



Development of Site-Specific RNA Labeling Strategies to Probe Alternative RNA Splicing

Tilemachos D. Kosmidis

A thesis submitted to the Department of Pure and Applied Chemistry, University of Strathclyde in
part fulfillment of the requirements for the degree of Doctor of Philosophy

2016

Academic Supervisor: Dr Glenn A. Burley

Industrial Supervisor: Dr Sabine Fenner

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not previously been submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed

Date

Contents

Declaration	ii
Acknowledgments.....	v
Abstract	vii
Abbreviations	ix
1 Introduction.....	1
1.1 Biodiversity – Result of Molecular Complexity	2
1.2 Primary Structure of RNA.....	2
1.3 RNA Secondary Structure	3
1.4 RNA Contribution to Complexity – mRNA Processing	6
1.4.1 5' Capping of mRNA.....	6
1.4.2 Polyadenylation	7
1.4.3 RNA Editing	8
1.5 RNA Splicing	9
1.6 Alternative RNA Splicing	15
1.7 Alternative RNA Splicing and Disease	16
1.8 Splice Site Selection	18
1.9 Fluorescent Labeling of Biomolecules.....	21
1.10 Current State-of-the-Art in RNA Labeling.....	24
1.11 Site-specific RNA Labeling	25
1.12 The Development of Orthogonal Base Pairs – Expanding Nature’s Genetic Code Beyond Watson-Crick	
1.12.1 Base Pairs Developed by the Benner Group.....	30
1.12.2 Base Pairs Developed by the Kool Group	32
1.12.3 Base Pairs Developed by the Hirao Group	33
1.12.4 Base Pairs Developed by the Romesberg Group	36
1.13 Orthogonal Base Pairs in RNA Labeling	37
1.14 Challenges in the Application of Orthogonal Base Pairs for Site-Specific RNA Labeling	41

1.15	Hypothesis to be Tested.....	43
1.16	Aims and Objectives.....	44
2	Synthesis of C-ribonucleoside Analogues as Unnatural Base Pairs Directed by Hydrogen Bonding	
2.1	Introduction	46
2.1.1	The Heck Reaction in the Synthesis of C-deoxyribonucleosides.....	47
2.1.2	Current Challenges in the Synthesis of C-ribonucleosides.....	50
2.2	Aims of this Study.....	52
2.3	Results and Discussion	53
2.3.1	Design of the Synthetic Strategy Involving the Heck Reaction	53
2.3.2	Synthesis of Heck Coupling Partners	54
2.3.3	Investigation of Heck Coupling Conditions Towards Ribo-Z Synthesis	54
2.3.4	Attempts to Prepare Z analogues	57
2.4	Summary.....	58
2.5	Experimental.....	58
3	Establishing a Modular Synthetic Methodology for the Preparation of C6-heteroaryl 2-aminopurines	
3.1	Introduction	67
3.2	Aims of this Study.....	69
3.3	Results and Discussion	70
3.3.1	Retrosynthetic Analysis of C6-heteroaryl 2-aminopurines	70
3.3.2	Synthesis of C6-ethynyl 2-aminopurine	70
3.3.3	Synthesis of C6-triazole Functionalized 2-aminopurines.....	73
3.3.4	Synthesis of C6-isoxazole Functionalized 2-aminopurines.....	80
3.3.5	Synthesis of C6-pyrazole Functionalized 2-aminopurines	83
3.4	Summary.....	88
3.5	Experimental.....	88
4	Synthesis of C6-functionalized 2-aminopurine Nucleoside Triphosphates	
4.1	Introduction	110
4.2	Aims of this Study.....	112
4.3	Results and Discussion	112

4.3.1	Initial Attempts to Synthesize Nucleoside Triphosphates	112
4.3.2	Strategy to Access Pure Nucleoside Triphosphates.....	115
4.4	Summary.....	123
4.5	Experimental.....	124
5	Transcriptional Evaluation of C6-functionalized 2-aminopurine Nucleoside Triphosphates	
5.1	Introduction	139
5.2	Aims of this Study	140
5.3	Results and Discussion	140
5.4	Summary.....	146
5.5	Experimental.....	147
6	Conclusions.....	148
7	References.....	149
	Appendix	158

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Glenn Burley, for giving me this opportunity to work in this project (RNA can indeed do amazing things!) and his advice during these years. Oh and his 30-min long pep-talks that boosted up your confidence! I want to thank my industrial supervisor in GSK Dr Sabine Fenner for her advice on chemistry and the German language and also for her lovely banter. Vielen Dank Sabine! Du bist sehr Nett! Also, special thanks go to Prof Ian Eperon at Leicester, for his patience in teaching a chemist how to do biochemistry experiments.

I would also like to thank Mr Craig Irving and Mrs Patricia Keating for their enormous help with NMR and MS experiments.

In addition, I would like to thank the members of the Burley group (past and present). Especially John May, with whom we started this PhD journey to the unknown together. Also Steven Pauff, for his enormous help and amazing chat. I should also not forget the members of the Tomkinson group (with whom we shared the same office for two years) and the members of the Eperon group.

Finally, a “big thank you” goes to my family and friends in Greece for their support, as well as all the people (that’s a long list to put in here...) I met in Glasgow for this joyful experience during these years.

Abstract

The objective of this project was to develop a method for the labeling of RNA at specific locations using non-natural base pairs. This thesis details the efforts made towards this objective through the development of a modular synthetic platform for an expanded genetic alphabet and its evaluation using transcriptional assays.

Chapter 1 introduces the structure and function of RNA, followed by a description of RNA processing events with emphasis placed on alternative RNA splicing and how aberrations in splicing can lead to disease. The concept of RNA labeling follows and examples from the literature are presented along with challenges associated with these techniques to elucidate key mechanistic splicing concepts. The concept of orthogonal base pairs is then introduced along with representative examples from the literature. The application of this concept in RNA labeling is then presented along with key examples from the literature. The limitations of these methods are highlighted followed by the specific aims behind this project.

Chapter 2 presents the efforts made towards the development of a modular synthetic route for the synthesis of C-ribonucleosides, with a Heck reaction as the key step. Specifically, its application in the synthesis of the **Z** ribonucleoside developed by the Benner group is investigated. The chapter starts with re-introducing the **Z/P** pair developed by the Benner group and highlighting the reasons that placed it as our chosen base pair for RNA labeling. The synthesis of C-deoxyribonucleosides utilizing the Heck reaction is presented, followed by the challenges associated with C-ribonucleoside synthesis.

The syntheses of the alkene and halide coupling partners of the proposed Heck reaction are firstly presented. Investigations on the Heck reaction are then presented, followed by investigations on the transformation of nitro group in **Z** to moieties suitable for further derivatization.

The key outcome of this work is that a synthetic route involving the Heck reaction for the synthesis of C-ribonucleosides is not feasible. This is due to long reaction times required to drive the Heck reaction to completion and difficulties encountered with elucidation of its stereoselectivity. Furthermore, PCR experiments conducted by our collaborators within the Eperon group at the University of Leicester revealed significant misincorporation of **Z** opposite G. Consequently, a change in strategy towards the use of nucleotides developed by the Hirao group was made.

Chapter 3 describes the development of a robust synthetic route for the synthesis of a library of C6-functionalized 2-aminopurines as potential candidates for an expanded genetic alphabet in RNA. The **s/Pa** pair developed by the Hirao group is re-introduced.

The installation of an alkyne functionality on the guanosine scaffold *via* a Sonogashira reaction is described, followed by investigations of [3+2] cycloaddition reactions between the alkyne and azides or aldehyde oximes. Finally, the development of a novel Suzuki-Miyaura protocol for the direct installation of heterocyclic substituents on the guanosine scaffold is also reported.

Key outcomes of this chapter are the following. A robust method was developed for the expedient synthesis of C6-functionalized **s** analogues. This method enabled access to various classes of analogues in three or six steps, including triazoles, isoxazoles, thiophenes and pyrazoles. Attempts to install an azide moiety to the guanosine were partly successful, but the strategy was abandoned due to reaction reproducibility issues. In addition, a C-H activation strategy was not successful on installing an oxazole moiety

Chapter 4 details the efforts towards the synthesis of nucleoside triphosphates based on the **s** analogues described in Chapter 3. This chapter will begin with the presentation of the most common strategies for the synthesis of nucleoside triphosphates. Initial attempts to synthesize triphosphates by global deprotection of nucleosides synthesized in Chapter 3 are next described. Attempts to install a silyl ether group in the 5'-OH are presented, followed by the installation of acetates on the 2'/3' hydroxyls and attempts to protect the exocyclic amine on guanosine as the phenoxyacetamide. The installation of a pyrazole moiety *via* Suzuki-Miyaura protocol using the corresponding boronic acid pinacol ester described in Chapter 3 is presented. The installation of an alkyne moiety is also described and the synthesis of an isoxazole analogue using this alkyne precursor will follow. The chapter will end with the presentation of the synthesis of pyrazole and isoxazole triphosphate analogues. The key outcome of this chapter is that employing different protecting groups on the 5' and the 2'/3' hydroxyls enabled the synthesis of pure triphosphates.

Chapter 5 presents the evaluation of the nucleotide triphosphates synthesized in Chapter 4 regarding their transcriptional efficiency. Key outcome of this work are that high concentrations (>0.5 mM) are needed in order to observe significant incorporation of the analogues, compared to **s**, which needs 0.1 mM for efficient incorporation. At high concentration, the isoxazole moiety exhibits better incorporation efficiency compared to the pyrazole analogue. Furthermore, the addition of 0.5 mM of MnCl₂ resulted in increased incorporation efficiency of the pyrazole analogue, while **s** and the isoxazole exhibited reduced efficiency under these conditions.

Abbreviations

)))	Sonication
°C	Degree Celsius
aq.	Aqueous
ACE	2-O-[bis[2-(acetyloxy)ethoxy]methyl]
APA	Alternative polyadenylation
ATP	Adenosine triphosphate
Boc	<i>tert</i> -Butyloxycarbonyl
CDI	1,1'-carbonyldiimidazole
CFP	Cyan fluorescent protein
cm	Centimeter
COSY	Correlation spectroscopy
CTP	Cytosine triphosphate
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition
d	Doublet
DCE	1,2-Dichloroethane
DCM	Dichloromethane
DEAD	Diethyl azodicarboxylate
DIBAL-H	Diisobutylaluminium hydride
DMA	Dimethylacetamide
DMAP	4-(Dimethylamino)pyridine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
ds	Double-stranded
DTT	Dithiothreitol

e.g.	Exempli gratia
EDTA	Ethylenediaminetetraacetic acid
ESE	Exonic splicing enhancer
ESI-MS	Electrospray ionization-mass spectrometry
ESS	Exonic splicing silencer
equiv.	Equivalent
FAM	Carboxyfluorescein
FRET	Förster resonance energy transfer
GSK	GlaxoSmithKline
GTP	Guanosine triphosphate
HGPS	Hutchinson-Gilford progeria syndrome
HMDS	Hexamethyldisilylazane
HOBt	1-Hydroxybenzotriazole
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
hnRNP	Heterogeneous ribonucleoprotein particle
J	Coupling constant
Kf	Klenow fragment
LC-MS	Liquid chromatography-mass spectrometry
LDA	Lithium diisopropylamide
LMNA	Lamin A/C protein
MALDI	Matrix-assisted laser adsorbtion ionization
Me	Methyl
min	Minute
mL	Millilitres
mol	Mole
Ms	Methanesulfonyl
mRNA	Messenger ribonucleic acid

NHS	N-hydroxysuccinimidyl
NMO	N-methylmorpholine N-oxide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NPE	4-Nitrophenylethyl
Pac	Phenoxyacetyl
PCR	Polymerase chain reaction
PHA	Phytohaemagglutinin
PTB	Polypyrimidine-tract-binding-protein
Ph	Phenyl
Phe	Phenylalanine
ppm	parts per million
PAGE	Polyacrylamide gel electrophoresis
PPN	Tris {bis(triphenylphosphoranylidene) ammonium}
PTSA	p-Toluenesulfonic acid
q	Quartet
RNA	Ribonucleic acid
r. t.	Room temperature
SAM	S-adenosyl methionine
sex	Sextet
s	Singlet
SR	Serine arginine
SMA	Spinal Muscular Atrophy
SMN	Survival of motor neuron
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoprotein
SPAAC	Strain-Promoted Alkyne-azide cycloaddition
t	triplet

Taq	<i>Thermus aquaticus</i>
TAMRA	Tetramethylrhodamine
TBAF	Tetra- <i>n</i> -butylammonium fluoride
TBDMS	<i>Tert</i> -butyldimethylsilyl
TBDPS	<i>Tert</i> -butyldiphenylsilyl
TBTA	Tris[(1-benzyl-1 <i>H</i> -1,2,3-triazol-4-yl)methyl]amine
TCEP	Tris(2-carboxyethyl)phosphine
TEAAc	Triethylammonium acetate
THF	Tetrahydrofuran
TFA	Trifluoroacetic acid
TMS	Trimethylsilyl
TOF	Time of flight
TosMIC	Toluenesulfonylmethyl isocyanide
Trt	Triphenylmethyl
TLC	Thin layer chromatography
tRNA	Transfer ribonucleic acid
UTP	Uridine triphosphate
UTR	Un-translated Region
δ	Chemical shift

Chapter 1

1 Introduction

1.1 Biodiversity – Result of Molecular Complexity

The biomolecule termed RNA plays a central role in gene expression. RNA acts as an intermediary between transcription and translation. RNA also plays a key role in the catalysis of various biochemical processes, such as RNA splicing.^{1,2} Before presenting some of the most common mechanisms responsible for this complexity the structure of RNA will be presented. In higher eukaryotes, RNA is processed before being exported from the nucleus to the cytoplasm. This project is focused on developing tools to probe RNA splicing; an essential RNA processing event that controls protein diversity in eukaryotic cells.³

1.2 Primary Structure of RNA

RNA is a molecule composed of building blocks termed nucleotides. Each nucleotide consists of the same ribose core and a different heterocyclic base. Four different bases are used in RNA: Two purines (A and G) and two pyrimidines (C and U) (Figure 1-1a). These nucleotides form a strand in a 5' to 3' fashion and each nucleotide is connected through a phosphodiester group (Figure 1-1b). The term primary structure refers to the linear sequence of the four nucleotides in RNA (Figure 1-1b).

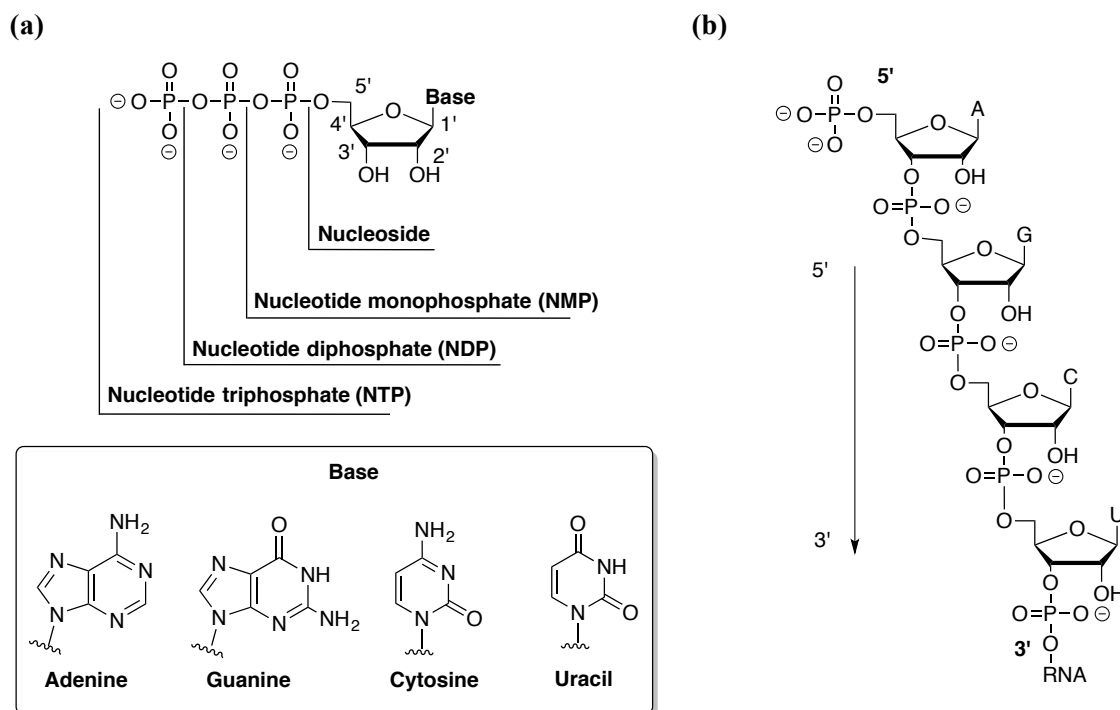


Figure 1-1. (a) Nucleoside and nucleotide structure showing the numbered sugar core and the heterocyclic bases. (b) An RNA tetramer illustrating the primary structure in RNA. A, G, C and U denote the four DNA nucleobases. The arrow shows the 5' to 3' direction.

It should be noted that various types of RNA molecules exist in a cell. This thesis will focus on messenger-RNA (mRNA), which is used as a template for the synthesis of proteins. Since different sequences encode

the synthesis of different proteins, the primary structure of RNA is critical. The fundamental recognition pattern used to encode this is the Watson-Crick base pairing (Figure 1-2). The basis of this is a specific hydrogen-bond profile for each nucleoside, allowing A to pair with U and G with C.

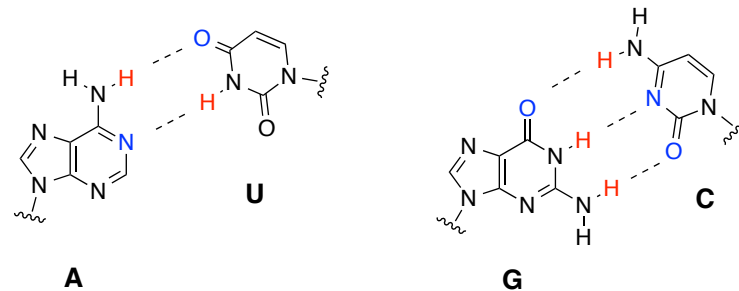


Figure 1-2: Watson-Crick base pairing between nucleotides in RNA. Sugar cores are omitted for clarity. Hydrogen-bond donors are marked in red. Hydrogen-bond acceptors are marked in blue.

1.3 RNA Secondary Structure

Secondary structure describes the 3D folding patterns in RNA. This molecule is highly dynamic and adopts a wide variety of spatial conformations. RNA is usually single stranded and thus, the nucleobases in its sequence can pair with each other through Watson-Crick and non-Watson-Crick interactions. Some of the most common non-Watson-Crick examples include the U/U, G/U, I/U, I/A, I/C and A/G pairs (Figure 1-3).⁴

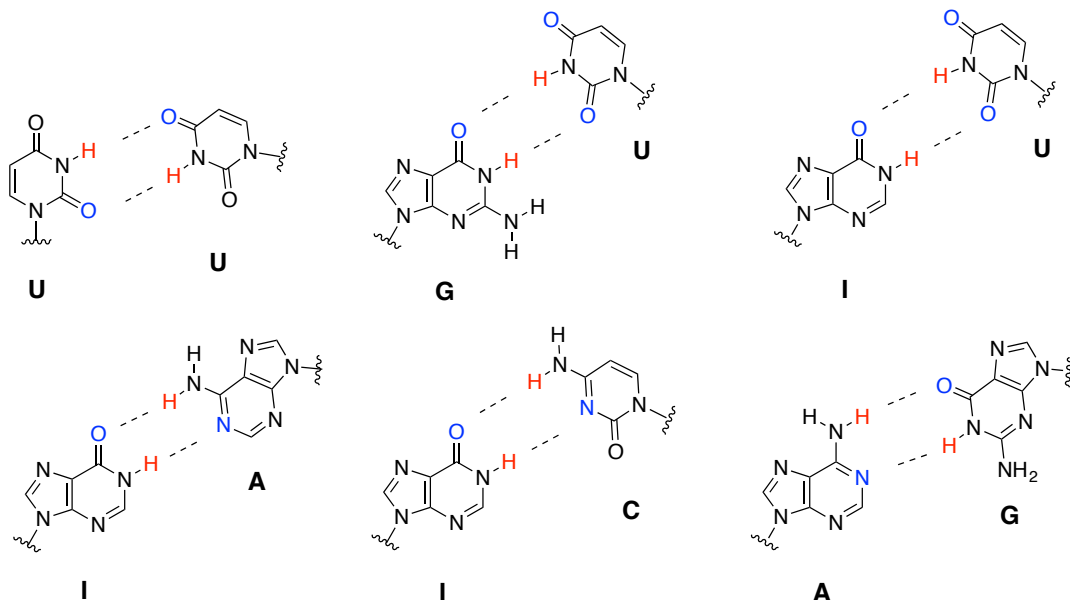


Figure 1-3: Common non-Watson-Crick base pairs encountered in RNA. Ribose sugars have been omitted for clarity. Hydrogen-bond donors are marked in red. Hydrogen-bond acceptors are marked in blue.

Consequently, depending on its primary sequence, RNA can adopt different secondary structures. Below the secondary structures of stem-loop, interior loop and pseudoknot encountered in RNA molecules are

presented to highlight the structural complexity encountered in RNA (Figure 1-4). A U/U base pair is also shown in Figure 1-4a.

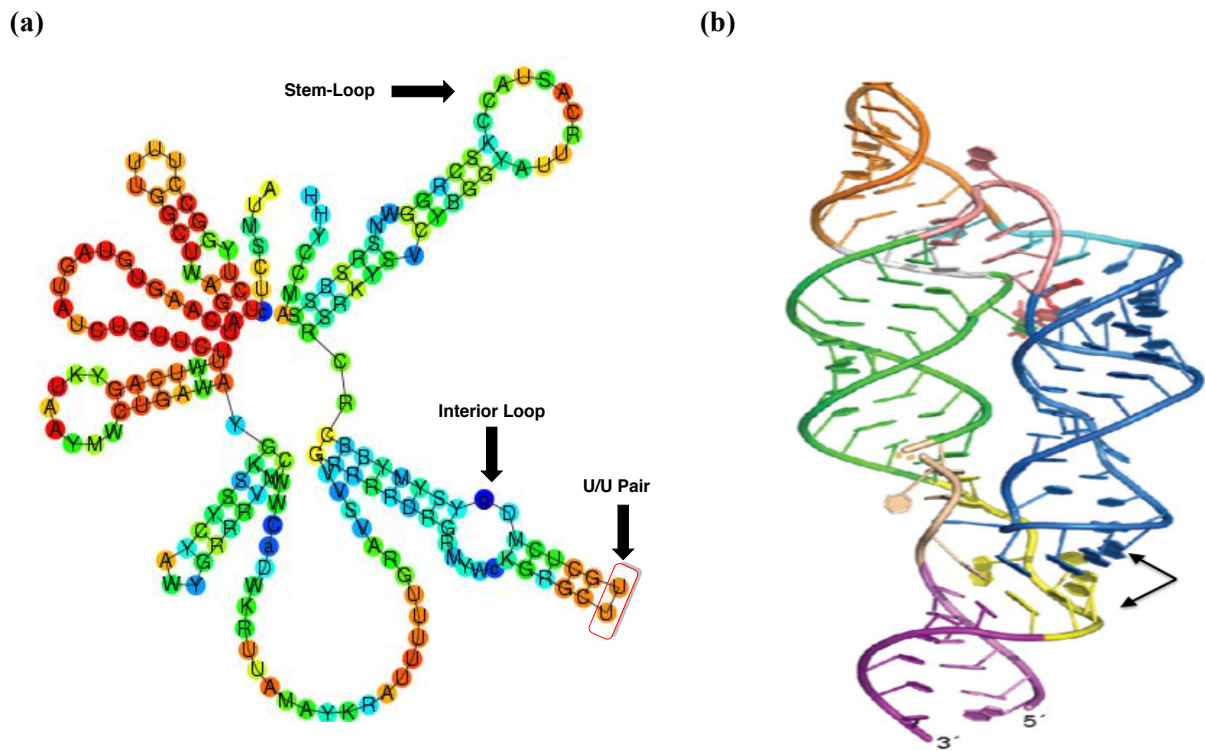


Figure 1-4: (a) Showing stem-loop and interior loop structures in U2 snRNA. The U/U non-canonical base pair is indicated by the red box. (b) Pseudoknot structure shown in a riboswitch. Arrows indicate interactions between two loops creating the pseudoknot.^{5,6} Reproduced with permission from reference 6. Copyright note not needed as stated by the journal’s editors.

In addition, purines feature a property not found in pyrimidines. Aside from Watson-Crick bonds, adenosine can also form hydrogen bonds through its N7 and C6-NH₂, while guanosine through its N7 and the C6 oxygen (Figure 1-5). This side of the heterocyclic base is termed the “Hoogsteen face”.

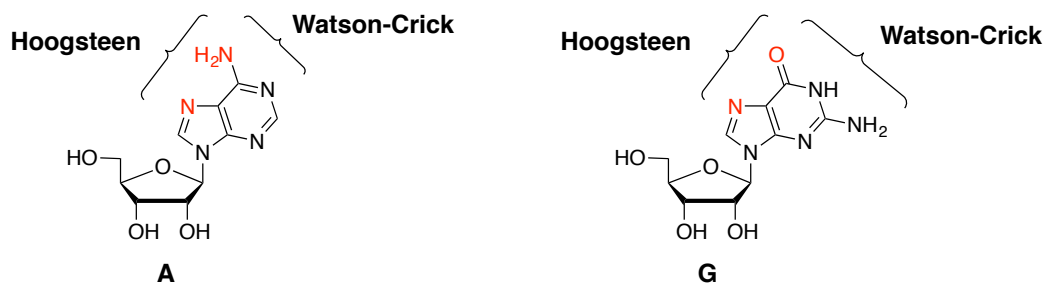


Figure 1-5. The Hoogsteen and Watson-Crick sides in purines. Atoms that can form Hoogsteen interactions are shown in red.

This property allows RNA to further form a variety of complicated structures, such as base triples. In base triples, a purine that participates in Watson-Crick interactions with its cognate nucleotide can further

interact with a purine or pyrimidine through its Hoogsteen face, thus creating a three-base junction. An example of this structure is the G•G/C triple (Figure 1-6).

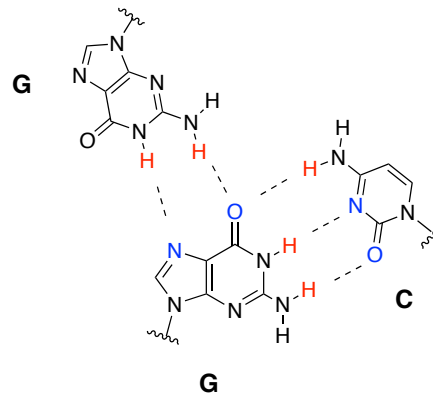


Figure 1-6. The G•G/C triple encountered in RNA. Ribose sugars have been omitted for clarity. Hydrogen-bond donors are marked in red. Hydrogen-bond acceptors are marked in blue.

Furthermore, the 2'-OH imparts unique properties to RNA. The 2'-OH has been demonstrated to aid in intramolecular interactions in tRNA. Substitution of the U33 in tRNA^{Phe} with the 2'-deoxy or 2'-O-Me analogue led to decrease in binding affinity with the 30S ribosomal subunit.⁷ Analogous but quantitatively different results were observed with different tRNAs, suggesting that this 2'-OH is important in aiding intramolecular interactions within each tRNA.⁷

Another example of the unique properties the 2'-OH imparts to RNAs is encountered in the A-motif, where an unpaired adenosine inserts itself between a G/C pair of a neighboring RNA helix (Figure 1-7). This structure forms through four non-Watson-Crick bonds, shown in Figure 1-7 from left to right. (i) Between the 2'-OH of A and the 2'-OH of C. (ii) Between the 2'-OH of A and the carbonyl on the C2 of C. (iii) Between the N1 of A and the amine on the C2 of G. (iv) Between the N3 of A and the 2'-OH of G. This motif has been reported to be extremely abundant in the ribosome (a protein-RNA structure responsible for protein synthesis), where it also aids in interactions with tRNA.⁸

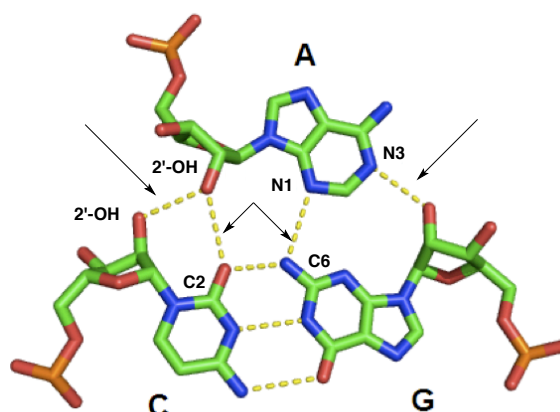


Figure 1-7. An example of the A-motif. Carbon atoms are colored green, nitrogen atoms blue, oxygen atoms red and phosphorus atoms orange. Yellow dotted lines indicate the hydrogen bonds formed. Arrows indicate the non-Watson-Crick bonds.⁹ Permission not needed to reproduce this figure.

A plethora of non-Watson-Crick pairs and base triples are also encountered in splicing, which is the study aim of this project. For example two G/U, three C/U pairs and a G•G/C triple are present in the active site area of the splicing machinery (see Figure 1-20).¹⁰ Furthermore, U/U pairs are also present in U2 snRNA, as shown in Figure 1-4a.

1.4 RNA Contribution to Complexity – mRNA Processing

Most RNA molecules are heavily processed post-synthetically. mRNA processing events occur extensively in eukaryotes and are critical for the proper function of an organism. They are also crucial in enabling the complexity of higher organisms. Firstly, some of these processes influence mRNA properties such as its stability and nuclear export. But most importantly, some result in the production of multiple mRNA molecules from a single mRNA precursor. Since mRNA is the template that an organism uses to synthesize its proteins, these processing events enable higher organisms to perform a wider variety of functions.

1.4.1 5' Capping of mRNA

mRNA capping is the installation of a modified guanosine on the 5' end of a transcript. This process occurs in all eukaryotic organisms and some viruses.^{11,12} Firstly, an RNA phosphatase cleaves the outermost phosphate group of the first nucleotide on the mRNA 5' end, producing a 5' diphosphate group. Next, the enzyme mRNA guanylyltransferase uses guanosine 5' triphosphate and ligates it to the first nucleotide in the transcript through a 5' to 5' triphosphate bridge. Finally, the enzyme mRNA (guanine-N7-)-methyltransferase mediates the methylation of the N7 of the recently installed guanosine (Figure 1-8). This modification protects the RNA against degradation by various enzymes, serves as a promoter for the initiation of protein synthesis and regulates the mRNA export from the nucleus.¹¹

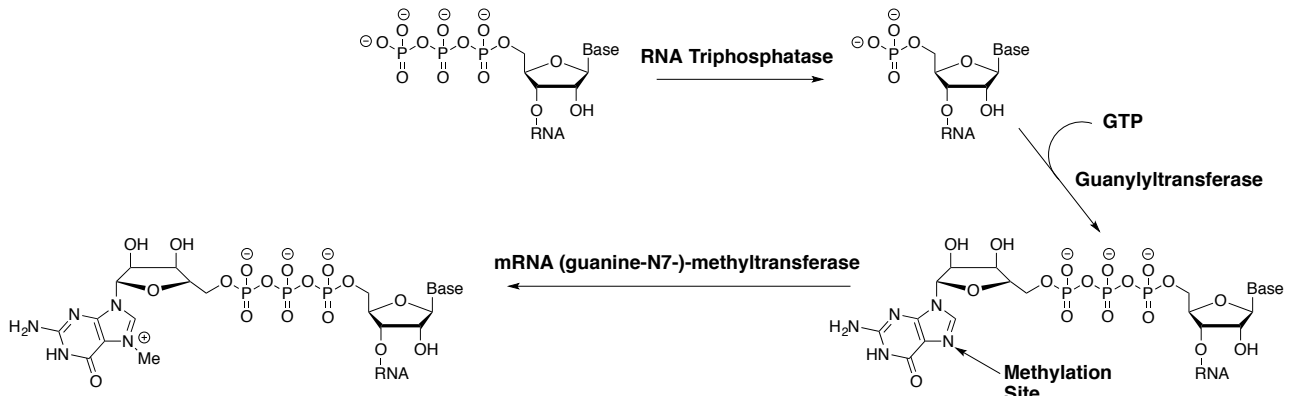


Figure 1-8. Schematic representation of the capping of RNA.

1.4.2 Polyadenylation

Polyadenylation is the ligation of a sequence of adenosines (the poly(A) tail) on the 3' end of an mRNA. This modification occurs in two steps. Firstly, the multi-protein complexes CPSF and CstF bind to two polyadenylation signaling sequences. These sequences are found in the 3' untranslated region of the mRNA (3'-UTR). Commonly encountered signals in humans are the AAUAAA for CPSF and the GU-rich sequence for CstF.^{13,14} This complex cleaves part of the 3'-UTR between the two signals. The enzyme polyadenylate polymerase then starts to add adenosines to the 3' end using ATP as source (Figure 1-9). The length of the poly(A) tail is specific to each species (e.g. ~70-90 adenosines in yeasts), and reaches ~250 adenosines in mammalian cells.^{14,15}

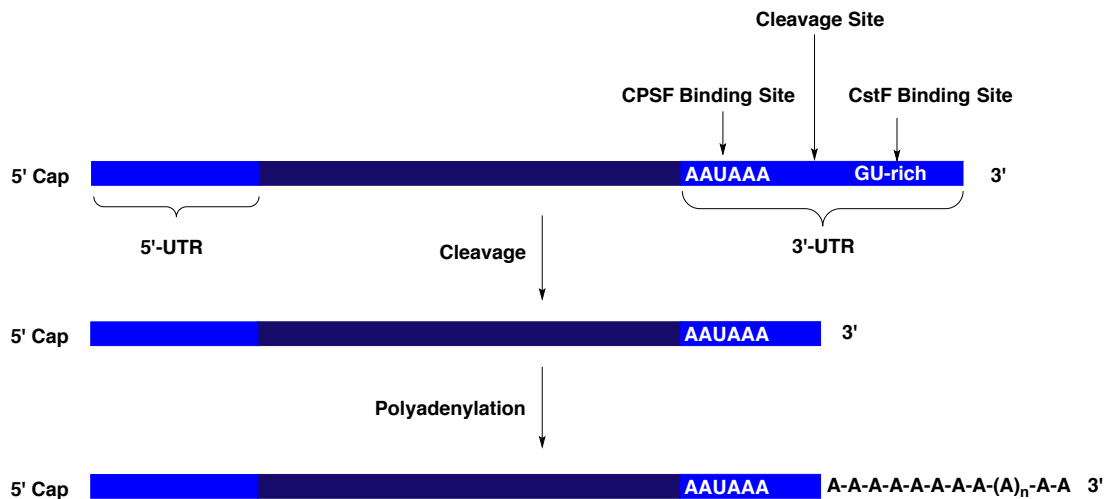


Figure 1-9. Schematic representation of the process of polyadenylation. UTR = Untranslatable region.

The importance of the poly(A) tail can be highlighted by its multiple roles. In nuclear export, enhancing translational efficiency and mRNA stability against enzymatic degradation by various cellular enzymes.^{14,16}

Polyadenylation is a process that occurs both in prokaryotes and eukaryotes; however, mRNAs of eukaryotic cells contain multiple polyadenylation signals. In organisms of lower complexity (such as the yeast *S. Cerevisiae*) 10-15% of genes possess multiple signals, while in most higher eukaryotes, >50% of their genome has been identified to undergo alternative polyadenylation (APA, Figure 1-10).¹⁷⁻¹⁹ Since the length of the 3'-UTR influences properties such as translational efficiency, mRNA stability and mRNA localization, different polyadenylation signals will lead to mRNA transcripts that differ in these properties.²⁰

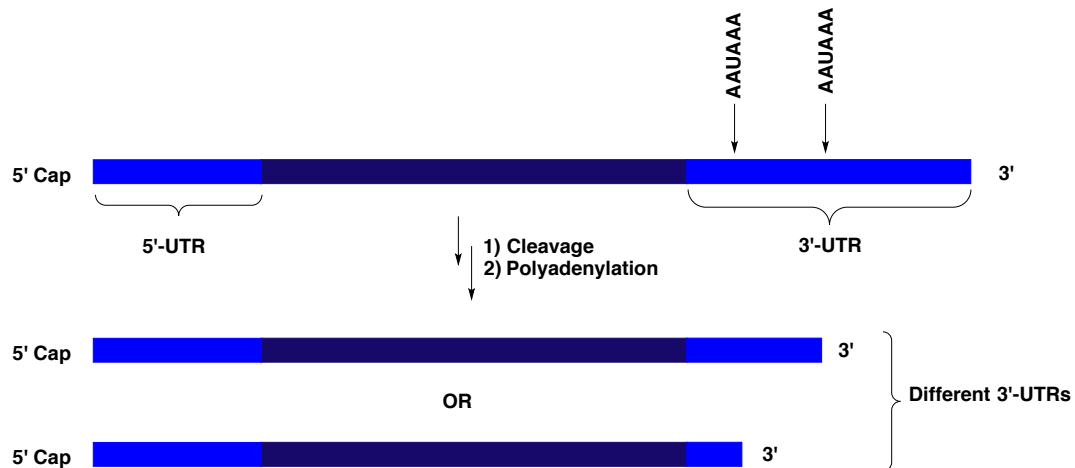


Figure 1-10. Alternative polyadenylation sites lead to mRNAs with different 3'-UTRs. UTR = Untranslatable region.

1.4.3 RNA Editing

This process can occur through multiple modes. Four of the most characterized are addition or deletion of a U residue and deamination of C to U or of A to I (Inosine).²¹ RNA editing is an often phenomenon and has multiple effects such as production of different proteins, tRNAs and miRNAs (miRNAs are RNA molecules involved in the downregulation of protein synthesis and it has been reported that ~16% of human miRNAs undergo RNA editing).²² An exemplar of RNA editing will follow. Deamination of C to U has been discovered to occur in the mRNA that codes for the protein apoB. This mRNA contains the nucleotide triplet CAA at position 6666-6669.²³ The protein synthesized based on this mRNA is the full-length apoB100. During RNA editing, a multi-protein complex called the editosome changes the triplet from CAA to UAA. This results in a premature stop signal during protein synthesis and yields a truncated protein called apoB48 (Figure 1-11). Both of these proteins are part of large protein complexes called lipoproteins. Lipoproteins that are associated with apoB100 increase susceptibility to atherosclerosis, while lipoproteins associated with apoB48 are not. This specific case of RNA editing is also tissue-specific; apoB100 is found in the liver and apoB48 is found in the small intestine. This observation highlights the importance of RNA processing events in the development of cells with tissue-specific roles.

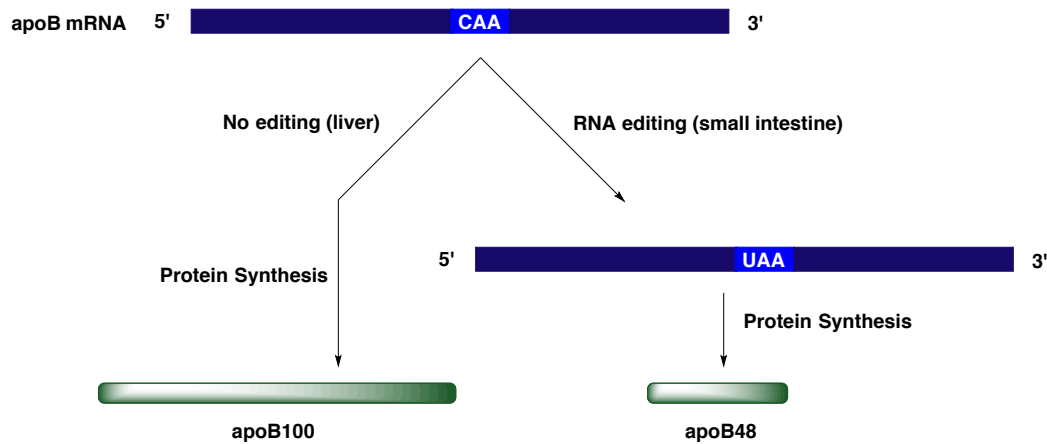


Figure 1-11. RNA editing in the expression of the apoB mRNA.

1.5 RNA Splicing

The pre-mRNAs synthesized in each eukaryotic cell undergo various processing events before exerting their biological role. One of the main processing events is the process of splicing. pre-mRNAs are composed of sections named exons and introns. Exons are sequences used to code for the protein's primary sequence, whereas introns are excised prior to translation (Figure 1-12).²⁴ Introns do serve other purposes though, for example increasing gene diversity through the process of alternative RNA splicing (see Section 1.6).²⁵ Splicing is an important process for proper cell function in higher eukaryotes. Errors in splicing affect protein synthesis and can lead to a variety of diseases ranging from muscular dystrophy to cancer.²⁶ Elucidating the key concepts of splicing and how aberrations in the splicing machinery lead to disease can provide new targets for therapeutic intervention.

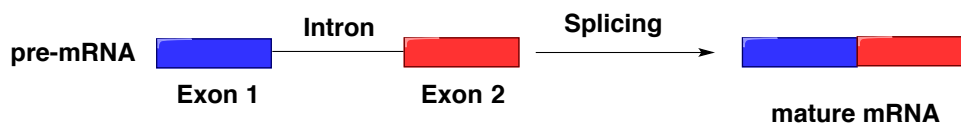


Figure 1-12. Schematic representation of the splicing process.

Figure 1-13 illustrates the key steps during splicing. Three key nucleotides are involved: (i) the terminal nucleotide in the 5' exon (5' splice site), (ii) an adenosine nucleotide present in the intron (the branch point) and (iii) the starting nucleotide in the 3' exon (3' splice site). During splicing, the branch point is brought in close proximity with the 5' splice site and the 2'-OH of an adenosine attacks the phosphodiester backbone. The 5' exon is excised out and then attacks the cyclic intermediate formed to furnish the mature mRNA and the cyclic lariat structure containing the intron.

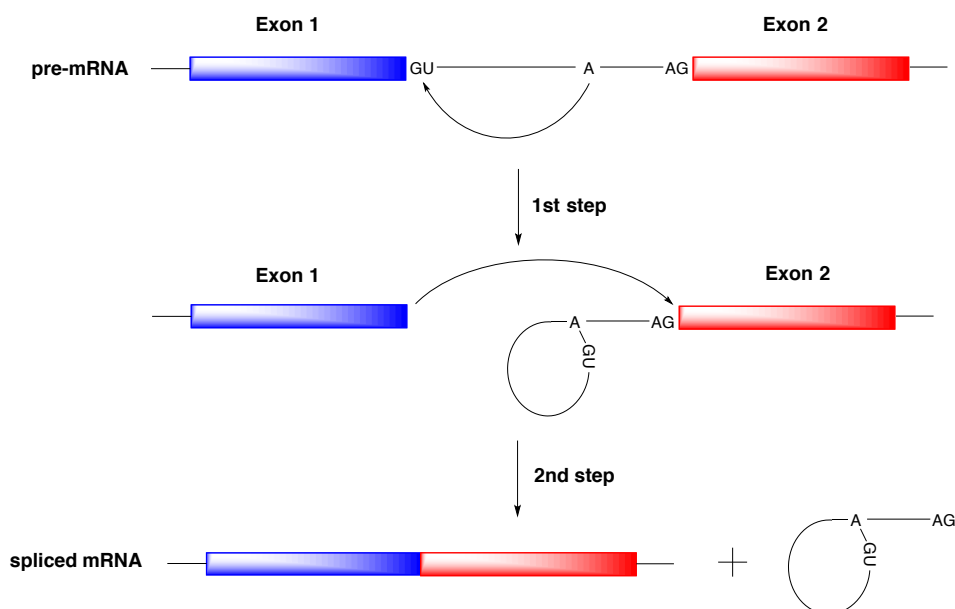


Figure 1-13. Key steps of the splicing mechanism.

Figure 1-14 gives an insight into the chemical mechanism. Splicing occurs through a sequence of two transesterification reactions. The 2'-OH of the Adenosine branch point (marked in green) attacks the phosphodiester bond between the 5' exon and the intron (marked in blue), thus displacing the 5' exon and forming a lariat structure. The 5' exon then attacks the phosphodiester bond between the intron and the 3' exon (marked in red). Consequently, the exons are joined together and the cyclic lariat intron is displaced.

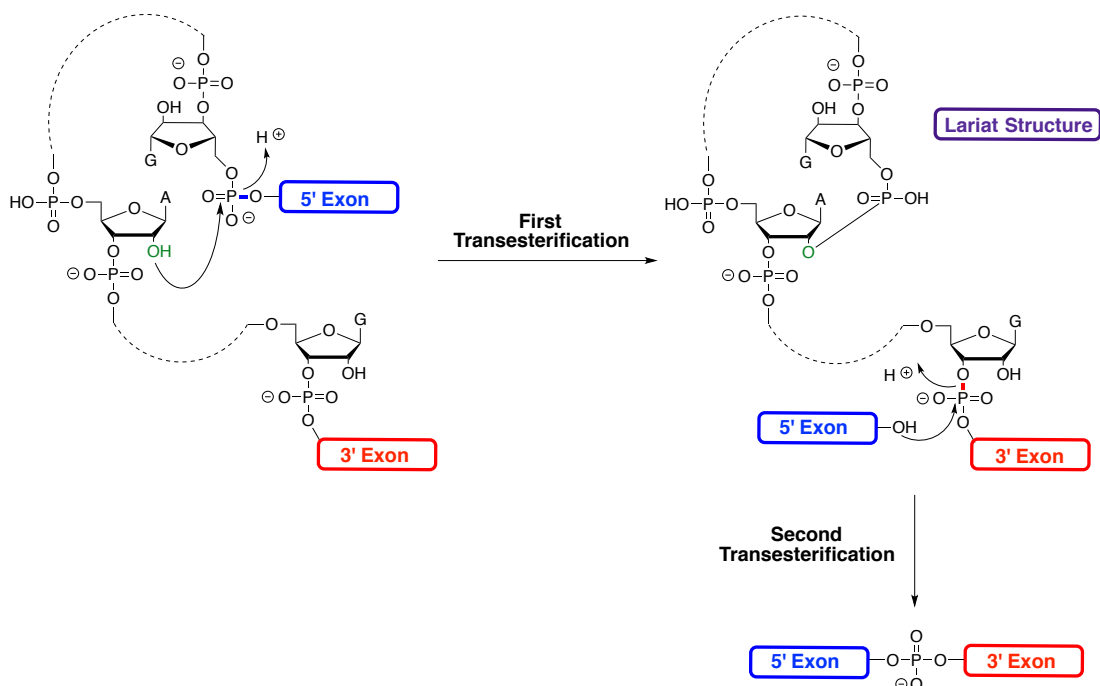


Figure 1-14. Chemical mechanism of the transesterification process occurring in splicing.

Splicing is catalyzed by the spliceosome (in some cases, introns auto-catalyze their excision out of the pre-mRNA), a macromolecular complex guided by five snRNPs (assemblies of snRNAs with proteins).²⁷ The five snRNPs that form the core of the spliceosome are U1, U2, U4, U5 and U6 and their role in splicing is summarized in Figure 1-15.

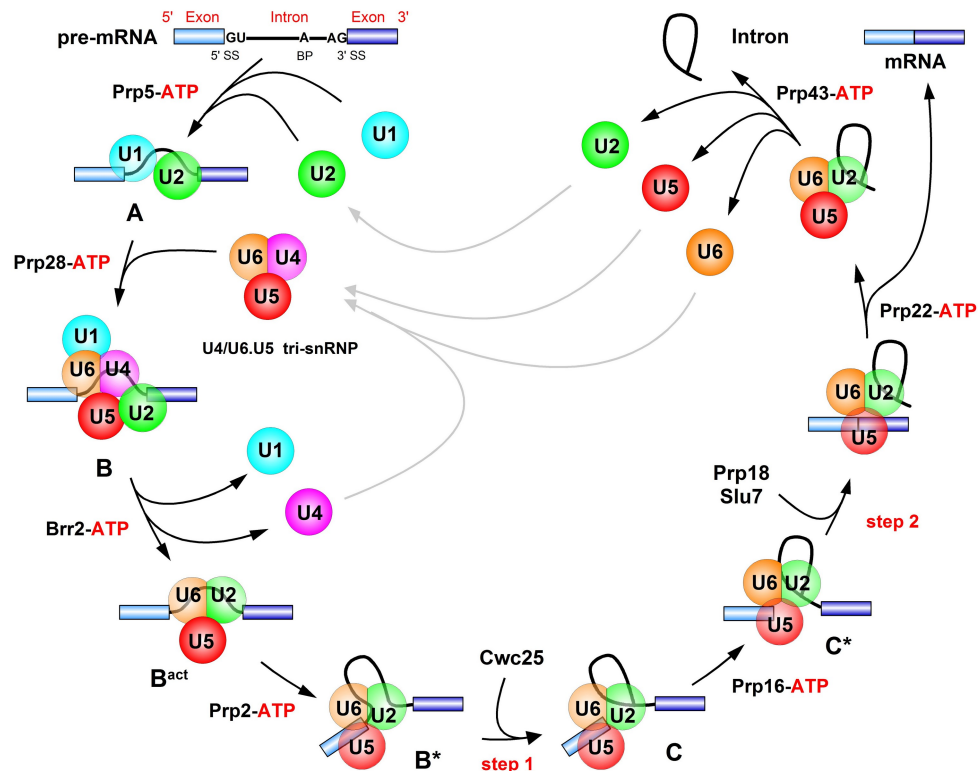


Figure 1-15. The RNA splicing catalytic cycle.²⁸ Permission not needed to reproduce this figure.

The process begins with the interaction of snRNP U1 and U2 with the pre-mRNA strand to form complex A. In complex A, U1 recognizes the 5' splice site and U2 binds to a specific point in the intron, known as the adenosine branch point. U2 causes the intron to form a bulge (Figure 1-15).^{29,30} Complex A interacts with the hetero-trimer of snRNP U4, U5 and U6 to form complex B. U1 and U4 are then ejected out of the spliceosome to take part in another splicing cycle; subsequently, complex B is activated, U6 binds to U2 thus forming the catalytic site, which brings the 5' and the branch point in close proximity.^{29,30} Then, the 2'-OH of the adenosine branch point attacks the 5' splice site, forming a cyclic lariat structure and creating complex C. Activation to C* brings the lariat structure and the 5' exon in close proximity. Finally, the last nucleotide (5' splice site) in the 5' exon attacks the 3' exon in the 3' splice site to create the mature mRNA.^{29,30} Afterwards, dissociation of the U2-U5-U6 trimer occurs, which breaks down to the three separate snRNP allowing those to take part in another cycle. The mature mRNA strand is then transported out of the nucleus and can be eventually translated into a protein.

In the following figures, the structures of the key spliceosomal complexes A, B, B* and C will be presented to further underscore the importance of RNA-RNA interactions present during this process.

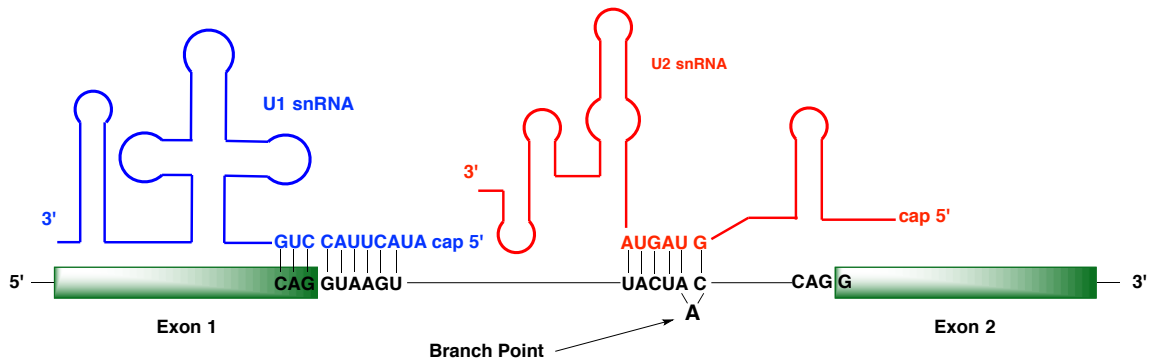


Figure 1-16. Key interactions between U1/U2 snRNAs and the pre-mRNA in complex A.

Figure 1-16 illustrates key interaction formed between the U1/U2 snRNAs and the pre-mRNA in complex A. U1 pairs through its 5' terminus with the 5' splice site. This facilitates the assembly of various other spliceosomal components; thus enhancing 5' splice site recognition. Subsequently, U2 binds to the adenosine branch point region. As shown, U2 snRNA is complementary to the pre-mRNA on either side of the adenosine branch point, but doesn't pair with it. This causes the adenosine to be flipped out of this RNA-RNA duplex, making its 2'-OH accessible and ready to perform the first transesterification reaction (see Figure 1-15).³⁰ This complex is then ready to recruit the U4/U5/U6 tripartite and form complex B. Key interactions in this complex are highlighted in Figure 1-17.

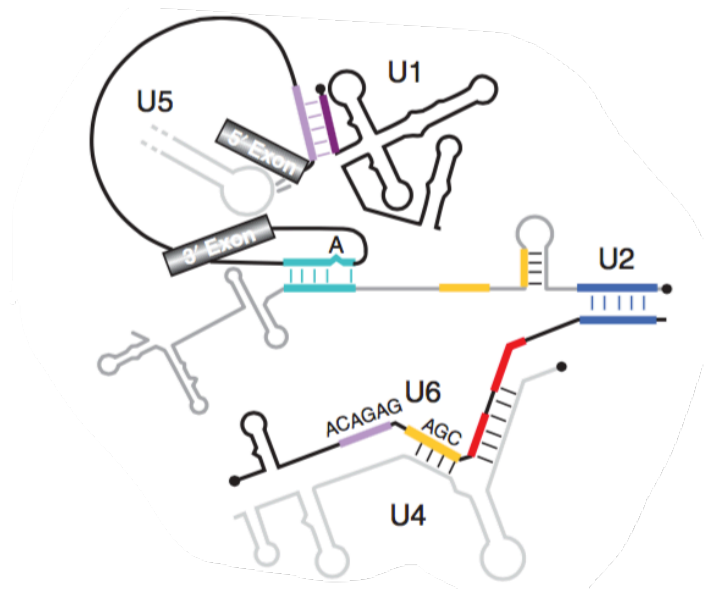


Figure 1-17. Key interactions between the spliceosome and the pre-mRNA in complex B.³⁰ Figure reproduced with permission from reference 30. Copyright 2011 Cold Spring Harbor Laboratory Press.

In complex B, the U4/U6 snRNA components pair extensively with each other. Furthermore, the 3' end of the U6 snRNA pairs with a complementary sequence found on the 5' of the U2 snRNA and a stem-loop formed in U5 snRNA positions itself between the 5' and 3' exons and pairs near the 5' splice site. Complex B is short-lived and rearranges itself displacing U1 and U4, forming the catalytically active complex B.

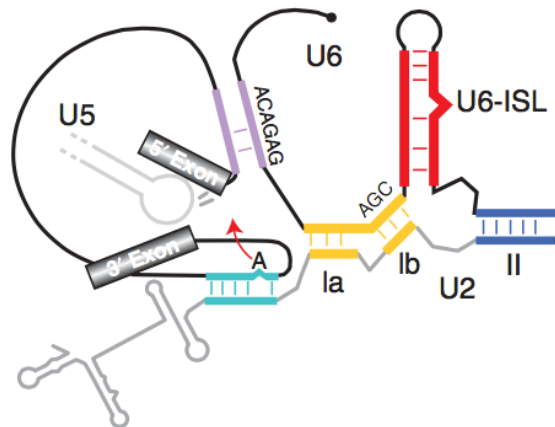


Figure 1-18. Key interactions between the spliceosome and the pre-mRNA in activated complex B.³⁰ Figure reproduced with permission from reference 30. Copyright 2011 Cold Spring Harbor Laboratory Press.

In the activate form of complex B, extensive base-pairing between U6 and U2 snRNA forms a “clamp”, positioning the 5' splice site towards the adenosine branch point. U2 maintains the interactions described in Figure 1-17 and forms three helical structures Ia, Ib and II. U6 binds to the pre-mRNA through its 5' end, while forming an intramolecular stem-loop (U6-ISL). The precise role of each of these structures is currently under debate, but data have suggested that these conformations are highly dynamic and change depending on the splicing step.^{30–32} Likewise, a highly dynamic structure is present in U2. Yeast U2 snRNA forms two mutually exclusive stem structures, IIa and IIc. It has been reported that U2 toggles between each structure, depending on the splicing step.^{30,33,34} Formation of stem IIa promotes the interaction between the branch point and U2. Conformational change of U2 results in stem IIc formation, which promotes the first catalytic step, while for the second catalytic step, U2 switches back to IIa (Figure 1-19). These data further underscore the conformational flexibility of RNA and its importance in mediating complex processes such as RNA splicing.

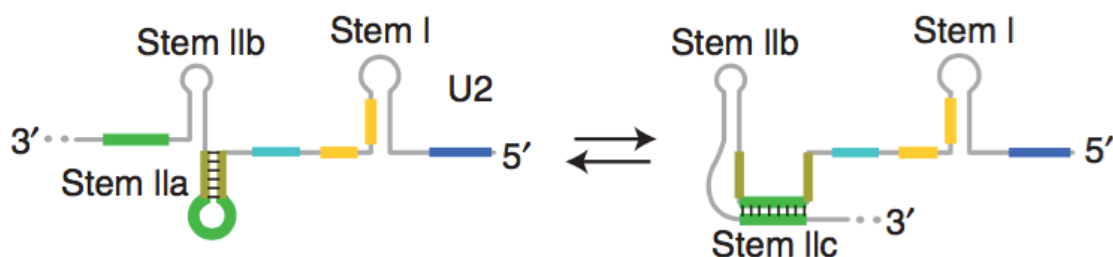


Figure 1-19. Conformational toggling of the stem IIa and IIc structures in yeast U2 snRNA.³⁰ Only the 5' end is shown for clarity. Figure reproduced with permission from reference 30. Copyright 2011 Cold Spring Harbor Laboratory Press.

After the first step, the spliceosome undergoes further rearrangement, forming complex C. A highly detailed structure of its active site has been recently reported by Nagai et al. and is presented in Figure 1-20.¹⁰

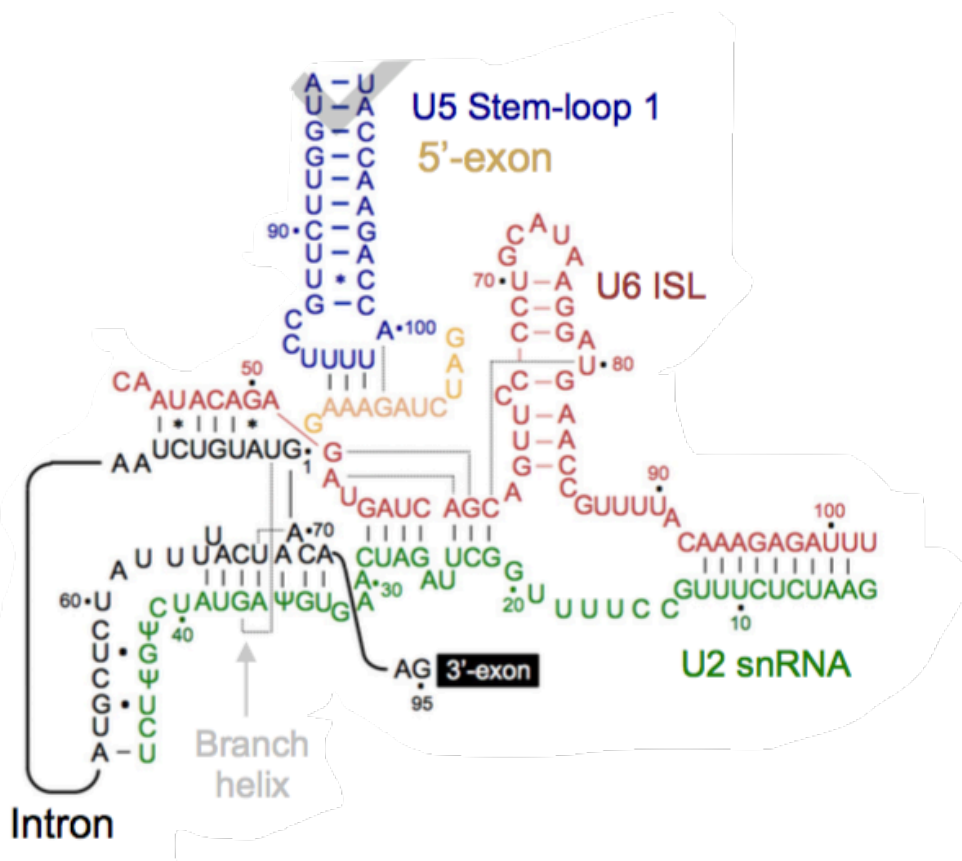


Figure 1-20. Detailed structure of the active site in complex C showing RNA-RNA interactions. Solid black and red and blue lines indicate Watson-Crick pairs. Dots and stars indicate non-canonical base pairs. Solid grey lines indicate base triples formed across the complex.¹⁰ Reproduced with permission from reference 6. Copyright 2016 Nature.

An extensive network of RNA-RNA interactions is present in this complex. Firstly, the 5' exon is still bound by the U5 stem-loop through three A-U pairs. Furthermore, many base triples are present in the active site, formed through non-canonical interactions (presented as solid grey lines). The adenine base of A70 interacts with both the sugar and the base of U68, while uridine 2 pairs with G37 of the U2 snRNA. According to Nagai et al., the U6 ISL forms three base triples with itself that stabilize the RNA catalytic site (both catalytic steps are performed by U6 snRNA and not by spliceosomal proteins).^{1,10} Subsequent activation of this complex results in the ligation of the 5' and 3' exon. A highly detailed structure has not been reported at the time this thesis was written but it has been proposed that U5 binds through its stem loop with the 5' and 3' exon, bridging them and enabling the second transesterification reaction (Figure 1-21).^{35,36}

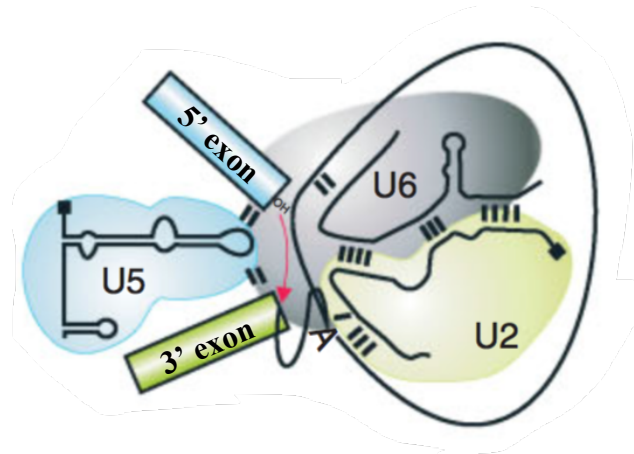


Figure 1-21. Proposed structure of activated complex C. The U5 stem loop bridges the 5' and 3' exons.³⁷ Adapted with permission from reference 37. Copyright 2012 Wiley.

1.6 Alternative RNA Splicing

Alternative RNA splicing occurs in higher eukaryotes. In alternative splicing, the same pre-mRNA can produce different mature mRNA sequences and consequently, different protein isoforms.¹⁵ This increases the diversity of proteins, which can be produced. A schematic of this process is presented in Figure 1-22.

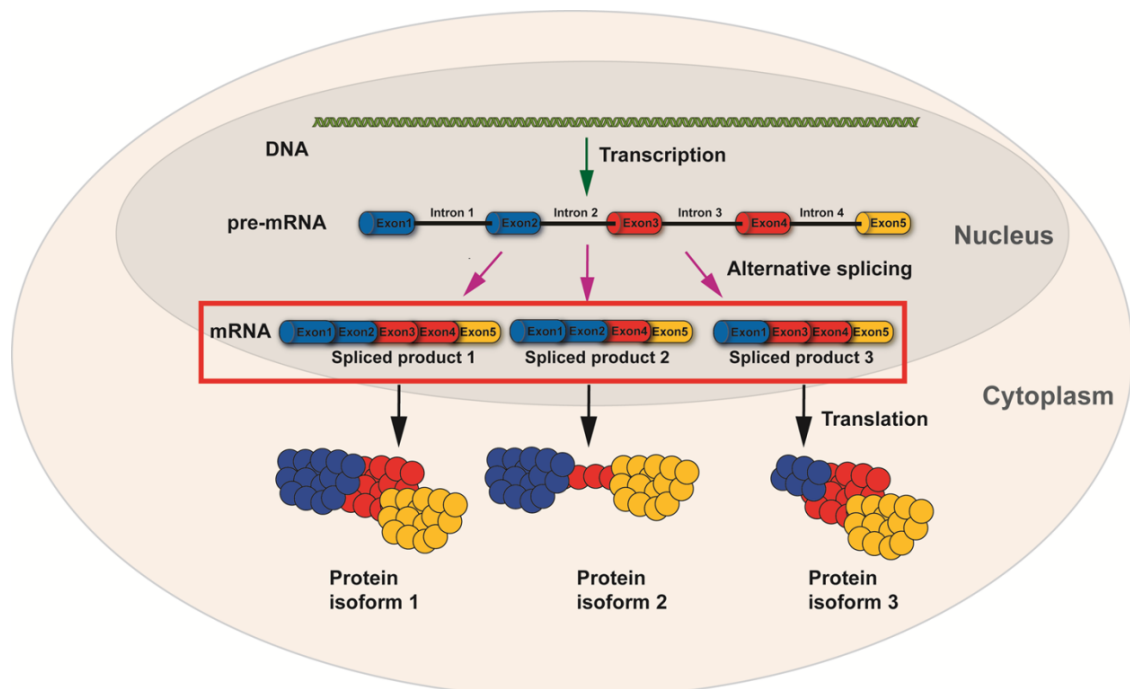


Figure 1-22. Schematic representation of alternative RNA splicing.

Alternative splicing is a common process in higher eukaryotes and it has been reported that >90% of the human genes undergo alternative RNA splicing, thus encoding for two or more mRNA isoforms.³⁸ There are eight ways through which alternative splicing occurs (Figure 1-23).³⁹⁻⁴¹ In mammals nearly 40% of alternative splicing events occur through the exon skipping mode (Figure 1-23A).^{40,42-44}

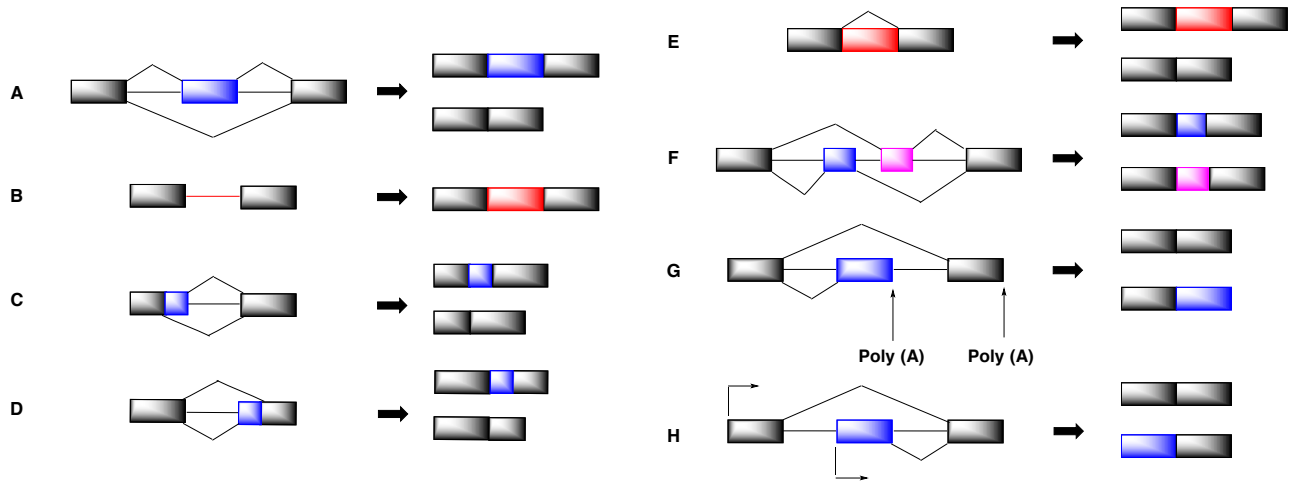


Figure 1-23. Alternative splicing patterns. Exons are presented as boxes, introns by lines and splicing patterns by lines joining the splicing sites. A) Cassette exon skipped or retained. B) Intron inclusion. C, D) Alternative 5' and 3' splice sites. E) Retention or exclusion of cryptic intron. F) Mutually exclusive splicing involves inclusion of one exon in a group of exons. G) Poly (A) sites terminate the transcription at different points. H) Alternative promoters leading to different start points of the 5' exon.

Alternative RNA splicing also has an important role in the differentiation of cells in embryos that will eventually lead them to become part of specific tissues in an adult organism. Burge et al. analyzed alternative splicing patterns that occur in different human tissues.⁴⁵ In their study they observed that the percentage of genes that undergo alternative splicing can vary substantially from tissue to tissue. For example, brain tissue had >40% of its genes undergoing alternative splicing, while muscle tissue had <20%. In addition, brain tissue cells exhibited the highest number of alternative splicing events that occur through the exon skipping mode (~30%). They also observed that the highest percentage of alternative 5' or 3' splice site selection events occurred in liver tissue. These results underscore the importance of alternative RNA splicing in tissue development and how important this process is in the complexity encountered in higher organisms.

1.7 Alternative RNA Splicing and Disease

Alternative splicing is a major source of protein diversity and its importance is emphasized through the fact that about 15% of disease-causing mutations are linked to aberrant splicing processes.⁴⁶ Even single-point mutations in pre-mRNA can result in major damage in the human body. Such is the case of Spinal Muscular Atrophy (SMA). The SMN protein is widely expressed in eukaryotic cells. Two genes encode this protein, SMN1 and SMN2. In SMN2, a C is mutated into a T in exon 7. This alteration leads to exon 7 being skipped during the splicing process and to the production of a truncated protein in 80-90% of the cases (Figure 1-24).⁴⁷⁻⁴⁹ In SMA affected individuals, SMN1 is unable to code for the functional protein due to various point-mutations or nucleotide deletion. The SMN2 gene still produces 10-20% of the functional protein but this reduced availability causes gradual death of the motor neurons.⁴⁷ This leads to symptoms ranging from muscle weakness and lack of head control to breathing and feeding difficulties.⁵⁰⁻⁵³

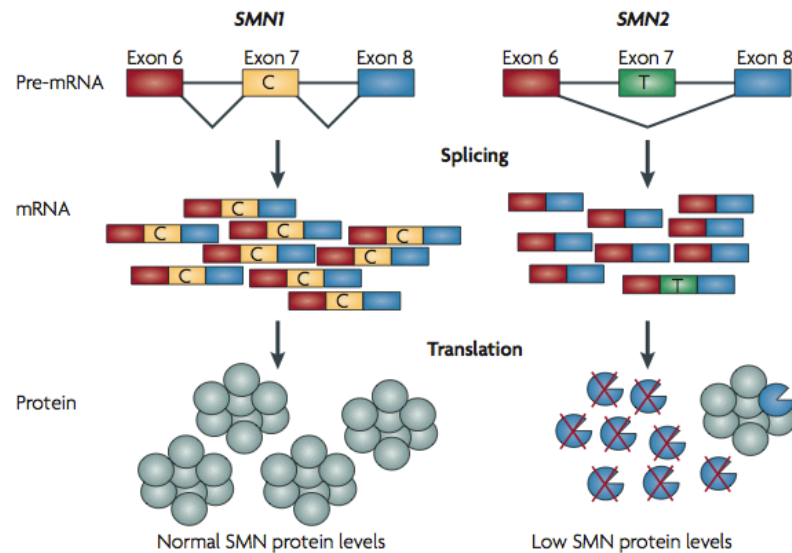


Figure 1-24. A mutation in the SMN2 gene causes the formation of a truncated protein.⁵³ Reproduced with permission from reference 53. Copyright 2019 Nature.

Another case of a single-point mutation causing aberrations in alternative splicing has been observed in the Hutchinson-Gilford progeria syndrome (HGPS).⁵⁴ HGPS is characterized by signs of premature ageing in children (growth retardation, loss of hair etc.). Ultimately, patients live an average of 13 years and die of complications of atherosclerosis and either stroke or cardiac disease.⁵⁵ The genotypical cause of HGPS has been located in the alternative RNA splicing pattern of the LMNA gene, which codes for the protein lamin A.⁵⁴ Lamin A is a component of the nuclear envelope, which encases the genetic material. In HGPS patients, a single nucleotide mutation of C to T in position 1824 of the LMNA gene results in activation of a cryptic splice site within exon 11, resulting in the production of a truncated isoform (Figure 1-25a).⁵⁴ This isoform causes an unstable nuclear envelope, which leads to loss of nucleus organization and ultimately a cell, which is unable to divide (Figure 1-25b and 1-25c).⁵⁶

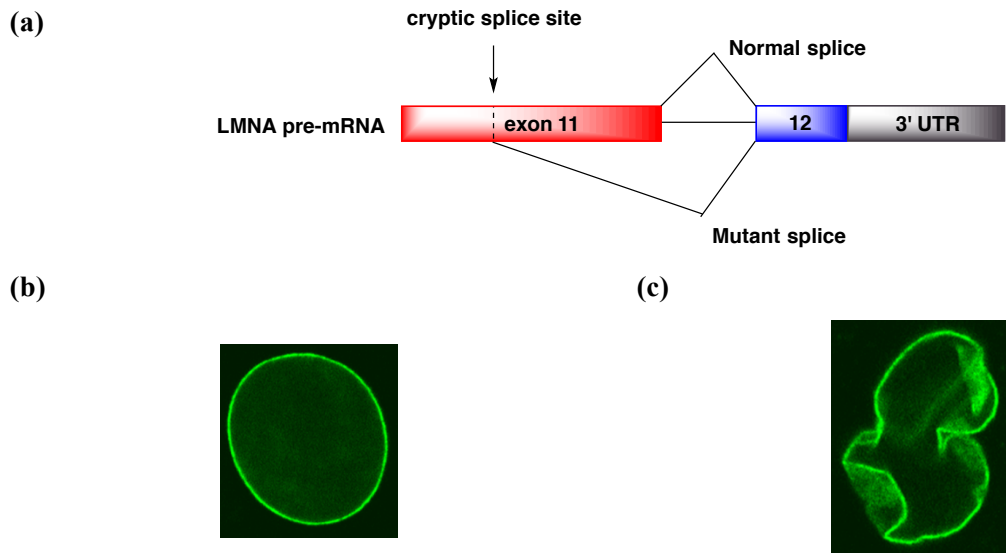
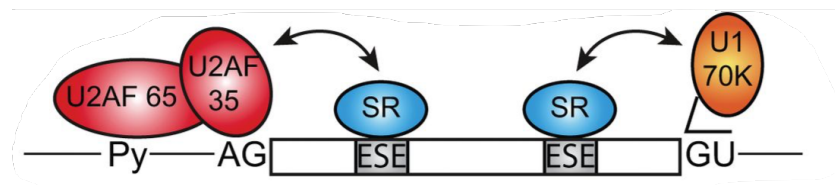


Figure 1-25. (a) Mutation of C to T in the LMNA gene causes activation of a cryptic splice site and production of a truncated lamin A protein. UTR: untranslated region. (b) Cell nucleus of a healthy cell.⁵⁶ (c) Cell nucleus of a progeric cell.⁵⁶ Permission not needed to reproduce Figure 1b and 1c.

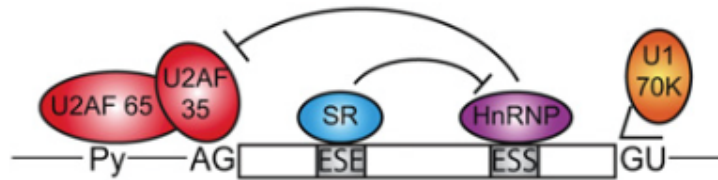
1.8 Splice Site Selection

Alternative RNA splicing can produce different mature mRNAs from the same pre-mRNA, through a number of different pathways (see Figure 1-23). The selection of a specific splice site can be directed by RNA sequences within both exons and in introns.⁵⁷ These sequences serve as recognition sites for various regulatory proteins that can either enhance or inhibit the recruitment of spliceosomal factors. One of the most understood auxiliary splicing elements are the exonic splicing enhancers (ESEs). ESEs are six nucleotide long motifs within an exon that direct pre-mRNA splicing. A key discovery regarding the role of ESEs was that they are recognized by proteins of the SR (Serine-Arginine) family of splicing factors.^{57,58} SR proteins interact directly with protein subunits of snRNPs to facilitate pre-spliceosomal assembly. Therefore, SR proteins bound to ESEs are thought to enhance spliceosomal recognition. Figure 1-26 shows three models that have been proposed to explain the mechanism by which SR proteins act.

(a)



(b)



(c)

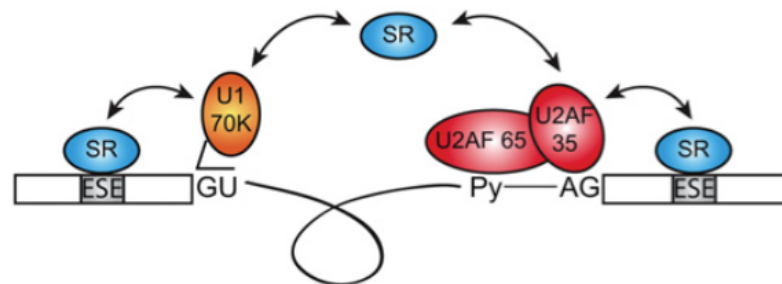


Figure 1-26. Schematic representation of two proposed models regarding the mechanism of SR proteins.⁵⁹ ESE = Exonic splicing enhancer. ESS = Exonic splicing silencer. hnRNP = heterogeneous ribonucleoprotein particles. **(a)** The “recruitment” model. **(b)** The “inhibitor” model. **(c)** The “bridging” model. Permission not needed to reproduce this figure as stated by the Portland Press.

In the “recruitment” model (Figure 1-26a), SR proteins bound to ESEs within the same exon interact with U1 and U2 components, directing them at the 3’ and 5’ splice sites respectively.⁶⁰ In the “inhibitor” model, SR proteins antagonize the repressing activity of the heterogeneous ribonucleoprotein particles (hnRNPs) (Figure 1-26b).⁶¹ Finally, in the “bridging” model an SR protein “bridges” two splicing factors bound onto the intron, while two more SR proteins are bound to ESEs within the two flanking exons, further stabilizing the splicing factors and driving splicing at these particular sites (Figure 1-26c).⁶² Furthermore, it has been demonstrated that different members of the SR family recognize different splice sites. Consequently, it has been hypothesized that cell-specific expression of SR proteins can drive splice site recognition.⁵⁸ The observation by Sreaton and co-workers that during phytohaemagglutinin (PHA) induced T-cell activation, levels of the CD44v6 and CD45RO isoforms of the CD44 glycoprotein increased, with a concomitant increase of SR proteins SRp30c and SRp55 supports this hypothesis.⁶³ Point mutations that lead to disease and most likely affect ESEs have been identified in a variety of genes.^{64,65} Another common regulatory element in alternative RNA splicing is the exonic splicing silencers (ESSs). While ESEs enhance splice site recognition, ESSs inhibit it. These sequences (usually 4-18 nts long) serve as binding sites for splicing repressors such as the hnRNPs.⁶⁶ The mechanism of inhibition depends on the specific hnRNP. For example, hnRNP A1 can prevent the association of SR proteins (which are necessary for splice site recognition), while

hnRNP I (also known as PTB) can block spliceosomal assembly by preventing interactions between spliceosomal components.⁶⁷

Exon skipping/inclusion in the Fas gene is an example of the second proposed mechanism. The Fas gene encodes for the Fas receptor (FasR), a receptor involved in programmed cell death (apoptosis).⁶⁸ There are eight splice variants for the FasR pre-mRNA but two of them are common.⁶⁹ In the case of the first isoform, the protein TIA-1 binds to an intronic splicing enhancer (ISE) between exons 6 and 7. TIA-1 enhances binding of U1 snRNP to the 5' splice.⁷⁰ U1 in turn recruits U2 auxiliary factor (U2AF), and eventually exon 6 gets spliced out. In the case of the second isoform, hnRNP I binds to a pyrimidine-rich ESS within exon 6. This disrupts the interactions between U2AF and U1, resulting in exon 6 inclusion.⁷⁰ The full-length isoform of FasR is membrane bound, while the truncated protein is found on the cytosol (Figure 1-27).⁶⁹

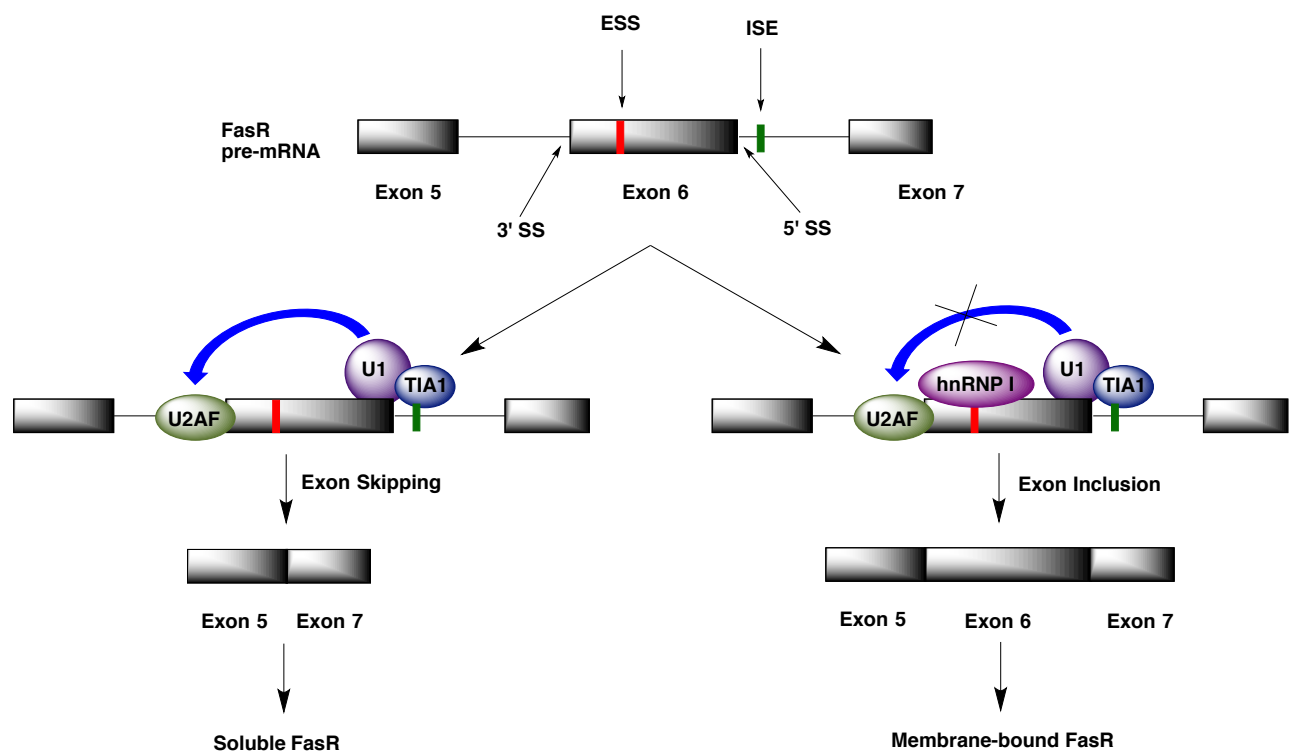


Figure 1-27. Inhibition of exon 6 skipping on the FasR pre-mRNA alternative splicing mediated by hnRNP I. ESS = Exonic splicing silencer. ISE = Intronic splicing enhancer.

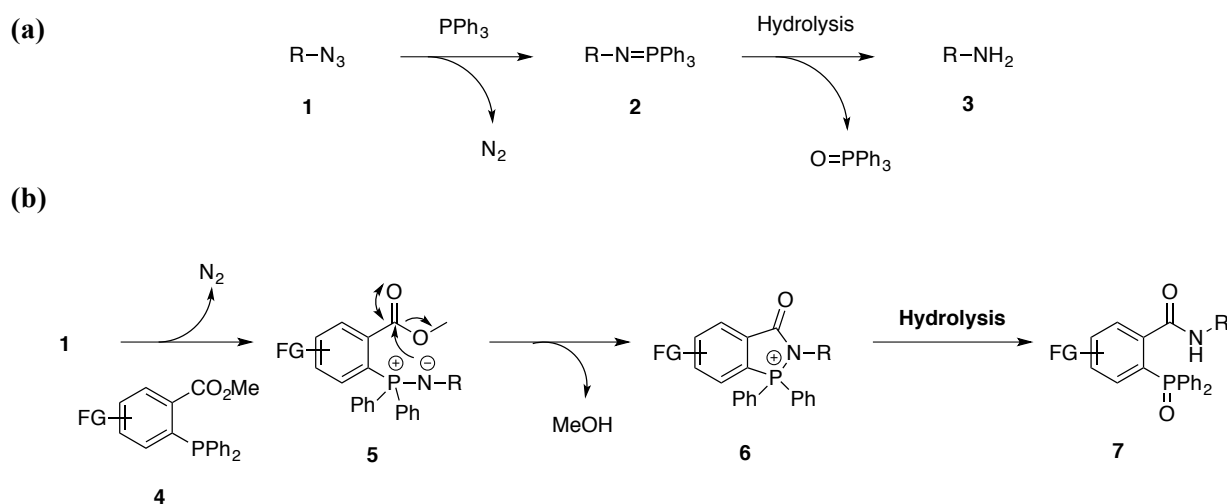
Elevated production of this cytosol-soluble FasR (sFas) has been discovered in many cancer lines.^{71,72} Membrane-bound Fas can be recognized by cytotoxic T cells that defend the organisms against viral infections or cancer cells.⁶⁸ However, sFas is not exposed and thus, cancer cells evade apoptosis.⁷³ A recent study has revealed 200 alternative splicing regulators involved in the balance between the two FasR isoforms, including the SR proteins SRSF6 and SRSF1.⁷⁴ These findings indicate that the alternative splicing pattern of FasR pre-mRNA can be an attractive target for therapeutic intervention.

It has become evident that studying the primary sequences of splicing regulators and how they can affect splice site selection can provide important insight into the splice selection process, its association with

disease and lead to the development of novel therapeutic strategies. This could be realized by labeling important sequences (e.g., ESEs) with fluorophores at specific locations within a pre-mRNA. In the following section general methods for biomolecule methods will be presented, followed by methods for RNA labeling and how FRET can be used to investigate the dynamics of pre-mRNA as this undergoes alternative RNA splicing.

1.9 Fluorescent Labeling of Biomolecules

One of the most common labeling strategies is the Staudinger ligation, which was developed by Bertozzi and co-workers.^{75,76} In the Staudinger reaction an azide reacts with a phosphine to form iminophosphorane **2**, which is then hydrolyzed to amine **3** and triphenylphosphine oxide (Scheme 1-1a). In Bertozzi's variant modified phosphine **4** forms iminophosphorane **5**. This intermediate is trapped by the ester moiety forming intermediate **6**, which is then hydrolyzed to form amide **7** (Scheme 1-1b).



Scheme 1-1: (a) The Staudinger reaction.⁷⁶ (b) The modification developed by Bertozzi et al. FG = Functional group.⁷⁵

Bertozzi et al. demonstrated the application of this method by labeling several cell lines (Jurkat, HL-60 etc.).⁷⁷ Firstly they prepared azide **9** from carbohydrate **8**. Jurkat cells were incubated with **9**, which was metabolized by the cells to introduce the modified sugar in cell-surface glycoproteins. These modified cells were treated with FLAG-tagged phosphine **10** to yield FLAG-tagged cells. Fluorescent anti-FLAG antibodies were then used to bind and label the cells, which were detected by flow-cytometry (Figure 1-28). A drawback of this method is the sensitivity of phosphines like **4** to air-oxidation and their synthetically challenging optimization for increased water solubility and reaction rate.⁷⁸

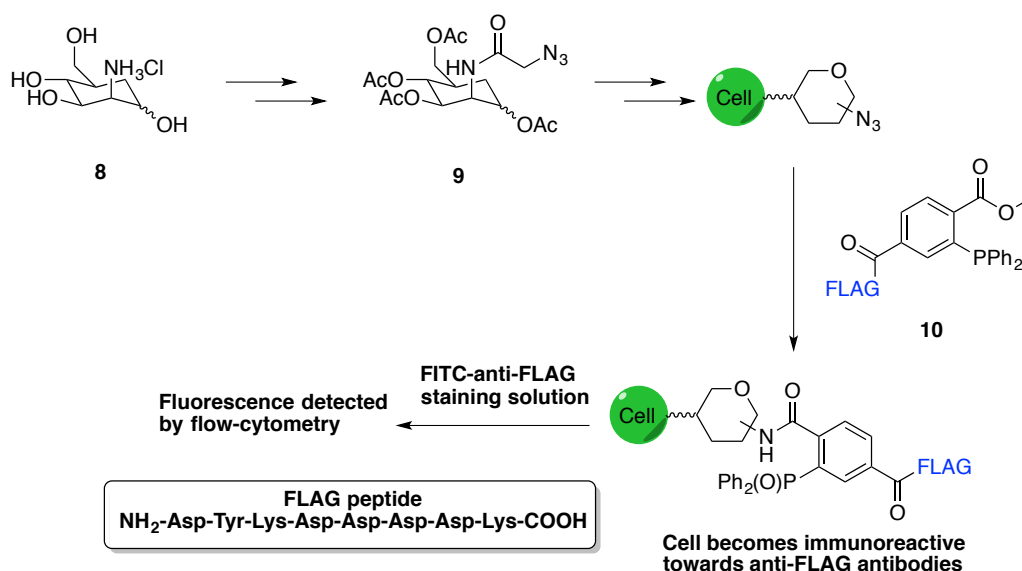
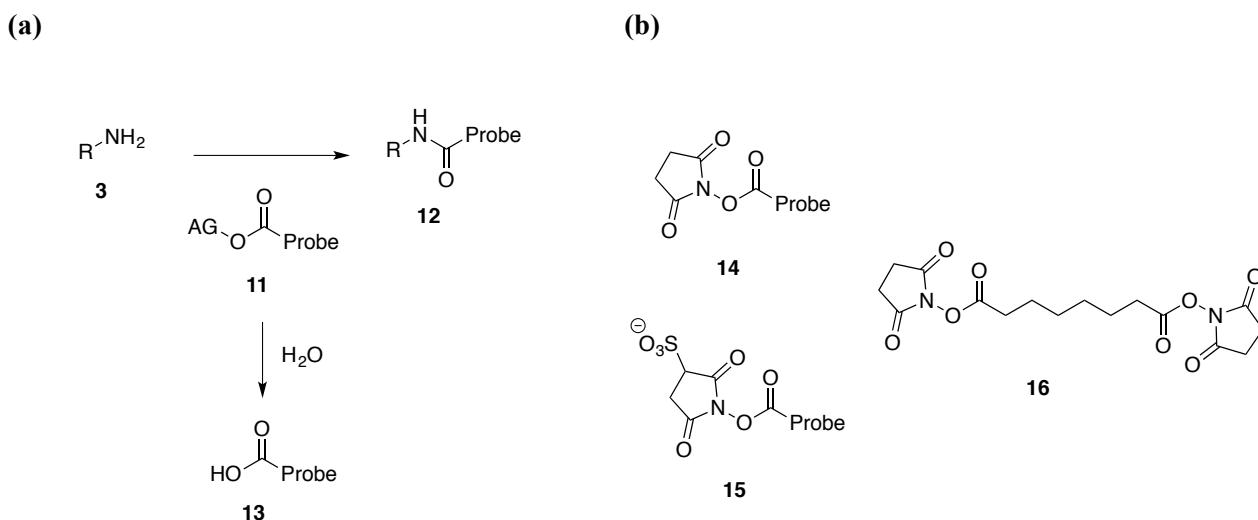


Figure 1-28. Labeling of Jurkat cells in vivo, using the Staudinger ligation.⁷⁷

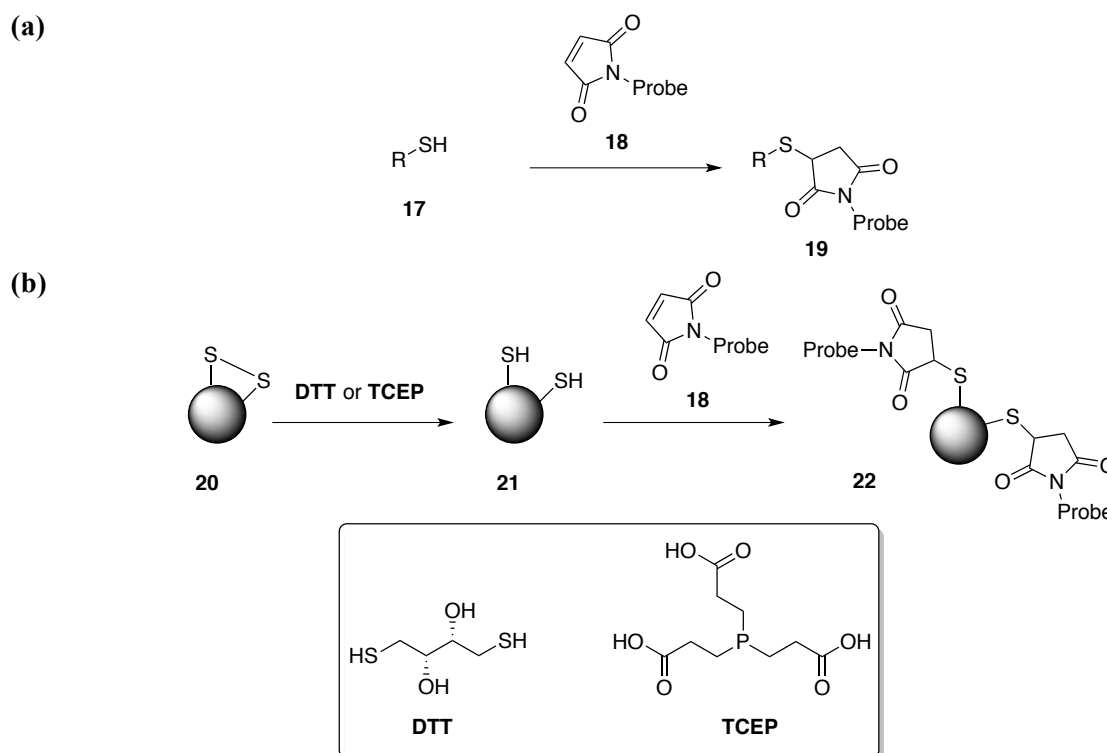
Amines can be derivatized by using activated ester **11** (Scheme 1-2a). The most common activating group used is the NHS (N-hydroxysuccinimide), as seen in derivative **14**. Other derivatives include **15**, which exhibits increased solubility in water, as well as bifunctional reagent **16** (Scheme 1-2b). The disadvantage of this method is competing hydrolysis of **11** to **13**.



Scheme 1-2: (a) Labeling of amines using activated esters. AG = Activating group. (b) Commercially available NHS esters. Probe = Various tags such as Alexa Fluor® 488, Biotin.

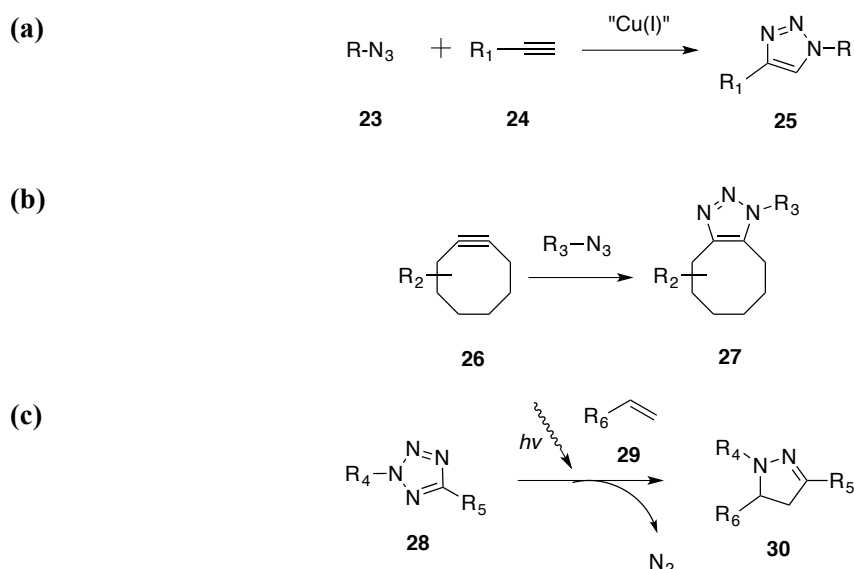
Thiols can be selectively labeled with maleimide **18** to furnish a stable thioether bond (Scheme 1-3a). Targets of maleimide probes are the thiols present in cysteine residues and in reduced disulfide bridges. Reduction can be accomplished using reagents such as DTT or TCEP (Scheme 1-3b). In addition, site-directed mutagenesis can be used to introduce cysteine residues to specific sites.⁷⁹ The reaction at pH 6.5-7.5 is fast (~30 min) and produces no by-products.⁸⁰ Unfortunately, thiol-containing reducing agents (e.g. DTT)

will also react with **18** and need to be completely removed before the coupling step. TCEP circumvents this problem. Furthermore, primary amines in proteins (e.g. lysine residues) can also react with maleimides like **18** at higher pH (~8.0).^{81,82}



Scheme 1-3: Thiol modification via maleimide coupling. (a) Thiol modification. (b) Disulfide reduction and subsequent labeling.

Another commonly used method is the copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) or “click” reaction.^{83,84} Azide **23** reacts with alkyne **24** affording triazole **25** in the presence of a Cu(II) salt and a reducing agent (e.g. sodium ascorbate) (Scheme 1-4a). The reaction can be carried out in water (or mixtures with miscible organic solvents) and does not produce any side-products, only requiring removal of the copper salt and the reducing agent. Furthermore, azides and alkynes are not naturally present in biomolecules, thus avoiding competing side reactions and can be introduced by chemical or biological means (e.g. protein engineering). Most importantly, azides and alkynes are small moieties, which do not perturb the natural behavior of biomolecules. Copper-free variants of the “click” reaction have also been developed. Bertozzi developed the strain-promoted azide-alkyne cycloaddition (SPAAC), which avoids the use of copper, by using strained cyclic alkyne **26** (Scheme 1-4b).⁷⁸ The driving force of this reaction is ring-strain relief. Photo-click reactions have also been reported, between tetrazole **28** and alkene **29**, which utilize nitrogen expulsion as the driving force (Scheme 1-4c).⁸⁵ These methods find use in *in vivo* applications by avoiding the use of copper(I), which is toxic for living cells.

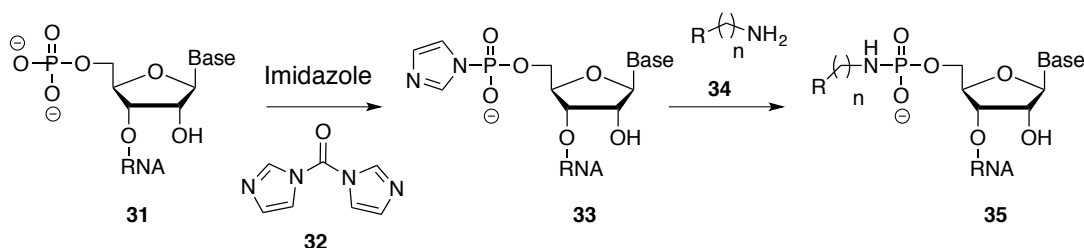


Scheme 1-4: (a) Copper catalyzed cycloaddition.^{83,84} (b) Strain-promoted cycloaddition.⁷⁸ (c) Photo-click variant.⁸⁵

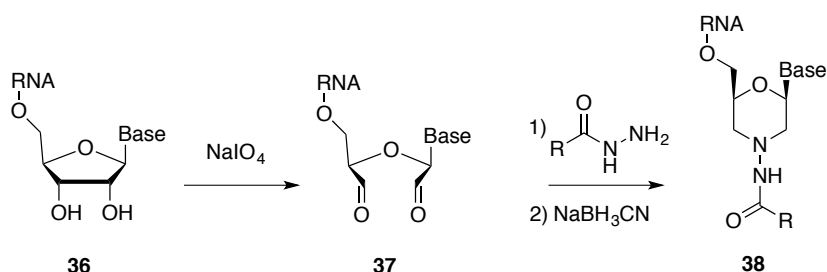
1.10 Current State-of-the-Art in RNA Labeling

Labeling RNA at specific sites enables the elucidation of biochemical processes that occur in living organisms. Fluorophore labeling of RNA has been widely used to study alternative RNA splicing.^{86,87} Site-specific labeling could provide a way to study the splice site recognition during alternative RNA splicing. The labeling method selected in each case depends on the desired location of the label and the length of the strand to be labeled.⁸⁸ For short RNA molecules (<40 nt long) solid-phase chemical RNA synthesis allows the incorporation of one or more fluorophores at a desired position.⁸⁸ A variety of fluorophore phosphoramidites can be incorporated by solid-phase synthesis. On the other hand, it is still challenging to prepare RNA molecules greater than 40 nucleotides by chemical synthesis. Labeling strategies for long RNA strands have focused on functionalizing either the 5' or 3' terminus.⁸⁸ Scheme 1-5 presents two of the most common methods for labeling RNA termini. In 1983 Wahl et al. reported that the 5' phosphate group can be derivatized using various amines ranging from ethylene diamine to poly-Lysine (Scheme 1-5a).⁸⁹ Treatment of **31** with imidazole and **32** furnishes phosphorimidazolide **33**. Condensation of **33** with **34** results in derivatized RNA **35**. This method leaves the nucleobase and backbone functionalities intact. Also, imidazolide **33** can be used directly without purification, but this requires lowering the concentration of CDI (by ~10 fold) to prevent internal site modification during the long incubation times with **34**.⁸⁸ The 3' RNA terminus can be labeled as shown in Scheme 1-17b. Treatment of RNA **36** with sodium periodate affords dialdehyde **37**. Condensation with hydrazine derivatives of fluorophores results in fluorophore attachment at the 3' end with almost 100% efficiency.^{88,90,91} The aforementioned methods cannot be applied in the case of labeling an internal site in long (<40 nt) RNA molecules, since they require either a free 5' phosphate group, or 3' terminal nucleotide.

(a)



(b)



Scheme 1-5: Labeling of RNA termini. (a) Labeling of the 5' end. (b) Labeling of the 3' end. R denotes a fluorescent tag.

1.11 Site-specific RNA Labeling

As shown in Figures 1-15 to 1-21, the catalytic cycle of splicing involves a variety of RNA-RNA and RNA-protein interactions. Furthermore, there are specific sequences in each snRNP that bind to conserved parts in mRNA. Thus, the need for site-specific labeling developed. This would allow for the labeling of a particular sequence of interest, elucidating the importance of specific sites. The research group of Moore developed one of the first strategies for site-specific RNA labeling.⁹² Moore and co-workers utilized a T4 DNA ligase and a DNA splint to join together modified RNA fragments constructing a 236 nt mRNA found in the eukaryote *Tetrahymena*. The 5' and 3' splice sites were modified with nucleosides **39** or **40** (Figure 1-29), which were incorporated into each segment by solid-phase synthesis. The 2'-OH group was methylated in **39** and removed in **40**. The authors used these modifications to elucidate the importance of a 2'-OH in the 3' splice site.⁹² Incorporation of either **39** or **40** in the 3' splice site did not affect the first step (5' exon excision), implying that this part of the mRNA is not recognized during spliceosome formation. However, termination of the second splicing step was observed. This highlights the importance of 2'-OH during the ligation of the 5' with the 3' exon, although it is unclear if it plays a direct role in the mechanism or serves as a recognition element.

analyze mRNA conformational dynamics during splicing using single molecule spectroscopy.^{95,96} Unfortunately, like Moore's method (see Figure 1-29) this method relies on the design and synthesis of a suitable DNA splint, which is substrate-specific and thus, will differ depending on the RNA target in hand.

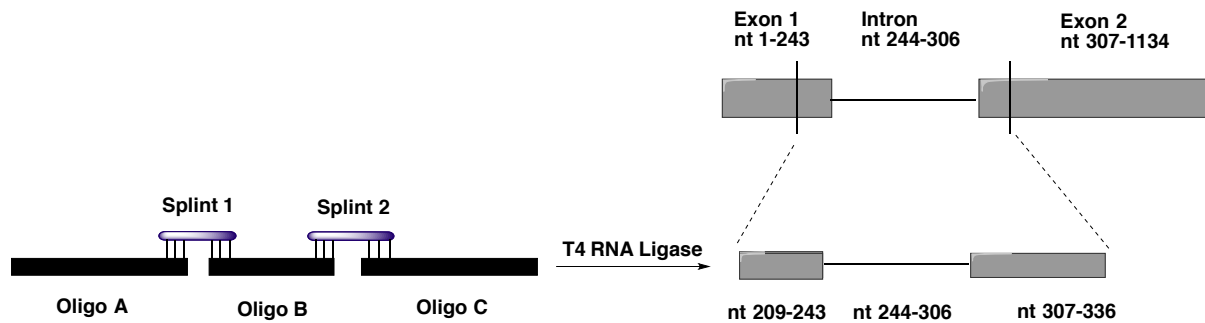


Figure 1-30. Stark's method allowed the construction of a 128mer, part of the YOL074c gene. nt = nucleotide

Indirect labeling of a large RNA molecule can also be achieved by hybridization with a labeled 18mer DNA sequence.⁹⁷ However, this method has only been tested in specific sequences of an RNase, which did not affect the tertiary structure of the ribozyme. The hybridization in other sequences in this ribozyme was not investigated. Furthermore, the studied ribozyme was enlarged in the cDNA binding position to increase sequence specificity.

