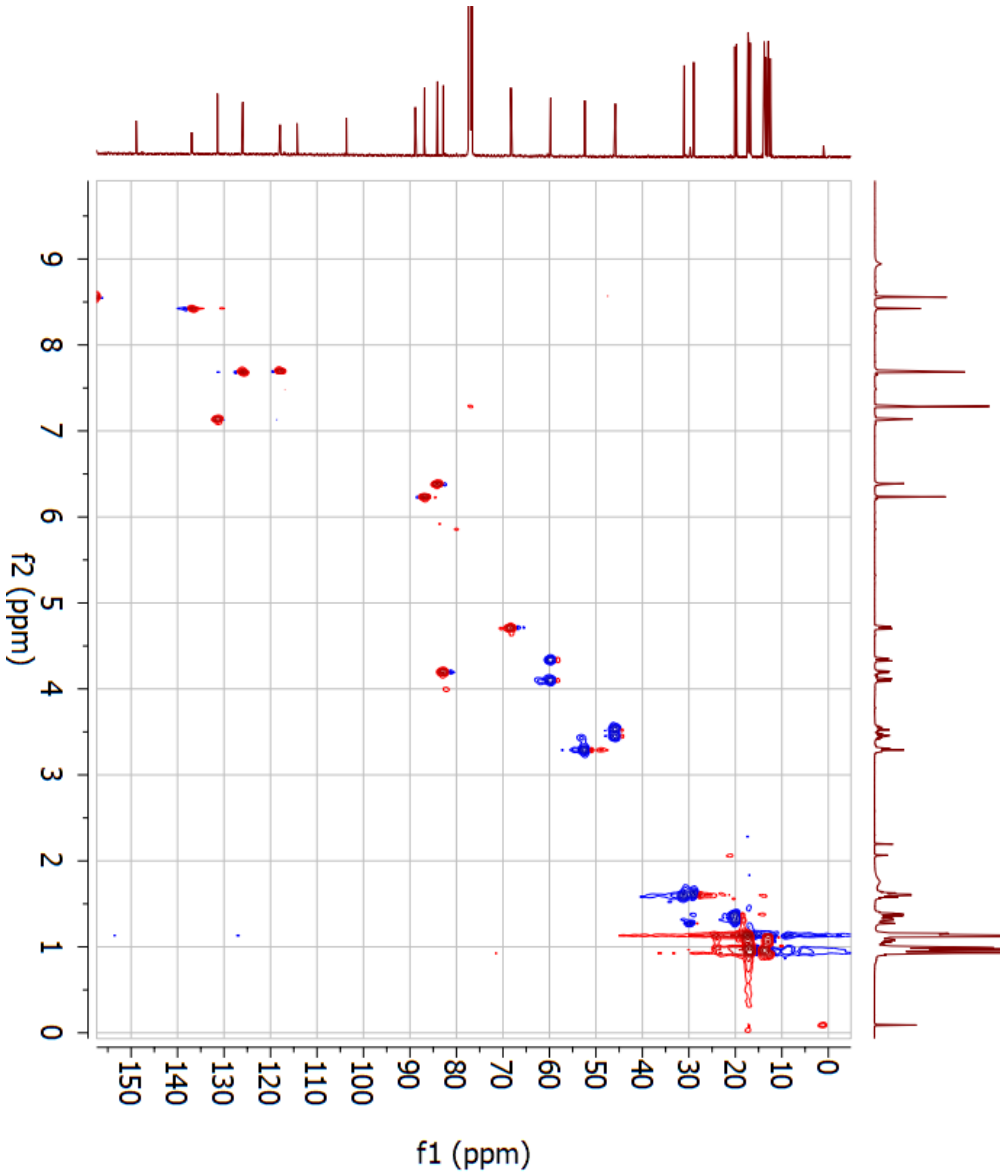

f2 (ppm)
f1 (ppm)

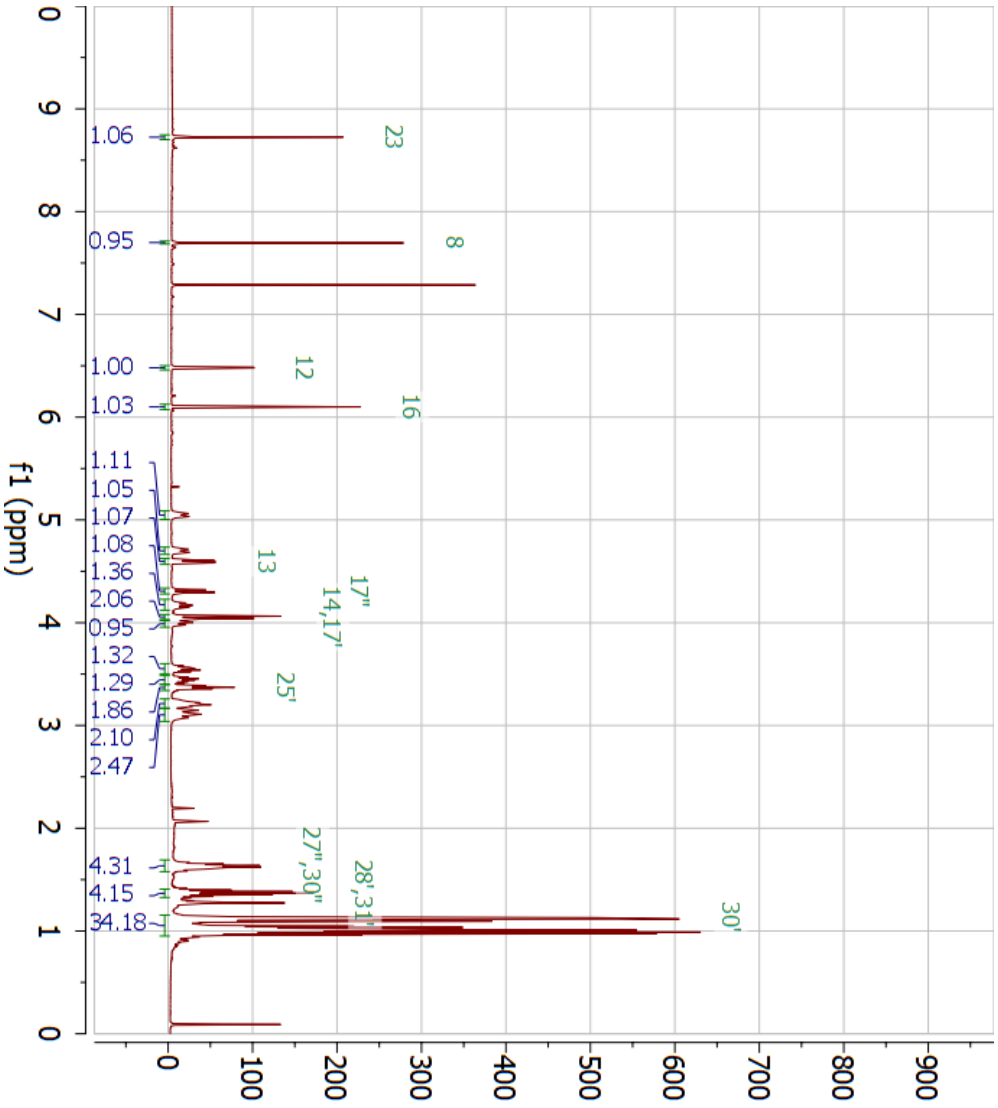

f1 (ppm)
1.03
0.96
0.92
0.91
0.81
0.88
0.98
0.92
1.01
1.01
1.02
1.10
1.17
1.17
1.71
4.03
5.10
12.54
2.08
18.90
8.56
8.43
8.42
8.42
7.70
7.70
7.69
7.29
7.14
7.14
7.14
7.14
6.39
6.23
1.61
1.60
1.59
1.16
1.14
1.14
1.13
1.13
1.12
1.01
0.99
0.99
0.98
0.97
0.97
0.96
0.95
0.94
0.94
0.93
0.93
0.92
0.09

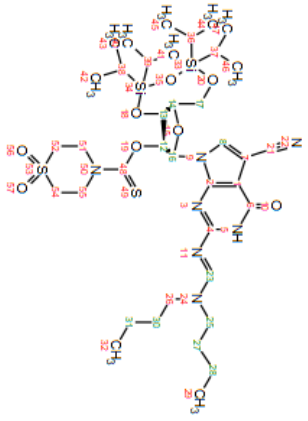


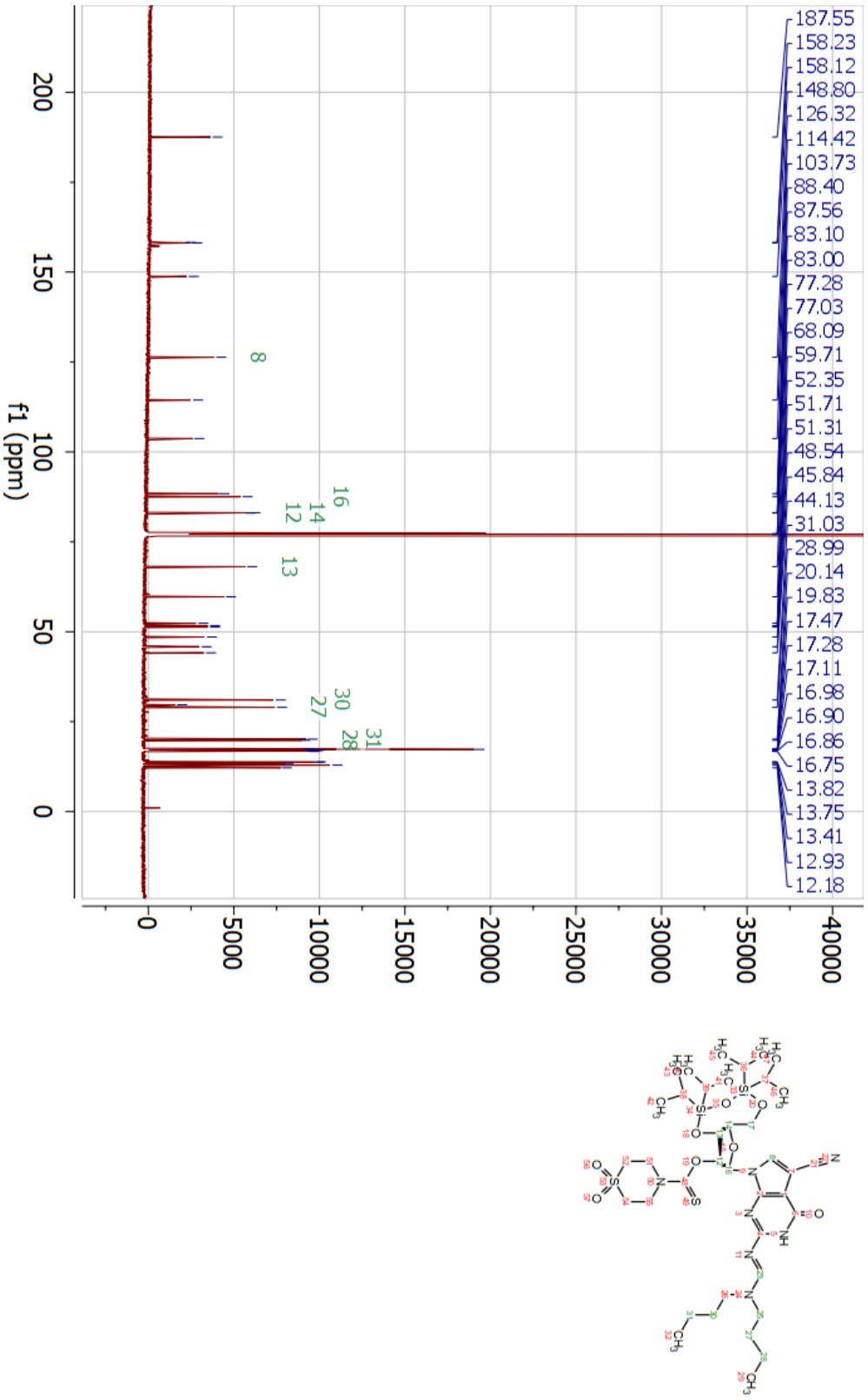

182.97
158.11
157.76
157.42
148.90
136.97
131.42
125.96
117.94
114.29
103.68
88.84
86.86
84.08
82.82
68.29
59.73
52.39
45.82
31.01
28.97
20.12
19.77
17.45
17.31
17.28
17.16
17.00
16.95
16.82
16.72
13.80
13.64
13.43
12.93
12.92
12.50
f1 (ppm)
200
150
100
50
0
0
10000
20000
30000
40000
50000
60000
23
54
52
8
51
16
14
12
13
17
26
25
27
30
31
28