Silverman et al. used deoxyribozyme 10DM24 to catalyze the coupling of a modified 19nt. long RNA at the 2'-OH position of adenosines, which were selected on the basis of their accessibility (based on the X-ray crystal structure) and their lack of contribution to tertiary interactions.⁹⁸ Firstly, RNA strand **44** was labeled using NHS ester **45** (Figure 1-31a). The unmodified amino-allyl RNA was also used. The yields in all cases were greater than 50% using biotin, fluorescein or TAMRA as tags. This method offers the advantage of labeling a variety of RNAs, since it has been observed that 10DM24 has considerable tolerance regarding its RNA substrates.⁹⁹ A drawback of this method is that 10DM24 shows specificity towards the 2'-OH of adenosines, thus providing little option for labeling of other nucleotides.⁹⁸ Furthermore, this method can only be used to label sites that do not participate in tertiary interactions, since a bulky 19mer is attached. In the same report the Silverman group reported the application of another deoxyribozyme, which recognizes the labeled RNA and cuts the tagging RNA down to eight nucleotides (Figure 1-31b). However, this introduces an extra step to an already laborious method.

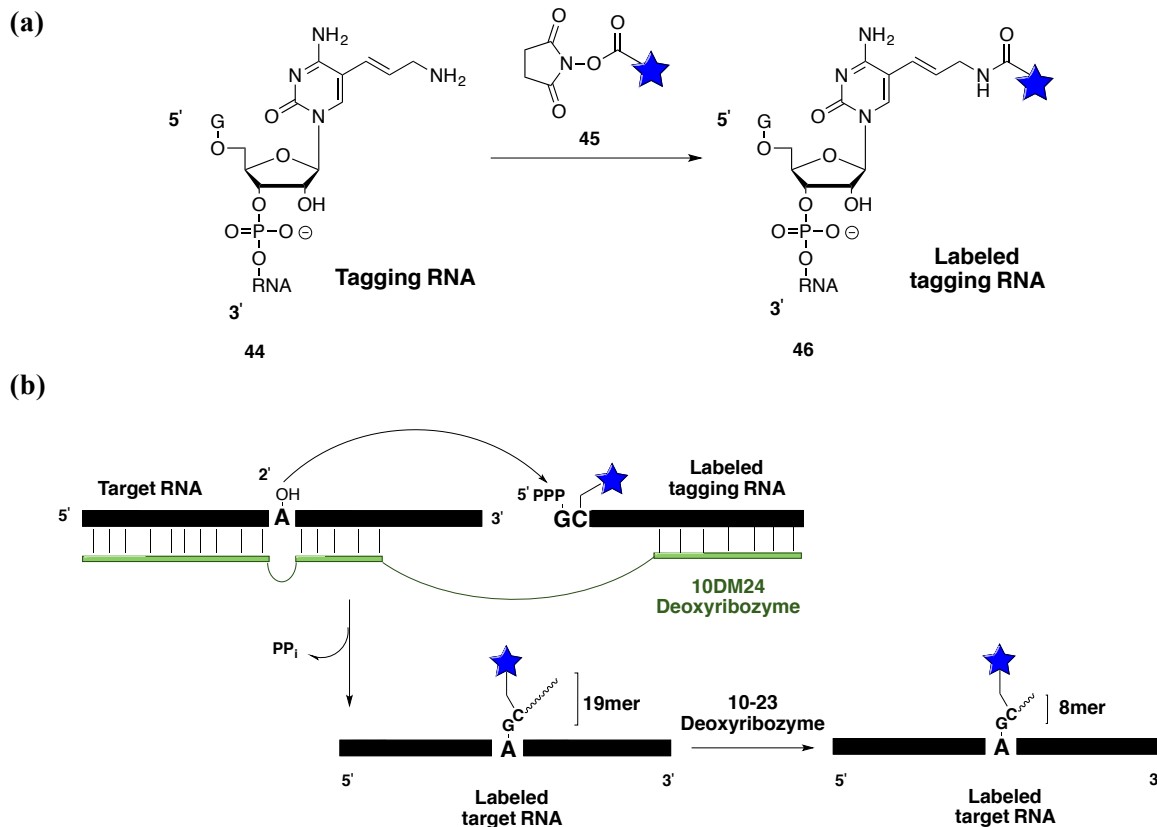


Figure 1-31. Silverman's labeling method using the deoxyribozyme 10DM24.⁹⁸ (a) Synthesis of labeled tagging RNA. (b) 10DM24 catalyzed labeling of a target RNA and cleavage of the 19mer tagging RNA to an 8mer. Blue asterisk denotes an affinity or fluorescent label (e.g. Biotin, TAMRA).

Efforts to improve this method were then initiated by the same group. In the same year, Höbartner and Silverman reported an engineered 10DM24 deoxyribozyme to accept a single nucleotide triphosphate.¹⁰⁰ This enzyme was later used by Höbartner (as an independent researcher at the Max-Planck Institute) to install labeled guanosine analogues at the 2'-OH of adenosines (Figure 1-32a).¹⁰¹ The applicability of this method was demonstrated by the sequential labeling of two positions of a SAM-binding riboswitch with a FRET pair (Figure 1-32b). The folding of this riboswitch could be monitored *via* FRET upon incubation with SAM. In spite of these improvements, this method suffers from the following: (i) The need to isolate the deoxyribozyme; (ii) Only adenosines can be labeled and (iii) The need to isolate a different ribozyme for each different adenosine target.

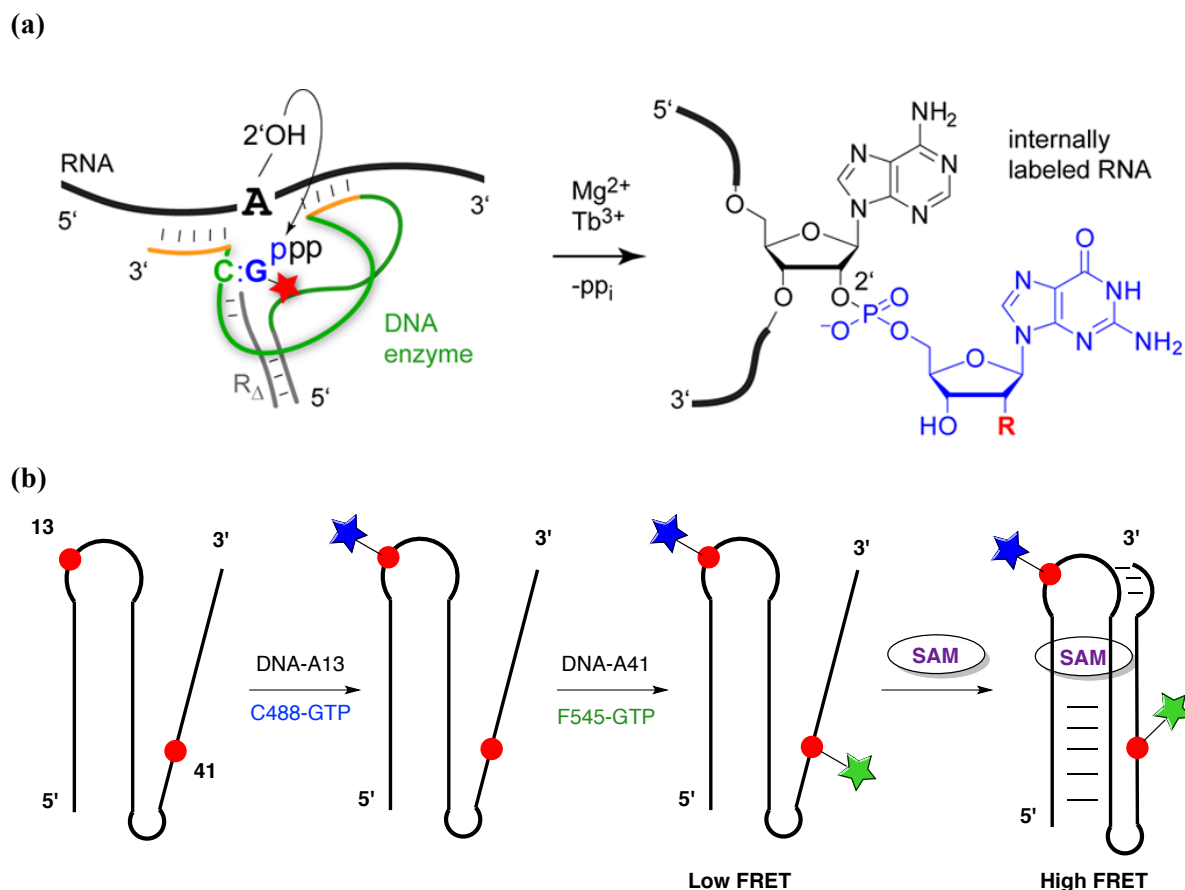


Figure 1-32. Höbartner's improvement of Silverman's deoxyribozyme-catalyzed RNA labeling.¹⁰¹ (a) Schematic representation. R denotes various tags (e.g. biotin, TAMRA) Figure reproduced with permission from Reference 101. Copyright (2014) American Chemical Society. (b) Application of this method to the double-labeling of a SAM-binding riboswitch.

Currently, the methods for efficient site-specific internal labeling of long RNA molecules are limited to the ligation of modified nucleotides or oligonucleotides *via* enzymatic means. These methods require the synthesis of a DNA splint or deoxyribozyme. These are designed to be substrate-specific, thus for every different RNA target a new oligonucleotide needs to be designed and synthesized. Indirect labeling has been achieved through hybridization of a labeled DNA sequence with a ribozyme; however, this method requires that the region of hybridization is enlarged to increase binding affinity to the label and is not involved in interactions with other RNAs or proteins and the enlargement. An alternative methodology that does not involve the ligation of modified oligonucleotides involves the incorporation of a synthetic pair of nucleotides that pairs selectively with each other into DNA or RNA. This method requires the synthesis of single nucleotides, thus bypassing the need for the synthesis of substrate-specific splints or large deoxyribozymes. This is the concept of orthogonal base pairs and will be described in detail in the next section.

1.12 The Development of Orthogonal Base Pairs – Expanding Nature’s Genetic Code Beyond Watson-Crick

In the late 1980’s, the concept of orthogonal base pairs was introduced independently by Rappaport and Benner.^{102,103} This concept encompasses the design of two artificial nucleotides that would pair selectively with each other but not with naturally-occurring nucleotides. This section will begin by introducing early examples of orthogonal base pairs and summarizing their evolution to the current state-of-the-art exemplars. In addition, the application of selected pairs in site-specific RNA labeling will be presented.

1.12.1 Base Pairs Developed by the Benner Group

The group of Benner synthesized the first artificial base pair in 1989. Benner and co-workers synthesized a pair analogous to the G/C base pair. These bases are isomers of guanosine and cytidine and were named *iso-G* and *iso-C* (Figure 1-33).¹⁰² The *iso-C/iso-G* base pair implements a hydrogen-bonding pattern similar to the G/C one. Both of these pairs form three hydrogen bonds. The C nucleotide implements a donor-donor-acceptor pattern, whereas *iso-C* implements an acceptor-acceptor-donor. In general, the properties of natural nucleosides must be taken into account. Pairing through hydrogen bonding; the shape complementarity to create a non-distorted double helix and the hypothesis that DNA polymerases seek electron density in the minor groove.^{104,105}

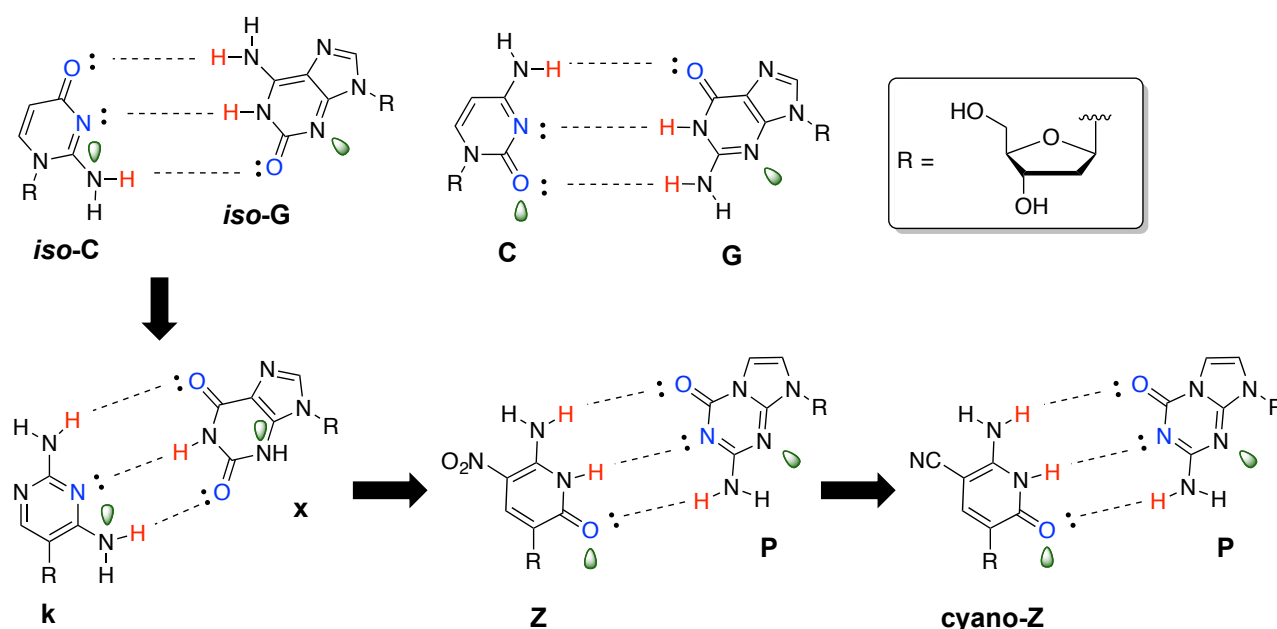


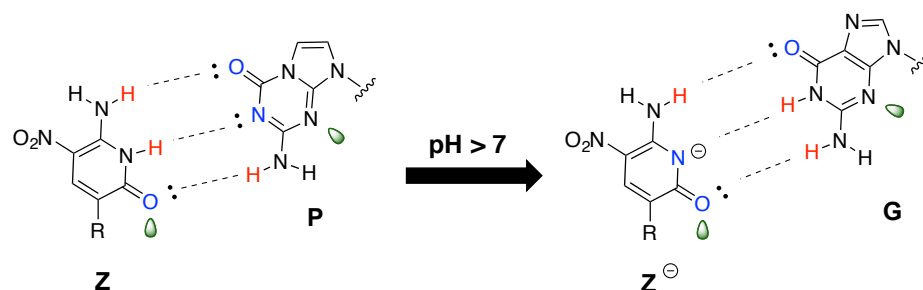
Figure 1-33: The G/C pair and evolution of base pairs developed by the Benner group. Hydrogen-bond donor moieties are marked in red. Hydrogen-bond acceptor moieties are marked in blue. Green lobes denote the electron density that DNA polymerases recognize in the active site.¹⁰⁴

Klenow fragment (Kf) polymerase efficiently incorporated the *iso-C/iso-G* base pair into a DNA strand via primer extension experiment. However, misincorporation of T opposite to *iso-G* was observed. Furthermore,

it was observed that *iso*-C undergoes deamination to 2'-deoxyuridine during oligonucleotide preparation by solid-phase synthesis.¹⁰⁶ In addition, *iso*-C triphosphate was observed to be more unstable than cytosine triphosphate, with 32% remaining after 6 weeks storage in -20 °C.¹⁰⁶ A base pair that was developed later was *x/k*.¹⁰⁷ Primer extension experiments showed that templates containing *k* were 70% as efficient at directing formation of full-length product as the templates containing natural nucleotides, but 5% misincorporation of *x* opposite C was observed.¹⁰⁸

Attempts to improve incorporation efficiency led to the development of the *Z/P* base pair.¹⁰⁹ This pair could be incorporated efficiently (*Z/P* retention averaged 99.8% per theoretical PCR cycle under the optimized conditions developed by the authors) by a variety of different polymerases (including *Taq* and Phusion). Even when four consecutive *Z* nucleotides were placed in a row, some polymerases (e.g. *Taq* polymerases) placed *Z* opposite to *P* selectively.¹¹⁰ Disadvantages of the *Z/P* pair are the tendency of *Z* to form mismatch with G under alkaline conditions (Figure 1-34a) and an increased epimerization rate under acidic conditions (Figure 1-34b).¹⁰⁹ Specifically, Benner et al. reported that upon incubation of an aqueous sample of *Z* at pH = 3, HPLC analysis revealed the formation of four compounds, which they presume to be compounds **47-50** (Figure 1-34b).¹⁰⁹

(a)



(b)

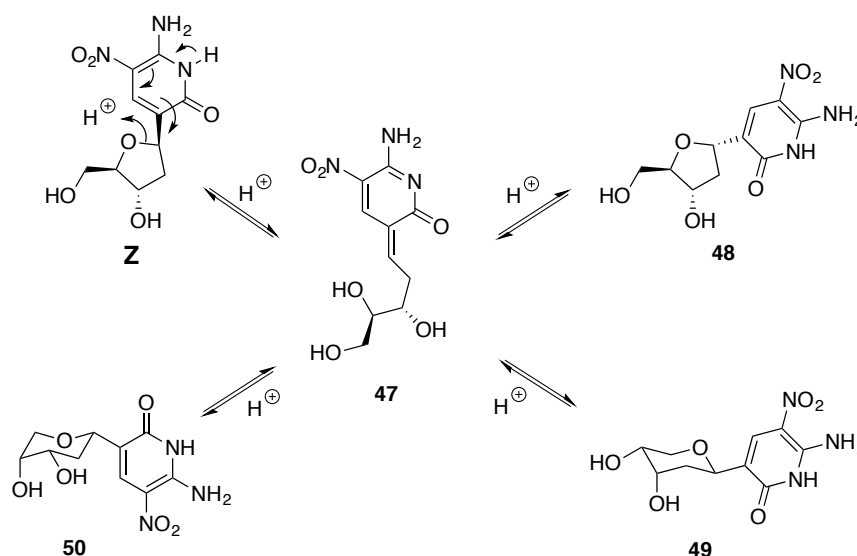


Figure 1-34. (a) Basic pH results in the formation of a deprotonated *Z*, which can then mismatch with G. Sugar cores have been omitted for clarity. Hydrogen-bond donor moieties are marked in red. Hydrogen-bond acceptor moieties are marked in blue.⁶⁵ (b) Proposed epimerization intermediates by Benner et al. that form under acidic conditions.¹⁰⁹

In order to mitigate the drawback of G-mismatch, the Benner group substituted the nitro group in **Z** with a cyano moiety (Figure 1-33).¹¹¹ It was expected that the less electron withdrawing nature of a cyano group would make the nucleoside less susceptible to deprotonation under alkaline conditions, thus preventing mismatch with G (Figure 1-34a). However, primer extension experiments conducted by the Benner group revealed that some polymerases do not sufficiently accept cyano-**Z**. Out of the six polymerases tested, only Terminator and Klenow fragment (exo⁻) were able to sufficiently incorporate cyano-**Z**.¹¹¹ Benner rationalized this result on the basis of a hydrogen bond forming between the nitro group and the exocyclic amine of **Z**, which could “lock” the amine in a planar position. Benner hypothesized that the absence of this hydrogen bond leads the amine to being forced out of the aromatic plane, conformation that might not be recognized by some polymerases (Figure 1-35b). Therefore, the **Z**/**P** pair still remains the current-state-of-the-art of Benner’s nucleotides and a promising candidate for site-specific RNA labeling if an RNA version of **Z** could be incorporated into an RNA strand.

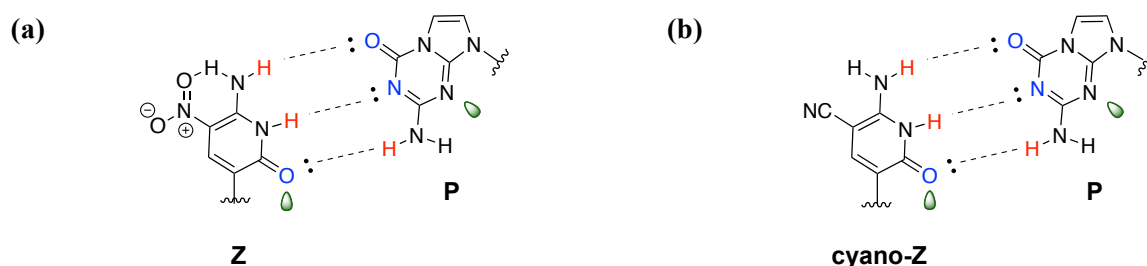


Figure 1-35. Proposed explanation by Benner for the observed inefficient incorporation of **cyano-Z** in DNA. Sugar cores have been omitted for clarity. Hydrogen-bond donor moieties are marked in red. Hydrogen-bond acceptor moieties are marked in blue. (a) Hydrogen bond forming in **Z**, between the nitro group and the exocyclic amine. (b) **Cyano-Z** lacks this feature.¹¹¹

1.12.2 Base Pairs Developed by the Kool Group

Erik Kool and co-workers have investigated the use of hydrophobic interactions in orthogonal base pairs, rather than the canonical hydrogen-bond pattern. An early example was the incorporation of abasic nucleotide ϕ opposite to pyrene substituted **P** (Figure 1-36). DNA 11-mer duplexes incorporating these artificial nucleotides opposite to each other are slightly less stable ($\Delta T_m = 1.6\text{--}2.2$ °C) than the A/T pair by UV-Vis melting studies.¹¹² This observation suggested that hydrogen bonding might not be necessary for base-pairing. This was the first example demonstrating pairing through hydrophobic interactions.

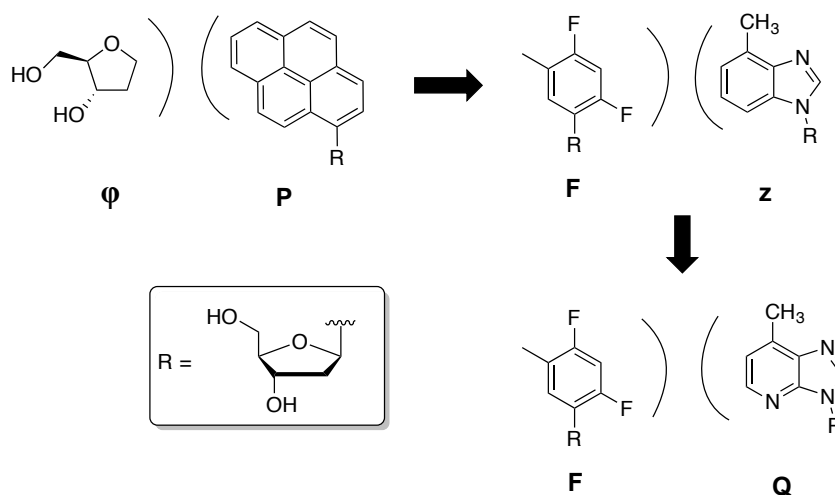


Figure 1-36. Examples of base pairs developed by the Kool group.

Further exploration of other base pairs included the **F/z** and **F/Q** base pairs. Specifically, **F** was designed as a **T** mimic and **Q** and **z** as **A** mimics.¹¹³ **Q** lacks Watson-Crick pairing but retained a nitrogen analogous to the N3 of adenosine, which sits in the minor groove and acts as a hydrogen bond acceptor, interacting with DNA polymerases.^{104,105} On the other hand, in **z** the nitrogen was removed to test the importance of a hydrogen bond acceptor during replication. Nucleotide insertion experiments showed efficient pair formation for both **z** and **Q** using a 23 nt primer. Misincorporation of **T** opposite **z** and **Q** were observed, as well as **A** opposite **F**, thus proving these artificial base pairs as efficient mimics (these are observations from qualitative data. Quantification of the misincorporation rate has not been reported). Overall, experimental data suggests that minor groove interactions of nucleobases with DNA polymerases have a relatively small effect on base pair formation.¹¹⁴ Additionally, Kool and co-workers performed experiments to test the extension efficiency of template already incorporating a **z/F** or **Q/F** pair.¹¹⁴ Fully elongated product was observed in both cases. In the case of **Q/F** complete elongation was observed after 2 min, but not in the template incorporating **z/F**. These observations suggest that minor groove interactions between polymerases and the 3' end of the primer are critical for the incorporation of the next base during replication.⁶⁸

1.12.3 Base Pairs Developed by the Hirao Group

The first non-natural base pair developed by Hirao and co-workers was **x/y** (Figure 1-37).¹¹⁵ Hirao showed that although **y** can be misincorporated opposite **A** or **G** (but not **C** and **T**), in the presence of **x**, selective incorporation was observed using Kf polymerase (with 3' exonuclease activity). When the variant lacking exonuclease activity was used, misincorporation of **T** opposite to **x** was observed, since a repair mechanism was absent. In 2004 Hirao reported an improved version, nucleotide **s**, which also pairs with **y**.¹¹⁶

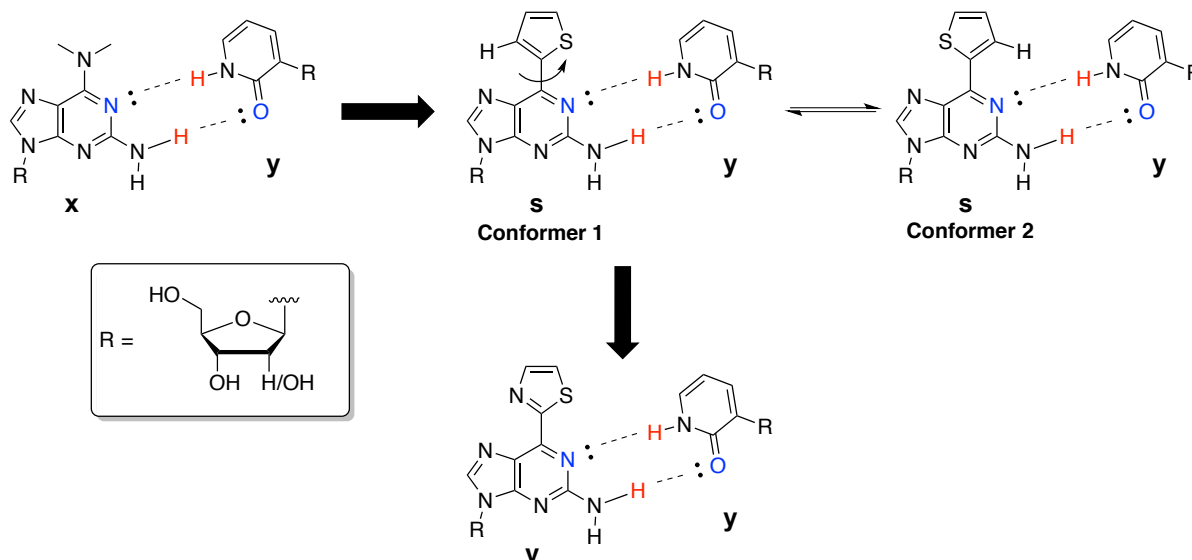


Figure 1-37: Early examples of base pairs developed by Ichiro Hirao and co-workers.

In **s** the dimethylamino group was replaced with a bulkier thiophene. According to the Hirao group, a thienyl group would increase the steric hindrance with the carbonyl group of T, thus preventing a mismatch; it was also assumed that the aromatic nature of this moiety could enhance base stacking with neighboring nucleobases. Indeed, primer extension experiments demonstrated that **s** exhibited improved incorporation efficiency compared to **x** (9.2×10^3 vs. 2.7×10^3 % $\text{min}^{-1} \times \text{M}^{-1}$), while also reducing misincorporation against T by 50%.¹¹⁶ However, the transcriptional efficiency of the **s**/y was observed by the Hirao group to be 50-60% lower than that of native transcription with natural pairs.¹¹⁷ Hirao and co-workers proposed that rotation of the bond between the thiophene and the guanosine scaffold in **s** could yield two different conformers. The thiophene C-H bond was hypothesized to resemble the methyl group on **x**. Since **x**/y exhibited lower transcriptional efficiency compared to **s**/y, they concluded that conformer 2 (where the C-H points towards the **y**), was responsible for this reduced efficiency (Figure 1-37). Efforts to overcome this drawback yielded the **v**/y pair.¹¹⁷ To avoid the potential problem of bond rotation, Hirao installed a thiazole moiety in **v**, where each conformer would have a heteroatom pointing towards the base pairing region. This modification resulted in increased transcriptional efficiency (98% and 110% relative yields with 0.5 mM and 1mM of yTP respectively) compared to 51% when **s** was placed on the template strand.¹¹⁷

Simultaneously, the Hirao group explored the utilization of the **Pa/Q** pair, based on Kool's work.¹¹⁸ **Pa** was designed to avoid mismatch of Kool's **F** with A.¹¹³ Still, **Q** is an efficient A mimic, and forms a mismatch with T (Figure 1-38).



Figure 1-38. (a) The **Pa/Q** base pair developed by the Hirao group.¹¹⁸ (b) The **F/Q** pair developed by the Kool group. Sugar cores have been omitted for clarity.¹¹³

Thus, Hirao developed the **Ds/Pa** pair (Figure 1-39).¹¹⁹ Derivatives of the **Pa/Ds** pair have also been used by Hirao to label an RNA transcript using “click” chemistry (see Section 1.13). In addition, Hirao observed that the previously developed **s** can also pair with **Pa**. However, **Ds** can form a stable self-pair; furthermore, a mismatch of **Pa** with A was observed. This hurdle was overcome by using modified γ -amidotriphosphates of **Ds** and A (**DsTP_N** and **ATP_N**), which express reduced incorporation reactivity in PCR experiments, but also limit *in vivo* applications.¹¹⁹ Efforts to overcome this drawback led to the synthesis of the **Ds/Pn** pair.¹²⁰ The nitro group in **Pn** was installed to electrostatically repel the N1 in A; thus, suppressing mismatching. According to single-nucleotide incorporation experiments performed by Hirao and co-workers, incorporation efficiency of **PnTP** opposite a **Ds**-modified template was 6-fold higher than with **PaTP** (3.7×10^5 vs. 6.2×10^4 % $\text{min}^{-1} \times \text{M}^{-1}$). Furthermore, the incorporation efficiency of **PnTP** opposite A was reduced by 2.5-fold (3.7×10^5 vs. 6.2×10^4 % $\text{min}^{-1} \times \text{M}^{-1}$).¹²⁰ Similar results were obtained when **Pn/Pa** were incorporated in the template.¹²⁰ The mutation rate during PCR amplification was determined to be less than 1% with **PnTP** compared to 3-4% when **PaTP** and the γ -amidotriphosphates were used.¹²⁰ Consequently, This modification overcame the use of **ATP_N**, but **DsTP_N** was still crucial to avoid the **Ds/Ds** mismatch.¹²⁰ In addition, **Pn** nucleotide appears to be less stable under strong basic conditions, presumably degrading to its sugar and 2-nitropyrrole components.¹²⁰ Further optimization efforts led to the **Ds/Px** pair.¹²¹ Based on literature precedent, Hirao speculated that increasing the hydrophobicity of the **Pn** would enhance the interactions with the hydrophobic side-chains of the polymerase, as well as the base of the 3' end of the primer.¹²¹ The **Ds/Px** pair was amplified with >99% selectivity and 10^7 -fold efficiency after 30 PCR cycles without the assistance of either the A or **Ds** amidotriphosphate, whereas the efficiency of the **Ds/Pa** decreases after 10 PCR cycles.¹²¹ On the other hand, the **Ds/Pa** is has higher selectivity in transcription and is more stable compared to the **Px**.¹²² Thus, Hirao retained the use of **Ds**, but varied its pairing partner regarding the application in hand.¹²²

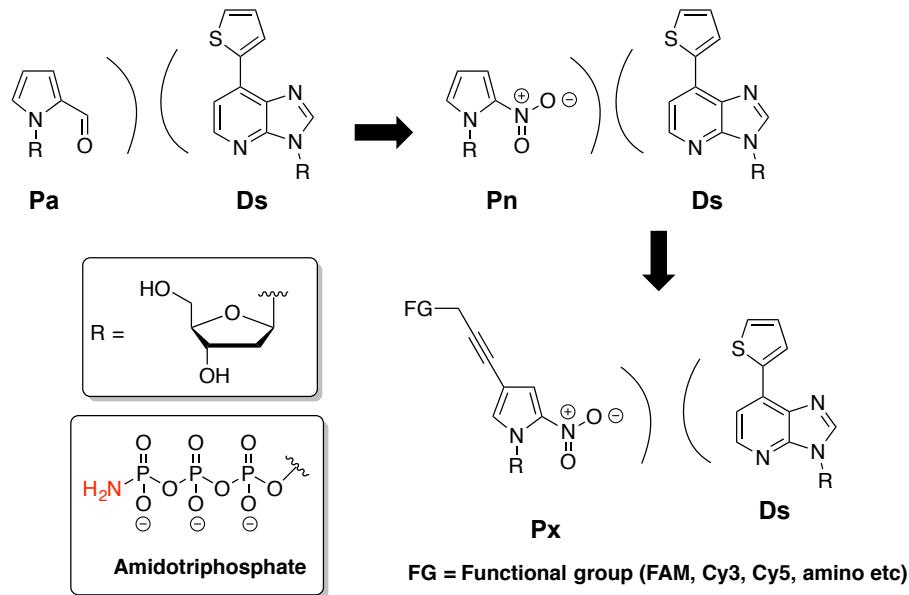


Figure 1-39. Evolution of the **Ds/Pa** base pair and γ -amidotriphosphate (shown at bottom) used by the Hirao group for PCR experiments.

1.12.4 Base Pairs Developed by the Romesberg Group

The Romesberg group has developed and optimized hydrophobic base pairs. In 2007, Romesberg reported the base pairs **5SICS/MMO2** and **5SICS/NaM**. The Romesberg group demonstrated that both pairs could be replicated and transcribed using Kf polymerase and T7 RNA polymerase respectively.^{123,124} Efforts towards a base-pair demonstrating better incorporation and extension efficiencies than the previous pairs led to the synthesis of the **FEMO** nucleoside as a pairing partner for **5SICS**.¹²⁵ Changing the toluoyl moiety in **5SICS** to a simple thiophene yielded nucleoside **TPT3** (Figure 1-40).¹²⁶ The **TPT3/NaM** pair displays increased incorporation in PCR amplification of a 45-mer DNA template.¹²⁶

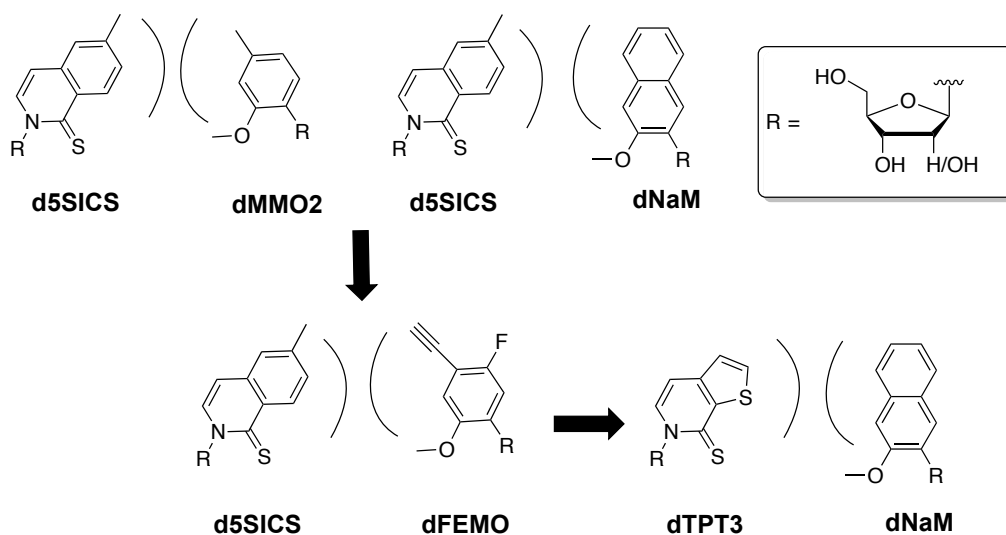


Figure 1-40. Latest base pairs developed by the Romesberg group.

Recently, Romesberg and co-workers reported the first example of *in vivo* replication of unnatural nucleotides.¹²⁷ A plasmid incorporating the **5SICS/NaM** pair was constructed and transformed into *E. coli* C41 (DE3) cells. Next, they transformed the cells with a plasmid expressing an *algae* nucleotide triphosphate transporter. In that way efficient uptake of the unnatural triphosphates was achieved, where the endogenous replication machinery used them to replicate the modified plasmid (Figure 1-41). The efficiency was determined to be 64% (determined by incubation with biotinylated **NaM** triphosphate and subsequent streptavidin gel-shift) and fidelity (retention per doubling) was determined to be 99.4%.¹²⁷

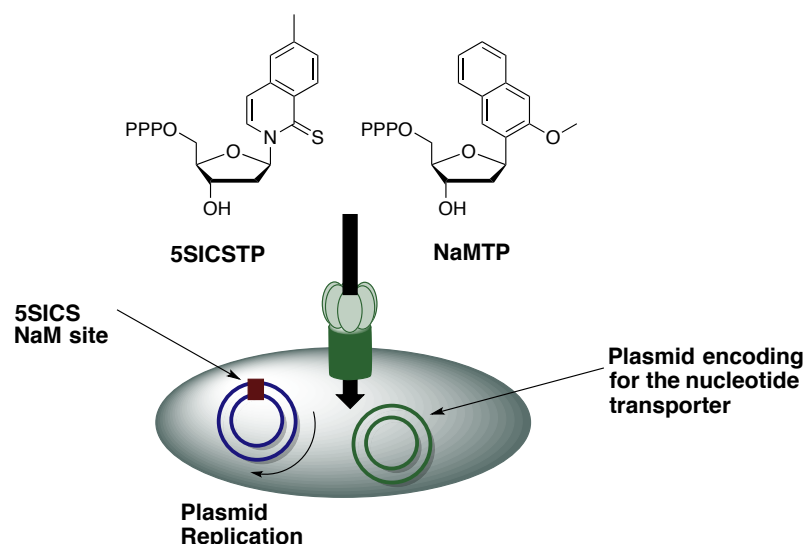


Figure 1-41. An *algae* nucleotide transporter facilitates uptake of the **5SICS/NaM**, where they are used by the cell to replicate a modified plasmid. PPP = Triphosphate.¹²⁷

1.13 Orthogonal Base Pairs in RNA Labeling

The development of orthogonal base-pair strategies has been mainly focused in DNA labeling. This is due to the complications encountered in the synthesis of both phosphoramidites required for solid-phase oligonucleotide synthesis and triphosphates required for transcription assays. For phosphoramidite synthesis, the 2'-OH has to be selectively protected, adding extra steps in the synthetic route for installation and deprotection. This has led to the development of a range of reagents for selective 2' protection.¹²⁸ For triphosphate synthesis, a common drawback is the formation of 3' and 2' triphosphate regioisomers. Many methods have been developed over the years, but most of them require extra protection steps, or highly reactive reagents.¹²⁹ Still, orthogonal base pairs have been used for site-specific RNA labeling and representative examples will be presented in this section. In one of the earliest examples, Hirao and co-workers used the **v/y** base pair to site specifically label an RNA strand.¹³⁰ Firstly, they synthesized a variety of fluorescently labeled variants of **y** (Figure 1-42).

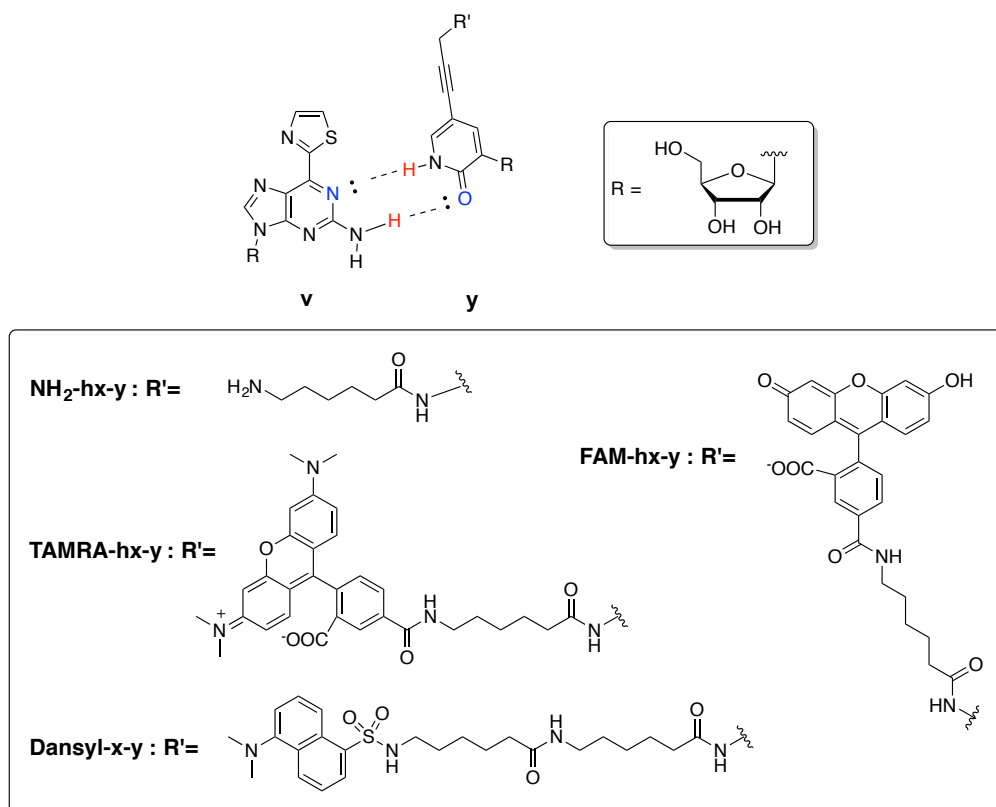


Figure 1-42. The v/y base pair and labeled variants of y used in Hirao's work.¹³⁰

These y analogues were used in transcription assays using T7 RNA polymerase to incorporate y variants within a 17 nt RNA using a 35 nt DNA template (Figure 1-43). **FAM-hx-y** gave the best relative yields (80%, average of three data sets), whereas dansyl-labeled y gave only 40% yield of incorporation efficiency.¹³⁰

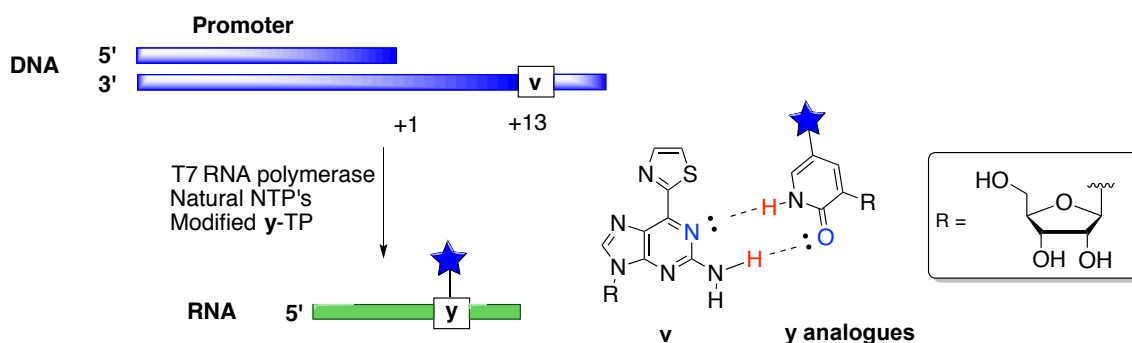


Figure 1-43. Hirao's strategy employing the v/y pair to site-specifically label 17-mer RNA transcripts. Transcription with modified y triphosphates furnished a labeled RNA.¹³⁰

In addition, Hirao examined the posttranscriptional modification of this 17-mer transcript containing **NH₂-hx-y**, using the NHS esters of TAMRA or FAM dyes. The FAM-NHS ester gave 40% yield only with 1000 molar equivalents. In sharp contrast, the TAMRA-NHS ester gave 66% yield with 100 molar equivalents and

99% with 1000 molar equivalents (Figure 1-44).¹³⁰ This work demonstrated the potential of the v/y pair for site-specific RNA labeling during transcription and post-transcription.

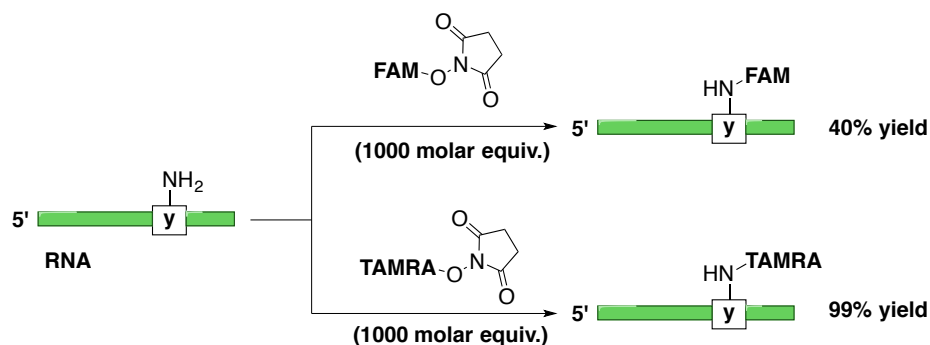


Figure 1-44. Hirao's strategy employing the v/y pair to site-specifically label 17-mer RNA transcripts. Post-transcriptional modification of the 17-mer RNA transcript with FAM or TAMRA NHS-esters.¹³⁰

In 2011, the Romesberg group published another example of site-specific RNA labeling.¹³¹ A modified DNA fragment was synthesized encoding a 77 nt tRNA (CUA anticodon) with **5SICS** instead of A in the anticodon. Transcription was carried out with the modified **MMO2^A** triphosphate. Romesberg used this modified tRNA to label the CU-**MMO2^A** anticodon with biotin (Figure 1-45). The incorporation efficiency was calculated using streptavidin gel-shift and was found to be greater than 75%.¹³¹

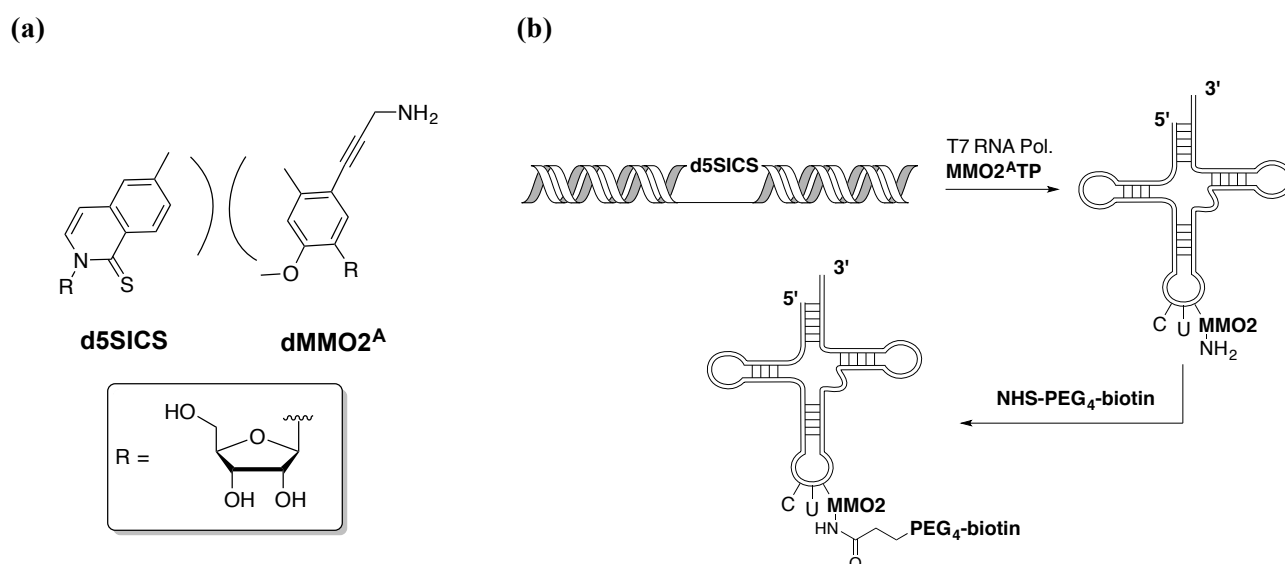


Figure 1-45. (a) The **5SICS** and **MMO2^A** pair used by Romesberg for site-specific RNA labeling. (b) Site-specific labeling of a tRNA in the third place of anticodon.¹³¹

In the same year, Hirao used the **Pa/Ds** base pair for the post-transcriptional modification of an RNA transcript.¹³² Hirao synthesized 35nt DNA templates containing the **Pa/Ds** pair. Transcription using T7 RNA polymerase in the presence of modified **Pa** triphosphates **Eth-C4-PaTP** and **Eth-PaTP** provided a 17nt RNA incorporating an alkyne functionality (Figure 1-46). Incorporation efficiency was calculated as high as 99% for both nucleosides after copper-catalyzed [3+2] cycloaddition with Biotin-N₃ and streptavidin gel-

shift assay.¹³² Labeling was also carried out with Alexafluor 488-N₃ or 593-N₃ dyes to produce fluorophore-labeled transcripts (Figure 1-32). Analysis of the products and comparison of band intensities using gel-electrophoresis revealed that labeling was more efficient with **Eth-C4-Pa** (97% for 488-N₃ and 98% for 593-N₃). Furthermore, Hiraio prepared a 75mer tRNA transcript containing **Eth-C4-Pa** in position 33, with an efficiency of 99%. Treatment with Alexafluor 488-N₃ or 593-N₃ dyes resulted in labeling of the transcript.

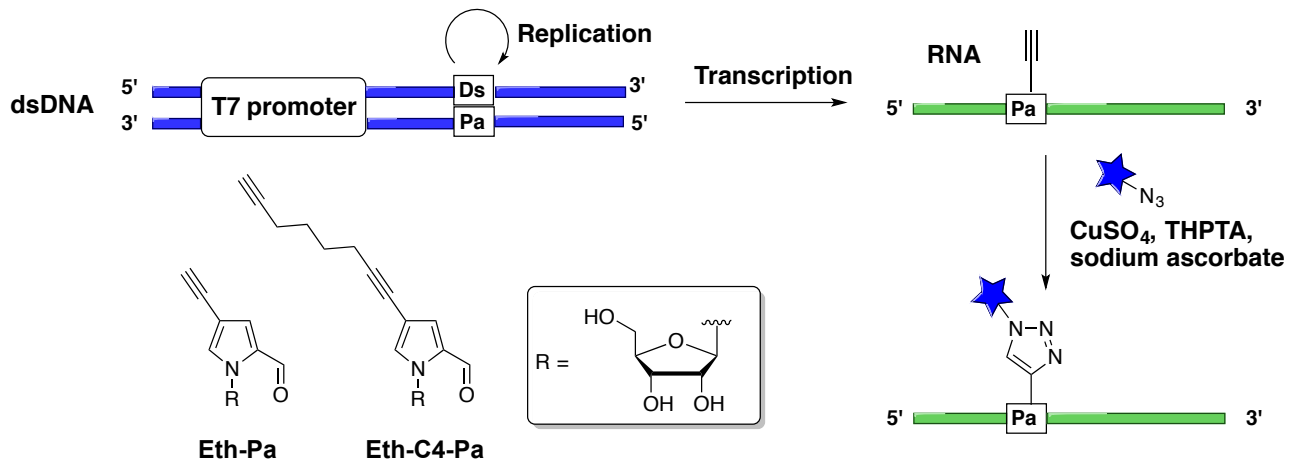


Figure 1-46. Eth-Pa and Eth-C4-Pa used in Hiraio's work and overview of Hiraio's labeling strategy. Fluorophore is denoted by a blue asterisk.¹³²

The Romesberg group reported recently the first example of a doubly labeled RNA strand.¹³³ Using modified versions of both the **5SISCS** and **NaM**, they succeeded in incorporating two different fluorophores in a 243-nt fragment of the *Thermus thermophilus* ribosome. For the synthesis of this long RNA used in their study, Romesberg and co-workers used the assembly PCR method to synthesize the corresponding DNA template. The template was divided into four strands. Each one was synthesized *via* solid-phase synthesis and had an overlapping sequence with another fragment. Specifically, 1 overlapped with 2, 2 with 3 and 3 with 4. Fragments 3 and 4 featured one of Romesbergs unnatural nucleotides. PCR amplification of fragments 1+2 and 3+4 yielded two strands that overlap with each other. Isolating each product and then co-amplifying them afforded the 243-nt long DNA template. (Figure 1-47).

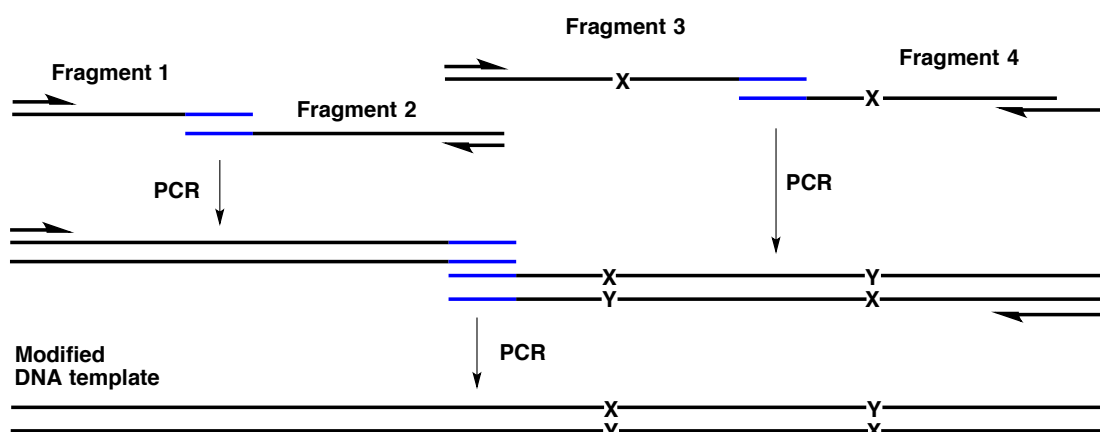


Figure 1-47. Assembly PCR method employed by the Romesberg group to incorporate two modifications in a DNA template. Half-ended arrows indicate the primers. Blue lines indicate overlapping sequences between fragments. X and Y indicate the unnatural nucleotides.¹³³

Subsequent transcription of this template with the unnatural nucleotide triphosphates yielded the modified RNA target, which was sequentially labeled with Cy3 and Cy5 fluorophores (Figure 1-48). This work represents a significant milestone in site-specific RNA labeling. Although, there are still some drawbacks. (i) Laborious synthesis of the required triphosphates (eight to 14 steps) (ii) The need for *de novo* synthesis of the base and subsequent coupling with a sugar donor (iii) The absence of a multiply functionalized long RNA molecule. Furthermore, this method has not been employed to elucidate aspects of splice-site selection during alternative RNA splicing.

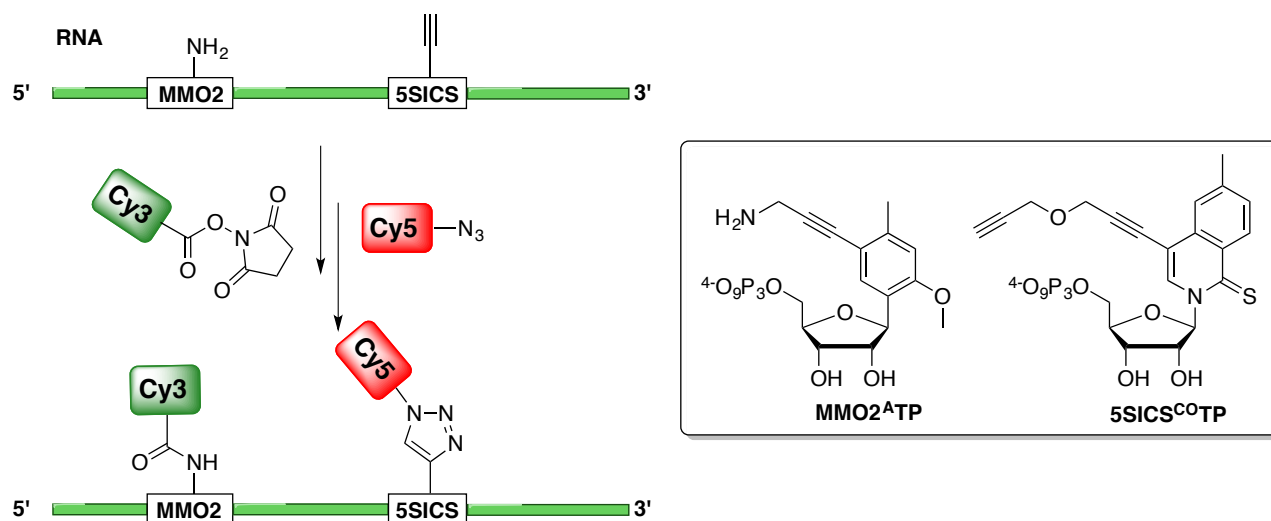


Figure 1-48. Functionalized nucleotides used in by the Romesberg group and overview of the double-labeling strategy.¹³³

1.14 Challenges in the Application of Orthogonal Base Pairs for Site-Specific RNA Labeling

At the time when this PhD project was initiated, the two studies reported by the Hirao group were the only examples of efficient site-specific labeling for both small (~17nt) and long (~75nt) RNA molecules using

one fluorescent tag. Labeling using two fluorescent tags in the same RNA has not been reported either at this point. The importance of doubly labeled RNA lies in studying RNA folding and dynamics using FRET. Förster Resonance Energy Transfer (FRET) describes the energy transfer between two chromophores (Figure 1-49). It is widely used in studying the dynamics of biomolecules, since the energy transfer efficiency is inversely proportional to the sixth power of the distance between the donor and acceptor molecule.

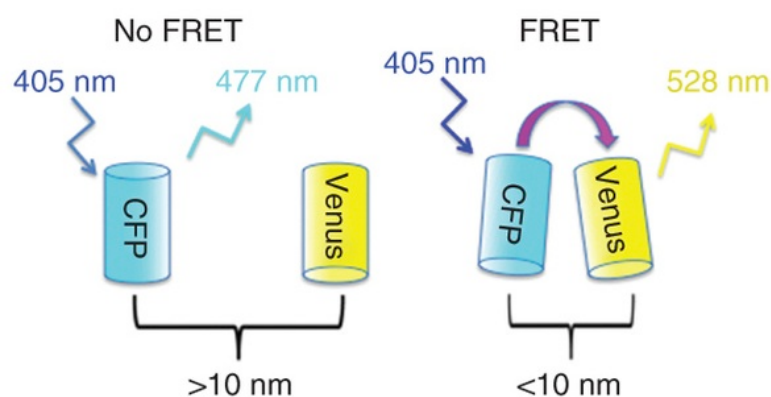


Figure 1-49. FRET between Cyan Fluorescent Protein (CFP) and Venus (yellow fluorescent protein).¹³⁴ Reproduced with permission from reference 85. Copyright (2013) Nature Protocols.

Using an orthogonal base-pair to incorporate two modifications at predetermined sites inside a long RNA combined with the ability for chemoselective labeling using biocompatible and orthogonal chemistry would provide a powerful methodology to study biological mechanisms. As reported in Section 1.13, Romesberg and coworkers published recently the first example of a doubly-labeled RNA and studied its folding *via* FRET. However, 14 steps are needed for **5SICS^{CO}** and 8 steps for **MMO2^A** triphosphates. Moreover, some of these steps employ sensitive organolithium reagents and cryogenic temperatures. Additionally, both bases have to be *de novo* synthesized and then coupled with a different sugar donor. Thus, from a synthetic chemistry perspective, this method is not robust.

1.15 Hypothesis to be Tested

The application of orthogonal base pairs for double site-specific RNA labeling hasn't been reported at the time this PhD project was initiated. Thus, the initial hypothesis of this work was that the development of an orthogonal base pair, where both nucleotides are modified with different functional groups, would provide a platform for site-specific RNA labeling by two different fluorescent molecules. This system could then be used to study the folding of a pre-mRNA model during alternative RNA splicing. The proposed strategy to be investigated is presented in Figure 1-49.

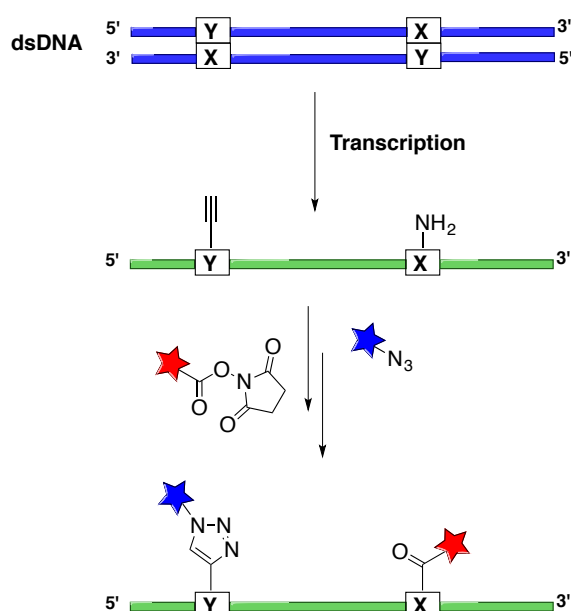


Figure 1-49. Proposed strategy for site-specific RNA labeling using orthogonal base pairs. X and Y indicate unnatural nucleotides. Blue and red stars indicate various fluorophores that can be used for labeling (e.g. Cy3, Cy5, AlexaFluor435 etc.)

A long dsDNA encoding for the pre-mRNA target can be synthesized by the assembly PCR method as reported by Romesberg (see Figure 1-47).¹³³ Transcription of this template employing unnatural nucleotide triphosphates featuring orthogonal reactive groups (e.g. an amine and an alkyne) would then afford an RNA strand, which could be labeled with two fluorophores in a chemoselective and sequential fashion (e.g. by a “click” reaction for the alkyne and then by an amine-NHS coupling (Figure 1-49).

1.16 Aims and Objectives

The overall objective of this thesis is to develop a novel unnatural base pair, in which both ribonucleotides can be easily synthesized and functionalized with reactive groups suitable for labeling. This thesis will focus towards addressing the following specific aims:

- 1) To develop a novel strategy utilizing the Heck reaction for the efficient synthesis of C-ribonucleosides and analogues of Benner's **Z** as potential candidates for a new generation of orthogonal base pairs (Figure 1-50).

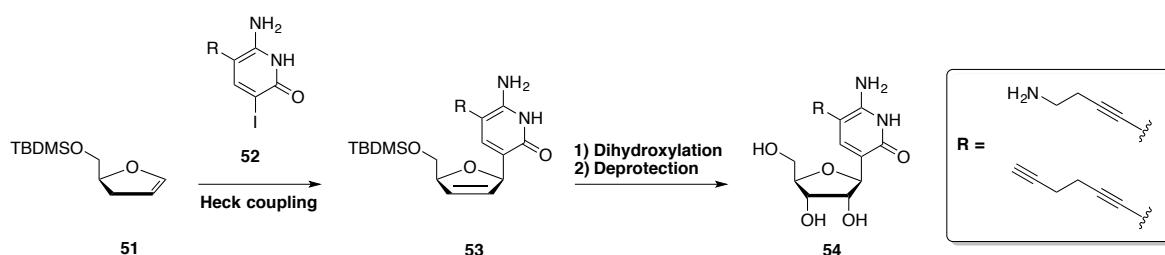


Figure 1-50. Proposed strategy to access C-ribonucleoside **Z** and functionalized analogues thereof.

- 2) To synthesize C-6 substituted analogues of Hiraio's **s** nucleotide as potential candidates for a new generation of orthogonal base pairs (Figure 1-51).

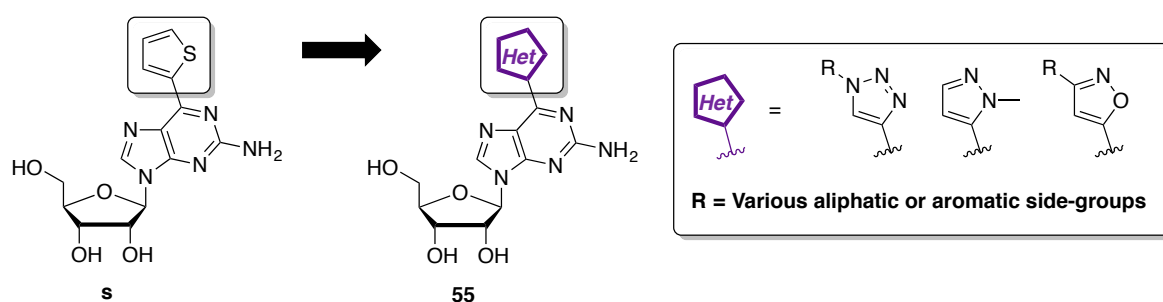


Figure 1-51. Proposed analogues of Hiraio's **s** nucleotide.

- 3) To evaluate the transcriptional efficiency and fidelity of these analogues using a DNA template modified with the cognate nucleotide (**Pa**, Figure 1-52).

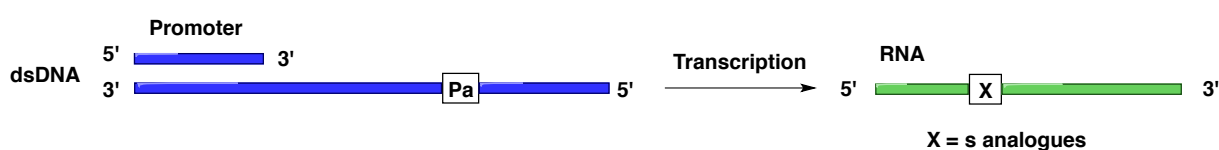


Figure 1-52. Schematic representation of the proposed assay to be employed for the transcriptional evaluation of unnatural nucleotides.

Chapter 2

2 Synthesis of C-ribonucleoside Analogues as Unnatural Base Pairs Directed by Hydrogen Bonding

2.1 Introduction

The orthogonal base pairing system initially chosen for the development of orthogonally modified RNA in this project is Benner's **Z/P** pair (Figure 2-1a). This is due to the following reasons: (i) Literature precedent showing efficient incorporation by PCR (**Z/P** retention averaged 99.8% per theoretical PCR cycle under the optimized conditions developed by the authors)¹¹⁰ and (ii) The presence of positions suitable for derivatization (Figure 2-1b).

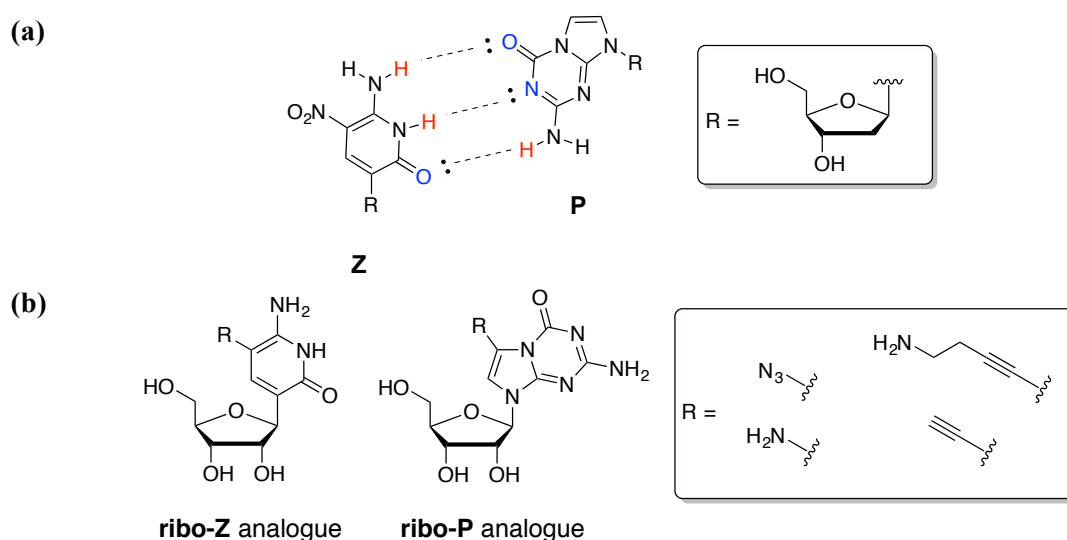


Figure 2-1. (a) The **Z/P** base pair. (b) Potential derivatives of **ribo-Z** and **ribo-P** suitable for labeling.

The **Z** is a C-nucleoside, unlike most naturally occurring nucleosides (which are N-nucleosides). In C-nucleosides the atom attached to the anomeric center is carbon, while in N-nucleosides is nitrogen (Figure 2-2a).

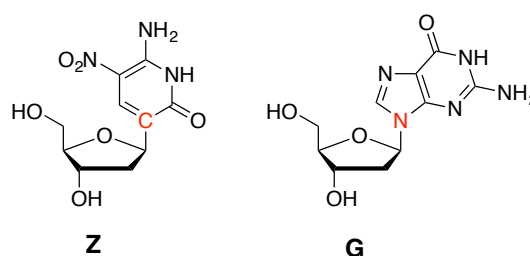


Figure 2-2. An example of a C-nucleoside is the **Z** (left). An example of an N-nucleoside is **G** (right). The atom attached to the anomeric center is marked in red.

The Benner group reported that this C-C bond is important in pyrimidine analogues implementing the hydrogen-bonding pattern of **Z**, because it results in an aromatic ring that enhances stacking (Figure 2-3a).¹³⁵ This is due to the heteroatom arrangement around the ring that if such an analogue was an N-nucleoside, it would feature only one double bond, disrupting the aromaticity in compound **57** (Figure 2-3b).

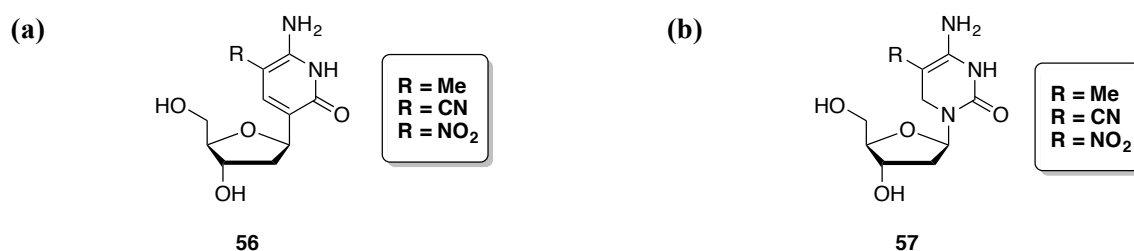
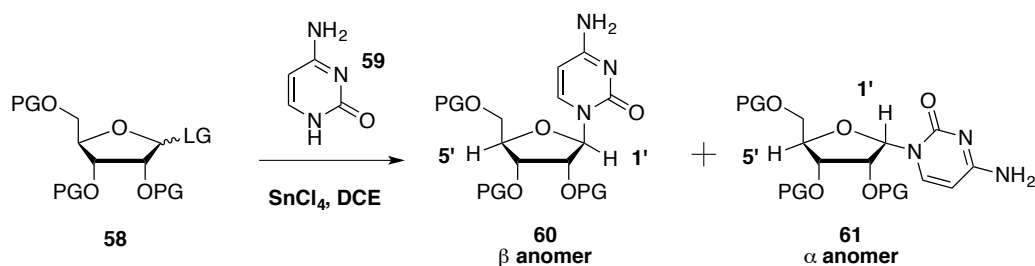


Figure 2-3. (a) C-nucleoside of **Z** analogues, which implement the donor-donor-acceptor hydrogen-bonding pattern. (b) N-nucleoside version of the same analogues.

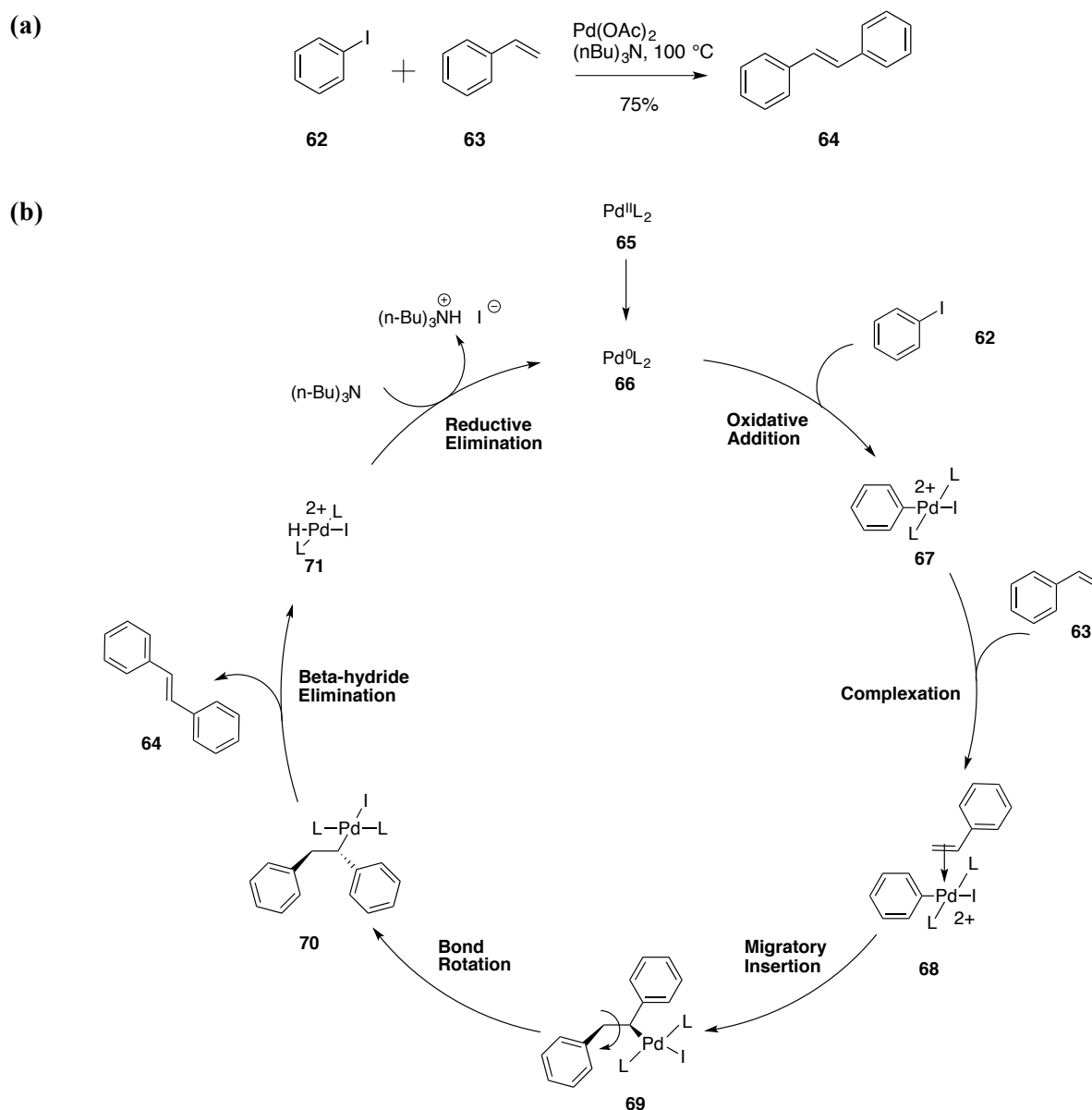
N-nucleosides are typically synthesized through glycosylation of a heterocycle such as **59** with sugar donor **58** using a Lewis acid such as TMSOTf or SnCl₄ in 1,2-dichloroethane (Scheme 2-1). This reaction can yield two different stereoisomers, β anomer **60** and α anomer **61**. In β anomers, the base is *cis* to the 5'-OH whereas in α anomers, it is *trans*. Since the β anomer corresponds to natural nucleotides, the α/β ratio should be taken into account when carrying out a glycosylation reaction. For C-nucleosides different approaches have been devised over the years, since unlike primary or secondary nitrogen atoms a carbon atom is not inherently nucleophilic.¹³⁶



Scheme 2-1. Schematic representation of a glycosylation reaction. LG = Leaving Group (e.g. -OAc, -Cl). PG = Protecting Group (e.g. TBDMS, Ac).

2.1.1 The Heck Reaction in the Synthesis of C-deoxyribonucleosides

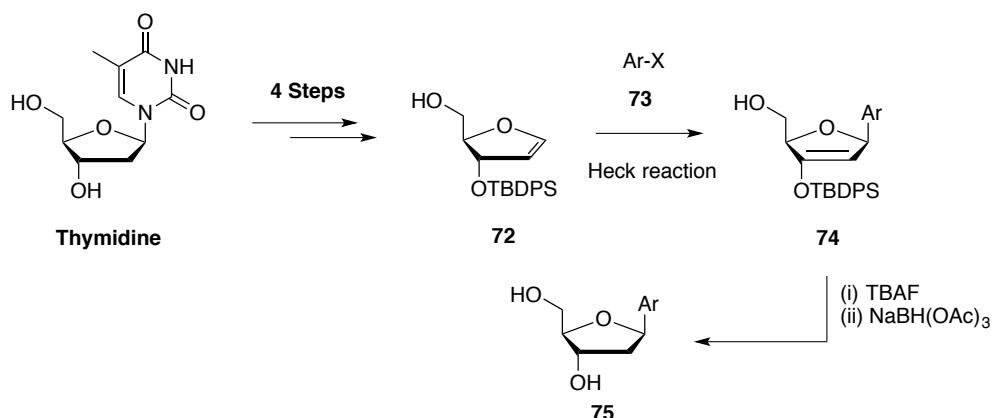
The Heck coupling is one of the most commonly employed strategies for C-deoxyribonucleoside synthesis.¹³⁷ This is a variation from Heck's original work.¹³⁸ Heck and co-workers coupled iodide **62** with alkene **63** using palladium(II) acetate and tributylamine as the base to afford **64** in 75% yield (Scheme 2-2a). The catalytic cycle is presented below (Scheme 2-2b).



Scheme 2-2. (a) First example reported by Heck et al.¹³⁷ (b) Generally accepted catalytic cycle of the Heck reaction.¹³⁹

The reaction begins with the reduction of Pd(II) complex **65** to Pd(0) complex **66**. **66** undergoes oxidative addition with iodide **62**, yielding Pd(II) complex **67**. Alkene **63** then coordinates to this complex through its π -bond to form intermediate **68**. The phenyl group adds to the coordinated alkene, forming **69**, which undergoes σ -bond rotation to **70** and subsequent beta-hydride elimination releasing the product and forming complex **71**. The final step is reductive elimination forming the active catalyst **66** and hydrogen iodide, which is scavenged by the base.

For the synthesis of C-deoxyribonucleosides, glycal **72** is synthesized in four steps from thymidine, and then coupled with the corresponding halogenated nucleobase **73**.^{137,140} Migration of the double bond is observed between carbons 2' and 3' and enol-ether **74** is obtained as the β -anomer (Scheme 2-3).¹⁴¹ Treatment of enol ether **74** with TBAF results in the formation of a ketone, which is directly reduced in a stereoselective fashion using $\text{NaBH}(\text{OAc})_3$ to give rise to **75**.¹⁴²



Scheme 2-3. The Heck reaction in the synthesis of deoxyribonucleosides.¹³⁷

The stereoselectivity in the Heck coupling step results from the bulky protecting group on the 3'-OH of intermediate **72** blocking the *re* face of the molecule and the free 5'-OH coordinating to the palladium catalyst, thus directing the addition of the aryl moiety to the alkene from the *si* face (Figure 2-4).¹⁴²

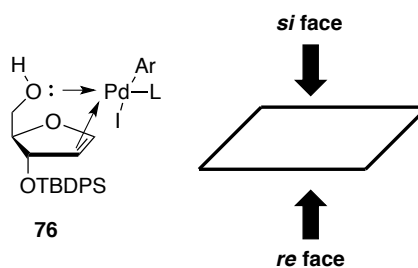
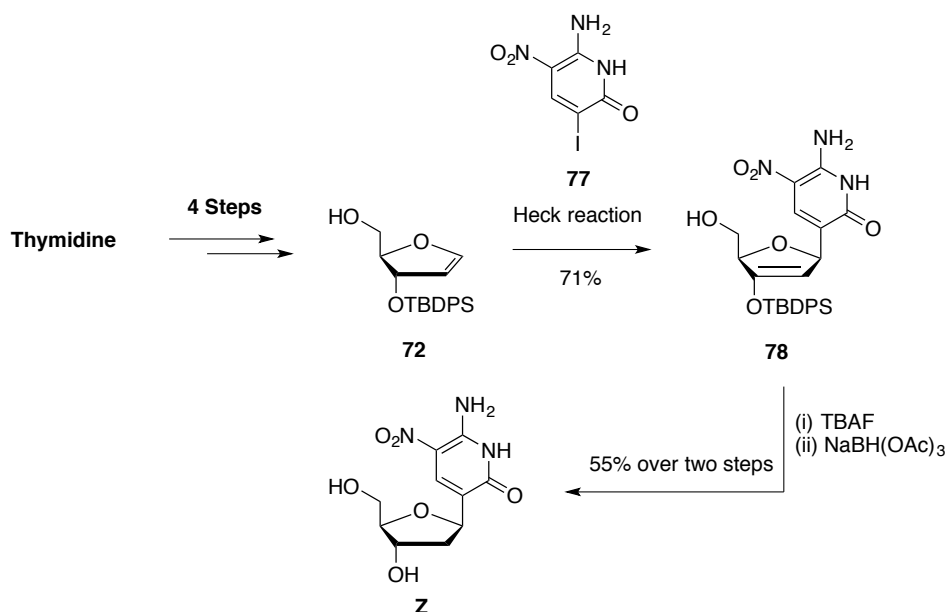


Figure 2-4. Postulated intermediate of the Heck reaction of an iodinated heterocycle with glycal **76**. L denotes a ligand (e. g. AsPh_3). Ar denotes an iodinated coupling partner.¹⁴¹

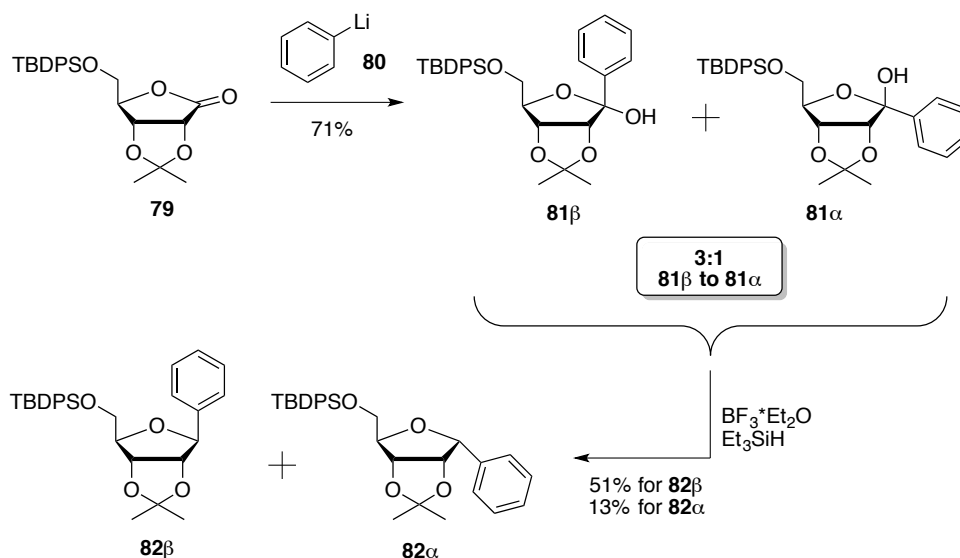
The deoxyribose version of **Z** has been already synthesized by Hutter and Benner using the Heck coupling strategy (Scheme 2-4).¹⁰⁹



Scheme 2-4. Benner's synthesis of nucleoside **Z**.¹⁰⁹

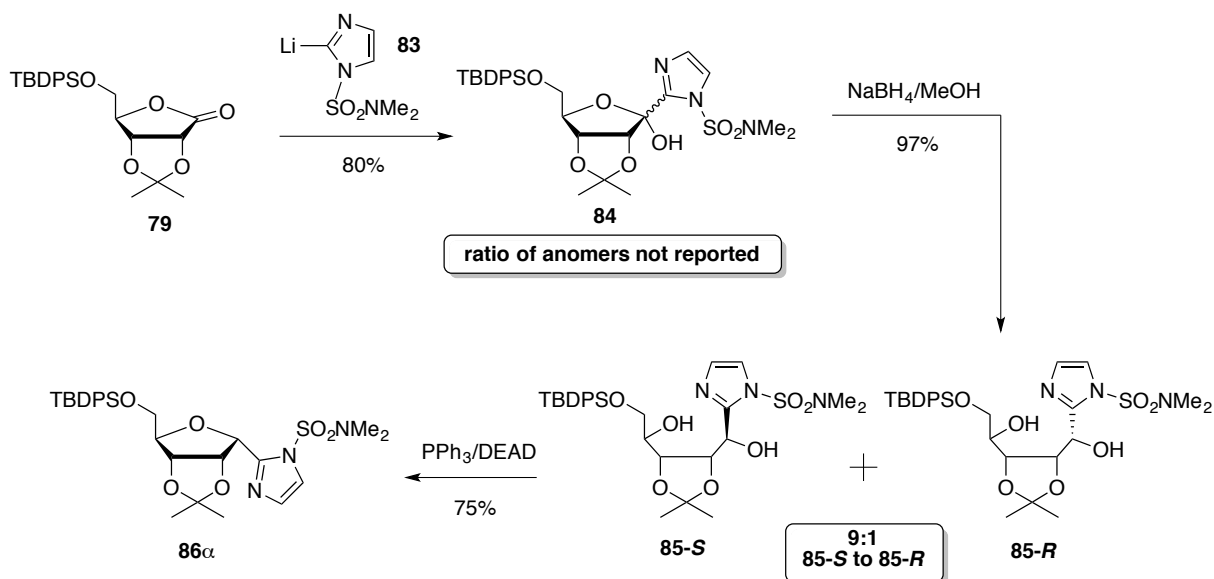
2.1.2 Current Challenges in the Synthesis of C-ribonucleosides

Methods for C-ribonucleoside synthesis have not been explored as widely as their deoxy-analogues. For the construction of C-ribonucleosides, the most common strategy is the nucleophilic addition of an organometallic reagent to a lactone. For example, lactone **79** reacted with organolithium reagent **80** affording an inseparable mixture of **81 β** and **81 α** in 3:1 ratio and 71% combined yield.¹⁴³ Treatment of this mixture with Et_3SiH and $\text{BF}_3 \cdot \text{Et}_2\text{O}$ resulted in 1' deoxygenation, affording a separable mixture of **82 β** and **82 α** , which were isolated in 51% and 13% yields respectively (Scheme 2-5).



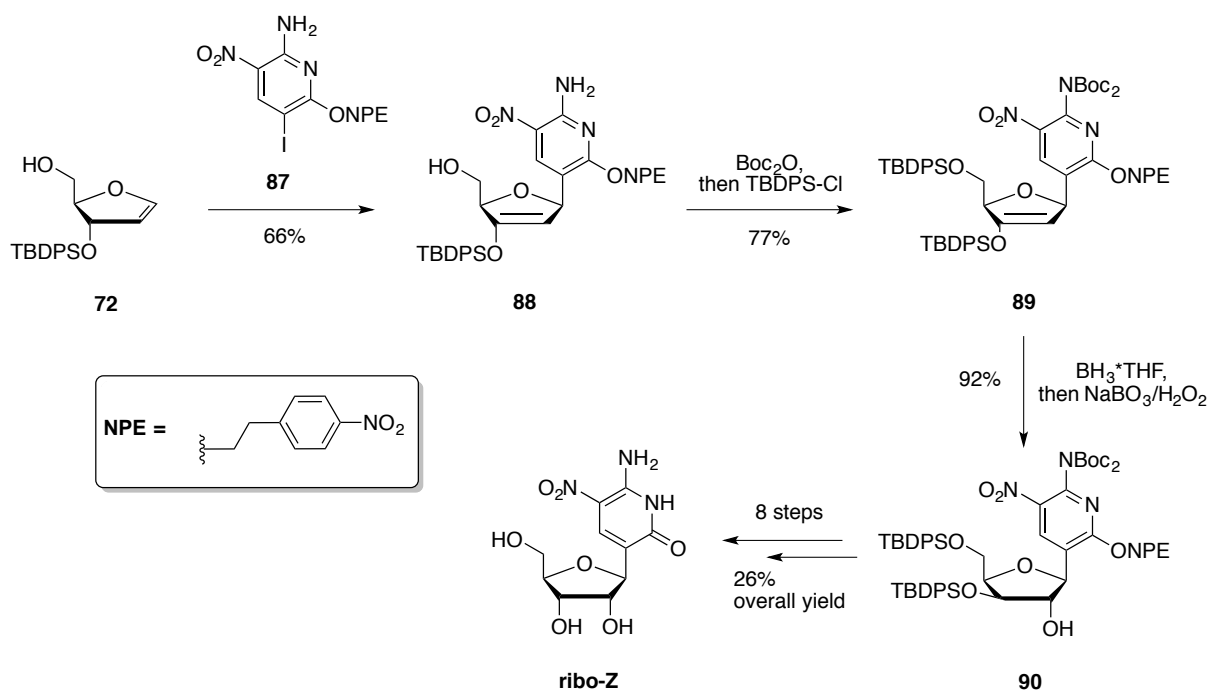
Scheme 2-5. Example of C-ribonucleoside synthesis by nucleophilic addition to a ribolactone and subsequent 1' deoxygenation.¹⁴²

Alternatively, deoxygenation can be achieved *via* reduction of hemiacetal **84** to a diol and subsequent Mitsunobu etherification. An exemplar of this strategy is described in Scheme 2-6. Addition of organolithium species **83** to ribolactone **79** furnished hemiacetal **84**. Reduction of **84** with NaBH₄ in MeOH afforded diol **85** as a mixture of *S* and *R* isomers in 97% combined yield and 9:1 *S/R* ratio. Cyclization of this mixture under Mitsunobu conditions afforded nucleoside **86** as the α anomer exclusively.¹⁴⁴



Scheme 2-6. Example of C-ribonucleoside synthesis by nucleophilic addition to a ribolactone, followed by hemiacetal reduction and Mitsunobu cyclization.¹⁴³

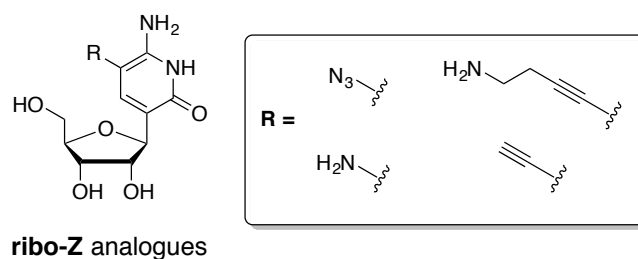
The two strategies described above are far from robust. The yield and anomeric ratio of the addition/deoxygenation reactions depend highly on the ribolactone protecting groups and the nature of the organometallic species.^{145,146} Therefore, experimentation is sometimes required to optimize the yield and $\alpha:\beta$ ratio, which can be time consuming.¹⁴⁵ Furthermore, inseparable mixtures of various products have been observed by some groups, while others reported problems generating the metallated heterocycle.^{147,148} Recently, Benner reported the synthesis of **ribo-Z**, using the Heck reaction.¹⁴⁷ The key step is the installation of the 2'-OH *via* a hydroboration/oxidation reaction of enol-ether **89** (Scheme 2-7). However, this step results in inversion of the 3'-OH stereochemistry and requires a lengthy synthetic strategy (9 steps from **89**) to install the correct stereochemistry on the 3'-OH and produce **ribo-Z** in 26% overall yield (Scheme 2-7).

Scheme 2-7. Benner's synthesis of **ribo-Z**.¹⁴⁷

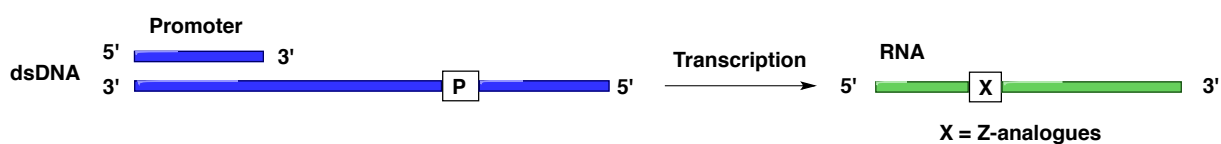
2.2 Aims of this Study

The aims of this chapter are to:

- 1) develop a robust synthetic methodology to access **ribo-Z** and analogues thereof (Figure 2-5).

Figure 2-5. Potential analogues of **ribo-Z** suitable for labeling.

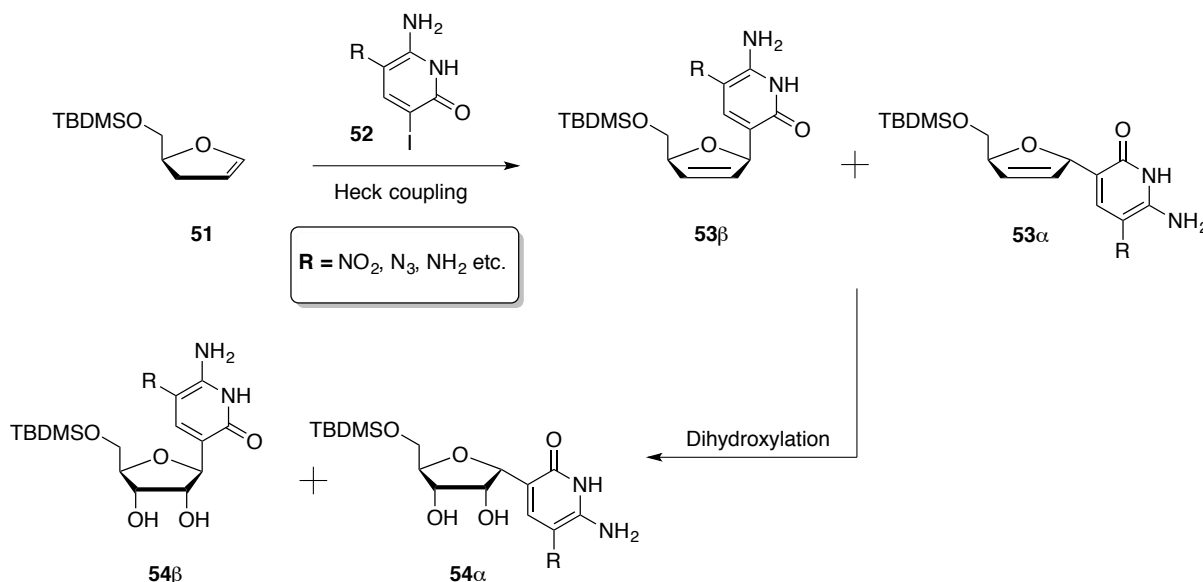
- 2) understand the base-pairing fidelity of **ribo-Z** in transcription (Figure 2-6).

Figure 2-6. Schematic representation of the proposed assay to be employed for the transcriptional evaluation of **Z**-analogues.

2.3 Results and Discussion

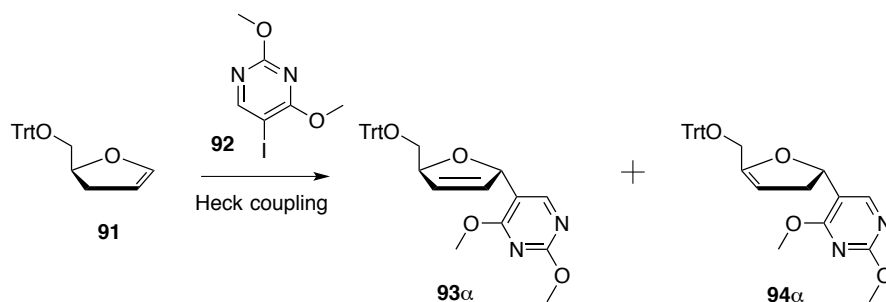
2.3.1 Design of the Synthetic Strategy Involving the Heck Reaction

It was envisaged that the Heck strategy could be modified as a robust methodology for ribonucleoside access. According to Scheme 2-8, glycal **51**.¹⁴⁹ could participate in a Heck coupling with heterocycle **52**. Alkene **53** could then be dihydroxylated *via* either Sharpless asymmetric dihydroxylation or Upjohn non-asymmetric conditions to furnish **54**.^{150,151}



Scheme 2-8. Proposed strategy for efficient access to derivatized C-ribonucleosides.

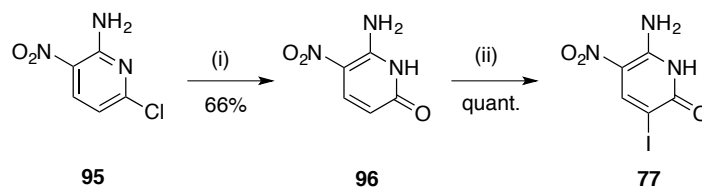
However, a bulky protecting group in the 3'-OH and a free 5'-OH are required for β -anomer selectivity; thus, using glycal **51** could potentially lead to the formation of the undesired α -anomer or a mixture of both. Indeed, Daves et al. reported that coupling of glycal **91** with pyrimidine **92** led to the formation of a mixture of regioisomeric α -anomers **93** and **94** (Scheme 2-9). Depending on the conditions used, the yields of **93 α** :**94 α** ranged from 53:30 to 40:0%.¹⁴¹



Scheme 2-9. Heck coupling of **88** with **89** led to the formation of a mixture of regioisomeric α -anomers.¹⁴⁰

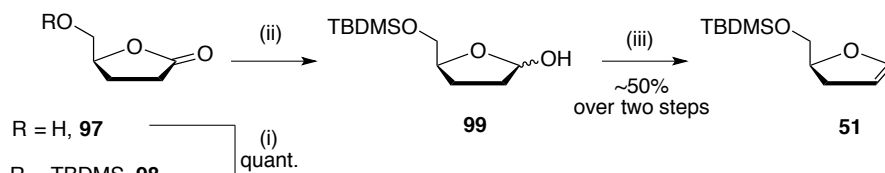
2.3.2 Synthesis of Heck Coupling Partners

The first step was to synthesize Heck coupling substrates **77** and **51**. Heterocycle **77** was prepared in three steps starting from commercially available pyridine **95** (Scheme 2-10).¹⁰⁹ Compound **95** was treated with NaOH in refluxing EtOH/H₂O furnishing pyridinone **96** in 66% yield. Iodination with NIS in DMF afforded iodide **77** in quantitative yield.



Scheme 2-10. Reagents and conditions: (i) 10 M aq. NaOH (8.0 equiv.), H₂O/EtOH (7:3), reflux, 5 min. (ii) NIS (1.5 equiv.), DMF, r.t., 4 days.

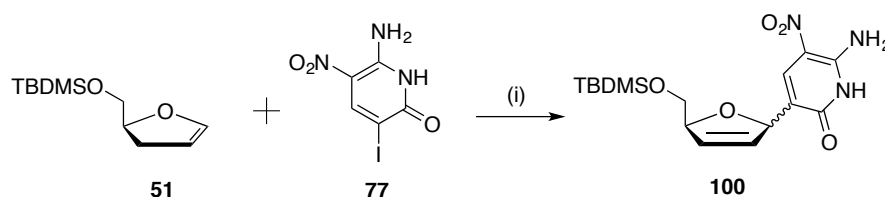
Glycol **51** was synthesized according to literature procedures in three steps (Scheme 2-11).^{149,152} The alcohol moiety of starting lactone **97** was protected as a TBDMS-ether with TBDMS-Cl and imidazole as a base in DMF affording lactone **98** quantitatively. Subsequent reduction with DIBAL-H yielded lactol **99**, which was directly treated with MsCl and Et₃N to yield glycol **51** (~50% over two steps).



Scheme 2-11. Reagents and conditions: (i) TBDMS-Cl (1.3 equiv.), imidazole (2.5 equiv.), DMF, 0 °C to r.t., 16 h. (ii) DIBAL-H (1.3 equiv.), DCM, -78 °C, 1 h. (iii) MsCl (1.3 equiv.), Et₃N (3.4 equiv.), DCM, -20 °C, 30 min; reflux, 2.5 h.

2.3.3 Investigation of Heck Coupling Conditions Towards Ribo-Z Synthesis

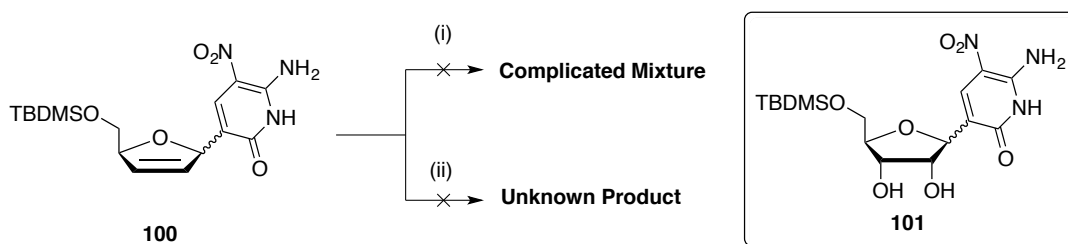
With **77** and **51** in hand, conditions for the Heck coupling were investigated (Scheme 2-12). Heck conditions that have been used by the Benner group in the synthesis of the DNA analogue of **Z** were investigated.¹⁰⁹ After 6 days at 60 °C, a single product was observed on TLC.



Scheme 2-12. Reagents and conditions: (i) glycol **51** (1.3 equiv.), Pd(OAc)₂ (0.1 equiv.), AsPh₃ (0.2 equiv.), Et₃N (1.4 equiv.), DMF, 60 °C, 6 days.

110 mg were isolated after column chromatography purification. ESI-MS analysis showed a signal corresponding to the expected product **100**; however, HRMS analysis attempts could not detect the desired

mass ion. $^1\text{H-NMR}$ analysis revealed the presence of the peaks corresponding to **100** (e.g. the proton of the heterocycle at 8.2 ppm and the $\text{H}1'$ at 5.9 ppm); however, the overall quality of the spectrum was poor due to the presence of various unidentifiable impurities in the sugar region and residual DMF. HPLC purification using the standard method (see the general experimental section) was unsuccessful and the product was not detected eluting from the column (as determined by the absence of a UV absorption signal). This could be due to the non-polar nature of **100**, which could strongly adhere to the stationary phase. Two different dihydroxylation strategies were then carried out to transform Heck product **100** to diol **101** (Scheme 2-13). The presence of the two extra hydroxyls was expected to increase the polarity of the compound and aid in HPLC purification. Sharpless dihydroxylation of **100** resulted in a complex mixture of products, which could not be purified. ESI-MS analysis of the crude mixture did not show formation of the desired product either.¹⁵³ The mixture of compounds could be a result of the K_2CO_3 present in AD-mix, which could cleave the TBDMS group under the presence of protic solvents. Furthermore, co-ordination of the osmium catalyst to the nitrogen-rich heterocycle might have hampered the catalytic cycle and resulted in the formation of various byproducts. Upjohn asymmetric dihydroxylation conditions (K_2OsO_4 , NMO) afforded a mixture of two products.¹⁵¹ The major product was isolated after column chromatography and 32 mg were obtained. ESI-MS analysis showed a signal corresponding to the expected product **101**; however, $^1\text{H-NMR}$ analysis failed to provide clear and identifiable signals that confirm the identity of the product.



Scheme 2-13. Reagents and conditions: (i) AD-mix- α (0.3 equiv. of K_2OsO_4), $t\text{-BuOH}/\text{H}_2\text{O}$ (1:1), r. t., 16 h. (ii) K_2OsO_4 (5 mol%), NMO (3.0 equiv.), $t\text{-BuOH}/\text{H}_2\text{O}/\text{Acetone}$ (2:1:0.5), r.t., 16 h.

It was decided to carry on with assigning the stereochemistry of **100** and a $^1\text{H-NOESY}$ NMR experiment was performed (Appendix 8). A clear correlation was observed between H^1 and H^2 and between H^1 and H^3 . In addition, no correlation between H^2 and the H^5 was observed. If the product was the β -anomer, H^2 would be on the same side as the H^5 and thus an NOE signal between them would be expected. These results suggested that the stereochemistry of compound **100** at $\text{H}1'$ was the undesired of an α -anomer (Figure 2-6). Due to this observation, as well as the difficulties encountered during the Heck and dihydroxylation reactions, this strategy was discontinued.

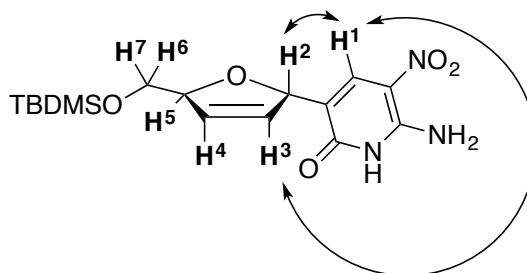
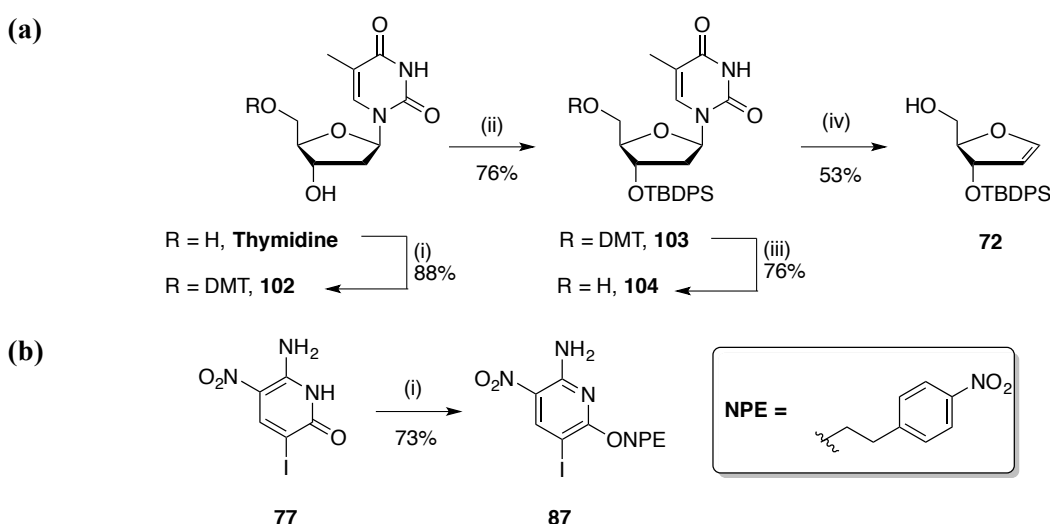


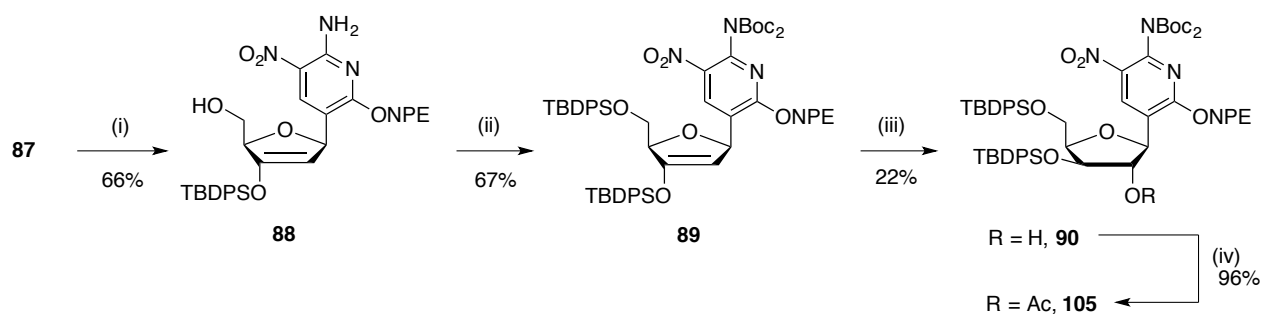
Figure 2-6. NOE correlations observed for **100**.

Benner's original synthetic route was next explored (Scheme 2-14).¹⁴⁷ Glycal **72** was synthesized in four steps according to the procedure reported by McLaughlin et al. (Scheme 2-14a).¹⁵⁴ Heterocycle **77** was protected as its NPE-ether (Scheme 2-14b).¹⁴⁶



Scheme 2-14. (a) Reagents and conditions: (i) DMT-Cl (1.5 equiv.), pyridine, 50 °C, 1 h. (ii) TBDPS-Cl (1.1 equiv.), imidazole (2.5 equiv.), DMF, 0 °C to r.t., 16 h. (iii) PTSA (3.2 equiv.), MeOH/DCM (2.3:1), 0 °C to r.t., 1 h. (iv) HMDS, (NH₄)₂SO₄ (0.25 equiv.), reflux, 3 h. **(b) Reagents and conditions:** (i) 2-(4-nitrophenyl)ethan-1-ol (1.5 equiv.), PPh₃ (1.5 equiv.), DEAD (1.5 equiv.), THF, 0 °C to r.t., 1 h.

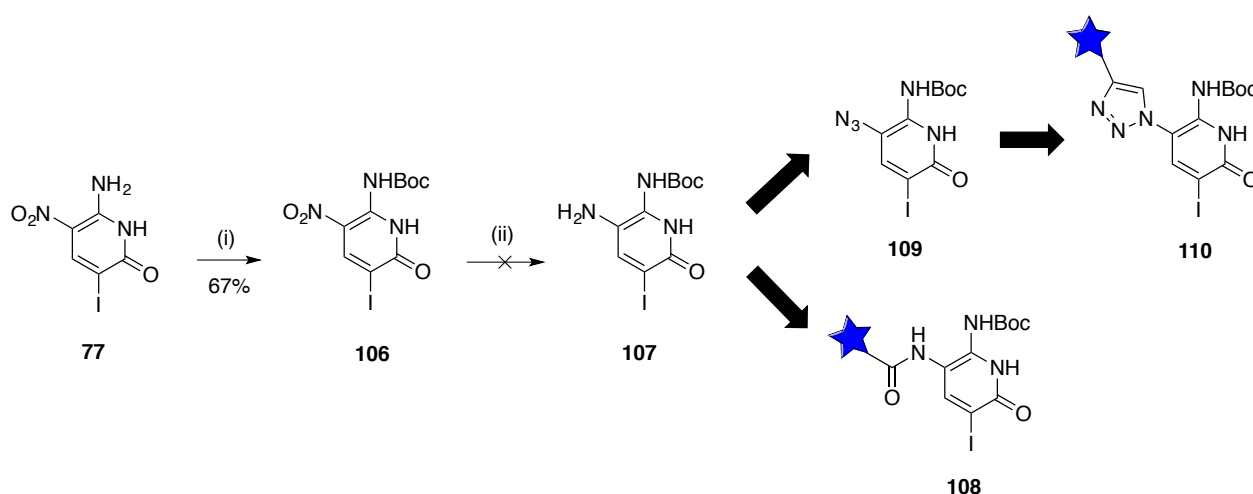
Heck coupling of **87** with **72** afforded **88** in 66% yield. The 5'-OH was then protected with a TBDPS group and the exocyclic amine as di-Boc carbamate affording **89**. The key step was a hydroboration reaction with BH₃•THF. Unfortunately, TLC analysis did not show product formation. Using the dimethylsulfide adduct, alcohol **90** was obtained in 22%. Subsequent acetylation with Ac₂O in DCM afforded acetate **105** in 96% yield (Scheme 2-15). At this stage, this strategy was paused to explore the transformation of the nitro moiety in **77** to moieties suitable for bioconjugation with fluorescent tags.



Scheme 2-15. *Reagents and conditions:* (i) glycal **72** (1.2 equiv.), Pd(OAc)₂ (0.1 equiv.), AsPh₃ (0.2 equiv.), Et₃N (2.0 equiv.), DMF, 70 °C, 16 h. (ii) TBDPS-Cl (1.5 equiv.), imidazole (3.1 equiv.), DMF, r.t., 16 h; then Boc₂O (3.0 equiv.), Et₃N (5.6 equiv.), DMAP (cat.), DCM, r.t., 1 h. (iii) BH₃•Me₂S (5.0 equiv.), THF, 0 °C to r.t., 16 h; then NaBO₃ (24 equiv.), H₂O, r.t., 2 h. (iv) Ac₂O (1.1 equiv.), pyridine (4.4 equiv.), DCM, r.t., 2 h.

2.3.4 Attempts to Prepare Z analogues

The investigation of analogues of **Z** bearing bioconjugable groups off the nucleobase at position 5 (Scheme 2-13) was undertaken in parallel to the preparation of **Z**. The reduction of the nitro group in **77** to an amine was explored. Amine **107** could participate in amide bond formation (NHS-coupling) or transformed into azide **109**, which could then participate in a [3+2] cycloaddition with an alkyne, affording triazole **110** (Scheme 2-13). The first step was to protect the exocyclic amine as Boc-carbamate. Using Benner's conditions for protection of **81** (see Scheme 2-12), mono-Boc protection was observed and **106** was obtained in 67% yield.¹⁴⁷ Reduction of **106** using sonication would avoid the use of strongly acidic conditions and high temperatures (which could result in Boc deprotection).¹⁵⁵ A single product was formed and isolated. ¹H-NMR analysis failed to provide any identifiable product peaks. LC-MS analysis did not reveal signals corresponding to the expected product.



Scheme 2-16. *Reagents and conditions:* (i) Boc₂O (3.0 equiv.), Et₃N (5.5 equiv.), DMAP (0.1 equiv.), DMF, r.t., 2 h. (ii) Fe powder (5.0 equiv.), AcOH/EtOH/H₂O, 40 °C, 2 h.

The AcOH present might have resulted in conditions sufficiently acidic for the cleavage of the Boc group, thus forming a highly polar diamine. This would result in loss of compound during the aqueous work-up of

the reaction. Switching to a different protecting group for the amine (e.g. Fmoc, acetamide) would circumvent this problem. These investigations were not initiated though; PCR experiments conducted by our collaborators at the University of Leicester revealed significant levels of misincorporation, arising from a **Z/G** mismatch. This mismatch would eventually lead to loss of the **Z/P** pair and replacement with a **C/G** pair. PCR is crucial for the synthesis and amplification of a modified long DNA template (see Figure 1-47). Consequently, PCR with a functionalized **Z** analogue could potentially lead to loss of the modification resulting in a natural DNA template incorporating the **C/G** pair. Due to this observation, the difficulties encountered in obtaining a pure sample of compound **100**, the difficulties encountered in the dihydroxylation of **100** to **101** (see Scheme 2-13) and the NOESY data suggesting the stereochemistry of compound **100** at H1' was the undesired of an α -anomer, a new strategy was pursued based on the base pairs developed by the Hirao group.

2.4 Summary

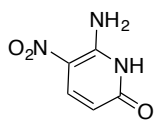
The investigations presented in this chapter lead to the following conclusions: The Heck coupling between **77** and **51** requires long reaction times (6 days) and yields a single product, but NOESY-NMR experiments suggested the H1' stereochemistry was that of the undesired α -anomer. Furthermore, dihydroxylation attempts of **100** afforded either a complex mixture of products, or yielded a major product that failed to provide peaks corresponding to **101** in $^1\text{H-NMR}$.

2.5 Experimental

Dry DCM and THF were obtained from an Innovative Technology Pure Solv. SPS-400-5 solvent purification system. All other solvents and reagents were supplied by Sigma Aldrich, Acros Organics, Alfa Aesar, Fisher Scientific and Fluorochem, unless stated otherwise. Thin layer chromatography (TLC) was performed on Merck aluminum backed silica plates coated with a fluorescent detector F₂₅₄. These were visualized using a Mineralight UVGL-25 lamp or developed with a KMnO₄ solution. Flash column chromatography was performed on Apollo Scientific ZEOPrep 60 silica (40-63 microns particle size). ^1H and $^{13}\text{C-NMR}$ spectra were obtained on a Bruker DPX 500, spectrometer at 500 MHz and 125 MHz respectively. Chemical shifts were recorded in parts per million (ppm, δ), J coupling constants are stated in Hertz (Hz). Low resolution mass spectra were recorded either on a Finnigan LCQ-DUO spectrometer or on an Agilent 1200 series LC-MS, equipped with a Zorbax 45mm x 150mm C18 column. A standard 45 min program using buffer C (5 mM ammonium acetate in water) and buffer D (5 mM ammonium acetate in acetonitrile) was used for all LC-MS analyses. The flow rate was 1.000 mL/min with a gradient as follows. 0-3 minutes held at 5% buffer D, then a linear gradient from 5-100% for 30 minutes, then the gradient was returned to 5% D and held at 5% buffer D for 3 minutes. The MS used was dual source with a capillary voltage of 4000 Volts for positive and negative modes and with a gas vaporizer temperature of 250 °C. HPLC analysis of **100** was carried out on a DIONEX RP-HPLC system equipped with a Phenomenex Clarity 5u Oligo RP 250

mm x 4.6 mm C18 column. A standard 55 min program using buffer A (0.1 M TEAAc in H₂O) and buffer B (0.1 M TEAAc in 80% MeCN/H₂O) was used. The flow rate was 1.0 mL/min with a gradient as follows. 0-3 minutes held at 5% buffer B, then a linear gradient from 5-95% for 40 minutes, then it was held at 95% buffer B for 5 minutes, then the gradient was returned to 5% B and held at 5% buffer B for 5 minutes.

6-amino-5-nitropyridin-2(1H)-one **96**¹⁰⁹

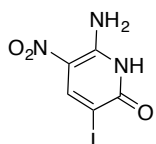


A suspension of 6-chloro-3-nitropyridin-2-amine (4 g, 23 mmol) in EtOH (34 mL) and H₂O (92 mL) was heated to reflux. A solution of 10 M aq. NaOH (18 mL, 184 mmol) was then added and the resulting red solution was stirred under reflux for 5 minutes. It was then cooled down to 0 °C and neutralized with HCl (conc). Filtration of the resulting yellow precipitate and drying over CaCl₂ under reduced pressure afforded **96** as a dark yellow solid (3.5 g, 84%).

¹H-NMR (500 MHz, DMSO-d₆) δ: 11.20 (br s, 1H), 8.60 (br s, 1H), 8.01 (d, *J* = 10.3 Hz, 1H), 7.60 (brs, 1H), 5.68 (d, *J* = 10.3 Hz, 1H).

ESI-MS (+ve mode): [M]⁺ 155.

6-amino-3-iodo-5-nitropyridin-2(1H)-one **77**¹⁰⁹

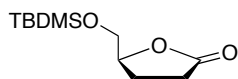


A solution of **96** (3.5 g, 22.5 mmol) and NIS (7.5 g, 33.8 mmol) in dry DMF (70 mL) was stirred in the dark under argon for four days at room temperature. The red solution was poured into vigorously stirred H₂O (3 L) and the yellow suspension stirred under air for 16 h. The suspension was filtered and dried to furnish **77** as a dark yellow solid (6 g, 94%).

¹H-NMR (500 MHz, DMSO-d₆) δ: 11.56 (brs, 1H), 8.49 (s, 1H).

ESI-MS (-ve mode): [M-H]⁻ 279.

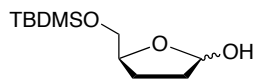
(S)-5-(((tert-butyldimethylsilyl)oxy)methyl)dihydrofuran-2(3H)-one **98**¹⁴⁹



A mixture of alcohol **97** (3 g, 24 mmol), TBDMS-Cl (0.031 mL, 31.20 mmol) and imidazole (4 g, 60 mmol) in DMF (12 mL) was stirred under an argon atmosphere for 16 h at room temperature. The reaction mixture was then poured into brine (60 mL) and extracted with hexane (2 × 40 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated *in vacuo*. Column chromatography purification (SiO₂) eluting with EtOAc/*n*-hexane (2:8) afforded **98** as a colorless oil (7.1 g, quant.).

¹H-NMR (500 MHz, CDCl₃) δ: 4.55 (m, 1H), 3.65 (dd, *J* = 11.2, 3.4 Hz, 1H), 3.65 (dd, *J* = 11.2, 3.4 Hz, 1H), 2.56 (m, 2H), 2.20 (m, 2H), 0.90 (s, 9H), 0.10 (s, 6H).

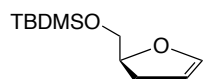
ESI-MS (+ve mode): [M+Na]⁺ 231.

(5S)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-ol 99¹⁵²

A solution of ketone **98** (1.55 g, 6.73 mmol) in DCM (8.5 mL) was cooled to -78°C . DIBAL-H (1.2 M in toluene, 7.3 mL, 8.8 mmol) was then added and the solution was stirred under argon for 2 hours under argon. The reaction mixture was poured into 0.2 M aq. HCl (50 mL) and extracted with DCM (2×30 mL). The organic phases were combined and washed with sat. NaHCO_3 (2×20 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. The crude material was used in the next step without further purification.

$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 5.40 (d, $J = 8$ Hz, 1H), 4.20 (m, 1H), 3.80 (dd, $J = 2.9, 10.7$ Hz, 1H), 3.76 (dd, $J = 1.9, 8.1$ Hz, 1H), 3.59 (m, 1H), 2.1 (m, 1H), 1.67-1.89 (m, 2H), 0.93 (s, 9H), 0.11 (6H, s).

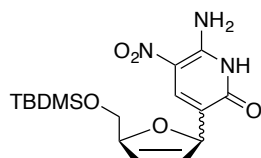
ESI-MS (-ve mode): $[\text{M}-\text{H}_2\text{O}]^-$ 215.

(S)-tert-butyl((2,3-dihydrofuran-2-yl)methoxy)dimethylsilane 51¹⁵²

A solution of lactol **99** (200 mg, 0.86 mmol) in DCM (2.15 mL) was cooled to -50°C . Et_3N was added (0.4 mL, 2.9 mmol), followed by the addition of MsCl (0.1 mL, 1.1 mmol). The solution was stirred at -50°C under argon for 1 hour and then refluxed for 2 hours. The solution was then poured into sat. NaHCO_3 (30 mL). The organic phases were combined and washed with sat. NH_4Cl (2×20 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Quick column chromatography purification (SiO_2) eluting with DCM afforded **51** as a colorless oil (92 mg, 50%).

$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 6.26 (q, $J = 2.4$ Hz, 1H), 4.65 (q, $J = 2.6$ Hz, 1H), 4.60 (dd, $J = 10.7, 6.5$ Hz, 1H), 3.71 (dd, $J = 10.6, 6.0$ Hz, 1H), 3.61 (dd, $J = 10.6, 6.0$ Hz, 1H), 2.63 (ddt, $J = 15.1, 10.3, 2.3$ Hz, 1H), 2.39 (ddt, $J = 15.1, 7.3, 2.4$ Hz, 1H), 0.91 (s, 9H), 0.06 (s, 3H), 0.07 (s, 3H).

ESI-MS (+ve mode): $[\text{M}]^+$ 215.

6-amino-3-(5-(((tert-butyldimethylsilyl)oxy)methyl)-2,5-dihydrofuran-2-yl)-5-nitropyridin-2(1H)-one 100¹⁰⁹

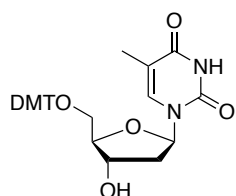
A mixture of $\text{Pd}(\text{OAc})_2$ (21 mg, 0.093 mmol) and AsPh_3 (57 mg, 0.2 mmol) in DMF (1.5 mL) was stirred for 30 minutes under argon at room temperature. A solution of glycol **51** (200 mg, 0.9 mmol) and pyridinone **77** (201 mg, 0.7 mmol) in DMF (1.5 mL) was dried over 3\AA molecular sieves overnight and then added to the catalyst suspension, together with Et_3N (0.2 mL, 1.3 mmol). The solution was then stirred under argon for 6 days at 60°C . It was then poured into brine (30 mL) and extracted with EtOAc (3×20 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with EtOAc/*n*-hexane (3:7 \rightarrow 6:4) afforded 110 mg of **100** as a brownish foam.

$^1\text{H-NMR}$ (400 MHz, CDCl_3). δ : 11.48 (brs, 1H), 8.21 (s, 1H), 5.88 (d, $J = 5.74$ Hz, 1H), 4.97 (q, $J = 4.53$, 4.72 Hz, 1H), 4.12 (m, 1H), 3.81 (m, 1H), 3.60 (m, 1H), 0.92 (s, 9H), 0.12 (s, 3H), 0.10 (s, 3H).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3). δ : 149.9, 134.3, 129.8, 127.5, 118.5, 114.6, 98.1, 86.6, 81.2, 4.9, 25.6, 17.8, -5.8, -5.1.

LC-MS (-ve mode): $[\text{M}]^-$ 366.

1-((2R,4S,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxytetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione **102**¹⁵⁴



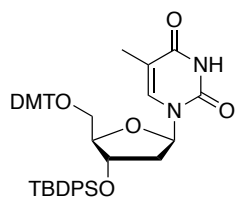
A mixture of thymidine (102 mg, 0.4 mmol), DMT-Cl (208 mg, 0.6 mmol) in pyridine (4.1 mL) was stirred under argon for 1 hour at 50 °C. The pyridine was evaporated *in vacuo* and the mixture diluted with EtOAc and washed with sat. NaHCO_3 (2 \times 20 mL).

The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with acetone/DCM (2:8 \rightarrow 6:4) afforded **102** as a white foam (197 mg, 88%).

$^1\text{H-NMR}$ (500 MHz, CDCl_3). δ : 7.56 (s, 1H), 7.25, 6.81 (m, 13H), 6.42 (t, $J = 6.78$ Hz, 1H), 4.58 (m, 1H), 4.03 (d, $J = 3.2$ Hz, 1H), 3.79 (s, 6H), 3.40 (m, 2H), 2.42 (m, 1H), 2.31 (m, 1H), 1.41 (s, 3H).

ESI-MS (+ve mode): $[\text{M}+\text{Na}]^+$ 567.

1-((2R,4S,5R)-5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-((tert-butylidiphenylsilyl)oxy)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione **103**¹⁵⁴

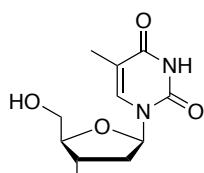


A solution of **102** (100 mg, 0.18 mmol) and imidazole (30 mg, 0.45 mmol) in DMF (0.4 mL) was cooled to 0 °C. TBDPS-Cl (0.05 mL, 0.20 mmol) was added dropwise and the mixture was stirred under argon for 16 hours at room temperature. The DMF was evaporated *in vacuo* and the residue dissolved in DCM and washed with brine (2 \times 20 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with DCM afforded **103** as a white foam (107 mg, 76%).

$^1\text{H-NMR}$ (500 MHz, CDCl_3). δ : 7.70–6.80 (m, 23H), 6.52 (t, $J = 7.0$ Hz, 1H), 4.55 (s, 1H), 4.08 (s, 1H), 3.8 (s, 6H) 3.35 (dd, $J = 2.9, 7.8$ Hz, 2H), 2.38 (m, 1H), 2.08 (m, 1H), 1.41 (s, 3H), 1.05 (s, 9H).

ESI-MS (+ve mode): $[\text{M}]^+$ 783.

1-((2R,4S,5R)-4-((tert-butylidiphenylsilyl)oxy)-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione **104**¹⁵⁴



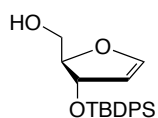
A solution of **103** (107 mg, 0.14 mmol) in DCM (7 mL) and MeOH (3 mL) was cooled

to 0 °C and treated with PTSA (77.2 mg, 0.45 mmol). The reaction was warmed to room temperature and stirred for 30 minutes under air. The solution was neutralized with solid Na₂CO₃ (29 mg, 0.28 mmol). The mixture was concentrated *in vacuo* and the residue dissolved in DCM and washed with aqueous NaHCO₃ (2 × 20 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated *in vacuo*. Column chromatography purification (SiO₂) eluting with acetone/DCM (1:9 → 5:5) afforded **104** as a white foam (51 mg, 76%).

¹H-NMR (500 MHz, CDCl₃). δ: 7.70-7.42 (m, 10H), 6.28 (t, *J* = 6.91 Hz, 1H), 4.45 (m, 1H), 4.00 (d, *J* = 2.6 Hz, 1H), 3.65 (dd, *J* = 1.9, 12.0 Hz, 1H), 3.25 (dd, *J* = 2.75, 12.0 Hz, 1H), 2.3 (m, 1H), 2.19 (m, 1H), 1.87 (s, 3H) 1.11 (s, 9H).

ESI-MS (+ve mode): [M+Na]⁺ 503.

((2R,3S)-3-((tert-butyldiphenylsilyloxy)-2,3-dihydrofuran-2-yl)methanol **72**¹⁵⁴

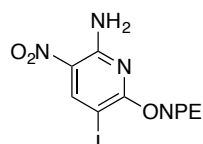


A solution of **104** (709 mg, 1.4mmol) in HMDS (7.6 mL), was heated to reflux and (NH₄)₂SO₄ (52 mg, 0.35 mmol) was added. The mixture was heated to reflux under argon for 2 hours until it turned dark yellow. The HMDS was evaporated *in vacuo* and the residue dissolved in DCM and washed with brine (50 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated. The brown gum obtained was dissolved in MeOH (17 mL) and cooled to 0 °C. K₂CO₃ (205 mg, 1.4 mmol) was added and the reaction was stirred for 30 minutes under air at room temperature. The mixture was concentrated, dissolved in DCM and washed with sat. NaHCO₃ (2 × 30 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated. Column chromatography purification (SiO₂) eluting with EtOAc/*n*-hexane (3:7) afforded the title compound as a pale yellow oil (270 mg, 53%).

¹H-NMR (500 MHz, CDCl₃). δ: 7.3-7.7 (m, 10H), 6.25 (t, *J* = 6.8 Hz, 1H), 4.4 (dt, *J* = 2.94 Hz, 6, 1H), 3.64 (d, *J* = 12.4 Hz, 1H), 3.26 (d, *J* = 12.4 Hz, 1H) 2.3 (m, 1H), 2.15(m, 1H), 1.11(s, 9H).

ESI-MS (+ve mode): [M]⁺ 354.

5-iodo-3-nitro-6-(4-nitrophenethoxy)pyridin-2-amine **87**¹⁴⁷

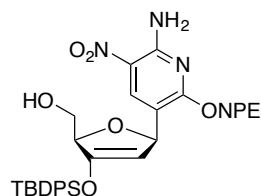


Pyridinone **77** (1 g, 3.5 mmol) was suspended in dry THF (10 mL). 2-(4-nitrophenyl)ethan-1-ol (892 mg, 5.2 mmol) and PPh₃ (1.4 g, 5.2 mmol) were added and the mixture was cooled to 0 °C, followed by the dropwise addition DEAD (0.85 mL, 5.2 mmol). The ice bath was removed, the mixture warmed to room temperature and stirred under argon for 1 hour. The mixture was concentrated, and loaded directly into a silica gel column. Elution with DCM afforded the title compound as a bright yellow powder (1 g, 73%).

¹H-NMR (500 MHz, CDCl₃) δ: 8.40 (s, 1H), 8.00 (d, *J* = 8.3 Hz, 2H), 7.60-7.70 (m, 8H), 7.30-7.50 (m, 15H), 5.60 (dd, *J* = 3.74, 1.52 Hz, 1H), 4.80 (m, 1H), 4.50 (m, 1H), 4.40 (t, *J* = 6.55 Hz, 2H), 4.00 (m, 1H), 3.90 (m, 1H), 2.90 (m, 2H).

LC-MS (-ve mode): $[M-C_8H_8NO_2]^-$ 281.

((2R,5R)-5-(6-amino-5-nitro-2-(4-nitrophenethoxy)pyridin-3-yl)-3-((tert-butyl-diphenylsilyl)oxy)-2,5-dihydrofuran-2-yl)methanol **88**¹⁴⁷

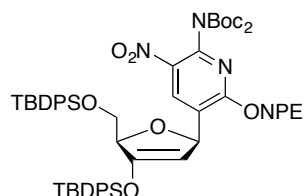


A solution of **72** (90 mg, 0.3 mmol), **87** (89 mg, 0.2 mmol), Pd(OAc)₂ (4.60 mg, 0.02 mmol) AsPh₃ (12.7 mg, 0.04 mmol) and Et₃N (0.06 mL, 0.4 mmol) in DMF (5 mL) was stirred under argon at 70 °C for 16 h. The solution was then poured into brine (10 mL) and extracted with EtOAc (3 × 10 mL). The organic phase was dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified via column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (3:7 → 6:4) afforded **88** as a yellow foam (90 mg, 66%).

¹H-NMR (500 MHz, CDCl₃) δ: 8.20 (s, 1H), 8.08 (d, *J* = 8.3 Hz, 2H), 7.70-7.80 (m, 4H), 7.40-7.50 (m, 6H), 7.30 (d, *J* = 8.3 Hz, 2H), 5.60 (dd, *J* = 3.74, 1.52 Hz, 1H), 4.70 (m, 1H), 4.40 (m, 2H), 4.20 (t, *J* = 1.65 Hz, 1H), 3.80 (m, 2H), 3 (m, 2H), 1.90 (brs, 1H), 1.00 (s, 9H).

LC-MS (-ve mode): $[M-C_8H_8NO_2]^-$ 506.

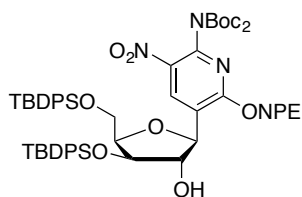
(N,N-Bis[(1,1-dimethylethoxy)carbonyl]-5-[2',5'-dihydro-5'-[[[(1,1-dimethylethyl)diphenylsilyl]oxy]methyl]-3'-[(1,1-dimethylethyl)diphenylsilyl]oxy-2'-furanyl]-3-nitro-6-[2-(4-nitrophenyl)ethoxy-2-pyridinamine **89**¹⁴⁷



A mixture of **88** (94 mg, 0.1 mmol), TBDPS-Cl (0.06 mL, 0.22 mmol) and imidazole (28 mg, 0.4 mmol) in DMF (1.4 mL) was stirred under argon at r.t. for 16 h. The solution was poured into brine (30 mL) and extracted with EtOAc (3 × 20 mL). The organic phase was dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified via column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (3:7) affording di-TBDPS ether as a yellow solid. This material was dissolved in DCM (9 mL). To the solution were added Boc₂O (0.2 mL, 0.4 mmol), Et₃N (0.18 mL, 0.67 mmol) and DMAP (2 mg, 0.013 mmol) and the reaction was stirred under argon at room temperature for 16 h. The solution was then washed with brine (30 mL). The organic phase was dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified via column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (3:7) to afford **89** as a yellowish foam (94 mg, 64% over two steps).

¹H-NMR (500 MHz, CDCl₃) δ: 8.40 (s, 1H), 8.0 (d, *J* = 8.3 Hz, 2H), 7.60-7.70 (m, 8H), 7.30-7.50 (m, 15H), 5.60 (dd, *J* = 3.74, 1.52 Hz, 1H), 4.80 (m, 1H), 4.50 (m, 1H), 4.40 (t, *J* = 6.55 Hz, 2H), 4.0 (m, 1H), 3.90 (m, 1H), 2.9 (m, 2H), 1.40 (s, 18H), 1.05 (s, 9H), 0.90 (s, 9H).

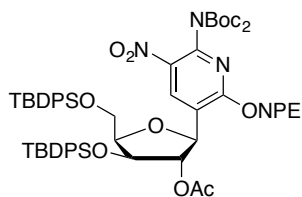
LC-MS (-ve mode): $[M-C_8H_8NO_2]^-$ 946.

(N,N-Bis[(1,1-dimethylethoxy)carbonyl]-5-[(3',5'-di-O-di-ter-butyl-diphenylsilyl)-β-D-xylofuranosyl]-3-nitro-6-[2-(4-nitrophenyl)ethoxy]-2-pyridinamine 90¹⁴⁷

A solution of **89** (130 mg, 0.12 mmol) in THF (2 mL) was cooled to 0 °C and treated with $\text{BH}_3 \cdot \text{Me}_2\text{S}$ (0.6 mL, 1.0M solution in THF, 0.6 mmol). The mixture was slowly warmed to room temperature over 2 hours and then stirred under argon for 16 h. The solution was quenched with water (0.5 mL) and stirred for 5 min. A suspension of NaBO_3 (285 mg, 2.9mmol) in water (1 mL) was added and the white suspension stirred under air at room temperature for 2 hours. It was then poured into brine (20 mL) and extracted with EtOAc (3×15 mL). The organic phase was dried over Na_2SO_4 and the solvents were evaporated *in vacuo*. The residue was purified via column chromatography (SiO_2) eluting with EtOAc/*n*-hexane (2:8 → 5:5) to afford **90** as a yellow solid (30 mg, 22%).

$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 8.70 (s, 1H), 8.10 (d, $J = 8.5$ Hz, 2H), 7.69 (d, $J = 6.5$ Hz, 2H), 7.73 (d, $J = 6.5$ Hz, 2H) 7.20-7.50 (m, 18H), 4.69 (brs, 1H), 4.60 (t, $J = 6.47$ Hz, 2H), 4.26 (m, 1H), 4.20 (d, $J = 3$ Hz, 1H), 3.66 (brs, 1H), 3.1 (q, $J = 6.28$ Hz, $J = 4$ Hz, 2H) 1.44 (s, 18H), 1.10 (s, 9H), 0.80 (s, 9H).

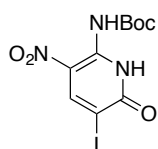
LC-MS (-ve mode): $[\text{M}-\text{C}_8\text{H}_8\text{NO}_2]^-$ 963.

(N,N-Bis[(1,1-dimethylethoxy)carbonyl]-5-[(3',5'-di-O-di-ter-butyl-diphenylsilyl)-2'-O-acetyl-β-D-xylofuranosyl]-3-nitro-6-[2-(4-nitrophenyl)ethoxy]-2-pyridinamine 105¹⁴⁷

A mixture of alcohol **90** (30 mg, 0.027 mmol), Ac_2O (28 μL , 0.03 mmol) and pyridine (9.6 μL , 0.12 mmol) in dry DCM (1.3 mL) was stirred under argon for 2 hours at room temperature. The mixture was poured into brine (20 mL). The organic phase was dried over Na_2SO_4 and the solvents were evaporated *in vacuo*. The residue was purified via column chromatography (SiO_2) eluting with EtOAc/*n*-hexane (3:7) to afford **105** as a yellowish foam (29mg, 96%).

$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 8.60 (s, 1H), 8.10 (d, $J = 8.5$ Hz, 2H), 7.69 (d, $J = 6.5$ Hz, 2H), 7.73 (d, $J = 6.5$ Hz, 2H) 7.20-7.50 (m, 18H), 5.0 (s, 1H), 4.80 (s, 1H) 4.50 (m, 2H), 4.14 (m, 2H), 3.80 (d, $J = 10.8$ Hz, $J = 3.8$ Hz, 1H) 3.07 (d, $J = 6.8$ Hz, 2H), 1.90 (s, 3H), 1.44 (s, 18H), 1.10 (s, 9H), 0.80 (s, 9H).

LC-MS (-ve mode): $[\text{M}-\text{C}_8\text{H}_8\text{NO}_2]^-$ 1004.

tert-butyl (5-iodo-3-nitro-6-oxo-1,6-dihydropyridin-2-yl)carbamate 106¹⁴⁷

Pyridone **77** (150 mg, 0.53 mmol) was dissolved in DMF (2.2 mL). To the solution were added Boc_2O (0.35 mL, 1.6 mmol), Et_3N (0.4 mL, 2.8 mmol) and DMAP (5.8 mg, 0.05

mmol) and the reaction was stirred under argon at room temperature for 16h. The solution was then washed with brine (40 mL). The organic phase was dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified via column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (6:4) to afford **106** as a yellow solid (130 mg, 64%).

¹H-NMR (500MHz, CDCl₃) δ: 12.07 (br s, 1H), 10.80 (s, 1H) 8.70 (s, 1H), 1.60 (s, 9H).

¹³C-NMR (125 MHz, CDCl₃). δ: 157.3, 152.1, 146.0, 144.0, 116.7, 87.0, 80.0, 27.9.

HRMS (micrOTOF-Q) m/z: [M+Na]⁺ calcd for C₁₀H₁₂IN₃NaO₅ 403.9714, found 403.9707.

Chapter 3

3 Establishing a Modular Synthetic Methodology for the Preparation of C6-heteroaryl 2-aminopurines

3.1 Introduction

Whereas base pairs developed by the Benner group such as the **Z/P** (see Figure 2-1) rely on pairing through hydrogen bonding, the Hirao nucleotides pair through hydrophobic interactions and shape complementarity. Among these, the current state-of-the-art are the **s/Pa**, **Ds/Pa** and **Ds/Px** (Figure 3-1a).^{119,121} The **Ds/Px** exhibits high PCR fidelity (>99%) and efficiency (10^7 -fold amplification after 30 cycles), whereas the **Ds/Pa** is better in transcription.¹²² Furthermore, alkyne-modified versions **111**, **112**, **113** have been synthesized and used for site-specific RNA labeling at a single site and DNA imaging tool (Figure 3-1b).^{132,156,157}

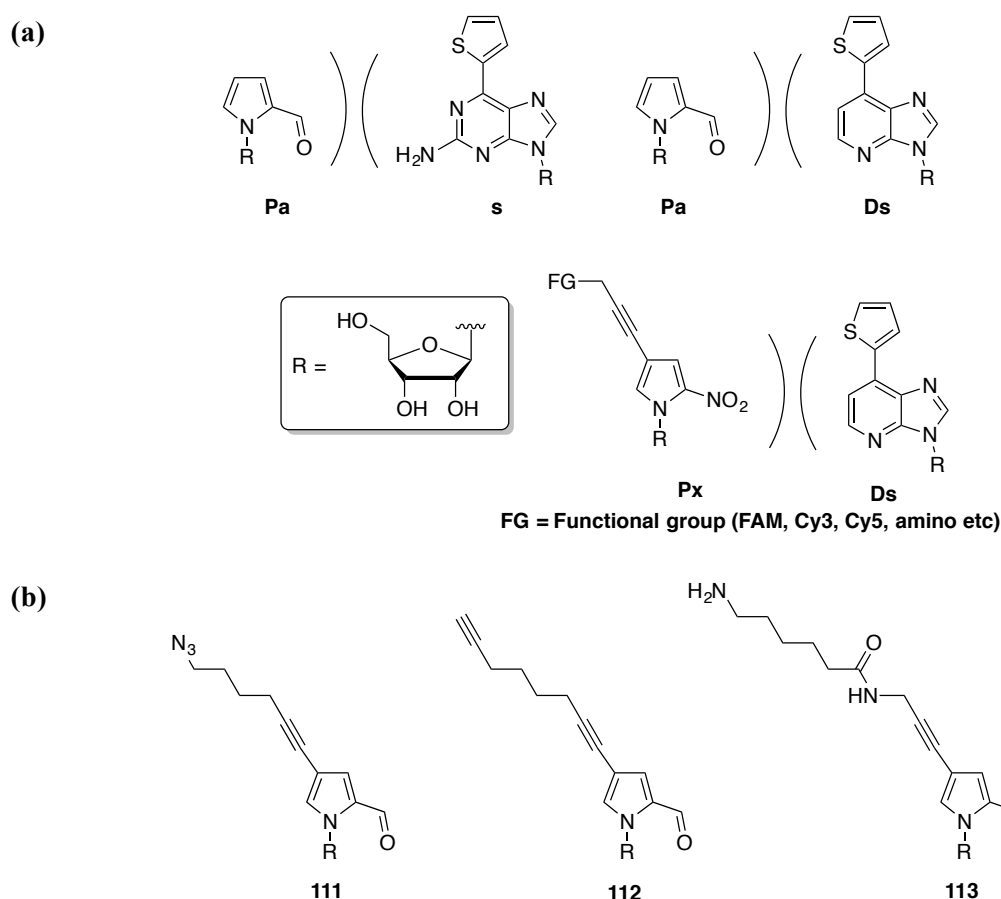


Figure 3-1. (a) The **Pa/s**, **Pa/Ds** and **Pn/Ds** base pairs. (b) Functionalized analogues of the **Pa** and **Pn** nucleotides synthesized by the Hirao lab.^{132,156,157}

At present, little is known regarding the structural determinants that define the transcriptional efficiency and fidelity of C-6 functionalized analogues of **s**. Early attempts of Hirao include the installation of a thiazole and a dimethylamine group (Figure 3-2).^{115,117} These nucleotides were developed before the synthesis of the **Pa** was reported and currently there is no literature precedent on investigations of their incorporation efficiency and fidelity with **Pa** as the cognate nucleotide.

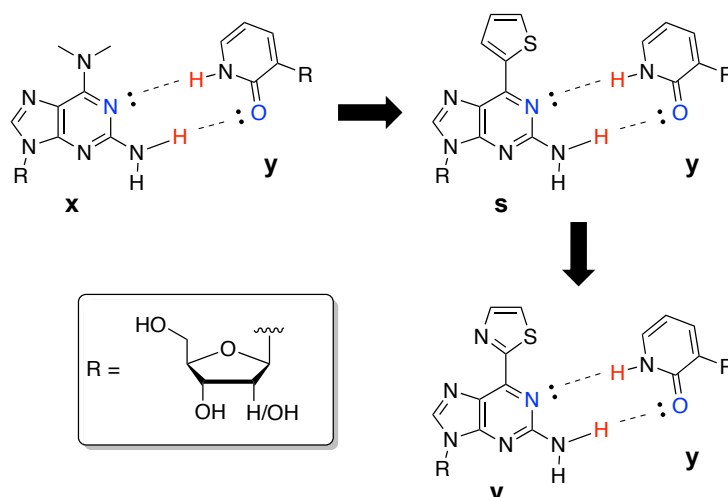


Figure 3-2. SAR investigations on the **x/y** pair conducted by the Hirao group.

In addition, a functionalized version of **s** or **Ds** suitable for fluorophore labeling has not been reported yet. Therefore, a structure-activity relationship (SAR) study investigating the factors that control the transcriptional efficiency/fidelity of **s** analogues could lead to the development of a doubly modified pair suitable for FRET studies. The **s/Pa** was selected as the basis of this study. Unlike the **Ds** that necessitates the *de novo* synthesis of the heterocyclic moiety and subsequent coupling with a suitable sugar donor, **s** is based on the guanosine (2-aminopurine) scaffold.¹¹⁹ Consequently, a robust synthesis could be developed with guanosine or a guanosine derivative as the starting material followed by direct functionalization of the desired position. In order to study the factors that govern the transcriptional efficiency/fidelity of the **s** family of nucleotides, it was proposed that 5-membered heterocycles could substitute the thiophene of **s** on the C6 position such as triazoles and isoxazoles (Figure 3-3). Firstly, this approach would enable the transcriptional evaluation of various side-groups since both classes of compounds can be obtained from [3+2] cycloaddition reactions of a common alkyne precursor with various azides (triazoles) or oximes (isoxazoles).

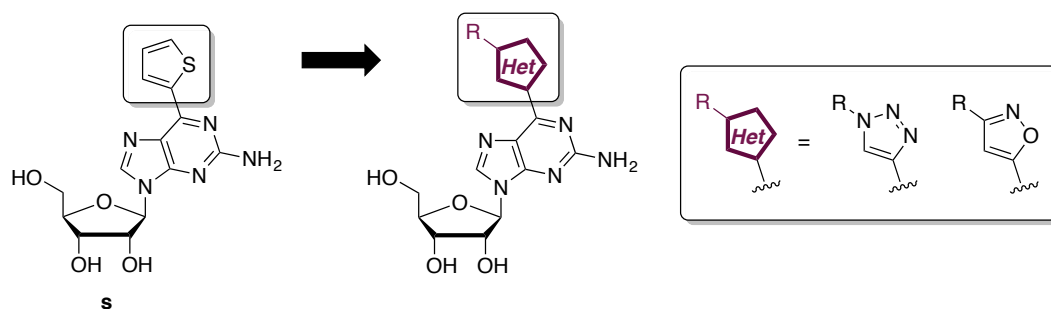


Figure 3-3. Substitution of the thiophene in **s** with triazoles or isoxazoles will provide analogues suitable for SAR studies regarding their transcriptional efficiency/fidelity. R indicates various side-groups.

In addition, despite the syntheses of an impressive number of triazole-functionalized nucleosides and nucleotides, C6-triazole purine analogues have not been widely explored.^{158,159} Likewise, there is a current dearth in the literature for C6-isoxazole purines. These reasons prompted the development of a modular synthetic platform in order to explore the utility of C6-triazole and C6-isoxazole functionalized 2-aminopurines (guanosines) as potential candidates for an expanded genetic alphabet.

3.2 Aims of this Study

The aims of this chapter are to:

- 1) design a robust route for accessing a library of C6 functionalized analogues of **s**.

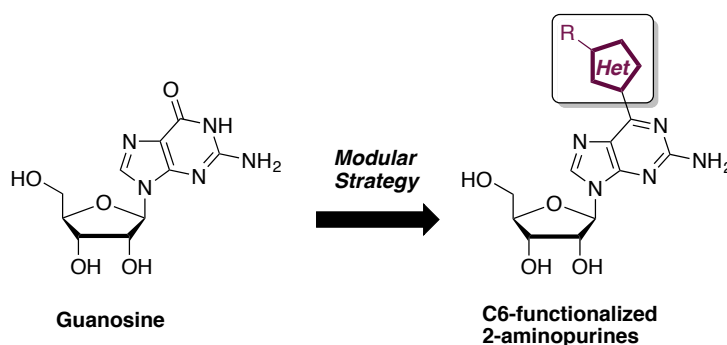


Figure 3-4. A modular synthetic strategy will enable the synthesis of a suite of **s** analogues.

- 2) synthesize a library of analogues of **s** to be evaluated for their transcriptional efficiency/fidelity.

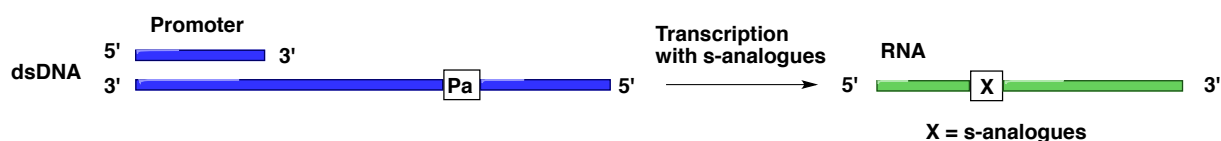


Figure 3-5. Schematic representation of the proposed assay to be employed for the transcriptional evaluation of **s**-analogues.

3.3 Results and Discussion

3.3.1 Retrosynthetic Analysis of C6-heteroaryl 2-aminopurines

The proposed retrosynthetic analysis is presented in Figure 3-4. Access to the desired triazole and isoxazole analogues was explored *via* [3+2] cycloaddition from the corresponding azides or oximes with alkyne **114**. Alkyne **114** would be synthesized *via* a Sonogashira coupling of **115** with commercially available (trimethyl)ethynylsilane followed by removal of the silyl group. This approach would also enable the installation of a pyrazole moiety through a Suzuki-Miyaura cross-coupling from **115** and the corresponding boronic acid.

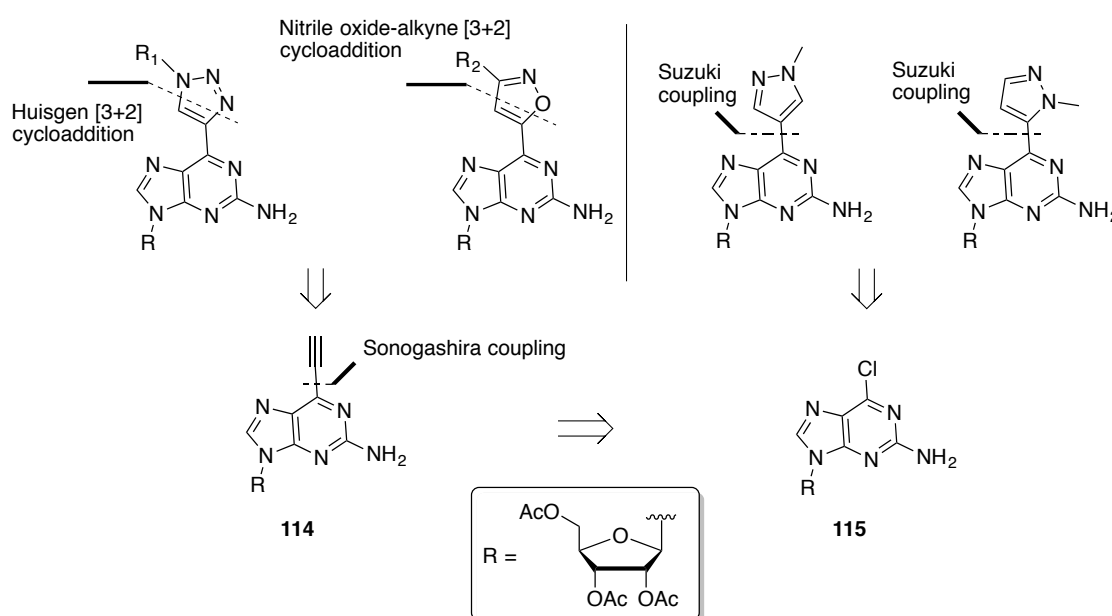
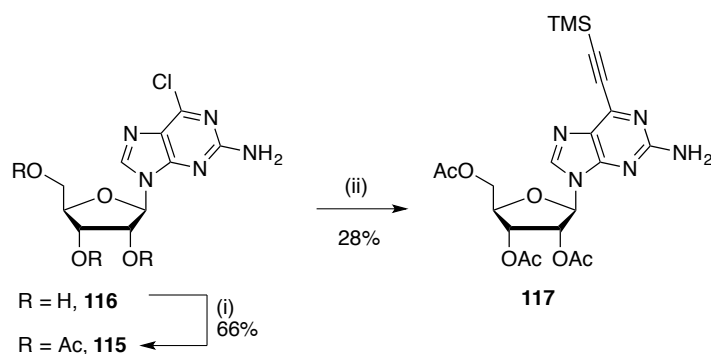


Figure 3-4. Retrosynthetic analysis of target compounds.

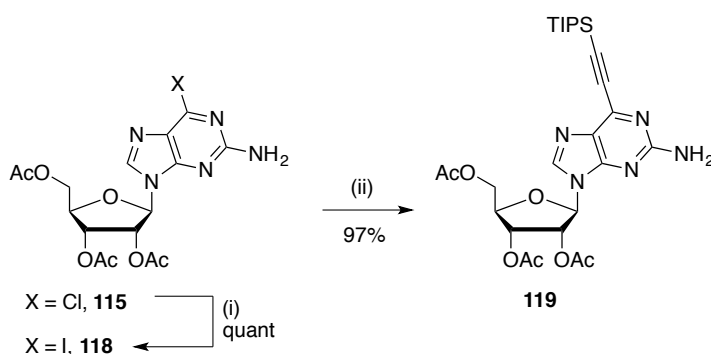
3.3.2 Synthesis of C6-ethynyl 2-aminopurine

To increase the solubility of 6-chloroguanosine riboside in organic solvents the sugar hydroxyl groups were protected as acetates, affording **115** in 66 % yield.¹⁶⁰ A Sonogashira coupling between **115** and ethynyltrimethylsilane was initially explored (Scheme 3-1). Since Sonogashira reactions are known to be susceptible to the presence of oxygen, the reaction mixture was purged with argon for 15 minutes.¹⁶¹



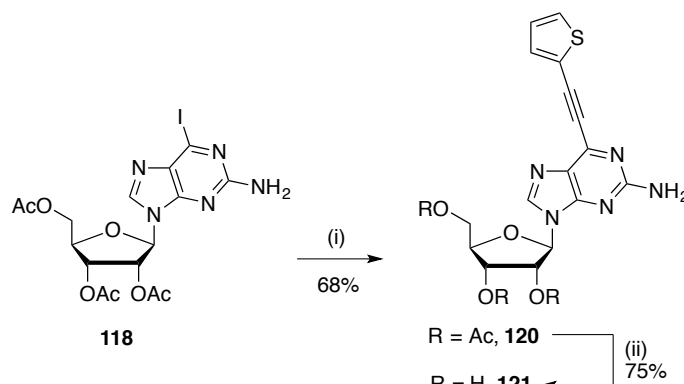
Scheme 3-1. Reagents and conditions: (i) Ac₂O (3.0 equiv.), Et₃N (6.0 equiv.), DMAP (0.1 equiv.), MeCN, 0 °C, 3 h. (ii) [Pd(PPh₃)₂]Cl₂ (2 mol%), CuI (6 mol %), DIPEA (3.0 equiv.), ethynyltrimethylsilane (1.2 equiv.), DMF, 60 °C, 16 h.

After overnight stirring at 60 °C, TLC analysis revealed formation of a new compound, along with starting material. Alkyne **117** was obtained in 28% yield and chloride **115** was recovered in 70% yield. Further optimization attempts included increasing the temperature to 100 °C and/or using the highly active Pd(OAc)₂/XPhos system, but these attempts resulted in either no conversion or 15% conversion to **107** according to LC-MS. Aryl iodides are known to be more reactive in Sonogashira couplings than aryl chlorides.^{161,162} Therefore, it was decided to synthesize the corresponding iodide and investigate its reactivity in the Sonogashira reaction. Employing Finklestein conditions (NaI, TFA) on chloride **115** afforded iodide **118** in quantitative isolated yield (Scheme 3-2).¹⁶³ Optimization attempts lead to the discovery that instead of carrying out the reaction at -40 °C and gradually warming it up to room temperature, the reaction can be conducted directly at room temperature. In addition aqueous sodium bicarbonate and sodium thiosulfate work-up can replace chromatographic purification, which allowed the reaction to be amenable to multi-gram scale. Iodide **118** was then subjected to the Sonogashira conditions (Scheme 3-2). Due to the low yield encountered in the previous Sonogashira attempt it was decided to pursue more stringent deoxygenated conditions by degassing the reaction mixture by employing the freeze-pump-thaw cycle (the cycle was repeated three times). (Triisopropylsilyl)acetylene was also used instead of the TMS-derivative to make the product less polar and assist with chromatographic purification. After overnight stirring at room temperature **119** was obtained in 97% yield.



Scheme 3-2. Reagents and conditions: (i) NaI (20.0 equiv.), TFA (2.0 equiv.), 2-butanone, r.t., 16 h, quant. (ii) [Pd(PPh₃)₂]Cl₂ (2 mol%), CuI (6 mol%), DIPEA (3.0 equiv.), (triisopropylsilyl)acetylene (1.3 equiv.), DMF, r.t., 16 h.

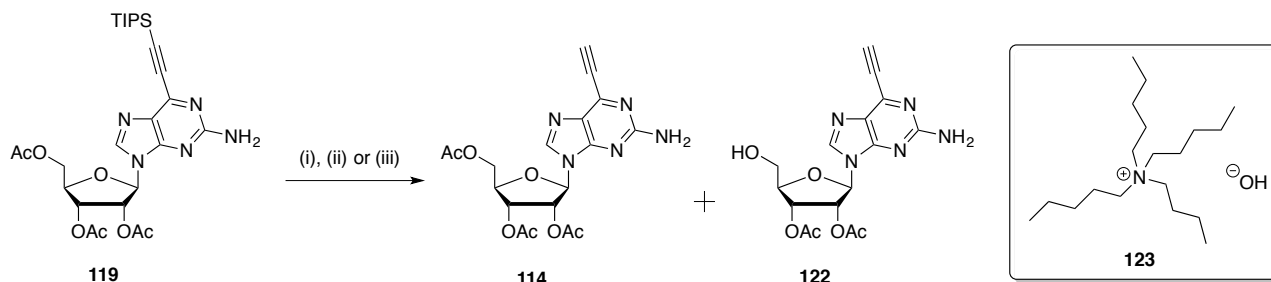
In addition, the installation of a 2-ethynylthiophene unit was explored to investigate the effect of having a linear ethynyl linker between the thiophene and the guanosine moieties. Utilizing the same Sonogashira conditions described for the synthesis of **119**, reaction of **118** with 2-ethynylthiophene afforded **120** in 68% yield (Scheme 3-3). Acetyl deprotection was achieved through exposure to K_2CO_3 in MeOH, affording free nucleoside **121** in 75% yield.



Scheme 3-3. Reagents and conditions: $[Pd(PPh_3)_2]Cl_2$ (2 mol%), CuI (6 mol%), DIPEA (3.0 equiv.), 2-ethynylthiophene (1.3 equiv.), r.t., 16 h. (ii) K_2CO_3 (0.3 equiv.), MeOH, r.t., 16 h.

With alkyne **119** in hand, conditions for the deprotection of the TIPS group were investigated (Table 3-1). Tetra-*n*-butylammonium fluoride (TBAF) in THF are commonly employed in deprotection of silyl ethers. Non-anhydrous conditions were firstly investigated (solid TBAF, normal-grade THF, non flame-dried glassware). Two products were observed on TLC after 30 min. The less polar was isolated in 49% yield and identified as desired alkyne **114** (Entry 1). The more polar spot was isolated as a mixture with **114**. ESI-MS analysis revealed the presence of side-product **122**. Presumably under non-anhydrous conditions the basic fluoride anions produce hydroxide **123**, which could cleave the less hindered 5' acetate. Indeed, running the reaction at $-10\text{ }^\circ\text{C}$ under dry conditions (anhydrous THF, flame-dried glassware) alkyne **114** was obtained in 91% yield (Entry 2). Unfortunately, this set of conditions proved to be not reproducible and even when the reaction was cooled to $-78\text{ }^\circ\text{C}$ and gradually warmed to $-20\text{ }^\circ\text{C}$, formation of **122** was observed on TLC. A literature survey revealed that AgF has been used for TIPS deprotection of alkynes bearing acetate esters, which remained intact.¹⁶⁴ Pleasingly, by employing this protocol, **114** was obtained in 60% without the formation of detectable amounts of **122** (Entry 3). This alkyne was then used to carry out the synthesis of the isoxazole and triazole sets of analogues, as presented in the next chapters.

Table 3-1. TIPS deprotection conditions: (i) TBAF (1.0 equiv.), non-anhydrous THF, r.t., 30 min. (ii) TBAF (1.0 equiv.), anhydrous THF, $-10\text{ }^{\circ}\text{C}$, 30 min. (iii) AgF (1.5 equiv.), MeCN, r.t., 16 h., then 1.0 M HCl (2.0 equiv.) r.t., 5 min.



Entry	Conditions	Yield of 114
1	(i)	49%
2	(ii)	91% (not reproducible)
3	(iii)	60%

3.3.3 Synthesis of C6-triazole Functionalized 2-aminopurines

Triazole-functionalized nucleosides and nucleotides have been widely explored for therapeutic or biotechnological applications.^{159,165–167} In addition, one of the two literature examples of C6-triazole adenosines has reported their ability to modulate cAMP production in mammalian cells.¹⁵⁹ Due to their biological potential and the current lack of triazole-functionalized 2-aminopurine examples, we sought to explore the synthesis of this type of compounds. Furthermore, triazoles are easy to access through copper-catalyzed azide-alkyne [3+2] cycloaddition (CuAAC). This straightforward synthesis is noteworthy, since reaction of various azides with alkyne **114** as the common precursor (Table 3-1) would furnish multiple triazole analogues. Three different triazoles were chosen for this study (Figure 3-5). Benzyl functionalized triazole **124** was chosen to investigate the tolerance of the RNA polymerase against bulky aromatic groups. Triazole **126** features two extra oxygen atoms that could potentially act as hydrogen-bond acceptors with the surrounding environment. Comparing the transcriptional efficiency and fidelity of **126** with the more hydrophobic triazole **125** could provide insight on such effect.

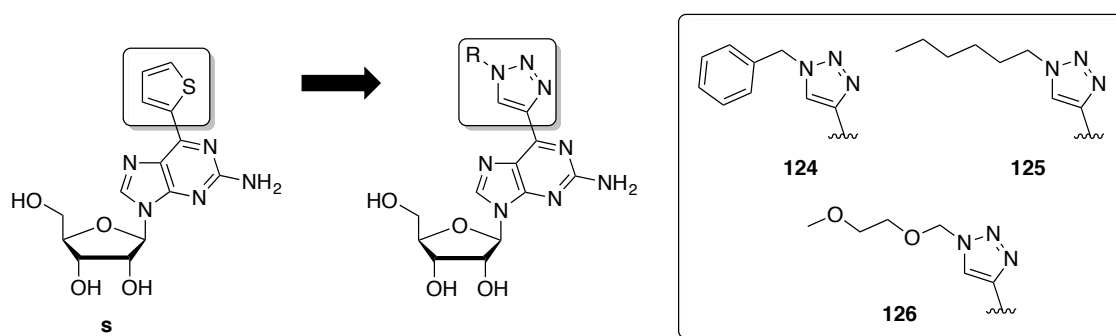
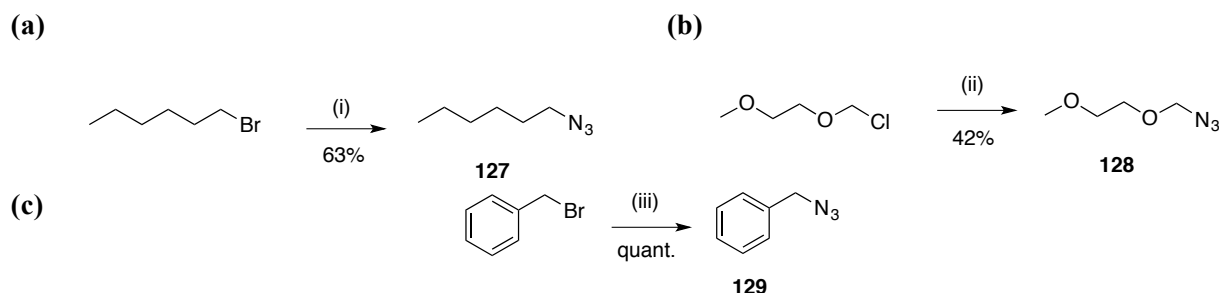


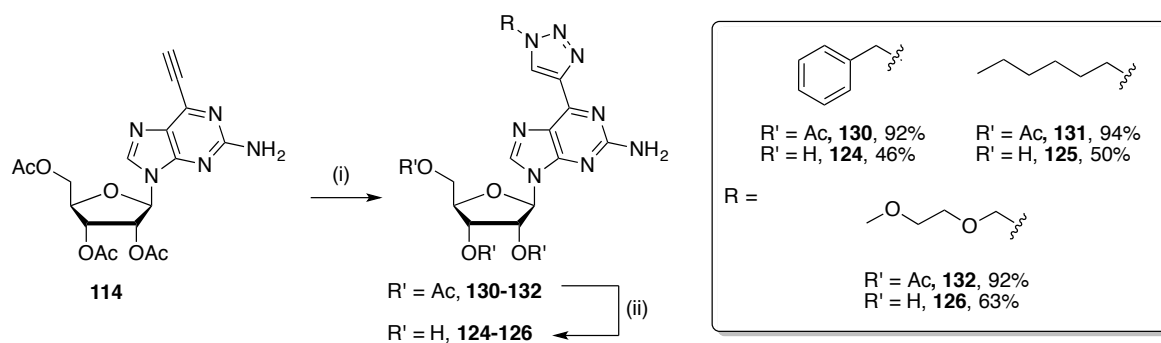
Figure 3-5. Proposed C6-triazole functionalized analogues **124-126**.

The syntheses of the corresponding azides were accomplished following literature procedures by treating the corresponding halide with sodium azide, affording azides **127-129** in 63% to quantitative yields (Scheme 3-4).¹⁶⁸⁻¹⁷⁰



Scheme 3-4. Reagents and conditions. (a) (i) NaN_3 (1.8 equiv.), DMSO, r. t., 16 h. (b) (ii) NaN_3 (2.0 equiv.) MeCN, reflux, 3 h. (c) (iii) NaN_3 , (1.5 equiv.), $\text{H}_2\text{O}/\text{acetone}$ (1:4 v/v), r.t., 16 h.

Reaction of alkyne **114** with each azide under conventional “click” conditions (CuSO_4 , TBTA, sodium ascorbate) furnished triazoles **130-132** in 90-94% yields (Scheme 3-5). Hexyltriazole **131** was isolated with residual TBTA used in the reaction, which was removed in the next step. Acetate removal was achieved *via* the same protocol for **120** affording free nucleosides **124-126** in 46-63% yield.



Scheme 3-5. Reagents and conditions: (i) Azide (1.3 equiv.), CuSO_4 (0.4 M in H_2O , 5 mol%), TBTA (0.1 equiv.), sodium ascorbate (1.0 equiv.), $\text{DMSO}/\text{H}_2\text{O}$ (3:1), r.t., 16 h. (ii) K_2CO_3 (0.3 equiv.), MeOH, r.t., 16 h.

The synthesis of triazole **133** was also initiated. This aryl-substituted triazole is structurally similar to Hirao’s bis-thiophene **Dss** nucleoside, a highly fluorescent derivative of the **Ds**. (Figure 3-6).¹⁷¹ Consequently, assaying **133** in transcription and comparing it with the **s** and the **Dss** could provide insight on the influence of each 5-membered heterocycle on efficiency and fidelity.

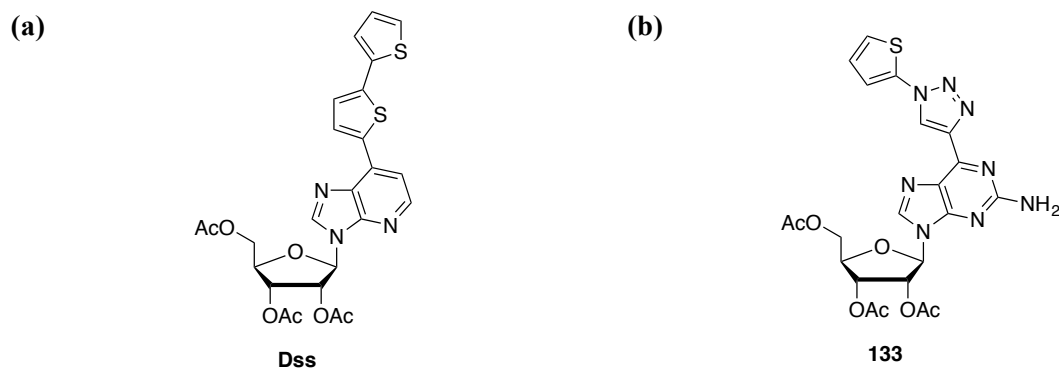
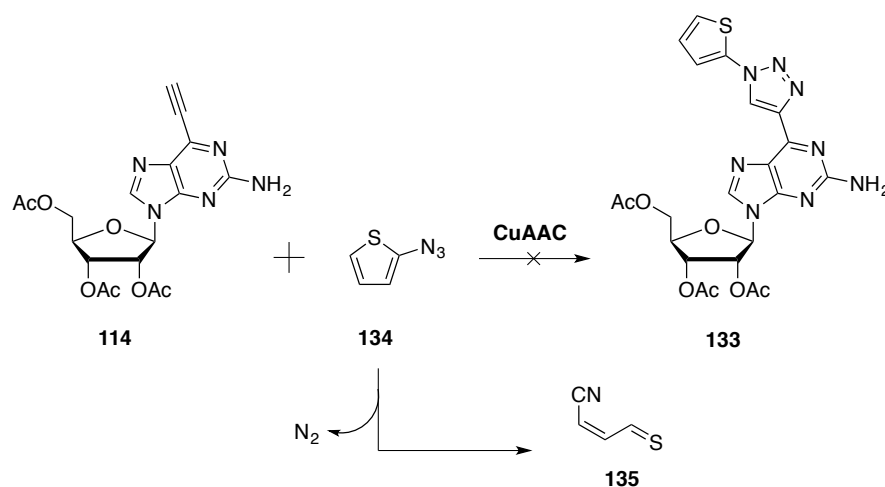


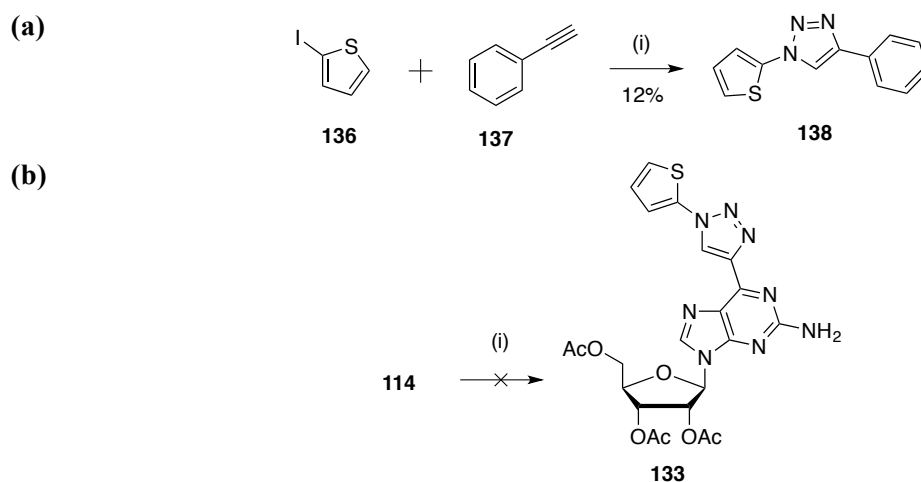
Figure 3-6. (a) Hirao's **Dss** nucleoside.¹⁶⁵ (b) Triazole **133**.

A straightforward method would be to react azidothiophene **134** with alkyne **114** under CuAAC conditions. However, **134** has been reported to be thermally unstable and is purported to decompose to nitrile **135** (Scheme 3-6).¹⁷²



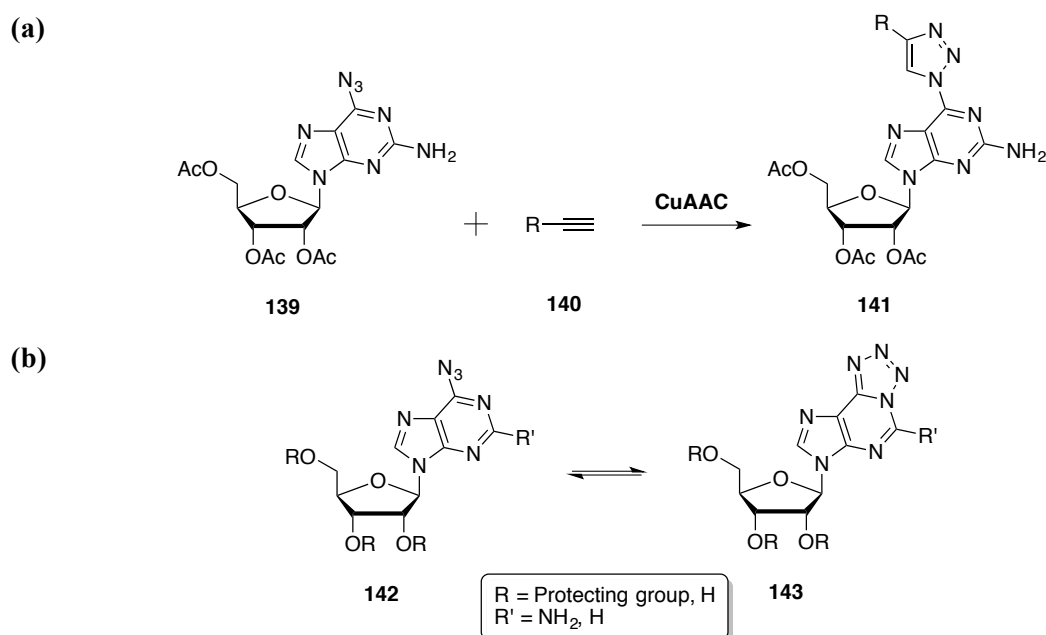
Scheme 3-6. Purported decomposition pathway of 2-azidothiophene.¹⁷²

Bäuerle et al. recently reported that iodide **136** can participate in a Cu(I)-mediated one-pot azidation-[3+2] cycloaddition reaction with alkyne **137**.¹⁷³ To investigate the feasibility of this procedure for our purposes, we attempted to reproduce this reaction. The reaction proceeded as expected and triazole **138** was obtained, albeit in low yield (12%). Unfortunately, employing these reaction conditions with 2-iodothiophene and alkyne **114** resulted in a complex mixture, as observed by TLC and confirmed by ¹H-NMR analysis following workup (Scheme 3-7b). The use of a base and protic solvent in combination with prolonged heating might have resulted in partial and/or full acetate cleavage of **114** and/or **133**, thus explaining the complex mixture observed. The use of a non-protic solvent (e.g. DMSO) could potentially circumvent this problem; however, the investigation of this hypothesis was not initiated in favor of expanding the diversity of the C6-functionalized analogues.



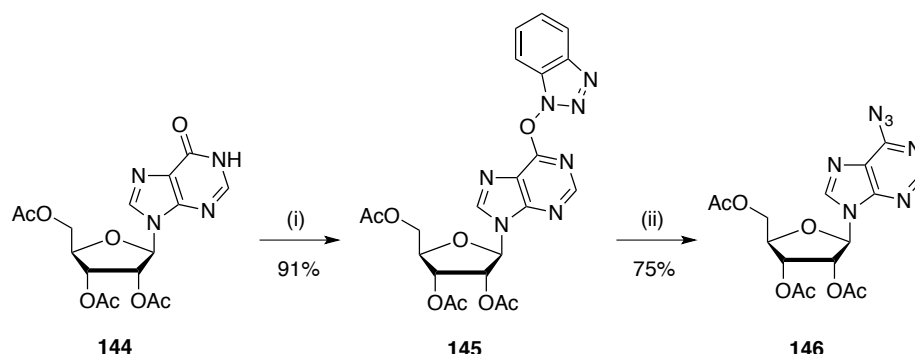
Scheme 3-7. (a) *Reagents and conditions:* (i) **137** (1.0 equiv.), CuI (10 mol%), NaAsc (10 mol%), DMEDA (0.2 equiv.), NaN₃ (2.0 equiv.), EtOH/H₂O (7:3), 50 °C, 16 h. (b) *Reagents and conditions:* (i) 2-Iodothiophene (1.0 equiv.), CuI (10 mol%), NaAsc (10 mol%), DMEDA (0.2 equiv.), NaN₃ (2.0 equiv.), EtOH/H₂O (7:3), 50 °C, 16 h.

Since a different heteroatom arrangement on the triazole scaffold could facilitate different interactions during transcription between the incoming nucleotide, the neighboring nucleotides and the polymerase, the synthesis of azide **139** was pursued (Scheme 3-8a). It was expected that azide **139** could undergo CuAAC with alkyne **140** furnishing triazole **141**, thus yielding another set of triazoles, with a different heteroatom arrangement compared to triazoles **130-132** (see Scheme 3-5). C6-azido nucleosides exist in equilibrium with their tetrazole tautomer **143** thus giving rise to an extra set of peaks in ¹H-NMR (Scheme 3-8b).^{174,175} This is particularly noteworthy for this reaction because this equilibrium is solvent dependent, and the tetrazole tautomers are unreactive towards CuAAC.^{174,175}



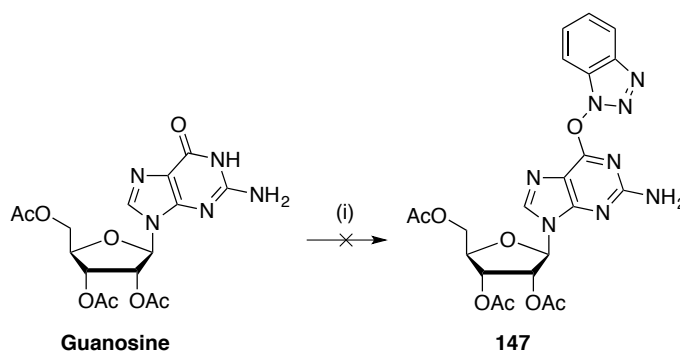
Scheme 3-8. (a) CuAAC of azide **139** with alkyne **140** can furnish triazole **141**. (b) Equilibrium of azide **142** with tetrazole **143**.^{174,175}

Recently, the Lakshman group reported that azide **146** can be obtained by synthesizing benzotriazolyl inosine **145** and reacting it with NaN_3 in DMSO (Scheme 3-9).¹⁷⁵



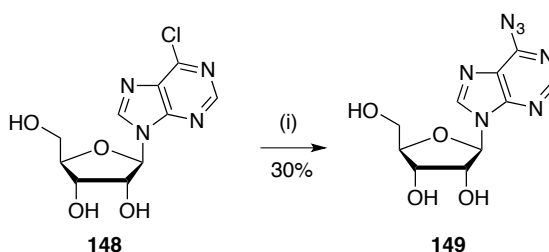
Scheme 3-9. Reagents and conditions: (i) BOP (2.0 equiv.), DIPEA (2.0 equiv.), DCM, r.t., 16 h. (ii) NaN_3 (3.0 equiv), DMSO, r.t., 16 h.¹⁷⁵

Unfortunately, an attempt to synthesize the corresponding benzotriazolyl guanosine **147** employing the same protocol resulted in a mixture of starting material and product as observed by LC-MS (Scheme 3-10). Due to the need to optimize the reaction conditions and to investigate the reactivity of **147** in the displacement step with an azide source, this strategy was discontinued in order to explore the direct installation of the azide group from the corresponding halide.



Scheme 3-10. Reagents and conditions: (i) BOP (2.0 equiv.), DIPEA (2.0 equiv.), DCM, r.t., 16 h.

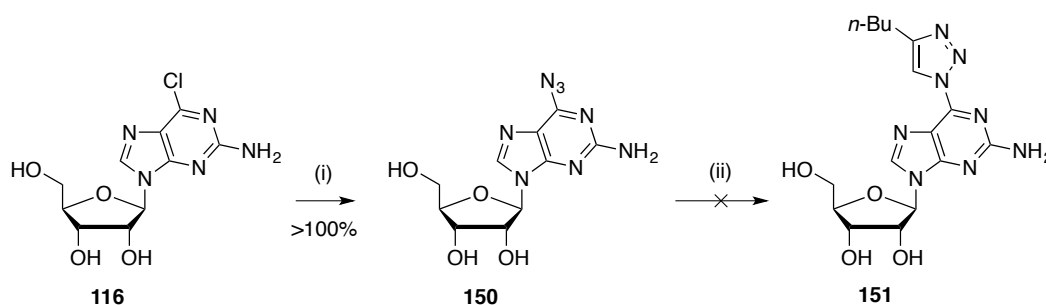
Previous studies have shown that NaN_3 in DMF can achieve the desired transformation, a more recent study by Guieu et al. reported that azido adenosine **149** could be obtained only by reacting unprotected **148** with LiN_3 in DMF (Scheme 3-11).^{159,174}



Scheme 3-11. Reagents and conditions: (i) LiN_3 (11.0 equiv.), DMF, r.t., 60 h.¹⁵⁹

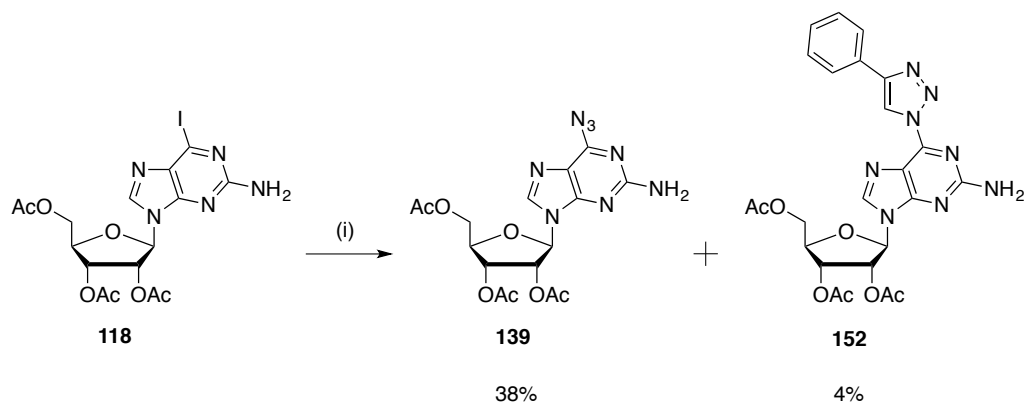
Thus, it was decided to employ this set of conditions with chloride **116**. After 2 days, LC-MS analysis revealed the presence of a single peak with the desired mass of azide **150**. After chromatographic purification, a compound was isolated in >100% yield, likely resulting from contamination of silica, LiN_3 , or DMF (Scheme 3-12). IR analysis revealed the presence of the product's azide peak at 2080 cm^{-1} . $^1\text{H-NMR}$ analysis revealed the presence of the H8 of the base (8.37 ppm) and several sugar protons (e.g. the H-1' as a doublet at 5.77 ppm). Further analysis of the spectrum was hindered by contamination from various impurities, DMF and the presence of an extra set of peaks, presumably due to the azide-tetrazole equilibrium (see Scheme 3-8b). A repeated normal-phase chromatographic purification was deemed risky due to the polarity of the compound, which could result in significant loss of material. Likewise, HPLC purification was concluded not to be feasible as the retention time of the compound in LC-MS was located at the solvent front.

To probe the presence of the azide moiety, a CuAAC reaction of **150** with 1-hexyne was performed (Scheme 3-12). Unfortunately, after 16 hours only starting material was observed on LC-MS. This observation suggests that either the desired product **150** has not formed, or it has, but the major tautomer is the unreactive tetrazole. Lakshman and co-workers have also studied the influence of solvent in CuAAC with C6-azido nucleosides. They suggested that the best solvent for CuAAC with this class of compounds is a mixture of DCM/ H_2O (which shifts the equilibrium to the azide tautomer).¹⁷⁵ However, we were concerned that azide **150** would not be soluble in that mixture. This strategy was deemed to be time-inefficient due to the projected amount of optimization needed and it was discontinued to pursue other strategies for the synthesis of C6-azido 2-aminopurine.



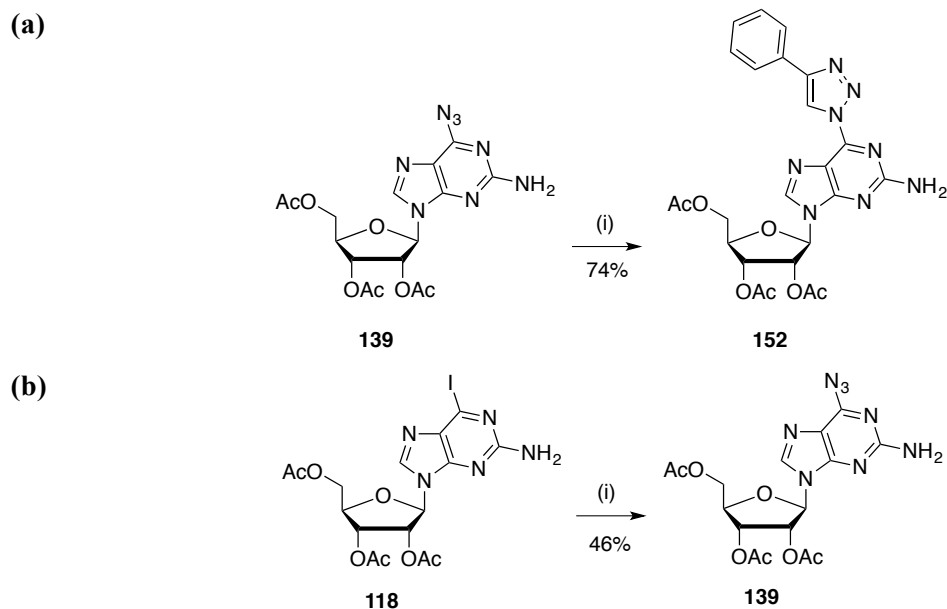
Scheme 3-12. Reagents and conditions: (i) LiN_3 (11.0 equiv.), DMF, r.t., 60 h. (ii) CuSO_4 (5 mol%), TBTA (0.1 equiv.), NaAsc (1.0 equiv.), 1-hexyne (1.3 equiv.), DMSO/ H_2O , r.t., 16 h.

The Cu(I)-mediated one-pot azidation-[3+2] cycloaddition reaction developed by Bäuerle et al. was investigated next (see Scheme 3-6).¹⁷³ Iodide **118** and phenylacetylene were employed as test substrates. The solvent was changed to non-protic DMSO due to concerns that the basic environment combined with heating would result in acetate cleavage (which might explain the complex mixture formed during the attempted synthesis of **133**). Two products were observed on TLC and isolated. The minor product was isolated in 4% yield and identified as triazole **152**, whereas the major product was isolated in 38% yield and was identified as intermediate azide **139** (Scheme 3-13).



Scheme 3-13. Reagents and conditions: (i) Phenylacetylene (1.0 equiv.), CuI (10 mol%), NaAsc (10 mol%), DMEDA (0.2 equiv.), NaN₃ (2.0 equiv.), DMSO/H₂O (2:1), 50 °C, 16 h.

Reaction of **139** with phenylacetylene under previously developed conditions (see Scheme 3-5) afforded triazole **152** 74% yield (Scheme 3-14a). Omitting the alkyne from the reaction mixture, the reaction afforded **139** in 46% yield (Scheme 3-14b). Unfortunately, this method proved to be not reproducible, as mixtures of starting material and product or complicated mixtures were observed upon TLC analysis of subsequent batches. Since amines are known to be susceptible to air oxidation, this poor reproducibility might be a result of the aged DMEDA batch employed. Repeating the reaction using fresh DMEDA would test the accuracy of this hypothesis. However, due to time constraints, the problems encountered with generating the C6-azido guanosine regardless of the method employed and its tendency to undergo solvent-dependent tautomerization, this strategy was paused in order to focus on the synthesis of C6-isoxazole 2-aminopurines.



Scheme 3-14. (a) *Reagents and conditions:* (i) CuSO_4 (5 mol%), TBTA (0.1 equiv.), NaAsc (1.0 equiv.), phenylacetylene (1.3 equiv.), DMSO/ H_2O (2:1), r.t., 16 h. (b) *Reagents and conditions:* (i) CuI (10 mol%), NaAsc (10 mol%), DMEDA (0.2 equiv.), NaN_3 (2.0 equiv.), DMSO/ H_2O (7:3), 50 °C, 16 h.

3.3.4 Synthesis of C6-isoxazole Functionalized 2-aminopurines

In an attempt to expand the nucleotide functionality, Hocek et al. prepared an adenosine modified with methylisoxazole at the C6 position, part of their studies towards novel cytostatic nucleosides.¹⁷⁶ To extend this methodology to 2-aminopurine analogues, it was decided to prepare isoxazole-modified guanosines bearing three different side-groups. Methyl and hexyl moieties would provide insight into the effects of side-chain length on transcriptional efficiency and fidelity, while a thiophene-functionalized isoxazole permits investigation into the effect of incorporating an extra 5-membered ring between the guanosine and the thiophene moieties.

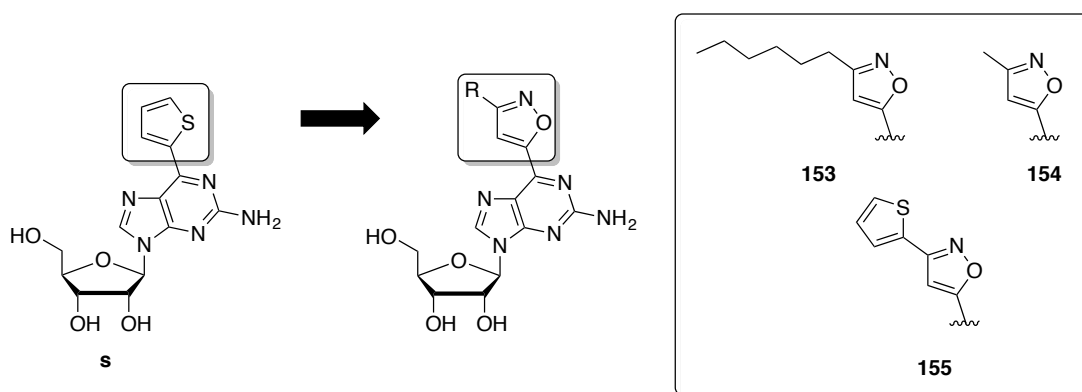
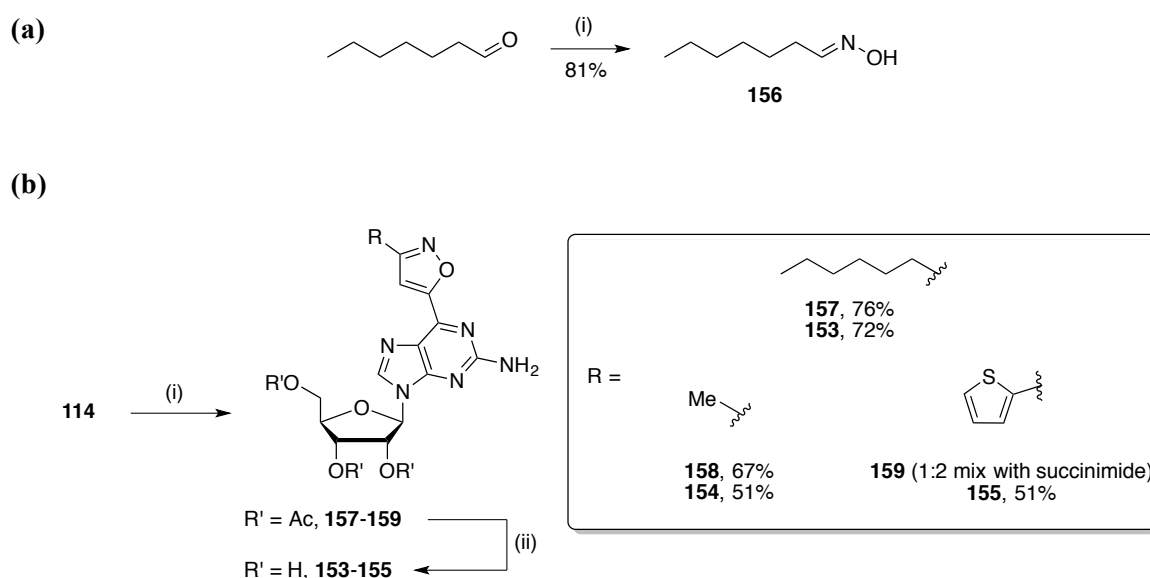


Figure 3-7. Proposed C6-triazole functionalized analogues **153-155**.

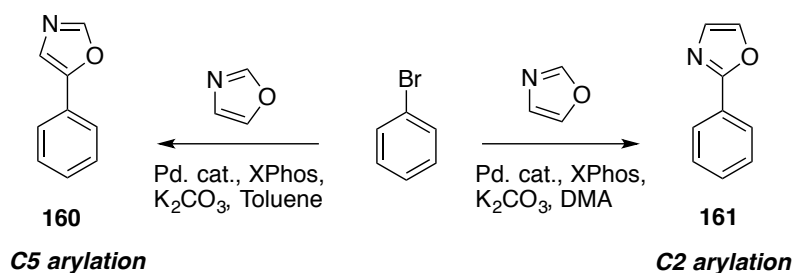
Hocek's isoxazole-functionalized adenosine was prepared *via* a [3+2] cycloaddition between an alkyne and a nitrile oxide, the latter being generated in situ from the corresponding oxime using *N*-chlorosuccinimide in dichloromethane. In accordance with this work, heptanal oxime **156** was firstly synthesized by employing a

literature procedure (Scheme 3-15a).¹⁷⁷ Using alkyne **114**, the cycloaddition protocol proceeded smoothly for the reaction with **156** and acetaldoxime providing the desired isoxazoles **157-158** in 67 and 76% yield respectively (Scheme 3-15b).¹⁷⁶ Subsequent deprotection of the acetates was then achieved under the same conditions reported for the triazole analogues furnishing free nucleosides **153-154** (Scheme 3-15b). The reaction of **114** with the corresponding thiophene carbaldehyde oxime was also investigated, which unfortunately afforded 46 mg of **159** as a 1:2 mixture with succinimide. Succinimide could have formed by the hydrolysis of N-chlorosuccinimide used in the reaction during the aqueous work-up. This material was carried forward to the deprotection step, in which the succinimide contamination was removed and **155** was obtained in 51% yield.



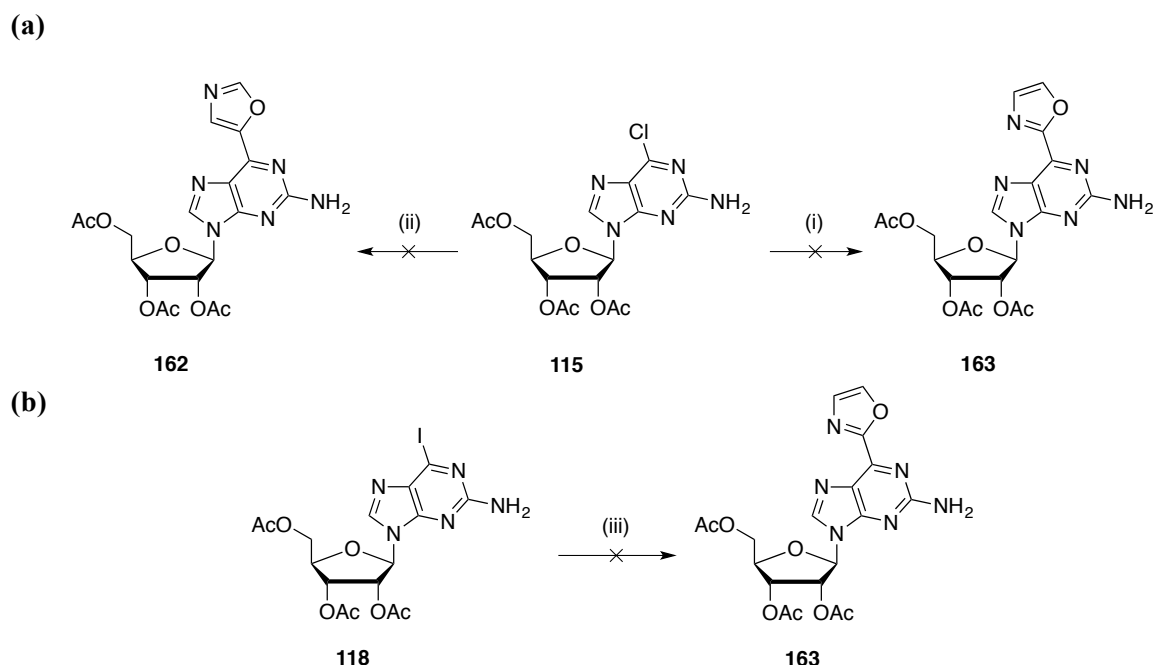
Scheme 3-15. (a) *Reagents and conditions:* (i) Hydroxylamine hydrochloride (2.0 equiv.), AcONa (3.0 equiv.), EtOH/H₂O r.t., 0°C to r.t., 3 h. (b) *Reagents and conditions:* (i) Oxime (1.2 equiv.), N-chlorosuccinimide (1.2 equiv.), pyridine (0.16 equiv.), r.t., 10 min. for **157** and **159** or 40 °C, 30 min. for **158**; then Et₃N (1.2 equiv.), DCM, r.t., 16 h. (ii) K₂CO₃ (0.3 equiv.), MeOH, r.t., 16 h.

To further expand our nucleoside palette the synthesis of C6-oxazole functionalized 2-aminopurines was explored. Oxazoles are similar moieties to isoxazoles. They both bear a nitrogen and an oxygen atom, but they differ on how these heteroatoms are arranged around the 5-membered ring. Comparing oxazole functionalized analogues with the previously synthesized isoxazole analogues would provide us with information regarding how this rearrangement can influence transcriptional efficiency and fidelity. A literature protocol was initially investigated.¹⁷⁸ The authors proposed that oxazole could undergo solvent-dependent, regioselective C-H activation and subsequent coupling with aryl halides, affording either C2 or C5 arylated oxazoles. When the authors run the reaction of bromobenzene and oxazole in toluene, a preference for C5-arylated product **160** was observed (23:1 ratio of C5 to C2), whereas solvent switch to DMA led to C2-arylated product **161** (21:1 ratio of C2 to C5) (Scheme 3-16).



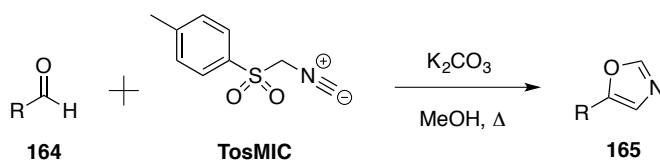
Scheme 3-16. Regioselective arylation of oxazole with bromobenzene according to Strotman et al.¹⁷⁸

This protocol would enable the installation of two regioisomeric oxazoles functionalities, from a single halide precursor, thus endowing our strategy with an extra point of modularity. Both conditions were investigated with chloride **115** as the halide coupling partner, but only starting material was observed upon TLC analysis (Scheme 3-17a). Iodide **118** was then subjected to the C2-selective conditions; again, only starting material was observed (Scheme 3-17b). These results suggest that the oxidative addition step is inefficient under these conditions. A systematic exploration of the catalyst/ligand/base combination would potentially achieve the desired transformation, but such investigations can be time consuming when three variables (catalyst, ligand, base) need to be explored. Therefore, the focus shifted on the investigation of alternative strategies for the installation of the oxazole moiety.



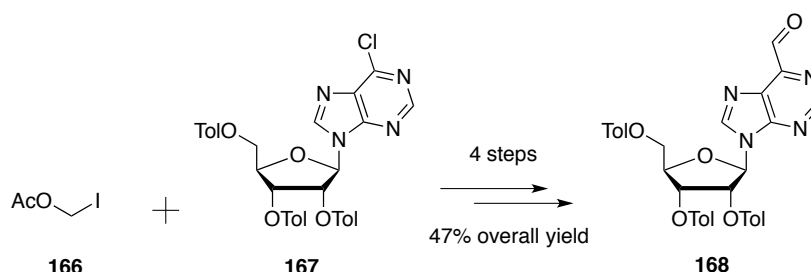
Scheme 3-17. (a) Reagents and conditions: (i) Pivalic acid (0.4 equiv.), K_2CO_3 (3.0 equiv.), $Pd(OAc)_2$ (5 mol%), Xphos (0.1 equiv.), oxazole (2.0 equiv.), toluene, 110 °C, 16 h. (ii) Pivalic acid (0.4 equiv.), K_2CO_3 (3.0 equiv.), $Pd(OAc)_2$ (5 mol%), Xphos (0.1 equiv.), oxazole (2.0 equiv.), DMA, 110 °C, 16 h. **(b) Reagents and conditions:** Pivalic acid (0.4 equiv.), Cs_2CO_3 (3.0 equiv.), $Pd(OAc)_2$ (5 mol%), Xphos (0.1 equiv.), oxazole (2.0 equiv.), toluene, 110 °C, 16 h.

Another well-known method for the synthesis of oxazoles is the Van Leusen reaction. In this reaction aldehyde **167** reacts with toluenesulfonylmethyl isocyanide (TosMIC) under basic conditions furnishing oxazole **168** (Scheme 3-18).¹⁷⁹



Scheme 3-18. The Van Leusen reaction in the synthesis of oxazoles.¹⁷⁸

It was envisioned that reaction of a C6-formyl 2-aminopurine with TosMIC would yield the desired oxazoles. A literature survey revealed that Hocek et al. have reported the synthesis of C6-formyl purines during their investigations on the synthesis of cytostatic nucleosides.^{180,181} However, access to formyl-purine **168** requires a four-step long synthesis starting from iodide **166** and halopurine **167** (Scheme 3-19). This strategy was concluded to be time-inefficient and our efforts shifted on the synthesis of C6-pyrazole functionalized analogues.



Scheme 3-19. Synthesis of aldehyde **160** according to Hocek et al.^{180,181}

3.3.5 Synthesis of C6-pyrazole Functionalized 2-aminopurines

Additionally, efforts to access pyrazole-functionalized guanosines were initiated. Pyrazole-functionalized nucleosides have been previously explored as potential antivirals, anticancers or as unnatural base pairs that pair through metal-chelation.¹⁸²⁻¹⁸⁴ Since literature precedent of C6-pyrazole functionalized nucleosides have been limited, the diversity of our analogues would be increased further by installing two different regioisomeric N-methylpyrazoles.¹⁷⁶ **169**, which has the methyl group facing towards the base-pairing area and **170**, where the methyl group is directed outwards (Figure 3-7). Comparing the transcriptional efficiency/fidelity of these distinct structures would provide insight regarding the importance of the methyl group direction and nitrogen arrangement around the pyrazole scaffold.

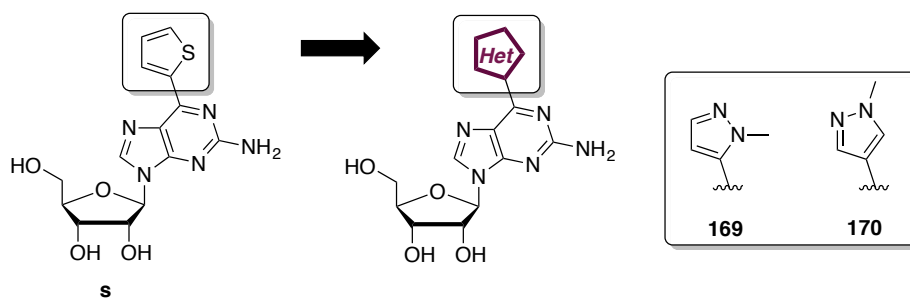
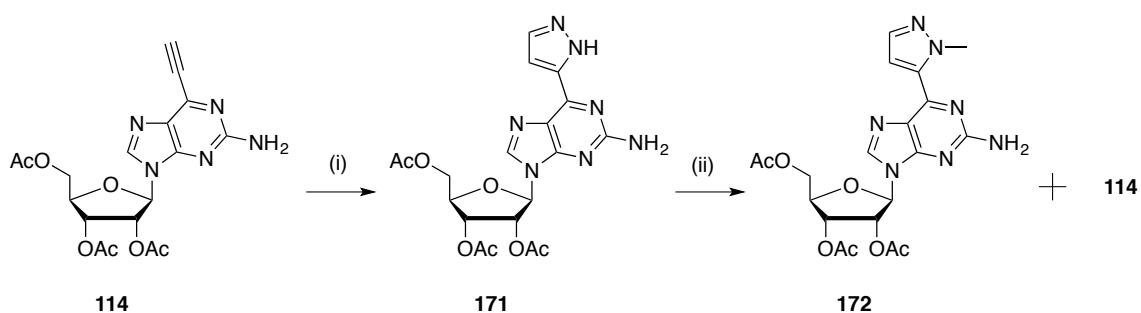


Figure 3-8. Proposed C6-pyrazole functionalized analogues.

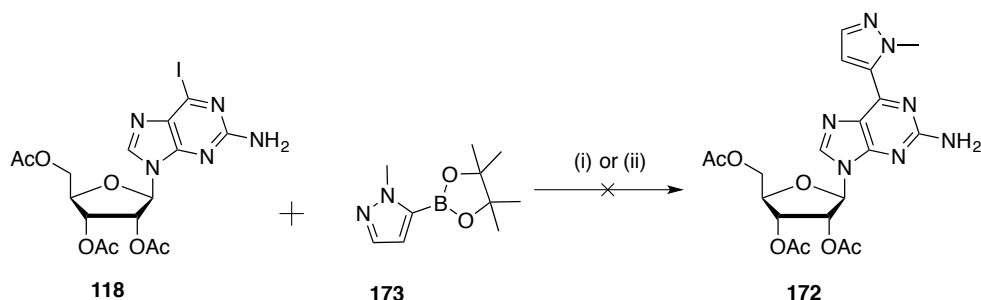
Initially, a two-step cycloaddition-methylation strategy was explored, as reported by Hocek et al (Scheme 3-20).¹⁷⁶ In the first step a CHCl_3 solution of **114** was treated with *in situ* generated diazomethane. TLC analysis of the crude reaction mixture revealed unreacted **114** along with formation of a polar spot that was identified as pyrazole **171** by LC-MS (Scheme 3-20). The residue was then concentrated and subjected to the next step without further purification. Methylation was carried out with methyl iodide and potassium carbonate in acetonitrile (Scheme 3-20). ESI-MS analysis identified the presence of expected product **172**. Three compounds were observed in TLC. One of them was isolated as a single spot and revealed to be alkyne **114** by $^1\text{H-NMR}$ analysis. The other two spots were isolated as a mixture and $^1\text{H-NMR}$ analysis revealed absence of the desired methyl moiety, although the characteristic pyrazole peaks were observed (two doublets at 7.3 and 7.7 ppm). Due to the complicated mixtures obtained throughout the scheme, this strategy was discontinued.



Scheme 3-20. Reagents and conditions: (i) CH_2N_2 (excess), CHCl_3 , r.t., 3 h. (ii) K_2CO_3 (2.0 equiv.), MeI (2.0 equiv.), MeCN, r.t., 16 h.

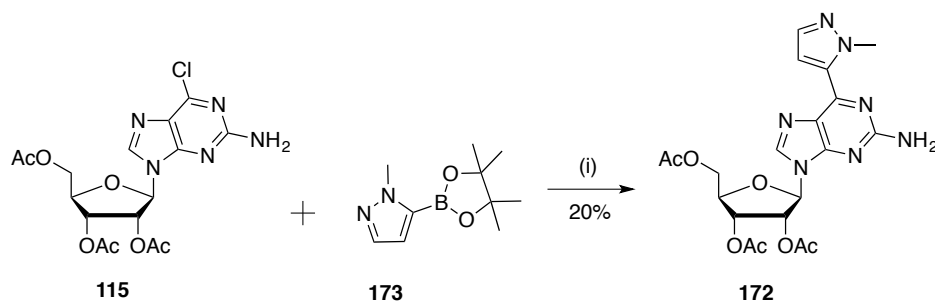
It was envisioned that a Suzuki-Miyaura coupling between iodide **118** or chloride **115** and commercially available boronate **173** could provide a straightforward strategy to access the desired compounds. The Hocek group has previously employed the combination of $\text{Pd}(\text{PPh}_3)_4/\text{K}_2\text{CO}_3/\text{toluene}$ to install various aryl or heteroaryl boronic acids on the C6 position of adenosine.¹⁷⁶ Furthermore, Lakshman et al. have reported the use of $\text{Pd}(\text{OAc})_2/\text{K}_3\text{PO}_4/\text{CyJohnPhos}/\text{toluene}$ combination to install various aryl moieties on the C6 position of 2'-deoxyguanosine.¹⁸⁵ Due to the observation that chloride **115** isn't readily soluble in toluene, these conditions were investigated with iodide **118**. Unfortunately, neither set of conditions provided pyrazole **172**

and only unreacted **118** was observed on TLC, suggesting that oxidative addition is not efficient under these conditions (Scheme 3-21).