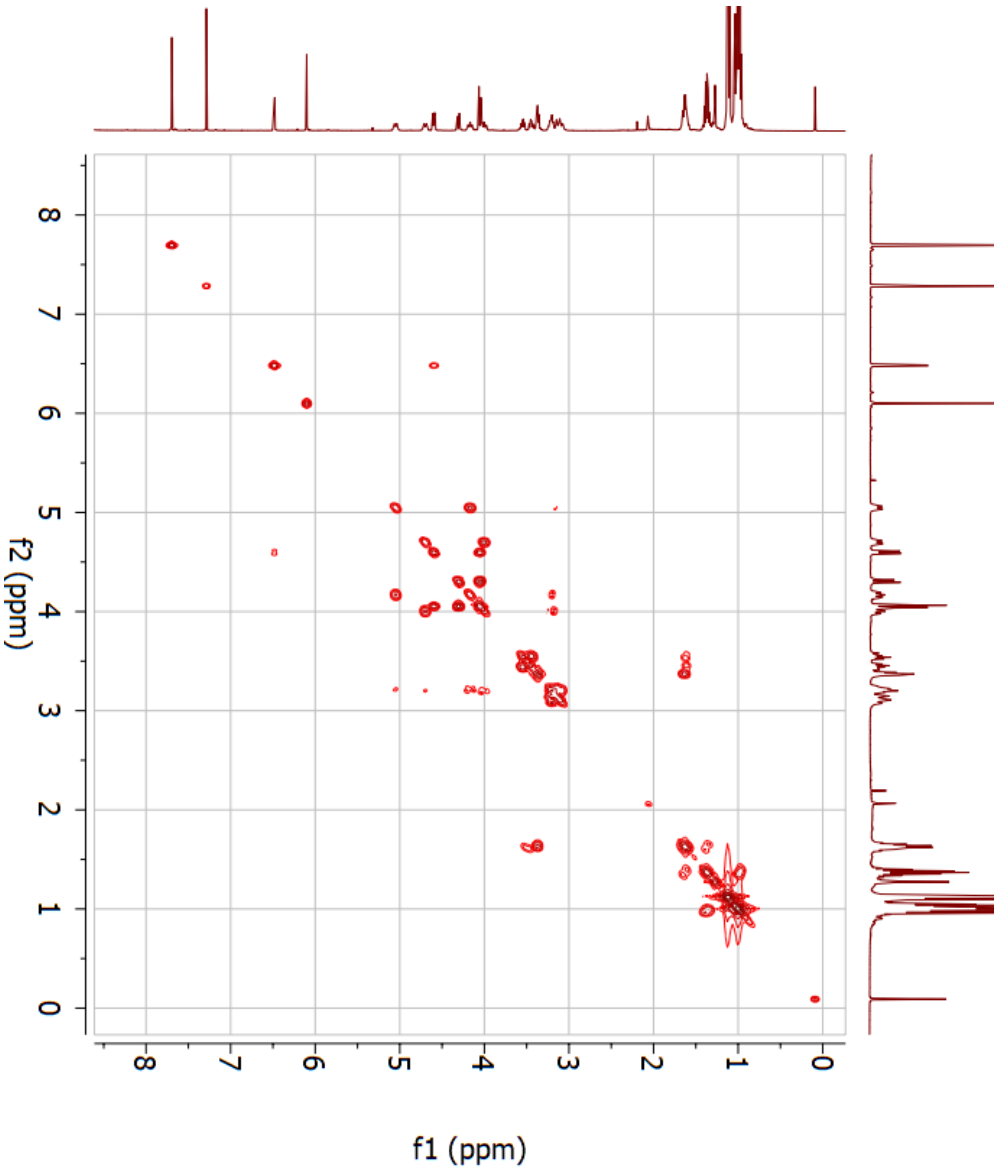

f2 (ppm)
f1 (ppm)

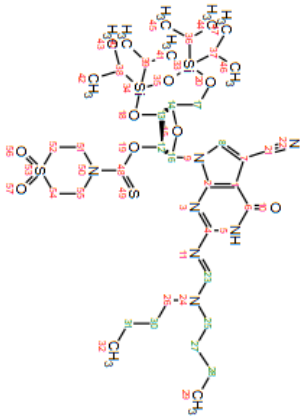

f2 (ppm)
f1 (ppm)

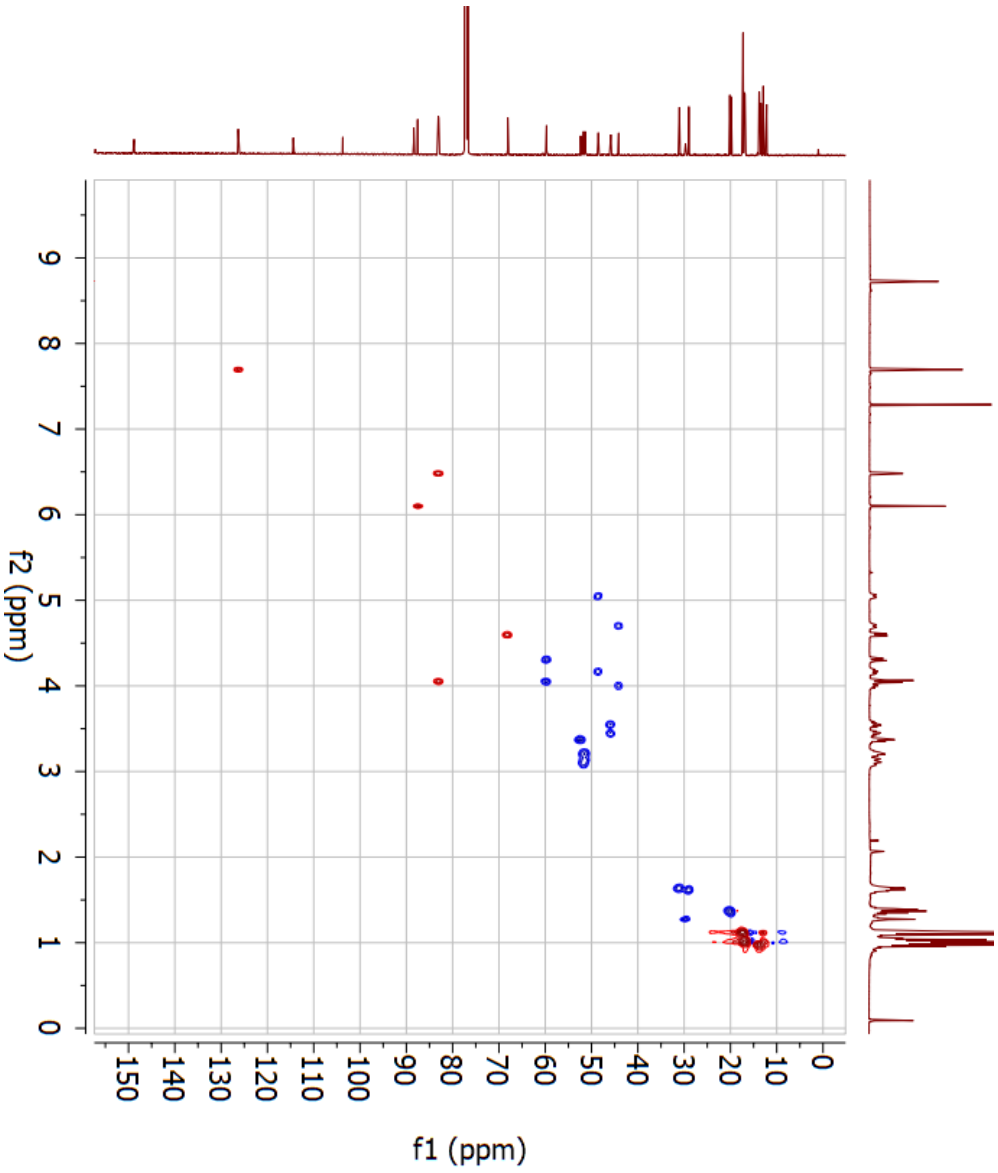

f1 (ppm)
1.06
0.95
1.00
1.03
1.11
1.05
1.07
1.08
1.36
2.06
0.95
1.32
1.29
1.86
2.10
2.47
4.31
4.15
4.18
23
8
12
16
13
17"
14,17'
25'
27",30"
28',31'
30'