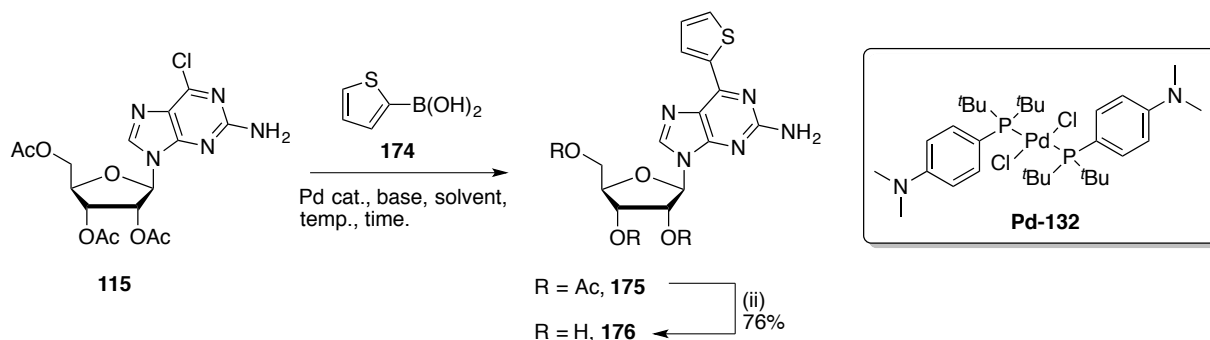
Scheme 3-21. Reagents and conditions: (i) Pd(PPh₃)₄ (5 mol%), K₂CO₃ (1.5 equiv.), boronate **173** (1.5 equiv.), toluene, 100 °C, 8 h. (ii) Pd(OAc)₂ (0.1 equiv.), CyJohnPhos (0.2 equiv.), K₃PO₄ (2.0 equiv.), boronate **173** (2.0 equiv.), toluene, 45 °C, 16 h.

A more recent study by the Hocek group has investigated the combination of Pd(dppf)Cl₂/K₂CO₃/DMF.¹⁸⁶ This was beneficial for this work since chloride **115** is readily soluble in DMF. Using boronate **173** and chloride **115**, the expected product was isolated in 20% yield (Scheme 3-22).



Scheme 3-22. Reagents and conditions: (i) Chloride **115** (1.1 equiv.), Pd(dppf)Cl₂ (5 mol%), K₂CO₃ (4.0 equiv.), DMF, 90 °C, 4 h.

To determine if the low yield was related to the chloride or the boronate substrate, the same set of conditions were explored with boronic acid **174**, which is commercially available and is also a key structural feature of Hirao's **s** nucleoside.¹¹⁶ RP-HPLC was employed to analyze the reaction conditions (Appendices 131-136). The results are summarized in Table 3-2.

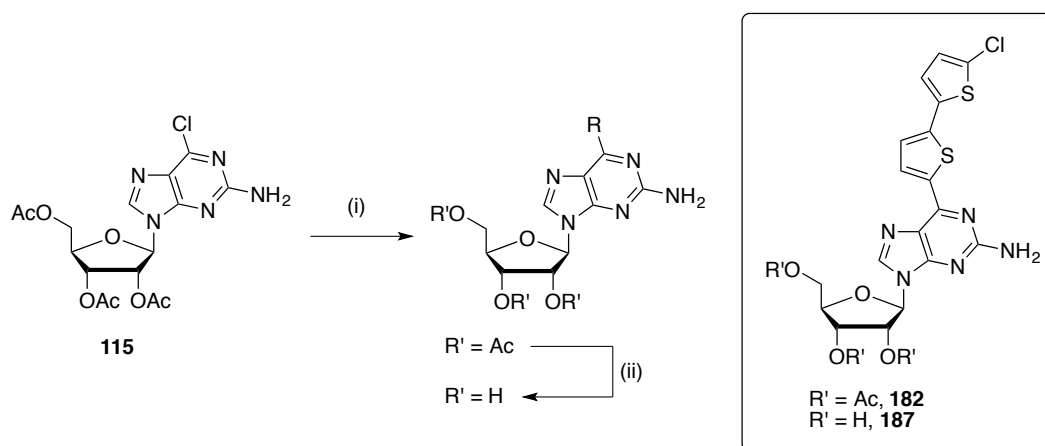
Table 3-2. Optimization of the Suzuki-Miyaura coupling of **115** with **174**. ^a Incomplete conversion or complicated mixture observed on HPLC. ^b 100% conversion observed on TLC. ^c Isolated yield.







Entry	Pd cat.	Base	Solvent	Temp.	Time	Yield
1	Pd(dppf)Cl ₂	K ₂ CO ₃	DMF	80 °C	4 h.	— ^a
2	Pd(dppf)Cl ₂	KF on alumina	DMF	80 °C	4 h.	— ^a
3	Pd(dppf)Cl ₂	KF on alumina	DMSO	80 °C	16 h.	43% ^{b,c}
4	Pd-132	KF on alumina	DMSO	80 °C	4 h.	81% ^{b,c}

After 4 hours, incomplete conversion was observed on HPLC, while increasing the reaction time to 16 hours resulted in a complex mixture of starting material, product and various unidentified side-products (Table 3-2, Entry 1). Previous studies have demonstrated that KF can be an efficient base for Suzuki-Miyaura couplings.^{187,188} In addition, it has been reported that due to the less basic nature of KF compared to K₂CO₃, KF is beneficial for Suzuki-Miyaura couplings involving compounds bearing base-labile groups, such as acetate esters.¹⁸⁹ Since solid KF can be difficult to handle (especially in older, wet samples), it was decided to employ alumina-supported KF, which is a free-flowing powder. Changing the base to KF didn't improve the conversion (Table 3-2, Entry 2); in addition, an un-identified side-product was observed. Finally, it was discovered that changing both the solvent and base to DMSO and KF respectively resulted in 100% conversion after 16 hours and **175** was isolated in 43% yield (Table 3-2, Entry 3). In an attempt to further enhance the yield, we decided to explore alternative Pd catalysts. Previous studies have shown that Pd-132 is highly effective in Suzuki-Miyaura couplings of heteroaryl boronic acids with heteroaryl chlorides.^{190,191} Using Pd-132 and KF on alumina as the catalyst-base combination resulted in 100% conversion after four hours and **175** was obtained in 81% yield. (Table 3-2, Entry 4).

With this set of optimized conditions in hand, various heteroaryl groups were successfully installed, albeit with impurities dependent on the reagent employed (Table 3-3).

Table 3-3. A novel Suzuki-Miyaura protocol enabled the direct installation of various heterocycles. (i) Pd-132 (0.1 equiv.), Het-M (1.3 equiv.), KF on alumina (4.4 equiv.), DMSO, 80 °C. (ii) K₂CO₃ (0.3 equiv.), MeOH, r.t. ^a Yield over two steps. ^b Isolated with 6% of **187**.



Entry	Het	M	Time	Yield ^a
1	 $\text{R}' = \text{Ac}, \mathbf{172}$ $\text{R}' = \text{H}, \mathbf{169}$	BPin	5 h	58%
2	 $\text{R}' = \text{Ac}, \mathbf{177}$ $\text{R}' = \text{H}, \mathbf{170}$	BPin	5 h	49%
3	 $\text{R}' = \text{Ac}, \mathbf{178}$ $\text{R}' = \text{H}, \mathbf{183}$	BPin	4 h	50 %
4	 $\text{R}' = \text{Ac}, \mathbf{179}$ $\text{R}' = \text{H}, \mathbf{184}$	B(OH) ₂	6 h	38%
5	 $\text{R}' = \text{Ac}, \mathbf{180}$ $\text{R}' = \text{H}, \mathbf{185}$	B(OH) ₂	8 h	44%
6	 $\text{R}' = \text{Ac}, \mathbf{181}$ $\text{R}' = \text{H}, \mathbf{186}$	B(OH) ₂	16 h	28% ^b

The use of boronic acid pinacol esters delivered products contaminated with pinacol (Entries 1-3), while coupling of a pyridine moiety afforded the desired compound contaminated with various unidentified impurities (Entry 4). Although these impurities were not separable from the acetyl-protected nucleosides, the

free nucleosides were isolated in pure form after acetate deprotection. In the case of (5-chlorothiophen-2-yl)boronic acid, the reaction yielded **181** contaminated with another compound, which was identified as **182** by LC-MS. Unfortunately, complete separation of these two compounds wasn't successful even after acetate deprotection and the amount of **187** was determined to be 6% by ¹H-NMR (Entry 5). Still this example demonstrates the potential of this methodology, which could be further optimized by increasing the equivalents of the boronic acid to form **163a** exclusively. Alternatively, further investigations on the temperature, catalyst loading and base could potentially yield **159a** exclusively.

3.4 Summary

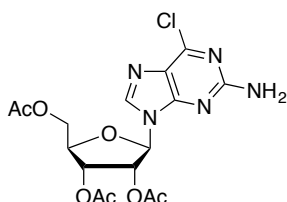
A robust synthetic route was developed, allowing quick access to a suite of C6-heteroaryl functionalized 2-aminopurines. Until the iodination step of **118**, no chromatographic purification was required and these were successfully carried out on a multi-gram scale. Employing common precursors **114** and **118** in [3+2] cycloadditions or Suzuki-Miyaura cross couplings a variety of C6-functionalized nucleosides with groups such as triazoles, isoxazoles, pyrazoles and thiophenes were obtained.

3.5 Experimental

LiN₃ has been prepared by another student within the Burley group (Clemence Mondielli, 2014) employing a literature protocol.¹⁹² All other solvents and reagents were supplied by Sigma-Aldrich, Acros Organics, Alfa Aesar, Fisher Scientific and Fluorochem and were used without further purification. ¹H and ¹³C NMR spectra were obtained on a Bruker DPX500 spectrometer at 500 MHz and 125 MHz respectively or on a Bruker DPX400 at 400 MHz and 100 MHz respectively. Low resolution mass spectra were recorded on an Agilent 1200 series LC-MS, equipped with a Zorbax 45mm x 150mm C18 column. A standard 45 min program using buffer C (5 mM ammonium acetate in water) and buffer D (5 mM ammonium acetate in acetonitrile) was used for all LC-MS analyses. The flow rate was 1.000 mL/min with a gradient as follows. 0-3 minutes held at 5% buffer D, then a linear gradient from 5-100% for 30 minutes, then the gradient was returned to 5% D and held at 5% buffer D for 3 minutes. The MS used was dual source with a capillary voltage of 4000 Volts for positive and negative modes and with a gas vaporizer temperature of 250 °C. HRMS spectra were recorded at the chemistry dept. of the University of Leicester using a Xevo QTOF mass spectrometer (Waters) coupled to an Acquity LC system (Waters) using an Acquity UPLC BEH C18 column (2.1 x 50 mm, Waters). The flow rate was 0.6 ml min⁻¹ and the gradient was as follows: 95% Solvent A (0.1% formic acid in water) with 5% solvent B (0.1% formic acid in acetonitrile) was held constant for 0.5 min, followed by a linear gradient to 100% B over the next 2.1 min. After 1 min at 100% solvent B, the gradient was returned to 95% solvent A and 5% solvent B over 0.2 min. The ESI capillary voltage was 3 kV, cone voltage 30 V and collision energy 4 eV. The MS acquisition rate was 10 spectra per second and m/z data ranging from 50 to 2000 Da was collected. Mass accuracy was achieved using a reference lock mass scan, once every 10 seconds. HPLC analysis of the Suzuki reaction conditions were carried out on a

DIONEX RP-HPLC system equipped with an Phenomenex Spherclone 5u ODS (2) 250 mm x 2.5 mm C18 column. A standard 44 min program using buffer A (0.1% TFA in H₂O) and buffer B (0.1% TFA in MeCN) was used. The flow rate was 1.0 mL/min with a gradient as follows. 0-3 minutes held at 5% buffer B, then a linear gradient from 5-75% for 30 minutes, then a linear gradient from 75-95% for 1 minute, then it was held at 95% buffer B for 5 minutes, then the gradient was returned to 5% B and held at 5% buffer B for 4 minutes.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-chloro-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **115¹⁶⁰**

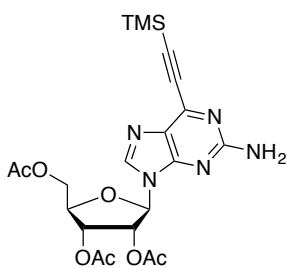


A suspension of 6-chloroguanosine riboside (100 mg, 0.33 mmol), DMAP (4 mg, 0.03 mmol) and Et₃N (0.28 mL, 2.0 mmol) in dry MeCN (1.3 mL) was cooled at 0 °C. Acetic anhydride was then added dropwise (0.1 mL, 1.0 mmol). The reaction was warmed to r.t. and the suspension was stirred at r.t. under argon for 2 hours. Any unreacted starting material was filtered off and the volatiles were removed *in vacuo*. The residue was recrystallized from boiling 2-propanol to afford the title compound as an off-white lumpy solid (93 mg, 66%).

¹H-NMR (500 MHz, acetone-d₆) δ: 8.2 (s, 1H), 6.19 (d, *J* = 5.2 Hz, 1H), 6.01 (t, *J* = 5.5 Hz, 1H), 5.70 (t, *J* = 5.2 Hz, 1H), 4.37-4.50 (m, 3H), 2.14 (s, 3H), 2.06 (s, 3H), 2.06 (s, 3H).

LCMS (+ve): [M+H]⁺ 428.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-((trimethylsilyl)ethynyl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **117**



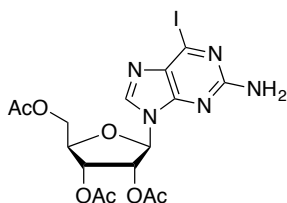
A mixture of **115** (40 mg, 0.09 mmol), [Pd(PPh₃)₂]Cl₂ (1.3 mg, 1.9 x 10⁻³ mmol), Et₃N (0.04 mL, 0.28 mmol) and CuI (1.1 mg, 5.6 x 10⁻³ mmol) was dissolved in dry and degassed DMF (1.3 mL). (Trimethylsilyl)acetylene (0.02 mL, 0.11 mmol) was then added and the yellow solution was stirred under argon at 60 °C for 16 h. It was then poured into brine (10mL) and extracted with EtOAc (2 x 10 mL). The organic phase was washed with brine (4 x 5 mL), dried over Na₂SO₄, filtered and the solvents were evaporated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with Et₂O/DCM (0:10 → 2:8) to afford **117** as a yellow solid (13mg, 28%). The starting material was recovered in 70% yield.

¹H-NMR (500 MHz, CDCl₃) δ: 8.04 (s, 1H), 6.03 (d, *J* = 5.1 Hz, 1H), 5.89 (t, *J* = 5.4 Hz, 1H), 5.69 (t, *J* = 4.8 Hz, 1H), 4.36-4.47 (m, 3H), 2.15 (s, 3H), 2.10 (s, 3H), 2.09 (s, 3H).

¹³C-NMR (125 MHz, CDCl₃) δ: 170.4, 169.6, 169.3, 158.6, 154.0, 142.3, 86.3, 80.2, 72.8, 70.5, 62.9, 34.8, 29.7, 20.7, 20.5, 20.4, 18.2, 11.7, 0.6.

LC-MS (+ve mode): $[M]^+$ 490.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-iodo-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate
118¹⁶³



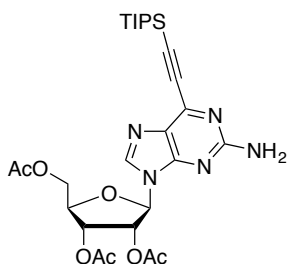
To a solution of **115** (150 mg, 0.36 mmol) in 2-butanone (7 mL) was added sodium iodide (1.0 g, 7.02 mmol) and TFA (0.06 mL, 1.77 mmol). The flask was covered with foil and the solution was stirred for 16 h under air. The resulting red solution was concentrated *in vacuo*. The residue was suspended in EtOAc, the suspension filtered through a pad of celite (the flask was rinsed with 3 x 10 mL EtOAc) and the solution was washed with sat. NaHCO₃ (1 x 20 mL) and sat. Na₂S₂O₃ (1 x 30 mL). The organic phase was dried over Na₂SO₄ and the solvent was removed *in vacuo*, affording iodide **118** as a yellow foam (185 mg, quant.).

¹H-NMR (500 MHz, CDCl₃) δ : 7.90 (s, 1H), 5.99 (d, $J = 5.0$ Hz, 1H), 5.95 (t, $J = 5.0$ Hz, 1H), 5.73 (t, $J = 4.7$ Hz, 1H), 5.31 (brs, 2H), 4.34-4.46 (m, 3H), 2.14 (s, 3H), 2.10 (s, 3H), 2.08 (s, 3H).

¹³C-NMR (125 MHz, CDCl₃) δ : 170.5, 169.6, 169.3, 158.8, 149.3, 140.1, 132.7, 123.3, 86.6, 80.0, 72.7, 70.5, 62.9, 20.7, 20.5, 20.4.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for C₁₆H₁₉IN₅O₇ 520.0329, found 520.0351.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-((triisopropylsilyl)ethynyl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **119**

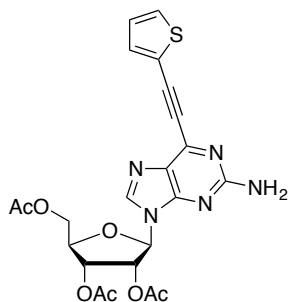


A mixture of **118** (40 mg, 0.08 mmol), [Pd(PPh₃)₂]Cl₂ (1.1 mg, 1.5 x 10⁻³ mmol), DIPEA (0.04 mL, 0.23 mmol) and CuI (0.87 mg, 4.62 x 10⁻³ mmol) was dissolved in dry DMF (1 mL). The mixture was subjected to the freeze-pump-thaw cycle three times. (Triisopropylsilyl)acetylene (0.02 mL, 0.09 mmol) was then added over 1 hour and the yellow solution was stirred under argon at r.t. for 16 h. It was then poured into brine (20 mL) and extracted with EtOAc (2 x 5mL). The organic phase was washed with brine (4 x 5 mL), dried over Na₂SO₄ and the solvents were removed *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with Et₂O/DCM (0:10 \rightarrow 2:8) to afford **119** as a yellow foam (43 mg, 97%).

¹H-NMR (500 MHz, CDCl₃) δ : 7.8 (s, 1H), 6.02 (d, $J = 5.2$ Hz, 1H), 5.96 (t, $J = 5.6$ Hz, 1H), 5.75 (t, $J = 4.9$ Hz, 1H), 5.50 (s, 2H), 4.33-4.44 (m, 3H), 2.14 (s, 3H), 2.08 (s, 3H), 2.07 (s, 3H), 1.18 (m, 21H).

¹³C-NMR (125 MHz, CDCl₃) δ : 170.0, 169.0, 168.7, 159.2, 152.7, 142.2, 140.4, 129.4, 101.0, 100.2, 85.5, 80.0, 72.1, 70.1, 62.5, 20.2, 20.0, 19.9, 18.2, 10.7.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for C₂₇H₄₀N₅O₇Si 574.2697, found 574.2721.

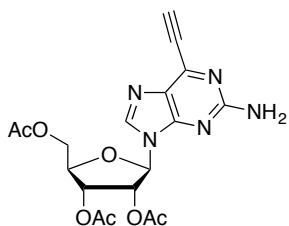
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(thiophen-2-ylethynyl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 120

A mixture of **118** (100 mg, 0.19 mmol), $[\text{Pd}(\text{PPh}_3)_2]\text{Cl}_2$ (2.6 mg, 3.8×10^{-3} mmol), DIPEA (0.1 mL, 0.57 mmol) and CuI (2.2 mg, 11.4×10^{-3} mmol) was dissolved in dry and degassed DMF (2.37 mL). The mixture was subjected to the freeze-pump-thaw cycle three times. 2-Ethynylthiophene (0.02 mL, 0.09 mmol) was then added over 1 hour and the brown solution was stirred under argon at r.t. for 16 h. It was then poured into brine (20 mL) and extracted with EtOAc (2 x 5 mL). The organic phase was washed with brine (4 x 5 mL), dried over Na_2SO_4 and the solvents were evaporated *in vacuo*. The residue was purified *via* column chromatography (SiO_2) eluting with $\text{Et}_2\text{O}/\text{DCM}$ (0:10 \rightarrow 2:8) to afford **120** as a brown foam (64 mg, 68%).

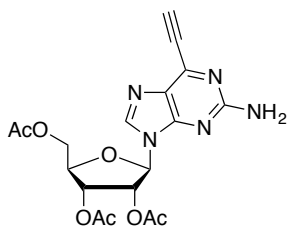
$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 7.92 (brs, 1H), 7.52 (d, $J = 3.8$ Hz, 1H), 7.41 (d, $J = 5.11$, 1H), 7.03 (dd, $J = 3.7, 5.3$, 1H), 6.07 (d, $J = 5.01$ Hz, 1H), 5.96 (t, $J = 5.2$ Hz, 1H), 5.75 (t, $J = 4.9$, 1H), 5.46 (brs, 2H), 4.45-4.39 (m, 2H), 4.37-4.32 (m, 1H) 2.11 (s, 3H), 2.07 (s, 3H), 2.06 (s, 3H).

$^{13}\text{C-NMR}$ (125 MHz, CDCl_3) δ : 170.5, 169.6, 169.4, 159.6, 153.1, 142.4, 141.1, 135.1, 130.1, 129.0, 127.4, 121.2, 91.0, 88.0, 86.4, 79.9, 72.8, 70.5, 63.0, 20.7, 20.5, 20.4.

HRMS (LC-MS) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{22}\text{H}_{22}\text{N}_5\text{O}_7\text{S}$ 500.1240, found 500.1262.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-ethynyl-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 114

Alkyne **119** (120 mg, 0.21 mmol) was dissolved in THF (3 mL). The yellow solution was cooled at 0 °C and TBAF (60 mg, 0.23 mmol) was added. The mixture was stirred at 0 °C for 30 min. The solvent was evaporated *in vacuo* and the residue purified *via* column chromatography (SiO_2) eluting with EtOAc affording **114** as a yellowish foam (43 mg, 49%).

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-ethynyl-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 114¹⁶⁴

Alkyne **118** (50 mg, 0.087 mmol) was added to a flame-dried round bottom flask and dissolved in dry MeCN (0.96 mL). The flask was covered with tin foil and AgF was then added (16.5 mg, 0.13 mmol). The mixture was stirred at r.t. under argon for 16 h. It was then treated with 1M HCl aq. (0.26 mL) and stirred for 5 min. The mixture was filtered through a celite pad (the flask was rinsed with

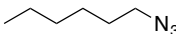
EtOAc 3 x 10 mL). The mixture was transferred to a separatory funnel and washed with sat. NaHCO₃ (1 x 20 mL) and brine (1 x 20 mL). The organic layer was dried over Na₂SO₄, filtered and concentrated in vacuo. Column chromatography purification (SiO₂) eluting with EtOAc/*n*-hexane (0:10 to 10:0) afforded the title compound as an off-white foam (22 mg, 60%).

¹H-NMR (400 MHz, CDCl₃) δ: 7.9 (s, 1H), 6.01 (d, *J* = 5.2 Hz, 1H), 6.00 (t, *J* = 5.6 Hz, 1H), 5.75 (t, *J* = 4.9 Hz, 1H), 5.30 (s, 2H), 4.32-4.44 (m, 3H), 3.6 (s, 1H), 2.11 (s, 3H), 2.08 (s, 3H), 2.06 (s, 3H).

¹³C-NMR (100 MHz, CDCl₃) δ: 170.0, 169.1, 169.0, 159.1, 152.6, 141.3, 140.7, 129.3, 85.8, 84.5, 79.4, 77.3, 72.2, 70.5, 69.9, 62.4, 20.2, 20.1, 20.0.

HRMS (LC-MS) *m/z*: [M+H]⁺ calcd for C₁₈H₂₀N₅O₇ 418.1363 found 419.2062.

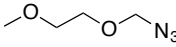
1-azidohexane **127**¹⁷⁰

 To a solution of 1-bromohexane (0.853 mL, 6 mmol) in DMSO (12 mL), NaN₃ (702 mg) was added and the colorless solution was stirred at ambient temperature and atmosphere for 16 h. Brine (10 mL) was then added, the solution was transferred to a separating funnel and extracted with Et₂O (2 x 10 mL). The organic layer was then washed with brine (4 x 10 mL). The organic layer was dried over Na₂SO₄ and concentrated *in vacuo* to afford 1-azidohexane as a yellow oil, which was used without further purification (485 mg, 63%).

¹H-NMR (500 MHz, CDCl₃) δ: 3.27 (t, *J* = 7.1 Hz, 2H), 1.61 (p, *J* = 7.1 Hz, 2H), 1.41-1.29 (m, 6H), 0.91 (t, *J* = 6.9 Hz, 3H).

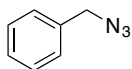
ESI-MS (+ve mode): [2M+Na]⁺ 277.

1-(azidomethoxy)-2-methoxyethane **128**¹⁶⁸

 A mixture of sodium azide (626.6 mg, 9.64 mmol), 1-(chloromethoxy)-2-methoxyethane (0.55 mL, 4.82 mmol), and dry MeCN (16 mL) was placed in a dry, two-necked round-bottomed flask, which was then heated to reflux under argon for 3 h. The reaction mixture was subsequently cooled to room temperature and mL of diethyl ether was added. This solution was transferred to a separating funnel and washed with brine (50 mL) solution. The organic layer was dried over anhydrous Na₂SO₄, filtered and concentrated *in vacuo* to give 267 mg (42%) of azide **128**, as a pale yellow oil. This product was used without further purification.

¹H-NMR (500 MHz, CDCl₃) δ: 4.70 (s, 2H), 3.76 (m, 2H), 3.58 (m, 2H), 3.39 (s, 3H).

GC-MS (-ve mode) [M-H]⁻ 130.

(azidomethyl)benzene 129¹⁶⁹

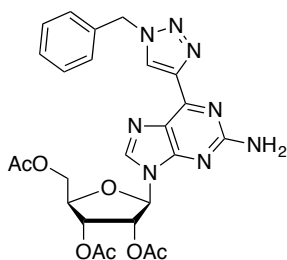
To a stirred solution of benzyl bromide (0.12 mL, 1.00 mmol), in a mixture of acetone/H₂O (50 mL, 4:1 v/v) sodium azide (95 mg, 1.5 mmol) was added and the mixture was stirred at r.t. under air for 16 h. The reaction was extracted with diethyl ether (3 x 20 mL), washed with brine (50 mL), and dried over Na₂SO₄. The organic phase was filtered and concentrated *in vacuo* to provide azide **129** as a pale yellow oil, which was used without further purification (133 mg, quant.).

¹H-NMR (500 MHz, CDCl₃) δ: 7.46-7.33 (m, 5H), 4.36 (s, 2H).

LC-MS (+ve mode): [M-N₂]⁺ 106.

GENERAL PROCEDURE FOR THE SYNTHESIS OF C6-TRIAZOLE 2-AMINOPURINES (GP1)

Alkyne **114** and the corresponding azide were dissolved in DMSO and H₂O in a screw cap vial equipped with a stir bar. In a separate vial, TBTA was dissolved in DMSO and a solution of CuSO₄ (0.4 M in H₂O) was added. The resulting green solution was vortexed and added to the alkyne solution. Sodium ascorbate was then added and the solution was stirred at r.t. under air for 16 h.

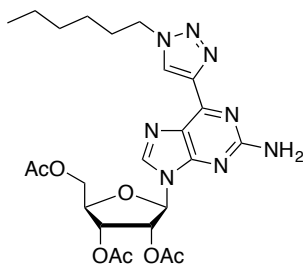
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(1-benzyl-1H-1,2,3-triazol-4-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 130

This compound was prepared following GP1 using alkyne **114** (45.5 mg, 0.11 mmol), benzyl azide (0.02 ml, 0.13 mmol), DMSO (0.73 mL), H₂O (0.88 mL), TBTA (5.9 mg, 0.01 mmol), DMSO (for TBTA, 0.11 mL), CuSO₄ (13.75 μL, 0.4M in H₂O, 5 x 10⁻³ mmol) and sodium ascorbate (22 mg, 0.11 mmol). The mixture was diluted with brine (5 ml) and EtOAc (5 mL) and the aqueous phase was extracted with EtOAc. The combined organic extracts were then washed with brine (4 x 10 mL) and the organic phases were concentrated, dried over Na₂SO₄, and concentrated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with acetone/EtOAc (0:10 → 2:8) to afford the desired compound as a yellow foam (56 mg, 92%).

¹H-NMR (400 MHz, CDCl₃) δ: 8.7 (s, 1H), 7.87 (s, 1H), 7.29-7.4 (m, 5H), 6.07 (d, *J* = 4.8 Hz, 1H), 5.98 (t, *J* = 5.4 Hz, 1H), 5.80 (t, *J* = 4.9 Hz, 1H), 5.65 (s, 2H), 4.33-4.47 (m, 3H), 2.10 (s, 3H), 2.09 (s, 3H), 2.07 (s, 3H).

¹³C-NMR (125 MHz, CDCl₃) δ: 170.5, 169.7, 169.4, 159.9, 153.6, 148.5, 143.5, 140.3, 134.5, 129.2, 128.8, 128.0, 127.0, 124.7, 86.2, 79.9, 72.8, 70.6, 62.3, 54.3, 20.7, 20.6, 20.4.

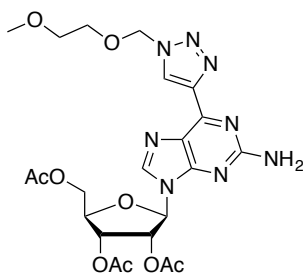
HRMS (LC-MS) *m/z*: [M+H]⁺ calcd for 550.1924, found 551.2021.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(1-hexyl-1H-1,2,3-triazol-4-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 131

This compound was prepared following GP1 using alkyne **114** (70 mg, 0.17 mmol), 1-azidohexane (27.8 mg, 0.22 mmol), DMSO (0.8 mL), H₂O (0.3 mL), TBTA (8.8 mg, 0.02 mmol), DMSO (0.16 mL), CuSO₄ (21.0 μL, 0.4M in H₂O, 8.5 x 10⁻³ mmol) and sodium ascorbate (33.1 mg, 0.17 mmol). The mixture was diluted with brine (5 mL) and EtOAc (5 mL). The aqueous phase was extracted with EtOAc and the combined organic extracts were washed with brine (4 x 10 mL). Column chromatography purification (SiO₂) eluting with a mixture of acetone/DCM (0:10 → 1:1) afforded **131** as an off-white solid, which was contaminated with TBTA and other impurities (86 mg, ~94%).
¹H-NMR (500 MHz, CDCl₃) δ: 8.70 (s, 1H), 7.88 (s, 1H), 6.06 (d, *J* = 4.9 Hz, 1H), 5.96 (t, *J* = 5.2 Hz, 1H), 5.77 (t, *J* = 4.97 Hz, 1H), 5.52 (brs, 2H), 4.42 (m, 4H), 4.33 (dd, *J* = 4.1, 11.1 Hz, 1H) 2.10 (s, 3H), 2.05 (s, 3H), 2.04 (s, 3H), 1.93 (m, 2H), 1.32-1.23 (m, 6H), 0.82 (t, *J* = 7.1 Hz, 3H).

¹³C-NMR (125 MHz, CDCl₃) δ: 170.5, 169.6, 169.4, 160.1, 153.4, 148.7, 143.0, 140.0, 129.0, 126.9, 86.1, 79.9, 72.9, 70.5, 63.0, 50.6, 31.1, 30.3, 21.1, 22.3, 20.7, 20.5, 20.4, 13.9.

HRMS (LC-MS) *m/z*: [M+H]⁺ calcd for C₂₄H₃₃N₈O₇ 545.2472, found 545.2487.

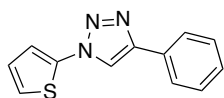
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(1-((2-methoxyethoxy)methyl)-1H-1,2,3-triazol-4-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 132

This compound was prepared following GP1 using alkyne **114** (100 mg, 0.24 mmol), azide **128** (41 mg, 0.31 mmol), DMSO (1.22 mL), H₂O (0.44 mL), TBTA (12.7 mg, 0.024 mmol), DMSO (0.22 mL), CuSO₄ (30.0 μL, 0.4M in H₂O, 8.5 x 10⁻³ mmol) and sodium ascorbate (47.6 mg, 0.24 mmol). The mixture was diluted with H₂O (10 mL) and lyophilized. Column chromatography purification (SiO₂) eluting with a mixture of acetone/DCM (0:10 → 1:1) afforded triazole **132** as an orange gum, which was then co-evaporated three times with DCM to collect the product as a yellow foam (118 mg, 90%).

¹H-NMR (400 MHz, CDCl₃) δ: 8.94 (s, 1H), 7.91 (s, 1H), 6.07 (d, *J* = 4.7 Hz, 1H), 5.99 (t, *J* = 5.1 Hz, 1H), 5.85 (s, 2H), 5.80 (d, *J* = 5.3 Hz, 1H), 5.46 (brs, 2H), 4.47-3.6 (m, 3H), 3.71 (t, *J* = 2.0 Hz, 2H), 3.50 (m, 2H), 3.33 (s, 3H) 2.13 (s, 3H), 2.10 (s, 3H), 2.09 (s, 3H), 2.06 (s, 3H).

¹³C-NMR (100 MHz, CDCl₃) δ: 170.0, 169.2, 168.9, 159.5, 153.0, 147.8, 143.4, 139.8, 128.6, 127.5, 126.7, 85.8, 79.4, 78.8, 72.4, 70.7, 68.5, 62.5, 58.5, 20.2, 20.0, 19.9.

HRMS (LC-MS) *m/z*: [M+H]⁺ calcd for C₂₂H₂₉N₈O₉ 549.2057, found 549.2049.

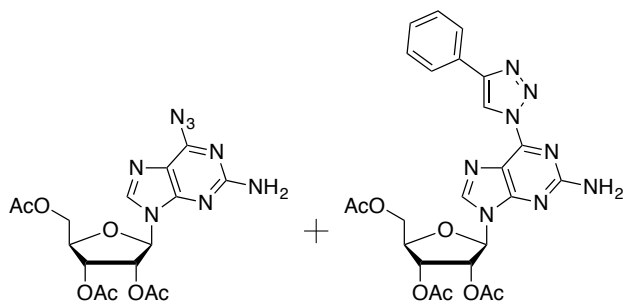
4-phenyl-1-(thiophen-2-yl)-1H-1,2,3-triazole 138¹⁷³

2-iodothiophene (52 μ L, 0.48 mmol), and phenylacetylene (50 μ L, 0.48 mmol), were dissolved in EtOH (1.4 mL) and H₂O (0.6 mL). NaN₃ (62 mg, 0.96 mmol) was added, followed by NaAsc (9.52 mg, 0.048 mmol), DMEDA (10.3 μ L, 0.096 mmol) and CuI (9 mg, 0.048 mmol). The yellow solution was stirred at 50 °C for 16 hours, turning dark brown. It was then treated with NH₄OH (5 mL). The mixture was then diluted with brine (5 mL) and extracted with EtOAc (3 x 20 mL). The organic phase was dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (1:9 \rightarrow 4:6) to afford **138** as a white solid (12 mg, 12%).

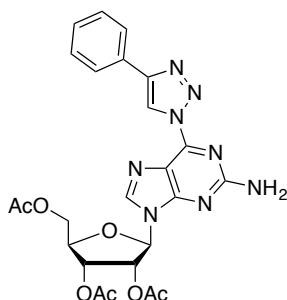
¹H-NMR (500 MHz, CDCl₃) δ : 8.11 (s, 1H), 7.90 (d, J = 7.2 Hz, 2H), 7.47 (t, J = 7.3 Hz, 2H), 7.40 (t, J = 7.3 Hz, 1H), 7.31 (dd, J = 3.7, 1.3 Hz, 1H), 7.25 (dd, J = 5.4, 1.2 Hz, 1H).

LC-MS (+ve mode): [M+H]⁺ 228.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(4-phenyl-1H-1,2,3-triazol-1-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 152 and (2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-azido-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 139¹⁷³



In a vial iodide **118** (200 mg, 0.38 mmol) and phenylacetylene (41 μ L, 0.38 mmol) were dissolved in DMSO (1.12 mL) and H₂O (0.48 mL). NaN₃ (50 mg, 0.76 mmol) was added, followed by NaAsc (7.6 mg, 0.038 mmol), DMEDA (10.3 μ L, 0.076 mmol) and CuI (7.2 mg, 0.038 mmol). The yellow solution was stirred at 50 °C for 16 hours, turning dark green. It was then diluted with EtOAc (20 mL) and washed with brine (4 x 10 mL). The aqueous phase was back extracted with EtOAc (10 mL). The organic phases were combined, dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with acetone/DCM (0:10 \rightarrow 2:8) to afford **152** as an off-white solid (8.2 mg, 4%). Azide **139** was obtained as the major product as an off-white solid (63 mg, 38%). Spectroscopic data were in accordance with the data presented below for **152** and **139**.

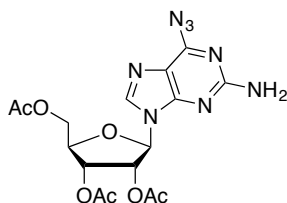
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(4-phenyl-1H-1,2,3-triazol-1-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 152

In a screw cap vial azide **139** (63 mg, 0.15 mmol) and phenylacetylene (0.02 ml, 0.19 mmol) were dissolved in DMSO (0.71 mL) and H₂O (1.1 mL). In a separate vial, TBTA (7.7 mg, 0.015 mmol) was dissolved in DMSO (0.14 mL) and a solution of CuSO₄ (18.1 μ L, 0.4M in H₂O, 7 x 10⁻³ mmol) was added. The resulting green solution was vortexed and added to the alkyne solution. Sodium ascorbate was then added (28.7 mg, 0.15 mmol) and the brown solution was stirred at r.t. under air for 16 h. It was then diluted with EtOAc and washed with brine (4 x 5 mL). The aqueous phase were back extracted with EtOAc (10 mL). The organic phases were concentrated, dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with acetone/EtOAc (0:10 \rightarrow 2:8) to afford **152** as an off-white solid (57 mg, 74%).

¹H-NMR (500 MHz, CDCl₃) δ : 9.30 (s, 1H), 8.00 (brs, 1H), 7.97 (d, J = 7.3 Hz, 1H), 7.45 (t, J = 7.6 Hz, 2H), 7.37 (t, J = 7.4 Hz, 1H), 6.11 (d, J = 4.8 Hz, 1H), 6.03 (t, J = 5.2 Hz, 1H), 5.81 (t, J = 5.2 Hz, 1H), 5.66 (s, 2H), 4.50-4.37 (m, 3H), 2.16 (s, 3H), 2.12 (s, 3H), 2.09 (s, 3H).

¹³C-NMR (125 MHz, CDCl₃) δ : 170.5, 169.7, 169.5, 159.5, 155.8, 148.1, 145.5, 141.4, 129.9, 128.9, 128.6, 126.2, 120.2, 117.0, 86.6, 80.0, 72.9, 70.5, 63.0, 20.7, 20.6, 20.5.

HRMS (microTOF-Q) m/z : [M+Na]⁺ calcd for C₂₄H₂₄N₈NaO₇ 559.1660, found 559.1642.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-azido-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 139¹⁷³

Iodide **118** (100 mg, 0.19 mmol) was dissolved in DMSO (0.56 mL) and H₂O (0.24 mL). NaN₃ (25 mg, 0.39 mmol) was added, followed by NaAsc (3.8 mg, 0.019 mmol), DMEDA (4.12 μ L, 0.038 mmol) and CuI (3.6 mg, 0.019 mmol). The yellow solution was stirred at 50 °C for 3 hours, turning dark green. It was then diluted with EtOAc and washed with brine (4 x 5 mL). The aqueous phase was back extracted with EtOAc (10 mL). The organic phases were concentrated, dried over Na₂SO₄ and the solvents were evaporated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with acetone/DCM (0:10 \rightarrow 2:8) affording azide **139** as an off-white solid, which exists in equilibrium with its tetrazole tautomer (38.5 mg, 46%).

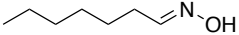
¹H-NMR (500 MHz, CDCl₃, only the peaks corresponding the major isomer are reported) δ : 8.01 (s, 1H), 6.96 (s, 2H), 6.13 (d, J = 4.9 Hz, 1H), 6.00 (t, J = 5.1 Hz, 1H), 5.72 (t, J = 4.8 Hz, 1H), 4.40-4.51 (m, 3H), 2.15 (s, 3H), 2.10 (s, 3H), 2.07 (s, 3H).

¹³C-NMR (125 MHz, CDCl₃, peaks corresponding to both tautomers are reported, due to difficulties distinguishing some of the peaks corresponding to the major isomer) δ : 170.8, 169.7, 169.5, 169.5, 159.4,

153.7, 146.1, 143.4, 142.8, 139.6, 138.9, 118.9, 114.7, 87.2, 87.2, 86.5, 80.2, 79.9, 72.9, 72.7, 70.6, 70.5, 63.1, 63.0, 20.8, 20.7, 20.6, 20.4.

HRMS (microTOF-Q) m/z : $[M+Na]^+$ calcd for $C_{16}H_{18}N_8NaO_7$ 457.1191, found 457.1170.

(*E*)-heptanaloxime **156**¹⁷⁷

 To a round bottom flask was added hydroxylamine hydrochloride (1.21 g, 17.5 mmol) and dissolved in EtOH (24 mL) and H₂O (5 mL). Heptanal (1.23 mL, 8.75 mmol) was then added and the solution was cooled to 0 °C. AcONa (2.15 g, 26.25 mmol) was then added portionwise. The reaction was warmed up to r.t. and stirred under air for 3 h. The mixture was then diluted with DCM (50 mL) and washed with H₂O (2 x 50 mL). The organic layer was dried over Na₂SO₄ and the solvent evaporated *in vacuo* affording oxime **156** as a white solid, which was used in the next step without further purification (920 mg, 81%).

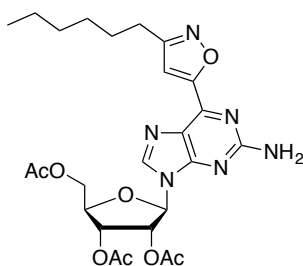
¹H-NMR (500 MHz, CDCl₃) δ : 8.29 (s, 1H), 6.74 (t, $J = 5.5$ Hz, 1H), 2.40 (dt, $J = 7.5, 5.5$ Hz, 2H), 1.56-1.46 (m, 2H), 1.42-1.28 (m, 6H), 0.94-0.88 (m, 3H).

GC-MS (-ve mode) $[M-OH]^-$ 112.1.

GENERAL PROCEDURE FOR THE SYNTHESIS OF C6-ISOXAZOLE 2-AMINOPURINES (GP2)¹⁷⁶

N-Chlorosuccinimide and pyridine were added into a screw-cap vial and dissolved in DCM. A solution of the corresponding oxime in DCM was added and the reaction was stirred at ambient conditions for 10 min (except for acetaldoxime, which was stirred at 40 °C for 20 min). A solution of **114** in DCM was added, followed by the dropwise addition of a solution of Et₃N in DCM. The solution was stirred under air at r.t. for 16 h. The solvent was evaporated *in vacuo* and the residue dissolved in EtOAc, washed with brine (2 x 30 mL), and dried over Na₂SO₄. The solvents were evaporated *in vacuo* and the residue purified *via* column chromatography (SiO₂) to furnish the corresponding isoxazole.

(2*R*,3*R*,4*R*,5*R*)-2-(acetoxymethyl)-5-(2-amino-6-(3-hexylisoxazol-5-yl)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **157**



157 was prepared according to GP2 using N-chlorosuccinimide (38.7 mg, 0.29 mmol), pyridine (1.92 μ L, 0.038 mmol) in DCM (1 mL), a solution of heptaldoxime (37.7 mg, 0.29 mmol) in DCM (1 mL), a solution of **114** (100 mg, 0.24 mmol) in DCM (1.92 mL) and a solution of Et₃N (40 μ L, 0.29 mmol) in DCM (1.92 mL). The residue was purified *via* column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (4:6 \rightarrow 10:0) affording the title compound as a

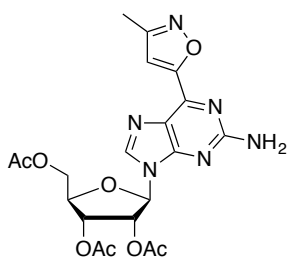
yellow gum, which could be obtained as yellow foam after three times co-evaporation with DCM (100 mg, 76%).

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 7.94 (s, 1H), 7.40 (s, 1H), 6.08 (d, $J = 4.9$ Hz, 1H), 6.01 (t, $J = 5.1$ Hz, 1H), 5.81 (t, $J = 5.0$ Hz, 1H), 5.66 (brs, 1H), 4.48-4.34 (m, 3H), 2.78 (t, $J = 7.6$ Hz, 1H), 2.15 (s, 3H), 2.12 (s, 3H), 2.08 (s, 3H), 1.74 (q, $J = 7.5$ Hz, 2H), 1.42-1.23 (m, 8H), 0.92-0.82 (m, 3H).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3) δ : 169.9, 169.1, 168.9, 164.4, 163.4, 159.4, 153.5, 144.0, 140.7, 124.1, 108.1, 85.8, 79.4, 72.4, 70.0, 62.4, 30.9, 28.3, 27.8, 25.6, 22.0, 20.2, 20.1, 19.9, 13.5.

HRMS (CI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{25}\text{H}_{33}\text{N}_6\text{O}_8$ 545.2360, found 545.2369.

(2*R*,3*R*,4*R*,5*R*)-2-(acetoxymethyl)-5-(2-amino-6-(3-methylisoxazol-5-yl)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **158**



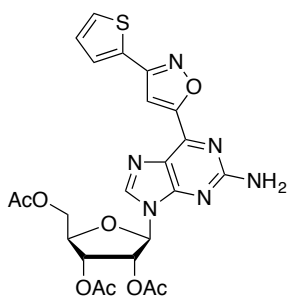
158 was prepared according to GP2 using N-chlorosuccinimide (38.7 mg, 0.29 mmol), pyridine (1.92 μL , 0.038 mmol) in DCM (1 mL), a solution of acetaldoxime (17.7 μL , 0.29 mmol) in DCM (1 mL), a solution of **114** (100 mg, 0.24 mmol) in DCM (1.92 mL) and a solution of Et_3N (40 μL , 0.29 mmol) in DCM (1.92 mL). The residue was purified *via* column chromatography (SiO_2) eluting with EtOAc/n -hexane (4:6 \rightarrow 10:0) affording the title compound as a yellowish foam (76 mg, 67%).

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 7.92 (s, 1H), 7.37 (s, 1H), 6.07 (d, $J = 4.9$ Hz, 1H), 6.02 (t, $J = 5.1$ Hz, 1H), 5.81 (t, $J = 5.1$ Hz, 1H), 5.59 (brs, 1H), 4.48-4.34 (m, 3H), 2.42 (s, 3H), 2.15 (s, 3H), 2.11 (s, 3H), 2.07 (s, 3H).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3) δ : 170.6, 169.7, 169.6, 164.1, 160.6, 159.9, 154.0, 144.4, 141.3, 124.7, 109.6, 86.4, 80.0, 72.9, 70.5, 63.0, 60.4, 20.8, 20.6, 20.5, 11.6.

HRMS (LC-MS) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{20}\text{H}_{23}\text{N}_6\text{O}_8$ 475.1577, found 475.1589.

(2*R*,3*R*,4*R*,5*R*)-2-(acetoxymethyl)-5-(2-amino-6-(3-(thiophene-2-yl)isoxazol-5-yl)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate (138a**)**



159 was prepared according to GP2 using N-chlorosuccinimide (38.7 mg, 0.29 mmol), pyridine (1.92 μL , 0.038 mmol) in DCM (1 mL), a solution of thiophene-2-carbaldehyde oxime (36.8 mg, 0.29 mmol) in DCM (1 mL), a solution of **114** (100 mg, 0.24 mmol) in DCM (1.92 mL) and a solution of Et_3N (40 μL , 0.29 mmol) in DCM (1.92 mL). The residue was purified *via* column chromatography (SiO_2) eluting with EtOAc/n -hexane (4:6 \rightarrow 10:0) affording 46 mg of **138a** as a 1:2 mixture with succinimide.

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 8.01 (s, 1H), 7.78 (s, 1H), 7.64-7.62 (dd, J = 3.67, 1.1 Hz, 1H), 7.48-7.46 (dd, J = 5.1, 1.1 Hz, 1H), 7.18-7.14 (dd, J = 5.2, 3.7 Hz, 1H), 6.11 (d, J = 4.9 Hz, 1H), 6.04 (t, J = 5.1 Hz, 1H), 5.81 (t, J = 5.1 Hz, 1H), 5.62 (brs, 2H), 4.52-4.36 (m, 3H), 2.17 (s, 3H), 2.13 (s, 3H), 2.11 (s, 3H).

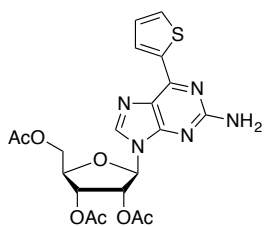
$^{13}\text{C-NMR}$ (100 MHz, CDCl_3) δ : 170.1, 169.3, 169.1, 164.2, 159.3, 157.9, 153.7, 143.5, 141.1, 129.8, 127.6, 127.5, 127.2, 124.2, 106.2, 86.9, 79.5, 72.4, 70.0, 62.5, 20.3, 20.1, 20.0.

HRMS (LC-MS) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{23}\text{H}_{23}\text{N}_6\text{O}_8\text{S}$ 543.1298, found 543.1309.

GENERAL PROCEDURE FOR THE SYNTHESIS OF C6-HETEROARYL 2-AMINOPURINES EMPLOYING THE SUZUKI PROTOCOL (GP3)

A flame-dried round bottom flask was charged with chloride **115** (100 mg, 0.23 mmol), the corresponding boronic acid or boronic acid pinacol ester (0.35 mmol), KF on alumina (153 mg, 1.1 mmol, 40% wt. loading) and Pd-132 (16 mg, 0.02 mmol). The solids were dissolved in DMSO (1.5 mL). The solution was purged with argon for 15 minutes and then heated at 80 °C for the respective time indicated. The mixture was diluted with EtOAc (25 mL) and brine (25 mL). The organic layer was washed with brine (4 x 25 mL), then collected, dried over Na_2SO_4 , filtered and concentrated *in vacuo*. The residue was purified *via* column chromatography (SiO_2) to afford the corresponding C6-heteroaryl nucleosides.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(thiophen-2-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **175**

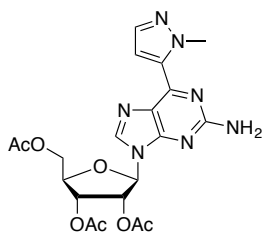


This compound was prepared following GP3 using thiophen-2-ylboronic acid. The reaction time was 4 hours. Purification *via* column chromatography (SiO_2) eluting with acetone:DCM (0:1 \rightarrow 2:8) afforded **175** as a yellow foam (89 mg, 81%).

$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 8.49 (d, J = 2.7 Hz, 1H), 7.8 (s, 1H), 7.49 (dd, J = 5.0, 0.8 Hz, 1H), 7.14 (dd, J = 5.0, 3.8 Hz, 1H), 5.99 (d, J = 5.0 Hz, 1H), 5.95 (t, J = 5.0 Hz, 1H), 5.76 (t, J = 5.0 Hz, 1H), 5.10 (brs, 2H), 4.41-4.29 (m, 3H), 2.08 (s, 3H), 2.03 (s, 3H), 2.02 (s, 3H).

$^{13}\text{C-NMR}$ (125 MHz, CDCl_3) δ : 170.5, 169.6, 169.4, 159.5, 153.4, 151.3, 140.0, 132.5, 130.32, 128.5, 124.1, 86.2, 79.9, 72.8, 70.6, 63.1, 29.7, 20.8, 20.6, 20.4.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{20}\text{H}_{21}\text{N}_5\text{NaO}_7\text{S}$ 498.1054, found 498.1030.

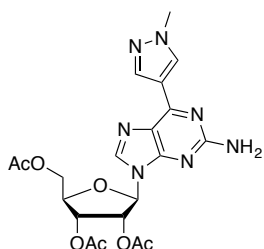
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(1-methyl-1H-pyrazol-5-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 172

This compound was prepared following GP3 using 1-methyl-5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazole. The reaction time was 5 hours. Purification *via* column chromatography (SiO₂) eluting with acetone:DCM (0:1 → 3:7) afforded **172** as an off-white foam contaminated with pinacol, that was removed in the next step.

¹H-NMR (500 MHz, CDCl₃) δ: 7.89 (s, 1H), 7.56 (d, *J* = 2.0 Hz, 1H), 7.48 (d, *J* = 2.0 Hz, 1H), 6.106 (d, *J* = 5.0 Hz, 1H), 6.00 (t, *J* = 5.2 Hz, 1H), 5.78 (t, *J* = 5.1 Hz, 1H), 5.26 (brs, 2H, partial overlap with DCM peak), 4.46-4.33 (m, 3H), 4.31 (s, 3H), 2.12 (s, 3H), 2.07 (s, 3H), 2.06 (s, 3H).

¹³C-NMR (100 MHz, acetone-d₆) δ: 170.8, 170.2, 170.0, 160.8, 154.8, 148.9, 141.5, 138.4, 136.9, 125.3, 112.8, 86.9, 81.0, 73.5, 71.7, 63.9, 41.2, 20.7, 20.5, 20.4

HRMS (LC-MS) *m/z*: [M+H]⁺ calcd for C₂₀H₂₄N₇O₇ 474.1737, found 474.1738.

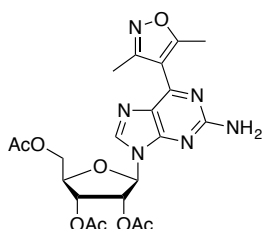
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(1-methyl-1H-pyrazol-4-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 177

This compound was prepared following GP3 using 1-methyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazole. The reaction time was 5 hours. Purification *via* column chromatography (SiO₂) eluting with acetone:DCM (0:1 → 1:1) afforded **177** as an off-white foam contaminated with a picanol related impurity, which was removed in the next step.

¹H-NMR (500 MHz, CDCl₃) δ: 8.41 (s, 1H), 8.32 (s, 1H), 7.76 (s, 1H), 6.0 (d, *J* = 5.1 Hz, 1H), 5.93 (t, *J* = 5.0 Hz, 1H), 5.75 (t, *J* = 5.0 Hz, 1H), 5.16 (brs, 2H), 4.40-4.26 (m, 3H), 3.90 (s, 3H), 2.06 (s, 3H), 2.02 (s, 3H), 2.00 (s, 3H).

¹³C-NMR (100 MHz, CDCl₃) δ: 170.5, 169.6, 169.4, 159.9, 152.8, 151.5, 140.1, 139.2, 132.6, 124.6, 119.1, 86.2, 79.8, 72.8, 70.6, 63.0, 39.2, 20.7, 20.5, 20.4.

HRMS (LC-MS) *m/z*: [M+H]⁺ calcd for C₂₀H₂₄N₇O₇ 474.1737, found 474.1750.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(3,5-dimethylisoxazol-4-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate 178

This compound was prepared following GP3 using (3,5-dimethylisoxazol-4-yl)boronic acid. The reaction time was 4 hours. Purification *via* column chromatography (SiO₂) eluting with acetone:DCM (0:1 → 2:8) afforded **178** as a

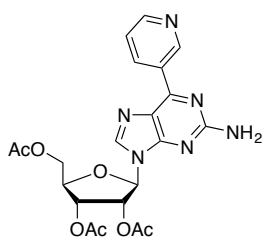
yellowish foam contaminated with a picanol related impurity, which was removed in the next step.

$^1\text{H-NMR}$ (500 MHz, CDCl_3) δ : 7.79 (s, 1H), 5.99 (d, 1H), 5.94 (t, 1H), 5.75 (t, 1H), 5.21 (brs, 2H), 4.43-4.29 (m, 3H), 2.55 (s, 3H), 2.39 (s, 3H), 2.08 (s, 3H), 2.05 (s, 3H), 2.01 (s, 3H).

$^{13}\text{C-NMR}$ (125 MHz, CDCl_3) δ : 170.5, 170.2, 169.6, 169.5, 159.7, 159.4, 153.2, 151.4, 139.9, 126.4, 112.4, 86.5, 82.7, 79.8, 75.0, 72.8, 70.5, 63.0, 20.7, 20.5, 20.4

HRMS (CI) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{21}\text{H}_{25}\text{N}_6\text{O}_8$ 489.1734, found 489.1749 (100%), 490.1751 (27%), 491.1884 (22%).

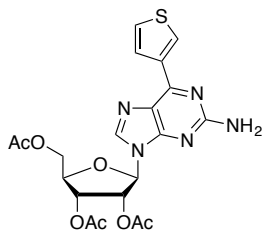
(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(pyridin-3-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **179**



This compound was prepared following GP3 using pyridine-3-ylboronic acid. The reaction time was 6 hours. Purification *via* column chromatography (SiO_2) eluting with acetone: CHCl_3 (0:1 \rightarrow 1:1) afforded **179** as a red solid, which was contaminated with various aromatic impurities and was used directly in the next step.

LC-MS (+ve mode): $[\text{M}+\text{H}]^+$ 471.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(thiophen-3-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **180**



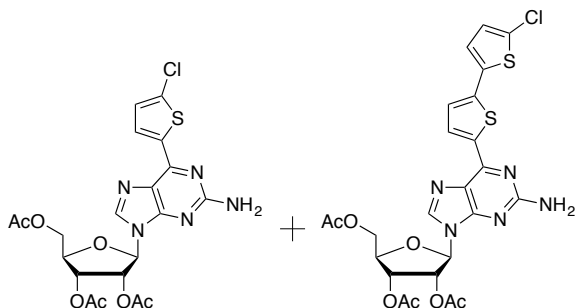
This compound was prepared following GP3 using thiophen-3-ylboronic acid. The reaction time was 8 hours. Purification *via* column chromatography (SiO_2) eluting with acetone: DCM (0:1 \rightarrow 2:8) afforded **180** as a yellowish foam (61 mg, 56%)

$^1\text{H-NMR}$ (600 MHz, CDCl_3) δ : 8.78 (d, $J = 2.2$ Hz, 1H), 8.16 (d, $J = 5.0$ Hz, 1H), 7.89 (s, 1H), 7.41 (dd, $J = 3.2, 1.9$ Hz, 1H), 6.08 (d, $J = 5.0$ Hz, 1H), 6.04 (t, $J = 5.1$ Hz, 1H), 5.84 (t, $J = 5.23$ Hz, 1H), 5.17 (brs, 2H), 4.49-4.43 (m, 2H), 4.39 (dd, $J = 11.6, 4.4$ Hz, 1H), 2.15 (s, 3H), 2.11 (s, 3H), 2.10 (s, 3H).

$^{13}\text{C-NMR}$ (150 MHz, CDCl_3) δ : 170.5, 169.6, 169.4, 159.8, 153.5, 152.2, 139.8, 138.2, 130.4, 127.8, 125.6, 125.2, 86.2, 79.9, 72.8, 70.6, 63.1, 20.7, 20.6, 20.4.

HRMS (FAB) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{20}\text{H}_{21}\text{N}_7\text{NaO}_7\text{S}$ 457.1191, found 457.1170.

(2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(5-chlorothiophen-2-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **181 and (2R,3R,4R,5R)-2-(acetoxymethyl)-5-(2-amino-6-(5'-chloro-[2,2'-bithiophen]-5-yl)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **182****



This compound was prepared following GP3 using (5-chlorothiophen-2-yl)boronic acid. After 4 hours, incomplete conversion was observed. Another 1.5 equiv. of the boronic acid was added and the reaction was stirred for another 4 hours. Purification *via* column chromatography (SiO₂) eluting with acetone:DCM (0:1 → 2:8) afforded **181** as a bright yellow foam (46 mg, 39%) which was contaminated with 7% of side-product **182**. Only the peaks corresponding to **181** are reported. The peaks corresponding to **182** are shown in Appendix 87.

¹H-NMR (400 MHz, CDCl₃) δ: 8.32 (d, *J* = 4.1 Hz, 1H), 7.85 (s, 1H), 7.01 (d, *J* = 4.0 Hz, 1H), 6.05 (d, *J* = 4.8 Hz, 1H), 6.01 (t, *J* = 5.0 Hz, 1H), 5.82 (t, *J* = 5.0 Hz, 1H), 5.12 (brs, 2H), 4.48-4.34 (m, 3H), 2.15 (s, 3H), 2.10 (s, 3H), 2.08 (s, 3H).

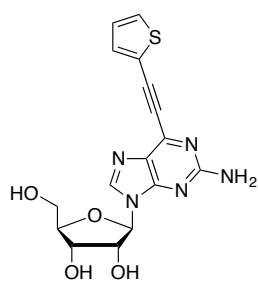
¹³C-NMR (100 MHz, CDCl₃) δ: 170.0, 169.1, 168.9, 158.9, 152.9, 149.8, 139.6, 138.2, 135.2, 131.5, 127.3, 123.4, 85.8, 79.4, 70.1, 62.5, 20.2, 20.1, 19.9.

HRMS (microTOF-Q) *m/z*: [M+Na]⁺ calcd for C₂₀H₂₀N₅NaO₇SCl 532.064, found 532.065.

GENERAL PROCEDURE FOR ACETATE DEPROTECTION (GP4)

In a screw cap vial equipped with a stir bar, the protected nucleoside was added and dissolved in MeOH. K₂CO₃ was then added and the solution was stirred at r.t. under air for 16 h, unless otherwise noted.

(2R,3R,4S,5R)-2-(2-amino-6-(thiophen-2-ylethynyl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **121**



121 was prepared according to GP4 using **120** (99 mg, 0.2mmol), K₂CO₃ (8.2 mg, 0.06 mmol) and MeOH (3.3 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as an off-white solid (56 mg, 75%).

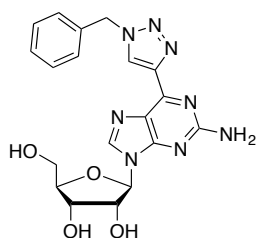
¹H-NMR (400 MHz, DMSO-d₆) δ: 8.39 (s, 1H), 7.85 (dd, *J* = 5.2, 1.1 Hz, 1H), 7.62 (dd, *J* = 3.7, 1.2 Hz, 1H), 7.22 (dd, *J* = 3.7, 5.2, 1H), 6.71 (brs, 2H), 5.84 (d, *J* = 5.8

Hz, 1H), 4.51 (t, $J = 5.35$ Hz, 1H), 4.14 (t, $J = 4.1$ Hz, 1H), 3.92 (q, $J = 3.9$ Hz, 1H), 3.67 (dd, $J = 11.2, 4.2$ Hz, 1H), 3.56 (dd, $J = 11.8, 4.2$ Hz, 1H).

^{13}C -NMR (125 MHz, CDCl_3) δ : 160.8, 154.2, 141.9, 140.9, 135.4, 131.7, 128.7, 127.9, 120.7, 89.4, 87.0, 85.8, 79.6, 73.4, 70.7, 61.7.

HRMS (LC-MS) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{16}\text{H}_{16}\text{N}_5\text{O}_4\text{S}$ 374.0923, found 374.0932.

(2R,3R,4S,5R)-2-(2-amino-6-(1-benzyl-1H-1,2,3-triazol-4-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 124



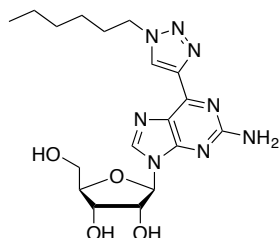
124 was prepared according to GP4 using **130** (94 mg, 0.17 mmol), K_2CO_3 (7.12 mg, 0.05 mmol) and MeOH (2.7 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a white solid (33.2 mg, 46%).

^1H -NMR (500 MHz, DMSO-d_6) δ : 9.04 (s, 1H), 8.36 (s, 1H), 7.39 (m, 4H), 7.40-7.35 (m, 5H), 6.62 (s, 2H), 5.86 (d, $J = 5.8$ Hz, 1H), 5.75 (s, 2H), 5.45 (brs, 1H), 5.16 (brs, 1H), 5.05 (t, $J = 5.5$ Hz, 1H), 4.52 (q, $J = 5.1$ Hz, 1H), 4.14 (quartet, 1H), 3.91 (q, $J = 4$ Hz, 1H), 3.65 (m, 1H), 3.55 (m, 1H).

^{13}C -NMR (125 MHz, DMSO-d_6) δ : 161.6, 154.9, 149.2, 144.4, 143.3, 136.5, 130.1, 129.7, 129.3, 128.4, 125.3, 90.5, 87.6, 75.1, 72.5, 63.3, 55.2.

HRMS (LC-MS) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{19}\text{H}_{21}\text{N}_8\text{O}_4$ 425.1686, found 425.1690.

(2R,3R,4S,5R)-2-(2-amino-6-(1-hexyl-1H-1,2,3-triazol-4-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 125



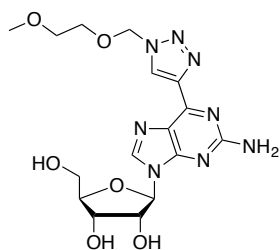
125 was prepared according to GP4 using **131** (86 mg, 0.16 mmol), K_2CO_3 (6.5 mg, 0.05 mmol) and MeOH (2.6 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a white solid (33.5 mg, 50%).

^1H -NMR (400 MHz, DMSO-d_6) δ : 8.99 (s, 1H), 8.39 (s, 1H), 6.61 (brs, 1H), 5.96 (d, $J = 5.96$ Hz, 1H), 4.55-5.48 (m, 3H), 4.16 (t, $J = 4.27$ Hz, 1H), 3.93 (q, $J = 3.78$ Hz, 1H), 3.68 (dd, $J = 12.00, 4.11$ Hz, 1H), 3.57 (dd, $J = 12.00, 4.11$ Hz, 1H), 1.89 (t, $J = 6.47$ Hz, 1H), 1.29 (m, 6H), 0.86 (m, 3H).

^{13}C -NMR (100 MHz, MeOD) δ : 159.7, 153.0, 147.4, 142.1, 141.4, 126.5, 123.4, 88.7, 85.7, 73.4, 70.6, 61.5, 49.8, 30.4, 29.3, 25.3, 21.6, 12.4.

HRMS (FAB) m/z : $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{18}\text{H}_{27}\text{N}_8\text{O}_4$ 419.2155, found 419.2172.

(2R,3R,4S,5R)-2-(2-amino-6-(1-((2-methoxyethoxy)methyl)-1H-1,2,3-triazol-4-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **126**



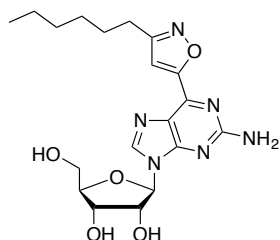
126 was prepared according to GP4 using **132** (100 mg, 0.18 mmol), K_2CO_3 (7.6 mg, 0.06 mmol) and MeOH (3 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a yellow hygroscopic solid (48 mg, 63%).

1H -NMR (400 MHz, MeOD) δ : 8.94 (s, 1H), 8.21 (s, 1H), 5.96 (d, $J = 6.15$ Hz, 1H), 5.91 (s, 2H), 4.80 (t, $J = 5.50$, 1H), 4.40 (triplet, 1H), 4.18 (quartet, 1H), 3.91 (dd, $J = 12.29$, 2.43 Hz, 1H), 3.82-3.76 (m, 3H), 3.55 (m, 2H), 3.37 (s, 3H).

^{13}C -NMR (100 MHz, MeOD) δ : 159.6, 153.0, 147.0, 142.6, 141.5, 126.94, 123.4, 88.5, 85.7, 78.7, 73.3, 70.6, 70.5, 68.4, 61.5, 57.2.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for $C_{16}H_{23}N_8O_6$ 423.1741, found 423.1756.

(2R,3R,4S,5R)-2-(2-amino-6-(3-hexylisoxazol-5-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **153**



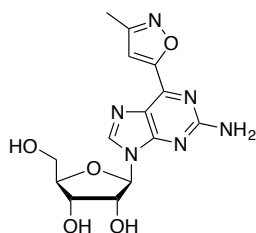
153 was prepared according to GP4 using **157** (100 mg, 0.184 mmol), K_2CO_3 (7.6 mg, 0.06 mmol) and MeOH (3 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a yellow solid (50 mg, 65%).

1H -NMR (400 MHz, MeOD) δ : 8.42 (s, 1H), 7.42 (s, 1H), 6.01 (d, $J = 5.40$ Hz, 1H), 4.73 (t, $J = 5.40$ Hz, 1H), 4.37 (dd, $J = 5.09$, 3.66 Hz, 1H), 4.15 (q, $J = 3.21$ Hz, 1H), 3.91 (dd, $J = 12.21$, 2.96 Hz, 1H), 3.79 (dd, $J = 12.25$, 3.27 Hz, 1H), 2.81 (t, $J = 7.52$ Hz, 1H), 1.78 (q, $J = 7.36$ Hz, 1H), 1.49 (m, 6H), 0.94 (m, 3H).

^{13}C -NMR (100 MHz, MeOD) δ : 164.5, 164.0, 159.7, 153.9, 143.1, 142.6, 123.4, 107.0, 88.4, 85.5, 73.9, 70.4, 61.2, 30.7, 28.0, 27.4, 24.9, 21.7, 12.5.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for $C_{19}H_{27}N_6O_8$ 419.2043, found 419.2055.

Synthesis of (2R,3R,4S,5R)-2-(2-amino-6-(3-methylisoxazol-5-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **154**



154 was prepared according to GP4 using **158** (170 mg, 0.35 mmol), K_2CO_3 (14.1 mg, 0.105 mmol) and MeOH (5.2 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a yellow solid (88 mg, 72%).

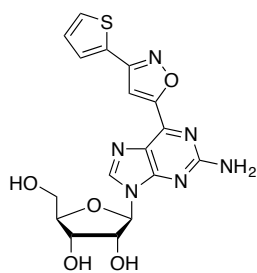
1H -NMR (400 MHz, DMSO- d_6) δ : 8.47 (s, 1H), 7.40 (s, 1H), 6.83 (s, 2H), 5.89 (d, J

= 5.82 Hz, 1H), 4.54 (t, $J = 5.57$ Hz, 1H), 4.16 (t, $J = 4.12$ Hz, 1H), 3.94 (quartet, 1H), 3.67 (dd, $J = 11.87$, 3.96 Hz, 1H), 3.57 (dd, $J = 11.87$, 4.06 Hz, 1H), 2.36 (s, 3H).

^{13}C -NMR (100 MHz, DMSO- d_6) δ : 164.3, 160.3, 160.2, 154.6, 142.9, 142.0, 123.2, 108.9, 86.5, 85.4, 73.5, 70.2, 61.2, 11.0.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for $\text{C}_{14}\text{H}_{17}\text{N}_6\text{O}_5$ 349.1260, found 349.1263.

(2R,3R,4S,5R)-2-(2-amino-6-(3-(thiophen-2-yl)isoxazol-5-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 155



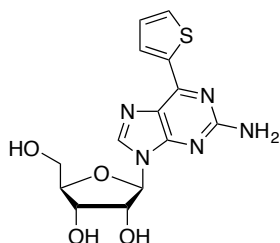
155 was prepared according to GP4 using **159** (169 mg, 0.31 mmol), K_2CO_3 (12.9 mg, 0.09 mmol) and MeOH (5.2 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a brownish solid (70 mg, 51%).

^1H -NMR (400 MHz, DMSO- d_6) δ : 8.52 (s, 1H), 7.89 (s, 1H), 7.84 (d, $J = 2.82$ Hz, 1H), 7.74 (d, $J = 5.18$ Hz, 1H), 7.22 (t, $J = 4.34$ Hz, 1H), 5.92 (d, $J = 5.68$ Hz, 1H), 4.56 (t, $J = 5.39$ Hz, 1H), 4.19 (obscured by 3 hydroxy protons), 3.96 (q, $J = 4.04$ Hz, 1H), 3.69 (dd, $J = 12.35$, 3.70 Hz, 1H), 3.59 (dd, $J = 11.99$, 3.88 Hz, 1H).

^{13}C -NMR (100 MHz, DMSO- d_6) δ : 165.4, 160.2, 157.9, 154.8, 142.5, 142.3, 129.3, 129.2, 128.8, 128.1, 123.4, 105.47, 86.7, 85.3, 73.6, 70.2, 61.1.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for $\text{C}_{17}\text{H}_{17}\text{N}_6\text{O}_8\text{S}$ 417.0981, found 417.0995.

(2R,3R,4S,5R)-2-(2-amino-6-(thiophen-2-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 176



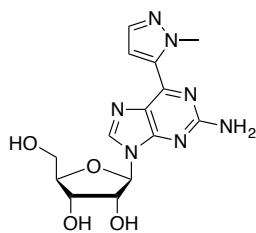
176 was prepared following GP4 using **175** (68 mg, 0.14 mmol), K_2CO_3 (6 mg, 0.04 mmol) and MeOH (1.4 mL). Reaction time was 16 hours. The mixture was then adsorbed into SiO_2 and loaded on a column. Purification was carried out eluting with MeOH/DCM (0:1 \rightarrow 1:9) affording the title compound as a white foam (38 mg, 76%).

^1H -NMR (500 MHz, MeOD) δ : 8.48 (m, 1H), 8.25 (d, $J = 1.3$ Hz, 1H), 7.68 (d, $J = 4.9$ Hz, 1H), 7.23 (m, 1H), 5.96 (dd, $J = 6.1$, 1.1 Hz, 1H), 4.78 (t, $J = 3.7$ Hz, 1H), 4.38 (m, 1H), 4.18 (m, 1H), 3.92 (d, $J = 12.3$ Hz, 1H), 3.8 (d, $J = 12.3$ Hz, 1H)

^{13}C -NMR (125 MHz, MeOD) δ : 159.8, 153.2, 151, 141.5, 139.9, 128.0, 123.2, 89.2, 86.2, 73.7, 71.2, 62.0.

HRMS (FAB) m/z : $[M+H]^+$ calcd for $\text{C}_{14}\text{H}_{15}\text{N}_5\text{O}_4\text{S}$ 349.0845, found 349.0828.

(2R,3R,4S,5R)-2-(2-amino-6-(1-methyl-1H-pyrazol-5-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 169



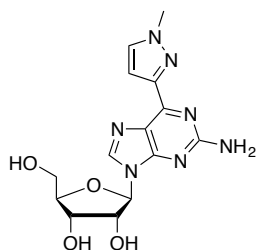
169 was prepared according to GP4 using **172** (90 mg, 0.19 mmol), K_2CO_3 (7.8 mg, 0.06 mmol) and MeOH (3.16 mL). After 16 hours, the mixture was adsorbed into silica and purified *via* column chromatography (SiO_2) eluting with MeOH/DCM (0:10→2:8) affording the title compound as a gray powder (42 mg, 63%).

1H -NMR (400 MHz, MeOD) δ : 8.28 (s, 1H), 7.58 (d, $J = 1.98$ Hz, 1H), 7.42 (d, $J = 1.98$ Hz, 1H), 5.97 (d, $J = 5.95$ Hz, 1H), 4.74 (t, $J = 5.65$ Hz, 1H), 4.35 (dd, $J = 3.20, 1.98$ Hz, 1H), 4.33 (s, 3H), 4.15 (q, $J = 2.9$ Hz, 1H), 3.92-3.88 (dd, $J = 12.29, 2.90$ Hz, 1H), 3.80-3.75 (dd, $J = 12.36, 3.05$ Hz, 1H).

^{13}C -NMR (100 MHz, MeOD) δ : 154.9, 149.3, 143.0, 139.1, 137.9, 112.6, 90.4, 87.5, 75.2, 72.4, 63.3, 40.6, 9.3.

HRMS (FAB) m/z : $[M+H]^+$ calcd for $C_{14}H_{18}N_7O_4$ 348.1420, found 348.1414.

(2R,3R,4S,5R)-2-(2-amino-6-(1-methyl-1H-pyrazol-4-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 170



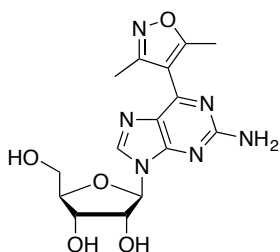
170 was prepared according to GP4 using **177** (264 mg, 0.56 mmol), K_2CO_3 (23 mg, 0.17 mmol) and MeOH (9.3 mL). After 16 hours, the mixture was concentrated *in vacuo* and the residue was recrystallized from MeOH/water, affording the free nucleoside as a grey solid (100 mg, 51%).

1H -NMR (400 MHz, MeOD) δ : 8.57 (s, 1H), 8.37 (s, 1H), 8.21 (s, 1H), 5.94 (d, $J = 6.41$ Hz, 1H), 4.77 (t, $J = 5.70$ Hz, 1H), 4.36 (dd, $J = 2.96, 2.17$ Hz, 1H), 4.17 (q, $J = 2.82$ Hz, 1H), 4.01 (s, 3H), 3.93-3.90 (dd, $J = 12.42, 2.82$ Hz, 1H), 3.80-3.77 (dd, $J = 12.42, 2.82$ Hz, 1H).

^{13}C -NMR (100 MHz, MeOD) δ : 159.6, 152.3, 150.5, 140.6, 139.2, 132.4, 123.2, 118.3, 88.7, 85.8, 73.2, 70.7, 61.6, 37.4.

HRMS (LC-MS) m/z : $[M+H]^+$ calcd for $C_{14}H_{18}N_7O_4$ 348.1420, found 348.1432.

(2R,3R,4S,5R)-2-(2-amino-6-(3,5-dimethylisoxazol-4-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol 183



183 was prepared following GP4 using **178** (74 mg, 0.15 mmol), K_2CO_3 (6 mg, 0.04 mmol) and MeOH (1.5 mL). Reaction time was 16 hours. The mixture was then adsorbed into SiO_2 and loaded on a column. Purification was carried out

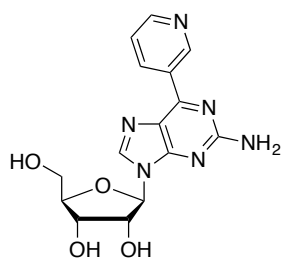
eluting with MeOH/CHCl₃ (0:1 → 2:8) affording the title compound as an off-white foam (41 mg, 50% over two steps).

¹H-NMR (500 MHz, MeOD) δ: 8.30 (s, 1H), 6.00 (d, *J* = 6.0 Hz, 1H), 4.78 (t, *J* = 5.5 Hz, 1H), 4.39 (dd, *J* = 3.0, 1.9 Hz, 1H), 4.18 (q, *J* = 2.8 Hz, 1H), 3.92 (dd, *J* = 12.4, 2.6 Hz, 1H), 3.8 (dd, *J* = 12.3, 2.8 Hz, 1H), 2.58 (s, 3H), 2.43 (s, 3H)

¹³C-NMR (125 MHz, MeOD) δ: 170.06, 160.28, 160.24, 159.5, 153.3, 150.4, 141.6, 125.6, 112.4, 89.0, 86.2, 73.8, 71.1, 61.9, 11.4, 9.9.

HRMS (CI) *m/z*: [M+H]⁺ calcd for C₁₅H₁₉N₆O₅ 363.1417, found 363.1409.

(2*R*,3*R*,4*S*,5*R*)-2-(2-amino-6-(10*H*pyridine-3-yl)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **184**



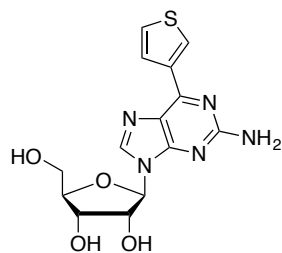
184 was prepared following GP4 using **179** (51 mg, 0.11 mmol), K₂CO₃ (5 mg, 0.03 mmol) and MeOH (1.1 mL). Reaction time was 30 min. The resulting red precipitate was filtered, washed with cold MeOH (3 x 5 mL) and dried to afford (**150b**) as a red solid. (30 mg, 38% over two steps).

¹H-NMR (400 MHz, DMSO-*d*₆) δ: 9.8 (dd, *J* = 2.2, 0.6 Hz, 1H), 8.92 (dt, *J* = 8.0, 2.0 Hz, 1H), 8.7 (dd, *J* = 4.8, 1.8 Hz, 1H), 8.44 (s, 1H), 7.59 (dd, *J* = 8.1, 0.7 Hz, 1H), 6.67 (brs, 2H), 5.9 (d, *J* = 5.9 Hz, 1H), 5.45 (brs, 1H), 5.16 (brs, 1H), 5.05 (t, *J* = 5.5 Hz, 1H), 4.54 (t, *J* = 5.2 Hz, 1H), 4.15 (brs, 1H), 3.92 (q, *J* = 3.9 Hz, 1H), 3.66 (dt, *J* = 11.8, 4.6 Hz, 1H), 3.56 (dt, *J* = 12.0, 4.8 Hz, 1H).

¹³C-NMR (100 MHz, DMSO-*d*₆) δ: 160.1, 154.5, 151.7, 151.0, 149.9, 141.2, 136.2, 131.4, 124.3, 123.7, 86.4, 85.3, 73.4, 70.3, 61.3.

HRMS (FAB) *m/z*: [M+H]⁺ calcd for C₁₅H₁₆N₆O₄ 344.1233, found 344.1224.

(2*R*,3*R*,4*S*,5*R*)-2-(2-amino-6-(thiophen-3-yl)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **185**



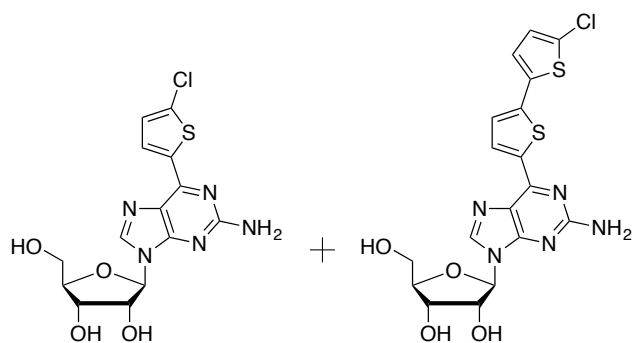
180 was prepared following GP4 using **185** (61 mg, 0.14 mmol), K₂CO₃ (5 mg, 0.04 mmol) and MeOH (1.3 mL). Reaction time was 16 hours. The mixture was then adsorbed into SiO₂ and loaded on a column. Purification was carried out eluting with MeOH/CHCl₃ (0:1 → 2:8) affording the title compound as a yellowish foam (35 mg, 77%).

¹H-NMR (500 MHz, MeOD) δ: 8.72 (dd, *J* = 3.3, 1.1 Hz, 1H), 8.23 (s, 1H), 8.09 (dd, *J* = 5.1, 1.1 Hz, 1H), 7.52 (dd, *J* = 5.1, 3.0 Hz, 1H), 5.96 (d, *J* = 6.2 Hz, 1H), 4.78 (t, *J* = 5.6 Hz, 1H), 4.37 (dd, *J* = 5.2, 3.0 Hz, 1H), 4.18 (q, *J* = 2.8 Hz, 1H), 3.92 (dd, *J* = 12.4, 2.7 Hz, 1H), 3.79 (dd, *J* = 12.3, 3.3 Hz, 1H).

^{13}C -NMR (100 MHz, MeOD) δ : 159.5, 152.8, 151.4, 140.9, 137.4, 129.4, 126.9, 124.8, 123.8, 88.7, 85.7, 73.2, 70.7, 61.5.

HRMS (EI) m/z : $[M]^+$ calcd for $\text{C}_{14}\text{H}_{15}\text{ClN}_5\text{O}_4\text{S}$ 349.0845, found 349.841 (100%), 385.0575 (21%), 350.0903 (20%).

(2R,3R,4S,5R)-2-(2-amino-6-(5-chlorothiophen-2-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **186 and (2R,3R,4S,5R)-2-(2-amino-6-(5'-chloro-[2,2'-bithiophen]-5-yl)-9H-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diol **187****



186 was prepared following GP4 using **181** (46 mg, 0.09 mmol), K_2CO_3 (4 mg, 0.03 mmol) and MeOH (0.9 mL). Reaction time was 16 hours. The mixture was then adsorbed into SiO_2 and loaded on a column. Purification was carried out eluting with MeOH/ CHCl_3 (0:1 \rightarrow 1:9) affording the title compound as a bright yellow foam (25 mg, 72%),

which was contaminated with 6% with deprotected side-product **187**. Only the peaks corresponding to **186** are reported. The peaks corresponding to **187** are shown in Appendix 90.

^1H -NMR (500 MHz, DMSO- d_6) δ : 8.41 (s, 1H), 8.34 (d, $J = 4.0$ Hz, 1H), 7.30 (d, $J = 4.0$ Hz, 1H), 6.60 (s, 2H), 5.86 (d, $J = 5.8$ Hz, 1H), 5.49 (brs, 1H), 5.2 (brs, 1H), 5.09 (brs, 1H), 4.52 (t, $J = 5.4$ Hz, 1H), 4.15 (t, $J = 4.4$ Hz, 1H), 3.93 (q, $J = 3.7$ Hz, 1H), 3.66 (d, $J = 12.4$ Hz, 1H), 3.56 (d, $J = 12.0$ Hz, 1H).

^{13}C -NMR (100 MHz, DMSO- d_6) δ : 159.9, 154.16, 148.1, 141.1, 139.4, 133.0, 131.5, 128.6, 122.1, 86.4, 85.3, 73.5, 70.3, 61.3.

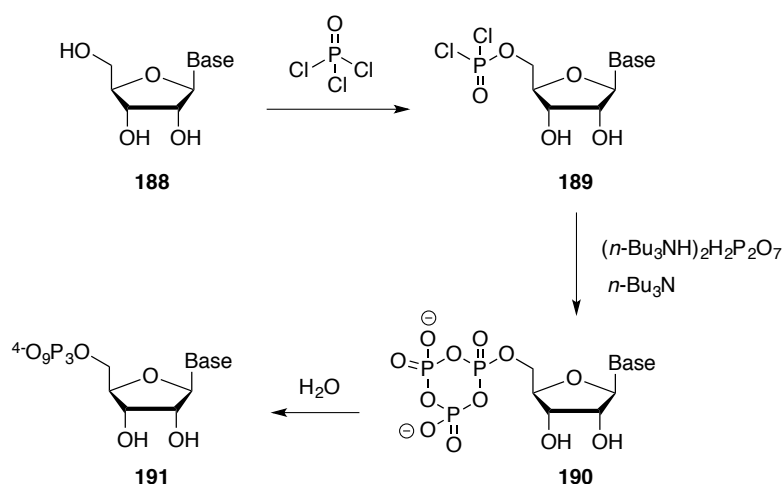
HRMS (CI) m/z : $[M+H]^+$ calcd for $\text{C}_{14}\text{H}_{15}\text{ClN}_5\text{O}_4\text{S}$ 384.0533, found 384.0539 (100%), 385.0575 (21%), 386.0543 (39%).

Chapter 4

4 Synthesis of C6-functionalized 2-aminopurine Nucleoside Triphosphates

4.1 Introduction

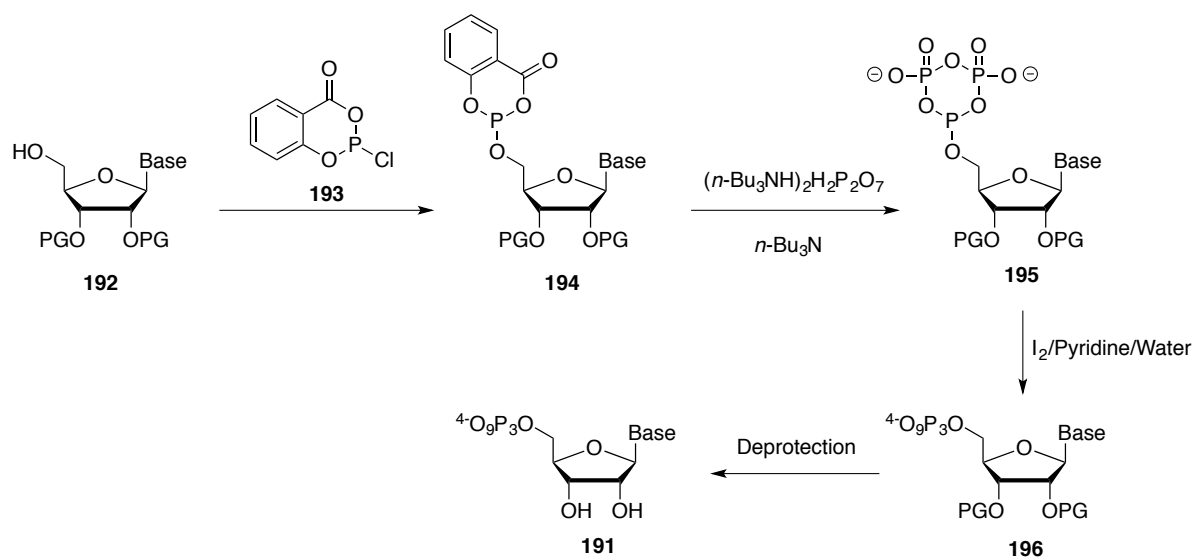
A primary aim of this project was the development of nucleoside analogues that could be assayed for *in vitro* transcription activity. As transcription requires 5' nucleoside triphosphate building blocks, it was necessary to phosphorylate the nucleosides prior to any biochemical applications. There are several established methods for synthesizing nucleoside triphosphates.^{129,193} One of the most common methods was developed in 1967 by Yoshikawa and co-workers.¹⁹⁴ In this protocol nucleoside **188** is treated with POCl₃ under basic conditions to generate phosphochloridate **189**. This intermediate is treated with a pyrophosphate salt and tributylamine furnishing **190**, which is hydrolyzed to yield triphosphate **191** (Scheme 4-1). The advantage of this method is its simplicity (three steps, one-pot procedure) and its reported selectivity towards the 5'-OH.¹⁹³ However, subsequent studies show that this reaction delivers both a poor regioselectivity and a variety of by-products, including 5'/3'-diphosphates and 3'-monophosphates.¹⁹⁵ Moreover, there have been conflicting findings concerning the role of reaction pH in regulating the 5' regioselectivity; e.g. various groups have reported the need for either acidic or basic conditions to obtain the desired product.¹⁹³



Scheme 4-1. The Yoshikawa triphosphate synthesis.¹⁹⁴

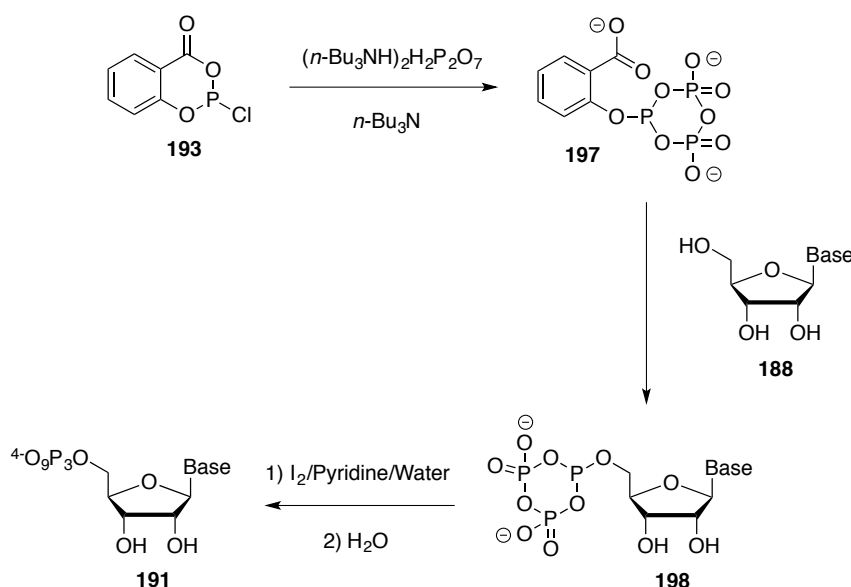
Another “three steps, one-pot” procedure commonly employed for triphosphate synthesis was developed in 1989 by Ludwig and Eckstein.¹⁹⁶ In this method salicyl chlorophosphite **193** reacts with nucleoside **192** generating intermediate **194**, which is then treated with tributylammonium pyrophosphate, furnishing cyclic intermediate **195** (Scheme 4-2). P(III) to P(V) oxidation and hydrolysis mediated by an iodine/pyridine/water mixture furnishes triphosphate **196**. Finally, cleavage of the protecting groups produces triphosphate **191**.

The disadvantage of this method is the required protection of the 2'/3'-OH moieties to avoid their phosphorylation, thus introducing extra synthetic steps.¹²⁹



Scheme 4-2. The Ludwig-Eckstein triphosphate synthesis.¹⁹⁶ PG = Protecting Group.

Recently, Huang and co-workers developed a modified Ludwig-Eckstein synthesis for the selective 5'-triphosphorylation of unprotected nucleosides.¹⁹⁷ According to the authors, bulky reagent **197** would selectively react with the 5'-OH of **188**, furnishing intermediate **198**, which is then oxidized and hydrolyzed to generate triphosphate **191**. If applicable, this method will be invaluable for this work as the robustness of the strategy would be bolstered by preparing pure triphosphates without the need for 2'/3' protection.



Scheme 4-3. The Huang triphosphate synthesis.¹⁹⁷

4.2 Aims of this Study

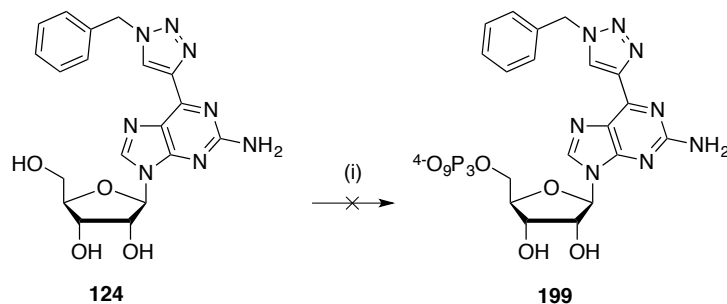
The aims of this chapter are to:

- 1) Synthesize C6-functionalized 2-aminopurine nucleoside triphosphates suitable for transcription assays.

4.3 Results and Discussion

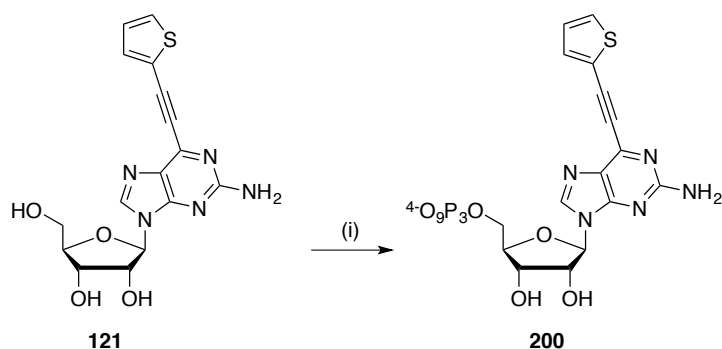
4.3.1 Initial Attempts to Synthesize Nucleoside Triphosphates

Due to its reported 5'-OH selectivity, the Huang method was explored initially.¹⁹⁷ Unfortunately, using **124**, only starting material was observed upon analysis of the crude reaction mixture by TLC and ¹H-NMR (Scheme 4-4). As there was no literature precedent regarding the utilization of the Huang method for the synthesis of artificial nucleoside triphosphates the possibility that the Huang conditions might only be applicable for natural nucleosides was considered. This could suggest that modification of the Huang method is required for use with this substrate class. Indeed, after the successful synthesis of the triphosphate analogues *via* a different method (*vide infra*), it was discovered that Balzarini and co-workers attempted to employ this method to synthesize ATP analogues.¹⁹⁸ The authors reported that the addition of 2.0 equivalents of triethylamine was necessary for this method to yield the desired triphosphates.



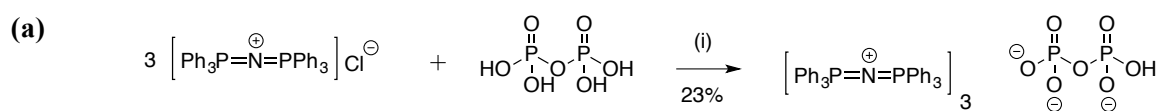
Scheme 4-4. *Reagents and conditions:* Salicyl chlorophosphite (1.2 equiv.), tetrabutylammonium pyrophosphate (2.0 equiv.), tributylamine, DMF, 1 h.; then compound **124** (1.0 equiv.), 0 °C, 3 h. I₂/py/water r.t.; water (excess), 1.5 h.

Next, the Yoshikawa conditions were explored.¹⁹⁴ Using nucleoside **121**, triphosphate **200** was successfully obtained following HPLC purification, although ³¹P-NMR analysis revealed that it contained 20% of the corresponding 5'-monophosphate (Scheme 4-5).

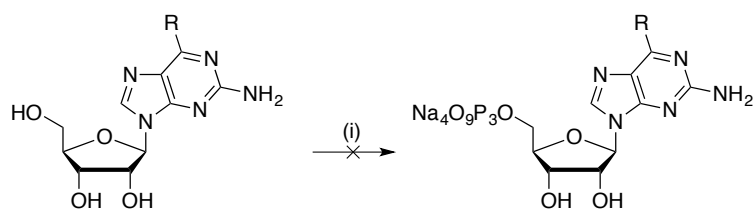


Scheme 4-5. *Reagents and conditions:* Triethylphosphate, POCl₃ (1.3 equiv.), Proton sponge® (1.5 equiv.), 0 °C, 2h.; then tributylamine (5.0 equiv.), tributylammonium pyrophosphate (5.0 equiv.), r.t., 10 min.

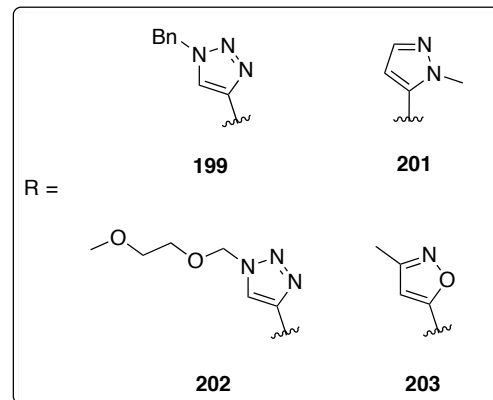
This residual intermediate likely results from the use of the hygroscopic and unstable tributylammonium pyrophosphate.¹⁹⁹ Recently, Hodgson *et al.* published the synthesis and application of a new pyrophosphate reagent (PPN pyrophosphate) in the Yoshikawa procedure.²⁰⁰ PPN pyrophosphate is prepared *via* a simple salt exchange, is non-hygroscopic, and is stable to both air and moisture (Scheme 4-6a). Using this reagent, triphosphates **199** and **201-203** were isolated following RP-HPLC purification (Scheme 4-6b).



(b)



R = Bn-triazole, **124**
 R = MEM-triazole, **126**
 R = Me-isoxazole, **154**
 R = 5-pyrazole, **169**



Scheme 4-6. (a) *Reagents and conditions:* (i) [bis(triphenylphosphoranylidene)ammonium] chloride (2.0 equiv.), H₂O, r.t., 1 min., 23%. (b) *Reagents and conditions:* (i) Triethylphosphate, POCl₃ (1.5 equiv.), 2,4,6-collidine (1.0 equiv.), -10 °C, 2 h.; then tris{[bis(triphenylphosphoranylidene)ammonium]} pyrophosphate (2.0 equiv. 0.4 M solution in MeCN), tri-n-octylamine, 0 °C, 2 h.; r.t., 20 h.

Although the compounds were found to have the correct *m/z* *via* MALDI-TOF MS, ³¹P-NMR analysis revealed the presence of various extra peaks (Figure 4-1, ³¹P-NMR of **199** is shown).

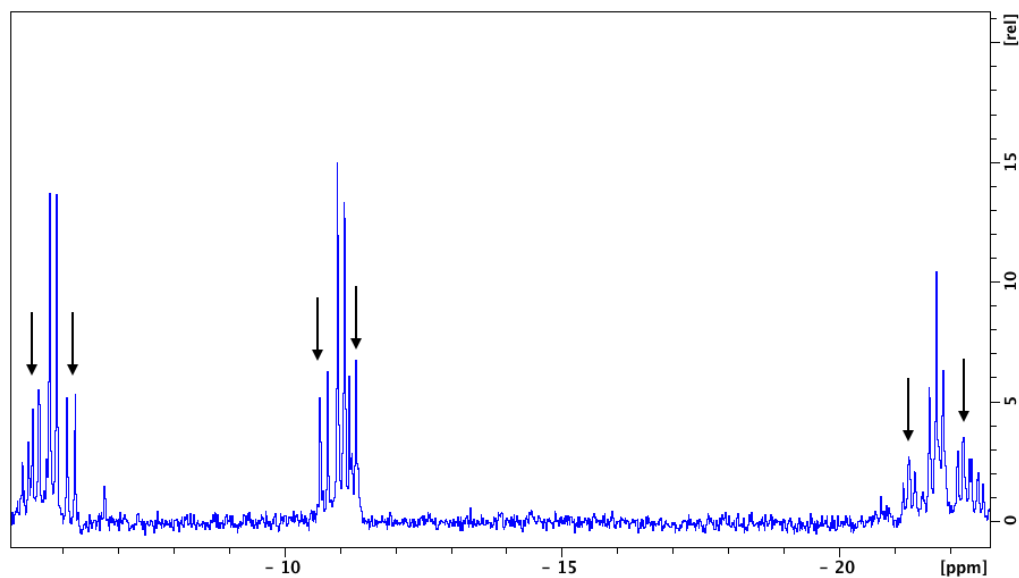
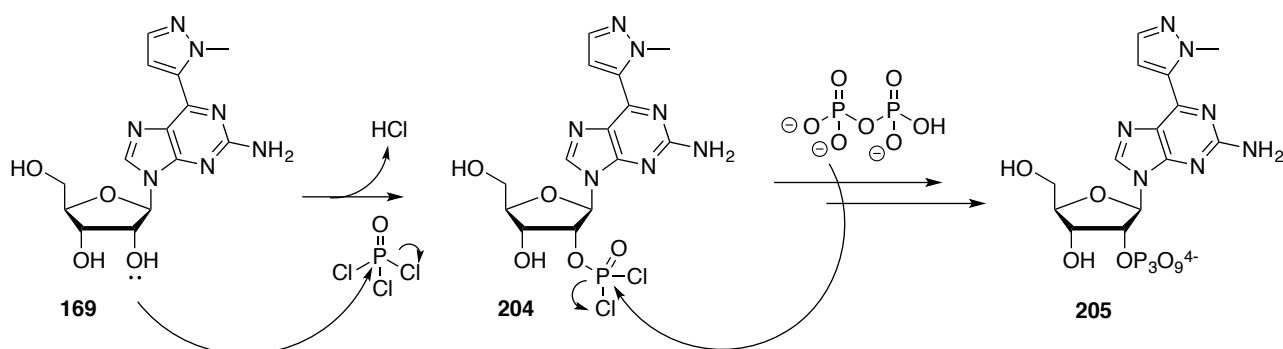


Figure 4-1. ^{31}P -NMR spectrum of **199**. Arrows indicate the presumed extra triphosphate peaks.

For each major peak, extra sets of peaks were observed, composed of two doublets and a triplet. Since this pattern is characteristic for triphosphate functionalities, this suggested the presence of triphosphates on alternate positions of the sugar (i.e. 2' and 3'). This is a common problem when NTP synthesis is attempted on completely free nucleosides. These byproducts can form through the POCl_3 -mediated phosphorylation of the free 2' and/or 3' sugar hydroxyls, which then reacts with the pyrophosphate reagent forming 2' and/or 3' triphosphates. Scheme 4-7 shows an example of triphosphorylation of the 2'-OH instead of the desired 5'-OH.



Scheme 4-7. Presumable mechanism of formation of impurity **205**. Cations have been omitted for clarity.

A sample of triphosphate **201** was analyzed by RP-HPLC using a slow gradient to resolve close-running or overlapping peaks. While the previously single isolated peak appears to begin resolving into two peaks, there was still a great extent of overlap and further purification by RP-HPLC was deemed unfeasible (Figure 8-4). Attempted NTP synthesis with the other free nucleosides that have been synthesized in Chapter 3 under Hodgson's conditions also yielded mixtures that could not be purified by RP-HPLC. Moreover, ^{31}P -NMR analysis of these compounds revealed complex mixtures containing only a small amount of the nucleoside triphosphate. Consequently, this strategy was discontinued.

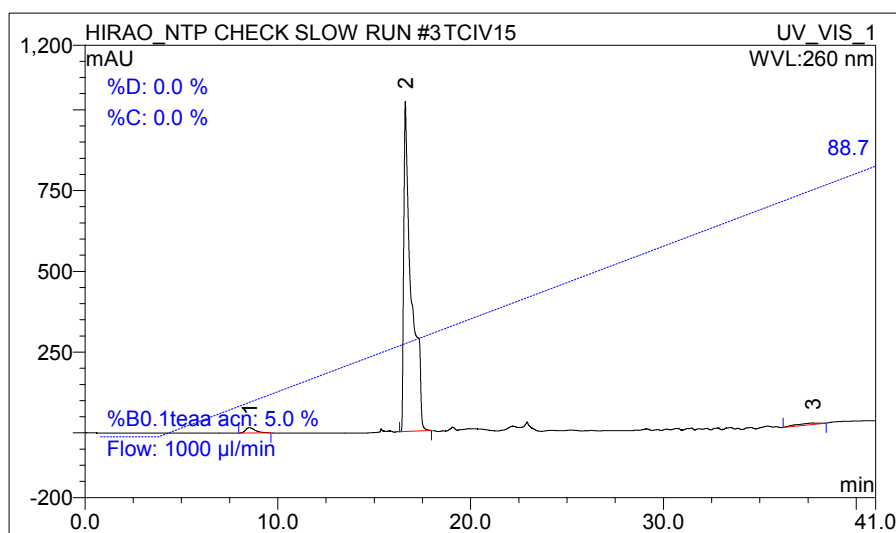
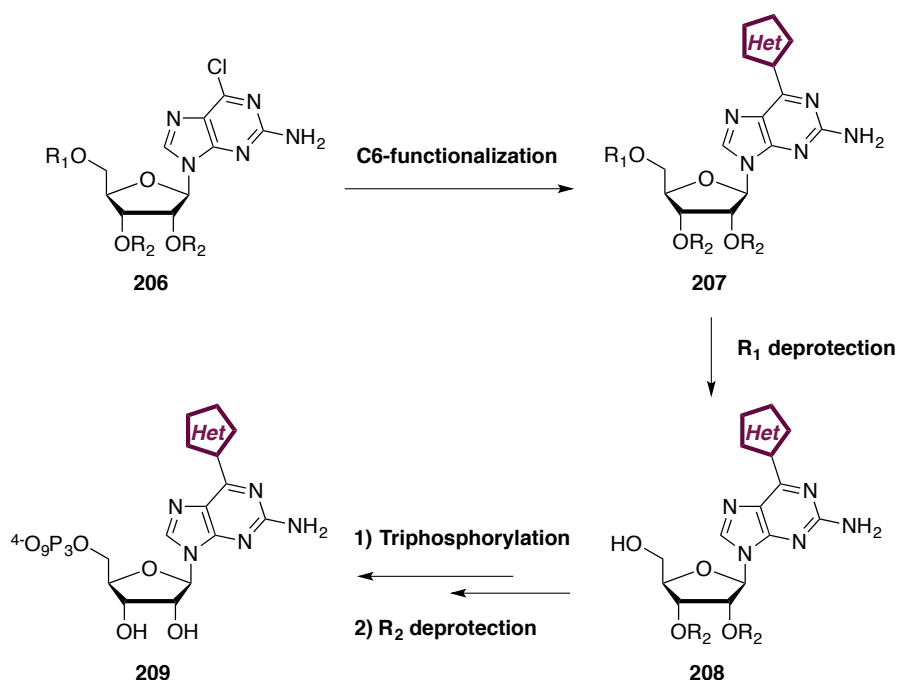


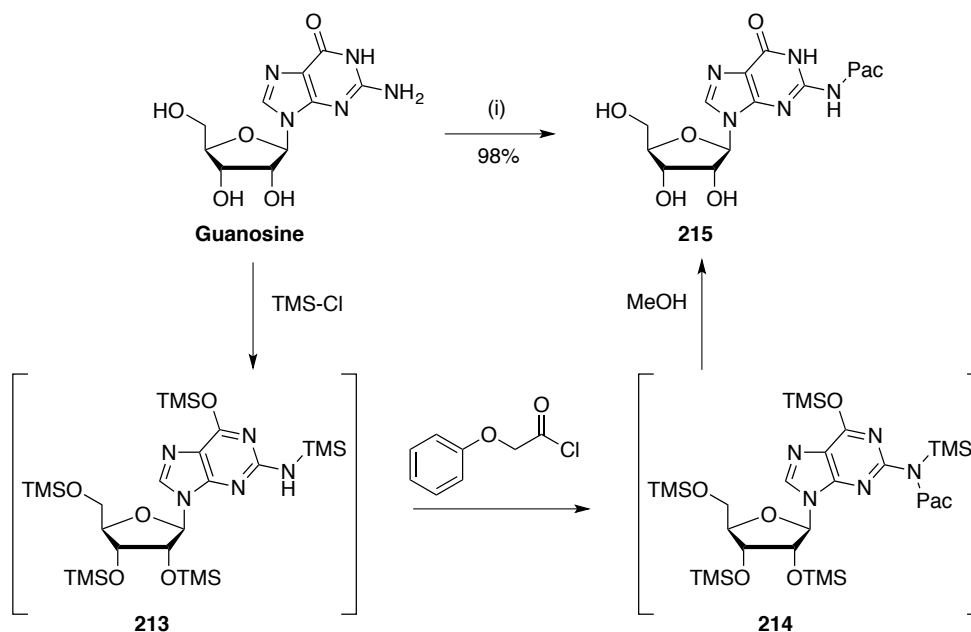
Figure 4-2. Analytical HPLC chromatogram of **201** using a slow-gradient method.

4.3.2 Strategy to Access Pure Nucleoside Triphosphates

Based on the assumption that the NTP mixtures result from phosphorylation of the free 2'/3' hydroxyls, it was decided to develop a novel route in order to access the desired triphosphates. It was envisaged that nucleoside **206**, bearing a 5' hydroxyl protecting group distinct from that at the 2'/3' positions, could be prepared and could function as the starting point for our NTPs. Following C6-functionalization, selective 5' deprotection of **207** would yield **208**, which could then be triphosphorylated solely at the 5'-OH. Subsequent deprotection of the 2'/3' protecting groups would then yield the desired triphosphate **209** (Scheme 4-8).

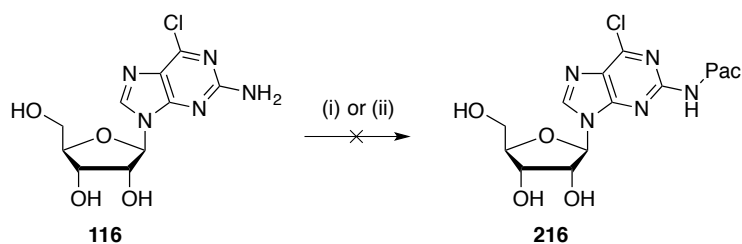


Scheme 4-8. Proposed strategy to access pure triphosphates. Cations have been omitted for clarity.



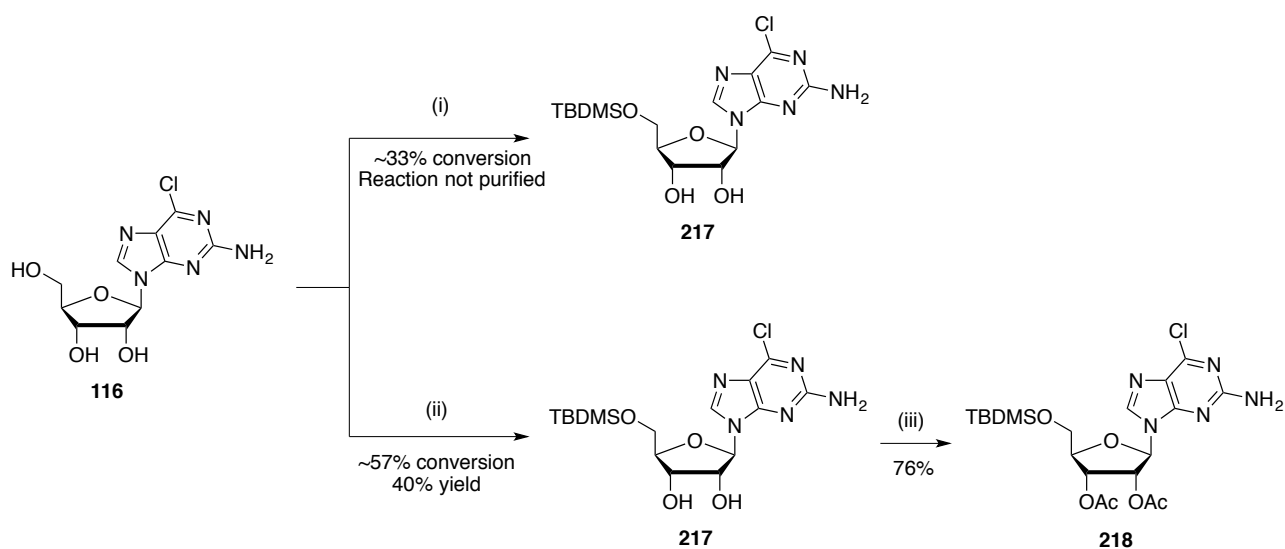
Scheme 4-10. Reagents and conditions: (i) TMS-Cl (9.0 equiv.), pyridine/DCM (1:4), 0 °C to r.t., 2 h; Pac-Cl (1.1 equiv.), 0 °C, 2 h; MeOH, r.t., 16 h.²⁰²

Using this method with chloride **116** produced a dark mixture and, following methanolysis, formation of a precipitate. Both the precipitate and filtrate were isolated and analyzed by ¹H-NMR. Unfortunately, neither of these collections contained the expected product and the spectra displayed only a complicated mixture of various unidentifiable peaks. A similar method reported by Chaix et al., used MeCN as a solvent and HOBT as an additive.²⁰³ Again, a dark mixture was observed along with a precipitate and ¹H-NMR analysis of both filtrate and precipitate failed to reveal peaks corresponding to **216**. Since unlike the C6 carbonyl of guanosine (Scheme 4-10) the C6 chloride group cannot be silylated, it was presumed that its electron-withdrawing nature would reduce the nucleophilicity of the exocyclic amine, thus suppressing the reaction with Pac-Cl. Subsequent hydrolysis of the remaining TMS-Cl and Pac-Cl with methanol would form HCl, which due to the prolonged reaction time, would result in nucleoside degradation, thus explaining the dark mixtures obtained. Increasing the reaction temperature and/or increasing the amount of the acylating agent could potentially circumvent the problem of the amine nucleophilicity, while careful monitoring of the methanolysis step could avoid the problem of nucleoside degradation. However, it was decided that the Pac protection step could simply be delayed until after the installation of the sugar protecting groups, thus avoiding potentially time-consuming optimization attempts.



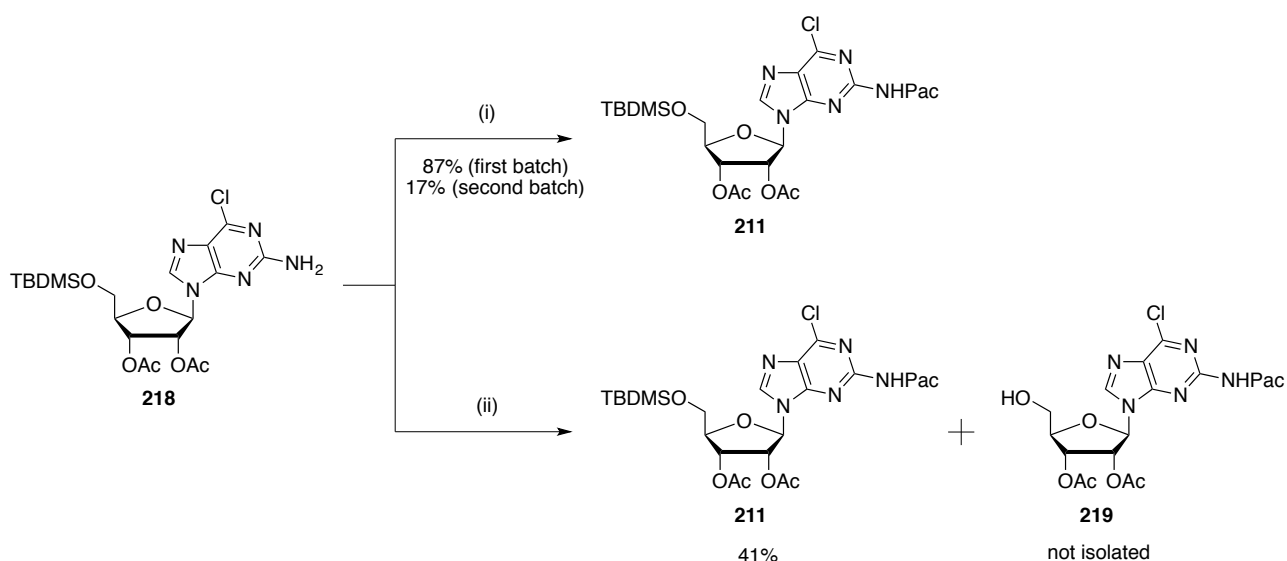
Scheme 4-11. *Reagents and conditions:* (i) TMS-Cl (9.0 equiv.), pyridine/DCM (1:4), 0 °C to r.t., 2 h; Pac-Cl (1.1 equiv.), 0 °C, 2 h; MeOH, r.t., 16 h, 98%. (ii) TMS-Cl (9.0 equiv.), HOBt (1.8 equiv.), Pac-Cl (1.5 equiv.), pyridine/MeCN (1:1), 0 °C to r.t., 16 h; then MeOH, r.t., 16 h.

Beginning with commercially available **116**, TBDMS protection of the 5'-OH was attempted using a combination of TBDMS-Cl, imidazole and DMF. Unfortunately, this method was deemed inadequate as only a ~33% conversion was observed via LC-MS (Scheme 4-12). Moreover, the reaction could not be improved either by increasing the reaction temperature to 60 °C or through the addition of an extra equivalent of TBDMS-Cl. An alternative TBDMS protection reported by Schinazi et al. suggested that DMAP/Et₃N could replace imidazole as the base and DMSO/DCM rather than DMF could be used as the solvent.²⁰⁴ The rationale behind employing this specific set of conditions has not been disclosed by the authors, with the exception of DMSO, which was used due to the low solubility of guanosine in other solvents.²⁰⁴ This method resulted in 56% conversion after 3 hours (determined by LC-MS). Increasing the reaction time to 16 hours did not improve the conversion and **217** was obtained in 40% yield. Despite the modest yield, an adequate amount of **217** could be obtained through this route to allow for the progression of the synthetic plan. Acetylation of the 2'/3' hydroxyls was then carried out following the same protocol described in Section 3.3.1, furnishing **218** in 76% yield (Scheme 4-12).¹⁶⁰



Scheme 4-12. *Reagents and conditions:* (i) TBDMS-Cl (1.3 equiv.), imidazole (2.0 equiv.), DMF, r.t. or 60 °C, 16 h. (not isolated) (ii) TBDMS-Cl (2.0 equiv.), Et₃N (2.2 equiv.), DMAP (0.1 equiv.), DMSO/DCM (1:4), r.t., 3 h. (iii) Acetic anhydride (2.1 equiv.), Et₃N (4.0 equiv.), DMAP (0.1 equiv.), MeCN, 0 °C to r.t., 30 min.

With **218** in hand, conditions to install the phenoxyacetyl group were explored. As described in the previous attempts to install the Pac group (see Scheme 4-10), literature protocols employ pyridine as the base for this reaction. Initially, **218** was treated with phenoxyacetyl chloride in neat pyridine.²⁰⁵ 112 mg of amide **211** were obtained through this method, but ¹H-NMR analysis revealed contamination with unidentified impurities in the aromatic region. Scaling up the reaction from 100 mg to a 600 mg scale did allow for the isolation of pure **211**, albeit in 17% yield (Scheme 4-13). During the reaction, formation of an unidentified yellow precipitate following the addition of the acyl chloride was observed. Consequently, it was decided to investigate the importance of the precipitate's solubility. Since MeCN has been used as a solvent for Pac protection of guanosine (see Scheme 4-11), the reaction was repeated with MeCN as a solvent and 1.3 equivalents of pyridine as the base. Interestingly, no precipitate was formed upon acyl chloride addition and **211** was obtained in 41% yield (Scheme 4-13). However, TBDMS-cleaved side-product **219** and an unidentified side-product were observed during LC-MS analysis, although neither of these compounds were isolated.

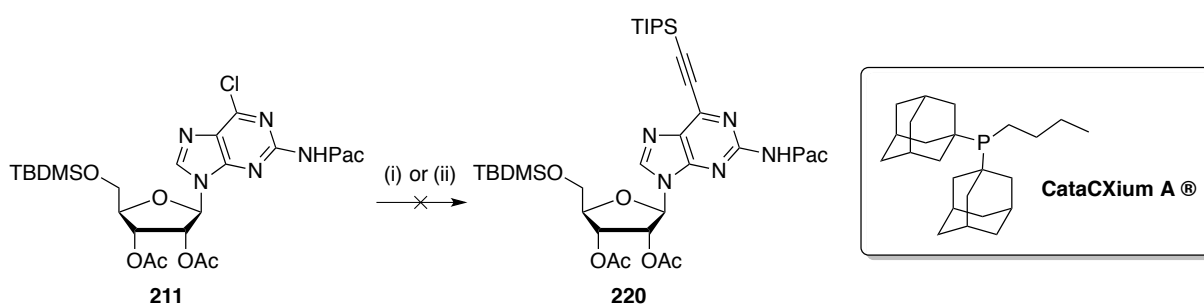


Scheme 4-13. *Reagents and conditions:* (i) Pac-Cl (1.2 equiv.), pyridine, r.t., 20 min. (ii) Pac-Cl (1.2 equiv.), pyridine (1.3 equiv.), r.t., MeCN, 16 h.

A sufficient amount of **211** was obtained in order to continue on to a Sonogashira reaction with (triisopropylsilyl)acetylene; however, there are a few ways in which this reaction could be optimized for future work. The formation of side product **219** could be a result of using only 1.3 equivalents of pyridine. Since 1.0 equivalent of acidic pyridinium chloride is generated during the reaction, only 0.3 equivalents of pyridine remains to buffer the reaction pH. Plans for further optimization of this reaction include using a larger amount of pyridine (e.g. 10 equivalents) and attempting the reaction in neat pyridine with the respective anhydride instead of the acid chloride.

To generate **220**, the Sonogashira conditions described in Section 3.3.1 were employed. After stirring overnight at room temperature, 50% conversion was observed upon LC-MS analysis (Scheme 4-14).

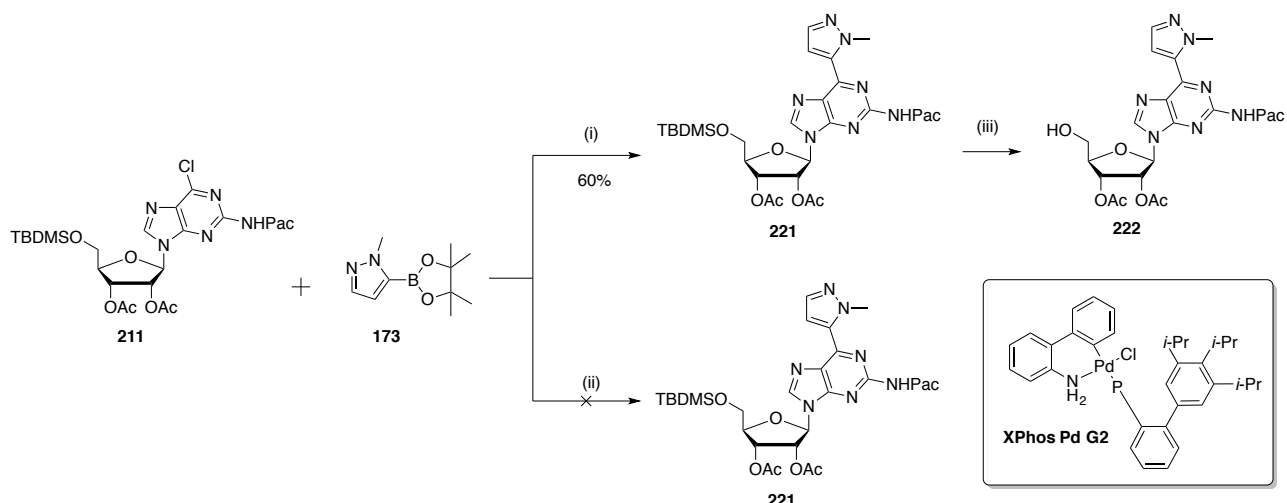
Increasing the temperature to 80 °C did not improve the conversion and this method was abandoned. A combination of Pd(OAc)₂/CataCXium A[®]/diisopropylamine was suggested by colleagues at GlaxoSmithKline (GSK). After overnight stirring at room temperature, only 3% conversion was observed *via* LC-MS. Once again, increasing the reaction temperature to 80 °C did not improve the conversion and this method was discontinued as well (Scheme 4-14). This is not surprising, since similar results were observed when chloride **115** was employed as the halide coupling partner (see Scheme 3-1). It was concluded that the respective iodide should be synthesized and employed in the Sonogashira coupling. However, these efforts were not initiated immediately in order to explore the installation of the pyrazole moiety.



Scheme 4-14. Reagents and conditions: (i) [Pd(PPh₃)₂]Cl₂ (2 mol%), CuI (6 mol%), DIPEA (3.0 equiv.), (triisopropylsilyl)acetylene (1.2 equiv.), DMF, r.t., 16 h; 60 °C, 5 h. (ii) Pd(OAc)₂ (2 mol%), CataCXium A[®] (4 mol%), CuI (6 mol%), Diisopropylamine, r.t., 16 h; 80 °C, 16 h.

4.3.2.3 Synthesis of C6-pyrazole 2-aminopurine 222

In parallel with the above Sonogashira reactions, the Suzuki protocol described in Section 3.3.5 was applied to chloride **211**. In Section 3.3.5, solubility issues with the starting chloride necessitated the use of DMSO. Fortunately, the increased solubility of chloride **211** allowed the reaction to be conducted in 2-MeTHF and, after 4 hours, pyrazole **221** was obtained in good yield (61%). In an attempt to optimize the reaction, a combination of K₃PO₄/XPhos Pd G2/1,4-dioxane was investigated as suggested by colleagues at GSK, but after 2 hours at 80 °C, LC-MS analysis indicated formation of various unidentified side-products along with 50% of unreacted **211** (Scheme 4-15). Consequently, this method was abandoned in favor of the previous protocol. TBAF-mediated deprotection of the TBDMS group afforded alcohol **222** in 91% yield. Due to the previous observation that a solution of TBAF in THF can potentially cleave acetate esters groups (Table 3-1), the TBAF solution in THF was first buffered with 1.0 equivalent of acetic acid.

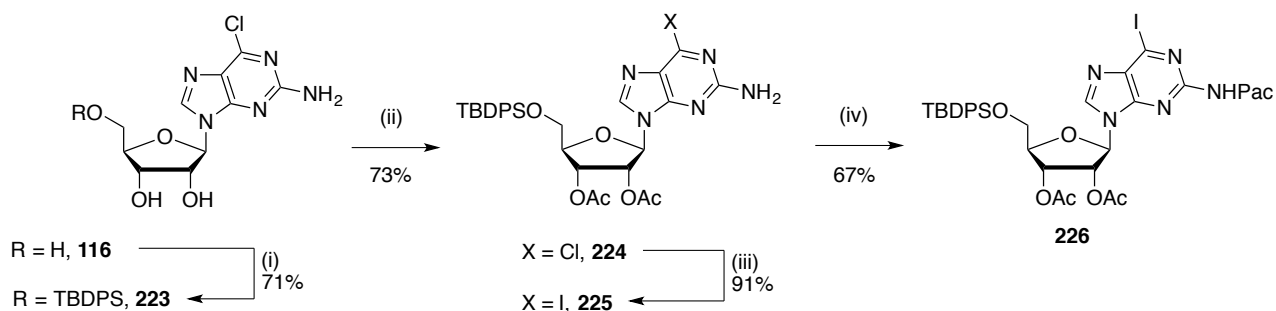


Scheme 4-15. Reagents and conditions: (i) Pd-132 (0.1 equiv.), **173** (1.3 equiv.), KF on alumina (4.4 equiv.), 2-MeTHF, 80 °C, 4 h. (ii) **173** (1.3 equiv.), K₃PO₄ (2.0 equiv.), XPhos Pd G2 (5 mol%), 1,4-dioxane, 80 °C, 2 h. (iii) TBAF (1.0 M in THF, 1.1 equiv.), AcOH (1.1 equiv.), THF, r.t., 90 min.

4.3.2.4 Synthesis of C6-alkynyl 2-aminopurine **210**

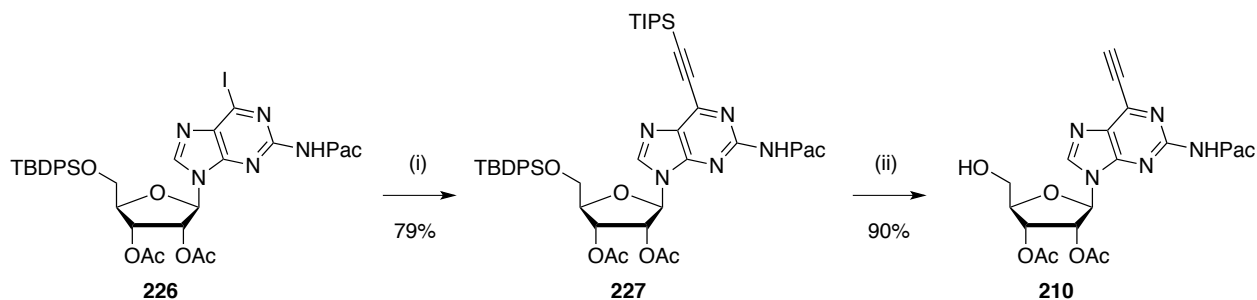
Despite the success with the Suzuki coupling and TBDMS deprotection, isolation of precursor **217** was proving problematic and was limiting the progress of the synthetic route. Chromatographic purification of **217** failed to yield pure compound regardless of the eluent system employed. Crystallization conditions utilising a range of one or two solvent systems were explored including: DCM, toluene, acetone, toluene/EtOAc, toluene/acetone; yet, none proved successful. One cause of these failed purifications could be traced to residual DMSO. The complete removal of DMSO became inefficient during scale-up as the volume of DMSO increased in proportion to the reagents. As a differentially protected 5'-OH precursor is essential for our strategy, alternative 5'-OH protecting groups were investigated.

Another common silyl ether used in protecting group chemistry is the TBDPS ether.²⁰¹ TBDPS groups are more stable than TBDMS groups in acidic conditions, which could prevent its cleavage during the Pac protection step (see Scheme 4-13). Employing a combination of TBDPS-Cl/imidazole/DMAP in DMF afforded silyl ether **223** in 71% yield. Acetyl protection of the remaining hydroxyls using the conditions described in Scheme 4-12 afforded diacetate **224** in 73% yield. Since attempts to install the ethynyl moiety using chloride **211** were unsuccessful (see Scheme 4-11), the respective iodide was decided to be employed, as described in Scheme 3-2. Iodination of **224** was carried out under the conditions described in Scheme 3-3, affording iodide **225** in 91% yield. Pac protection was attempted using acetonitrile as a solvent, as described in Scheme 4-13, but with 3.0 equivalents of pyridine rather than 1.3. Experiments conducted by other members of the Burley group suggested that cooling down the reaction to 0 °C before chloride addition could improve the yield. Additionally, the acyl chloride was added slowly over a period of ten minutes. These modifications afforded **226** in 67% yield.



Scheme 4-16. *Reagents and conditions:* (i) TBDPS-Cl (2.1 equiv.), imidazole (2.5 equiv.), DMAP (0.1 equiv.), DMF, r.t., 3 h. (ii) Acetic anhydride (2.1 equiv.), Et₃N (4.0 equiv.), DMAP (0.1 equiv.), MeCN, 0 °C to r.t., 30 min. (iii) NaI (20.0 equiv.), TFA (2.0 equiv.), 2-butanone, r.t. (iv) Pac-Cl (1.2 equiv.), pyridine (3.0 equiv.), MeCN, 0 °C to r.t., 16 h.

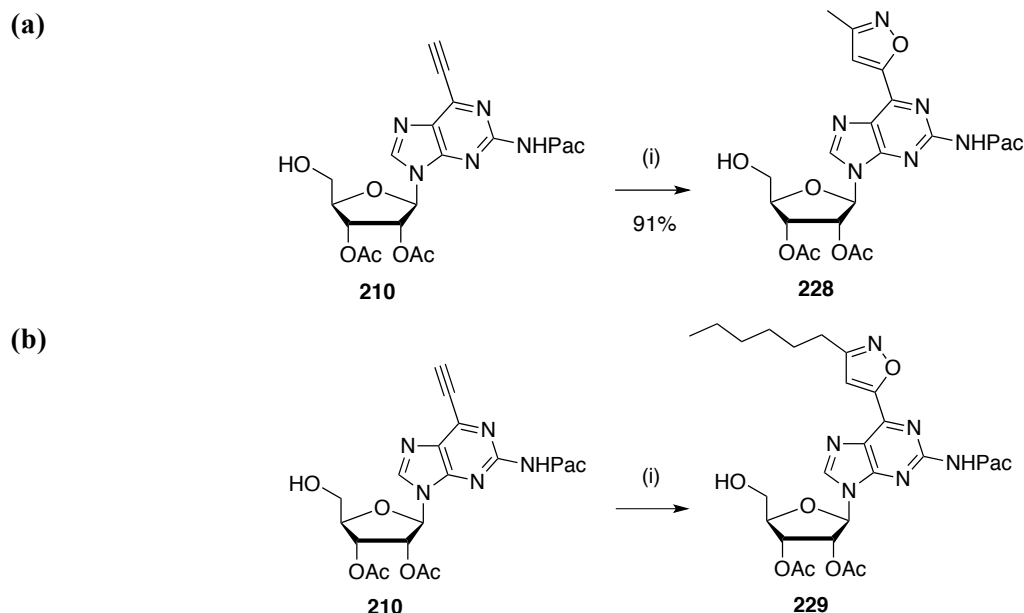
Sonogashira coupling of iodide **226** was then attempted using the standard conditions described in Scheme 3-2. After overnight stirring, 20% conversion was observed by LC-MS. In an attempt to drive the reaction to completion, a further 1.2 equivalents of the alkyne were added and the reaction temperature was increased to 30 °C. Following overnight stirring under these conditions, complete conversion was attained and alkyne **227** was isolated in 79% yield. Cleavage of the silyl protecting group was achieved through treatment with a TBAF/AcOH mixture in THF, affording **210** in 90% yield. With alkyne **210** in hand, the installation of isoxazole moieties on the C6 position was explored.



Scheme 4-17. *Reagents and conditions:* (i) [Pd(PPh₃)₂]Cl₂ (2 mol%), CuI (6 mol%), DIPEA (3.0 equiv.), (triisopropylsilyl)acetylene (1.2 equiv.), DMF, r.t., 16 h; then (triisopropylsilyl)acetylene (1.2 equiv.), 30 °C, 16 h. (ii) TBAF (1.0 M in THF, 2.1 equiv.), AcOH (2.1 equiv.), THF, r.t., 90 min.

4.3.2.5 Synthesis of C6-isoxazole 2-aminopurines **228** and **229**

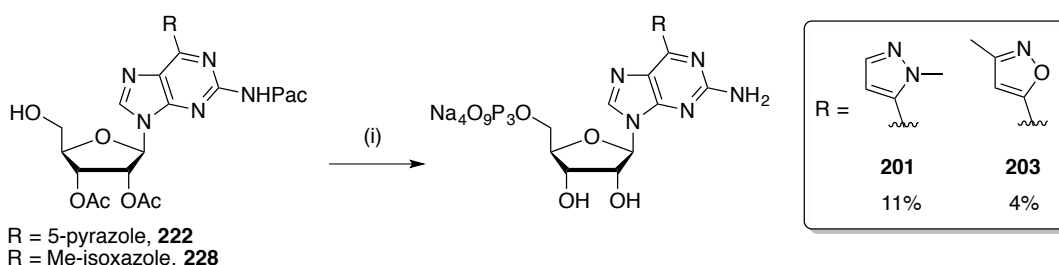
The Hocek method presented in Scheme 3-12 was employed for installation of the isoxazole moiety.¹⁷⁶ Using acetaldoxime, isoxazole **228** was obtained in 91% yield. Although ¹H-NMR analysis revealed a contamination of 7% succinimide, the material was deemed pure enough to be used as a substrate for triphosphate synthesis as the impurity is not expected to interfere with the triphosphorylation reaction (see Scheme 4-19). Likewise, synthesis of the hexyl-isoxazole using heptaldoxime, resulted in 20 mg of a mixture composed of isoxazole **229** and succinimide in a 1:2.7 ratio as determined by ¹H-NMR. In addition, integration of the aliphatic isoxazole peaks of **229** revealed the presence of extra protons, presumably from residual heptaldoxime. Due to these impurities, the focus was shifted on synthesizing triphosphates from the more pure pyrazole **222** and isoxazole **228** and investigating their transcriptional efficiency/fidelity.



Scheme 4-18. (a) *Reagents and conditions:* (i) Acetaldoxime (1.2 equiv.), N-chlorosuccinimide (1.2 equiv.), pyridine (0.16 equiv.), 40 °C, 30 min; then Et₃N (1.2 equiv.), DCM, r.t., 16 h. (b) *Reagents and conditions:* (i) Heptaldoxime (1.2 equiv.), N-chlorosuccinimide (1.2 equiv.), pyridine (0.16 equiv.), r.t., 10 min; then Et₃N (1.2 equiv.), DCM, r.t., 16 h, 1:2 mixture with succinimide.

4.3.2.6 Synthesis of Pyrazole and Isoxazole Functionalized Triphosphates 201 and 203

With pyrazole **222** and isoxazole **228** in hand, the synthesis of the corresponding triphosphates was initiated. To avoid the time-consuming preparation of the pyrophosphate reagent or the preparation of PPN pyrophosphate employed by Hodgson et al., the original Yoshikawa conditions were employed using commercially available tris(tetra-*n*-butylammonium) pyrophosphate.^{194,199,200} Global deprotection was achieved *via* treatment of the crude mixture with 28% NH₄OH at 55 °C for 3 hours. Ion-exchange and subsequent RP-HPLC purification afforded triphosphates **201** and **203** in 11% and 4% yield respectively (Scheme 4-19) in pure form as confirmed by ¹H-NMR and ³¹P-NMR.



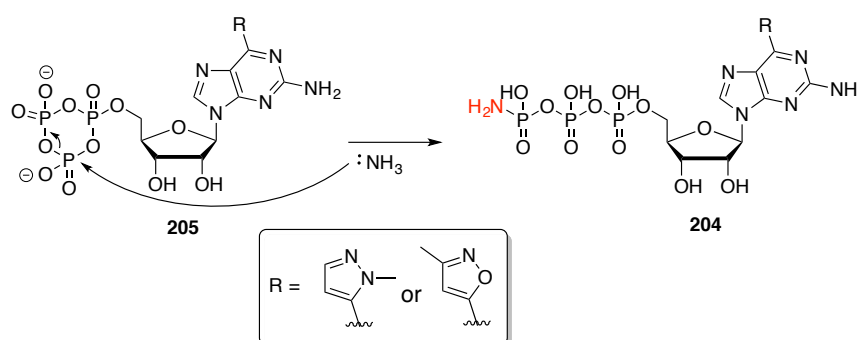
Scheme 4-19. *Reagents and conditions:* (i) Trimethylphosphate, POCl₃ (1.3 equiv.), Proton Sponge® (1.5 equiv.), 0 °C, 2 h; tris(tetrabutylammonium) pyrophosphate (0.5 M in DMF, 5.0 equiv.), tributylamine (5.0 equiv), r.t., 10 min; TEAB (excess), 28% NH₄OH (excess), 55 °C, 3 h.

4.4 Summary

This chapter described the design and execution of a strategy for the synthesis of C6-functionalized triphosphates based on the guanosine scaffold. The synthesis of 5'-OH free nucleosides **222** and **228** was enabled in 5 and 8 steps respectively. Triphosphorylation of these compounds under the standard Yoshikawa

conditions afforded triphosphates **201** and **203** free of any phosphate contaminations. The transcriptional efficiency and fidelity of these compounds was next investigated and the results are presented in the next chapter.

Post-Submission Update: HRMS acquisition of triphosphates **201** and **203** revealed an exact mass that does not correspond to the expected molecular formula. The masses detected correspond to molecular formulas that suggest substitution of an OH group with an NH₂ group (assuming the presence of three phosphorus atoms, which is confirmed by the detection of three signals in ³¹P-NMR). A possible explanation for this outcome is presented below. The last intermediate before quenching the reaction with TEAB is cyclic trimetaphosphate **205**.¹⁹⁴ If its hydrolysis was incomplete then the addition of ammonia could have resulted in competing ammonolysis of the trimetaphosphate moiety, thus forming compounds of the general structure **204**, which match the exact mass observed in HRMS (Scheme 4-20).



Scheme 4-20. Presumed structure of the isolated triphosphate **204** and mechanism of formation.

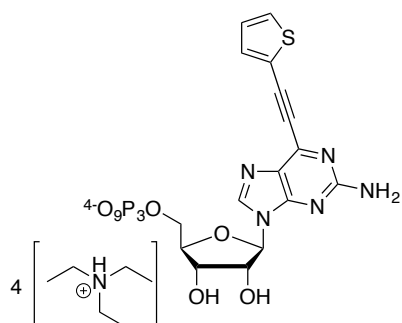
A ¹⁴N-NMR experiment could provide further information regarding the identity of these compounds. If the presence of an NH₂ group is verified, possible solutions to this drawback would be extending the time of TEAB treatment to allow for complete hydrolysis of **205**; or, if complete conversion is not feasible, carrying out the ammonia treatment at ambient conditions instead of 55 °C.

4.5 Experimental

Acetonitrile was dried for two days over activated 3 Å molecular sieves. Trimethyl phosphate and triethyl phosphate were distilled under CaH₂ and stored over 3 Å molecular sieves. 2,4,6-Collidine, tri-*n*-octylamine and tri-*n*-butylamine were distilled and stored over KOH pellets. tris(tetra-*n*-butylammonium) pyrophosphate was supplied by Sigma-Aldrich and was stored as a 0.5 M solution in DMF under 3 Å molecular sieves. All other solvents and reagents were supplied by Sigma-Aldrich, Acros Organics, Alfa Aesar, Fisher Scientific and Fluorochem and were used without further purification. ¹H-NMR and ¹³C-NMR spectra were obtained on Bruker DPX400 at 400 MHz and 100 MHz respectively. HPLC purification was carried out on a DIONEX RP-HPLC system equipped with an ACE 250 mm x 2.5 mm C18 column. A standard 31 min program using buffer A (0.1 M TEAAc in H₂O) and buffer B (0.1 M TEAAc in 80% MeCN/H₂O) was used. The flow rate was 4.5 mL/min with a gradient as follows. 0-3 minutes held at 5% buffer B, then a linear gradient from 5-90% for 20 minutes, then it was held at 90% buffer B for 3 minutes,

then the gradient was returned to 5% B and held at 5% buffer B for 4 minutes. HPLC analysis of triphosphate **201** was carried out on a DIONEX RP-HPLC system equipped with a Phenomenex Clarity 5u Oligo RP 250 mm x 4.6 mm C18 column. A standard 55 min program using buffer A (0.1 M TEAAc in H₂O) and buffer B (0.1 M TEAAc in 80% MeCN/H₂O) was used. The flow rate was 1.0 mL/min with a gradient as follows. 0-3 minutes held at 5% buffer B, then a linear gradient from 5-95% for 40 minutes, then it was held at 95% buffer B for 5 minutes, then the gradient was returned to 5% B and held at 5% buffer B for 5 minutes. Ion exchange chromatography was carried out on a GE Healthcare AKTA explorer system equipped with a GE Healthcare custom source 15Q 17 ml column using buffer A (0.2 M aq. NaCl) and buffer B (3.0 M aq. NaCl). The flow rate was 15 mL/min with a gradient as follows. 4 minutes at 0% buffer B. Then a linear gradient from 0-2% buffer B for 2 minutes, then it was held at 2% buffer B for 5 minutes, then a linear gradient from 2-10% for 3 minutes, then it was held at 10% for 2 minutes and then a linear gradient from 10-20% for 1.5 minutes.

triethylammonium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(thiophen-2-ylethynyl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate **200¹⁹⁴**



A flame-dried flask was charged with nucleoside **121** (20 mg, 0.05 mmol) and Proton Sponge® (17.3 mg, 0.81 mmol). The solids were dissolved in trimethyl phosphate (0.27 mL) and cooled to 0 °C. POCl₃ (6.5 µL, 0.07 mmol) was then added and the solution was stirred at 0 °C under argon for 2 hours. Tributylamine (0.06 mL, 0.27 mmol) was added followed by bis(tri-*n*-butylammonium) pyrophosphate (0.5 M in DMF, 0.54 mL, 0.27 mmol) and the solution was stirred at 0 °C under argon

for 5 min. The reaction was quenched with 0.1 M TEAB buffer (pH = 8.5, 1 mL), stirred for 5 min. and then lyophilized. The residue was purified by RP-HPLC using the standard method described in the general experimental section. The appropriate fractions were combined and lyophilized to yield 8.2 mg of a mixture of triphosphate **200** and the corresponding 5'-monophosphate.

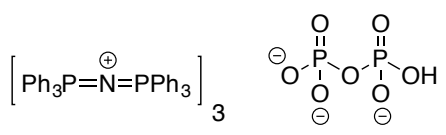
¹H-NMR (400 MHz, D₂O, only the peaks corresponding to **200** are reported.) δ: 8.4 (s, 1H), 7.58 (d, *J* = 5.1 Hz, 1H), 7.53 (d, *J* = 3.8 Hz, 1H), 7.06 (dd, *J* = 5.3, 3.9 Hz, 1H), 5.96 (d, *J* = 6.1 Hz, 1H), 4.79 (t, *J* = 5.5 Hz, 1H), 4.52 (dd, *J* = 5.3, 3.4 Hz, 1H), 4.31 (m, 1H, obscured by various impurities), 4.19 (m, 2H, obscured by various impurities).

³¹P-NMR (162 MHz, D₂O, only the peaks corresponding to **200** are reported.) δ: -10.9 (d, *J*_{P-P} = 19.2 Hz), -11.4 (d, *J*_{P-P} = 21.3 Hz), -23.0 (t, *J*_{P-P} = 19.2 Hz).

MALDI-TOF *m/z*: [M-C₆H₁₆+4H]⁺ 612.

HPLC retention time (semi-preparative): 15.8 min.

tris{[bis(triphenylphosphoranylidene)]ammonium} pyrophosphate²⁰⁰



Tetrasodium pyrophosphate decahydrate (0.79 g, 1.77 mmol) was dissolved in distilled water (25 mL) at 50 °C and the pH was adjusted to 5 with 1 M hydrochloric acid. This solution was added dropwise, with stirring, to a solution of [bis(triphenylphosphoranylidene)ammonium] chloride (2.0 g, 3.47 mmol) in distilled water (260 mL) at 50 °C. The mixture was transferred to six 50 mL centrifuge tubes, allowed to cool to room temperature and the tubes were centrifuged at 4000 rpm for 20 minutes. The supernatant was removed and the solid residues were combined, recrystallized from distilled water (45 mL) at 80 °C, collected by centrifugation at 4000 rpm for 20 minutes, dried, ground to a powder and further dried in a vacuum desiccator over P₂O₅ to give PPN pyrophosphate as a white powder (1.45 g, 23%). The compound was stored as a 0.4 M solution in MeCN over 3 Å mol. sieves.

¹H-NMR (400 MHz, DMSO-d₆) δ: 7.70-7.51 (m, phenyl H).

¹³C-NMR (100 MHz, DMSO-d₆) δ: 133.7, 132-131.8 (m), 129.7-129.3 (m), 126.7 (d, *J*_{C-P} = 107 Hz).

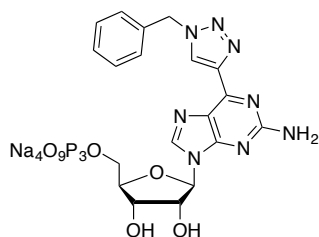
³¹P-NMR (162 MHz, DMSO-d₆) δ: 20.7 (s, PPh₃), -6.1 (s, pyrophosphate).

LC-MS (+ve mode): [PPN]⁺ 538.

GENERAL PROCEDURE FOR TRIPHOSPHATE SYNTHESIS USING THE HODGSON METHOD (GP1)²⁰⁰

The nucleoside was dried overnight under high vacuum in a 5 mL round bottom flask. It was then dissolved in triethyl phosphate and cooled at -10 °C. POCl₃ was then added, followed by *sym*-collidine and the resulting suspension was stirred under argon at -10 °C for 2 h. PPN pyrophosphate was added (0.4 M in MeCN), followed by tri-*n*-octylamine and the mixture was vigorously stirred at 0 °C for 2 h. and at r.t. for 20 h. TEAB (0.1 M, pH = 8.5) and CHCl₃ were then added and the mixture was vigorously stirred under air for 30 min. It was then transferred to a separatory funnel and the aqueous layer was washed twice with CHCl₃. The aqueous layer was transferred to a falcon tube. NaI and acetone were added. The mixture was vigorously shaken for 30 min. and then cooled to 0 °C for 30 min. The precipitate was collected by centrifugation, re-dissolved in TEAB (0.1 M, pH = 8.5) and cooled to 0 °C. A 0.1 M NaI solution in acetone was added dropwise, with stirring. The precipitate was collected by centrifugation, the pellet was lyophilized and purified by RP-HPLC using a standard method (see general information) to afford the corresponding triphosphate as the triethylammonium salt, which was exchanged to the sodium salt by dissolving in TEAB, (0.1 M, pH = 8.5) cooling to 0 °C and adding a 0.1 M NaI in acetone solution dropwise while shaking.

tetrasodium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(1-benzyl-1*H*-1,2,3-triazol-4-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate 199

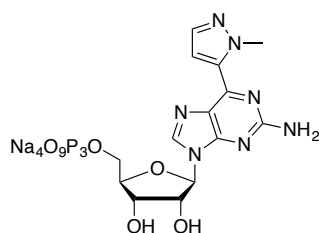


Preparation of **199** was attempted according to GP1 using **124** (26 mg, 0.06 mmol), triethyl phosphate (0.3 mL), POCl₃ (8.4 uL, 0.09 mmol), 2,4,6-collidine (8 uL, 0.06 mmol), PPN pyrophosphate (0.4 mL, 0.12 mmol, 0.4 M in MeCN) and tri-*n*-octylamine (0.05 mL, 0.12 mmol). It was quenched with TEAB (1.5 mL) and washed with CHCl₃ (2 × 6 mL). The first precipitation was performed using NaI (24 mg) and acetone (6 mL). The second precipitation and cation exchange after RP-HPLC were performed using TEAB (0.1 M, pH = 8.5, 0.9 mL) and a 0.1 M solution of NaI in acetone (1.49 mL) respectively. 3.8 mg of a mixture of triphosphate species were obtained.

MALDI-TOF m/z: [M-4Na+3H]⁻ 663.

HPLC retention time (semi-preparative): 13.8 min.

tetrasodium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(1-methyl-1*H*-pyrazol-5-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate **201**

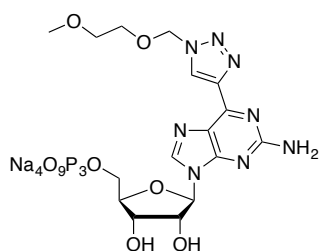


Preparation of **201** was attempted according to GP1 using **126** (28 mg, 0.08 mmol), triethyl phosphate (0.44 mL), POCl₃ (11.2 uL, 0.12 mmol), 2,4,6-collidine (10.5 uL, 0.08 mmol), PPN pyrophosphate (0.4 mL, 0.16 mmol, 0.4 M in MeCN) and tri-*n*-octylamine (0.07 mL, 0.16 mmol). It was quenched with TEAB (2.0 mL) and washed with CHCl₃ (2 × 7 mL). The first precipitation was performed using NaI (31 mg) and acetone (7.8 mL). The second precipitation was performed using TEAB (0.1 M, pH = 8.5, 1.2 mL) and a 0.1 M solution of NaI in acetone (1.9 mL). The final sodium exchange was performed using TEAB (0.1 M, pH = 8.5, 0.4 mL) and a 0.1 M solution of NaI in acetone (0.74 mL). 20 mg of a mixture of triphosphate species were obtained.

MALDI-TOF m/z: [M-4Na+3H]⁻ 586.

HPLC retention time (semi-preparative): 11.3 min.

tetrasodium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(1-((2-methoxyethoxy)methyl)-1*H*-1,2,3-triazol-4-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate **202**



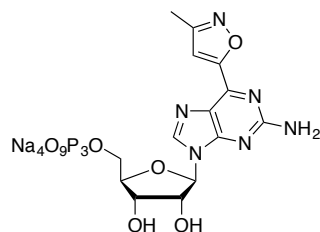
Preparation of **202** was attempted according to GP1 using **154** (48 mg, 0.11 mmol), triethyl phosphate (0.56 mL), POCl₃ (15.9 uL, 0.09 mmol), 2,4,6-collidine (15 uL, 0.11 mmol), PPN pyrophosphate (0.57 mL, 0.23 mmol, 0.4 M in MeCN) and tri-*n*-octylamine (0.1 mL, 0.23 mmol). It was quenched with TEAB (2.9 mL) and washed with CHCl₃ (2 × 10 mL). The first precipitation was performed using NaI (45 mg) and acetone (11 mL). The second precipitation was performed using TEAB (0.1 M, pH = 8.5, 1.6 mL) and a 0.1 M solution of NaI in acetone (2.76 mL). The final sodium

exchange was performed using TEAB (0.1 M, pH = 8.5, 0.6 mL) and a 0.1 M solution of NaI in acetone (1.0 mL). 7.7 mg of a mixture of triphosphate species were obtained.

MALDI-TOF m/z: $[M-4Na+3H]^-$ 661.

HPLC retention time (semi-preparative): 12.0 min.

tetrasodium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(3-methylisoxazol-5-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate **203**



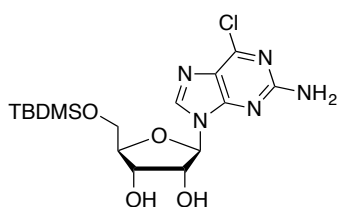
Preparation of **199** was attempted according to GP1 using **169** (30 mg, 0.09 mmol), triethyl phosphate (0.45 mL), POCl₃ (12 uL, 0.13 mmol), 2,4,6-collidine (11.3 uL, 0.09 mmol), PPN pyrophosphate (0.42 mL, 0.17 mmol, 0.4 M in MeCN) and tri-*n*-octylamine (0.08 mL, 0.17 mmol). It was quenched with TEAB (2.3 mL) and washed with CHCl₃ (2 × 9 mL). The first precipitation was

performed using NaI (34 mg) and acetone (8.6 mL). The second precipitation was performed using TEAB (0.1 M, pH = 8.5, 1.3 mL) and a 0.1 M solution of NaI in acetone (2.15 mL). The final sodium exchange was performed using TEAB (0.1 M, pH = 8.5, 0.4 mL) and a 0.1 M solution of NaI in acetone (0.74 mL). 14 mg of a mixture of triphosphate species were obtained.

MALDI-TOF m/z: $[M-4Na+3H]^-$ 587.

HPLC retention time (semi-preparative): 11.8 min.

(2*R*,3*R*,4*S*,5*R*)-2-(2-amino-6-chloro-9*H*-purin-9-yl)-5-(((*tert*-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-3,4-diol **217²⁰⁴**



A flame-dried round-bottom flask was charged with **116** (600 mg, 1.98 mmol) and DMAP (24 mg, 0.2 mmol). The solids were dissolved in DMSO (2.8 mL) and DCM (11.8 mL). Et₃N (0.6 mL) was added, followed by *tert*-butyldimethylsilyl chloride (600 mg, 3.96 mmol). The reaction was stirred under argon at ambient temperature for 3 hours. It was then

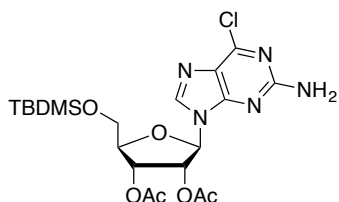
quenched with MeOH (5 mL). The volatiles were evaporated *in vacuo*. The residue was dissolved in EtOAc (40 mL) and washed with sat. aq. NH₄Cl (1 × 30 mL) and brine (4 × 20 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated *in vacuo*. Column chromatography purification (SiO₂) eluting with a mixture of MeOH/DCM (0:10 → 1:9) afforded **217** as a white foam (332 mg, 40%).

¹H-NMR (400 MHz, CDCl₃) δ: 8.04 (s, 1H), 6.03 (brs, 2H), 5.92 (d, *J* = 5.3 Hz, 1H), 5.88 (brs, 1H), 4.48 (triplet, *J* = 4.2 Hz, 1H), 4.42 (brs, 1H), 4.36 (triplet, *J* = 4.2 Hz, 1H), 4.15 (doublet, *J* = 2.7 Hz, 1H), 3.80 (m, 2H), 0.80 (s, 9H), 0.01 (s, 3H), 0.00 (s, 3H)

^{13}C -NMR (100 MHz, CDCl_3) δ : 159.1, 153.0, 150.8, 140.5, 124.5, 88.3, 85.8, 75.9, 71.4, 63.1, 25.9, 18.4, -5.4, -5.5

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{16}\text{H}_{26}\text{N}_5\text{NaO}_4\text{SiCl}$ 438.1335, found 438.1322.

(2R,3R,4R,5R)-2-(2-amino-6-chloro-9H-purin-9-yl)-5-(((tert-butyl)dimethylsilyloxy)methyl)tetrahydrofuran-3,4-diyl diacetate **218¹⁶⁰**



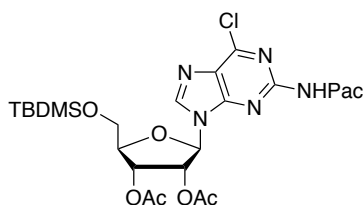
A flame-dried round-bottom flask was charged with silylated nucleoside **217** (200 mg, 0.48 mmol) and DMAP (5.8 mg, 0.048 mmol). The solids were dissolved in MeCN (4.8 mL). Et_3N (0.28 mL, 1.92 mmol) was added and the solution was cooled at 0 °C. Acetic anhydride (0.1 mL, 0.1 mmol) was then added dropwise. After the addition the ice-bath was removed and the reaction was stirred under argon at ambient temperature for another 30 minutes. The volatiles were evaporated *in vacuo* and the residue was dissolved in EtOAc (30 mL) and washed with sat. aq. NH_4Cl (1 \times 30 mL) and brine (1 \times 30 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with a mixture of acetone/DCM (0:10 \rightarrow 2:8) afforded **218** as a white foam (184 mg, 76%).

^1H -NMR (400 MHz, CDCl_3) δ : 8.02 (s, 1H), 5.92 (d, $J = 6.8$ Hz, 1H), 5.64 (m, 1H), 5.52 (brs, 2H), 5.42 (dd, $J = 8.0, 2.6$ Hz, 1H), 4.16 (quartet, $J = 2.5$ Hz, 1H), 3.76 (qd, $J = 12, 2.2$ Hz, 2H), 2.03 (s, 3H), 1.90 (s, 3H), 0.81 (s, 9H), 0.01 (s, 3H), 0.00 (s, 3H).

^{13}C -NMR (100 MHz, CDCl_3) δ : 169.8, 169.3, 159.4, 153.8, 151.5, 140.1, 125.3, 84.6, 83.9, 73.8, 71.8, 63.1, 25.9, 20.7, 20.4, 18.4, -5.50, -5.5.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{20}\text{H}_{30}\text{N}_5\text{NaO}_6\text{SiCl}$ 522.1546, found 522.1532.

(2R,3R,4R,5R)-2-(((tert-butyl)dimethylsilyloxy)methyl)-5-(6-chloro-2-(2-phenoxyacetamido)-9H-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **211**



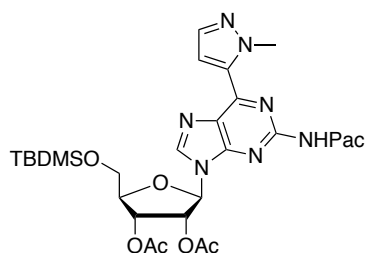
A flame-dried round-bottom flask was charged with nucleoside **218** (600 mg, 1.12 mmol). The solids were dissolved in pyridine (1 mL). Phenoxyacetyl chloride (0.17 mL, 1.44 mmol) was then added and the mixture was stirred at r.t. under argon for 20 min. The volatiles were evaporated *in vacuo* and the residue was dissolved in EtOAc (60 mL) and washed with sat. aq. NH_4Cl (1 \times 60 mL) and brine (1 \times 60 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with a mixture of EtOAc/n-hexane (0:10 \rightarrow 1:1) afforded **211** as a yellow foam (135 mg, 17%).

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 9.00 (s, 1H), 8.30 (s, 1H), 7.20 (m, 2H), 6.94-6.85 (m, 3H), 6.21 (d, $J = 6.4$ Hz, 1H), 5.65 (t, $J = 5.7$ Hz, 1H), 5.45 (dd, $J = 8.0, 2.7$ Hz, 1H), 4.6 (s, 2H), 4.20 (quartet, $J = 2.2$ Hz, 1H), 3.80 (qd, $J = 12.0, 2.6$ Hz, 2H), 1.91 (s, 3H), 1.89 (s, 3H), 0.80 (s, 9H), 0.00 (s, 6H).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3) δ : 175.6, 169.9, 169.5, 166.3, 157.0, 152.6, 151.4, 151.3, 143.1, 129.8, 128.8, 122.4, 114.9, 85.7, 84.5, 74.3, 71.9, 67.9, 63.2, 25.9, 20.7, 20.4, 18.4.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{28}\text{H}_{36}\text{N}_5\text{NaO}_8\text{SiCl}$ 656.1914, found 656.1887.

(2*R*,3*R*,4*R*,5*R*)-2-(((*tert*-butyldimethylsilyloxy)methyl)-5-(6-(1-methyl-1*H*-pyrazol-5-yl)-2-(2-phenoxyacetamido)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **221**



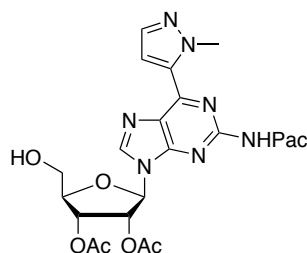
A flame-dried round bottom flask was charged with chloride **211** (100 mg, 0.16 mmol), boronate **173** (49 mg, 0.24 mmol), KF on alumina (136 mg, 0.72 mmol, 5.5 mmol F⁻/g loading) and Pd-132 (11 mg, 0.052 mmol). The solids were dissolved in 2-MeTHF (1 mL). The solution was purged with argon for 15 minutes and then heated at 80 °C for 4 hours. The mixture was diluted with EtOAc (50 mL) and brine (50 mL). The organic layer was washed with brine (4 × 25 mL), then collected, dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with a mixture of acetone/DCM (1:9 → 1:1) afforded **221** as an off-white foam (275 mg, 77%).

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 8.92 (s, 1H), 8.24 (s, 1H), 7.58 (d, $J = 2.0$ Hz, 2H), 7.49 (d, $J = 2.1$ Hz, 1H), 7.24 (t, $J = 8.7$ Hz, 1H), 6.94, (t, $J = 7.8$ Hz, 3H), 6.22 (d, $J = 6.5$ Hz, 1H), 5.73 (triplet, $J = 5.9$ Hz, 1H), 5.49 (dd, $J = 5.3, 2.8$ Hz, 1H), 4.61 (brs, 2H), 4.40 (s, 3H), 4.20 (triplet, $J = 2.2$ Hz, 1H), 3.80 (dd, $J = 20.2, 2.3$ Hz, 1H), 3.78 (dd, $J = 20.1, 2.4$ Hz, 1H), 2.05 (s, 3H), 1.92 (s, 3H), 0.81 (s, 9H), 0.01 (s, 3H), 0.00 (s, 3H)

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3) δ : 169.9, 169.4, 165.9, 157.1, 153.0, 151.2, 148.0, 142.2, 138.5, 135.5, 129.9, 127.7, 122.5, 115.0, 112.9, 84.9, 84.1, 73.9, 71.7, 68.0, 63.1, 40.9, 25.9, 20.7, 20.4, 18.4, -5.4, -5.5.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{32}\text{H}_{41}\text{N}_7\text{NaO}_8\text{Si}$ 702.2673, found 702.2681.

(2*R*,3*R*,4*R*,5*R*)-2-(hydroxymethyl)-5-(6-(1-methyl-1*H*-pyrazol-5-yl)-2-(2-phenoxyacetamido)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **222**



A round-bottom flask was charged with **221** (200 mg, 0.29 mmol). A mixture of TBAF (1.0 M in THF, 0.15 mL, 0.32 mmol), THF (1.8 mL) and AcOH (9.2 μL , 0.32 mmol) was prepared in a vial and added to the nucleoside. The mixture was stirred under air for 90 min and concentrated *in vacuo*. The mixture was diluted with EtOAc (30 mL). The organic layer was washed with brine (1 × 30 mL),

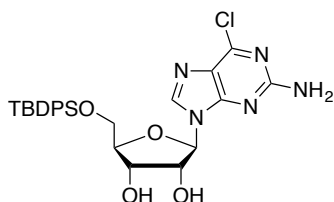
then collected, dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with a mixture of acetone/DCM (1:9 \rightarrow 2:8) afforded **222** as an off-white foam (135 mg, 90%).

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 9.18 (s, 1H), 8.13 (s, 1H), 7.67 (d, $J = 2.0$ Hz, 1H), 7.59 (d, $J = 2.0$ Hz, 1H), 7.36 (m, 2H), 7.06 (d, $J = 8.2$ Hz, 1H), 6.1 (d, $J = 6.9$ Hz, 1H), 6.02 (m, 1H), 5.82 (dd, $J = 5.3, 2.5$ Hz, 1H), 4.70 (s, 2H), 4.47 (s, 3H), 4.38 (quartet, $J = 2.2$ Hz, 1H), 4.04 (dd, $J = 12.9, 1.7$ Hz, 1H), 3.93 (d, $J = 12.6$ Hz, 1H), 2.18 (s, 2H), 2.16 (s, 3H).

$^{13}\text{C-NMR}$ (100 MHz, acetone- d_6) δ : 168.9, 168.5, 165.7, 157.5, 152.3, 151.1, 147.0, 143.2., 137.3, 134.8, 129.0, 126.7, 121.0, 114.4, 112.2, 85.7, 84.2, 72.9, 71.2, 67.2, 61.2, 40.2.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{26}\text{H}_{27}\text{N}_7\text{NaO}_8$ 588.1813, found 588.1788.

(2R,3R,4S,5R)-2-(2-amino-6-chloro-9H-purin-9-yl)-5-(((tert-butyl)diphenylsilyloxy)methyl)tetrahydrofuran-3,4-diol **223**



A flame-dried round-bottom flask was charged with **116** (600 mg, 1.98 mmol), DMAP (24 mg, 0.2 mmol) and imidazole (337 mg, 4.95 mmol). The solids were dissolved in DMF (21 mL) and TBDPS-Cl was then added (1.02 mL, 3.96 mmol). The reaction was stirred under argon at ambient temperature

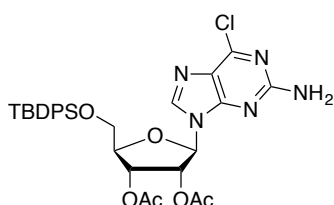
for 3 hours. The volatiles were evaporated *in vacuo* and the residue was dissolved in EtOAc (30 mL) and washed with sat. aq. NH_4Cl (1×100 mL) and brine (1×100 mL). The organic phase was dried over Na_2SO_4 , filtered and concentrated *in vacuo*. Column chromatography purification (SiO_2) eluting with a mixture of acetone/DCM (2:8 \rightarrow 1:1) afforded **223** as a white foam (764 mg, 71%).

$^1\text{H-NMR}$ (400 MHz, CDCl_3) δ : 7.97 (s, 1H), 7.60 (m, 4H), 7.43-7.32 (m, 6H), 5.91 (d, $J = 5.8$ Hz, 1H), 5.47 (triplet, 1H), 5.39 (brs, 2H), 4.67 (q, $J = 5.3$ Hz, 1H), 4.5 (m, 1H), 4.28 (quartet, $J = 3.0$ Hz, 1H), 3.88 (dd, $J = 11.5, 3.5$ Hz, 1H), 3.82 (dd, $J = 11.6, 3.2$ Hz, 1H), 3.52 (brs, 1H), 1.00 (s, 9H).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3) δ : 158.8, 152.9, 151.5, 140.5, 135.5, 135.4, 132.6, 132.5, 130.1, 130.0, 128.0, 127.9, 125.3, 89.2, 86.1, 75.4, 71.7, 63.8, 26.9, 19.2.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{30}\text{H}_{34}\text{N}_5\text{NaO}_4\text{SiCl}$ 562.1648, found 562.1633.

(2R,3R,4R,5R)-2-(2-amino-6-chloro-9H-purin-9-yl)-5-(((tert-butyl)diphenylsilyloxy)methyl)tetrahydrofuran-3,4-diol diacetate **224¹⁶⁰**



A flame-dried round-bottom flask was charged with alcohol **223** (300 mg, 0.56 mmol) and DMAP (6.9 mg, 0.06 mmol). The solids were suspended in MeCN (8.4 mL). Et_3N (0.3 mL, 2.22 mmol) was added and the suspension was cooled at 0 °C. Acetic anhydride (0.1 mL, 1.17 mmol) was then added

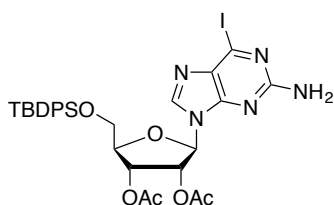
dropwise, at which point the suspension turned into a clear solution. After the addition the ice-bath was removed and the reaction was stirred under argon at ambient temperature for 30 minutes. The volatiles were evaporated *in vacuo* and the residue was dissolved in EtOAc (100 mL) and washed with sat. aq. NH₄Cl (1 × 100 mL) and brine (1 × 100 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated *in vacuo*. Column chromatography purification (SiO₂) eluting with a mixture of acetone/DCM (0:10 → 2:8) afforded **224** as a white foam (308 mg, 89%).

¹H-NMR (400 MHz, CDCl₃) δ: 7.97 (s, 1H), 7.66 (m, 4H), 7.45-7.32 (m, 6H), 6.11 (d, *J* = 6.5 Hz, 1H), 6.01 (t, *J* = 5.4 Hz, 1H), 5.79 (dd, *J* = 5.3, 3.3 Hz, 1H), 5.28 (brs, 2H), 4.29 (q, *J* = 3.2 Hz, 1H), 3.98 (dd, *J* = 11.6, 3.3 Hz, 1H), 3.85 (dd, *J* = 11.6, 3.3 Hz, 1H), 2.14 (s, 3H), 2.07 (s, 3H), 1.09 (s, 9H).

¹³C-NMR (100 MHz, CDCl₃) δ: 169.7, 169.3, 159.3, 153.7, 151.6, 140.4, 135.6, 135.5, 132.6, 132.3, 130.1, 130.0, 128.0, 127.9, 125.5, 85.0, 83.3, 72.8, 71.0, 63.3, 26.9, 20.7, 20.4, 19.2.

HRMS (micrOTOF-Q) *m/z*: [M+Na]⁺ calcd for C₃₀H₃₄N₅NaO₆SiCl 646.1859, found 646.1834.

(2*R*,3*R*,4*R*,5*R*)-2-(2-amino-6-iodo-9*H*-purin-9-yl)-5-(((*tert*-butyldiphenylsilyl)oxy)methyl)tetrahydrofuran-3,4-diyl diacetate **225¹⁶³**



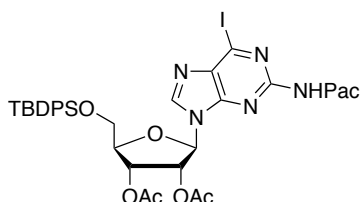
To a solution of chloride **224** (1.3 g, 2.0 mmol) in 2-butanone (40 mL) was added sodium iodide (6.0 g, 40.0 mmol) and TFA (0.03 mL, 4 mmol). The flask was covered with foil and the solution was stirred for 16 h under air. The resulting red solution was concentrated *in vacuo*. The residue was suspended in EtOAc, the suspension filtered through a pad of celite (the flask was rinsed with 3 × 50 mL EtOAc) and the solution was washed with sat. NaHCO₃ (1 × 80 mL) and sat. Na₂S₂O₃ (1 × 80 mL). The organic phase was dried over Na₂SO₄ and the solvent was removed *in vacuo*, affording iodide **225** as an off-white foam (1.3 g, 91%).

¹H-NMR (400 MHz, CDCl₃) δ: 7.97 (s, 1H), 7.68-7.63 (m, 4H), 7.46-7.33 (m, 6H), 6.08 (d, *J* = 6.7 Hz, 1H), 5.78 (dd, *J* = 5.3, 3.2 Hz, 1H), 5.03 (brs, 2H), 4.27 (q, *J* = 3.1 Hz, 1H), 4.00 (dd, *J* = 11.6, 3.2 Hz, 1H), 3.83 (dd, *J* = 11.6, 3.2 Hz, 1H), 2.14 (s, 3H), 2.07 (s, 3H), 1.08 (s, 9H).

¹³C-NMR (100 MHz, CDCl₃) δ: 169.7, 169.3, 158.9, 149.9, 139.8, 135.6, 135.5, 132.6, 132.5, 132.4, 130.1, 130.0, 128.0, 127.9, 123.05, 85.0, 83.4, 72.7, 71.1, 63.3, 26.9, 20.7, 20.5, 19.2.

HRMS (micrOTOF-Q) *m/z*: [M+Na]⁺ calcd for C₃₀H₃₄IN₅NaO₆Si 738.1215, found 738.1196.

(2*R*,3*R*,4*R*,5*R*)-2-(((*tert*-butyldiphenylsilyl)oxy)methyl)-5-(6-iodo-2-(2-phenoxyacetamido)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **226**



A flame-dried round-bottom flask was charged with iodide **225** (500 mg, 0.69 mmol). The solids were dissolved in MeCN (3.2 mL) and pyridine was

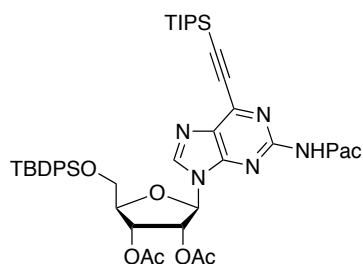
added (0.16 mL, 2.1 mmol). The solution was cooled at 0 °C. Phenoxyacetyl chloride (0.1 mL, 0.8 mmol) was added over a period of 10 minutes. After the addition the reaction was allowed to warm to r.t. and the mixture was stirred at this temperature under argon for 16 h. The volatiles were evaporated *in vacuo* and the residue was dissolved in EtOAc (60 mL) and washed with sat. aq. NH₄Cl (1 x 50 mL) and brine (1 x 50 mL). The organic phase was dried over Na₂SO₄, filtered and concentrated *in vacuo*. Column chromatography purification (SiO₂) eluting with a mixture of acetone/DCM (0:1 → 1:9) afforded **226** as a yellowish foam (397 mg, 67%).

¹H-NMR (400 MHz, CDCl₃) δ: 9.08 (s, 1H), 8.25 (s, 1H), 7.63 (td, *J* = 8.1, 1.3 Hz, 1H), 7.43-7.26 (m, 8H), 7.04 (t, *J* = 7.4 Hz, 1H), 7.00 (d, *J* = 8.0 Hz, 2H), 6.27 (d, *J* = 5.6 Hz, 1H), 5.9 (t, *J* = 5.6 Hz, 1H), 5.81 (dd, *J* = 4.2 Hz, 1H), 4.66 (ABq, *J* = 15.5 Hz, 3H), 4.35 (q, *J* = 3.6 Hz, 1H), 4.62 (dd, *J* = 11.9, 2.4 Hz, 1H), 3.95 (dd, *J* = 11.7, 3.9 Hz, 1H), 2.15 (s, 3H), 2.07 (s, 3H), 1.07 (s, 9H).

¹³C-NMR (100 MHz, CDCl₃) δ: 169.7, 169.5, 165.8, 157.0, 152.4, 151.5, 151.2, 143.1, 135.7, 135.5, 132.7, 132.4, 130.0, 129.9, 129.1, 127.9, 127.8, 122.5, 114.9, 86.1, 84.0, 73.9, 71.1, 67.8, 63.7, 26.9, 20.7, 20.4, 19.2.

HRMS (microTOF-Q) *m/z*: [M+Na]⁺ calcd for C₃₈H₄₀N₅NaO₈Si 872.1583, found 872.1555.

(2*R*,3*R*,4*R*,5*R*)-2-(((*tert*-butyldiphenylsilyl)oxy)methyl)-5-(2-(2-phenoxyacetamido)-6-(triisopropylsilyl)ethynyl)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **227**



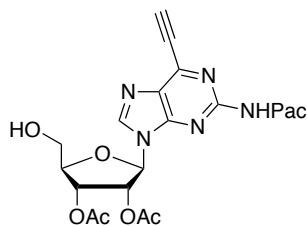
A mixture of **226** (200 mg, 0.24 mmol), [Pd(PPh₃)₂]Cl₂ (3.3 mg, 4.7 x 10⁻³ mmol), DIPEA (0.12 mL, 0.71 mmol) and CuI (2.7 mg, 0.14 mmol) was dissolved in dry DMF (2.9 mL). The mixture was subjected to the freeze-pump-thaw cycle three times. (Triisopropylsilyl)acetylene (0.07 mL, 0.31 mmol) was then added over 1 hour and the yellow solution was stirred under argon at r.t. for 16 h. Another 0.07 mL of (triisopropyl)silylacetylene was added and the solution was stirred under argon at 30 °C for 16 h. It was then poured into brine (50 mL) and extracted with EtOAc (2 x 50 mL). The organic phase was washed with brine (4 x 50 mL), dried over Na₂SO₄ and the solvents were removed *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with EtOAc/*n*-hexane (0:10 → 4:6) to afford **227** as a yellow foam (168 mg, 79%).

¹H-NMR (400 MHz, CDCl₃) δ: 9.10 (brs, 1H), 8.22 (s, 1H), 7.65 (qd, *J* = 8.1, 1.3 Hz, 4H), 7.43-7.27 (m, 8H), 7.06-6.98 (m, 3H), 6.29 (d, *J* = 5.7 Hz, 1H), 5.90 (t, *J* = 5.6 Hz, 1H), 5.83 (dd, *J* = 5.7, 4.2 Hz, 1H), 4.69 (m, 2H), 4.34 (q, *J* = 3.7 Hz, 1H), 4.01 (dd, *J* = 11.8, 2.7 Hz, 1H), 3.94 (dd, *J* = 11.8, 3.8 Hz, 1H), 2.14 (s, 3H), 2.07 (s, 3H), 1.22 (m, 21H), 1.09 (s, 9H).

¹³C-NMR (100 MHz, CDCl₃) δ: 169.7, 169.5, 157.2, 152.6, 151.9, 143.2, 142.3, 135.7, 135.5, 132.8, 132.7, 132.5, 130.0, 129.9, 129.8, 127.9, 127.8, 122.3, 114.9, 104.02, 100.3, 85.6, 83.8, 73.6, 71.0, 67.9, 63.7, 53.5, 27.0, 20.7, 20.4, 19.2, 18.71, 11.3.

HRMS (microTOF-Q) *m/z*: [M+Na]⁺ calcd for C₄₉H₆₁N₅NaO₈Si₂ 926.3951, found 926.3912.

(2*R*,3*R*,4*R*,5*R*)-2-(6-ethynyl-2-(2-phenoxyacetamido)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diyl diacetate **210**



A round-bottom flask was charged with alkyne **227** (140 mg, 0.15 mmol). A mixture of TBAF (1.0 M in THF, 0.32 mL, 0.32 mmol) and AcOH (18.3 μ L, 0.32 mmoles) was prepared in THF (1.2 mL). The mixture was briefly shaken and added to the nucleoside. The solution was stirred at r.t. under air for 90 min.

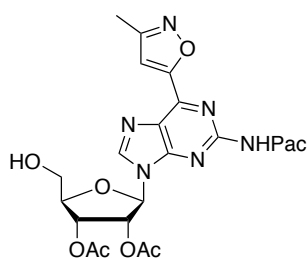
The mixture was concentrated *in vacuo*. The residue was purified *via* column chromatography (SiO₂) eluting with acetone/DCM (0:10 \rightarrow 4:6) to afford **210** as a brownish foam (69 mg, 90%)

¹H-NMR (400 MHz, CDCl₃) δ : 9.27 (s, 1H), 8.16 (s, 1H), 7.35 (t, J = 7.9 Hz, 2H), 7.06 (J = 7.5 Hz, 1H), 7.01 (d, J = 4.2 Hz, 2H), 6.07 (d, J = 3.5 Hz, 1H), 5.99 (t, J = 6.1 Hz, 1H), 5.86 (dd, J = 5.4, 2.8 Hz, 1H), 4.69 (ABq, J = 15.3 Hz, 2H), 4.36 (triplet, J = 2.3 Hz, 1H), 4.06 (dd, J = 13.0 Hz, 2.0 Hz, 1H), 3.94 (dd, J = 13.0 Hz, 2.2 Hz, 1H), 3.76 (s, 1H), 2.17 (s, 3H, overlap with acetone), 2.03 (s, 3H).

¹³C-NMR (100 MHz, CDCl₃) δ : 169.7, 169.2, 166.5, 156.9, 152.1, 151.6, 145.1, 141.86, 133.7, 129.9, 122.5, 114.9, 88.1, 87.0, 85.3, 77.3, 73.1, 71.9, 67.7, 61.8, 20.7, 20.4.

HRMS (micrOTOF-Q) m/z : [M+Na]⁺ calcd for C₂₄H₂₃N₅NaO₈ 532.1444, found 532.1363.

(2*R*,3*R*,4*R*,5*R*)-2-(hydroxymethyl)-5-(6-(3-methylisoxazol-5-yl)-2-(2-phenoxyacetamido)-9*H*-purin-9-yl)tetrahydrofuran-3,4-diyl diacetate **228¹⁷⁶**



N-Chlorosuccinimide (49.6 mg, 0.37 mmol) and pyridine (4 μ L, 0.05 mmol) were added into a round-bottom flask and dissolved in DCM (1.24 mL). A solution of acetaldoxime (22.6 μ L, 0.37 mmol) in DCM (1.24 mL) was added and the reaction was stirred at 40 °C for 20 min. A solution of **210** (160 mg, 0.31 mmol) in DCM (2.47 mL) was added, followed by the dropwise addition of

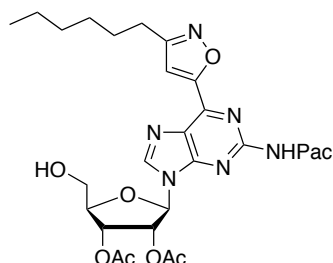
a solution of Et₃N (51 μ L, 0.37 mmol) in DCM (2.47 mL). The solution was stirred under air at r.t. for 16 h. The solvent was evaporated *in vacuo* and the residue dissolved in EtOAc (40 mL), washed with brine (2 \times 40 mL), and dried over Na₂SO₄. The solvents were evaporated *in vacuo* and the residue purified *via* column chromatography (SiO₂) eluting with a mixture of acetone/DCM (0:1 \rightarrow 1:1) to furnish isoxazole **228** as a white foam (161 mg, 91%).

¹H-NMR (400 MHz, CDCl₃) δ : 9.26 (s, 1H), 8.16 (s, 1H), 7.35 (t, J = 7.9 Hz, 2H), 7.06 (t, J = 7.4 Hz, 1H), 7.01 (d, J = 8.4 Hz, 1H), 6.07 (d, J = 6.8 Hz, 1H), 5.98 (t, J = 6.0, 1H), 5.86 (dd, J = 5.4, 2.7 Hz, 1H), 4.69 (ABq, J = 15.3 Hz, 2H), 4.36 (triplet, J = 2.3 Hz, 1H), 4.05 (dd, J = 13.0, 2.0 Hz, 1H), 3.94 (dd, J = 13.0, 2.2 Hz, 1H), 3.76 (s, 3H), 2.17 (s, 3H, overlap with acetone), 2.03 (s, 3H).

^{13}C -NMR (100 MHz, CDCl_3) δ : 169.8, 169.5, 167.0, 163.4, 160.7, 157.1, 153.0, 151.7, 145.2, 143.9, 129.8, 128.0, 122.2, 114.9, 110.3, 87.6, 85.0, 73.4, 71.8, 67.8, 61.8, 20.7, 20.4, 11.5.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{26}\text{H}_{26}\text{N}_6\text{NaO}_9$ 589.1637, found 589.1637.

(2*R*,3*R*,4*R*,5*R*)-2-(6-(3-hexylisoxazol-5-yl)-2-(2-phenoxyacetamido)-9*H*-purin-9-yl)-5-(hydroxymethyl)tetrahydrofuran-3,4-diyl diacetate **229¹⁷⁶**



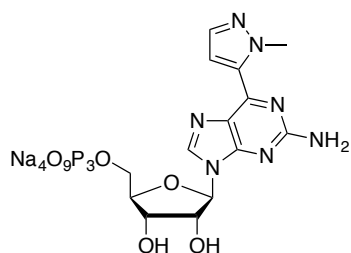
N-Chlorosuccinimide (21 mg, 0.16 mmol) and pyridine (1.6 μL , 0.02 mmol) were added into a round-bottom flask and dissolved in DCM (0.5 mL). A solution of heptaldoxime (20 mg, 0.16 mmol) in DCM (0.5 mL) was added and the reaction was stirred at r.t. for 10 min. A solution of **210** (68 mg, 0.13 mmol) in DCM (1.0 mL) was added, followed by the dropwise addition of a solution of Et_3N (20 μL , 0.16 mmol) in DCM (1.0 mL). The solution was stirred under air at r.t. for 16 h. The solvent was evaporated *in vacuo* and the residue dissolved in EtOAc (40 mL), washed with brine (2×40 mL), and dried over Na_2SO_4 . The solvents were evaporated *in vacuo* and the residue purified *via* column chromatography (SiO_2) eluting with a mixture of acetone/DCM (0:1 \rightarrow 1:9). 20 mg of isoxazole **299** were obtained as a mixture with succinimide in a 1:2.7 ratio.

^1H -NMR (400 MHz, CDCl_3) δ : 9.45 (s, 1H), 8.30 (s, 1H), 7.47 (s, 1H), 7.36 (t, $J = 8.2$ Hz, 1H), 7.10-7.03 (m, 3H), 6.16 (d, $J = 6.5$ Hz, 1H), 6.02 (t, $J = 6.0$ Hz, 1H), 5.86 (dd, $J = 5.4, 2.7$ Hz, 1H), 4.79 (ABq, $J = 15.2$ Hz, 2H), 4.61 (triplet, $J = 6.2$ Hz, 1H), 4.38 (quartet, $J = 2.3$ Hz, 1H), 4.10-3.91 (m, 3H), 2.81 (t, $J = 7.6$ Hz, 2H), 2.18 (s, 3H), 2.05 (s, 3H), 1.76 (q, $J = 7.6$ Hz, 2H), 1.44-1.23 (m, 6H), 0.89 (m, 3H).

^{13}C -NMR (100 MHz, CDCl_3) δ : 178.1, 169.8, 169.4, 167.1, 165.1, 163.3, 157.1, 153.0, 151.7, 145.3, 144.4, 129.8, 128.2, 122.3, 115.0, 109.4, 87.9, 85.0, 73.2, 71.8, 67.8, 61.8, 31.4, 28.8, 28.3, 26.0, 22.5, 20.7, 20.4, 14.0.

HRMS (microTOF-Q) m/z : $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{31}\text{H}_{36}\text{N}_6\text{O}_9\text{Na}$ 659.2436, found 659.2437.

tetrasodium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(1-methyl-1*H*-pyrazol-5-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate **201¹⁹⁴**



A flame-dried flask was charged with nucleoside **222** (30 mg, 0.05 mmol) and Proton Sponge® (17 mg, 0.08 mmol). The solids were dissolved in trimethyl phosphate (0.26 mL) and cooled to 0 °C. POCl_3 (6.4 μL , 0.07 mmol) was then added and the solution was stirred at 0 °C under argon for 2 hours. Tributylamine (0.06 mL, 0.27 mmol) was added followed by bis(tributylammonium) pyrophosphate (0.5 M in DMF, 0.53 mL, 0.27 mmol) and the solution was stirred at 0 °C under argon for 10 min. The reaction was quenched with 0.1 M TEAB buffer (pH = 8.5, 1.3 mL). The solution was transferred to a microwave vial and the flask was rinsed with 28%

NH₄OH (2 × 5 mL). The vial was crimped and ammonolysis was carried out for 3 hours at 55 °C. The residual ammonia was evaporated *in vacuo* and the residue was purified by ion-exchange chromatography using the method described in the general experimental section. The appropriate fractions were combined and lyophilized. The residue was further purified *via* RP-HPLC using the standard method described in the general experimental section. The appropriate fractions were combined and lyophilized. The residue was dissolved in TEAB (0.1 M, pH = 8.5, 0.8 mL) and cooled to 0 °C. A 0.1 M NaI/acetone solution (3 mL) was added dropwise while shaking. The suspension was centrifuged at 6000 rpm for 10 min. The precipitate was collected, washed with acetone (2 mL), centrifuged at the same speed and collected. The precipitate was lyophilized to afford triphosphate **201** as a white powder (4 mg, 11%).

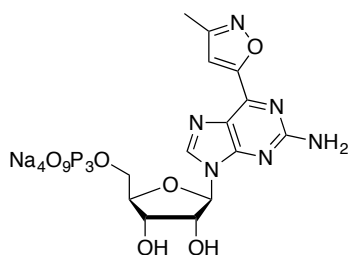
¹H-NMR (500 MHz, D₂O) δ: 8.42 (s, 1H), 7.65 (doublet, *J* = 1.8 Hz, 1H), 6.98 (doublet, *J* = 1.8 Hz, 1H), 6.05 (d, *J* = 6.0 Hz, 1H), 4.83 (t, *J* = 5.6 Hz, 1H), 4.55 (triplet, *J* = 3.5 Hz, 1H), 4.34 (brs, 1H), 4.24-4.16 (m, 2H), 4.03 (s, 3H).

³¹P-NMR (202 MHz, D₂O) δ: -1.1 (d, *J*_{P-P} = 18.4 Hz), -11.4 (dt, *J*_{P-P} = 19.6, 4.6 Hz), -22.7 (t, *J*_{P-P} = 19.7 Hz).

HRMS (microTOF-Q) *m/z*: [M-H]⁻ detected 585.0143, corresponds to C₁₄H₂₀N₈O₁₂P₃ instead of the expected C₁₄H₁₉N₇O₁₃P₃.

HPLC retention time (semi-preparative): 13.2 min.

tetrasodium ((2*R*,3*S*,4*R*,5*R*)-5-(2-amino-6-(3-methylisoxazol-5-yl)-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate **203¹⁹⁴**



A flame-dried flask was charged with nucleoside **228** (50 mg, 0.09 mmol) and Proton Sponge® (28 mg, 0.13 mmol). The solids were dissolved in trimethyl phosphate (0.4 mL) and cooled to 0 °C. POCl₃ (11 μL, 0.13 mmol) was then added and the solution was stirred at 0 °C under argon for 2 hours. Tributylamine (0.1 mL, 0.44 mmol) was added followed by bis(*tri-n*-butylammonium) pyrophosphate (0.5 M in DMF, 0.88 mL, 0.44 mmol) and the solution was stirred at 0 °C under argon for 10 min. The reaction was quenched with 0.1 M TEAB buffer (pH = 8.5, 2.1 mL). The solution was transferred to a microwave vial and the flask was rinsed with 28% NH₄OH (2 × 8 mL). The vial was crimped and ammonolysis was carried out for 3 hours at 55 °C. The residual ammonia was evaporated *in vacuo* and the residue was purified by ion-exchange chromatography using the method described in the general experimental section. The appropriate fractions were combined and lyophilized. The residue was further purified *via* RP-HPLC using the standard method described in the general experimental section. The appropriate fractions were combined and lyophilized. The residue was dissolved in TEAB (0.1 M, pH = 8.5, 1.3 mL) and cooled to 0 °C. A 0.1 M NaI/acetone solution (6 mL) was added dropwise while shaking. The suspension was centrifuged at 6000 rpm for 10 min. The precipitate was collected, washed with acetone (3 mL), centrifuged at the same speed and collected. The precipitate was lyophilized to afford triphosphate **203** as a white powder (2.5 mg, 4%)

$^1\text{H-NMR}$ (500 MHz, D_2O) δ : 8.47 (s, 1H), 7.21 (s, 1H), 6.03 (d, $J = 6.0$ Hz, 1H), 4.81 (t, $J = 5.6$ Hz, 1H), 4.54 (t, $J = 4.4$ Hz, 1H), 4.34 (triplet, $J = 2.3$ Hz, 1H), 4.24-4.16 (m, 2H), 2.37 (s, 3H).

$^{31}\text{P-NMR}$ (202 MHz, D_2O) δ : -1.0 (d, $J_{\text{P-P}} = 19.9$ Hz), -11.4 (d, $J_{\text{P-P}} = 18.4$ Hz), -22.7 (t, $J_{\text{P-P}} = 19.9$ Hz).

HRMS (micrOTOF-Q) m/z : $[\text{M-H}]^-$ detected 586.0256, corresponds to $\text{C}_{14}\text{H}_{19}\text{N}_7\text{O}_{13}\text{P}_3$ instead of the expected $\text{C}_{14}\text{H}_{18}\text{N}_6\text{O}_{14}\text{P}_3$.

HPLC retention time (semi-preparative): 13.5 min.

Chapter 5

5 Transcriptional Evaluation of C6-functionalized 2-aminopurine Nucleoside Triphosphates

5.1 Introduction

As discussed in Section 1.12.3, the Hirao group developed the unnatural base pair, **s/Pa**. This pair exhibits excellent transcriptional efficiency (92%); however, the structural factors that contribute to this attribute have not been investigated thus far. Consequently, evaluating the effect of various **s** analogues on their transcriptional efficiency and fidelity would establish a structure-activity profile, enabling the development of an improved base pair. In order to perform this SAR study, triphosphates **201 (FpTP)** and **203 (MxTP)** were synthesized (see Section 4.3). These analogues retain the guanosine scaffold present in **s**, but the thiophene has been substituted with an N-methylpyrazole and a methylisoxazole moiety, respectively. These two analogues would further our limited knowledge regarding the tolerance that RNA polymerases have for non-native functional groups.

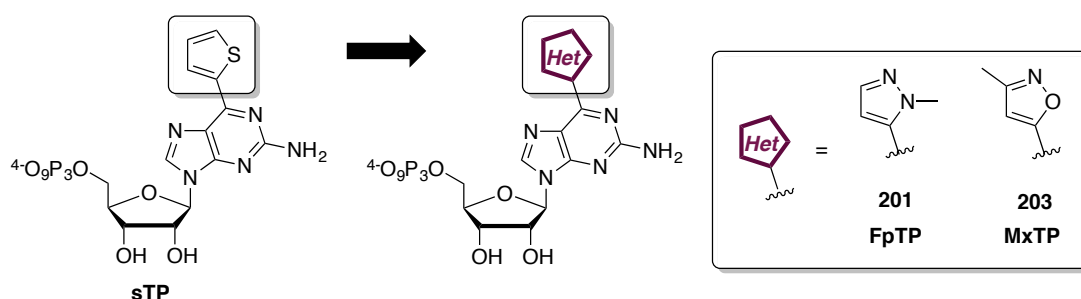


Figure 5-1. Triphosphates **201** and **203** tested in transcription assays. Cations have been omitted for clarity.

A schematic representation of our transcription assay is presented below (Figure 5-2). The **Pa**-modified DNA template was designed to be transcribed from an oligonucleotide containing a T7 RNA polymerase promoter, as in Hirao's assay.²⁰⁶ The full sequences of the DNA strands used in this study are shown in Section 5.5.1; briefly, either an **APaA** or a control **ATA** sequence were included in the template strand. Transcription experiments were carried out with the radioactive nucleotide [α -³²P]GTP, thus enabling quantitation of the transcripts. Full-length transcripts would produce a 17-mer, while termination events occurring just before the **Pa** or after the **Pa** site would produce a 12-mer and 13-mer, respectively. Gel-electrophoresis resolves full-length transcripts from early termination events; analysis of these data allows us to evaluate the incorporation efficiency of our analogues compared to **sTP**.

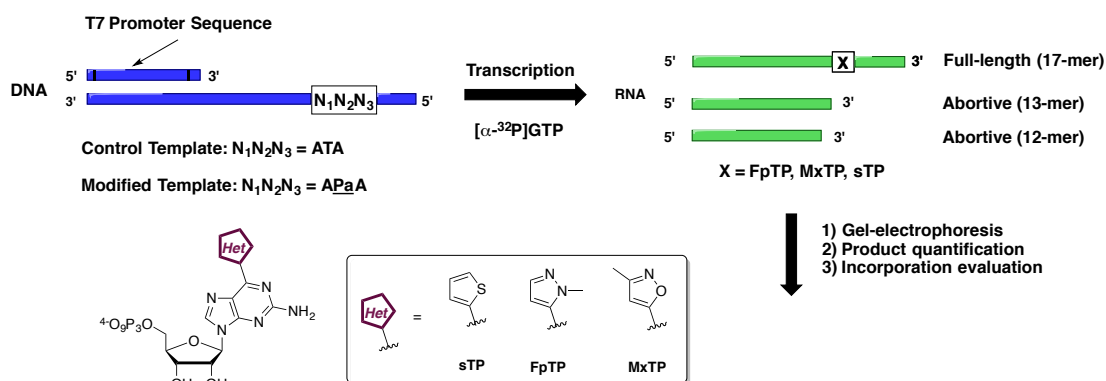


Figure 5-2. Schematic representation of the transcription assay used in this study to evaluate the incorporation efficiency of **sTP**, **MxTP** and **FpTP**.

5.2 Aims of this Study

The aim of this chapter are to:

- 1) Evaluate triphosphates **FpTP** and **MxTP** in transcription opposite a DNA template containing the **Pa** and determine the impact of each moiety in incorporation efficiency.

5.3 Results and Discussion

In order to examine the specificity and efficiency of incorporation of the unnatural triphosphates **FpTP**, **sTP**, and **MxTP**, *in vitro* transcription of both natural and **Pa**-modified template DNA was performed. The resultant transcripts were then visualized by purification on a denaturing polyacrylamide gel (Figure 5-3 and Appendix 127-128). Our measure for specificity is the generation of the full-length transcript for **Pa**-modified template, only when the reaction was supplemented with the unnatural analogue. The efficiency of **Pa** transcription is compared to that of the natural template via the production of full-length RNA.

The successful production of full-length transcripts from the natural template in presence of even up to 2 mM **FpTP** indicates that the unnatural triphosphate does not inhibit transcription (Figure 5-3, lanes 4-8). However, the production of full-length product from the **Pa** template in the absence of **FpTP**, shows that **Pa** does not necessitate specific incorporation of the unnatural analogue (Lane 10). Similar results were observed for **sTP** and **MxTP** (Appendices 127 and 128). This is in contrast to the stalled transcription observed in the absence of UTP or CTP seen for both the templates (Lanes 1, 3, 9, and 11). It is still possible that **Pa** has a preference, if not a strict requirement, for the incorporation of **FpTP**; in order to determine this, we quantified the generation of full-length transcription in the presence of the analogue (lanes 12-16) compared to its absence (lane 10). An increase in the percentage of full-length transcripts upon the addition of our analogues indicates some selectivity for these unnatural triphosphates. The results from this analysis and those performed with **MxTP** and **sTP** are summarized in Chart 5-1.

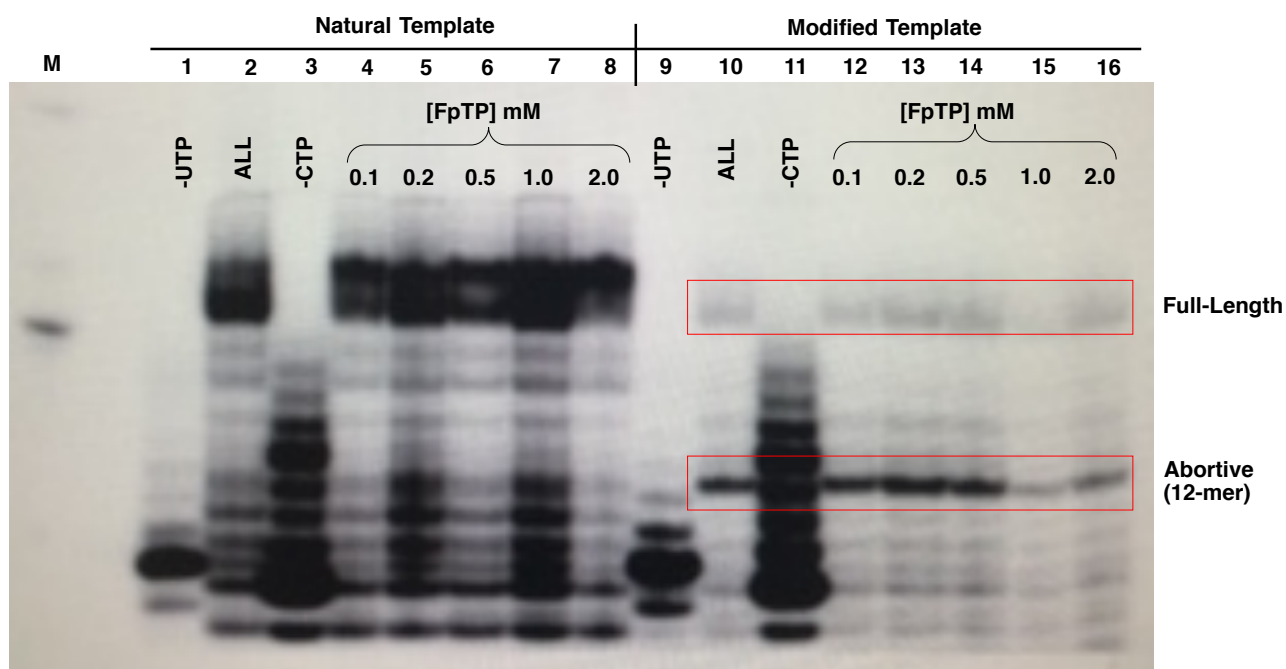


Figure 5-3. Transcription experiments run with FpTP. M = Marker. ALL = All natural nucleotides.

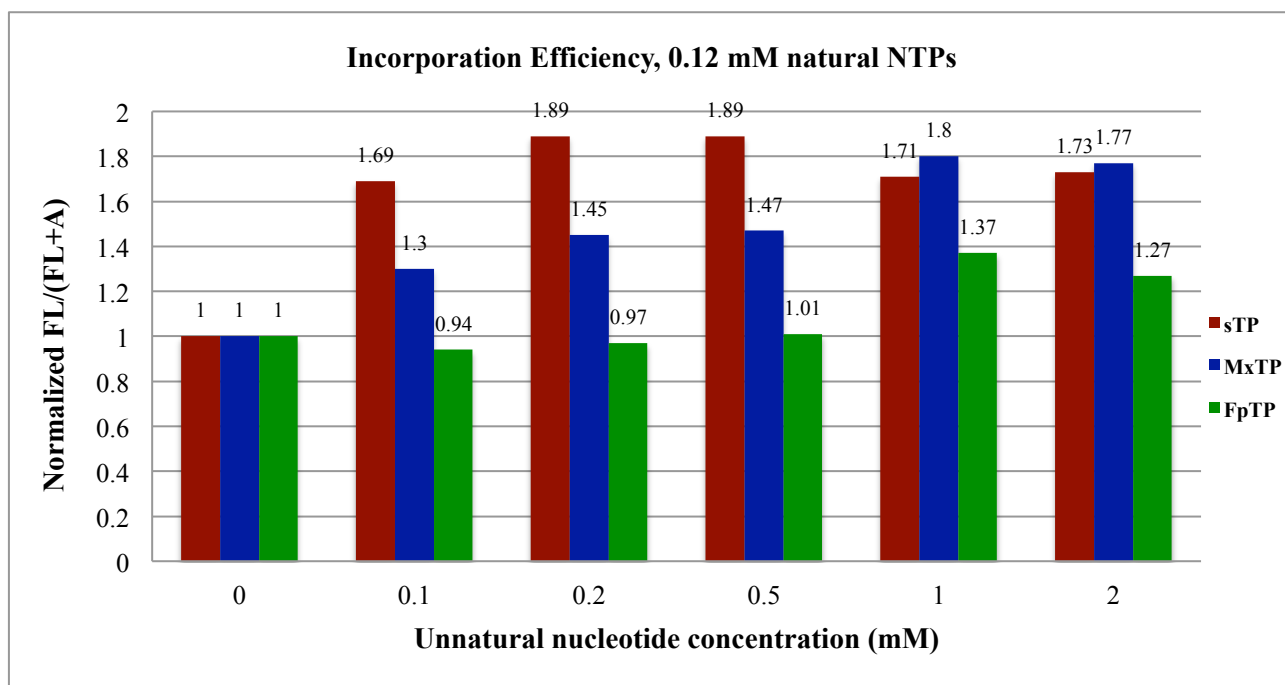


Chart 5-1. Quantification of the transcription results with 0.12 mM natural NTPs. FL = Full-length product. A = Abortive transcript.

FpTP incorporation is observed only at the high concentrations (1 and 2 mM), indicating that the selectivity for this analogue is not particularly strong. On the other hand, sTP is more strongly preferred substrate, as it is incorporated even at the lowest concentration employed (0.1 mM). A similar result is observed for MxTP, although its efficiency is lower than Hirao's sTP at concentrations up to 0.5 mM. These results suggest that the isoxazole functionality is better tolerated than the pyrazole functionality, although the thiophene remains

the gold standard (Figure 5-4). These results also suggest that the sulfur atom has a central role in incorporation efficiency. Substitution of the sulfur atom with oxygen in **MxTP**, as well as the installation of a nitrogen group on the five-membered ring might have resulted in reduced incorporation efficiency. The low incorporation efficiency of **FpTP** might be attributed to the nitrogen atoms present on the pyrazole moiety, as well as the methyl group, which could prevent recognition by **Pa** and/or the polymerase.

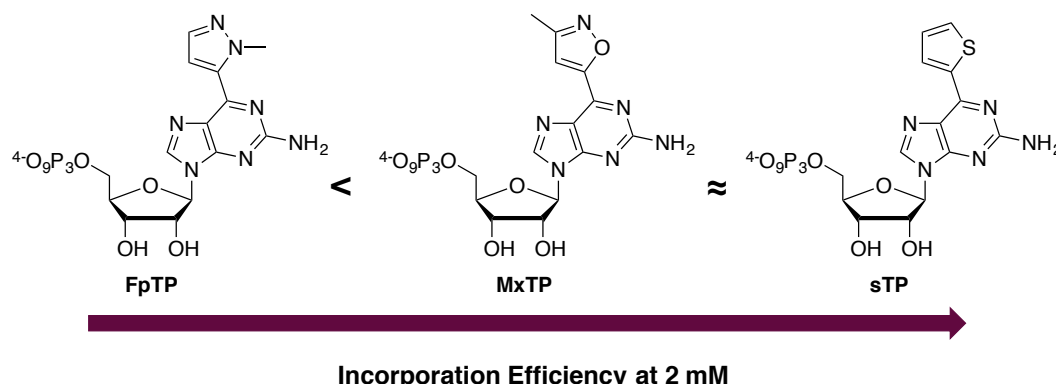


Figure 5-4. Incorporation efficiency scale at 1 and 2 mM concentration of the unnatural nucleotide.

It is well established that RNA polymerases require two divalent metal ions as co-factors.^{207,208} Magnesium salts are commonly used for *in vitro* transcriptions. However, previous studies suggest that MnCl₂ can enhance incorporation efficiency of unnatural nucleotides as compared to the standard MgCl₂.^{209–211} For example, the incorporation of C8-azido ATP was greatly improved by the utilization of 2.5mM MnCl₂ in addition to MgCl₂.²⁰⁹ The reason behind this effect of Mn²⁺ ions is not completely clear. However, it has been suggested that Mn²⁺ ions result in greater conformational flexibility of the active site of the enzyme, thus reducing its discrimination against non-natural substrates.²⁰⁹ The addition of MnCl₂ to the transcription experiments was explored. Since MnCl₂ concentrations higher than 2.5 mM are highly inhibitory to the T7 RNA pol., a low concentration of 0.5 mM MnCl₂ was maintained.²¹⁰ The transcripts generated with **FpTP** are presented below (Figure 5-5).