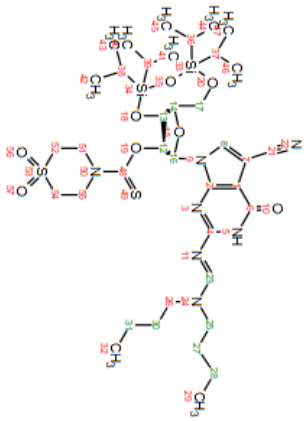


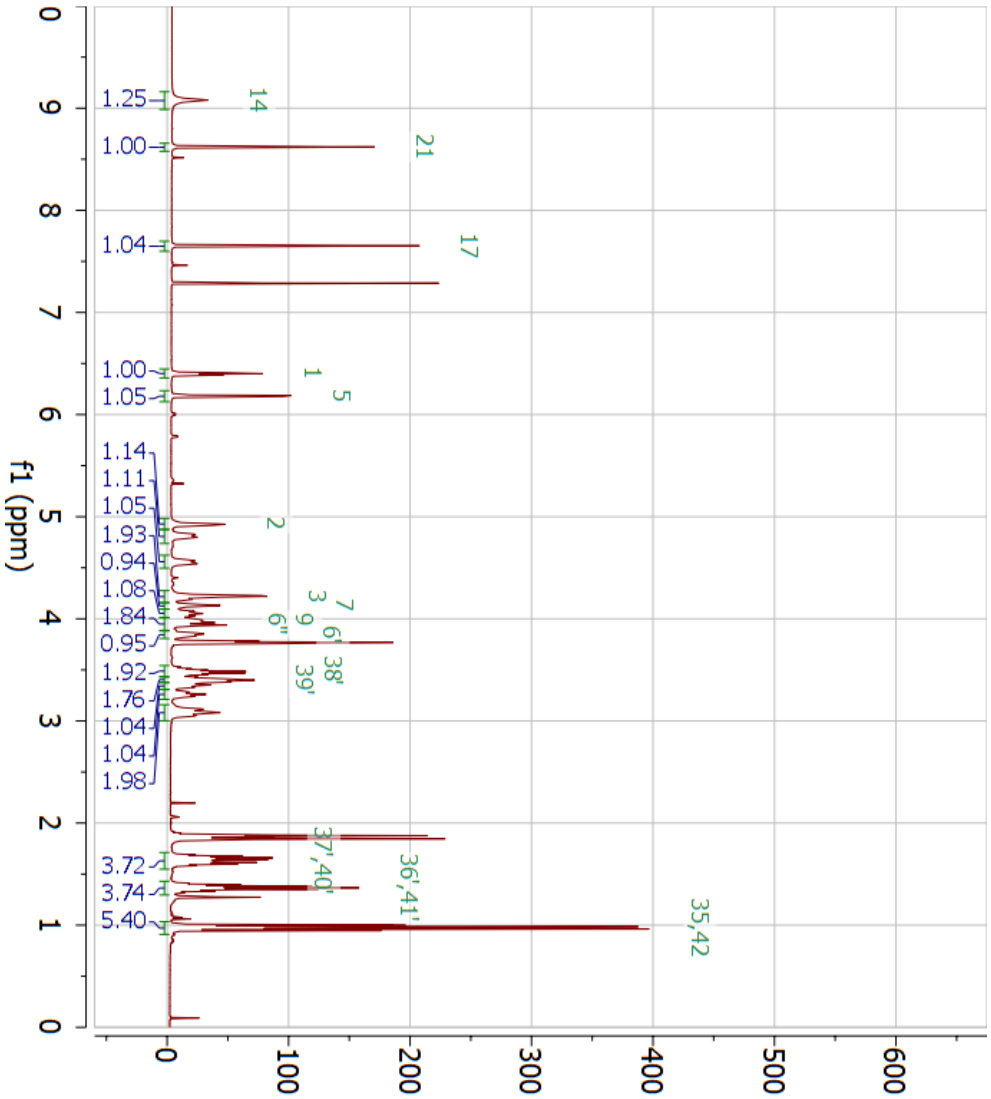

187.55
158.23
158.12
148.80
126.32
114.42
103.73
88.40
87.56
83.10
83.00
77.28
77.03
68.09
59.71
52.35
51.71
51.31
48.54
45.84
44.13
31.03
28.99
20.14
19.83
17.47
17.28
17.11
16.98
16.90
16.86
16.75
13.82
13.75
13.41
12.93
12.18
8
16
14
12
13
27
30
28
31
f1 (ppm)
0
50
100
150
200
0
5000
10000
15000
20000
25000
30000
35000
40000

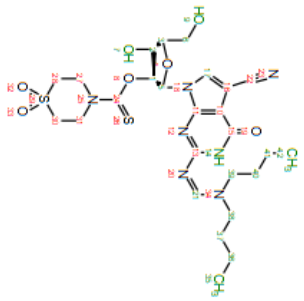

f2 (ppm)
f1 (ppm)

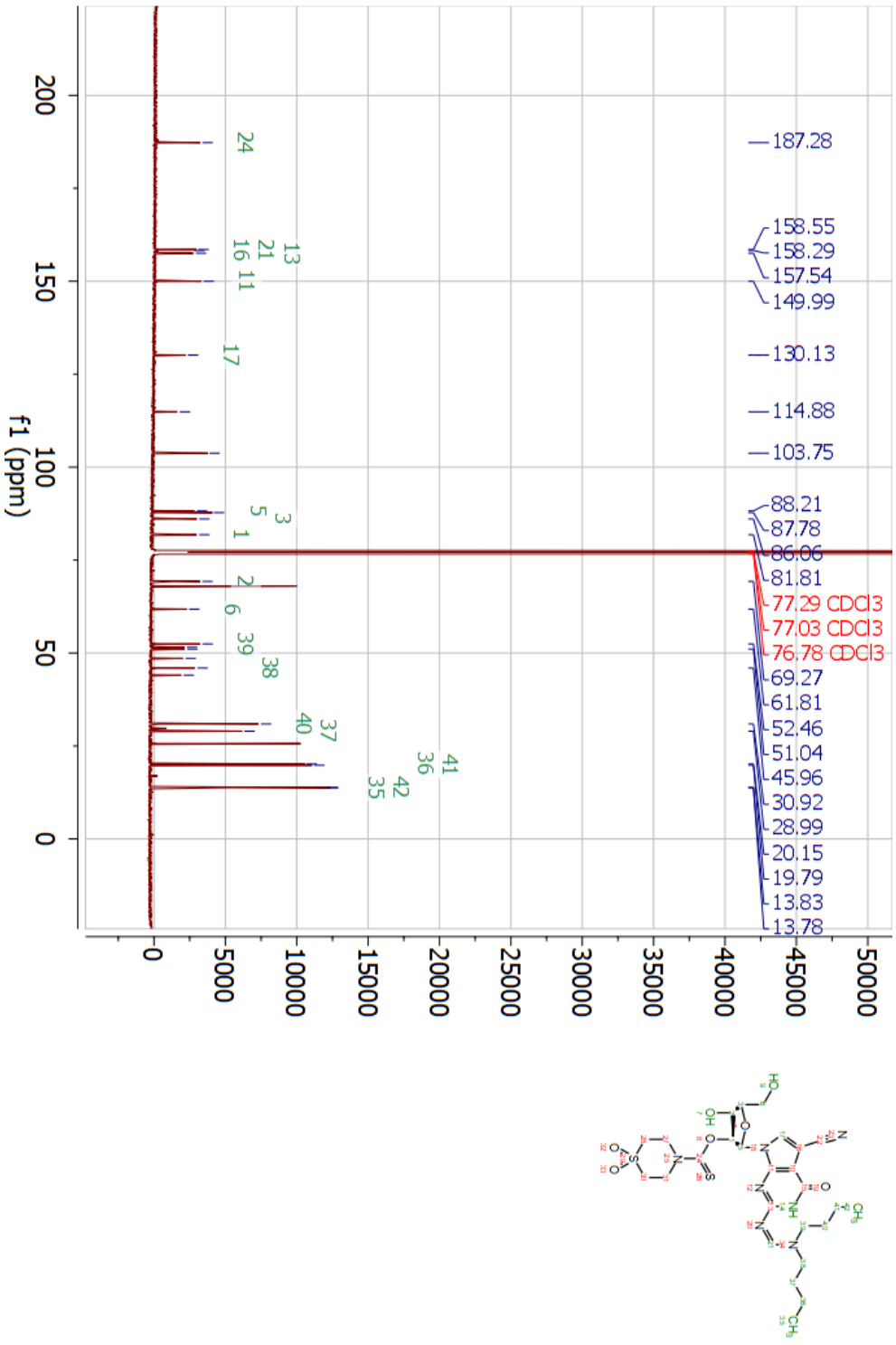


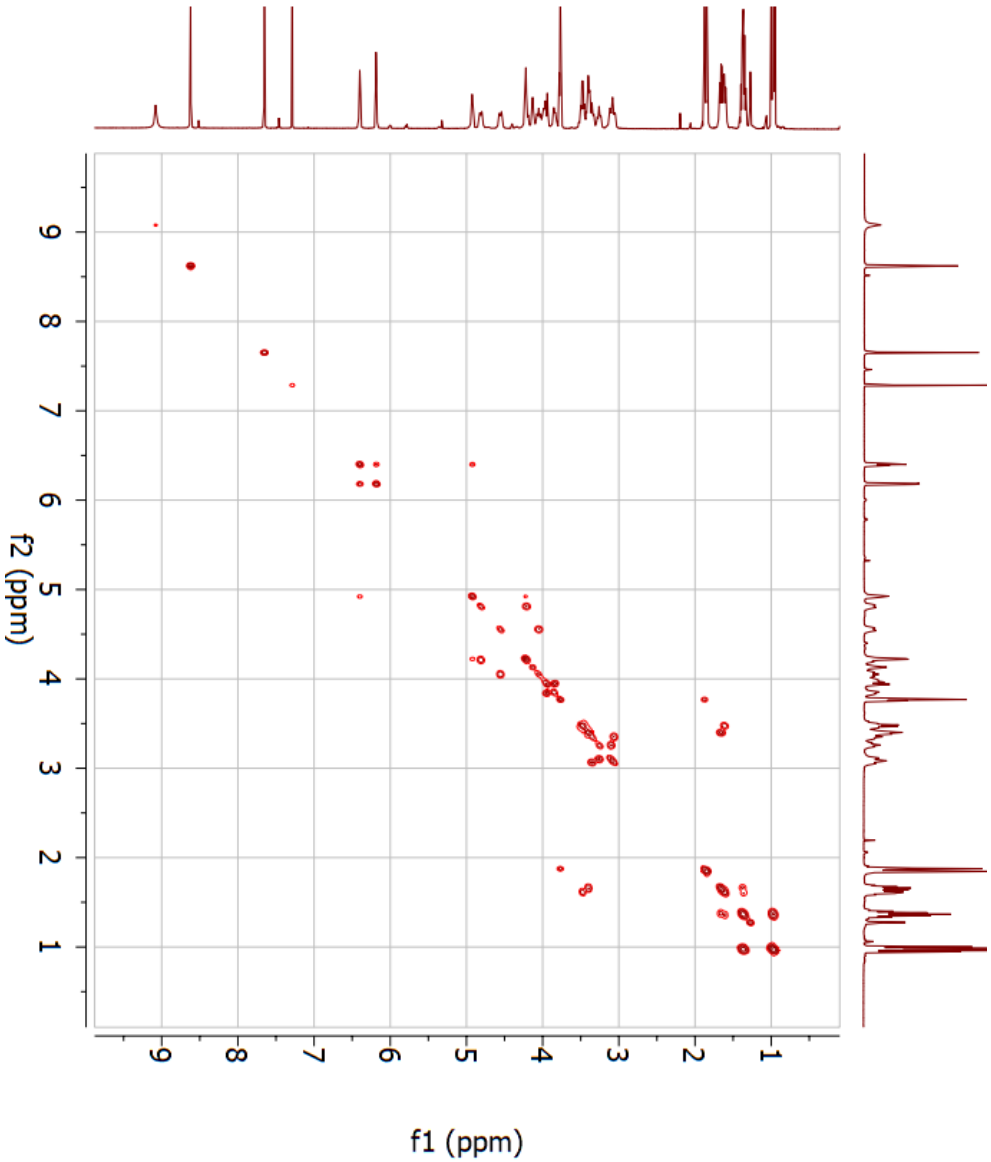

f2 (ppm)
f1 (ppm)