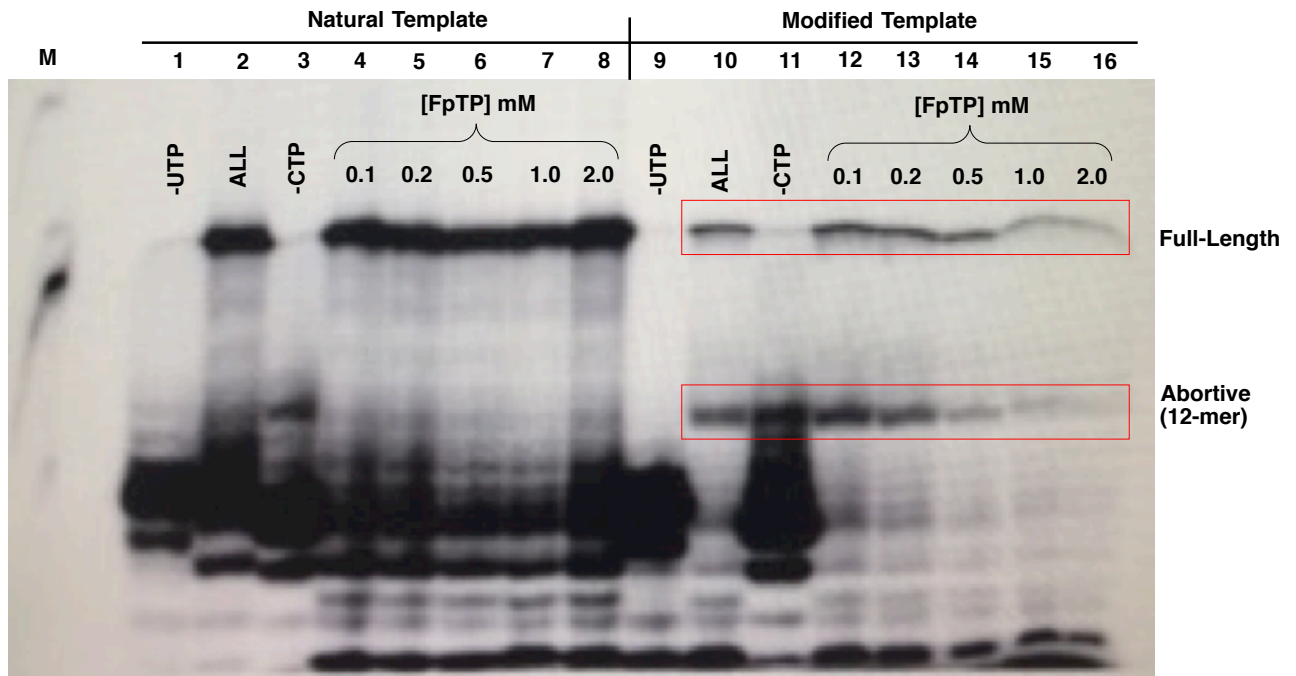


Figure 5-5. Transcription experiments run with **FpTP** in the presence of 0.5 mM MnCl_2 . M = Marker. ALL = All natural NTPs present.

As observed previously, the presence of **FpTP** at concentrations up to 2.0 mM does not inhibit native transcription of the control template with natural nucleotides, as full-length product is observed (Lanes 4-8). Productive transcription is also achieved on the modified template (Figure 5-5: lanes 12-16), unfortunately, we still observed misincorporation of natural nucleotides for the **Pa** template in the absence of **FpTP** (Lane 10). Similar results were observed for **sTP** and **MxTP** (Appendix 129 and 130, respectively). As before, the percent of full-length transcripts in lanes 10 and 12-16 were quantified in order to compare transcriptional efficiency for incorporation of each analogue over natural nucleotides. The results are presented below (Chart 5-2).

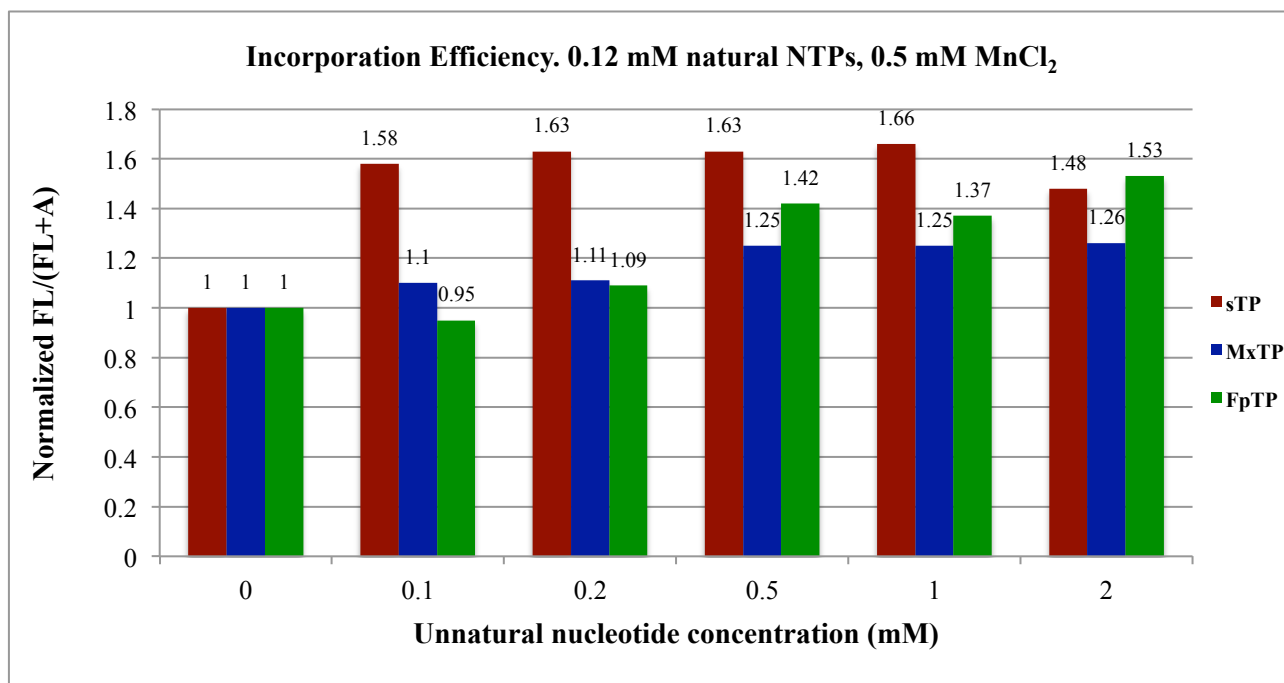


Chart 5-2. Quantification of the transcription results with 0.12 mM natural NTPs and 0.5 MnCl₂. FL = Full-length product. A = Abortive transcript.

In general, **sTP** incorporation remains largely unaffected, although a small decrease was observed in the presence of MnCl₂ (e.g. 1.63 vs. 1.89 at 0.5 mM). The incorporation efficiency of **MxTP** was also slightly reduced by the addition of MnCl₂; for example at 1 mM, a reduction of ~0.6 was observed (1.25 vs. 1.8). Interestingly, enhanced incorporation efficiency was observed for **FpTP** with MnCl₂, and for concentrations >0.5 mM, its efficiency was greater than **MxTP**. This suggests that MnCl₂ foments selectivity between these nucleotides. Complexation of Mn²⁺ with **MxTP** and **sTP** would form complexes **230** and **231**. The presence of a metal ion on the base pairing side of **MxTP** or **sTP** could prevent their recognition by **Pa** and/or the polymerase, thus explaining their reduced incorporation efficiency. Due to the presence of the N-Me group on the pyrazole, chelation of **FpTP** is unlikely, thus allowing its recruitment by the enzyme and recognition by **Pa**, so that it might benefit from the promiscuity afforded by MnCl₂.

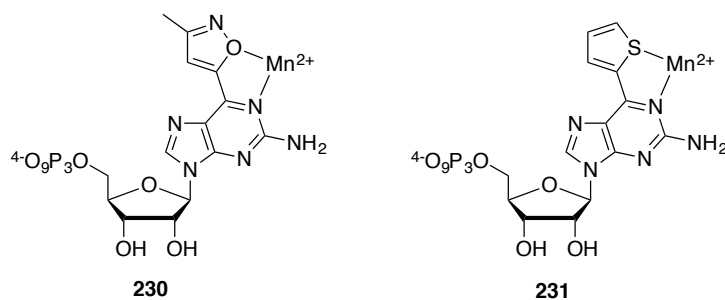


Figure 5-6. Potential coordination complexes **230** and **231**.

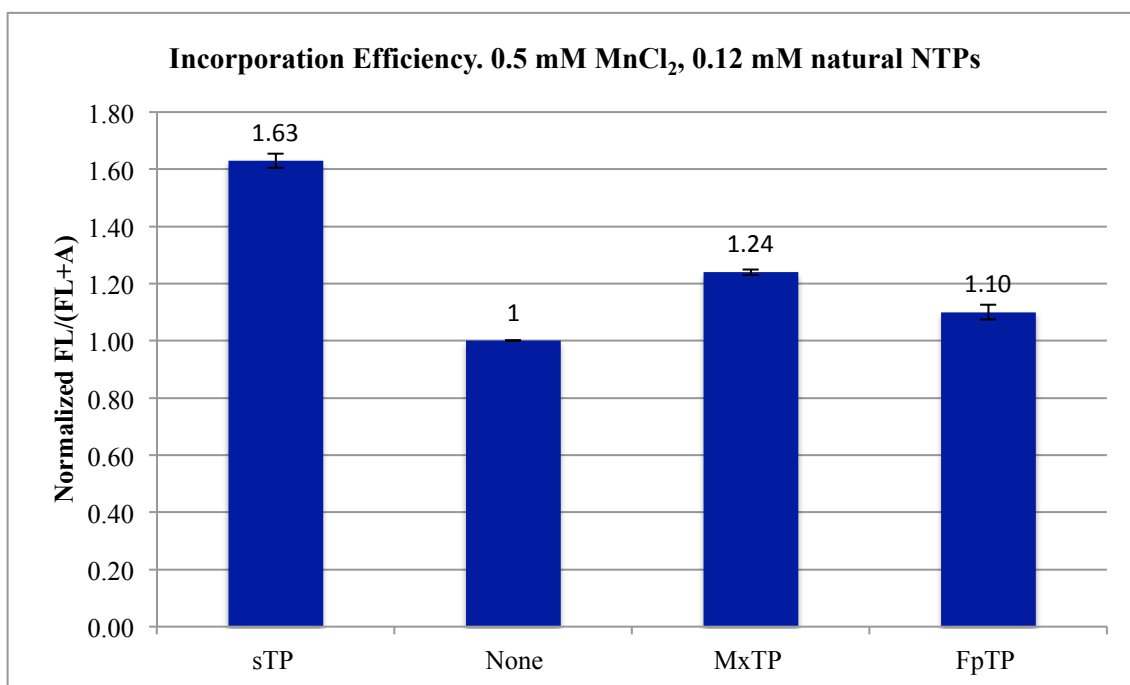


Chart 5-3. Quantification of the transcription results with 0.12 mM natural NTPs, 0.5 mM MnCl_2 and 0.5 mM unnatural nucleotide. Values are averages of three individual experiments. FL = Full-length product. A = Abortive transcript.

Having surveyed several reaction conditions for our analogues, we selected a sampling to perform in triplicate. As utilization of 0.5 mM MnCl_2 had improved **FpTP** incorporation, this was investigated further (Chart 5-3). While the results for **sTP** and **MxTP** remained consistent (1.63 and 1.24, respectively), the incorporation of **FpTP** was not as improved by MnCl_2 as initially perceived (1.10 vs. 1.42). This suggests that the pyrazole moiety has some detrimental effect on incorporation into RNA.

To enhance our understanding of the role of the different functional groups on unnatural analogue incorporation, we further investigated the effects of nucleotide concentration and ionic salt identity on their incorporation. Two concentrations of the unnatural nucleotides were chosen (0.2 mM and 2 mM), and the reactions were then run in the presence and absence of MnCl_2 (Chart 5-4).

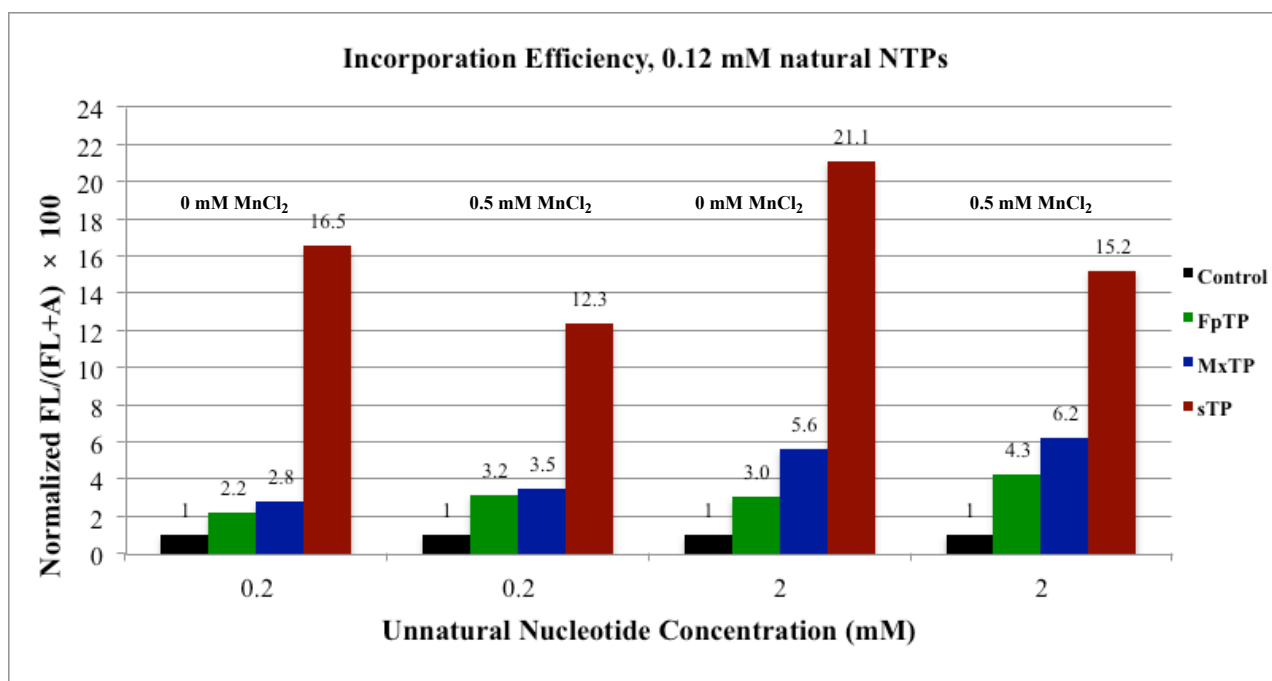


Chart 5-4. Quantification of the transcription results with 0.12 mM natural NTPs. FL = Full-length product. A = Abortive transcript. Control = Reactions run without unnatural NTPs. Values are the average of three individual experiments. Experiments were conducted by Andrew Jobbins and are reprinted with his permission.

Based on this work, neither **FpTP** nor **MxTP** compare to **sTP** transcription efficiency; however, **MxTP** outperforms **FpTP** regardless of the presence of MnCl₂. Interestingly, while MnCl₂ appears to enhance incorporation efficiency for both of our analogues, it was detrimental to incorporation efficiency of **sTP**. This result is in accordance with our previous results, where reduced efficiency was observed upon addition of manganese ions (Charts 5-2 and 5-1). These data further suggest potential chelation of manganese ions to the sulfur atom of **sTP**, resulting in a coordination complex that is not active in transcription. While neither **MxTP** nor **FpTP** matched **sTP**, since the isoxazole moiety appears to be better tolerated compared to the pyrazole functionality, further amelioration of **MxTP** could potentially result in an unnatural triphosphate with efficiency comparable to that of **sTP**.

5.4 Summary

This chapter described the design and execution of a transcriptional assay to evaluate our previously synthesized C6-functionalized 2-aminopurine triphosphates and establish a structure-activity profile correlating C6-substitution and transcriptional efficiency. Our analogues included C6-isoxazole functionalized nucleotide **MxTP** and C6-pyrazole functionalized nucleotide **FpTP**. Key outcomes of this work are the following. Transcriptional efficiency increases with increasing concentration of the unnatural nucleotide. The addition of 0.5 MnCl₂ enhances the incorporation efficiency for **FpTP** and **MxTP** but reduces it for **sTP**, presumably due to chelation of the thiophene moiety of **sTP** to the manganese ions. Finally, the isoxazole functionality is preferable to the pyrazole moiety, likely because it is better tolerated by T7 RNA polymerase.

Post-submission Edit: As stated in section 4.4, HRMS analysis of triphosphates **FpTP** and **MxTP** suggested replacement of a hydroxyl with an amine, thus forming an terminal amidotriphosphate (see Scheme 4-20). Consequently, the reduced incorporation of **MxTP** and **FpTP** observed (compared to commercially obtained **sTP**) can be rationalized by the presence of this unnatural triphosphate moiety. Even if this is the case, incorporation of both analogues was observed. Consequently, this can lead to the hypothesis that the isoxazole and pyrazole moieties are not essentially detrimental to incorporation efficiency.

5.5 Experimental

The DNA sequences used in this study are presented below.

Promoter: 5'-ATAATACGACTCACTATAGGG-3'

Pa-modified template: 3'-TATTATGCTGAGTGATATCCCTCGAAGGGAPaACTC-5'

Unmodified template (control): 3'-TATTATGCTGAGTGATATCCCTCGAAGGGATACTC-5'

The promoter and **Pa**-modified template were bought from BioTeZ Berlin-Buch GmbH. The unmodified template was bought from Eurofins UK. All DNA strands were used without further purification. **MxTP** and **FpTP** were synthesized as described in Chapter 4, dissolved in ultrapure water and stored as 10 mM stock solutions. **sTP** was bought from Glen Research as a 10 mM solution in ultrapure water. The 10 mM stock solutions were then used to make stock solutions of 5, 2.5, 1 and 0.5 mM for each unnatural nucleotide. ATP, CTP and UTP were purchased from Promega. [α - 32 P]GTP was purchased from Perkin-Elmer. T7 RNA polymerase was made in-house by the Eperon group.

Annealing of the promoter with each template was carried out in a buffer containing 10mM HEPES (pH = 8.0) and 10mM NaCl, by heating at 95 °C and slow cooling to 4 °C.

Transcription was performed in a reaction buffer (2 μ L) containing 80 mM HEPES buffer (pH 8.0), 8 mM MgCl₂, 0.5 mM or 0 mM MnCl₂ 10 mM DTT, 5 U RNase OUT™ and 0.05% NP40 in the presence of 0.12 mM natural NTPs, 0.5 μ Ci [α - 32 P]GTP, 2 μ M DNA and 0.1 μ L of T7 RNA polymerase. Experiments were performed as follows. Each template was transcribed under 8 conditions/nucleotide analogue. Two reactions lacking either UTP or CTP forced the formation of stalled intermediates and served as negative controls. One reaction contained a mixture of ATP/CTP/[α - 32 P]GTP/UTP as either the positive control (for the unmodified template) or to indicate misincorporation of a natural nucleotide (for the modified template). Five reactions contained the natural triphosphates plus the unnatural triphosphate in increasing concentrations (0.1, 0.2, 0.5 1.0 and 2.0 mM). For the unmodified template, this would indicate if the unnatural triphosphate is inhibiting the polymerase. For the **Pa**-modified template, it would indicate if incorporation occurs. After incubation at 37 °C for 30 min, the reaction was quenched by the addition of an

E dye solution (9:1 formamide:EDTA mixture containing small amounts of xylene cyanol and bromophenol blue) and the products were analyzed on a 15% polyacrylamide-7.0 M urea gel. Visualization was carried out by exposing the gel to an X-ray film for 16 h and scanning the film on a Perkin-Elmer phosphorimager. The data were analyzed with the Perkin-Elmer OptiQuant software.

6 Conclusions

The flexible nature of RNA enables it to adopt a variety of complicated secondary and tertiary structures. This complexity is important for various biological processes, such as the process of alternative RNA splicing. Furthermore, splice site selection is dictated by mechanisms involving various proteins that recognize key sequences with a pre-mRNA molecule. The ability to site-specifically label key RNA sequences that can dictate splice-site selection within a long RNA molecule is important in order to elucidate mechanistic aspects of alternative RNA splicing. However, most efforts towards efficient site-specific RNA labeling have mainly focused on enzymatic methods necessitating the isolation of various enzymes. Recently, the concept of orthogonal base pairs has been applied for site-specific RNA labeling, but most methods have focused on labeling at a single site, or employ unnatural triphosphates that require laborious syntheses.

The work presented in this thesis has led to the following conclusions:

- 1) The **Z/P** base pair developed by the Benner group is not a suitable candidate for site-specific labeling due to PCR experiments indicating significant mismatch of **Z** with **G**. Furthermore, the synthesis of the ribonucleoside version of **Z** proved to be problematic, either by employing our proposed strategy or Benner's method.¹⁴⁷
- 2) A robust method to access C6-functionalized analogues of the Hirao **s** nucleoside was developed, enabling the synthesis of a library of pyrazole, triazole and isoxazole functionalized analogues in three or six steps.
- 3) This strategy was modified in order to enable the synthesis of triphosphate analogues suitable for transcription studies. Pyrazole functionalized analogue **FpTP** and isoxazole functionalized analogue **MxTP** were synthesized in five and eight steps respectively.
- 4) A transcriptional assay for the evaluation of **FpTP**, **MxTP** and Hirao's **sTP** was carried out. The data acquired suggested that the thiophene moiety results in the best incorporation efficiency, followed by the isoxazole moiety and the pyrazole moiety. The addition of $MnCl_2$ increased the incorporation efficiency for **FpTP** and **MxTP** and reduced it for **sTP**, potentially due to chelation of the thiophene moiety to the manganese ions.

7 References

- (1) Fica, S. M.; Tuttle, N.; Novak, T.; Li, N.-S.; Lu, J.; Koodathingal, P.; Dai, Q.; Staley, J. P.; Piccirilli, J. A. *Nature* **2013**, *503* (7475), 229–234.
- (2) Lilley, D. M. J. *Philos. Trans. R. Soc. B Biol. Sci.* **2011**, *366* (1580), 2910–2917.
- (3) Graveley, B. R. *Trends Genet* **2001**, *17*.
- (4) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30* (16), 3497–3531.
- (5) U2 snRNA https://en.wikipedia.org/wiki/U2_spliceosomal_RNA (accessed Dec 6, 2013).
- (6) Huang, L.; Ishibe-Murakami, S.; Patel, D. J.; Serganov, A. *Proc. Natl. Acad. Sci.* **2011**, *108* (36), 14801–14806.
- (7) von Ahsen, U.; Green, R.; Schroeder, R.; Noller, H. F. *RNA* **1997**, *3* (1), 49–56.
- (8) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci.* **2001**, *98* (9), 4899–4903.
- (9) A-Motif http://eternawiki.org/wiki/index.php5/User:ElNando888/Blog/A_minor_glitch%253F (accessed Aug 19, 2016).
- (10) Galej, W. P.; Wilkinson, M. E.; Fica, S. M.; Oubridge, C.; Newman, A. J.; Nagai, K. *Nature* **2016**, *advance on*.
- (11) Ramanathan, A.; Robb, G. B.; Chan, S.-H. *Nucleic Acids Res.* **2016**, *44* (16), 7511–7526.
- (12) Shuman, S. *Nat Rev Mol Cell Biol* **2002**, *3* (8), 619–625.
- (13) Proudfoot, N. J. *Genes Dev.* **2011**, *25* (17), 1770–1782.
- (14) Lutz, C. S.; Moreira, A. *Wiley Interdiscip. Rev. RNA* **2011**, *2* (1), 23–31.
- (15) Kühn, U.; Gündel, M.; Knoth, A.; Kerwitz, Y.; Rüdell, S.; Wahle, E. *J. Biol. Chem.* **2009**, *284* (34), 22803–22814.
- (16) Colgan, D. F.; Manley, J. L. *Genes Dev.* **1997**, *11* (21), 2755–2766.
- (17) Tian, B.; Hu, J.; Zhang, H.; Lutz, C. S. *Nucleic Acids Res.* **2005**, *33* (1), 201–212.
- (18) Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. *Science* **2008**, *320* (5881), 1344–1349.
- (19) Gruber, A. R.; Martin, G.; Keller, W.; Zavolan, M. *Wiley Interdiscip. Rev. RNA* **2014**, *5* (2), 183–196.
- (20) Lutz, C. S. *ACS Chem. Biol.* **2008**, *3* (10), 609–617.
- (21) Brennicke, A.; Marchfelder, A.; Binder, S. *FEMS Microbiol. Rev.* **1999**, *23* (3), 297 LP-316.
- (22) Kawahara, Y.; Megraw, M.; Kreider, E.; Iizasa, H.; Valente, L.; Hatzigeorgiou, A. G.; Nishikura, K. *Nucleic Acids Res.* **2008**, *36* (16), 5270–5280.
- (23) Maas, S.; Rich, A. *BioEssays* **2000**, *22* (9), 790–802.
- (24) Berg, Jeremy Mark; Tymocko, John L.; Stryer, L. *Biochemistry*, 5th ed.; W H Freeman % Co Ltd., 2002.
- (25) Chorev, M.; Carmel, L. *Front. Genet.* **2012**, *3*, 55.
- (26) Faustino, N. A.; Cooper, T. A. *Genes Dev* **2003**, *17*.

- (27) Cech, T. R. *Annu. Rev. Biochem.* **1990**, *59* (1), 543–568.
- (28) Lührmann, R. Cellular Biochemistry <http://www.mpibpc.mpg.de/luehrmann> (accessed Jul 27, 2013).
- (29) Matera, A. G.; Wang, Z. *Nat Rev Mol Cell Biol* **2014**, *15* (2), 108–121.
- (30) Will, C. L.; Lührmann, R. *Cold Spring Harb. Perspect. Biol.* **2011**, *3* (7).
- (31) Madhani, H. D.; Guthrie, C. *Cell* **2016**, *71* (5), 803–817.
- (32) Sashital, D. G.; Cornilescu, G.; Butcher, S. E. *Nat Struct Mol Biol* **2004**, *11* (12), 1237–1242.
- (33) Hilliker, A. K.; Mefford, M. A.; Staley, J. P. *Genes Dev.* **2007**, *21* (7), 821–834.
- (34) Perriman, R. J.; Ares, M. *Genes Dev.* **2007**, *21* (7), 811–820.
- (35) Teigelkamp, S.; Newman, A. J.; Beggs, J. D. *EMBO J.* **1995**, *14* (11), 2602–2612.
- (36) Keefe, R. T.; Newman, A. J. *EMBO J.* **1998**, *17* (2), 565 LP-574.
- (37) Turunen, J. J.; Niemelä, E. H.; Verma, B.; Frilander, M. J. *Wiley Interdiscip. Rev. RNA* **2013**, *4* (1), 61–76.
- (38) Wang, E. T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S. F.; Schroth, G. P.; Burge, C. B. *Nature* **2008**, *456* (7221), 470–476.
- (39) Black, D. L. *Annu. Rev. Biochem.* **2003**, *72* (1), 291–336.
- (40) Xing, Y.; Lee, C. *Nat Rev Genet* **2006**, *7* (7), 499–509.
- (41) Matlin, A. J.; Clark, F.; Smith, C. W. J. *Nat Rev Mol Cell Biol* **2005**, *6* (5), 386–398.
- (42) Keren, H. Lev-Maor, G., A. G. *Nat. Rev. Genet.* **2010**, *11* (5), 345–355.
- (43) Alekseyenko, A. V.; Kim, N.; Lee, C. J. *RNA* **2007**, *13* (5), 661–670.
- (44) Sugnet, C. W.; Kent, W. J.; Ares, M.; Haussler, D. Altman, R. B., Dunker, A. K., Hunter, L., Jung, T. A., Klein, T. E., Eds.; World Scientific: Singapore, 2004.
- (45) Yeo, G.; Holste, D.; Kreiman, G.; Burge, C. B. *Genome Biol.* **2004**, *5* (10), 1–15.
- (46) Grabowski, P. J.; Black, D. L. *Prog. Neurobiol.* **2001**, *65* (3), 289–308.
- (47) Lefebvre, S.; Bürglen, L.; Reboullet, S.; Clermont, O.; Burlet, P.; Viollet, L.; Benichou, B.; Cruaud, C.; Millasseau, P.; Zeviani, M.; Le Paslier, D.; Frézal, J.; Cohen, D.; Weissenbach, J.; Munnich, A.; Melki, J. *Cell* **2016**, *80* (1), 155–165.
- (48) Lorson, C. L.; Hahnen, E.; Androphy, E. J.; Wirth, B. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (11), 6307–6311.
- (49) Monani, U. R.; Lorson, C. L.; Parsons, D. W.; Prior, T. W.; Androphy, E. J.; Burghes, A. H. M.; McPherson, J. D. *Hum. Mol. Genet.* **1999**, *8* (7), 1177–1183.
- (50) Gubitza, A. K.; Feng, W.; Dreyfuss, G. *Exp. Cell Res.* **2004**, *296* (1), 51–56.
- (51) Zhang, Z.; Lotti, F.; Dittmar, K.; Younis, I.; Wan, L.; Kasim, M.; Dreyfuss, G. *Cell* **2008**, *133* (4), 585–600.
- (52) Rino, J.; Martin, R. M.; Carvalho, T.; Carmo-Fonseca, M. *Methods* **2014**, *65* (3), 359–366.
- (53) Burghes, A. H. M.; Beattie, C. E. *Nat Rev Neurosci* **2009**, *10* (8), 597–609.
- (54) Eriksson, M.; Brown, W. T.; Gordon, L. B.; Glynn, M. W.; Singer, J.; Scott, L.; Erdos, M. R.; Robbins, C. M.; Moses, T. Y.; Berglund, P.; Dutra, A.; Pak, E.; Durkin, S.; Csoka, A. B.; Boehnke,

- M.; Glover, T. W.; Collins, F. S. *Nature* **2003**, *423* (6937), 293–298.
- (55) Gordon, Leslie B; Brown, Ted W; Collins, F. S. *GeneReviews*; Pagon RA, Adam MP, Ardinger HH, et al., E., Ed.; University of Washington, Seattle, 2003.
- (56) Scaffidi, P.; Gordon, L.; Misteli, T. *PLoS Biol* **2005**, *3* (11), e395.
- (57) Cooper, T. A.; Mattox, W. *Am. J. Hum. Genet.* **1997**, *61* (2), 259–266.
- (58) Zahler, A. M.; Neugebauer, K. M.; Lane, W. S.; Roth, M. B. *Science (80-.)*. **1993**, *260* (5105), 219 LP-222.
- (59) Long, J. C.; Cáceres, J. F. *Biochem. J.* **2008**, *417* (1), 15 LP-27.
- (60) Graveley, B. R.; Hertel, K. J.; Maniatis, T. *RNA* **2001**, *7* (6), 806–818.
- (61) Zhu, J.; Mayeda, A.; Krainer, A. R. *Mol. Cell* **2001**, *8* (6), 1351–1361.
- (62) Wu, J. Y.; Maniatis, T. *Cell* **2016**, *75* (6), 1061–1070.
- (63) Sreaton, G. R.; Cáceres, J. F.; Mayeda, A.; Bell, M. V; Plebanski, M.; Jackson, D. G.; Bell, J. I.; Krainer, A. R. *EMBO J.* **1995**, *14* (17), 4336–4349.
- (64) Llewellyn, D. H.; Scobie, G. A.; Urquhart, A. J.; Whatley, S. D.; Roberts, A. G.; Harrison, P. R.; Elder, G. H. *J. Med. Genet.* **1996**, *33* (5), 437–438.
- (65) Jin, Y.; Dietz, H. C.; Montgomery, R. A.; Bell, W. R.; McIntosh, I.; Collier, B.; Bray, P. F. *J. Clin. Invest.* **98** (8), 1745–1754.
- (66) Goren, A.; Ram, O.; Amit, M.; Keren, H.; Lev-Maor, G.; Vig, I.; Pupko, T.; Ast, G. *Mol. Cell* **2016**, *22* (6), 769–781.
- (67) Sharma, S.; Falick, A. M.; Black, D. L. *Mol. Cell* **2005**, *19* (4), 485–496.
- (68) Krammer, P. H. *Nature* **2000**, *407* (6805), 789–795.
- (69) David, C. J.; Manley, J. L. *Genes Dev.* **2010**, *24* (21), 2343–2364.
- (70) Izquierdo, J. M.; Majós, N.; Bonnal, S.; Martínez, C.; Castelo, R.; Guigó, R.; Bilbao, D.; Valcárcel, J. *Mol. Cell* **2005**, *19* (4), 475–484.
- (71) Kondera-Anasz, Z.; Mielczarek-Palacz, A.; Sikora, J. *Apoptosis* **2005**, *10* (5), 1143–1149.
- (72) Sheen-Chen, S.-M.; Chen, H.-S.; Eng, H.-L.; Chen, W.-J. *World J. Surg.* **2003**, *27* (1), 10–13.
- (73) Cheng, J.; Zhou, T.; Liu, C.; Shapiro, J. P.; Brauer, M. J.; Kiefer, M. C.; Barr, P. J.; Mountz, J. D. *Science (80-.)*. **1994**, *263* (5154), 1759 LP-1762.
- (74) Tejedor, J. R.; Papasaikas, P.; Valcárcel, J. *Mol. Cell* **2015**, *57* (1), 23–38.
- (75) Saxon, E.; Bertozzi, C. R. *Science (80-.)*. **2000**, *287* (5460), 2007 LP-2010.
- (76) Staudinger, H.; Meyer, J. *Helv. Chim. Acta* **1919**, *2* (1), 635–646.
- (77) Saxon, E.; Luchansky, S. J.; Hang, H. C.; Yu, C.; Lee, S. C.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2002**, *124* (50), 14893–14902.
- (78) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046–15047.
- (79) Kim, Y.; Ho, S. O.; Gassman, N. R.; Korlann, Y.; Landorf, E. V; Collart, F. R.; Weiss, S. *Bioconjug. Chem.* **2008**, *19* (3), 786–791.
- (80) Toseland, C. P. *J. Chem. Biol.* **2013**, *6* (3), 85–95.

- (81) Brewer, C. F.; Riehm, J. P. *Anal. Biochem.* **1967**, *18* (2), 248–255.
- (82) Sulfhydryl-reactive Crosslinker Chemistry <https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/sulfhydryl-reactive-crosslinker-chemistry.html> (accessed Sep 27, 2016).
- (83) Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67* (9), 3057–3064.
- (84) Sharpless, B. K., Finn, M. G. and Kolb, H. C. *Angew. Chem. Int. Ed. Engl.* **2001**, *40* (11), 2004–2021.
- (85) Song, W.; Wang, Y.; Qu, J.; Lin, Q. *J. Am. Chem. Soc.* **2008**, *130* (30), 9654–9655.
- (86) van Amstel, J. K. P.; Bergman, A. J. I. W.; van Beurden, E. A. C. M.; Roijers, J. F. M.; Peelen, T.; van den Berg, I. E. T.; Poll-The, B. T.; Kvittingen, E. A.; Berger, R. *Hum. Genet.* **1996**, *97* (1), 51–59.
- (87) Blanco, A. M.; Rausell, L.; Aguado, B.; Perez-Alonso, M.; Artero, R. *Nucleic Acids Res.* **2009**, *37* (17), e116–e116.
- (88) Qin, P. Z.; Pyle, A. M. *Methods* **1999**, *18* (1), 60–70.
- (89) Chu, B. C.; Wahl, G. M.; Orgel, L. E. *Nucleic Acids Res.* **1983**, *11* (18), 6513–6529.
- (90) Stephenson, M. L.; Zamecnik, P. C. In *Nucleic Acids, Part A*; Enzymology, B. T.-M. in, Ed.; Academic Press, 1967; Vol. Volume 12, pp 670–678.
- (91) Griffin Jr, E. A.; Qin, Z.; Michels Jr, W. J.; Pyle, A. M. *Chem. Biol.* **2016**, *2* (11), 761–770.
- (92) Moore, M. J.; Sharp, P. A. *Sci.* **1992**, *256* (5059), 992–997.
- (93) Stark, M. R.; Pleiss, J. A.; Deras, M.; Scaringe, S. A.; Rader, S. D. *RNA* **2006**, *12* (11), 2014–2019.
- (94) Scaringe, S. A.; Wincott, F. E.; Caruthers, M. H. *J. Am. Chem. Soc.* **1998**, *120* (45), 11820–11821.
- (95) Abelson, J.; Blanco, M.; Ditzler, M. A.; Fuller, F.; Aravamudhan, P.; Wood, M.; Villa, T.; Ryan, D. E.; Pleiss, J. A.; Maeder, C.; Guthrie, C.; Walter, N. G. *Nat Struct Mol Biol* **2010**, *17* (4), 504–512.
- (96) Abelson, J.; Hadjivassiliou, H.; Guthrie, C. In *Single Molecule Tools: Fluorescence Based Approaches, Part A*; Enzymology, N. G. W. B. T.-M. in, Ed.; Academic Press, 2010; Vol. Volume 472, pp 31–40.
- (97) SMITH, G. J.; SOSNICK, T. R.; SCHERER, N. F.; PAN, T. A. O. *RNA* **2005**, *11* (2), 234–239.
- (98) Baum, D. A.; Silverman, S. K. *Angew. Chemie Int. Ed.* **2007**, *46* (19), 3502–3504.
- (99) Zelin, E.; Wang, Y.; Silverman, S. K. *Biochemistry* **2006**, *45* (9), 2767–2771.
- (100) Höbartner, C.; Silverman, S. K. *Angew. Chemie Int. Ed.* **2007**, *46* (39), 7420–7424.
- (101) Büttner, L.; Javadi-Zarnaghi, F.; Höbartner, C. *J. Am. Chem. Soc.* **2014**, *136* (22), 8131–8137.
- (102) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111* (21), 8322–8323.
- (103) Rappaport, H. P. *Nucleic Acids Res.* **1988**, *16* (15), 7253–7267.
- (104) Joyce, C. M.; Steitz, T. A. *Annu. Rev. Biochem.* **1994**, *63* (1), 777–822.
- (105) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73* (3), 804–808.
- (106) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32* (39), 10489–10496.
- (107) Piccirilli, J. A.; Benner, S. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343* (6253),

- 33–37.
- (108) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci.* **1995**, *92* (14), 6329–6333.
- (109) Hutter, D.; Benner, S. A. *J. Org. Chem.* **2003**, *68* (25), 9839–9842.
- (110) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133* (38), 15105–15112.
- (111) Kim, H.-J.; Chen, F.; Benner, S. A. *J. Org. Chem.* **2012**, *77* (7), 3664–3669.
- (112) Matray, T. J.; Kool, E. T. *J. Am. Chem. Soc.* **1998**, *120* (24), 6191–6192.
- (113) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399* (6737), 704–708.
- (114) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121* (10), 2323–2324.
- (115) Ishikawa, M.; Hirao, I.; Yokoyama, S. *Tetrahedron Lett.* **2000**, *41* (20), 3931–3934.
- (116) Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2001**, *11* (16), 2221–2223.
- (117) Mitsui, T.; Kimoto, M.; Harada, Y.; Yokoyama, S.; Hirao, I. *J. Am. Chem. Soc.* **2005**, *127* (24), 8652–8658.
- (118) Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125* (18), 5298–5307.
- (119) Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S. *Nat Meth* **2006**, *3* (9), 729–735.
- (120) Hirao, I.; Mitsui, T.; Kimoto, M.; Yokoyama, S. *J. Am. Chem. Soc.* **2007**, *129* (50), 15549–15555.
- (121) Kimoto, M.; Kawai, R.; Mitsui, T.; Yokoyama, S.; Hirao, I. *Nucleic Acids Res.* **2009**, *37* (2), e14–e14.
- (122) Kimoto, Michiko; Yamashige, Rie; Yokoyama, Shigeyuki and Hirao, I. *J. Nucleic Acids* **2012**, *2012* (Article ID: 230943), 8 pages.
- (123) Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E. *J. Am. Chem. Soc.* **2008**, *130* (7), 2336–2343.
- (124) Seo, Y. J.; Matsuda, S.; Romesberg, F. E. *J Am Chem Soc* **2009**, *131* (14), 5046–5047.
- (125) Lavergne, T.; Degardin, M.; Malyshev, D. A.; Quach, H. T.; Dhimi, K.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.* **2013**, *135* (14), 5408–5419.
- (126) Li, L.; Degardin, M.; Lavergne, T.; Malyshev, D. A.; Dhimi, K.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.* **2014**, *136* (3), 826–829.
- (127) Malyshev, D. A.; Dhimi, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R.; Romesberg, F. E. *Nature* **2014**, *509* (7500), 385–388.
- (128) Somoza, A. *Chem. Soc. Rev.* **2008**, *37* (12), 2668–2675.
- (129) Hollenstein, M. *Molecules* . 2012.
- (130) Kawai, R.; Kimoto, M.; Ikeda, S.; Mitsui, T.; Endo, M.; Yokoyama, S.; Hirao, I. *J. Am. Chem. Soc.* **2005**, *127* (49), 17286–17295.
- (131) Seo, Y. J.; Malyshev, D. A.; Lavergne, T.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.*

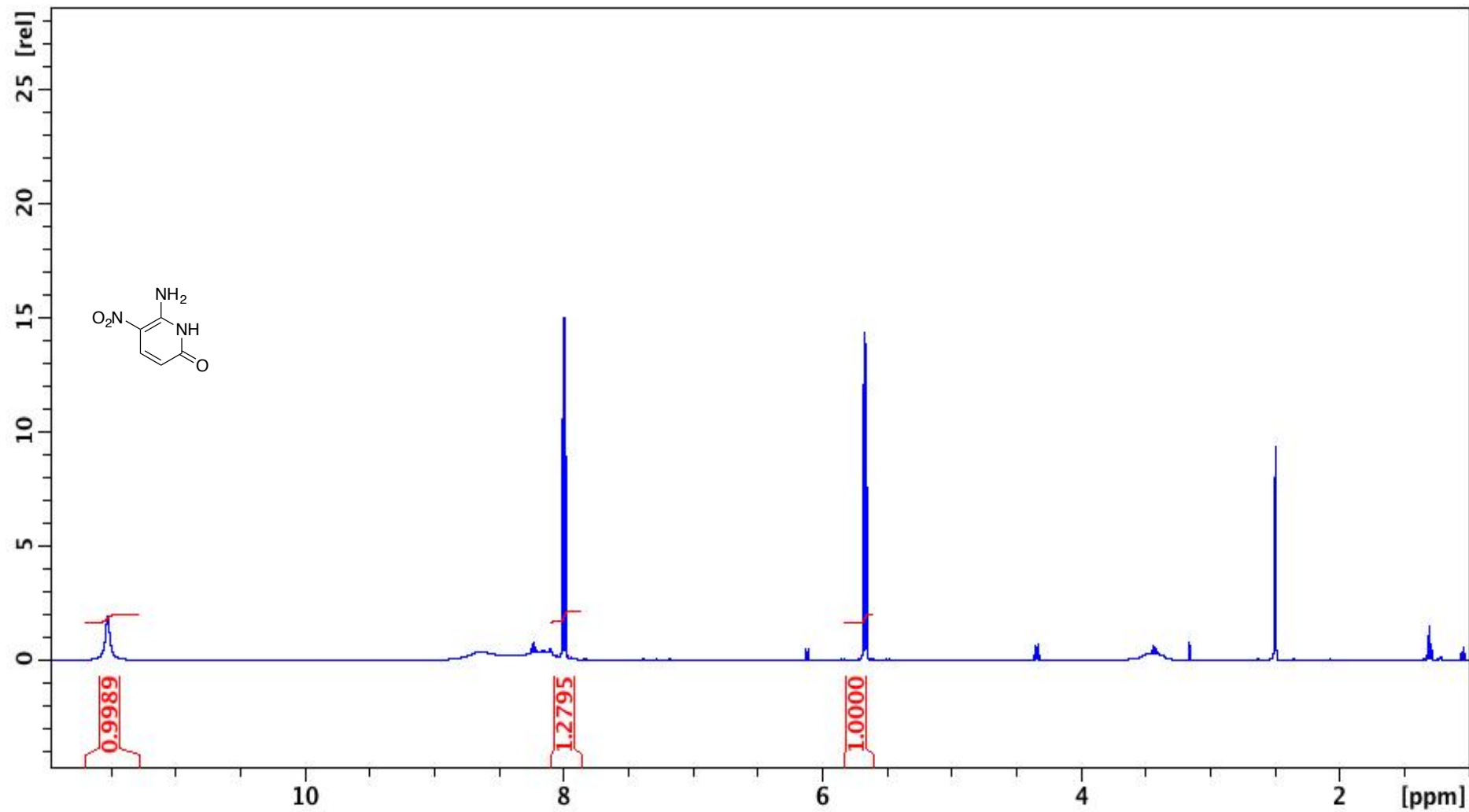
- 2011**, *133* (49), 19878–19888.
- (132) Ishizuka, T.; Kimoto, M.; Sato, A.; Hirao, I. *Chem. Commun.* **2012**, *48* (88), 10835–10837.
- (133) Lavergne, T.; Lamichhane, R.; Malyshev, D. A.; Li, Z.; Li, L.; Sperling, E.; Williamson, J. R.; Millar, D. P.; Romesberg, F. E. *ACS Chem. Biol.* **2016**, *11* (5), 1347–1353.
- (134) Broussard, J. A.; Rappaz, B.; Webb, D. J.; Brown, C. M. *Nat. Protoc.* **2013**, *8* (2), 265–281.
- (135) Voegel, J. J.; Benner, S. A. *J. Am. Chem. Soc.* **1994**, *116* (15), 6929–6930.
- (136) Hocek, Michal; Stambasky, Jan; Kocovsky, P. *Chem. Rev.* **2009**, *109* (12), 6729–6764.
- (137) Wellington, K. W.; Benner, S. A. *Nucleosides, Nucleotides and Nucleic Acids* **2006**, *25* (12), 1309–1333.
- (138) Heck, R. F. *J. Am. Chem. Soc.* **1969**, *91* (24), 6707–6714.
- (139) Heck Reaction <http://www.name-reaction.com/heck-reaction> (accessed Sep 27, 2016).
- (140) Erik Larsen, Per T. Jørgensen, Mamdouh A. Sofan, E. B. P. *Synthesis (Stuttg.)* **1994**, *1994* (10), 1037–1038.
- (141) Zhang, H. C.; Daves, G. D. *J. Org. Chem.* **1993**, *58* (9), 2557–2560.
- (142) Cheng, J. C. Y.; Hacksell, U.; Daves, G. D. *J. Org. Chem.* **1986**, *51* (16), 3093–3098.
- (143) Matulic-Adamic, J.; Beigelman, L.; Portmann, S.; Egli, M.; Usman, N. *J. Org. Chem.* **1996**, *61* (11), 3909–3911.
- (144) Guianvarc'h, D.; Fourrey, J.-L.; Tran Huu Dau, M.-E.; Guérineau, V.; Benhida, R. *J. Org. Chem.* **2002**, *67* (11), 3724–3732.
- (145) Piccirilli, J. A.; Krauch, T.; MacPherson, L. J.; Benner, S. A. *Helv. Chim. Acta* **1991**, *74* (2), 397–406.
- (146) Matulic-adamic, J.; Beigeiman, L. **1997**, *38* (10), 1669–1672.
- (147) Kim, H.-J.; Leal, N. A.; Hoshika, S.; Benner, S. A. *J. Org. Chem.* **2014**, *79* (7), 3194–3199.
- (148) Joubert, N.; Pohl, R.; Klepetářová, B.; Hocek, M. *J. Org. Chem.* **2007**, *72* (18), 6797–6805.
- (149) Takle, A.; Kocięński, P. *Tetrahedron Lett.* **1989**, *30* (13), 1675–1678.
- (150) Jacobsen, E. N.; Marko, I.; Mungall, W. S.; Schroeder, G.; Sharpless, K. B. *J. Am. Chem. Soc.* **1988**, *110* (6), 1968–1970.
- (151) VanRheenen, V.; Kelly, R. C.; Cha, D. Y. *Tetrahedron Lett.* **1976**, *17* (23), 1973–1976.
- (152) Paulo Roberto Rodrigues Meira, Angélica Venturini Moro, C. R. D. C. *Synthesis (Stuttg.)* **2007**, *2007* (15), 2279–2286.
- (153) Spivey, A. C.; Hanson, R.; Scoria, N.; Thorpe, S. J. *J. Chem. Educ.* **1999**, *76* (5), 655.
- (154) Lan, T.; McLaughlin, L. W. *Bioorg. Chem.* **2001**, *29* (4), 198–210.
- (155) Gamble, A. B.; Garner, J.; Gordon, C. P.; O'Conner, S. M. J.; Keller, P. A. *Synth. Commun.* **2007**, *37* (16), 2777–2786.
- (156) Kimoto, M.; Mitsui, T.; Yamashige, R.; Sato, A.; Yokoyama, S.; Hirao, I. *J. Am. Chem. Soc.* **2010**, *132* (43), 15418–15426.
- (157) Someya, T.; Ando, A.; Kimoto, M.; Hirao, I. *Nucleic Acids Res.* **2015**.

- (158) Amblard, F.; Cho, J. H.; Schinazi, R. F. *Chem. Rev.* **2009**, *109* (9), 4207–4220.
- (159) Mathew, S. C.; By, Y.; Berthault, A.; Virolleaud, M.-A.; Carrega, L.; Chouraqui, G.; Commeiras, L.; Condo, J.; Attolini, M.; Gaudel-Siri, A.; Ruf, J.; Rodriguez, J.; Parrain, J.-L.; Guieu, R. *Org. Biomol. Chem.* **2010**, *8* (17), 3874–3881.
- (160) Park, T.; Todd, E. M.; Nakashima, S.; Zimmerman, S. C. *J. Am. Chem. Soc.* **2005**, *127* (51), 18133–18142.
- (161) Chinchilla, R.; Najera, C. *Chem. Soc. Rev.* **2011**, *40* (10), 5084–5121.
- (162) Gottardo, C.; Kraft, T. M.; Hossain, M. S.; Zawada, P. V.; Muchall, H. M. *Can. J. Chem.* **2008**, *86* (5), 410–415.
- (163) Liu, J.; Janeba, Z.; Robins, M. J. *Org. Lett.* **2004**, *6* (17), 2917–2919.
- (164) Kim, Sanghee; Kim, Bogyong; In, J. *Synthesis (Stuttg.)* **2009**, No. 12, 1963–1968.
- (165) Ostergaard, M. E.; Guenther, D. C.; Kumar, P.; Baral, B.; Deobald, L.; Paszczynski, A. J.; Sharma, P. K.; Hrdlicka, P. J. *Chem. Commun.* **2010**, *46* (27), 4929–4931.
- (166) Sau, S. P.; Hrdlicka, P. J. *J. Org. Chem.* **2012**, *77* (1), 5–16.
- (167) Mutisya, D.; Selvam, C.; Kennedy, S. D.; Rozners, E. *Bioorg. Med. Chem. Lett.* **2011**, *21* (11), 3420–3422.
- (168) Vougioukalakis, G. C.; Roubelakis, M. M.; Orfanopoulos, M. *J. Org. Chem.* **2010**, *75* (12), 4124–4130.
- (169) Gann, A. W.; Amoroso, J. W.; Einck, V. J.; Rice, W. P.; Chambers, J. J.; Schnarr, N. A. *Org. Lett.* **2014**, *16* (7), 2003–2005.
- (170) Ngai, M. H.; Yang, P.-Y.; Liu, K.; Shen, Y.; Wenk, M. R.; Yao, S. Q.; Lear, M. J. *Chem. Commun.* **2010**, *46* (44), 8335–8337.
- (171) Kimoto, M.; Mitsui, T.; Yokoyama, S.; Hirao, I. *J. Am. Chem. Soc.* **2010**, *132* (14), 4988–4989.
- (172) Spinelli, D.; Zanirato, P. *J. Chem. Soc. Perkin Trans. 2* **1993**, No. 6, 1129–1133.
- (173) Potratz, S.; Mishra, A.; Bäuerle, P. *Beilstein J. Org. Chem.* **2012**, *8*, 683–692.
- (174) Mathé, C.; Lioux, T.; Gosselin, G. *Nucleosides, Nucleotides and Nucleic Acids* **2003**, *22* (5–8), 605–609.
- (175) Lakshman, M. K.; Singh, M. K.; Parrish, D.; Balachandran, R.; Day, B. W. *J. Org. Chem.* **2010**, *75* (8), 2461–2473.
- (176) Hocek, M.; Nauš, P.; Pohl, R.; Votruba, I.; Furman, P. A.; Tharnish, P. M.; Otto, M. J. *J. Med. Chem.* **2005**, *48* (18), 5869–5873.
- (177) Allen, C. L.; Davulcu, S.; Williams, J. M. J. *Org. Lett.* **2010**, *12* (22), 5096–5099.
- (178) Strotman, N. A.; Chobanian, H. R.; Guo, Y.; He, J.; Wilson, J. E. *Org. Lett.* **2010**, *12* (16), 3578–3581.
- (179) van Leusen, A. M.; Hoogenboom, B. E.; Siderius, H. *Tetrahedron Lett.* **1972**, *13* (23), 2369–2372.
- (180) Šilhár, P.; Pohl, R.; Votruba, I.; Hocek, M. *Org. Lett.* **2004**, *6* (19), 3225–3228.
- (181) Šilhár, P.; Pohl, R.; Votruba, I.; Hocek, M. *Synthesis (Stuttg.)* **2006**, *2006* (11), 1848–1852.

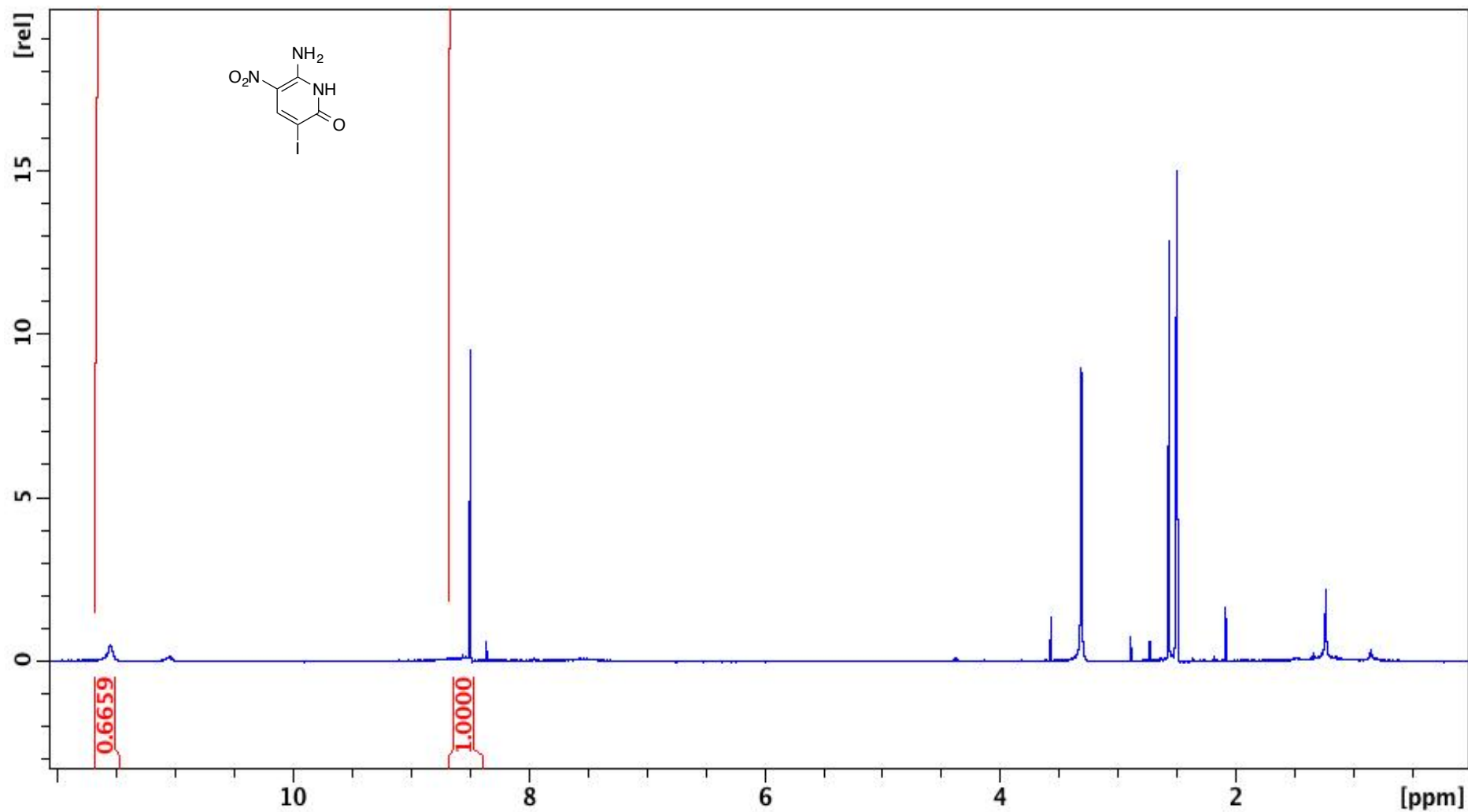
- (182) Elgemeie, G. H.; Zaghary, W. A.; Amin, K. M.; Nasr, T. M. *Nucleosides, Nucleotides and Nucleic Acids* **2005**, *24* (8), 1227–1247.
- (183) Manfredini, S.; Baraldi, P. G.; Bazzanini, R.; Durini, E.; Vertuani, S.; Pani, A.; Marceddu, T.; Demontis, F.; Vargiu, L.; La Colla, P. *Nucleosides, Nucleotides and Nucleic Acids* **2000**, *19* (4), 705–722.
- (184) Su, M.; Tomas-Gamasa, M.; Carell, T. *Chem. Sci.* **2015**, *6* (1), 632–638.
- (185) Lakshman, M. K.; Gunda, P.; Pradhan, P. *J. Org. Chem.* **2005**, *70* (25), 10329–10335.
- (186) Klecka, M.; Pohl, R.; Klepetarova, B.; Hocek, M. *Org. Biomol. Chem.* **2009**, *7* (5), 866–868.
- (187) Littke, A. F.; Dai, C.; Fu, G. C. *J. Am. Chem. Soc.* **2000**, *122* (17), 4020–4028.
- (188) Kabalka, G. W.; Namboodiri, V.; Wang, L. *Chem. Commun.* **2001**, No. 8, 775.
- (189) Wright, S. W.; Hageman, D. L.; McClure, L. D. *J. Org. Chem.* **1994**, *59* (20), 6095–6097.
- (190) Guram, A. S.; Wang, X.; Bunel, E. E.; Faul, M. M.; Larsen, R. D.; Martinelli, M. J. *J. Org. Chem.* **2007**, *72* (14), 5104–5112.
- (191) Fleury-Brégeot, N.; Oehlich, D.; Rombouts, F.; Molander, G. A. *Org. Lett.* **2013**, *15* (7), 1536–1539.
- (192) Kolganova, N. A.; Florentiev, V. L.; Chudinov, A. V.; Zasedatelev, A. S.; Timofeev, E. N. *Chem. Biodivers.* **2011**, *8* (4), 568–576.
- (193) Burgess, K.; Cook, D. *Chem. Rev.* **2000**, *100* (6), 2047–2060.
- (194) Yoshikawa, M.; Kato, T.; Takenishi, T. *Tetrahedron Lett.* **1967**, *8* (50), 5065–5068.
- (195) Borsenberger, V.; Kukwikila, M.; Howorka, S. *Org. Biomol. Chem.* **2009**, *7* (18), 3826–3835.
- (196) Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54* (3), 631–635.
- (197) Caton-Williams, J.; Lin, L.; Smith, M.; Huang, Z. *Chem. Commun.* **2011**, *47* (28), 8142–8144.
- (198) Toti, K. S.; Derudas, M.; Pertusati, F.; Sinnaeve, D.; Van den Broeck, F.; Margamuljana, L.; Martins, J. C.; Herdewijn, P.; Balzarini, J.; McGuigan, C.; Van Calenbergh, S. *J. Org. Chem.* **2014**, *79* (11), 5097–5112.
- (199) Murphy, P. J. *Organophosphorus Reagents*; Oxford University Press, 2004.
- (200) Korhonen, H. J.; Bolt, H. L.; Vicente-Gines, L.; Perks, D. C.; Hodgson, D. R. W. *Phosphorus. Sulfur. Silicon Relat. Elem.* **2015**, *190* (5–6), 758–762.
- (201) Peter G. M. Wuts, T. W. G. *Greene's Protective Groups in Organic Synthesis, Fourth Edition*, Fourth.; WILEY-VCH Verlag, 2006.
- (202) Fan, Y.; Gaffney, B. L.; Jones, R. A. *Org. Lett.* **2004**, *6* (15), 2555–2557.
- (203) Chaix, C.; Molko, D.; Téoule, R. *Tetrahedron Lett.* **1989**, *30* (1), 71–74.
- (204) Cho, J. H.; Coats, S. J.; Schinazi, R. F. *Org. Lett.* **2012**, *14* (10), 2488–2491.
- (205) Aibibuli, Z.; Wang, Y.; Tu, H.; Huang, X.; Zhang, A. *Molecules* . 2012.
- (206) Kimoto, M.; Mitsui, T.; Harada, Y.; Sato, A.; Yokoyama, S.; Hirao, I. *Nucleic Acids Res.* **2007**, *35* (16), 5360–5369.
- (207) Mildvan, A. S.; Loeb, L. A.; Wu, C.-W. *Crit. Rev. Biochem.* **1979**, *6* (3), 219–244.
- (208) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. *J. Chem. Theory Comput.* **2011**, *7* (4), 1177–1188.

- (209) GOPALAKRISHNA, S.; GUSTI, V.; NAIR, S.; SAHAR, S.; GAUR, R. K. *RNA* **2004**, *10* (11), 1820–1830.
- (210) Zhu, B.; Hernandez, A.; Tan, M.; Wollenhaupt, J.; Tabor, S.; Richardson, C. C. *Nucleic Acids Res.* **2015**, *43* (14), e94–e94.
- (211) Tabor, S.; Richardson, C. C. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (11), 4076–4080.

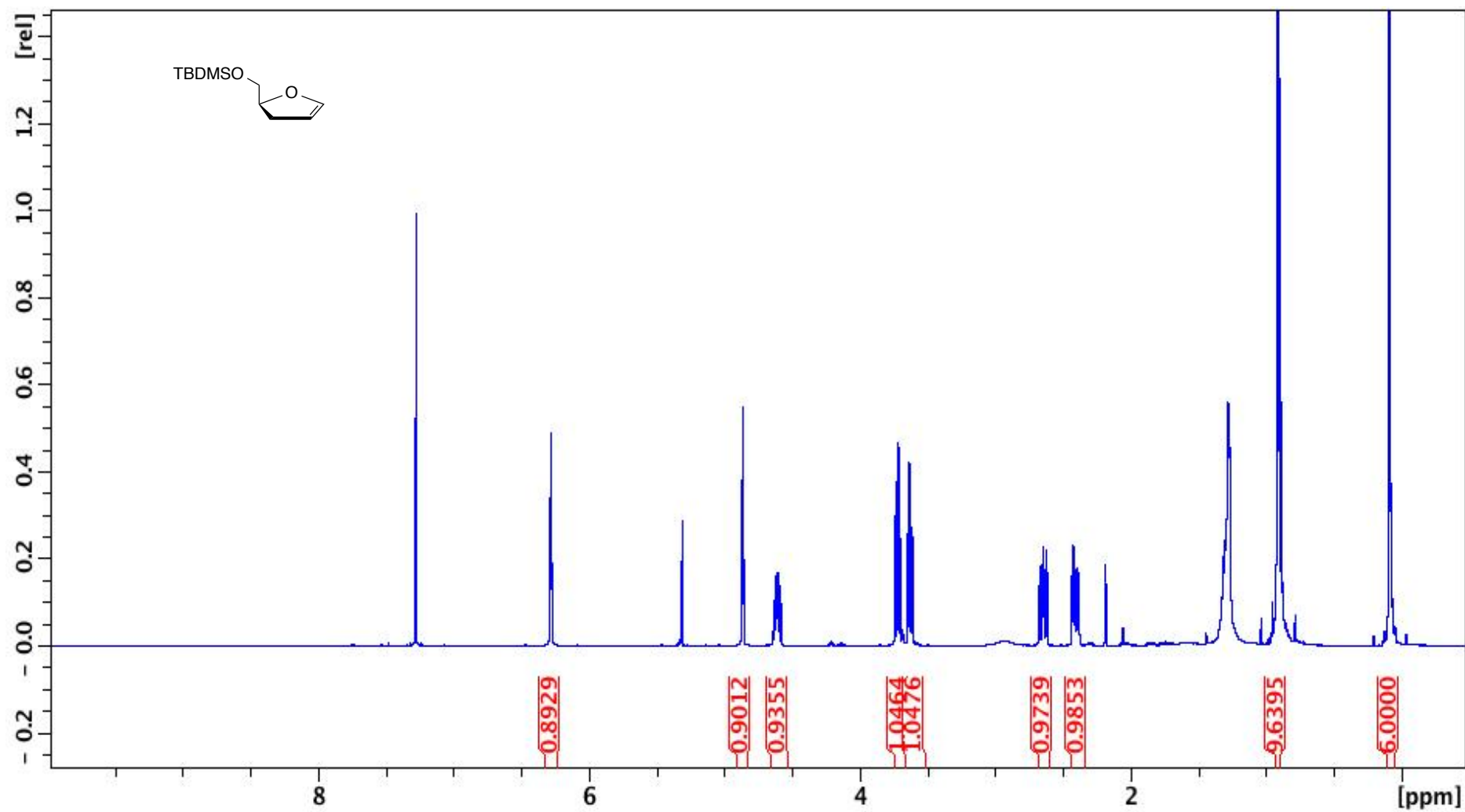
Appendix



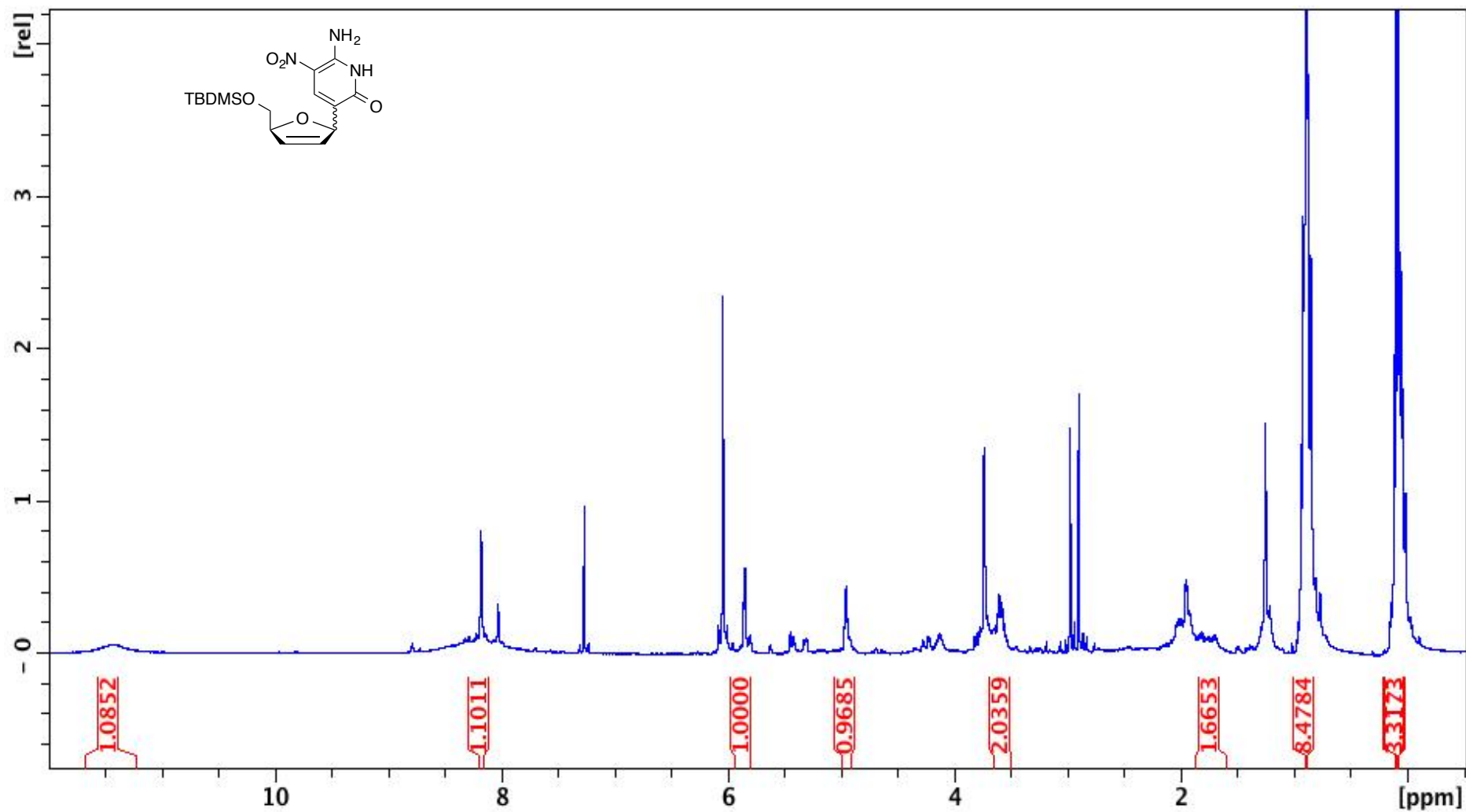
Appendix 1. ¹H-NMR spectrum of **99**. 500 MHz, DMSO-d₆. Peak at 2.50 ppm comes from DMSO.



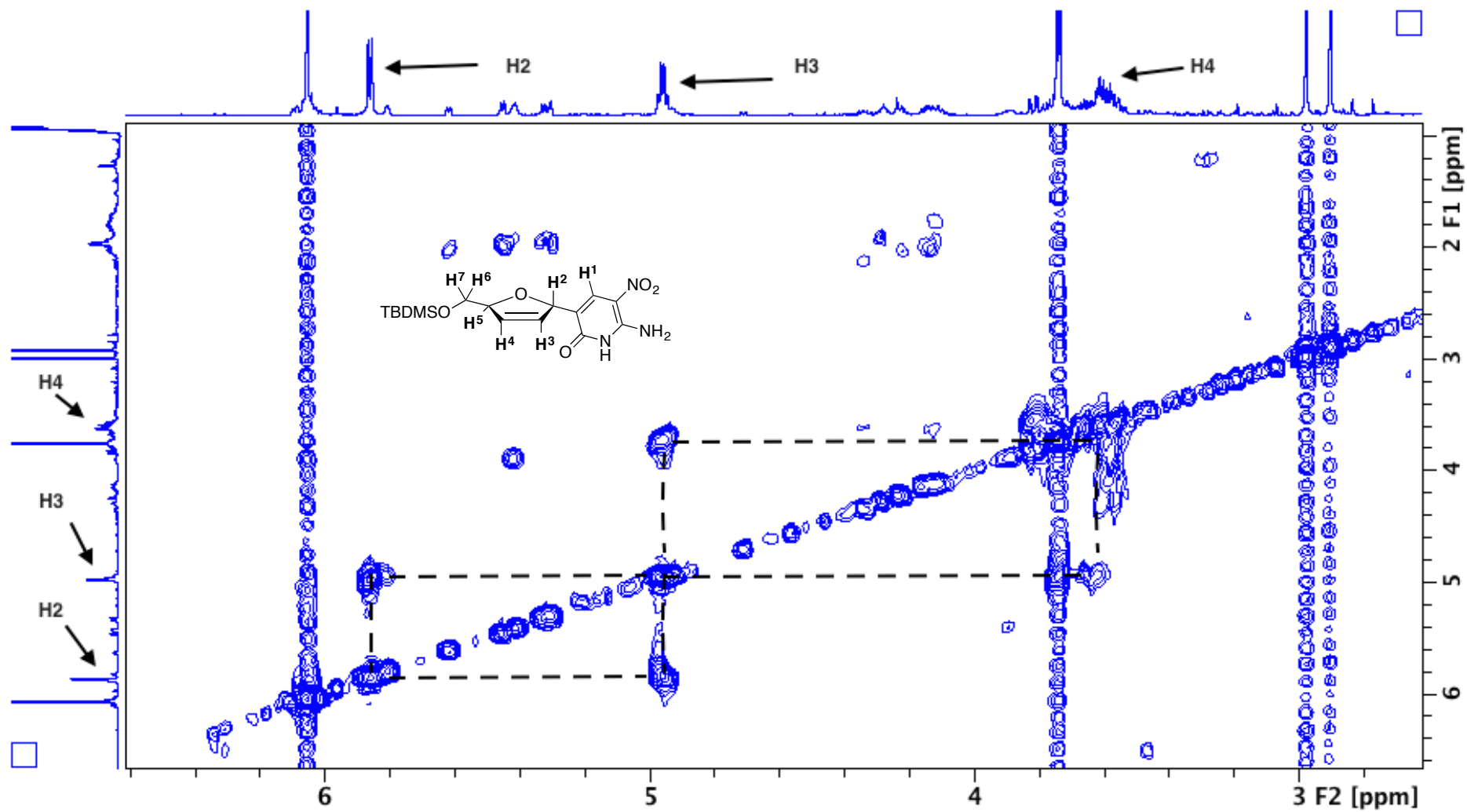
Appendix 2. ^1H -NMR spectrum of 77. 500 MHz, DMSO- d_6 . Peak at 2.50 comes from DMSO. Peak at 3.30 comes from H_2O . Peak at 2.54 ppm comes from an unknown impurity.



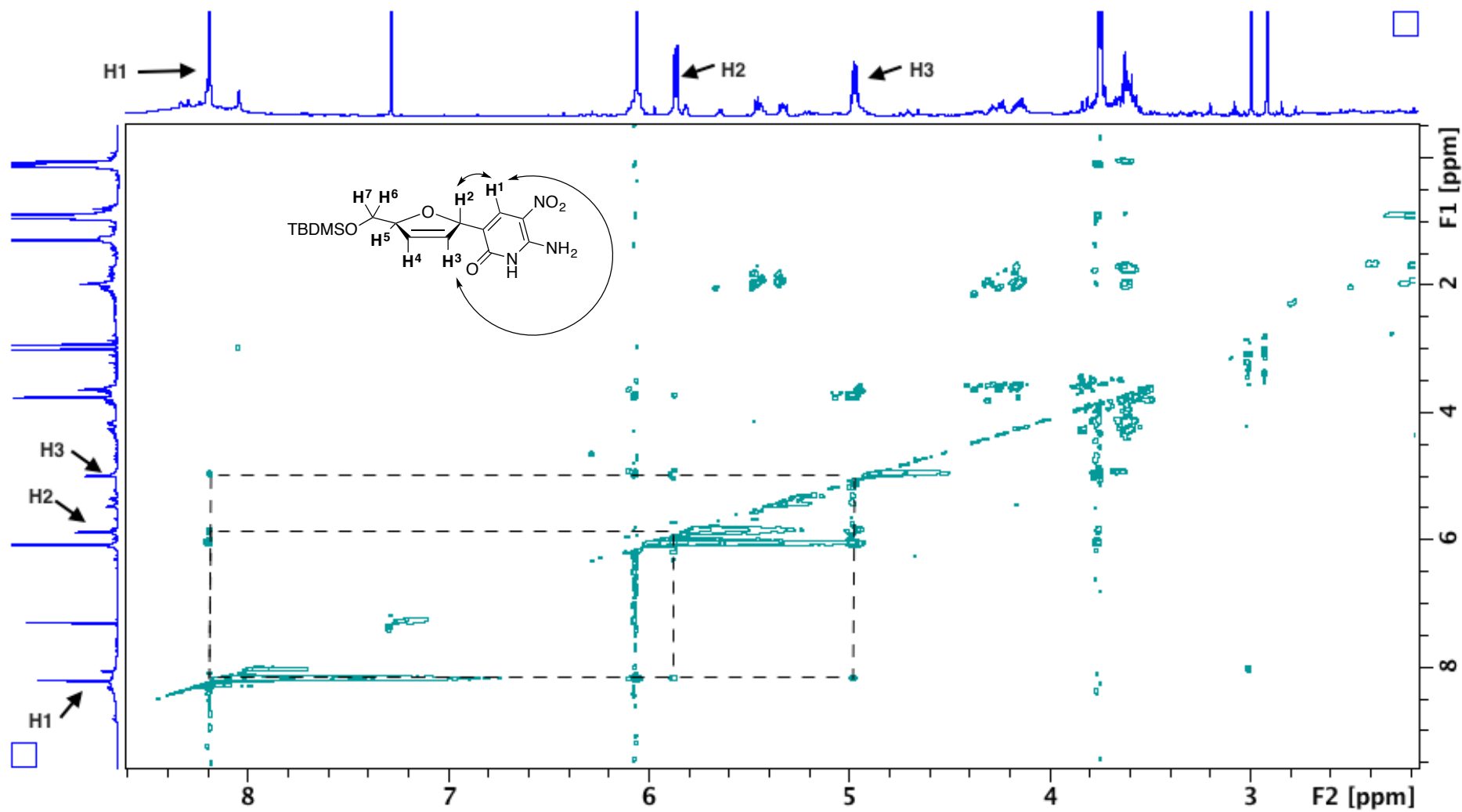
Appendix 5. ^1H -NMR spectrum of **59**. 500 MHz, CDCl_3 . Peak at 1.4 ppm comes from unknown impurity. Peak at 5.3 ppm comes from DCM.



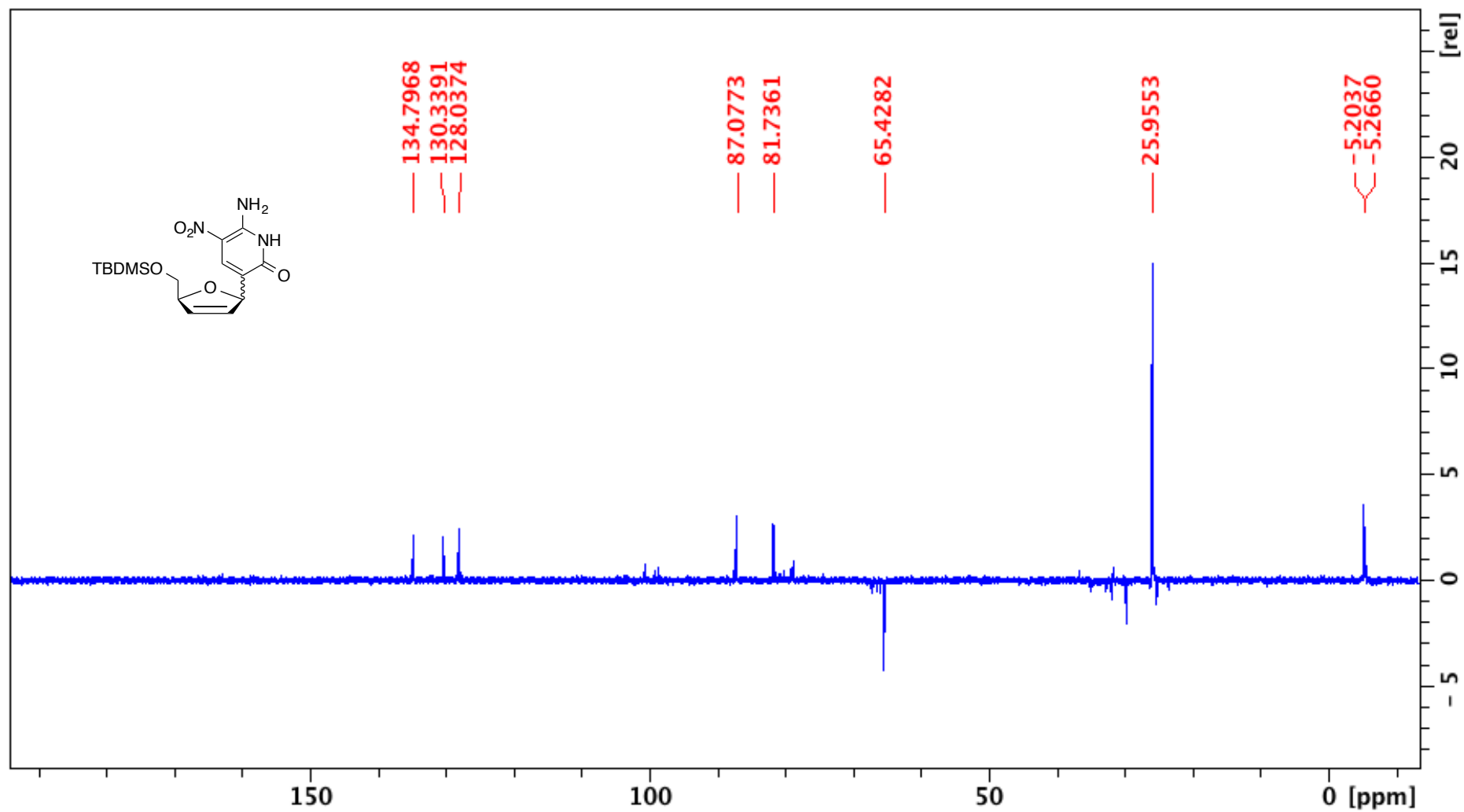
Appendix 6. ¹H-NMR spectrum of **100**. 400 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃. Peaks at 2.90 ppm and 2.99 ppm come from DMF.



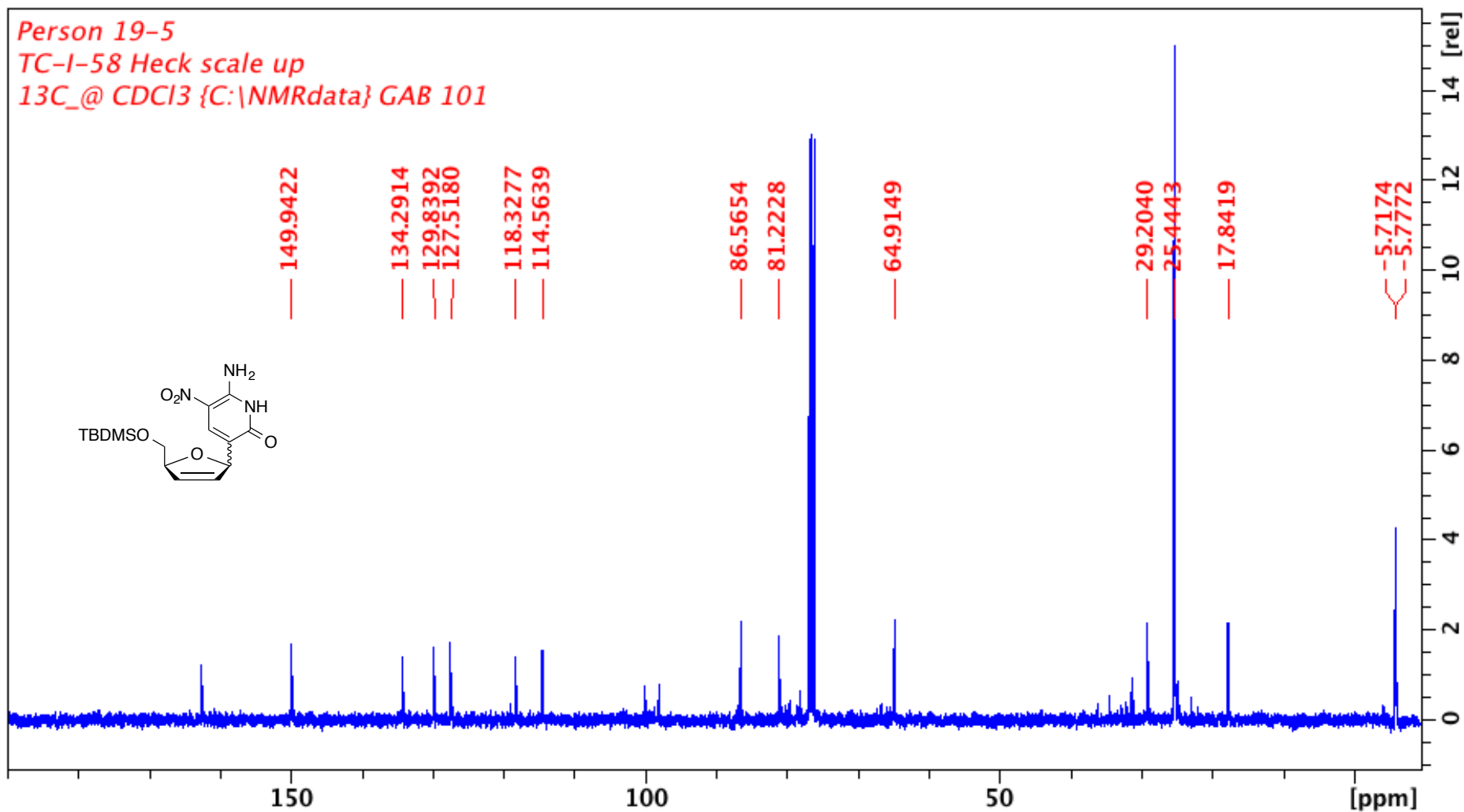
Appendix 7. ^1H -COSY NMR spectrum of **100** showing the correlations observed. 500 MHz, CDCl_3 .



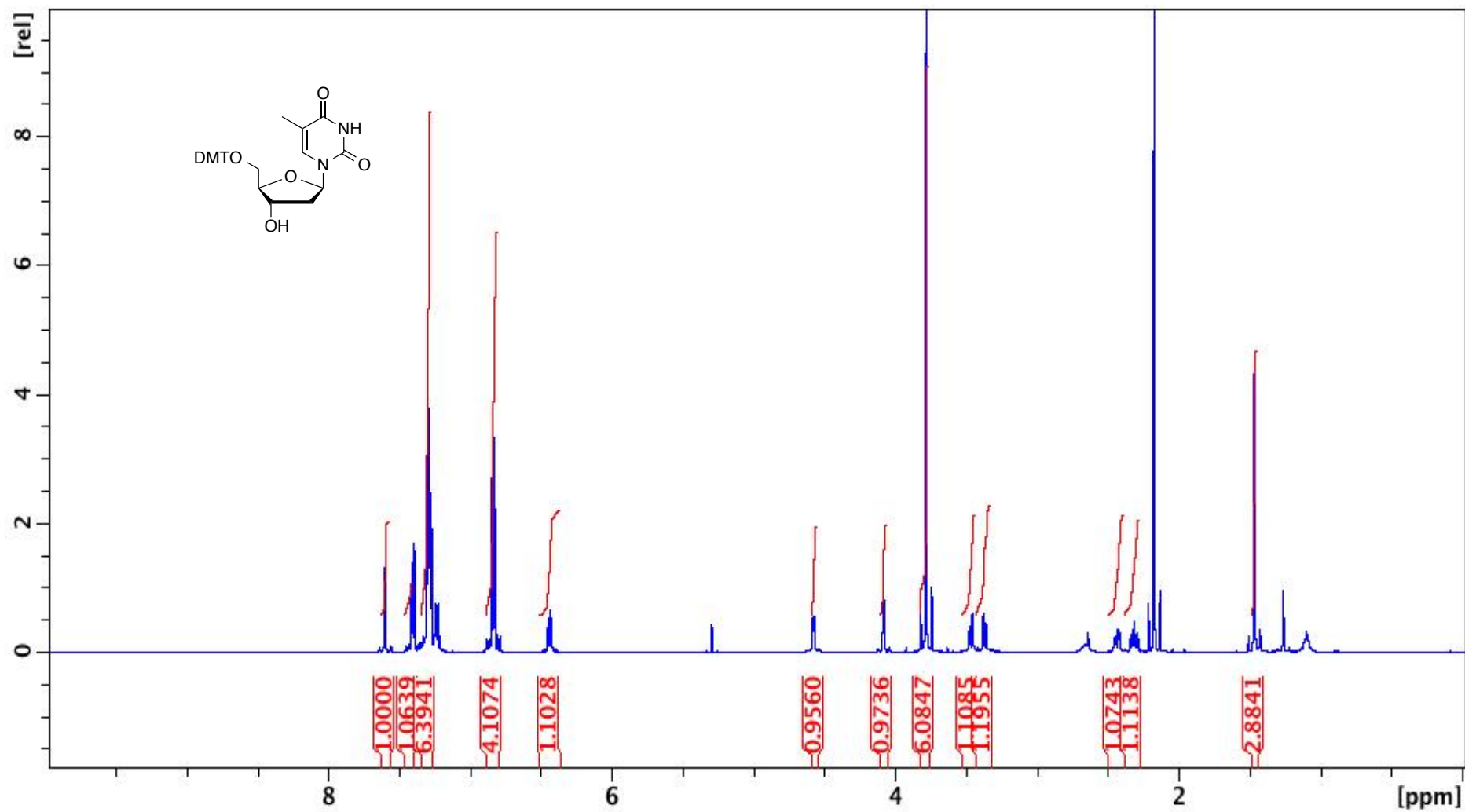
Appendix 8. ^1H -NOESY NMR spectrum of **100** showing the correlations observed. 500 MHz, CDCl_3 . Only negative contour signs are shown for clarity.

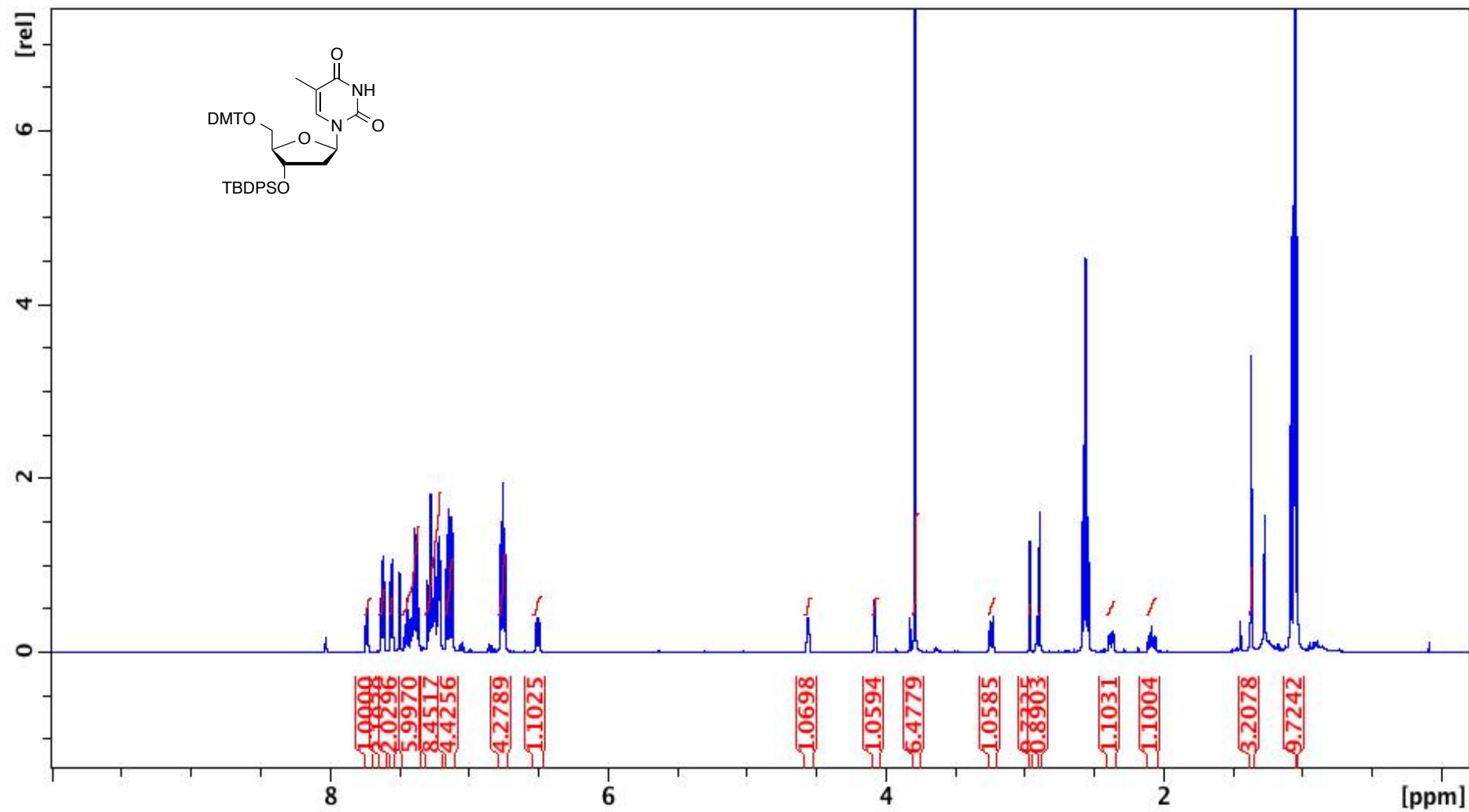


Appendix 9. ^{13}C -DEPT135 NMR spectrum of **100**. 100 MHz, CDCl_3 .

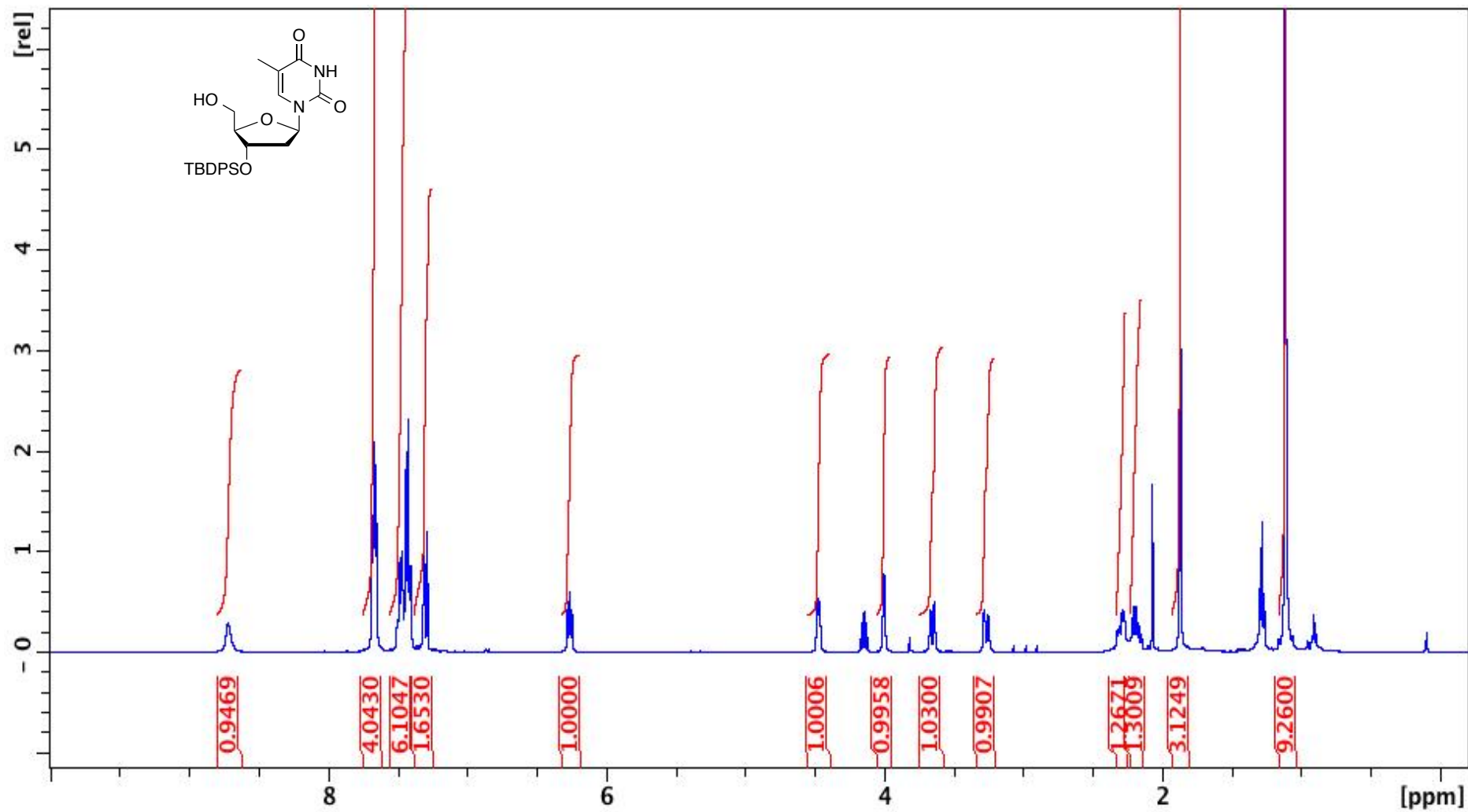


Appendix 10. ^{13}C -NMR spectrum of **100**. 100 MHz, CDCl_3 . Peak at 76 ppm comes from CHCl_3 . Peaks at 36, 31 and 162 ppm come from DMF.

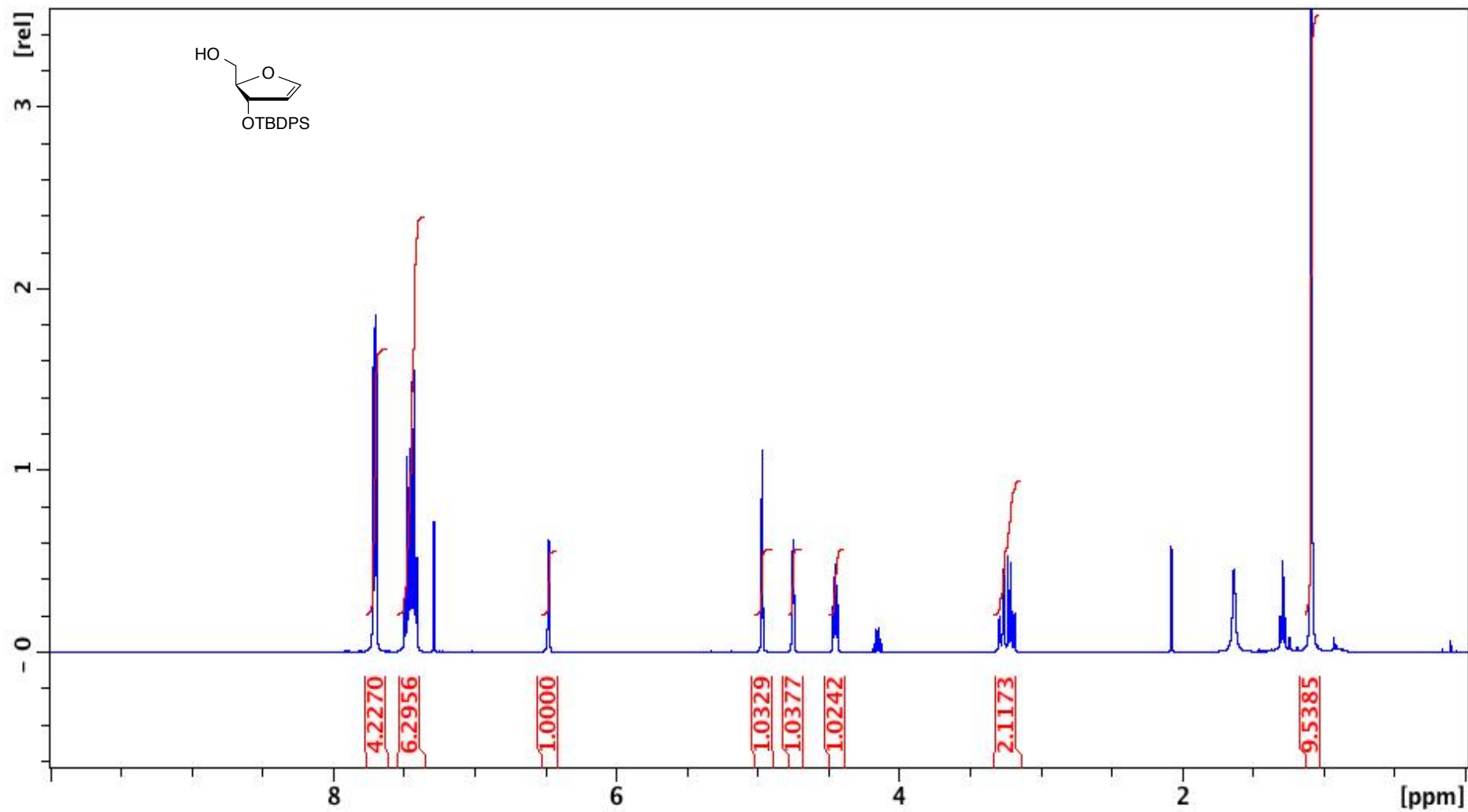
Appendix 11. ¹H-NMR spectrum of **102**. 500 MHz, CDCl₃



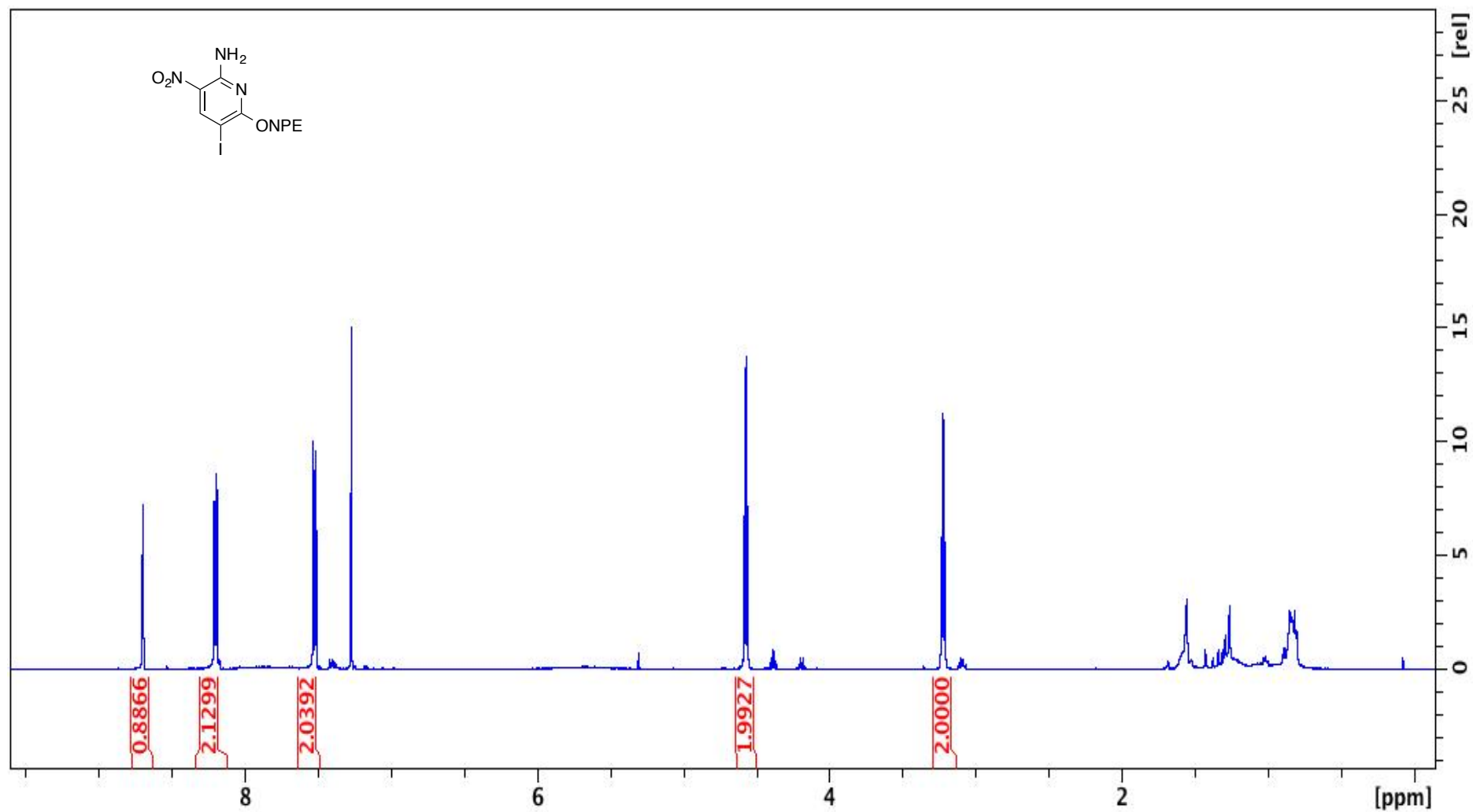
Appendix 12. ¹H-NMR spectrum of **103**. 500 MHz, CDCl₃. Peak at 2.5 ppm comes from unknown impurity.



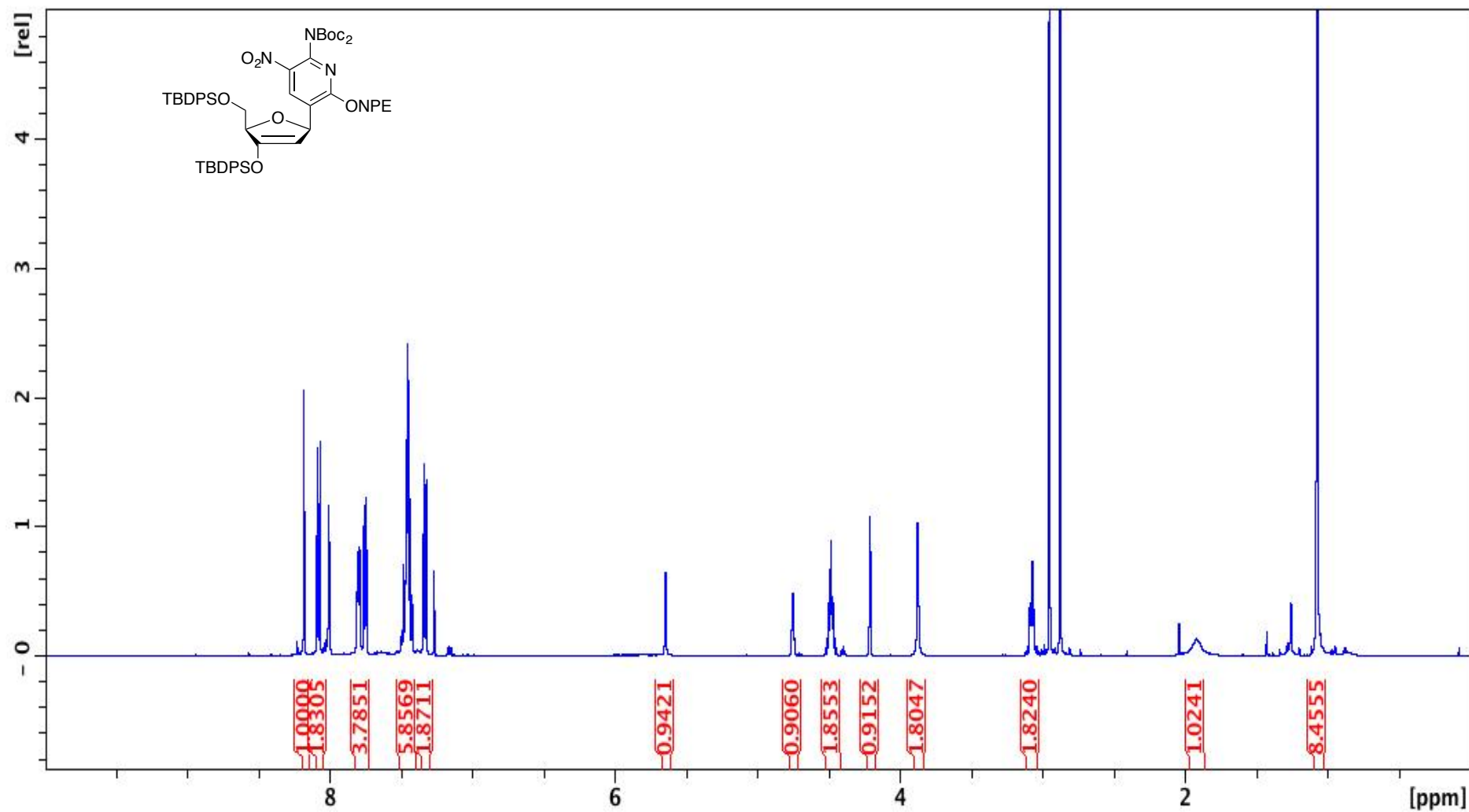
Appendix 13. ¹H-NMR spectrum of **104**. 500MHz, CDCl₃. Peaks at 1.26, 2.05 and 4.05 ppm come from ethyl acetate.