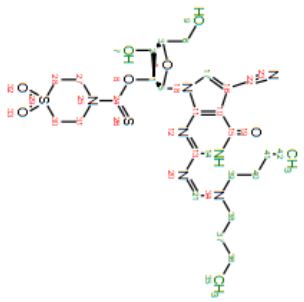


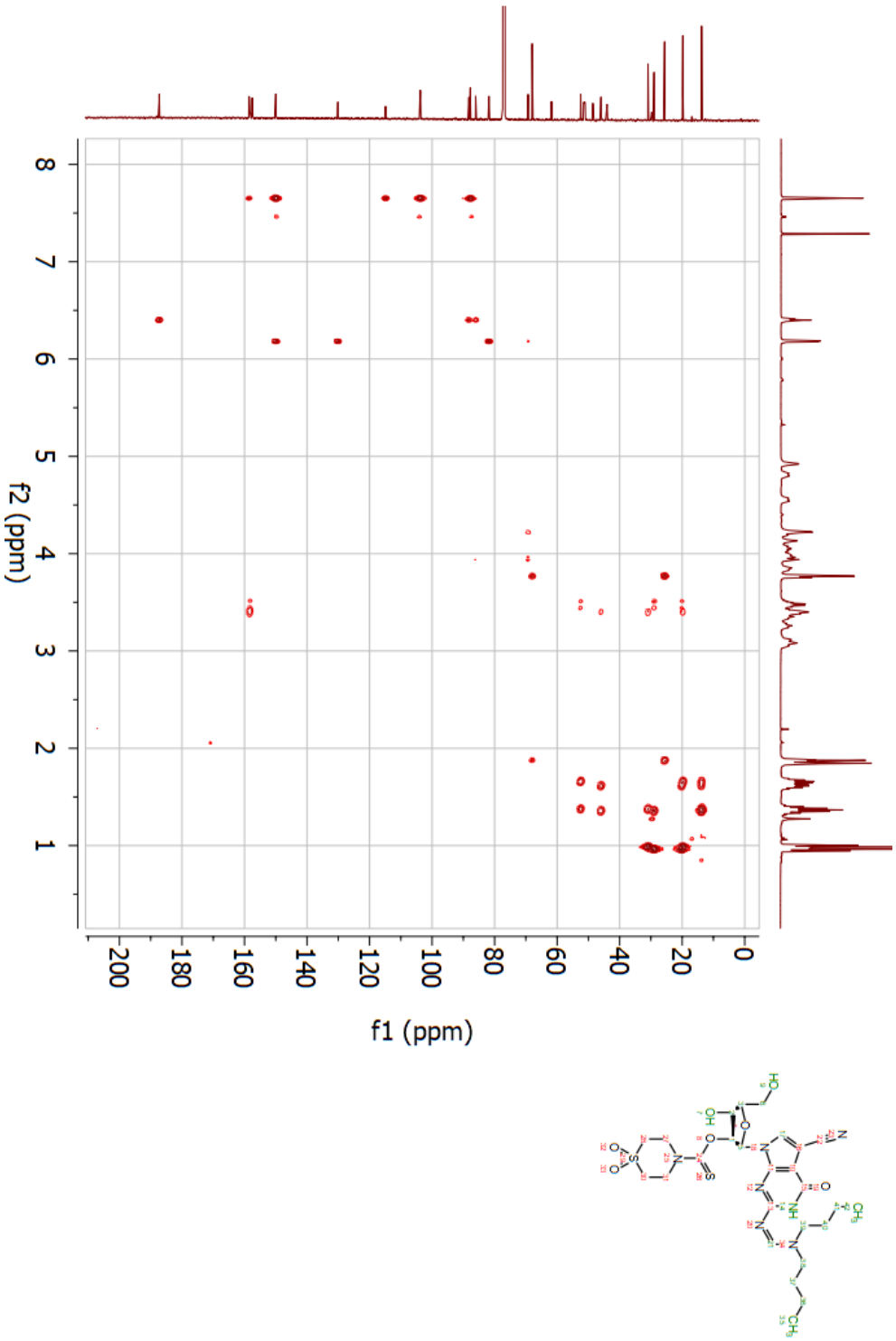

f1 (ppm)
1.25
1.00
1.04
1.00
1.05
1.14
1.11
1.05
1.93
0.94
1.08
1.84
0.95
1.92
1.76
1.04
1.04
1.98
3.72
3.74
5.40
14
21
17
1
5
2
7
3
9
6'
6''
38'
39'
37',40'
36',41'
35,42

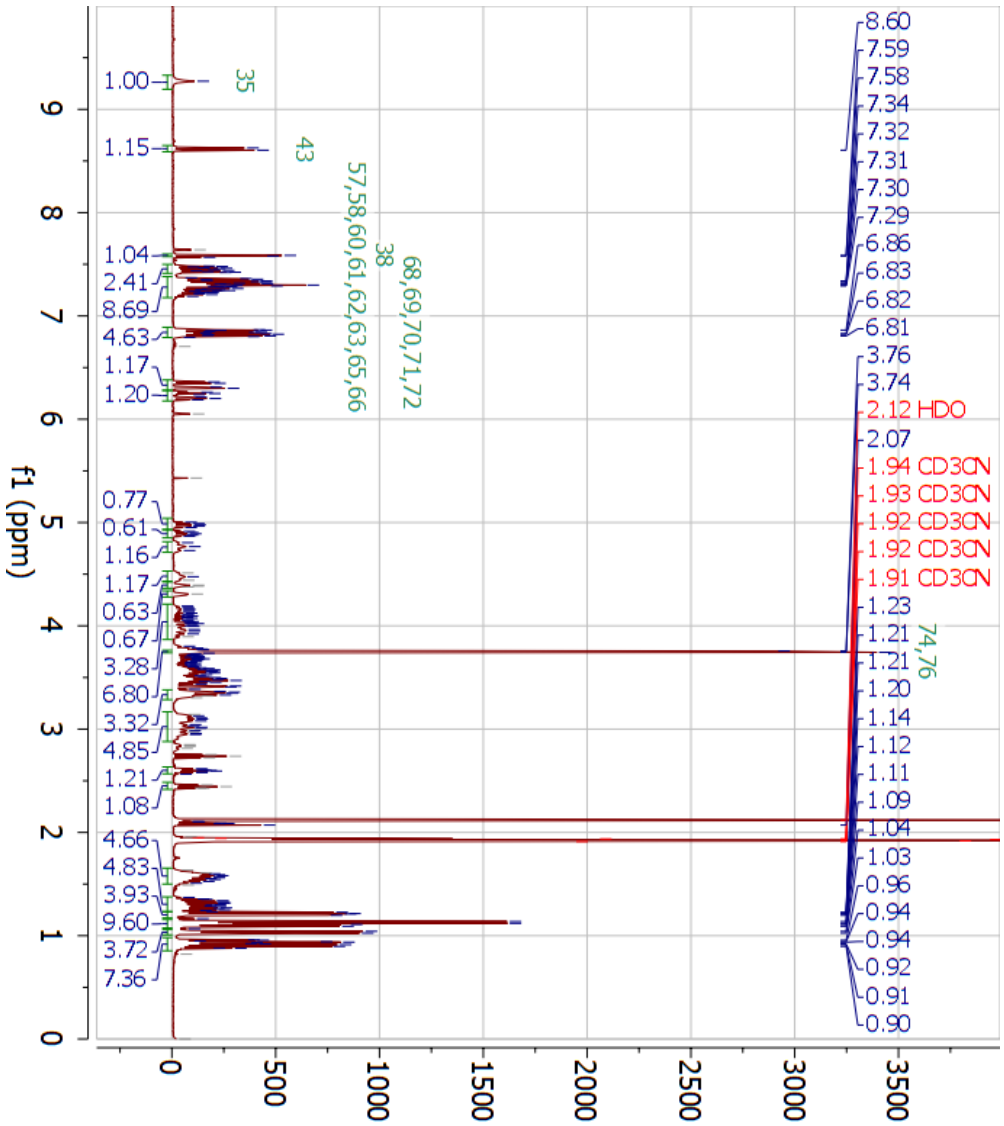


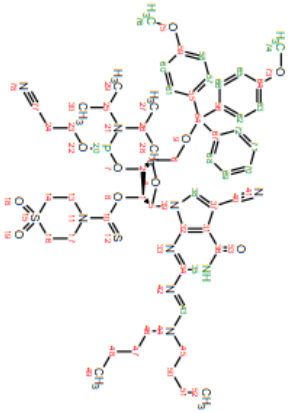

f1 (ppm)
200
150
100
50
0
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
187.28
158.55
158.29
157.54
149.99
130.13
114.88
103.75
88.21
87.78
86.06
81.81
77.29 CDCl3
77.03 CDCl3
76.78 CDCl3
69.27
61.81
52.46
51.04
45.96
30.92
28.99
20.15
19.79
13.83
13.78
24
13
21
16
11
17
3
5
1
2
6
39
38
37
40
41
36
42
35

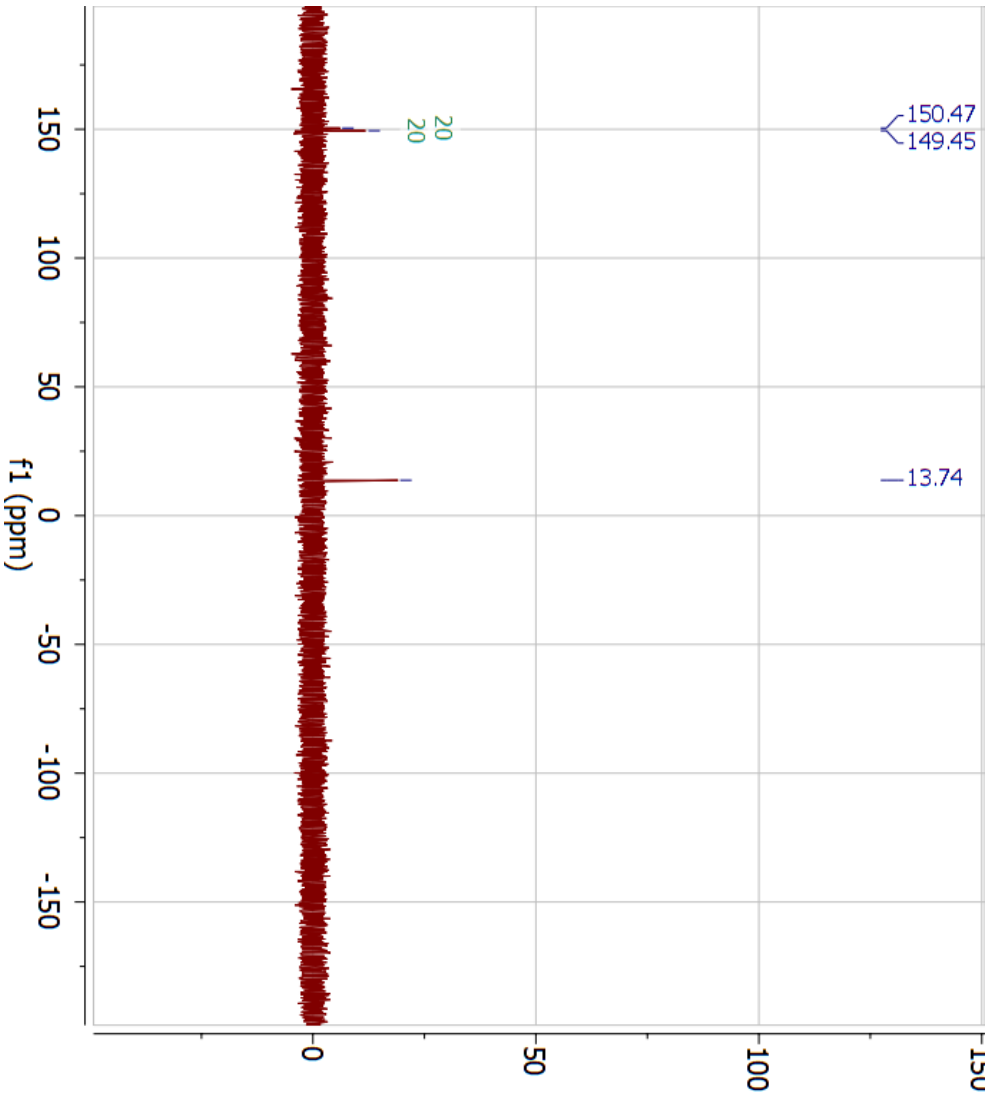

f2 (ppm)
f1 (ppm)

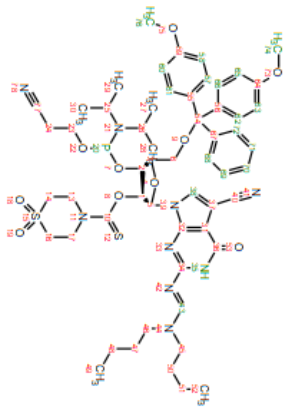


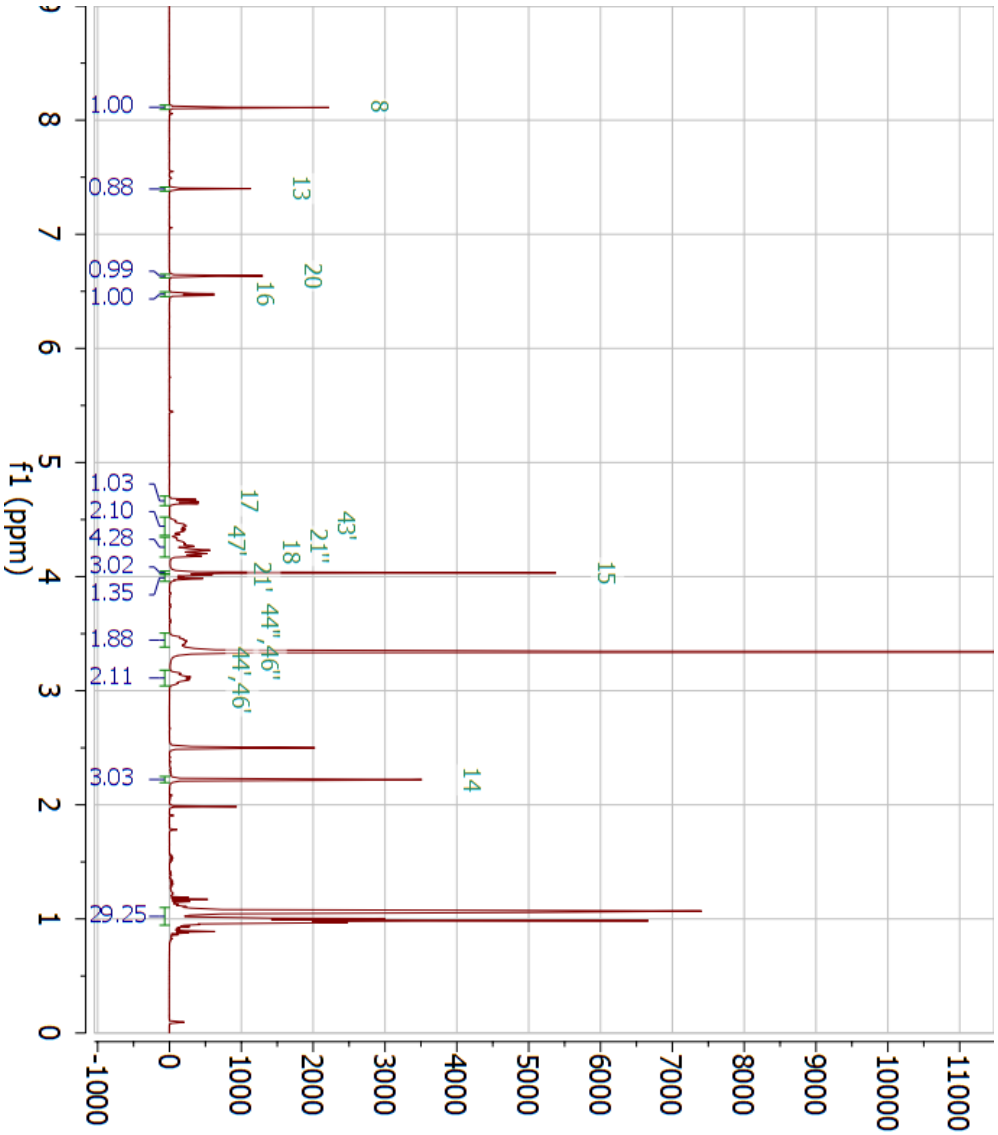

f2 (ppm)
f1 (ppm)

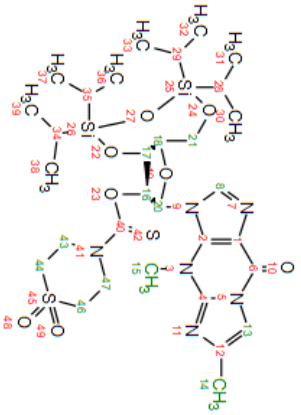

f1 (ppm)
0
1
2
3
4
5
6
7
8
9
0
500
1000
1500
2000
2500
3000
3500
1.00
1.15
1.04
2.41
8.69
4.63
1.17
1.20
0.77
0.61
1.16
1.17
0.63
0.67
3.28
6.80
3.32
4.85
1.21
1.08
4.66
4.83
3.93
9.60
3.72
7.36
35
43
38
57,58,60,61,62,63,65,66
68,69,70,71,72
74,76
8.60
7.59
7.58
7.34
7.32
7.31
7.30
7.29
6.86
6.83
6.82
6.81
3.76
3.74
2.12 HDO
2.07
1.94 CD3CN
1.93 CD3CN
1.92 CD3CN
1.92 CD3CN
1.91 CD3CN
1.23
1.21
1.21
1.20
1.14
1.12
1.11
1.09
1.04
1.03
0.96
0.94
0.94
0.92
0.91
0.90