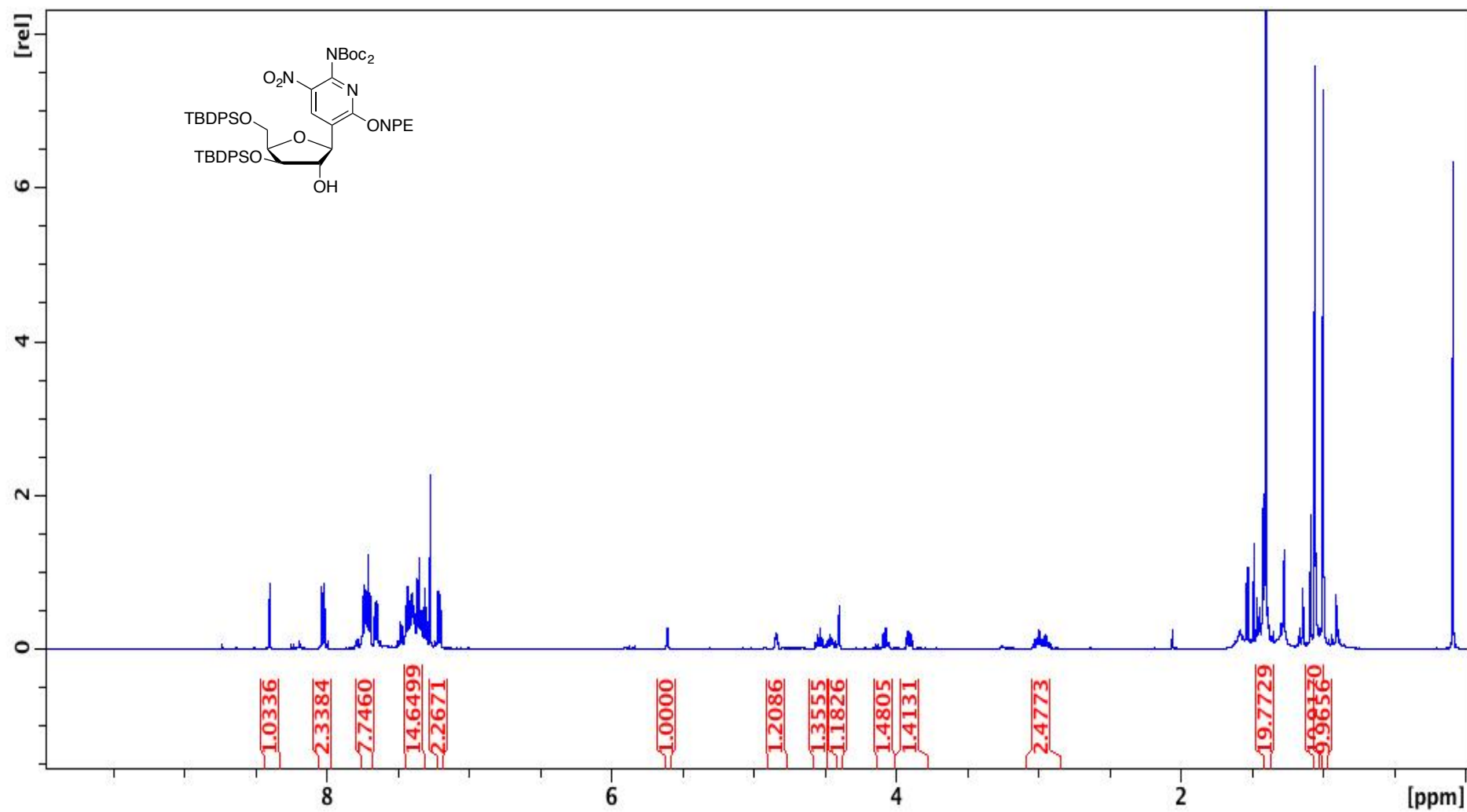
Appendix 14. ¹H-NMR spectrum of 72. 500 MHz, CDCl₃. Peak at 2.17 ppm comes from acetone. Peaks at 1.3, 2.1 and 4.1 ppm come from EtOAc.



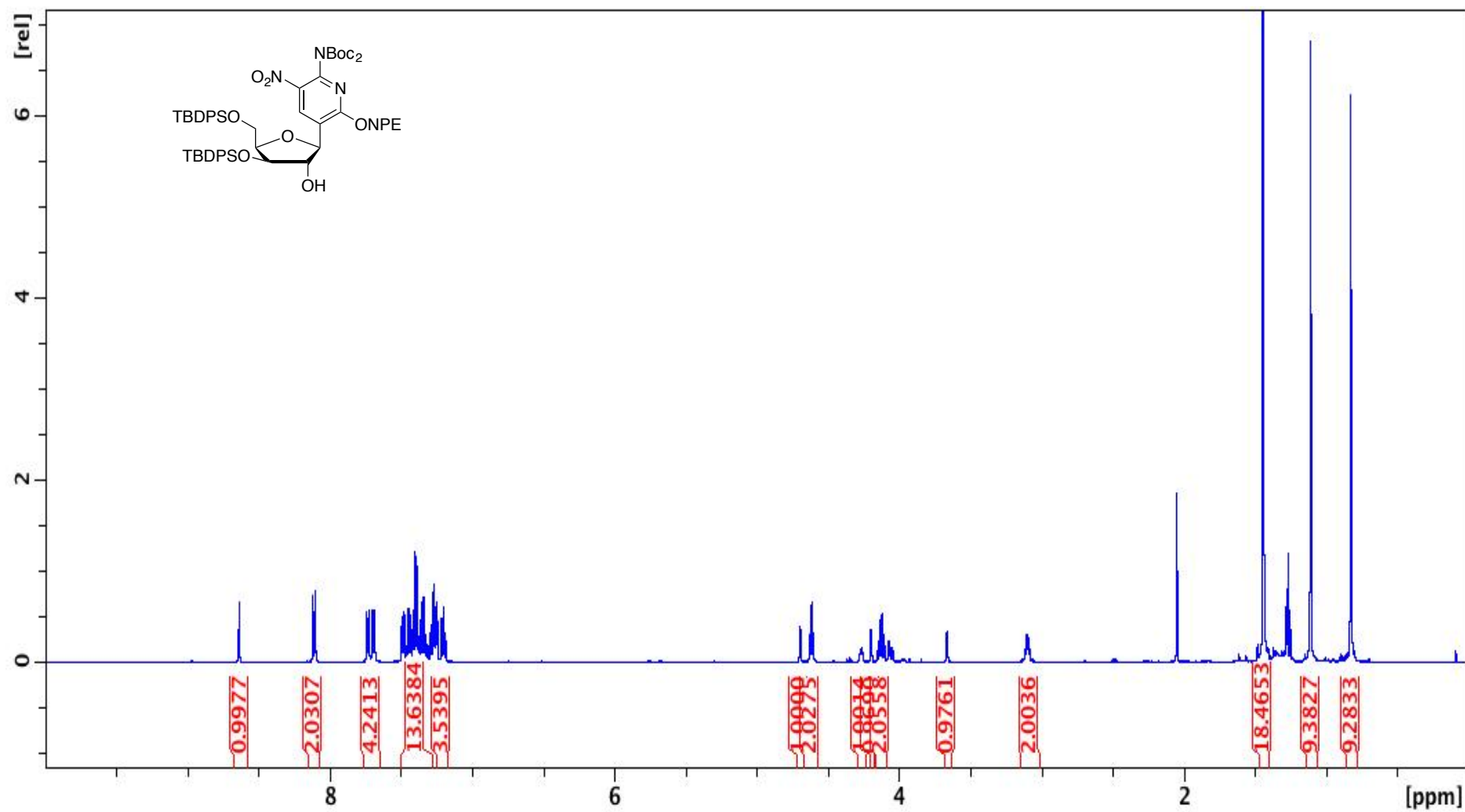
Appendix 15. ¹H-NMR spectrum of **87**. 500 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃.



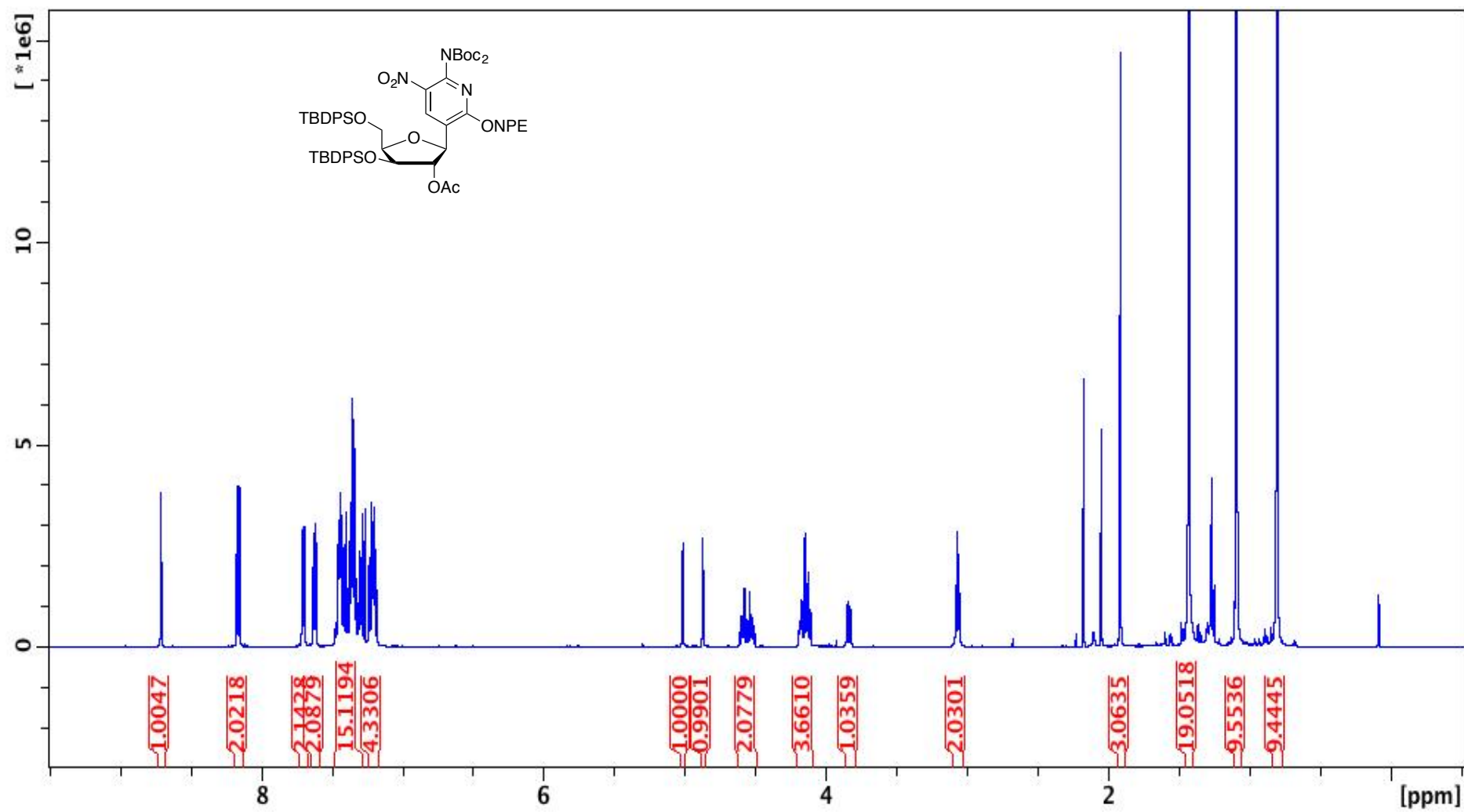
Appendix 16. ¹H-NMR spectrum of **88**. 500 MHz, CDCl₃. Peaks at 2.8, 2.9 and 8.0 ppm come from DMF.



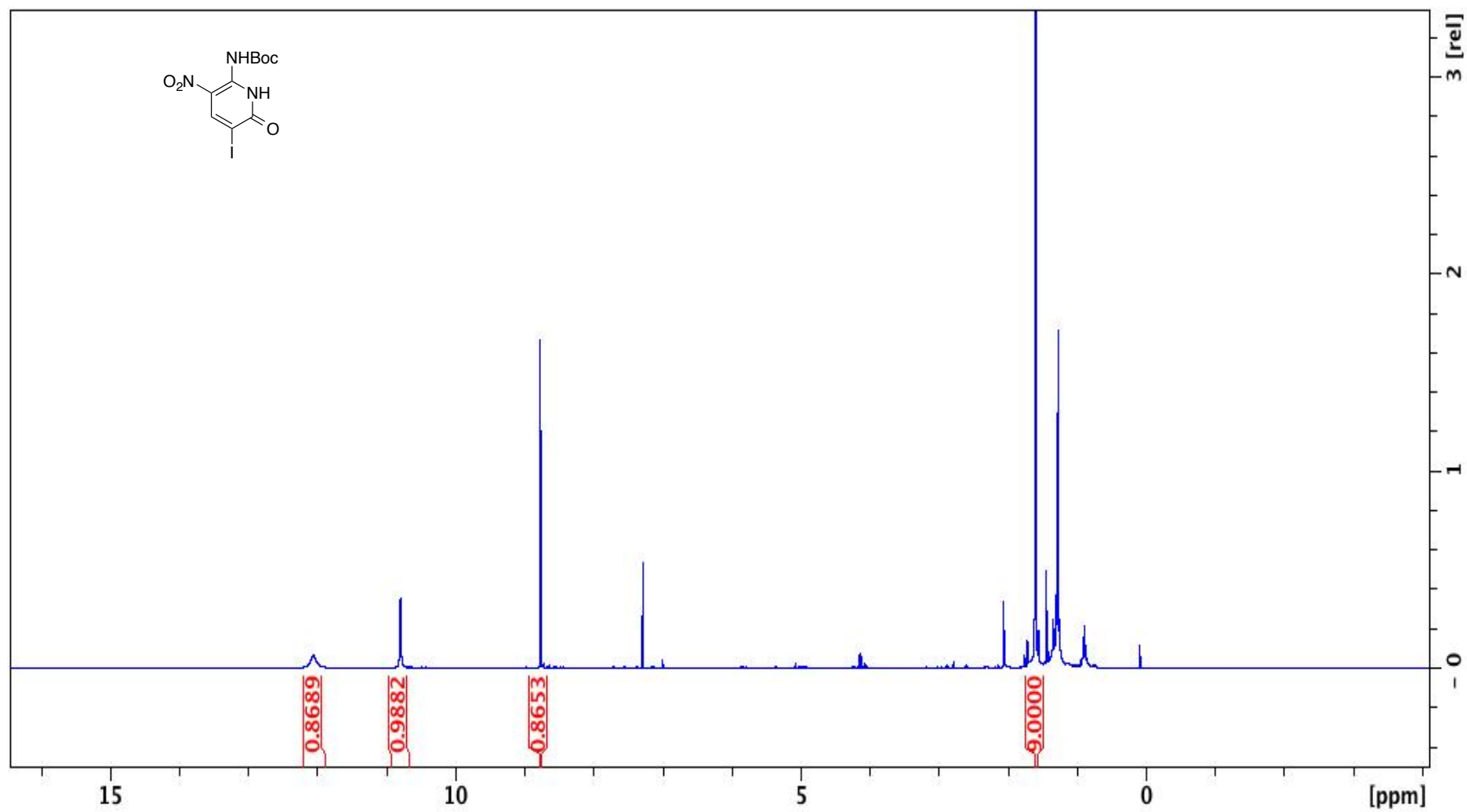
Appendix 17. ^1H -NMR spectrum of **89**. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peak at 0.08 ppm comes from silicon grease.



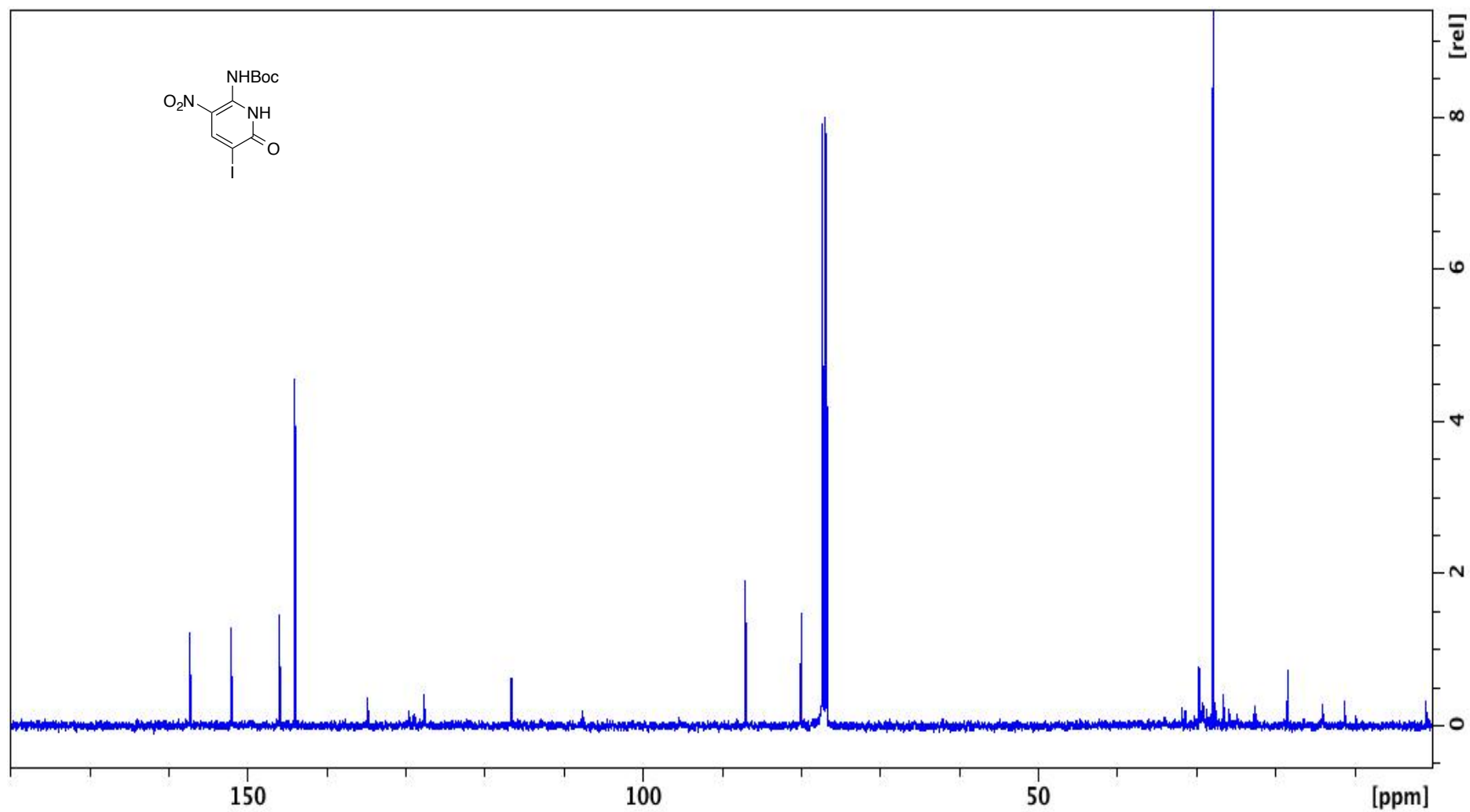
Appendix 18. $^1\text{H-NMR}$ spectrum of **90**. 500 MHz, CDCl_3 . Peak at 1.26, 2.0 and 4.1 ppm come from EtOAc.



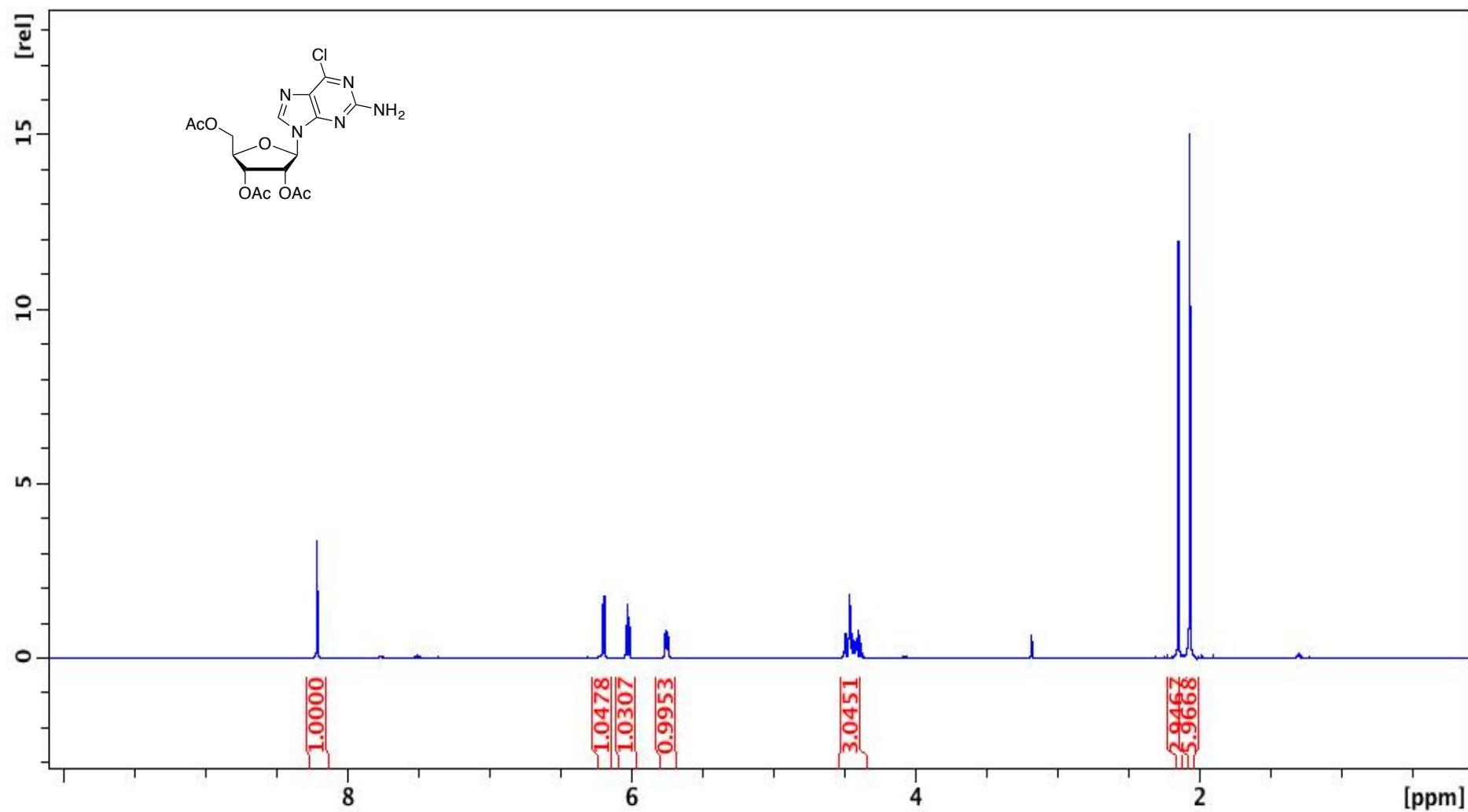
Appendix 19. ¹H-NMR spectrum of **105**. 500 MHz, CDCl₃.



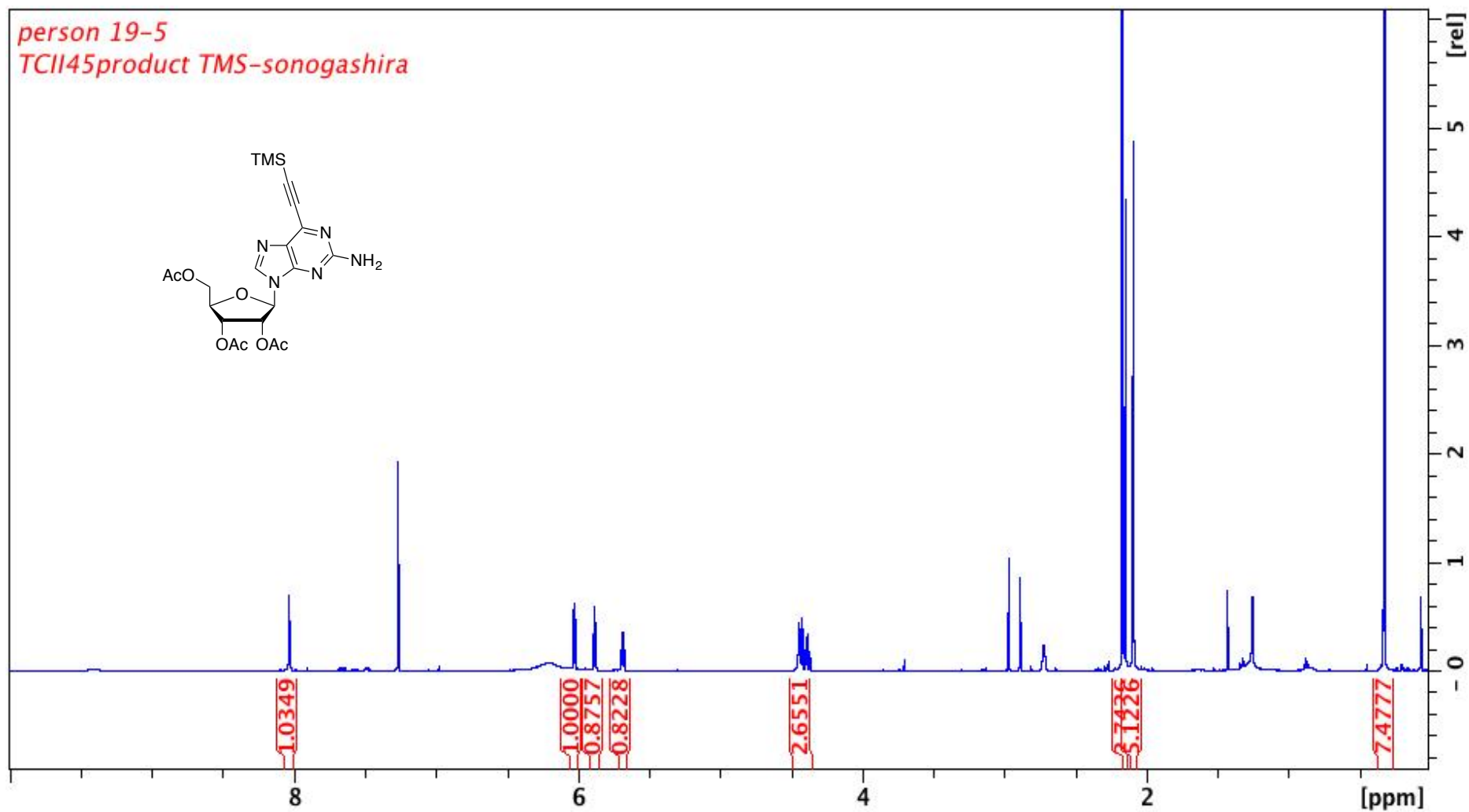
Appendix 20. ¹H-NMR spectrum of **106**. 500 MHz, CDCl₃. Peak at 1.27 ppm comes from unknown impurity.



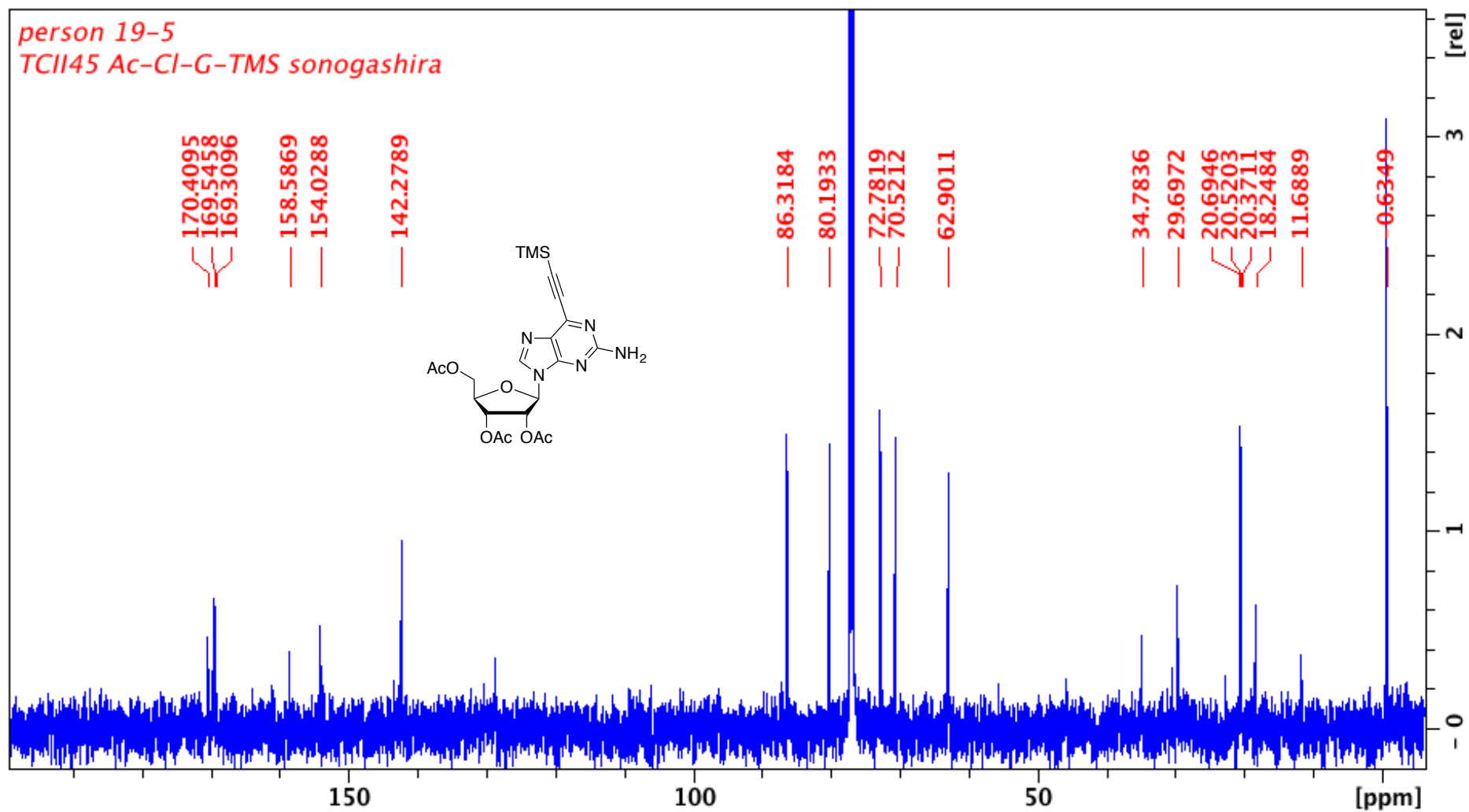
Appendix 21. ^{13}C -NMR spectrum of **106**. 125 MHz, CDCl_3 . Peak at 76 ppm comes from CHCl_3 .



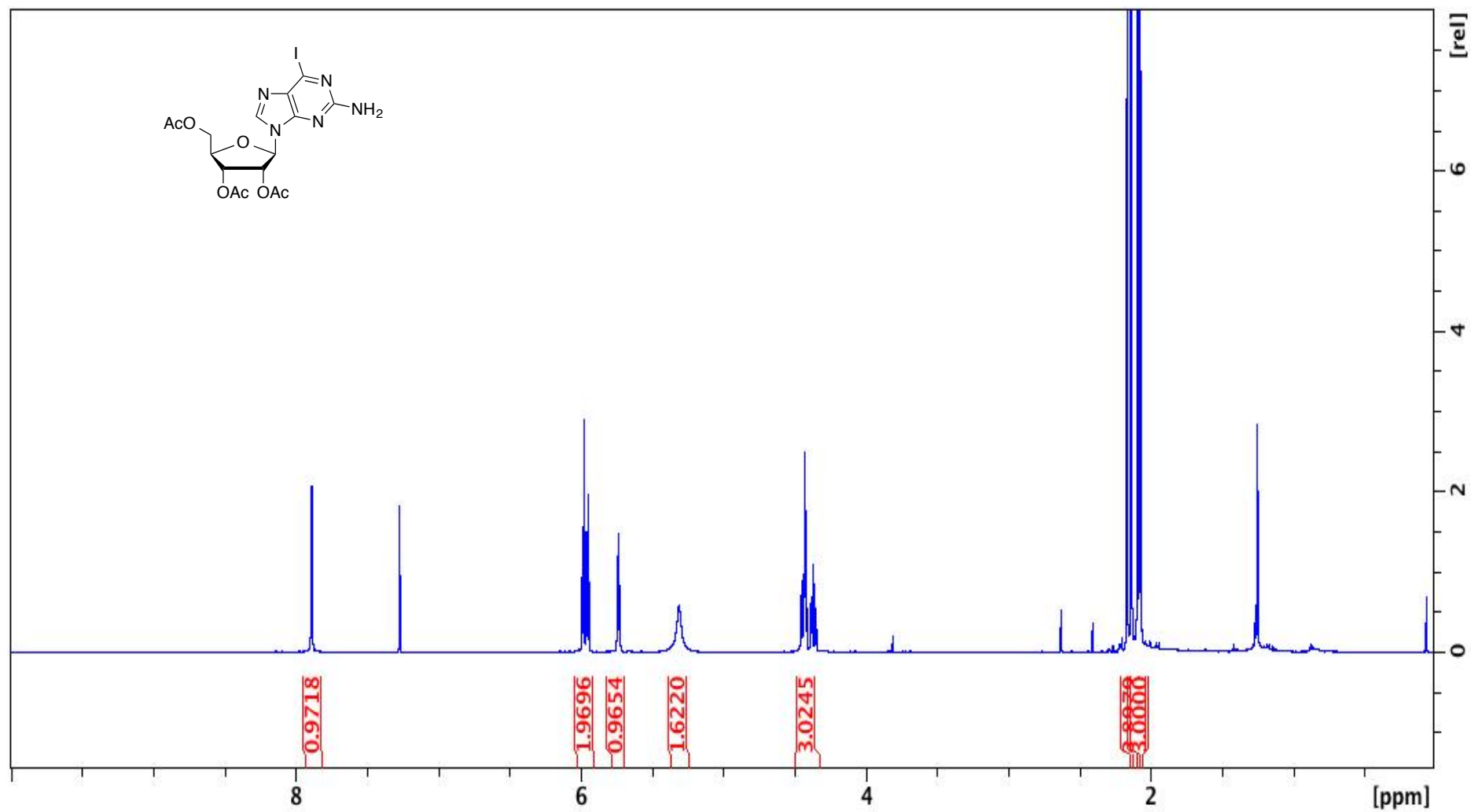
Appendix 22. ¹H-NMR spectrum of **115**. 500 MHz, acetone-d₆.



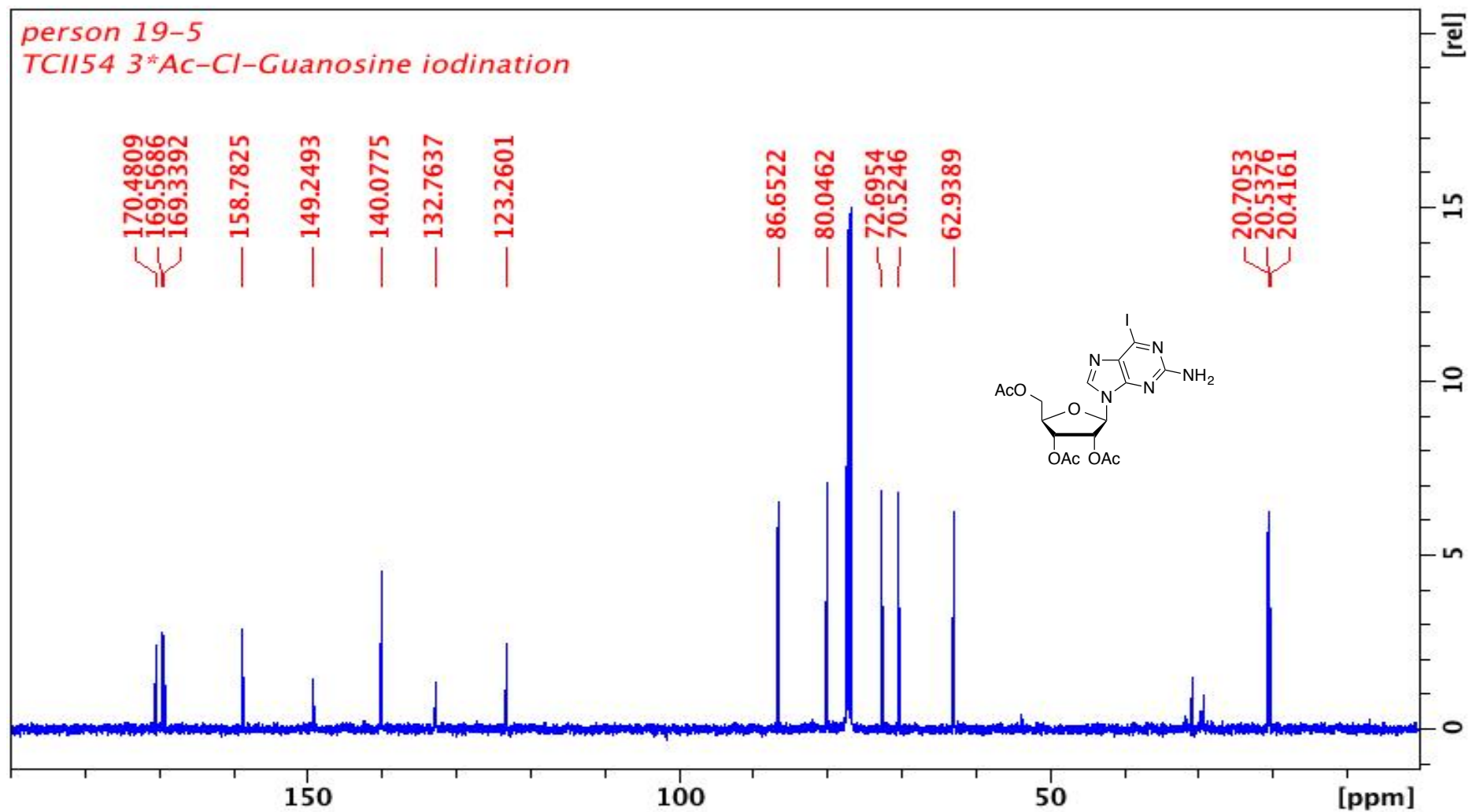
Appendix 23. $^1\text{H-NMR}$ spectrum of **117**. 500 MHz, CDCl_3 . Peaks at 2.9 and 3.0 ppm come from DMF. Peak at 7.26 comes from CHCl_3 .



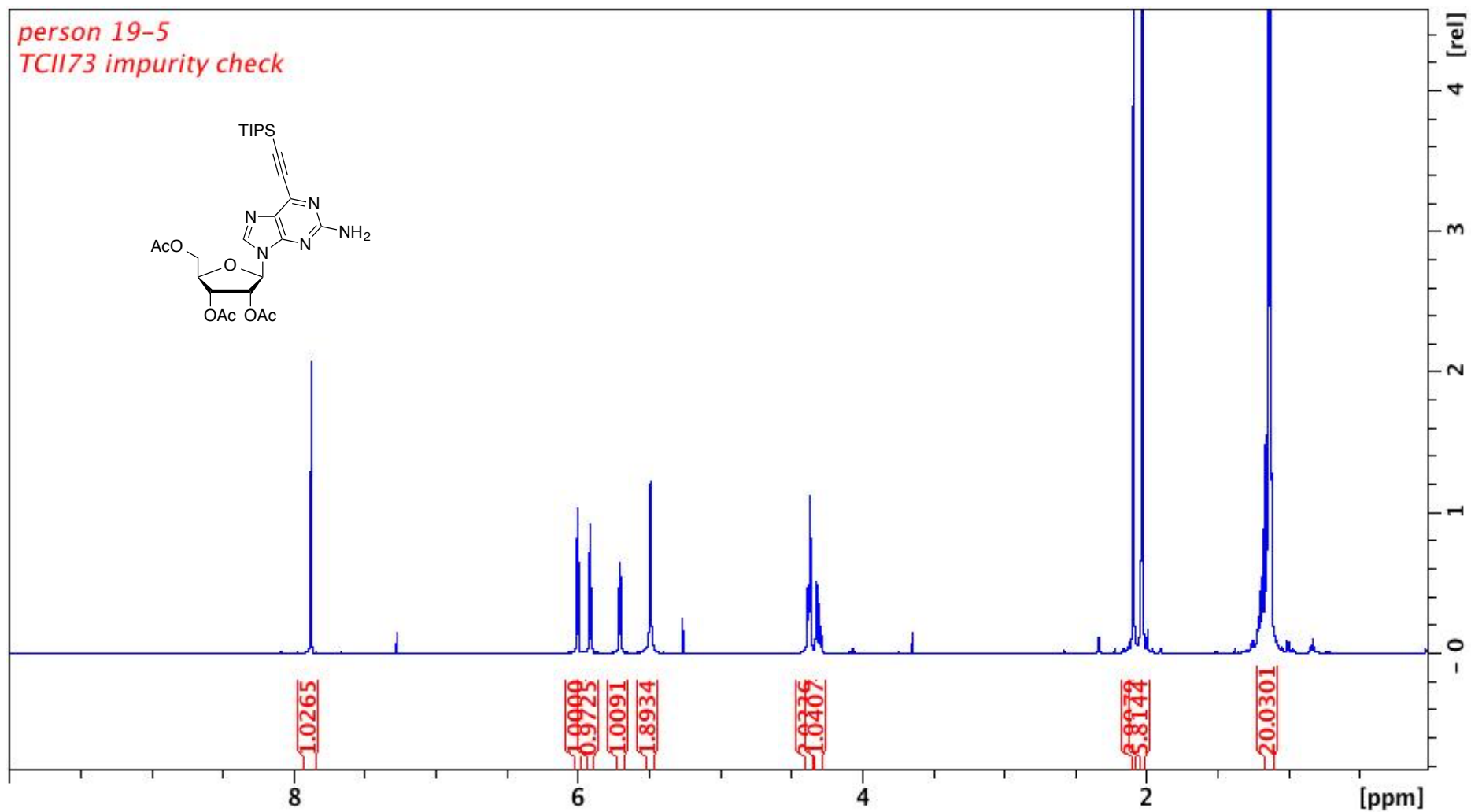
Appendix 24. ^{13}C -NMR spectrum of 117. 125 MHz, CDCl_3 . Peak at 7.26 comes from CHCl_3 .



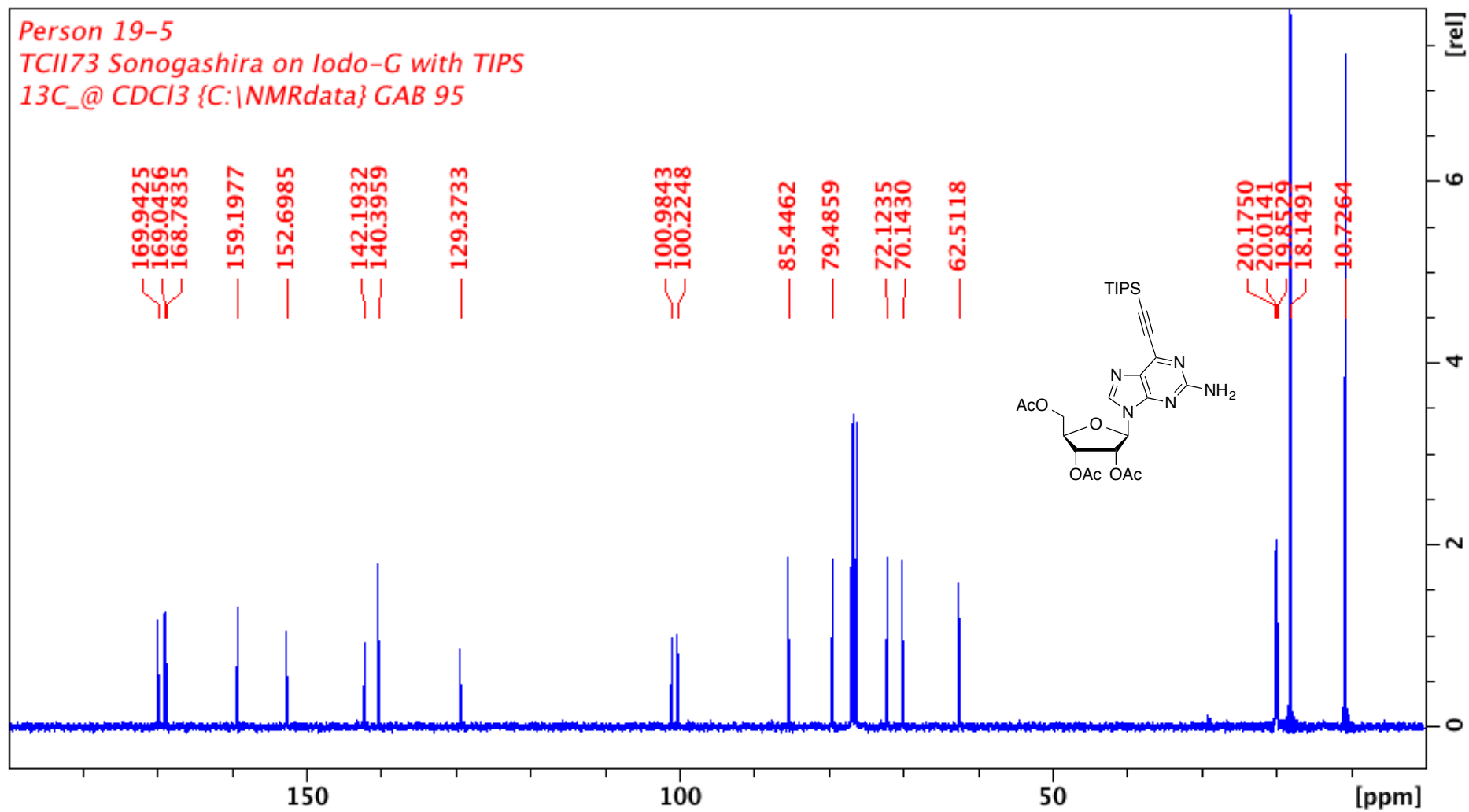
Appendix 25. ¹H-NMR spectrum of **118**. 500 MHz, CDCl₃. Peak at 1.25 and 0.86 come from grease. Peak at 7.26 comes from CHCl₃.



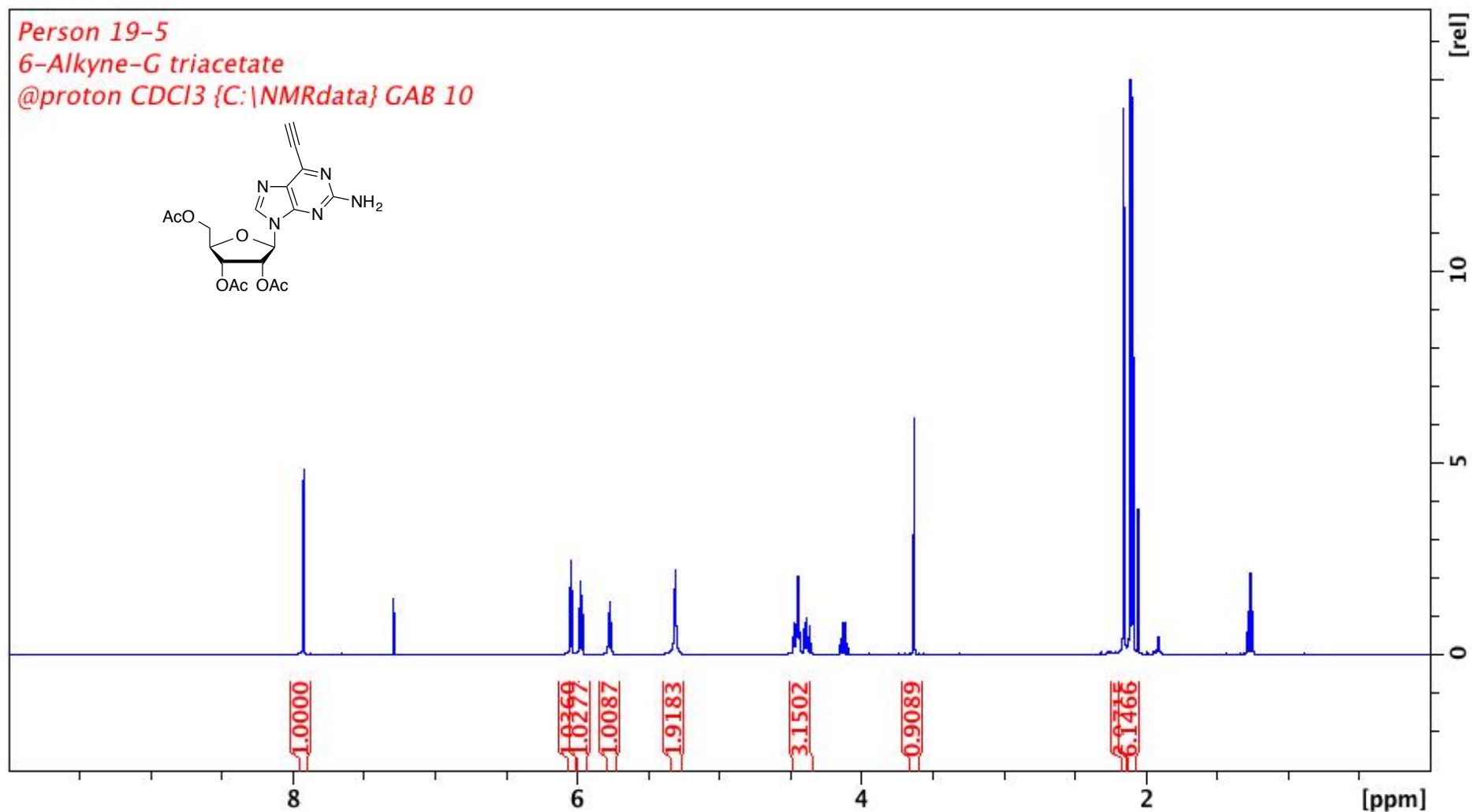
Appendix 26. ^{13}C -NMR spectrum of **118**. 125 MHz, CDCl_3 . Peak at 7.26 comes from CHCl_3 .



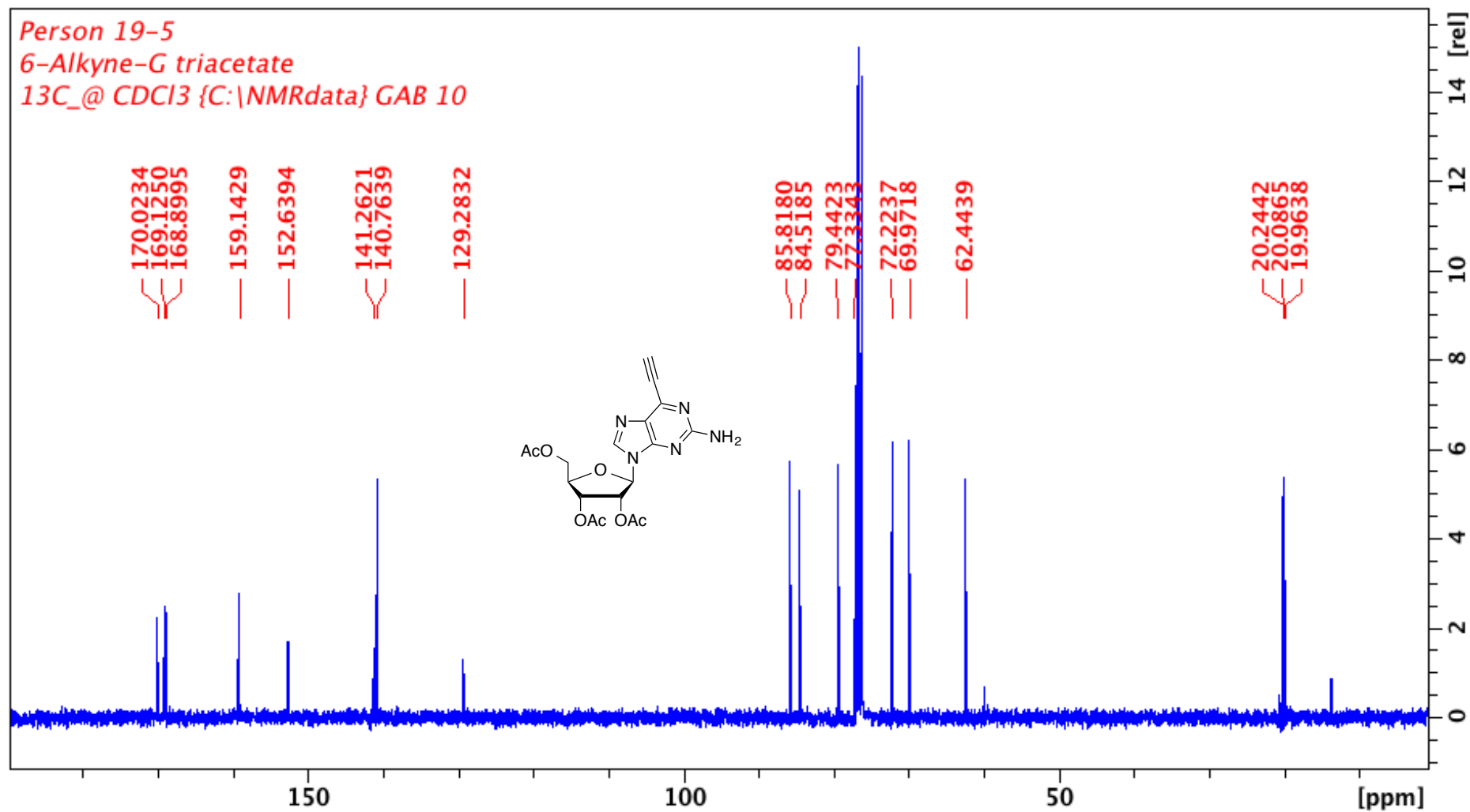
Appendix 27. ¹H-NMR spectrum of 119. 500 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃. Peak at 5.3 ppm comes from DCM.



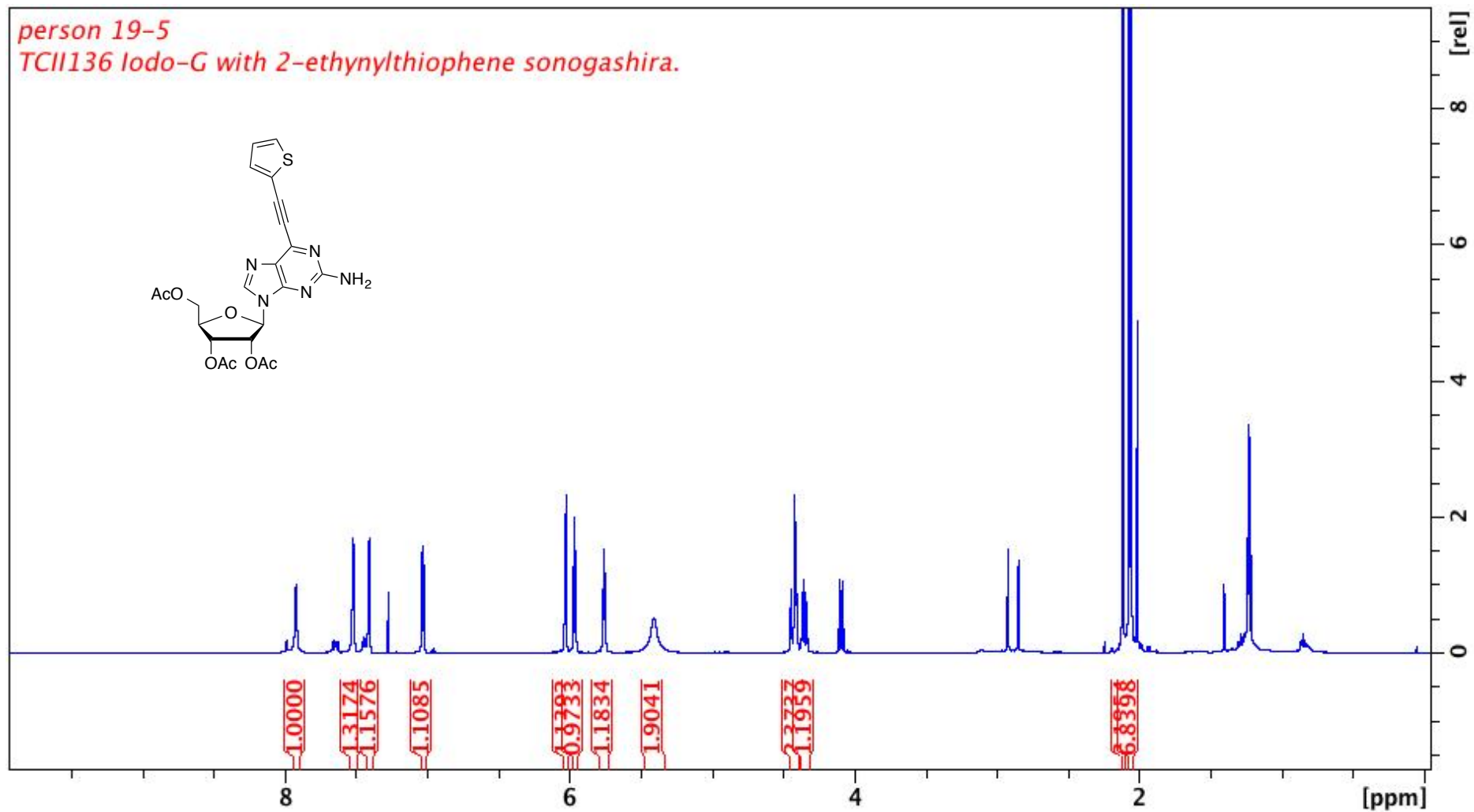
Appendix 28. ^{13}C -NMR spectrum of 119. 125 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .



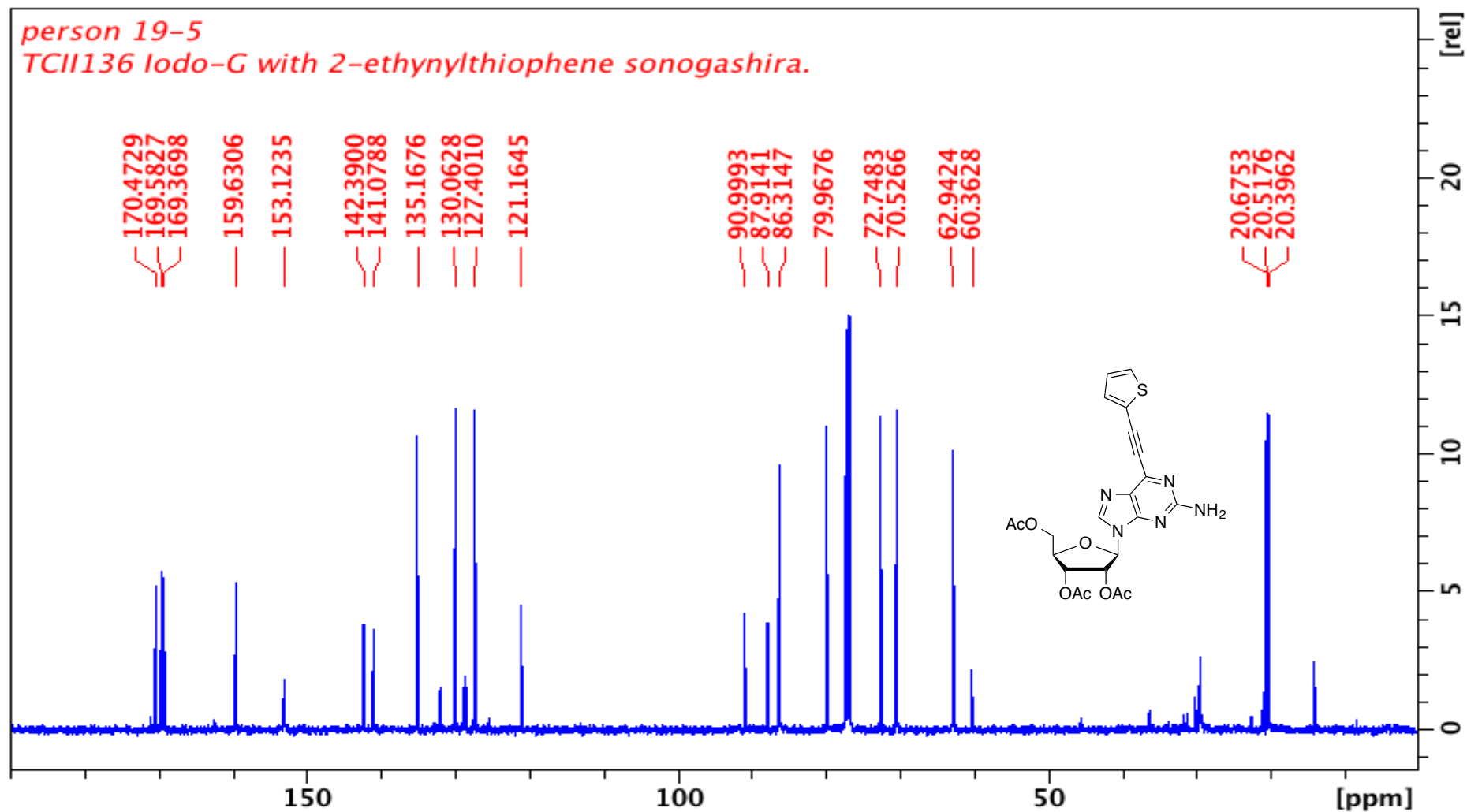
Appendix 29. ¹H-NMR spectrum of 114. 400 MHz, CDCl₃. Peaks at 1.26, 2.1 and 4.1 ppm come from EtOAc. Peak at 7.26 ppm comes from CHCl₃.



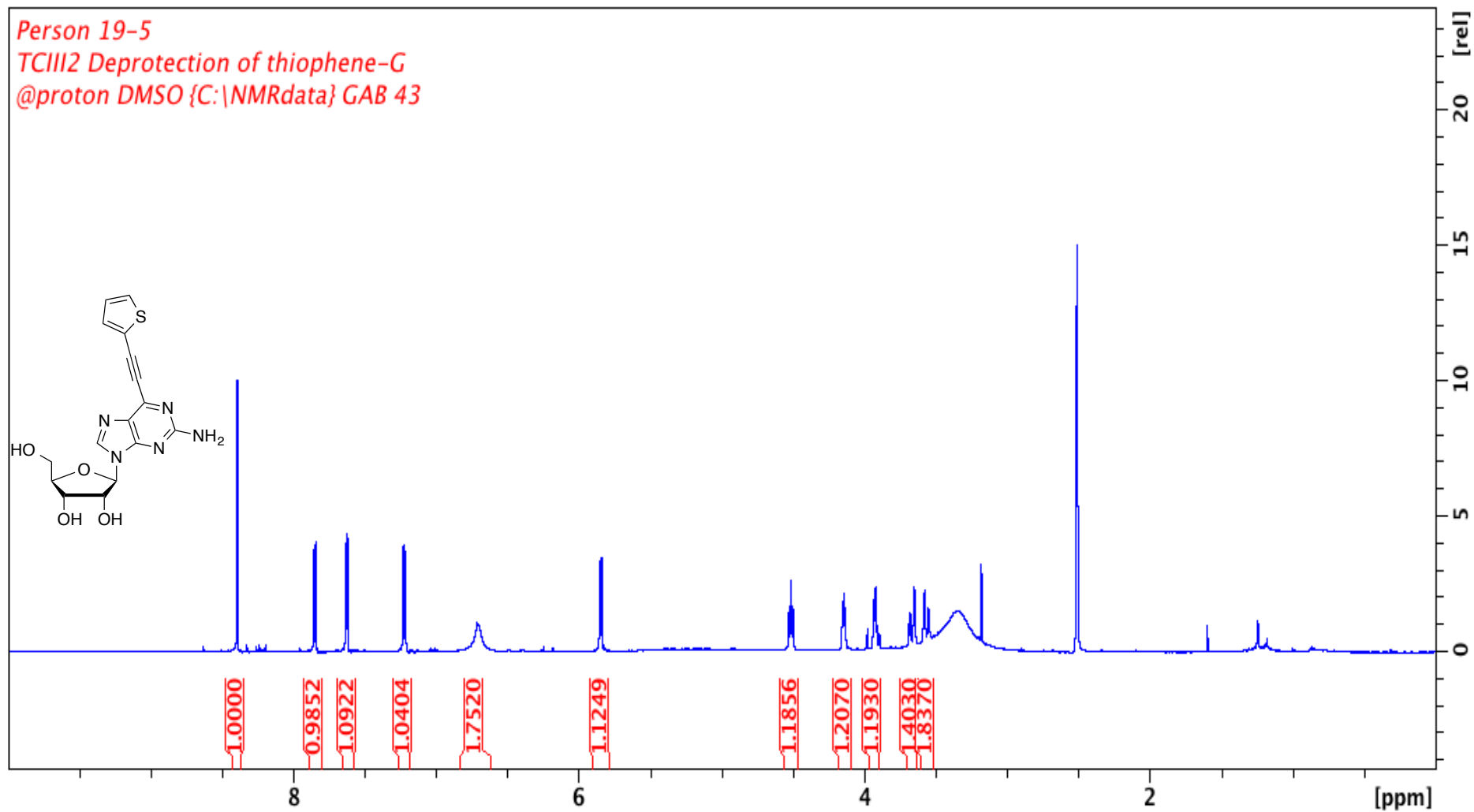
Appendix 30. ^{13}C -NMR spectrum of **114**. 100 MHz, CDCl_3 . Peaks at 14 and 60 ppm come from EtOAc. Peak at 78 ppm comes from CHCl_3 .



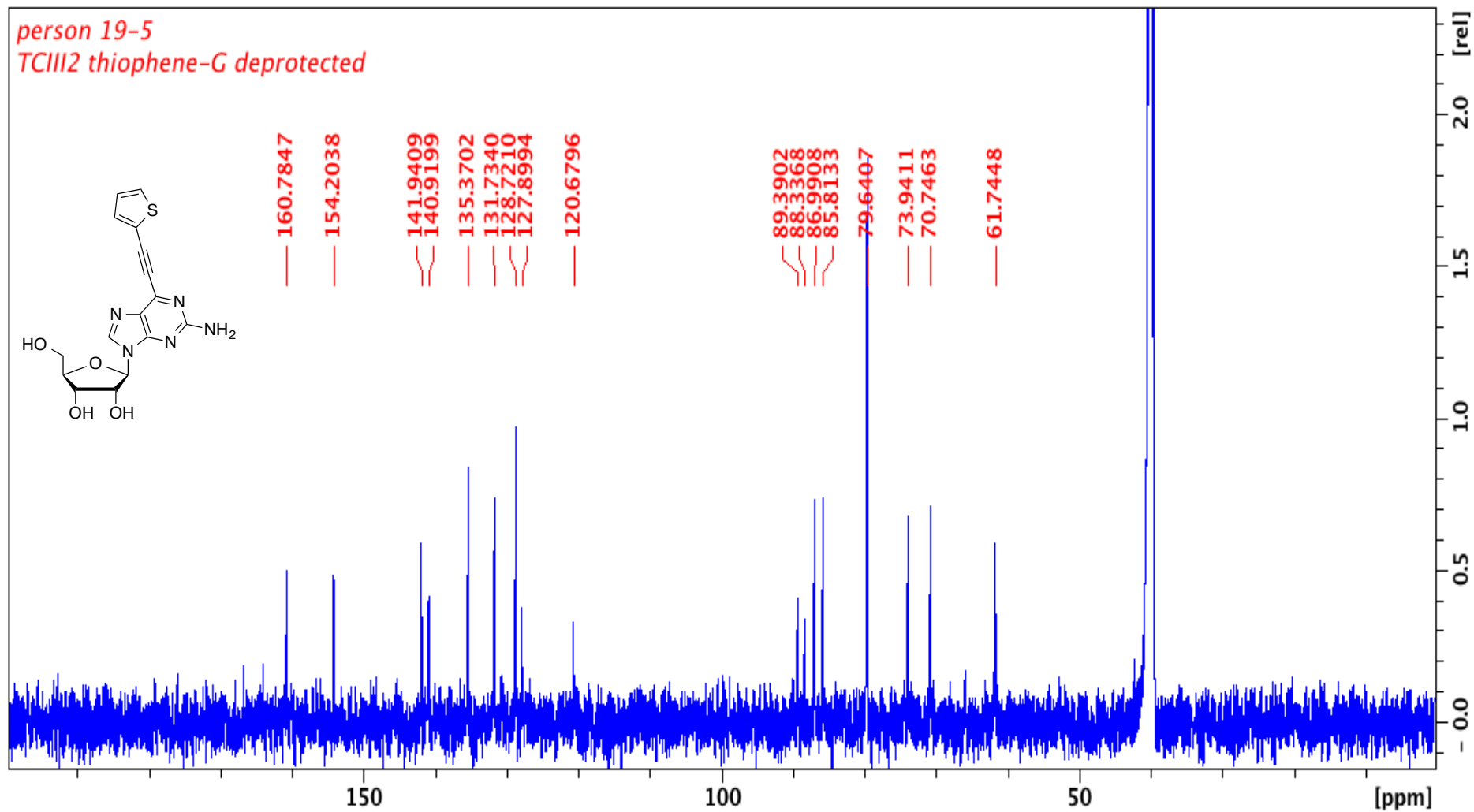
Appendix 31. $^1\text{H-NMR}$ spectrum of 120. 500 MHz, CDCl_3 . Peaks at 2.8, 2.9 and 8.0 ppm come from DMF. Peaks at 1.26, 2.1 and 4.1 come from EtOAc.



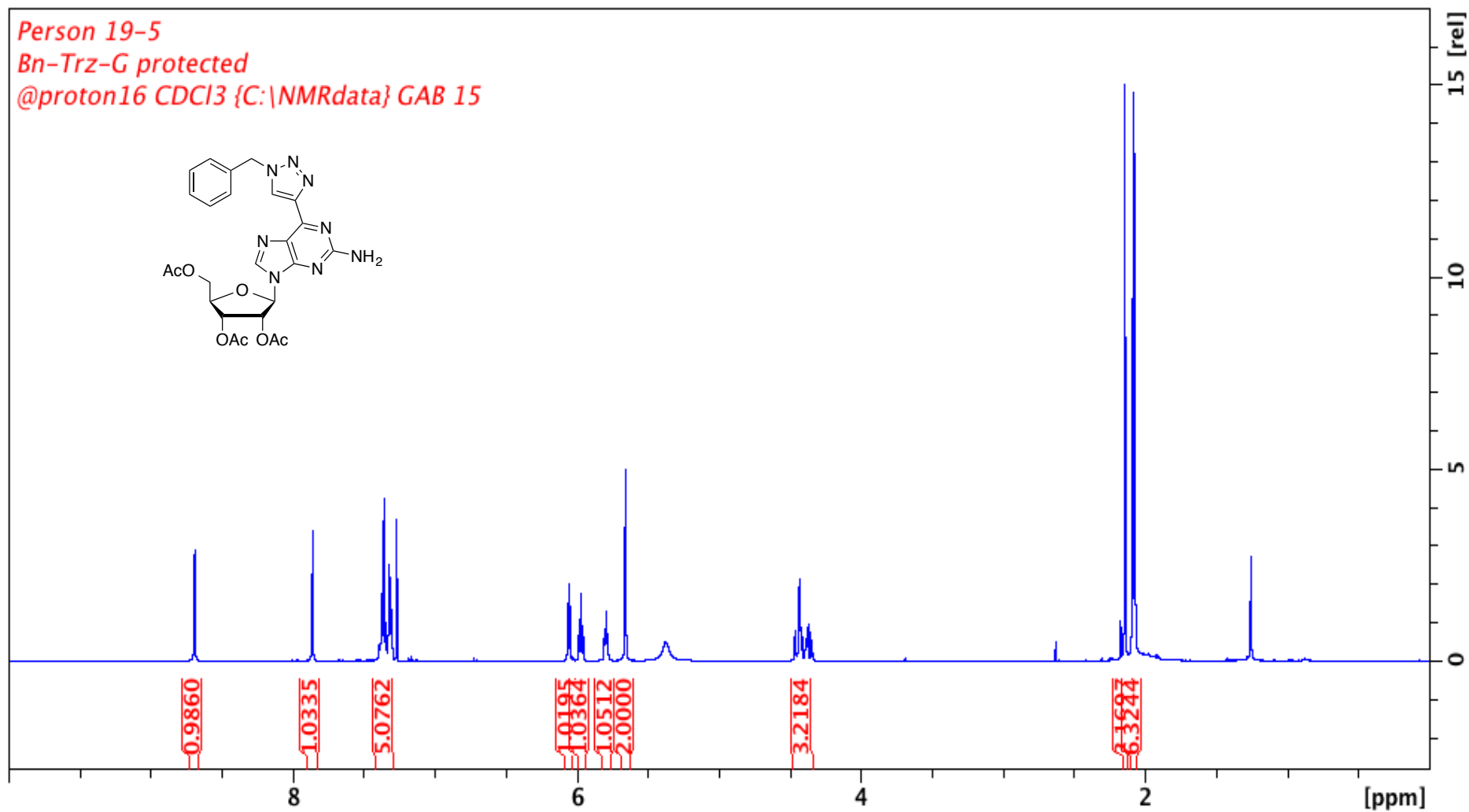
Appendix 32. ^{13}C -NMR spectrum of **120**. 125 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peaks at 14, 21 and 60 ppm come from EtOAc. Peak at 30 ppm comes from DMF.



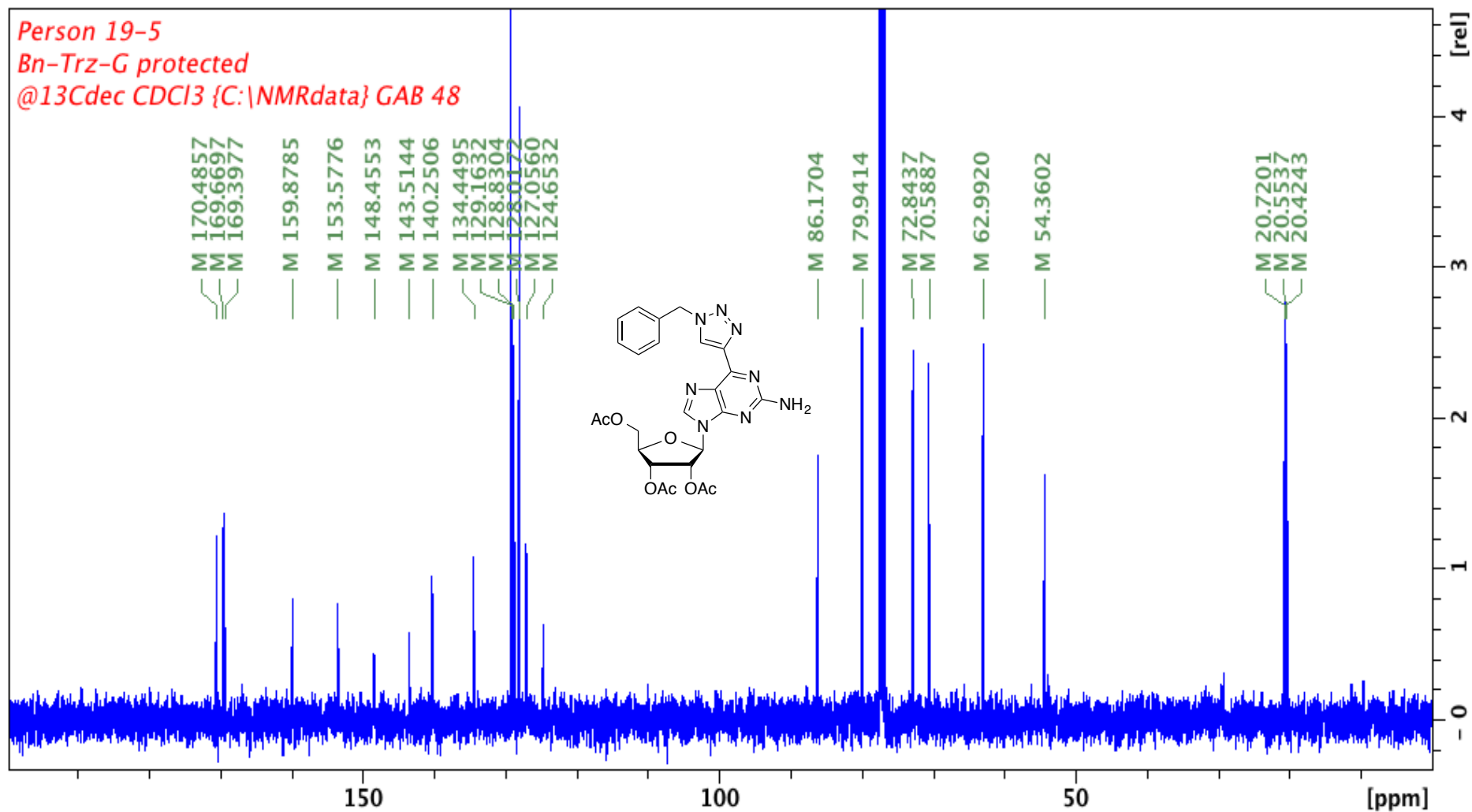
Appendix 33. $^1\text{H-NMR}$ spectrum of 121. 400 MHz, DMSO- d_6 . Peak at 2.5 ppm comes from DMSO. Peak at 3.33 ppm comes from water. Peak at 3.1 comes from MeOH.



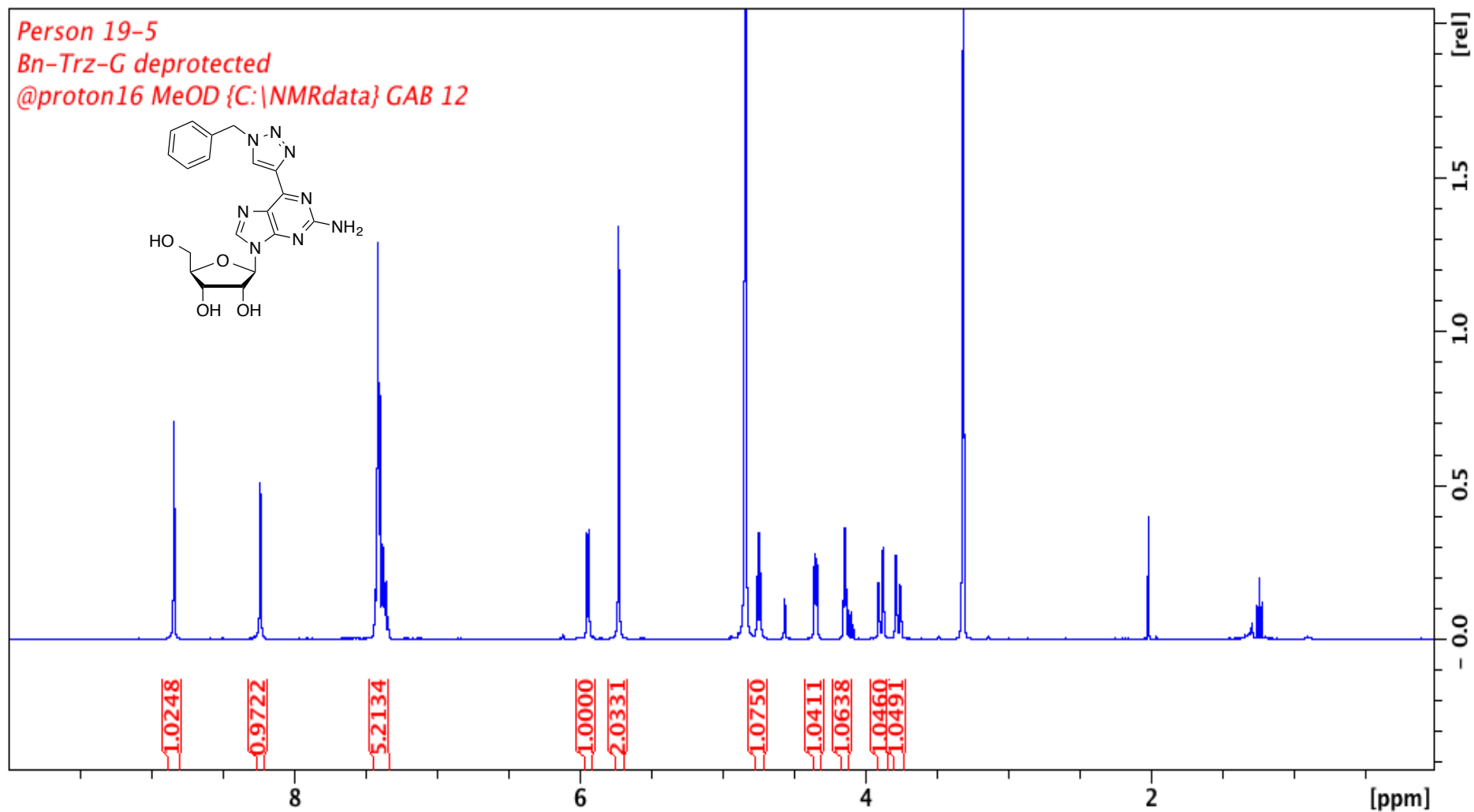
Appendix 34. ^{13}C -NMR spectrum of 121. 125 MHz, DMSO-d₆. Peak at 40 ppm comes from DMSO.



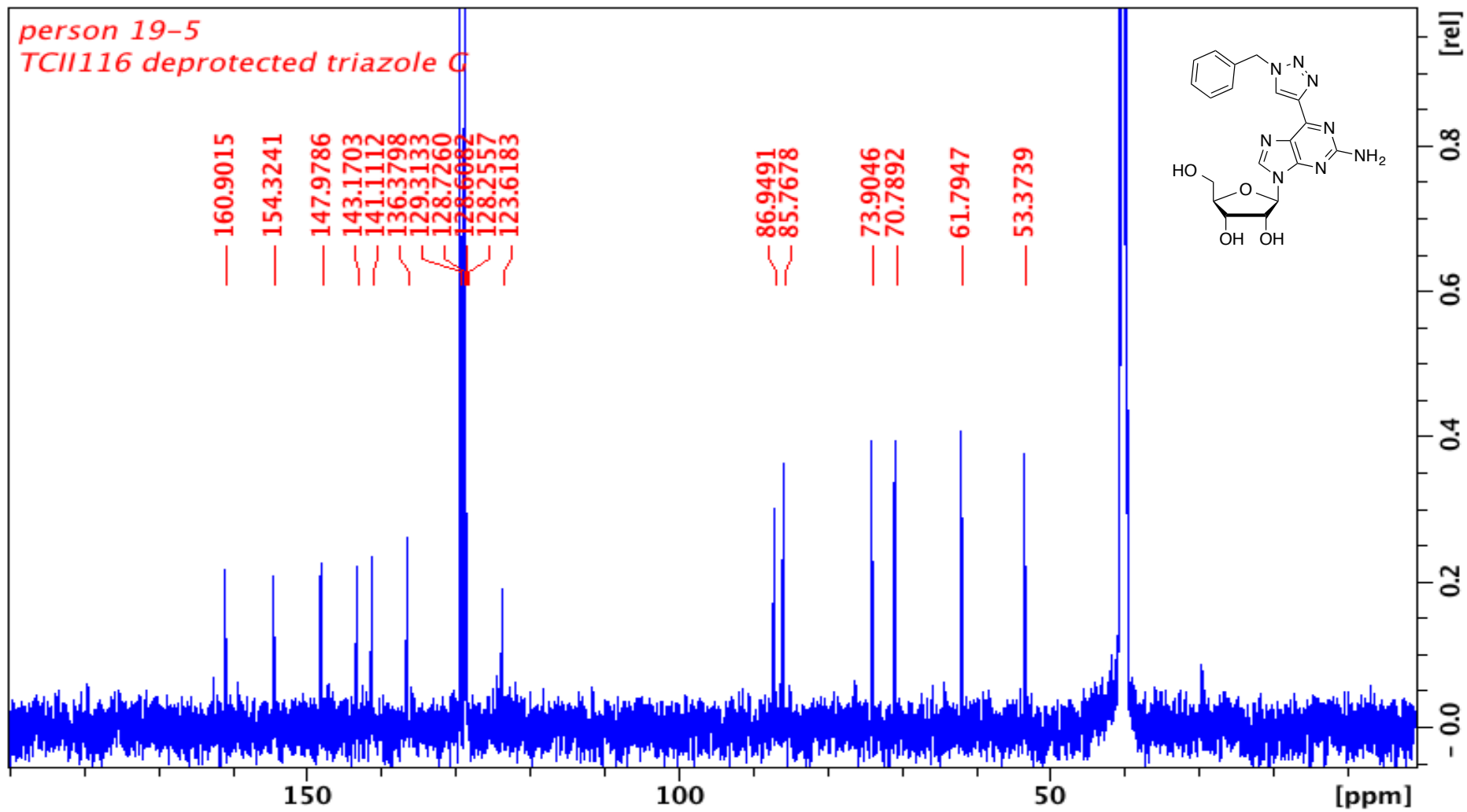
Appendix 35. ¹H-NMR spectrum of **130**. 400 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃. Peak at 5.3 comes from DCM. Peak at 1.25 ppm comes from grease.



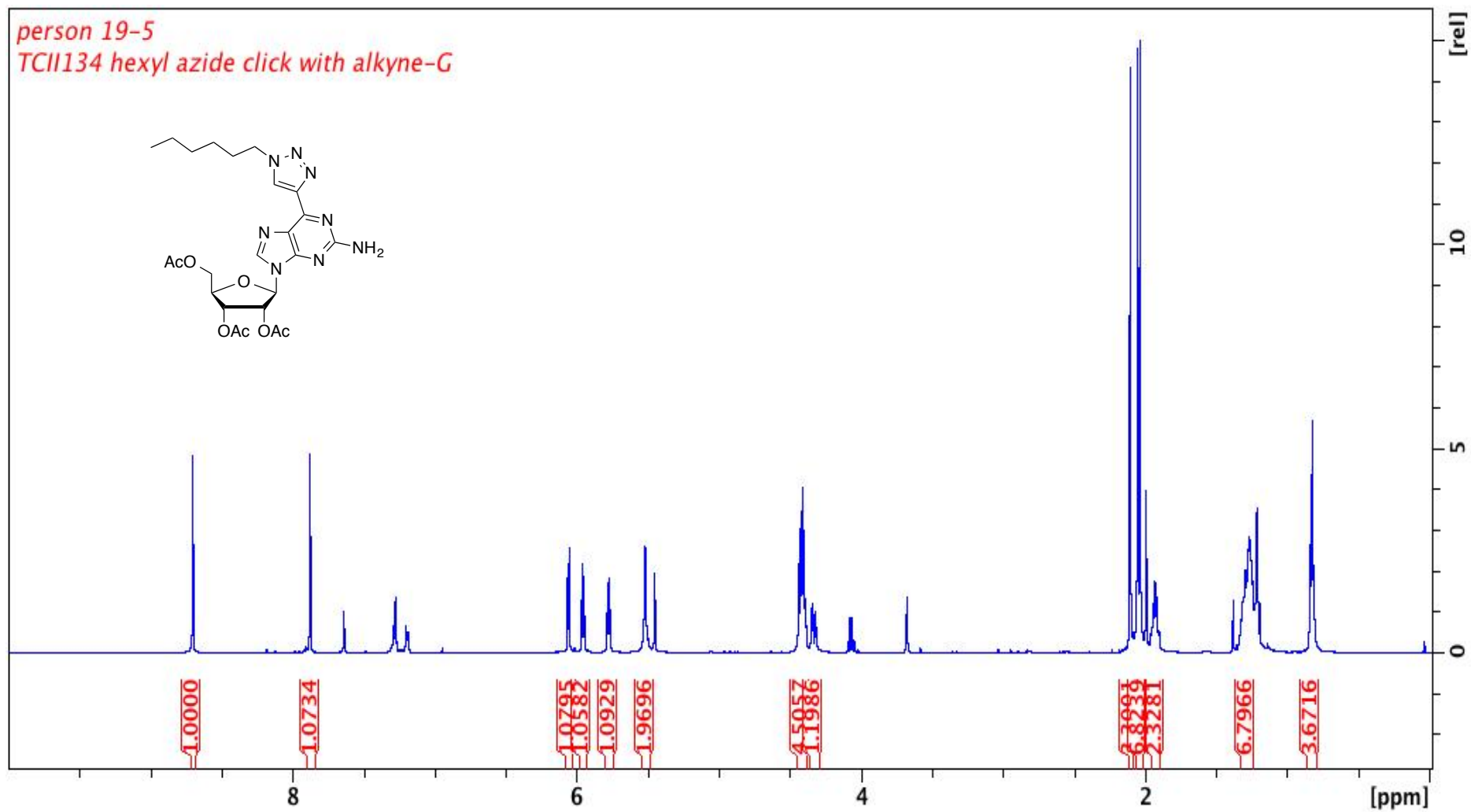
Appendix 36. ¹³C-NMR spectrum of **130**. 100 MHz, CDCl₃. Peak at 78 ppm comes from CHCl₃.



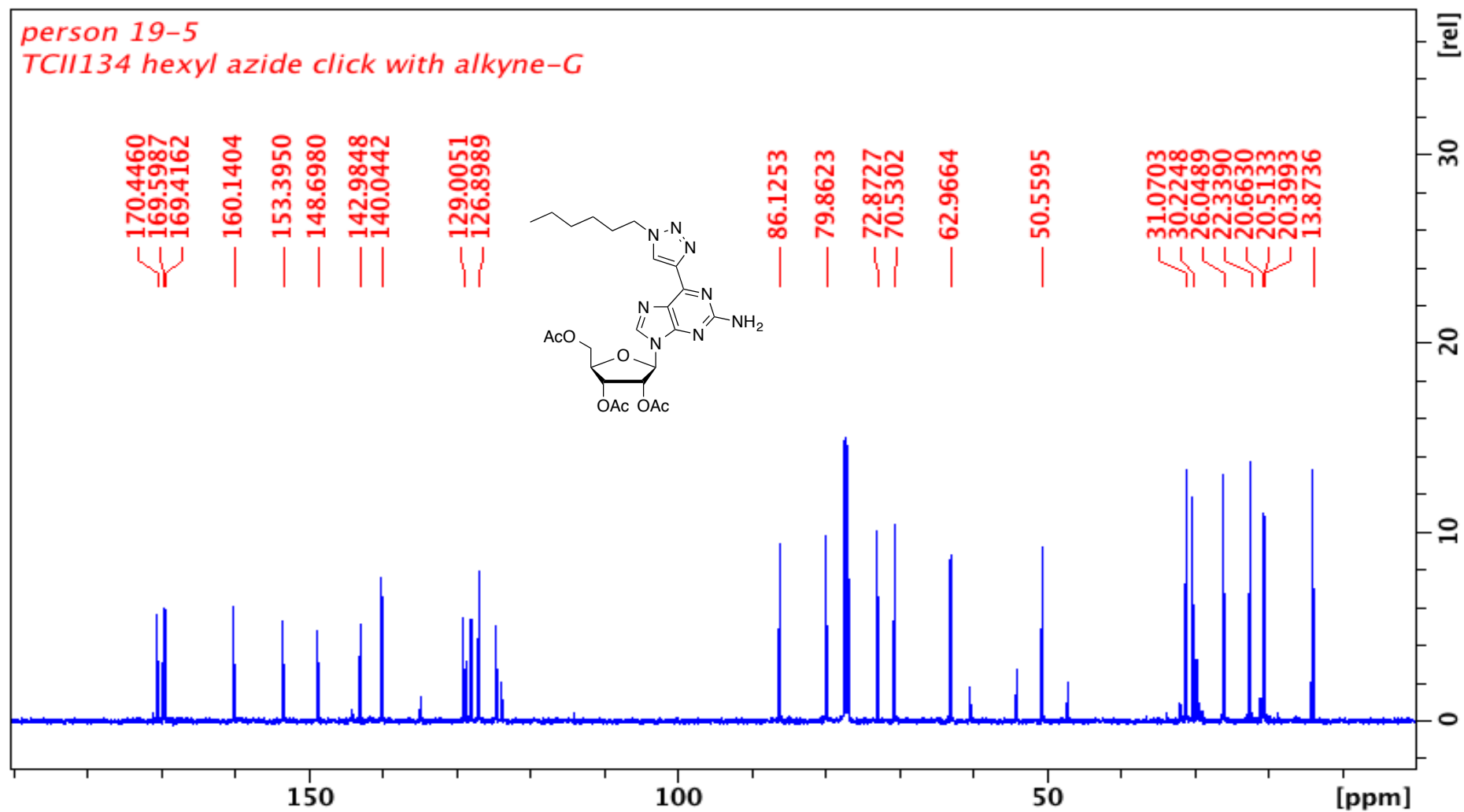
Appendix 37. $^1\text{H-NMR}$ spectrum of **124**. 400 MHz, MeOD. Peak at 4.8 ppm comes from water. Peak at 3.3 ppm comes from MeOH. Peak at 2.0 ppm comes from acetone. Peaks at 1.26 ppm comes from grease.



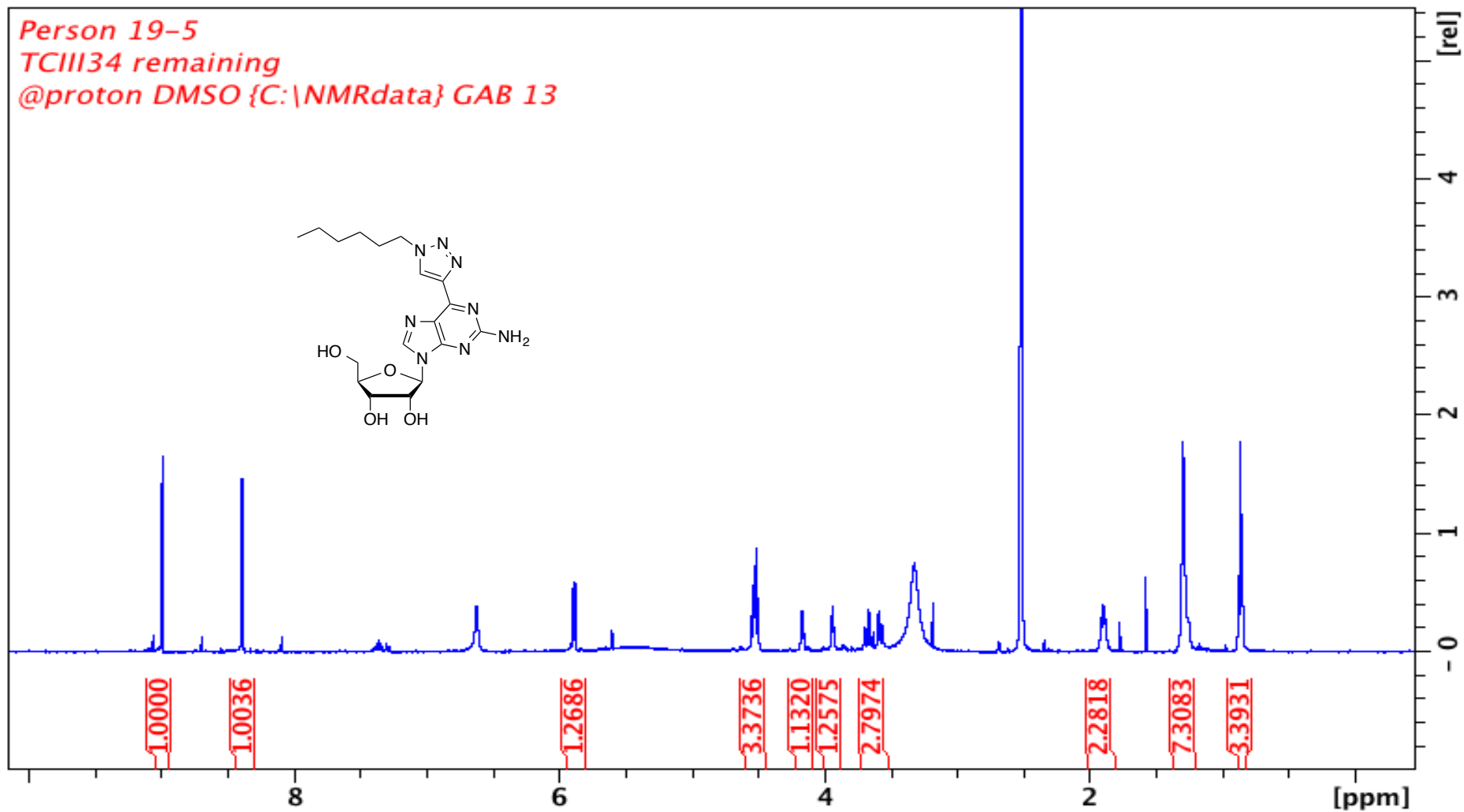
Appendix 38. ^{13}C -NMR spectrum of **124**. 125 MHz, DMSO- d_6 . Peak at 40 ppm comes from DMSO.



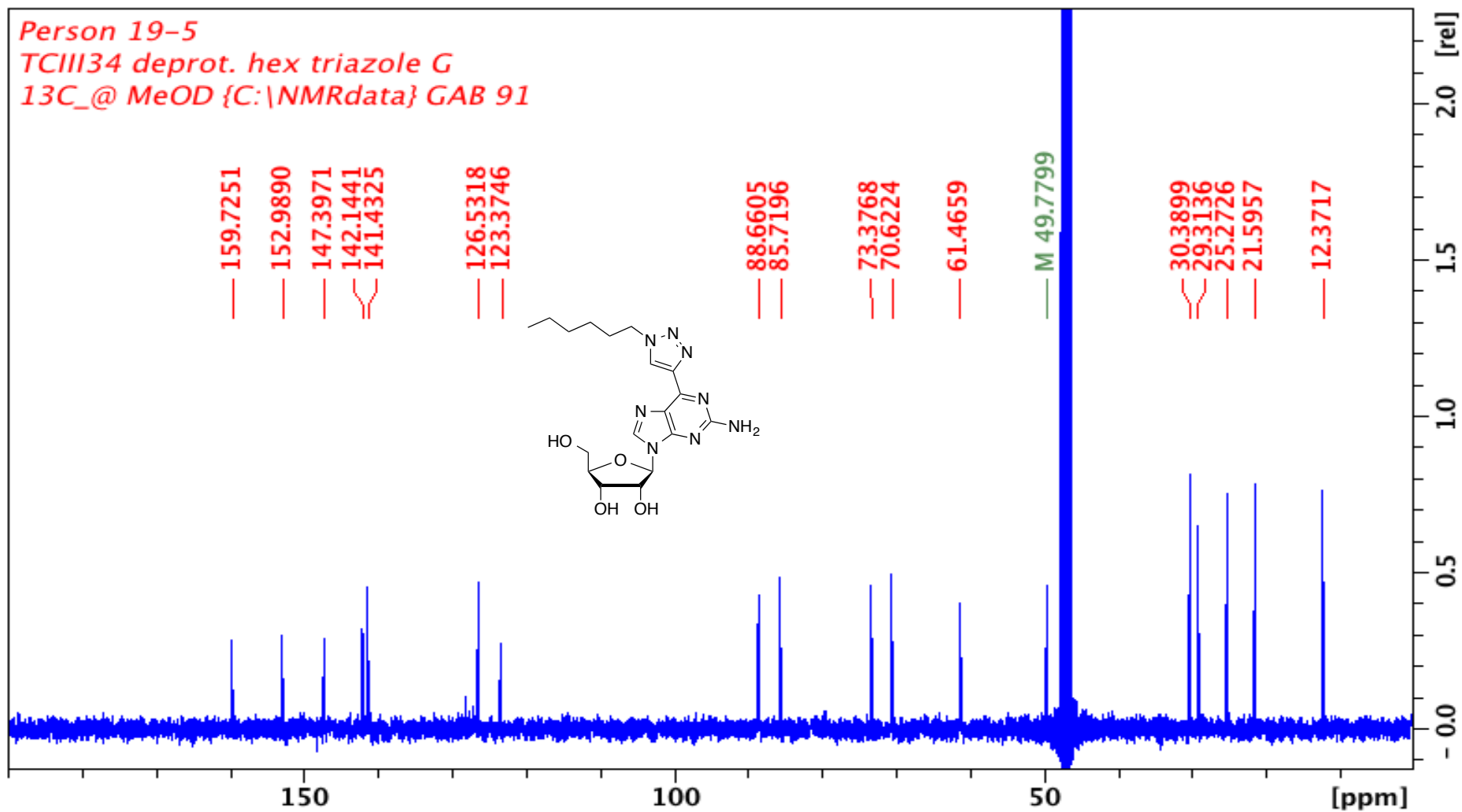
Appendix 39. $^1\text{H-NMR}$ spectrum of **131**. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peaks at 4.1 and 2.0 ppm come from EtOAc. Peaks at 7.25 come from TBTA.



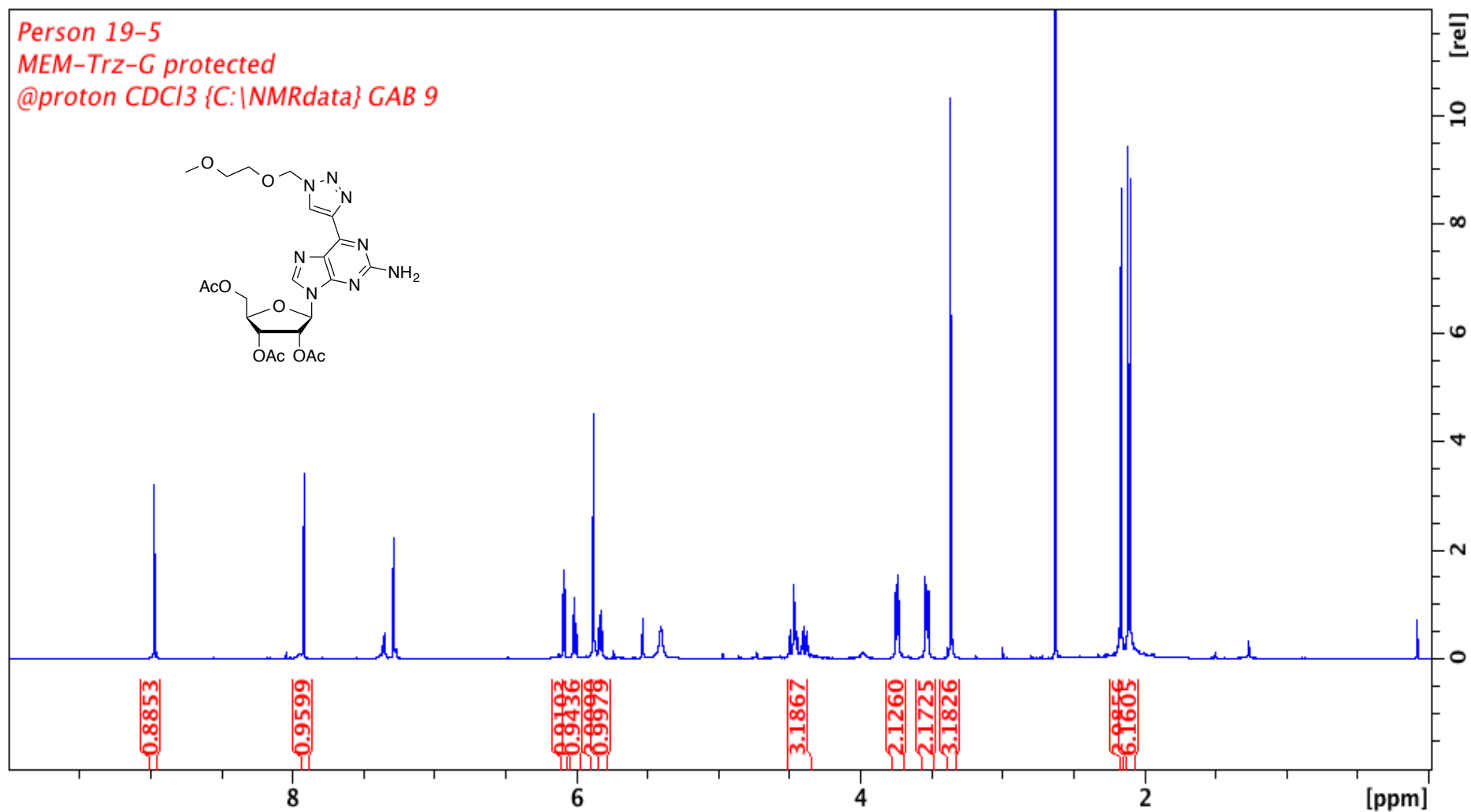
Appendix 40. ^{13}C -NMR spectrum of **131**. 125 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 62 ppm comes from EtOAc. Peaks at 127 ppm come from TBTA.