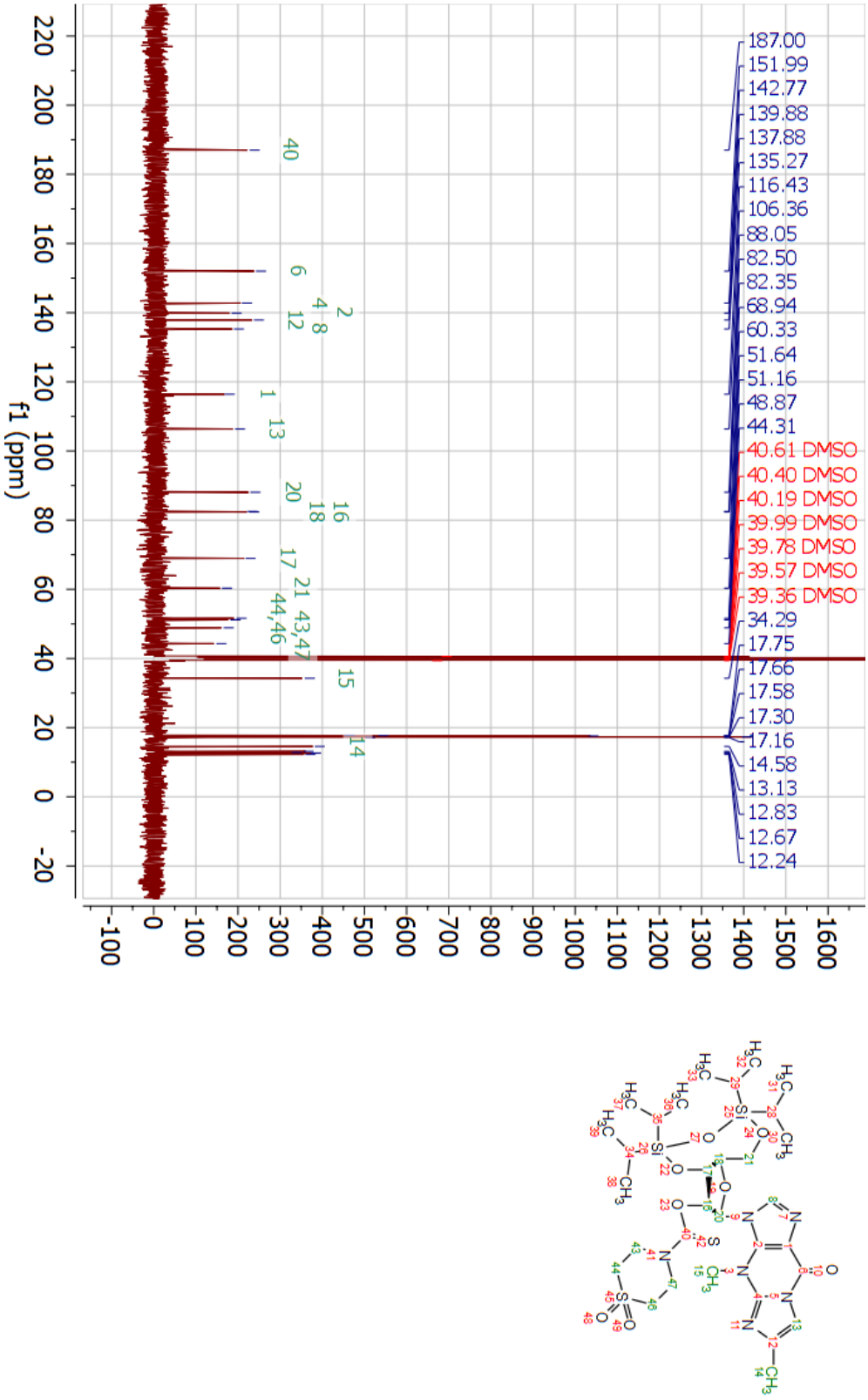


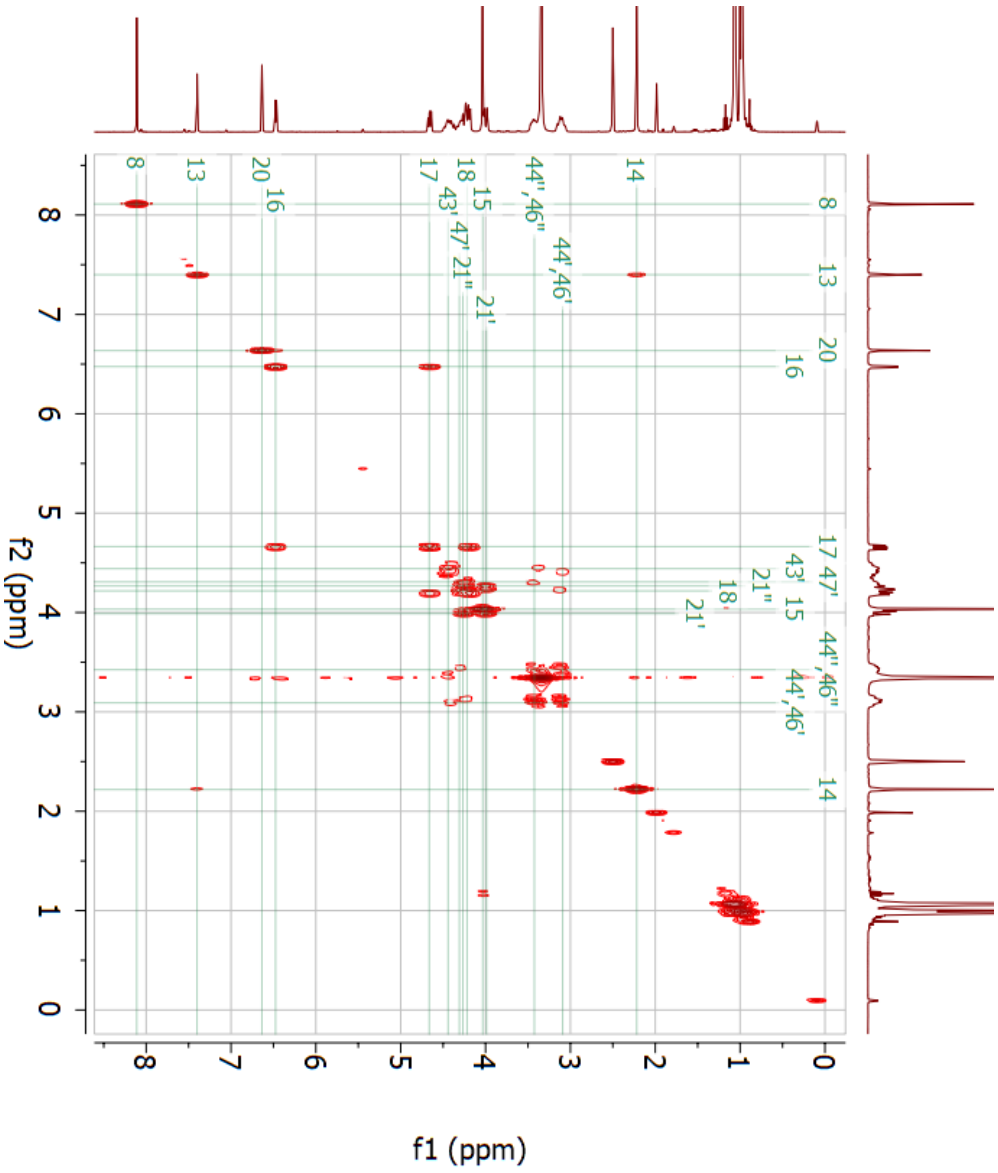

150.47
149.45
13.74
20
20
f1 (ppm)
150
100
50
0
-50
-100
-150
0
50
100
150

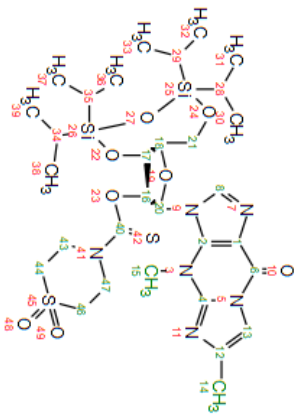


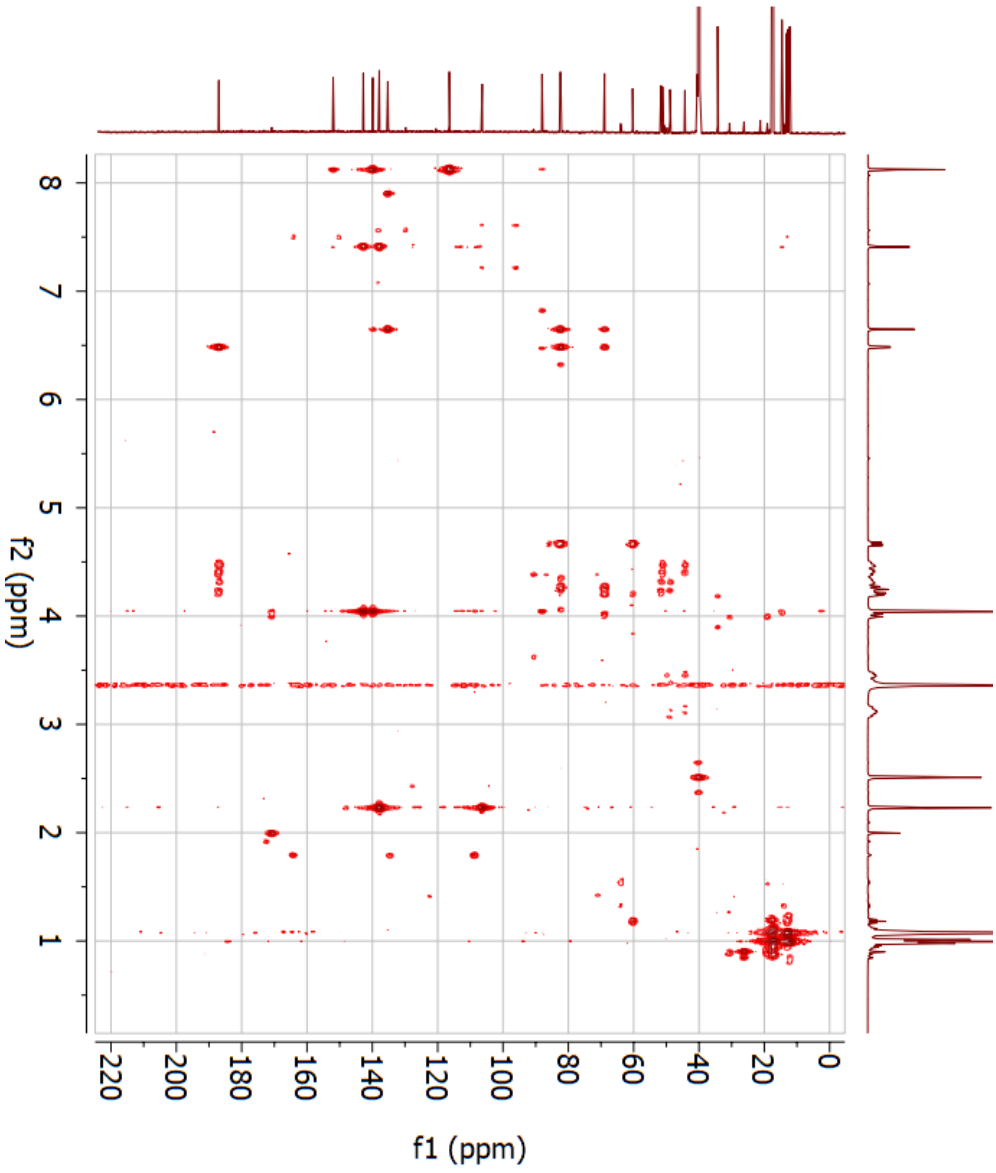

f1 (ppm)
1.00
0.88
0.99
1.00
1.03
2.10
4.28
3.02
1.35
1.88
2.11
3.03
29.25
8
13
20
16
17
47'
43'
21''
18
21'
15
44'',46''
44',46'
14

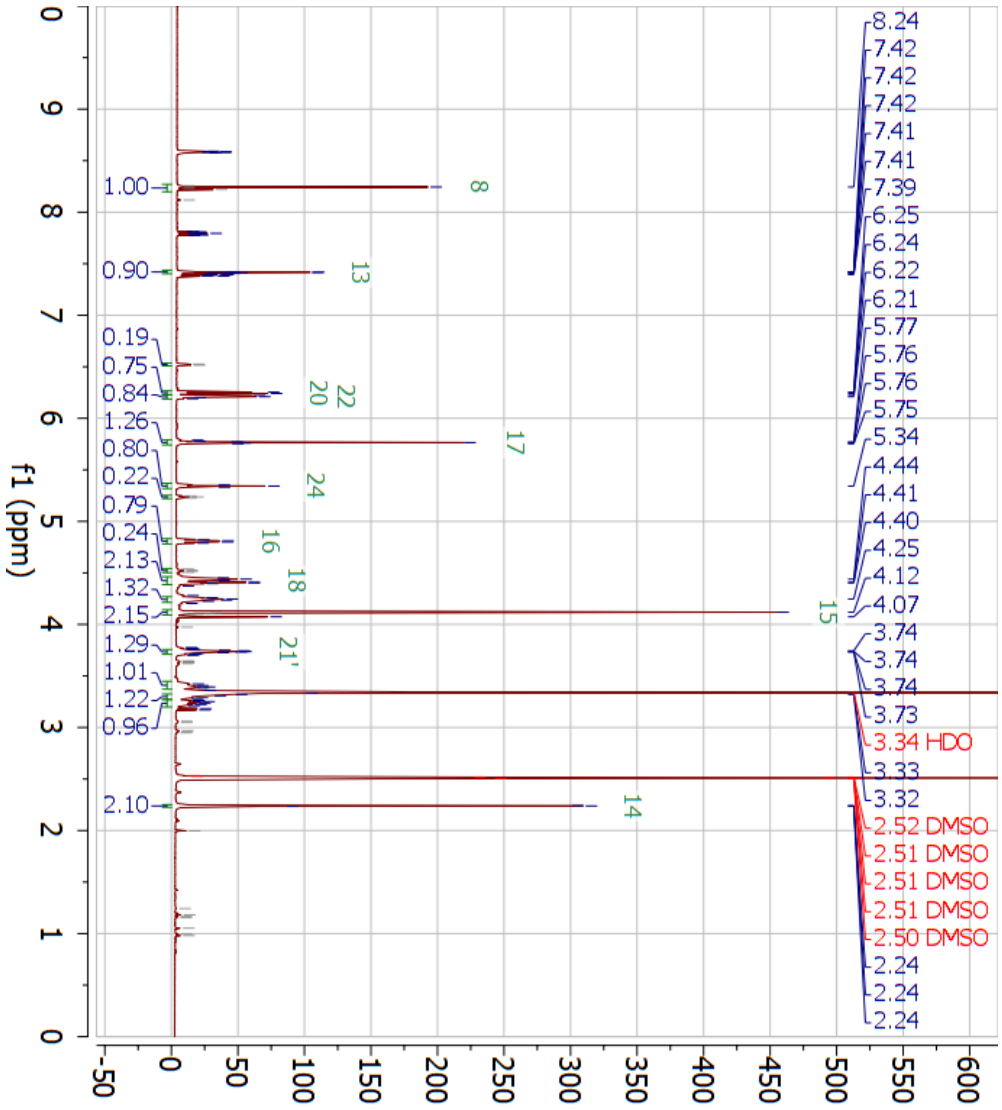


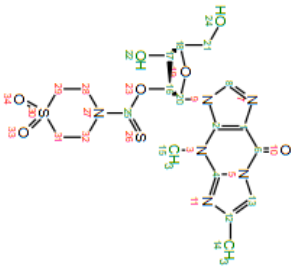

f1 (ppm)
187.00
151.99
142.77
139.88
137.88
135.27
116.43
106.36
88.05
82.50
82.35
68.94
60.33
51.64
51.16
48.87
44.31
40.61 DMSO
40.40 DMSO
40.19 DMSO
39.99 DMSO
39.78 DMSO
39.57 DMSO
39.36 DMSO
34.29
17.75
17.66
17.58
17.30
17.16
14.58
13.13
12.83
12.67
12.24

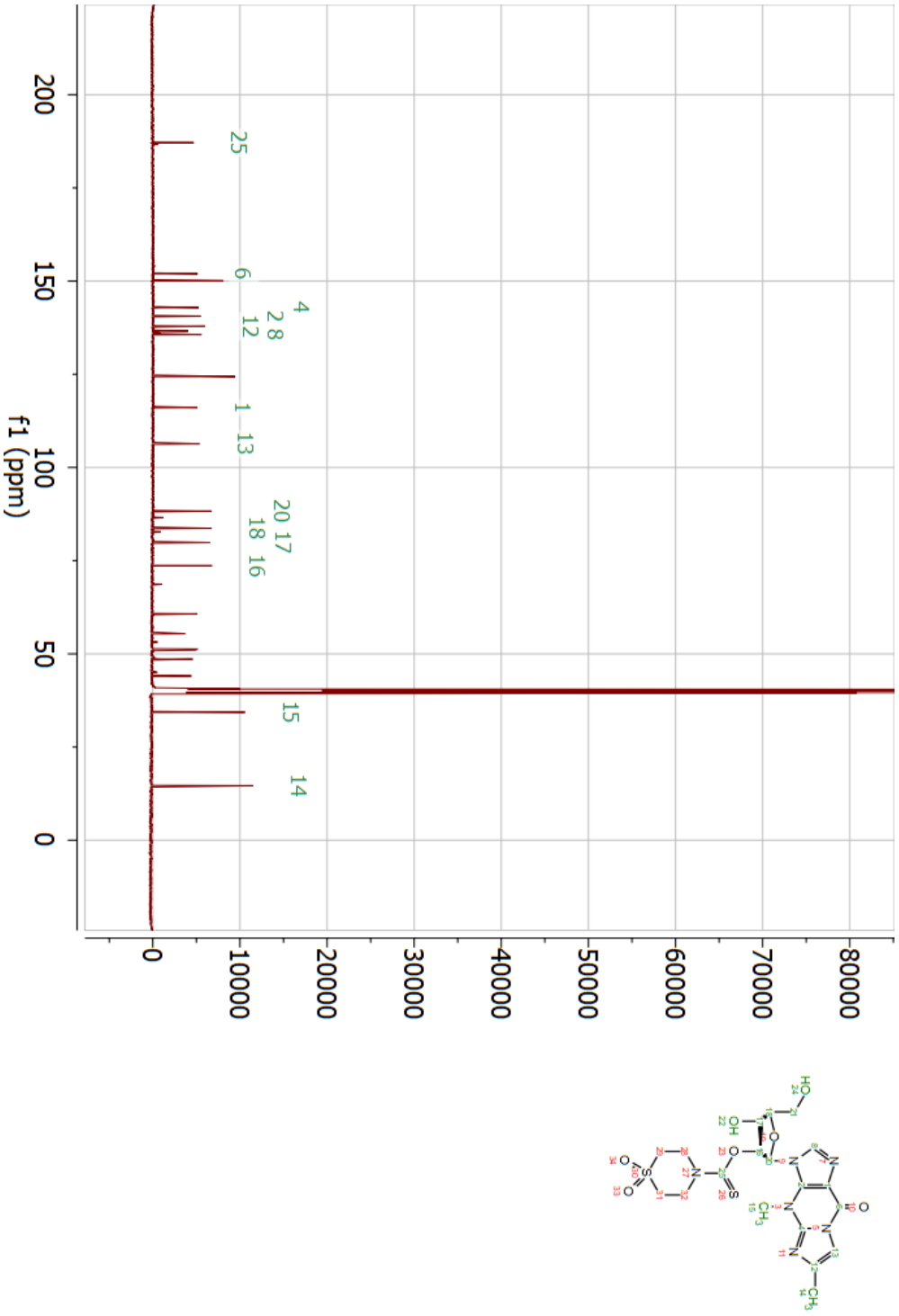

f2 (ppm)
f1 (ppm)

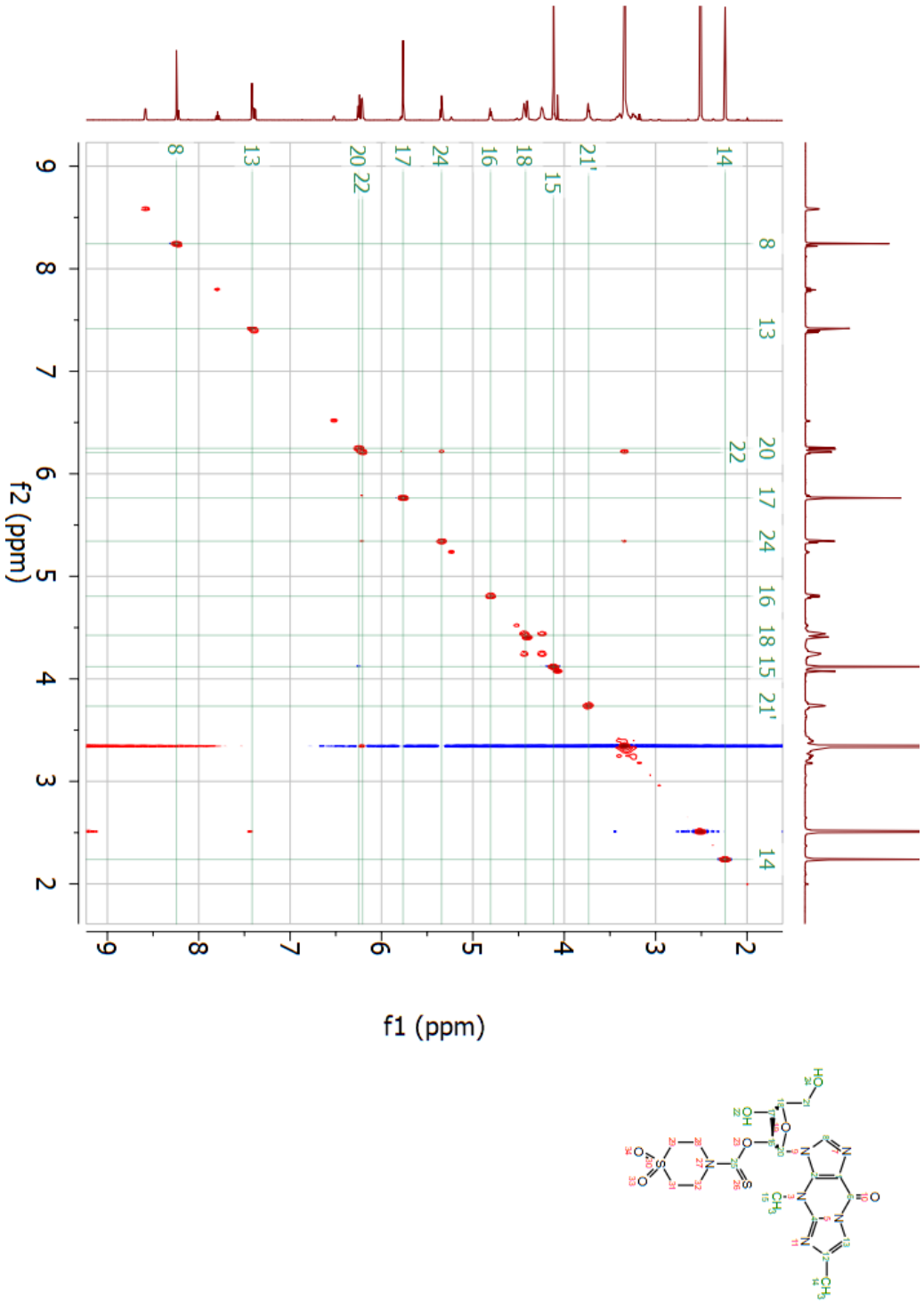


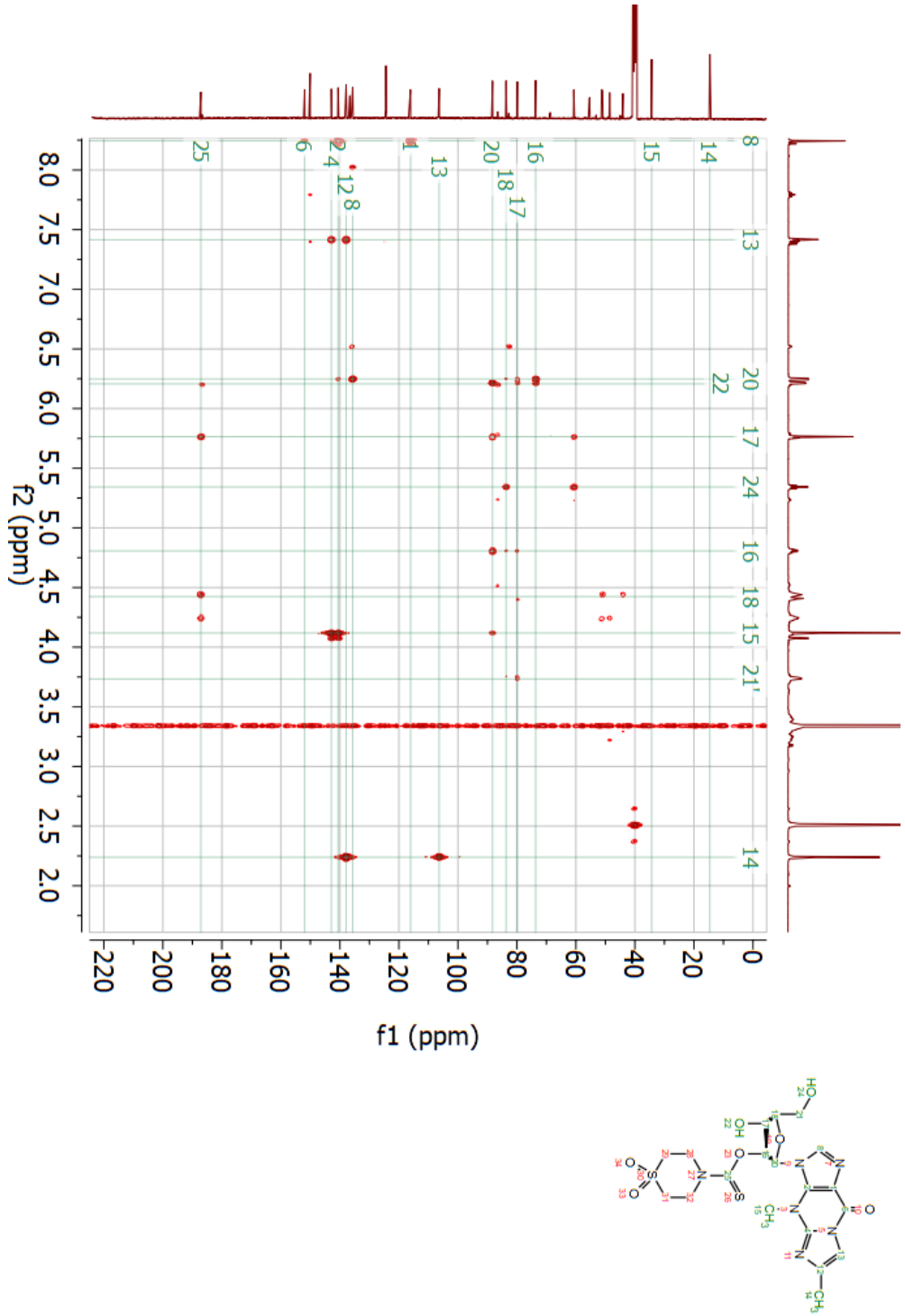

f2 (ppm)
f1 (ppm)

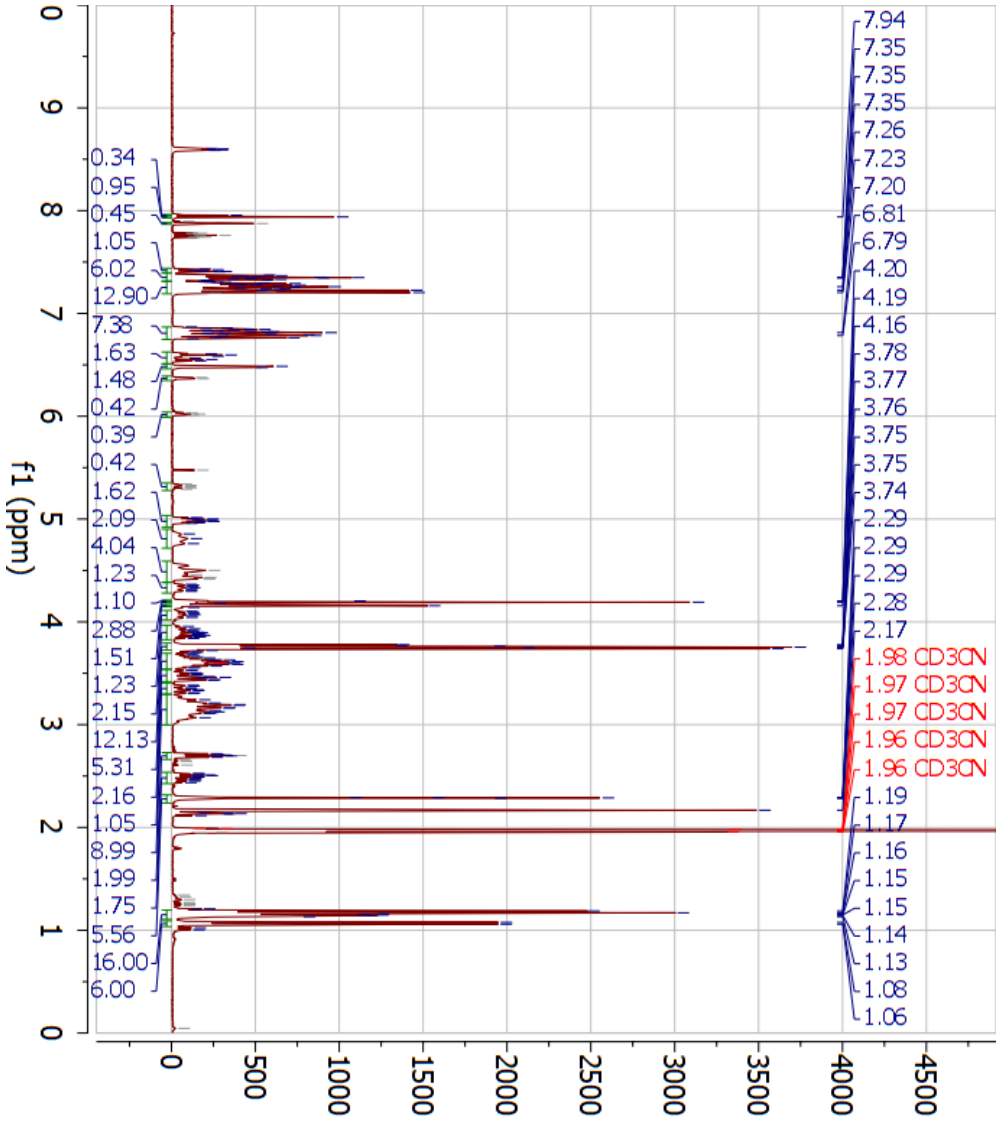

f1 (ppm)
1.00
0.90
0.19
0.75
0.84
1.26
0.80
0.22
0.79
0.24
2.13
1.32
2.15
1.29
1.01
1.22
0.96
2.10
8
13
22
20
17
24
16
18
21'
15
14
8.24
7.42
7.42
7.42
7.41
7.41
7.39
6.25
6.24
6.22
6.21
5.77
5.76
5.76
5.75
5.34
4.44
4.41
4.40
4.25
4.12
4.07
3.74
3.74
3.74
3.73
3.34 HDO
3.33
3.32
2.52 DMSO
2.51 DMSO
2.51 DMSO
2.51 DMSO
2.50 DMSO
2.24
2.24
2.24

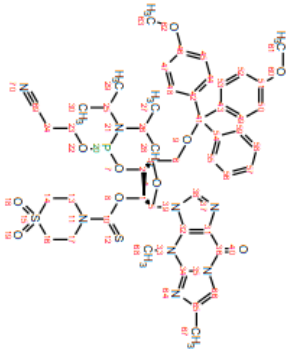


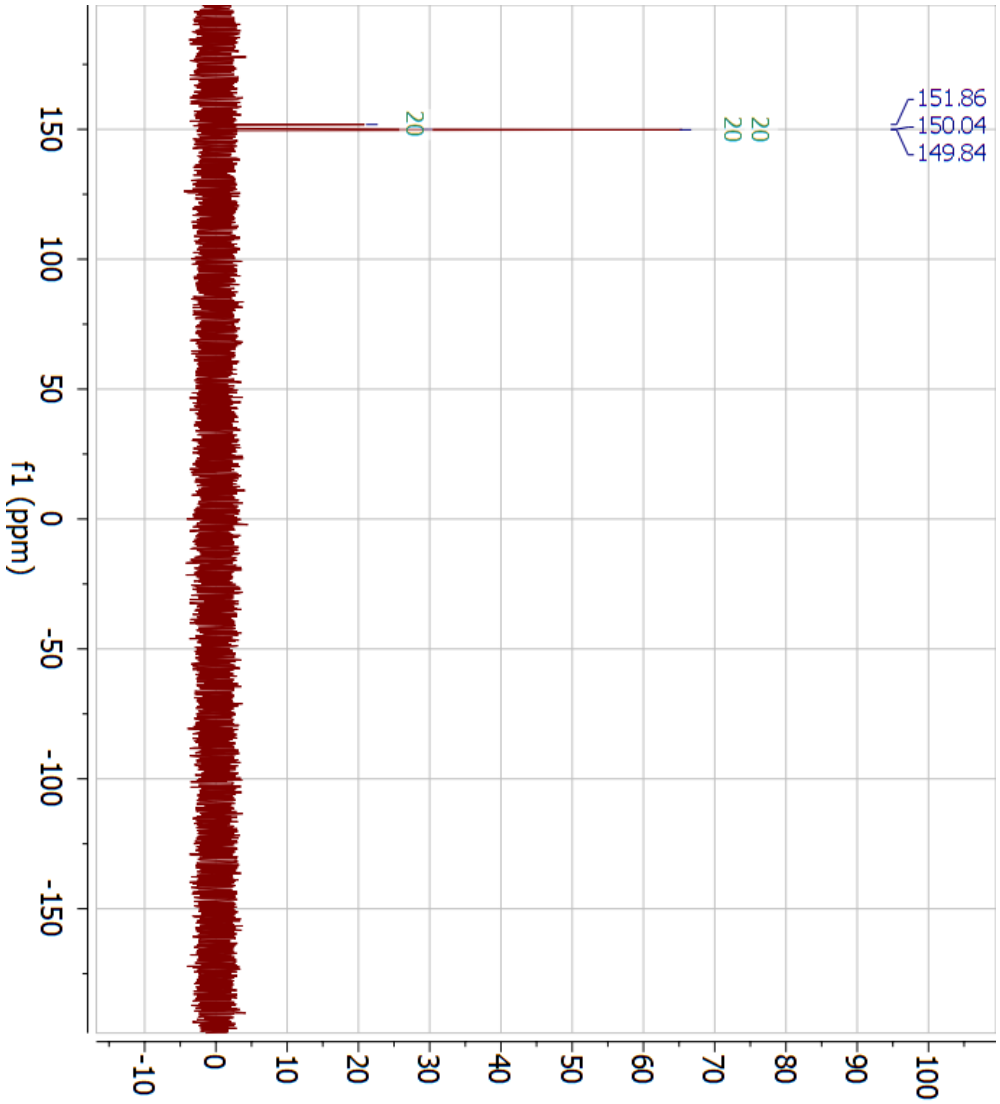

f1 (ppm)
200
150
100
50
0
25
6
4
2 8
12
1
13
20 17
18 16
15
14
0
10000
20000
30000
40000
50000
60000
70000
80000


f2 (ppm)
f1 (ppm)
8
13
20 22
17
24
16
18
15
21'
14


f2 (ppm)
f1 (ppm)


f1 (ppm)
0
1
2
3
4
5
6
7
8
9
0
500
1000
1500
2000
2500
3000
3500
4000
4500
0.34
0.95
0.45
1.05
6.02
12.90
7.38
1.63
1.48
0.42
0.39
0.42
1.62
2.09
4.04
1.23
1.10
2.88
1.51
1.23
2.15
12.13
5.31
2.16
1.05
8.99
1.99
1.75
5.56
16.00
6.00
7.94
7.35
7.35
7.35
7.26
7.23
7.20
6.81
6.79
4.20
4.19
4.16
3.78
3.77
3.76
3.75
3.75
3.74
2.29
2.29
2.29
2.28
2.17
1.98 CD3CN
1.97 CD3CN
1.97 CD3CN
1.96 CD3CN
1.96 CD3CN
1.19
1.17
1.16
1.15
1.15
1.14
1.13
1.08
1.06




151.86
150.04
149.84
20
20
20
f1 (ppm)
150
100
50
0
-50
-100
-150
-10
0
10
20
30
40
50
60
70
80
90
100