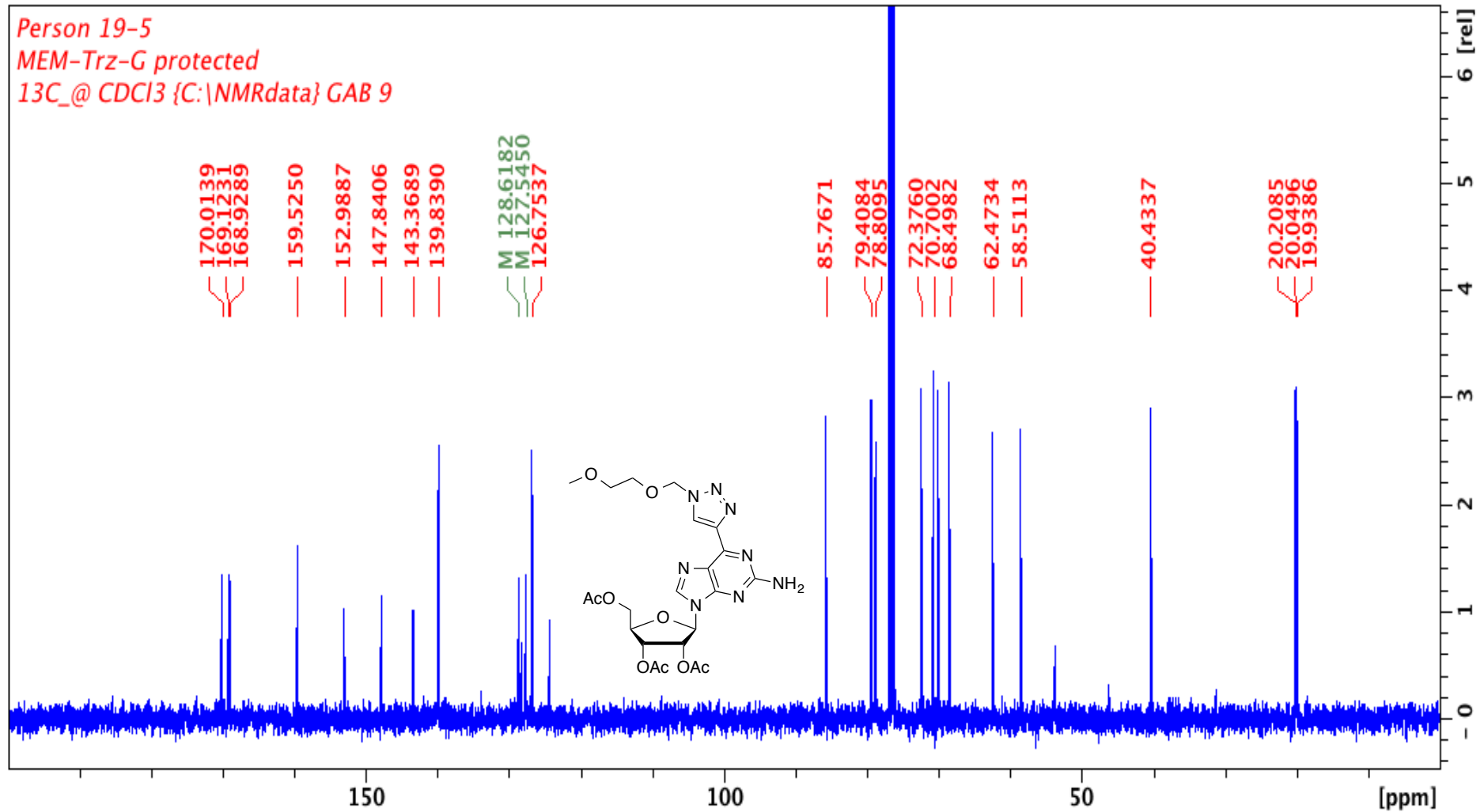
Appendix 41. ^1H -NMR spectrum of 125. 400 MHz, DMSO- d_6 . Peak at 2.5 ppm comes from DMSO. Peak at 3.3 ppm comes from water. Peak at 3.1 comes from MeOH.



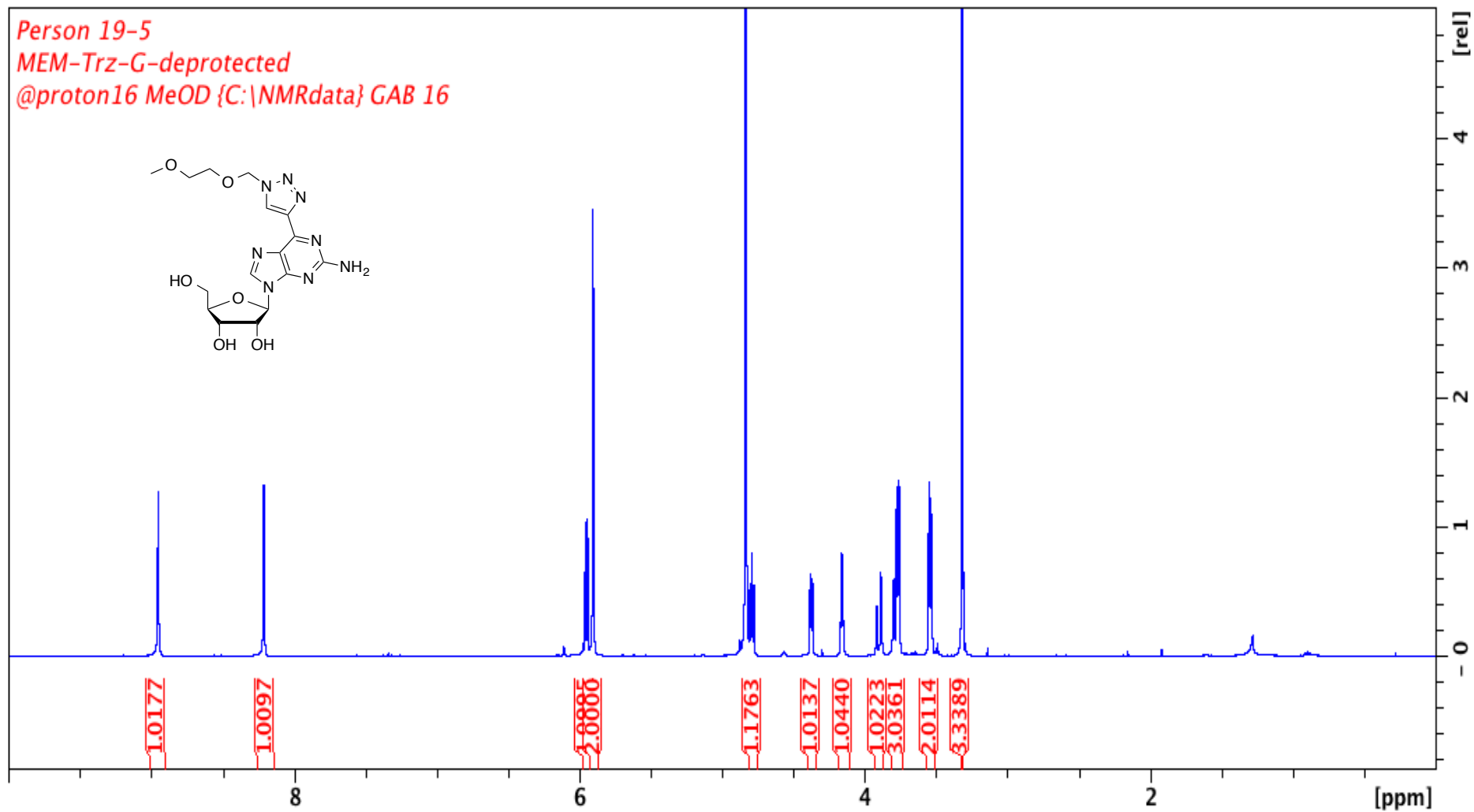
Appendix 42. ^{13}C -NMR spectrum of 125. 100 MHz, MeOD. Peak at 49 ppm comes from MeOH.



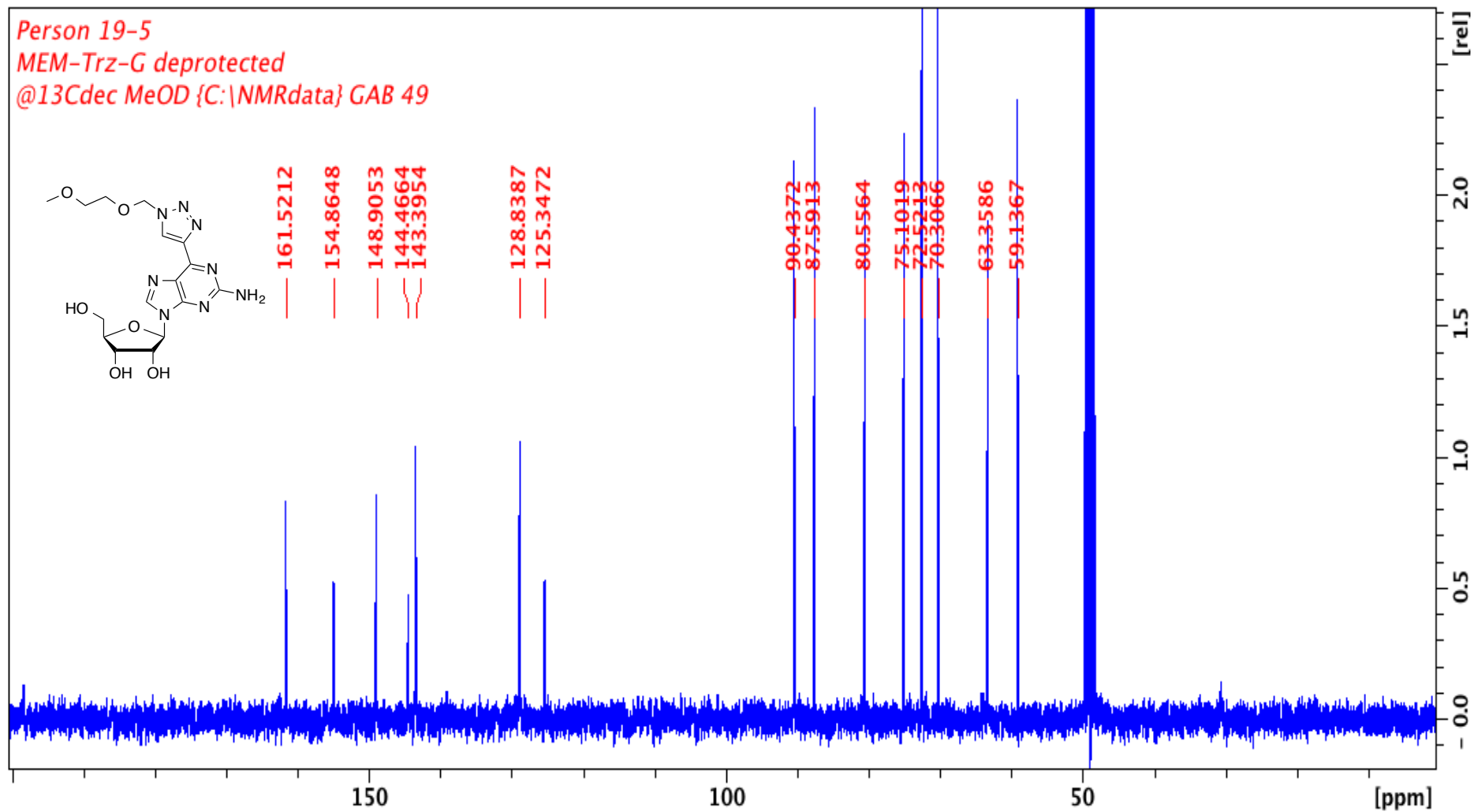
Appendix 43. ¹H-NMR spectrum of **132**. 400 MHz, CDCl₃. Peak at 2.5 ppm comes from DMSO. Peak at 5.3 ppm comes from DCM. Peaks at 7.2 ppm come from TBTA.



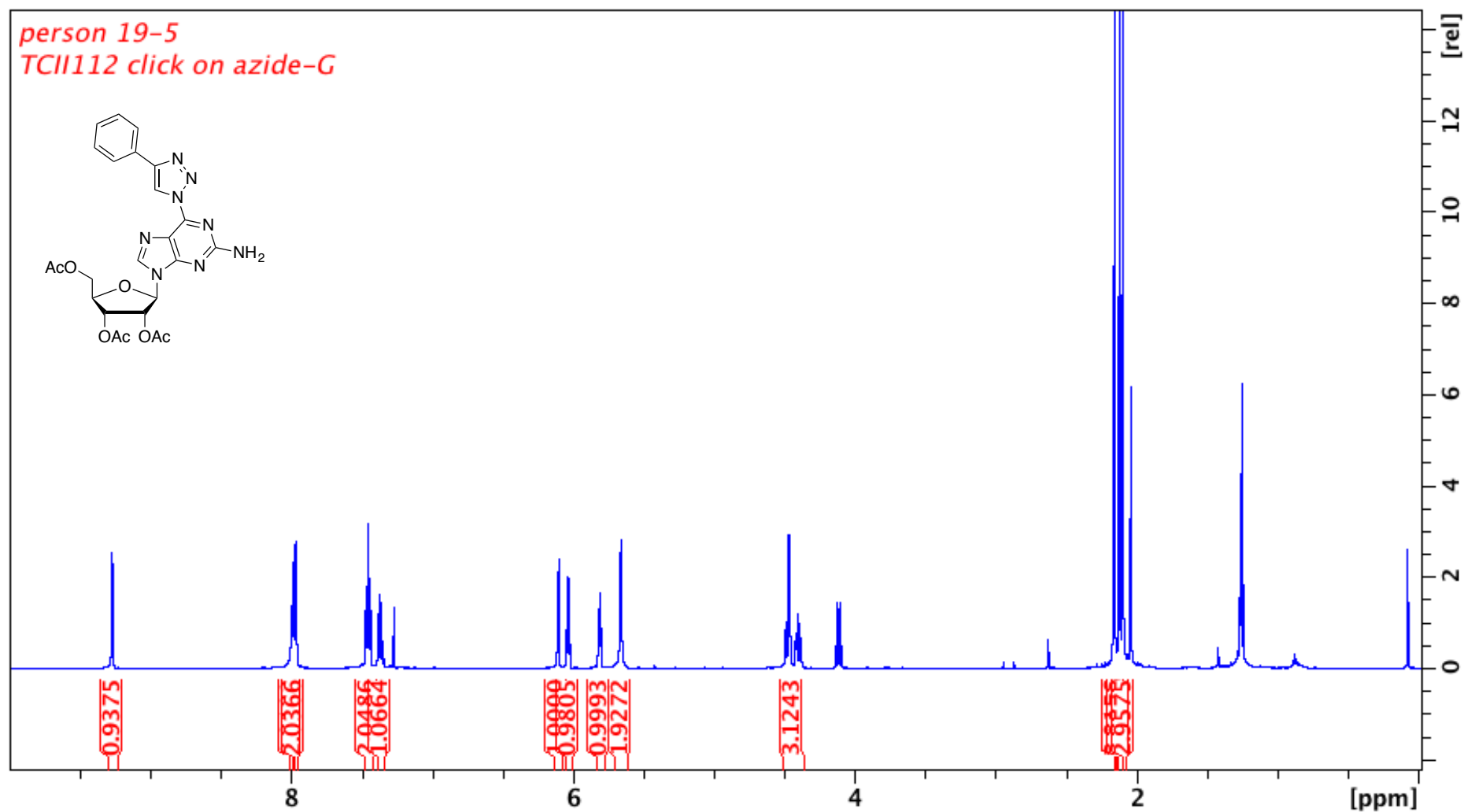
Appendix 44. ^{13}C -NMR spectrum of 132. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 125 and 53 ppm come from TBTA.



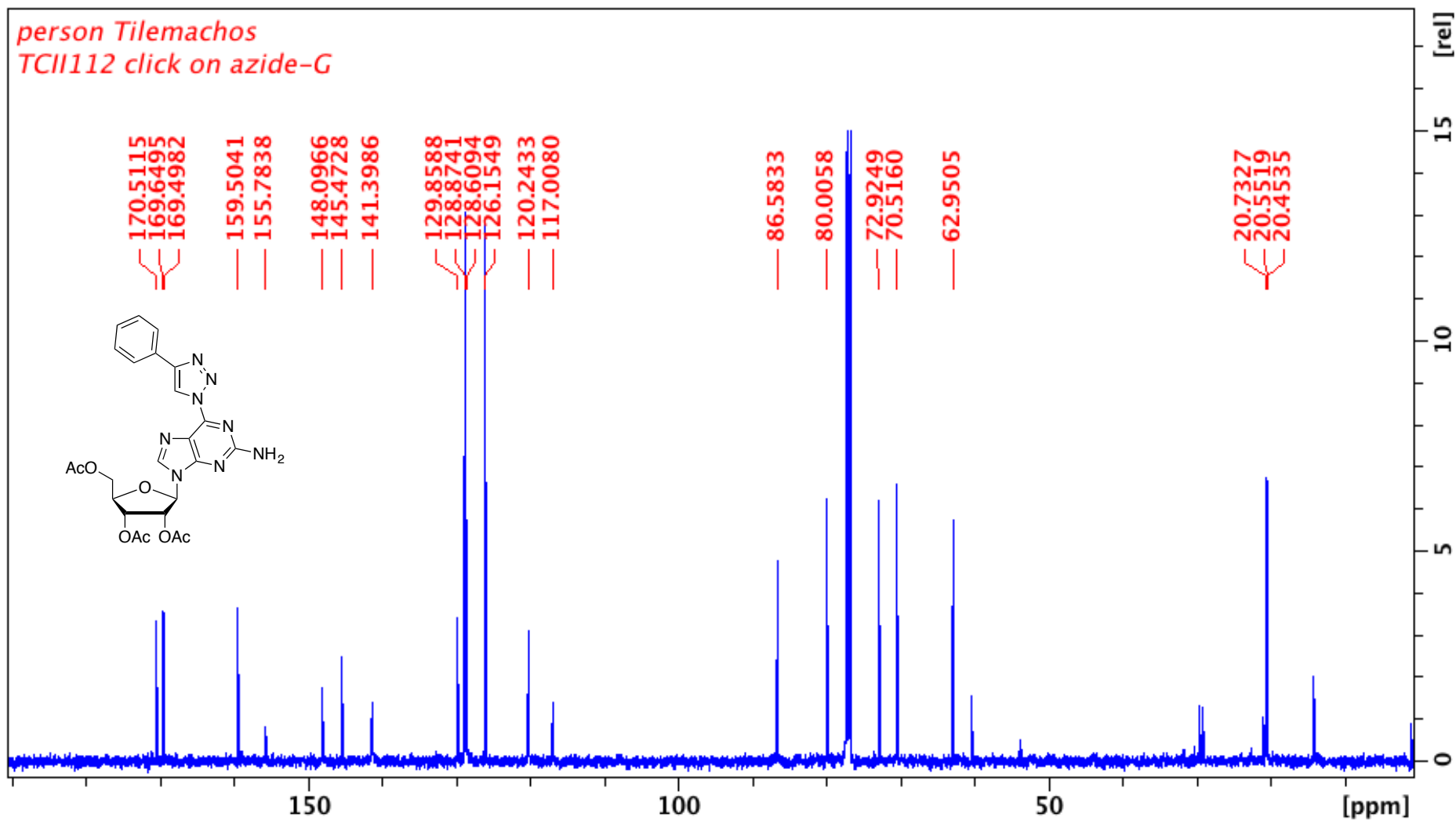
Appendix 45. ¹H-NMR spectrum of 126. 400 MHz, MeOD. Peak at 3.3 ppm comes from MeOH. Peak at 4.8 ppm comes from water.



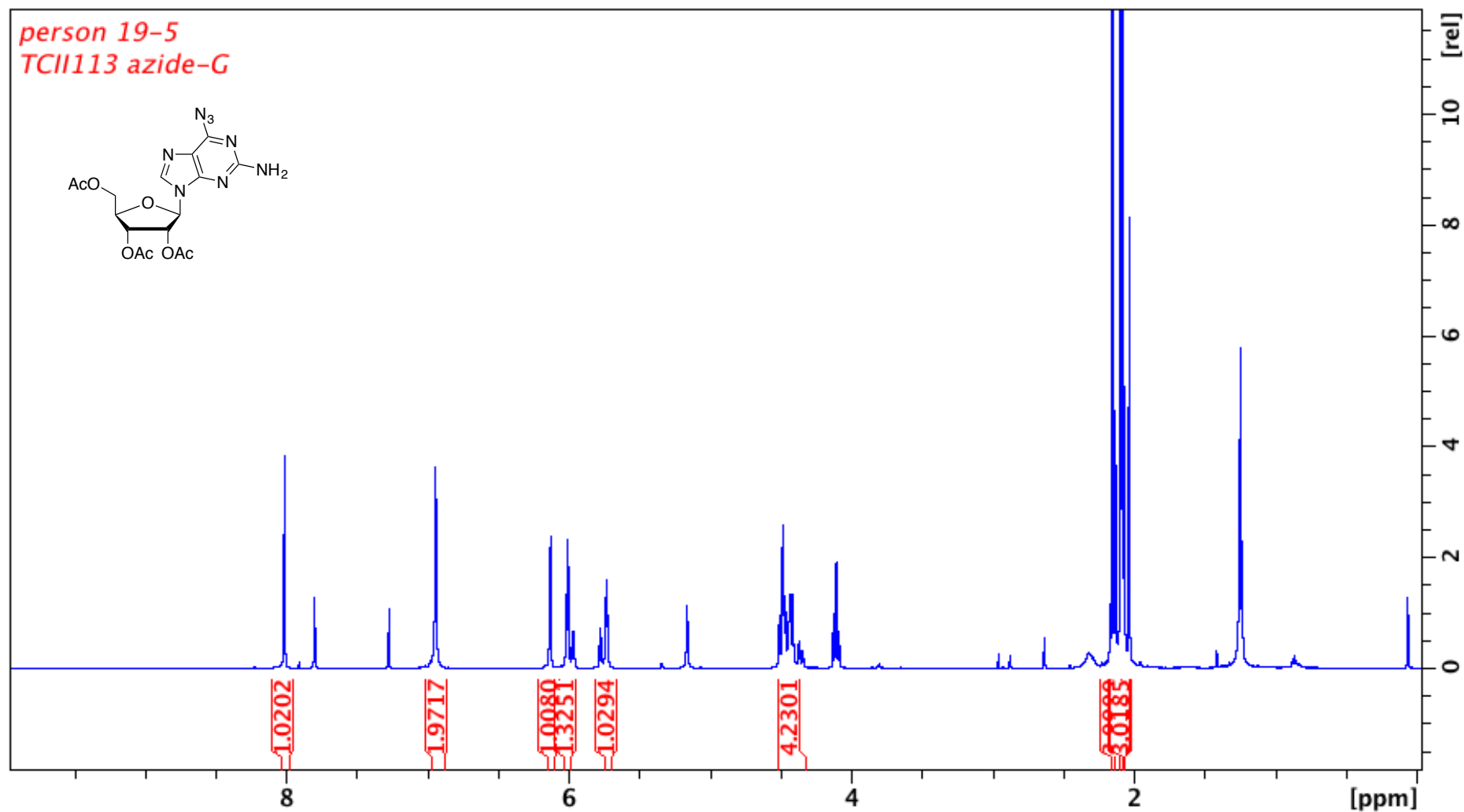
Appendix 46. ^{13}C -NMR spectrum of 126. 100 MHz, MeOD. Peak at 49 ppm comes from MeOH.



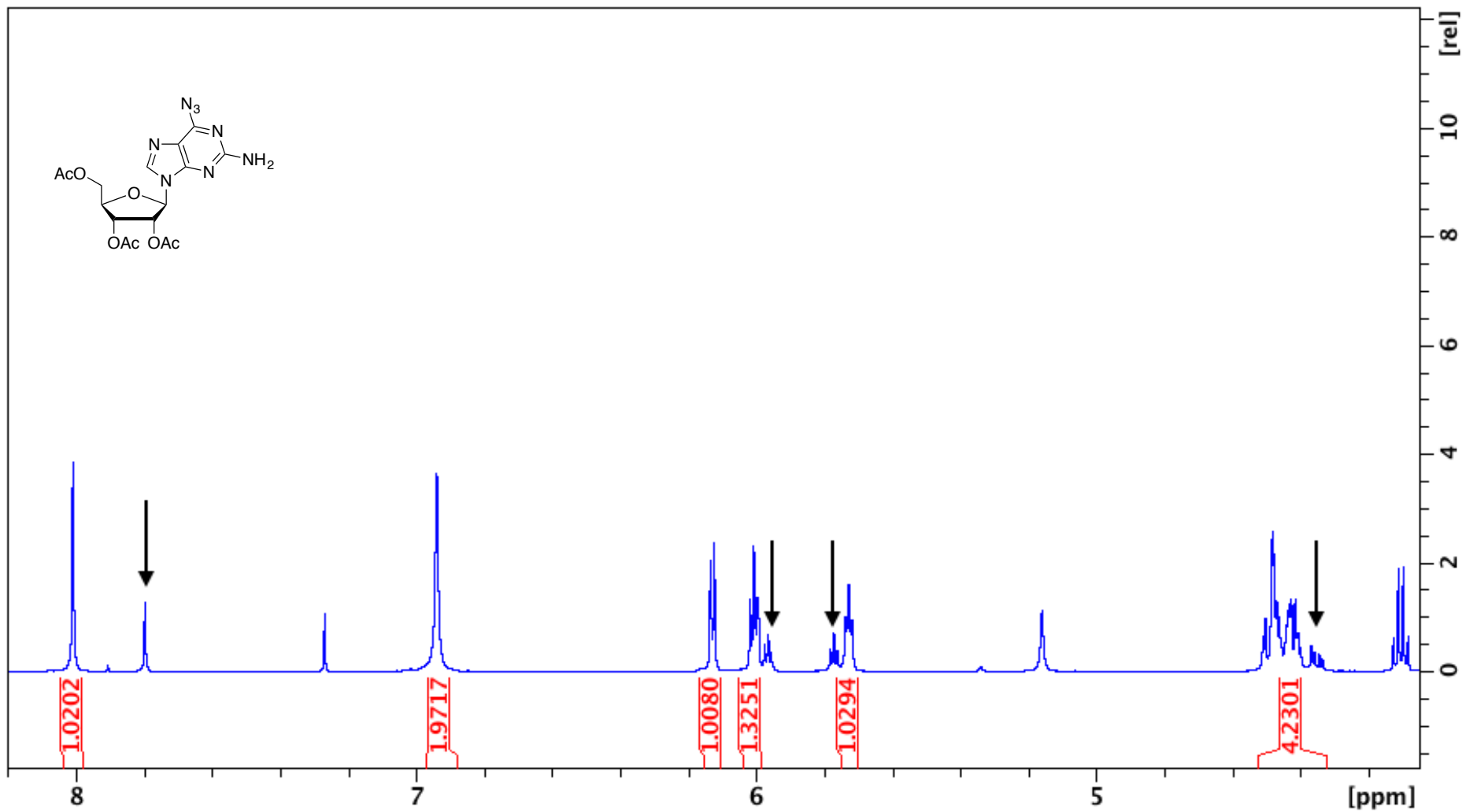
Appendix 47. $^1\text{H-NMR}$ spectrum of **152**. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peakst at 1.26 2.1 and 4.1 ppm come from EtOAc. Peaks at 0.08 and 0.84 ppm come from grease.



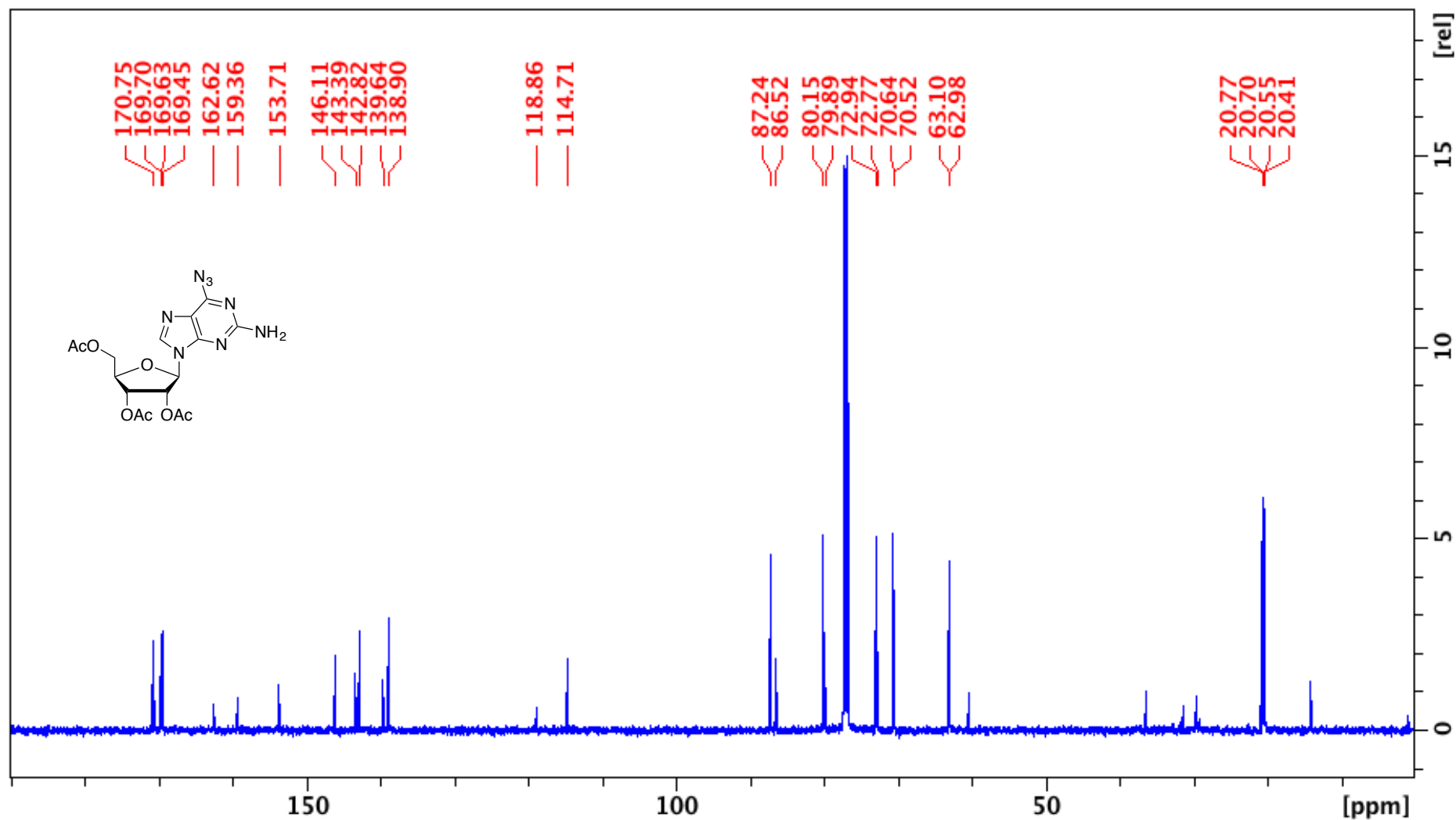
Appendix 48. ^{13}C -NMR spectrum of **152**. 125 MHz, CDCl_3 . Peak at 78 ppm comes from CDCl_3 . Peaks at 21, 171 and 60 ppm come from EtOAc. Peak at 1.2 and 29 ppm come from grease.



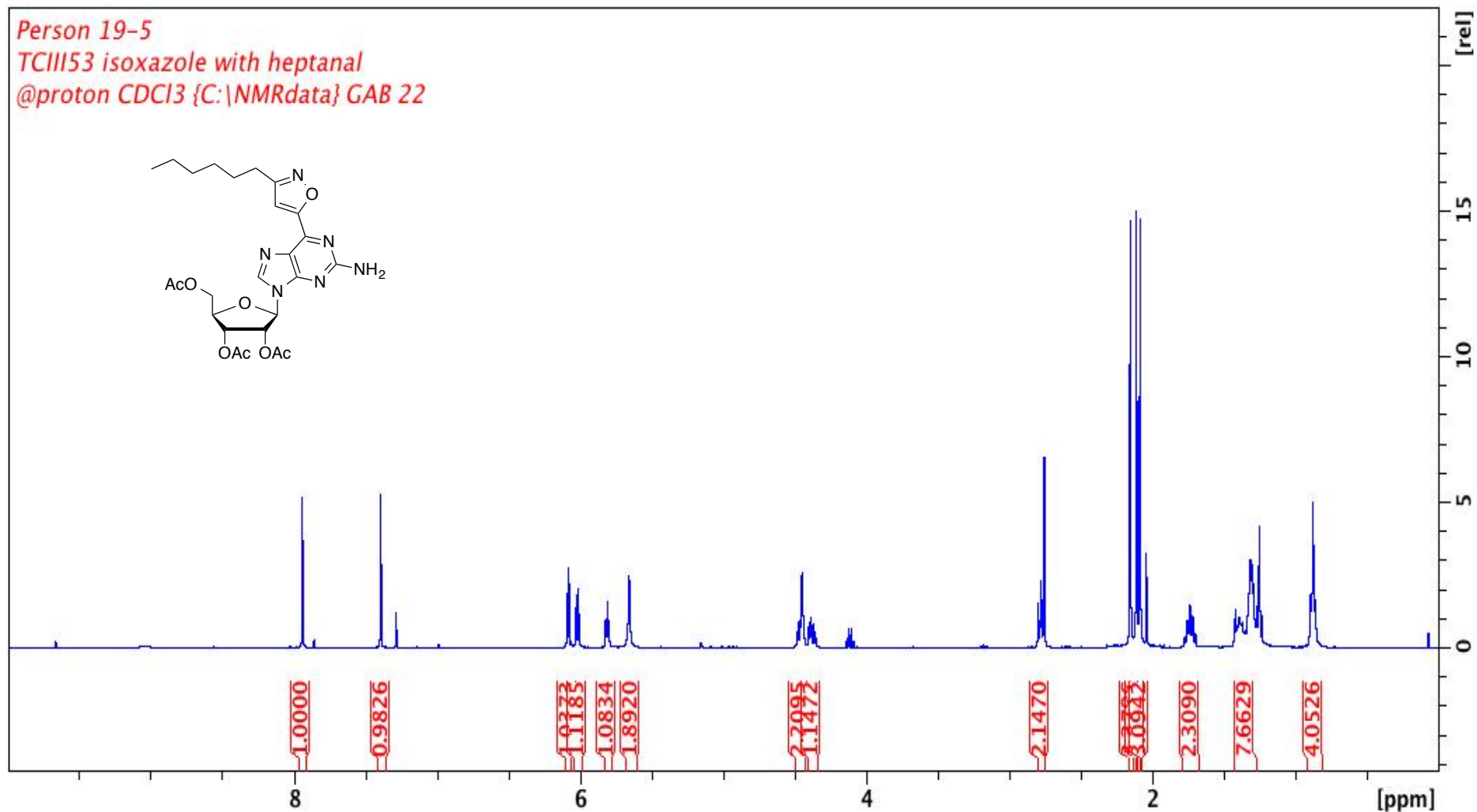
Appendix 49. $^1\text{H-NMR}$ spectrum of **139**. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peak at 5.3 comes from DCM. Peaks at 1.26, 2.1 and 4.1 ppm come from EtOAc. Peak at 0.08 ppm comes from grease.



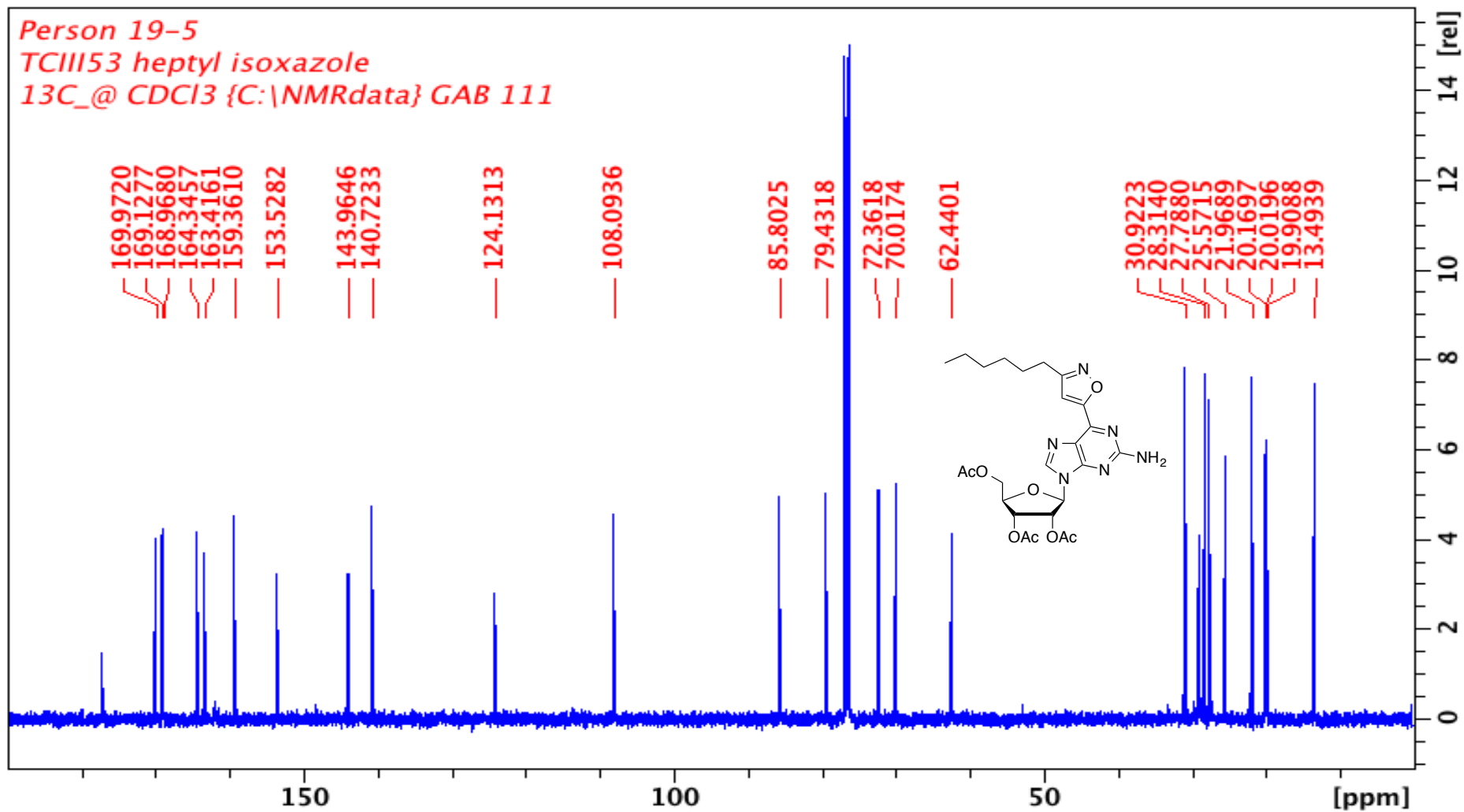
Appendix 50. Expanded ¹H-NMR spectrum of **139** showing the peaks corresponding to the minor isomer.



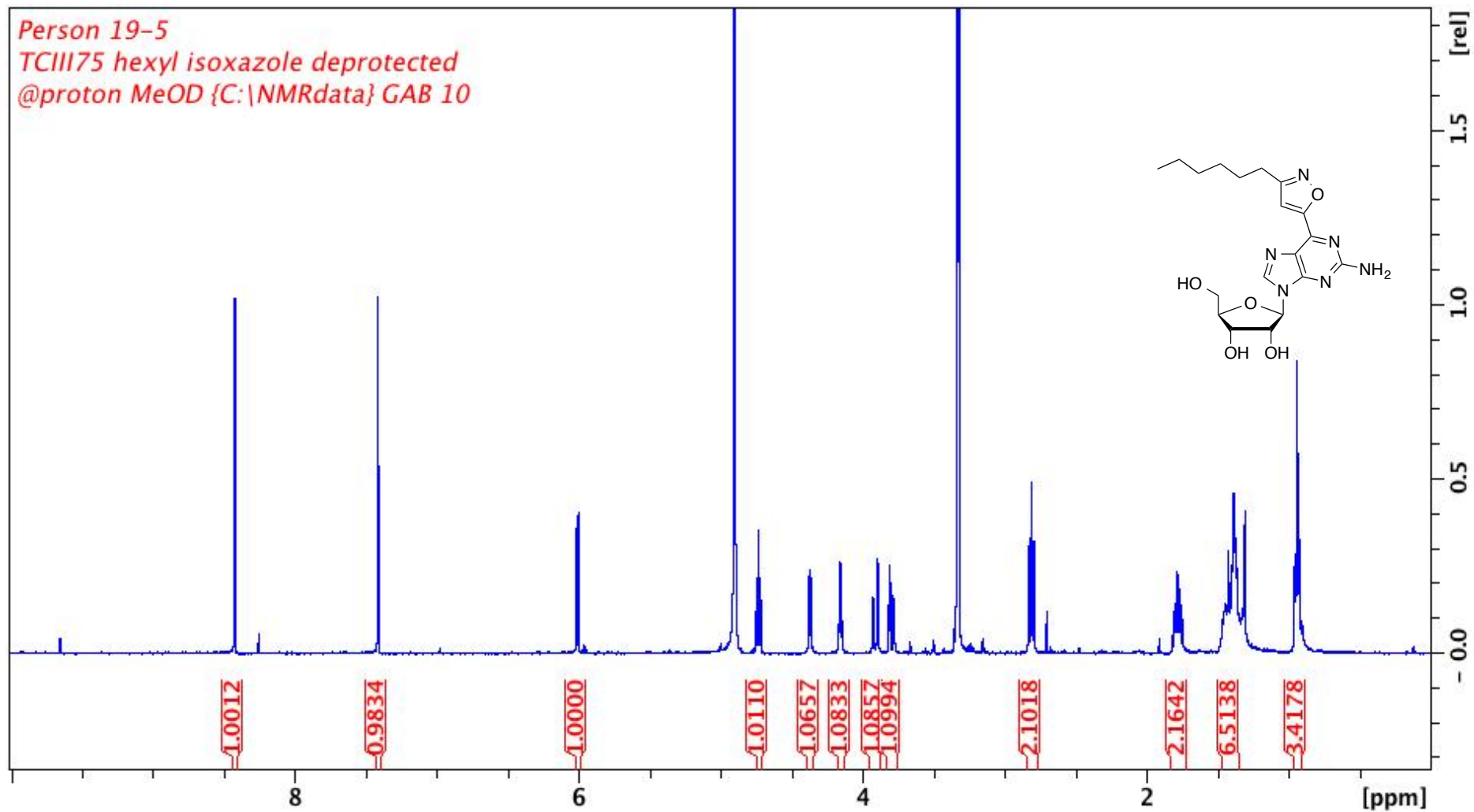
Appendix 51. ¹³C-NMR spectrum of **139**. 125 MHz, CDCl₃. Peak at 78 ppm comes from CDCl₃. Peaks at 21, 171 and 60 ppm come from EtOAc. Peak at 1.2 and 29 ppm come from grease.



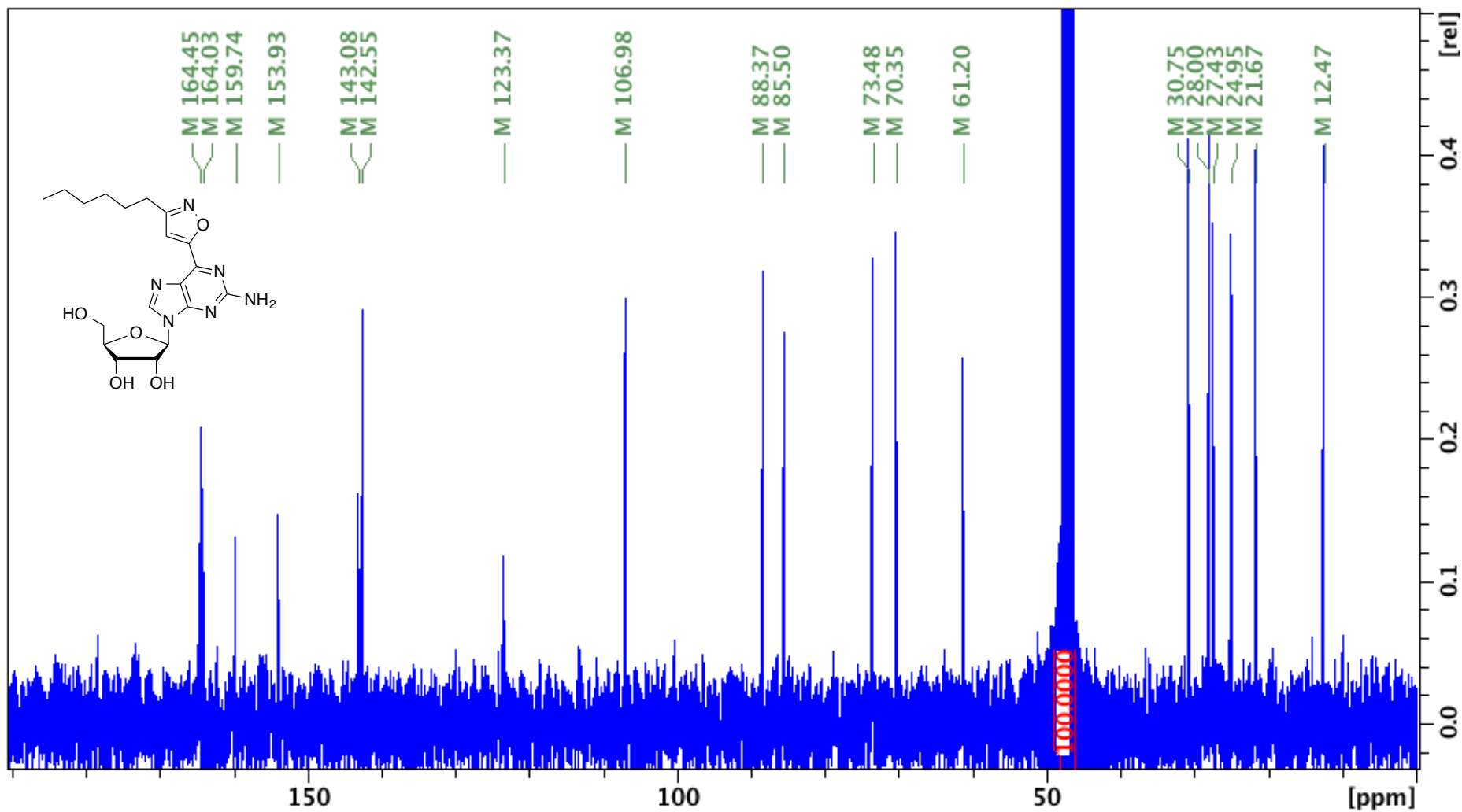
Appendix 52. ¹H-NMR spectrum of 157. 400 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃. Peaks at 1.26, 2.0 and 4.1 come from EtOAc.



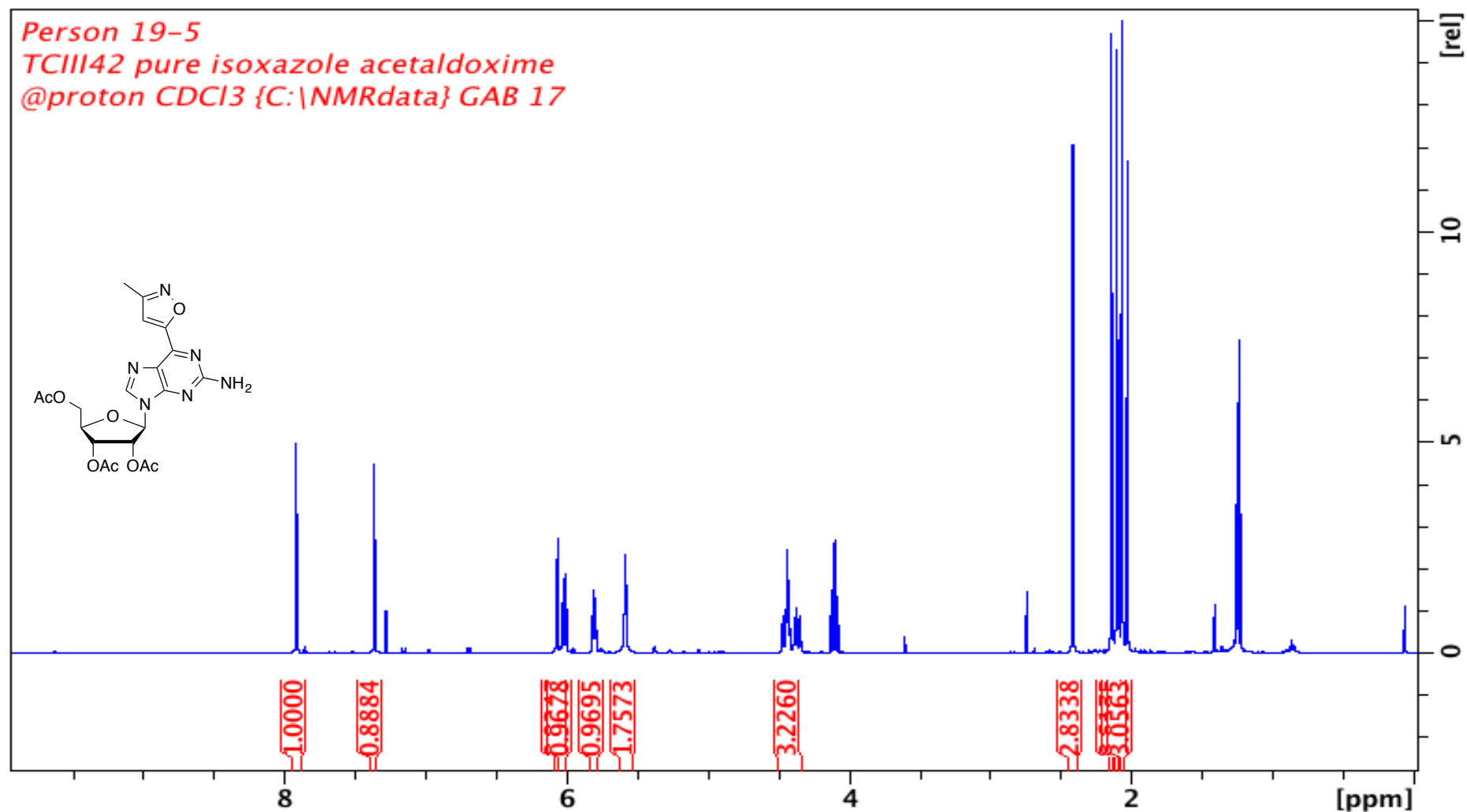
Appendix 53. ^{13}C -NMR spectrum of 157. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peaks at 29 and 177 ppm come from EtOAc.



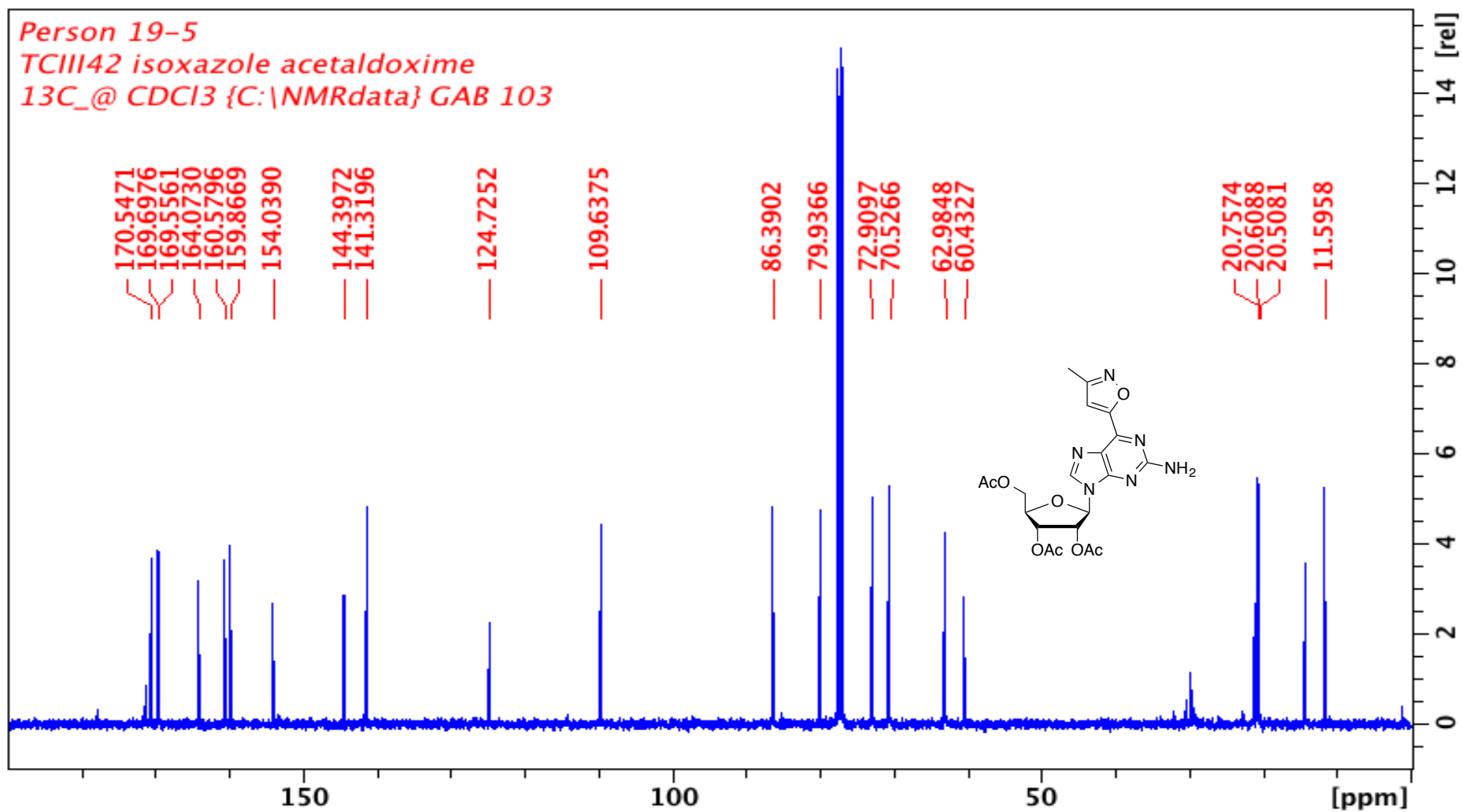
Appendix 54. ¹H-NMR spectrum of 153. 400 MHz, MeOD. Peak at 3.3 ppm comes from MeOH. Peak at 4.8 comes from water.



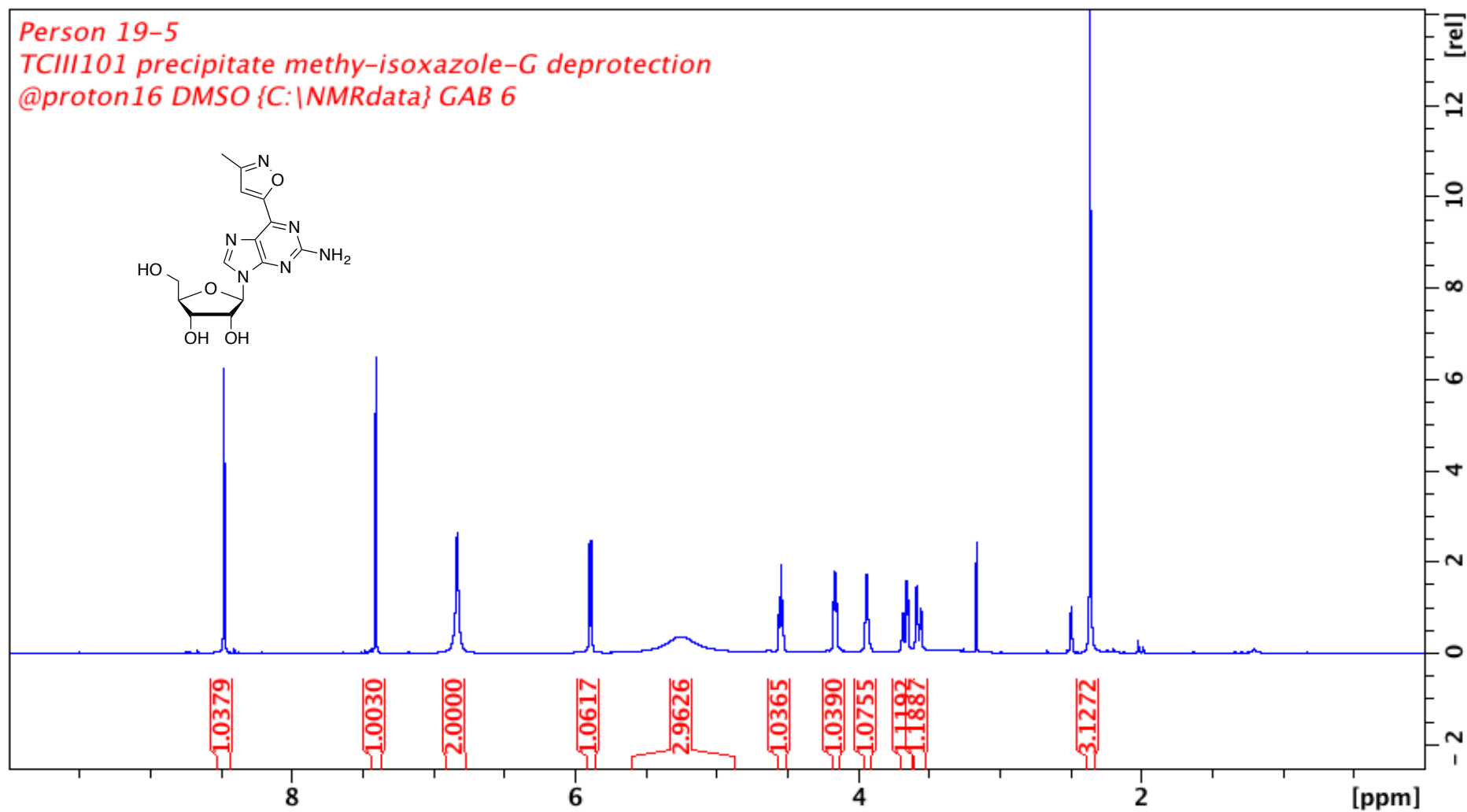
Appendix 55. ^{13}C -NMR spectrum of **153**. 100 MHz, MeOD. Peak at 49 ppm comes from MeOH.



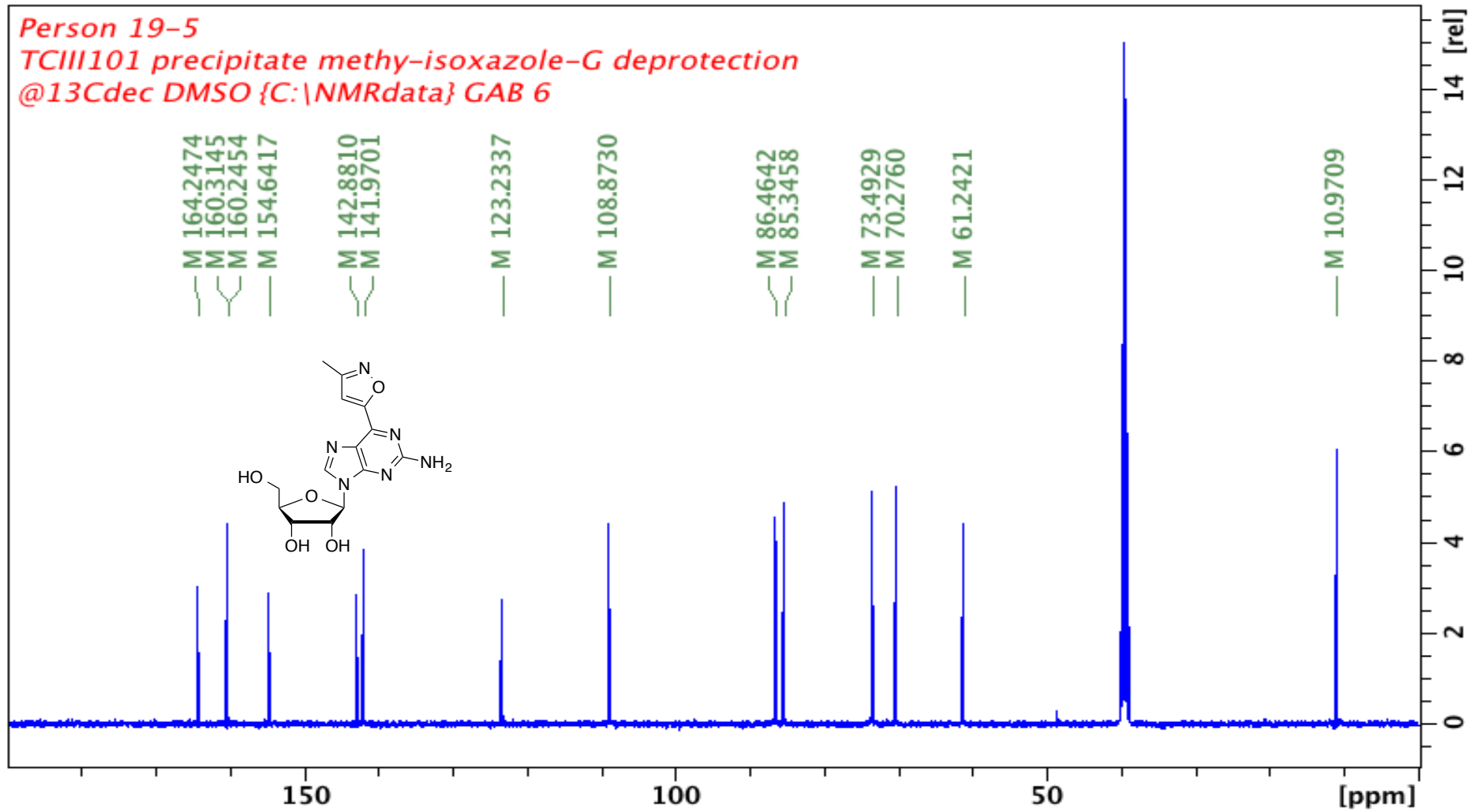
Appendix 56. ¹H-NMR spectrum of **158**. 400 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃. Peaks at 1.26, 2.1 and 4.1 ppm come from EtOAc.



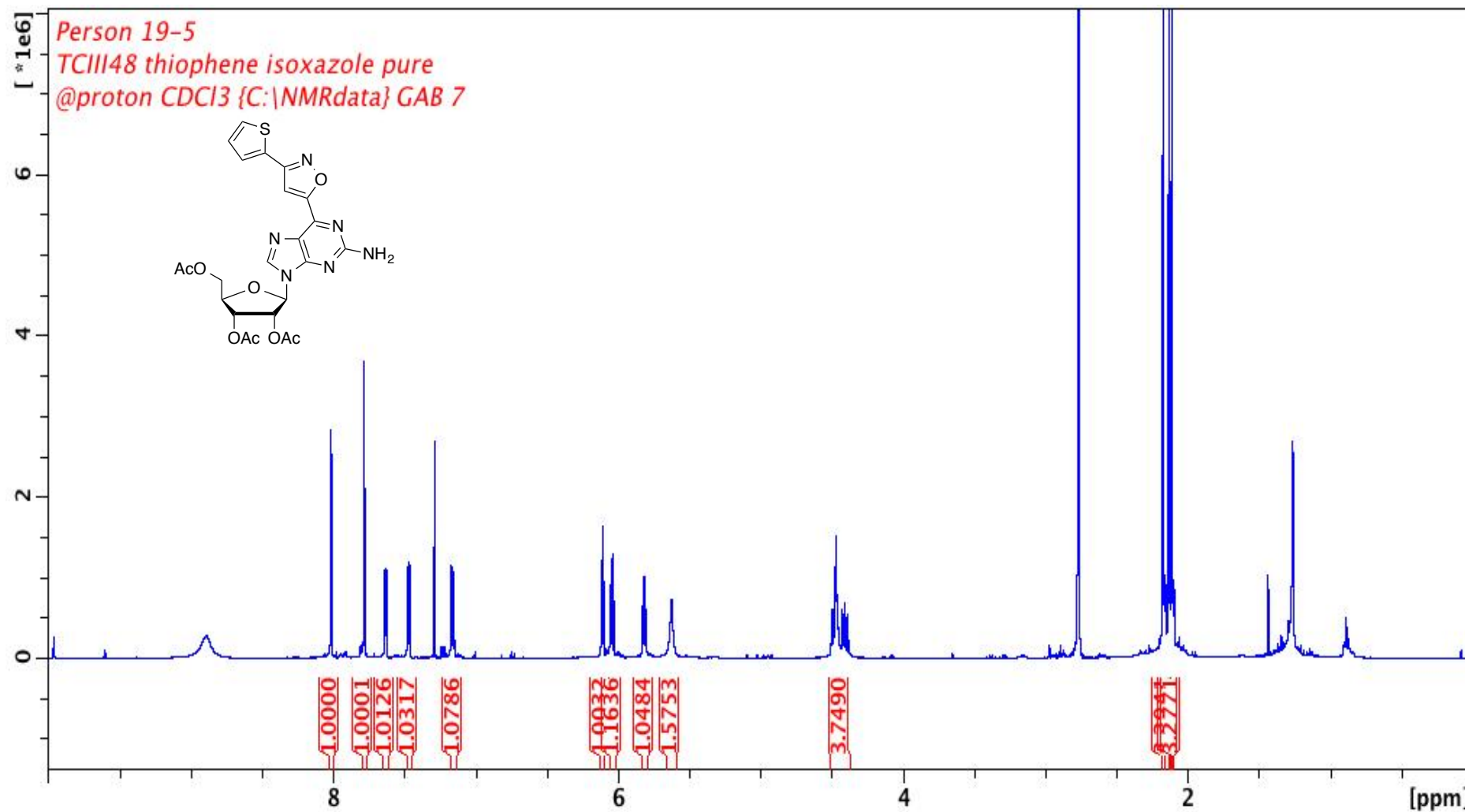
Appendix 57. ^{13}C -NMR spectrum of **158**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peaks at 14.2, 21, 61 and 171 ppm come from EtOAc.



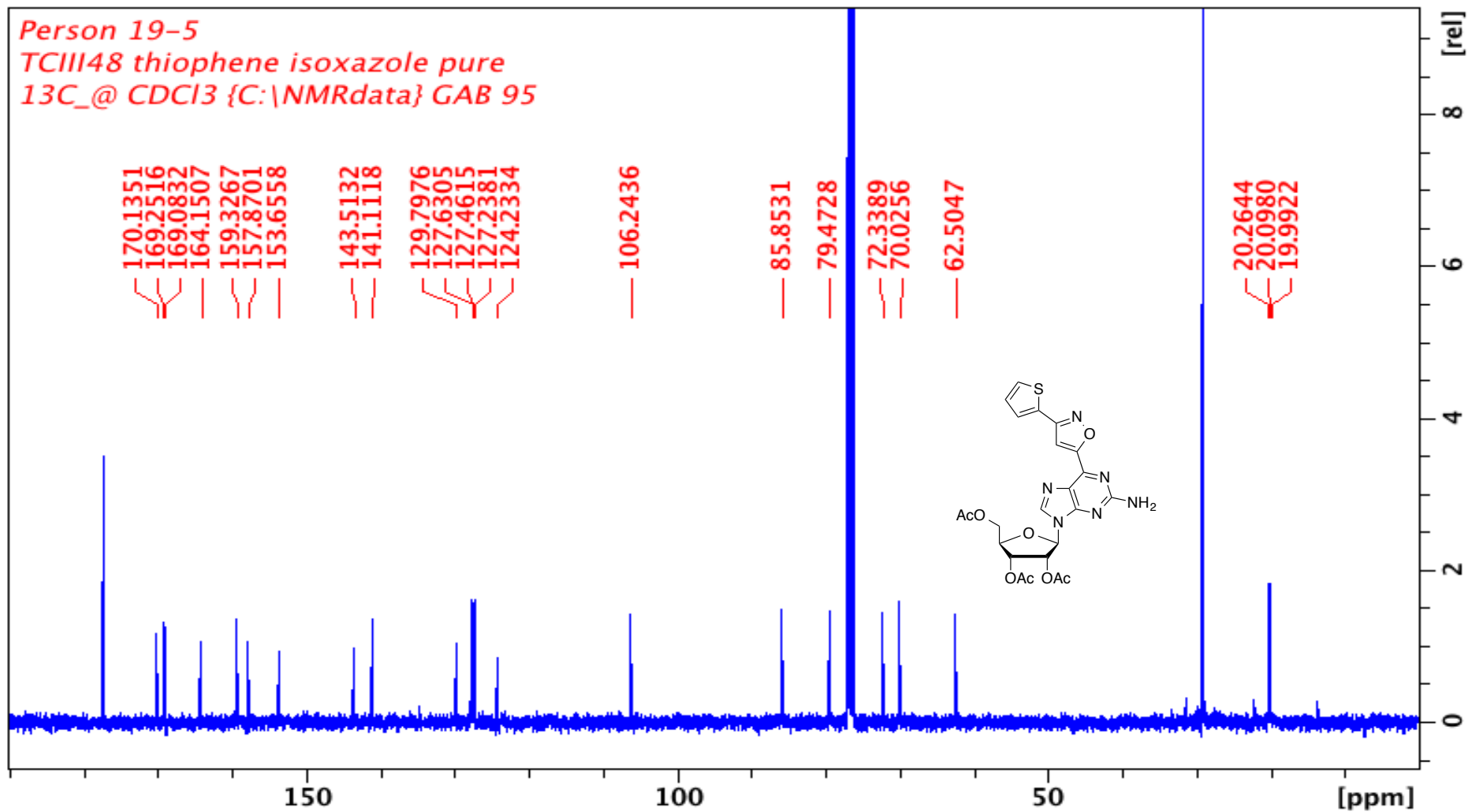
Appendix 58. $^1\text{H-NMR}$ spectrum of **154**. 400 MHz, DMSO- d_6 . Peak at 2.5 ppm comes from DMSO. Peak at 3.3 ppm comes from water.



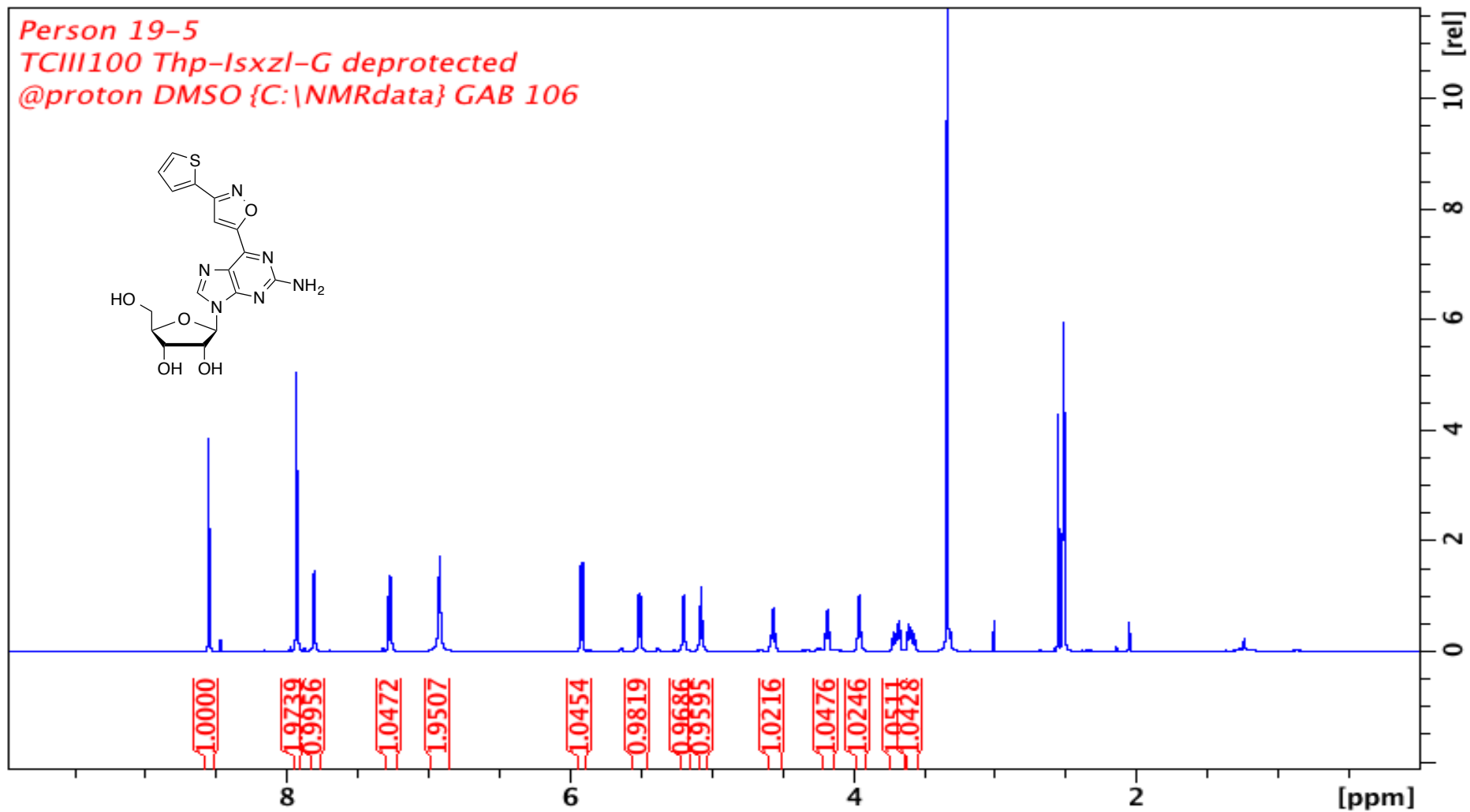
Appendix 59. ^{13}C -NMR spectrum of **154**. 100 MHz, DMSO- d_6 . Peak at 49 ppm comes from DMSO.



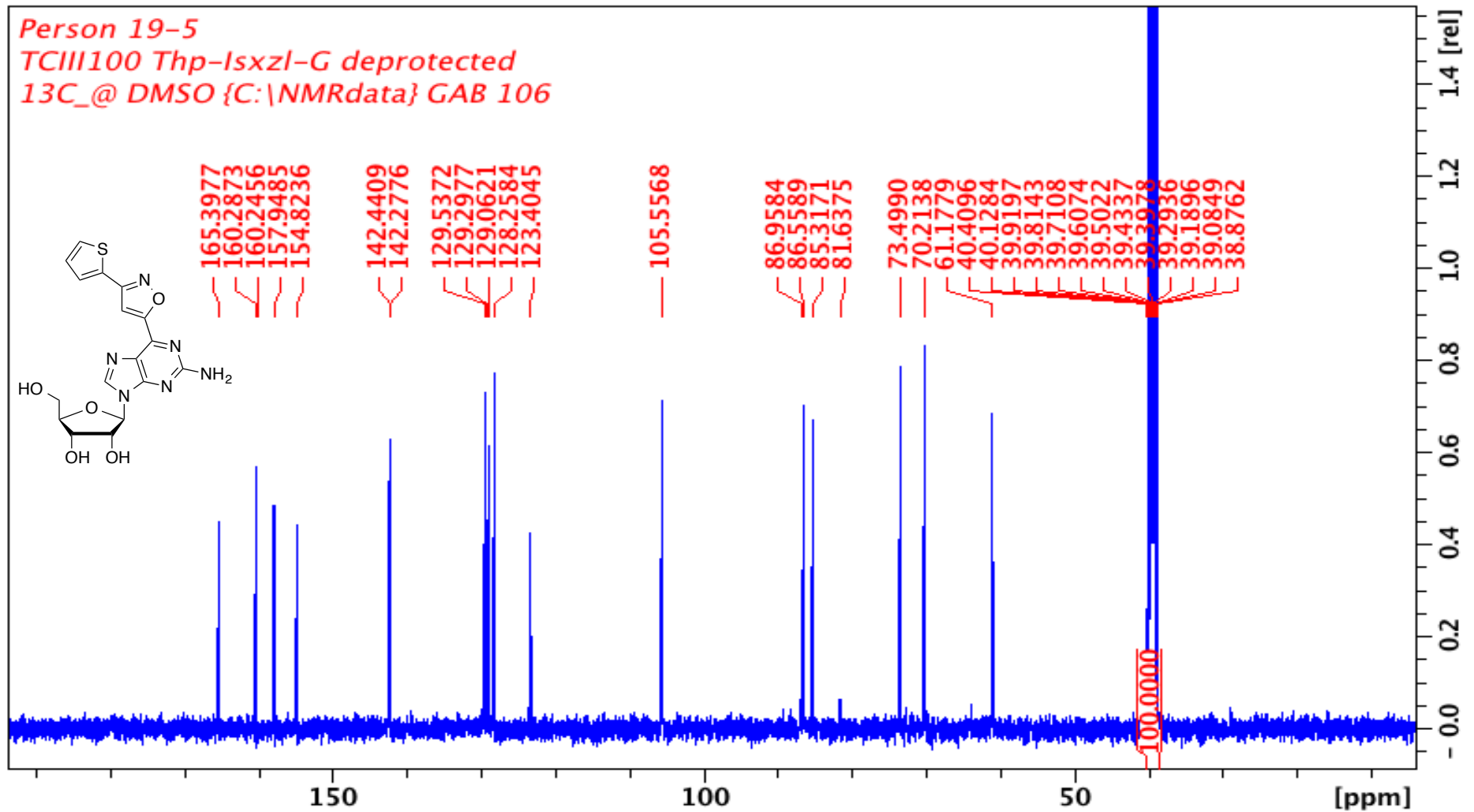
Appendix 60. $^1\text{H-NMR}$ spectrum of **159**. 400 MHz, CDCl_3 . Peak at 7.26 ppm comes from chloroform. Peak at 1.2 ppm comes from grease. Peak at 2.7 ppm comes from succinimide.



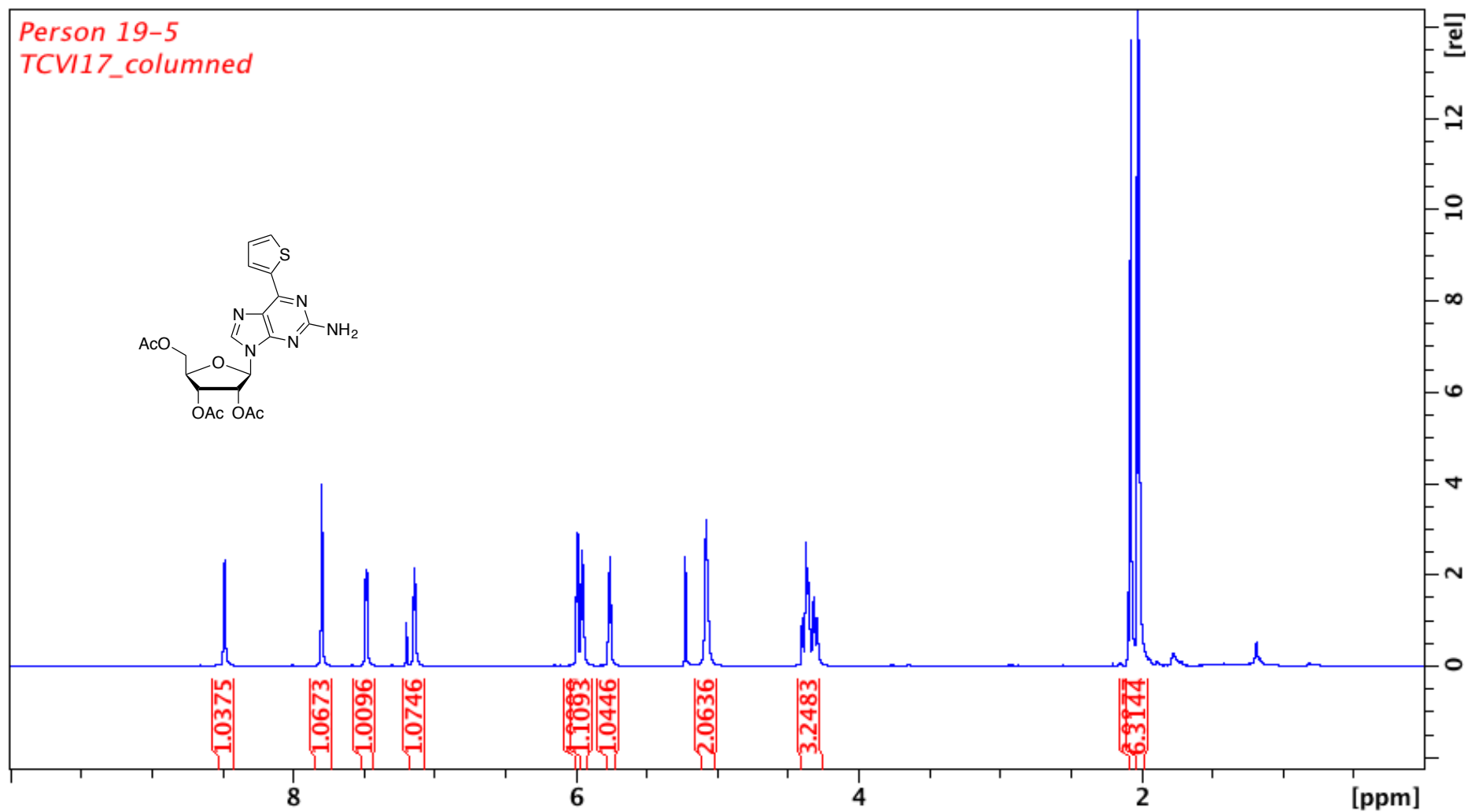
Appendix 61. ^{13}C -NMR spectrum of 159. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peaks at 29 and 177 ppm come from succinimide.



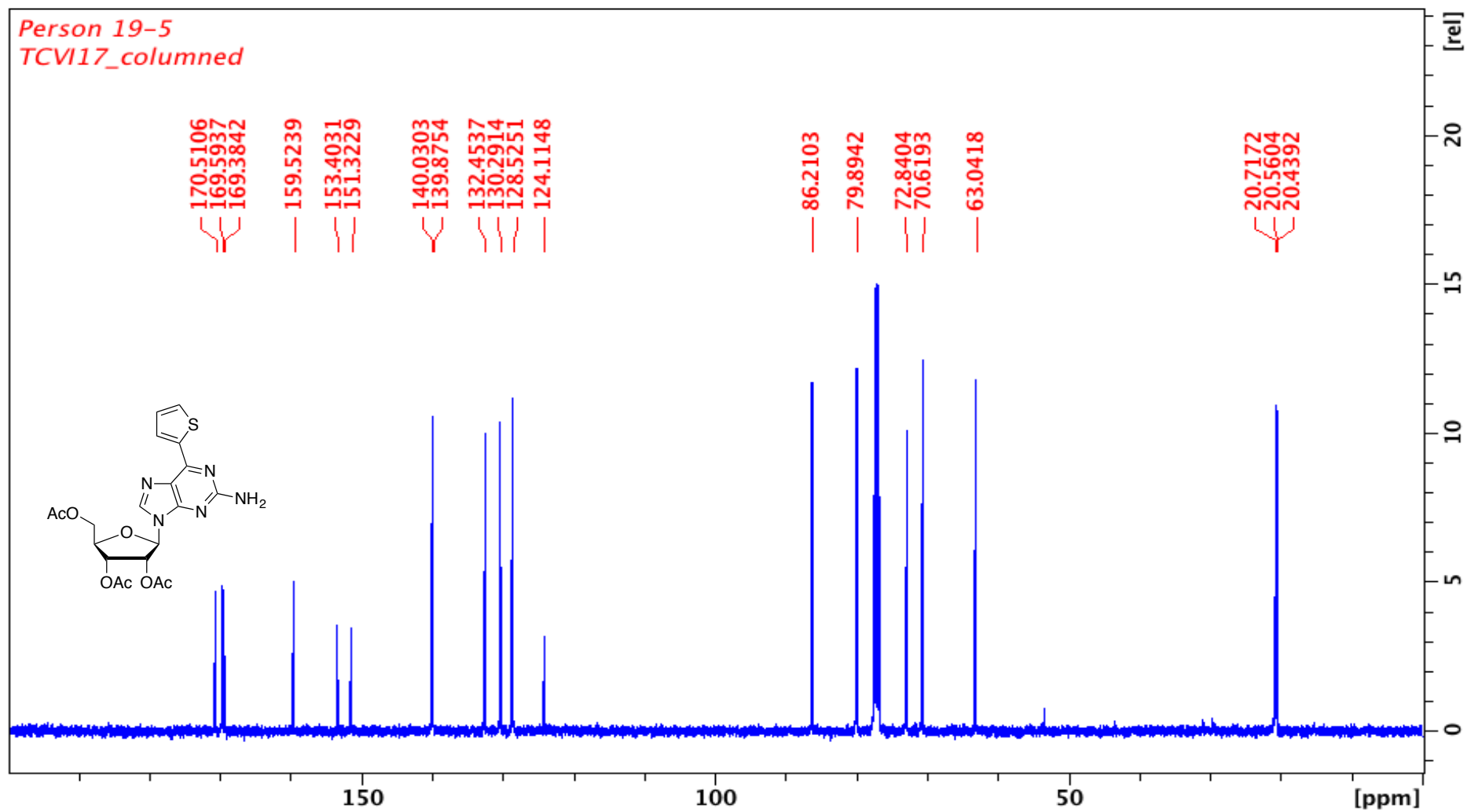
Appendix 62. $^1\text{H-NMR}$ spectrum of 155. 400 MHz, DMSO- d_6 . Peak at 3.3 ppm comes from water. Peak at 2.5 ppm comes from DMSO.



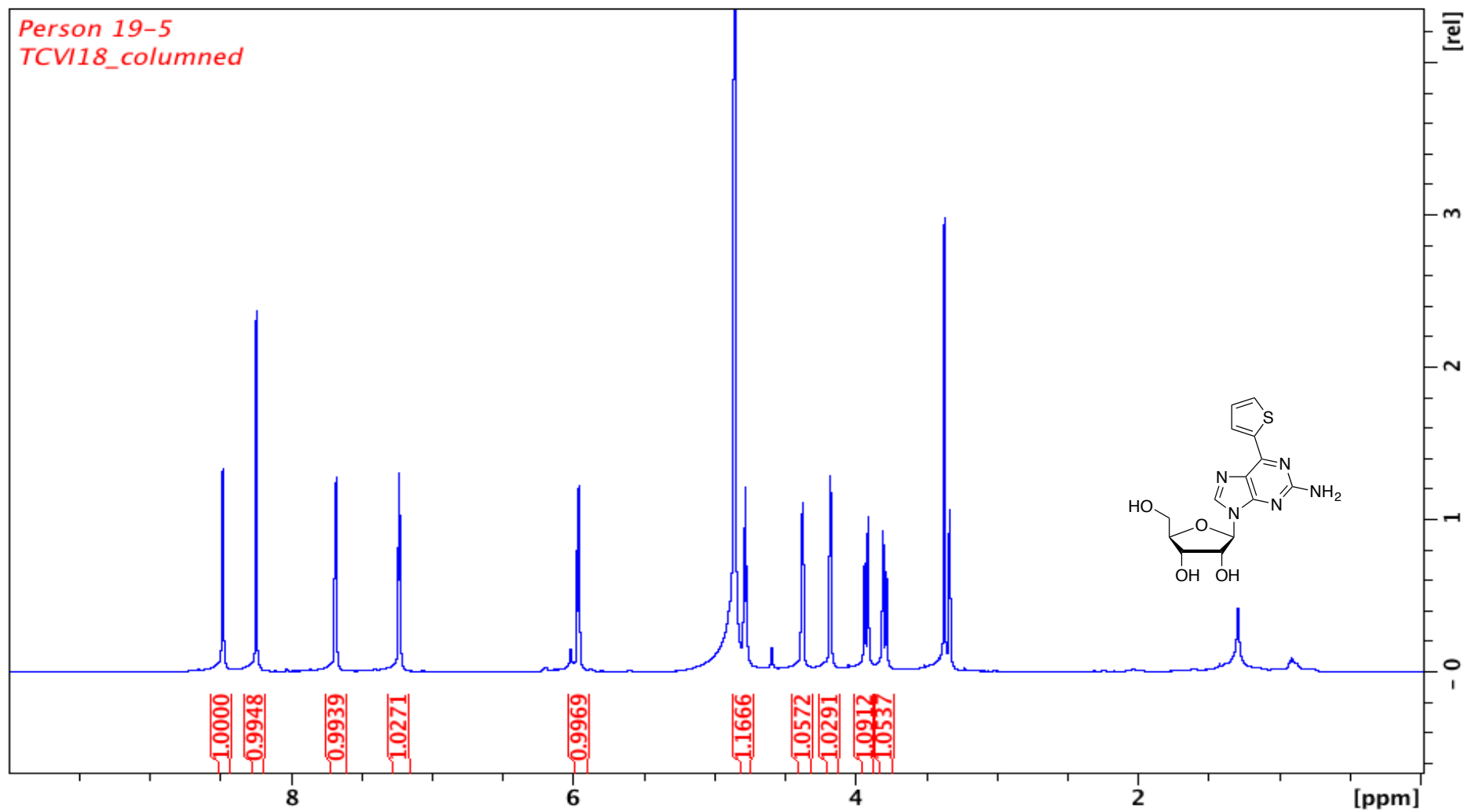
Appendix 63. ¹³C-NMR spectrum of 155. 100 MHz, DMSO-d₆. Peak at 39 ppm comes from DMSO.



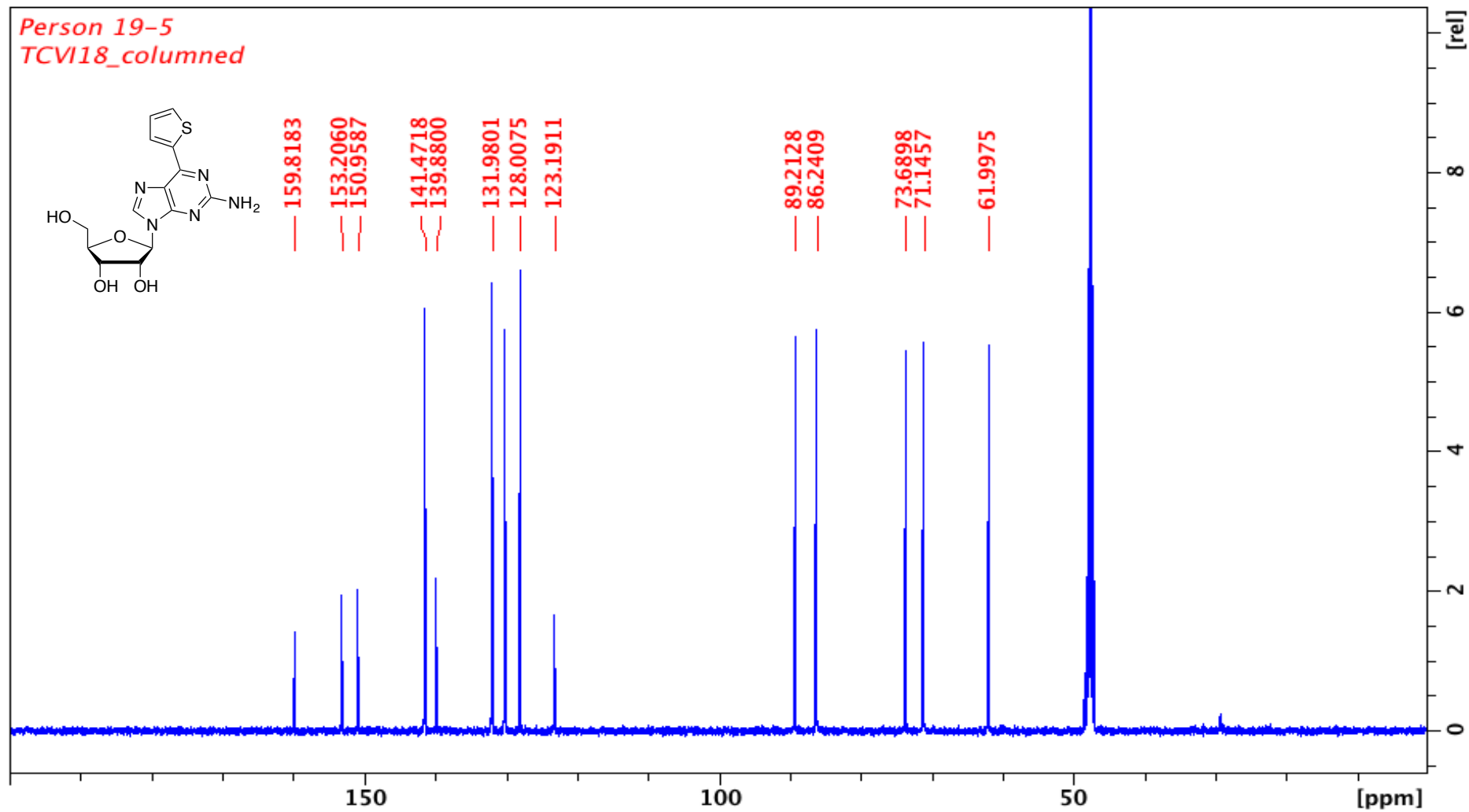
Appendix 64. $^1\text{H-NMR}$ spectrum of **175**. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peak at 5.25 ppm comes from DCM.



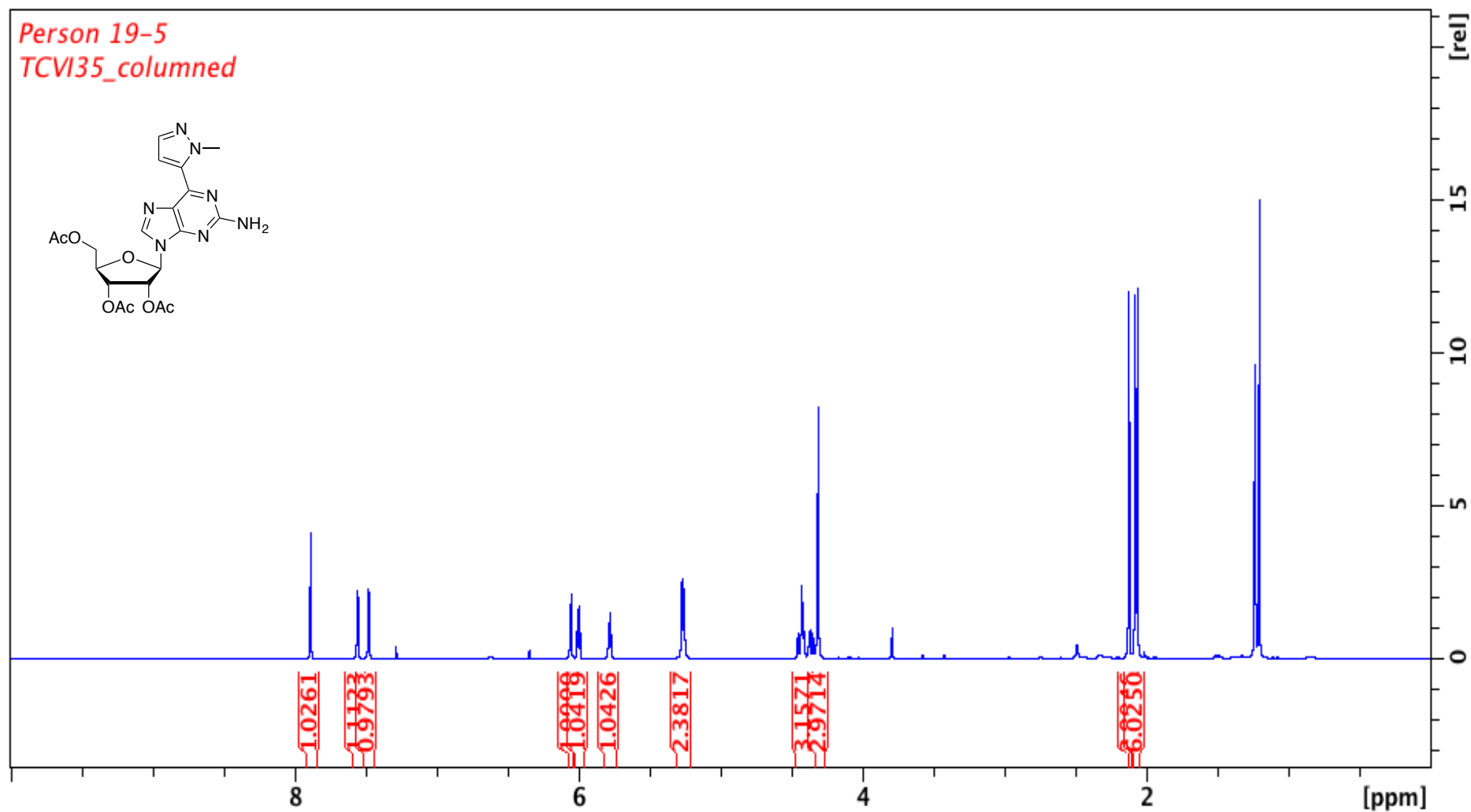
Appendix 65. ^{13}C -NMR spectrum of 175. 125 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .



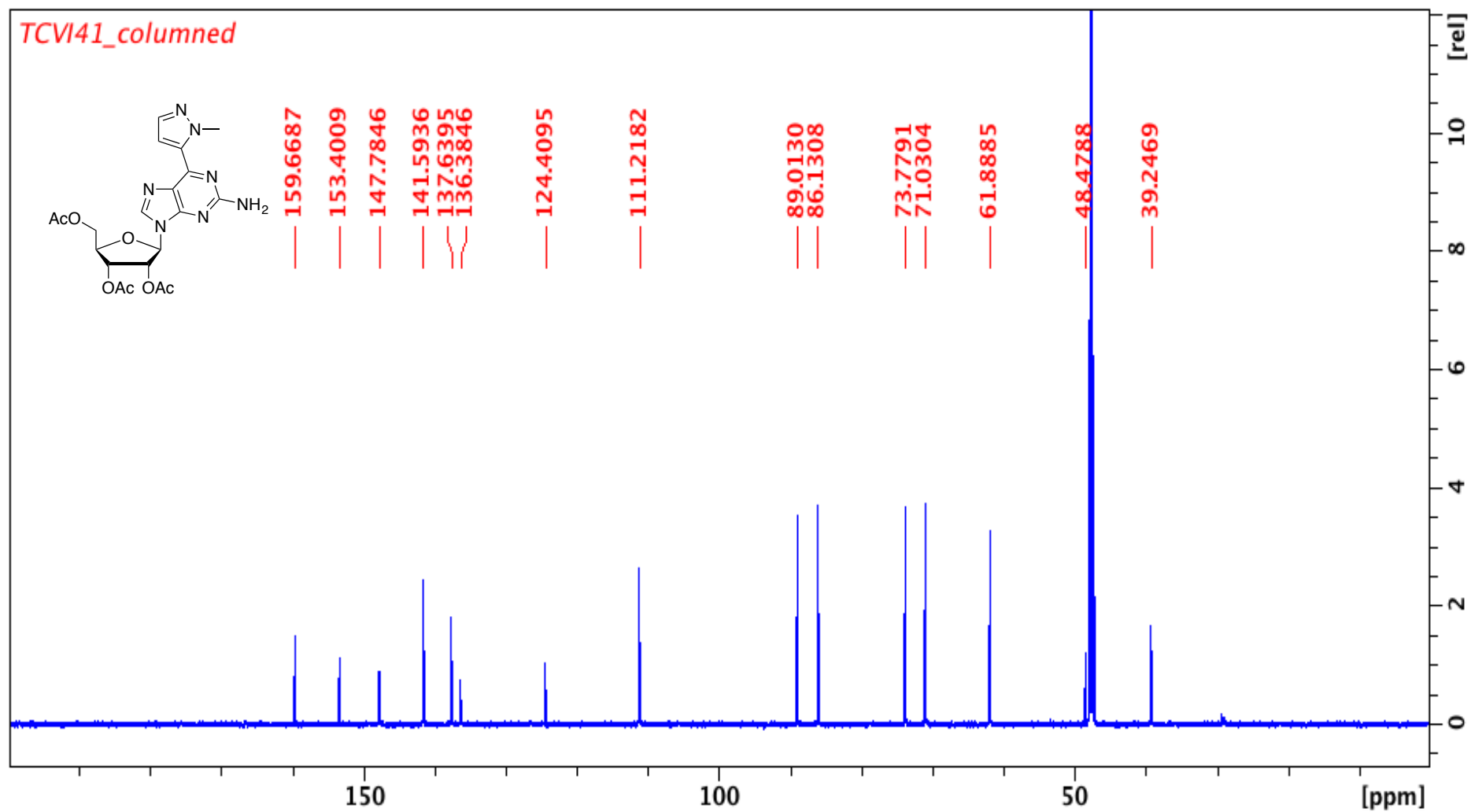
Appendix 66. ¹H-NMR spectrum of 176. 500 MHz, MeOD. Peak at 3.8 ppm comes from water. Peaks at 3.3 ppm come from MeOH.



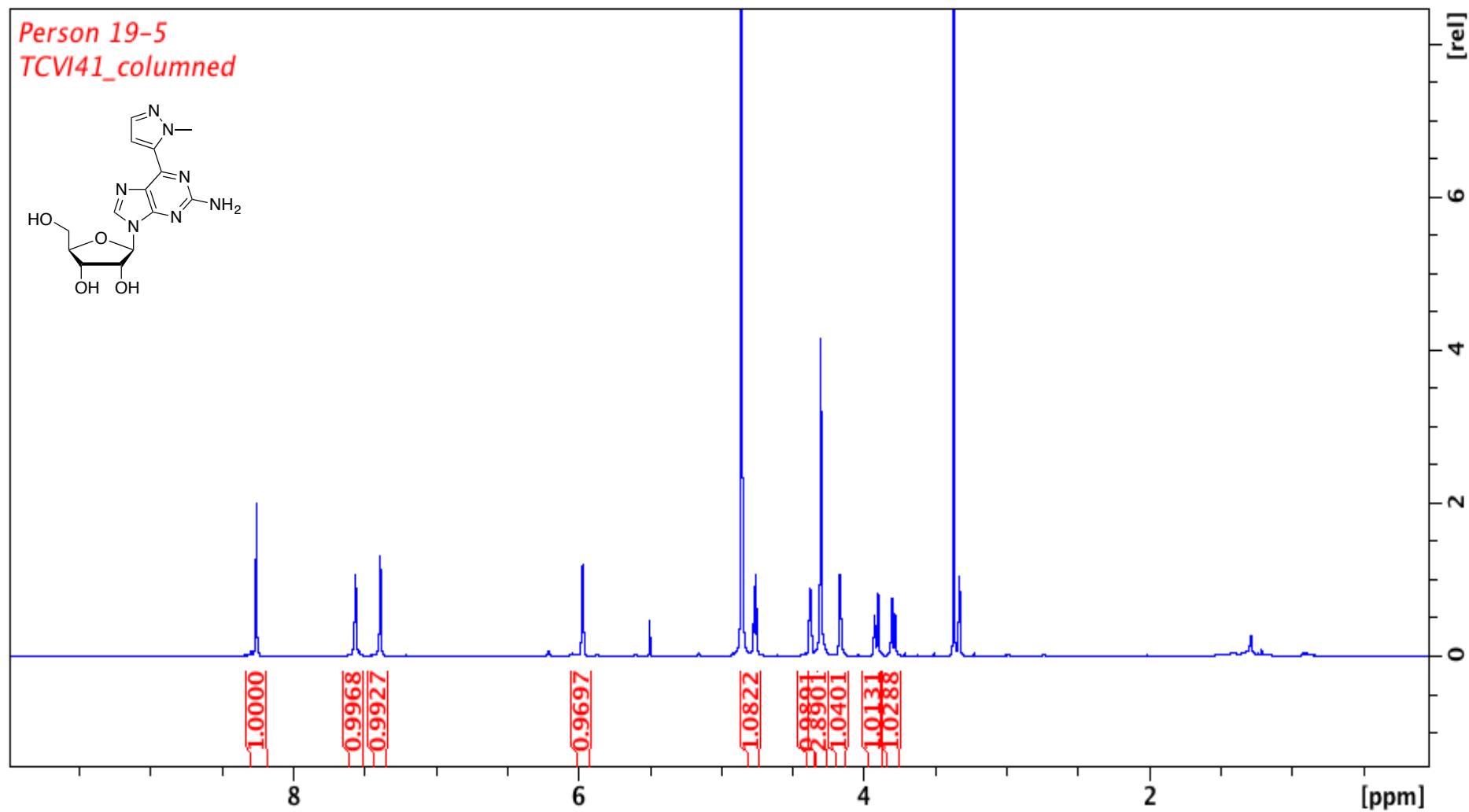
Appendix 67. ^{13}C -NMR spectrum of 176. 125 MHz, CDCl_3 . Peak at 49 ppm comes from MeOH.



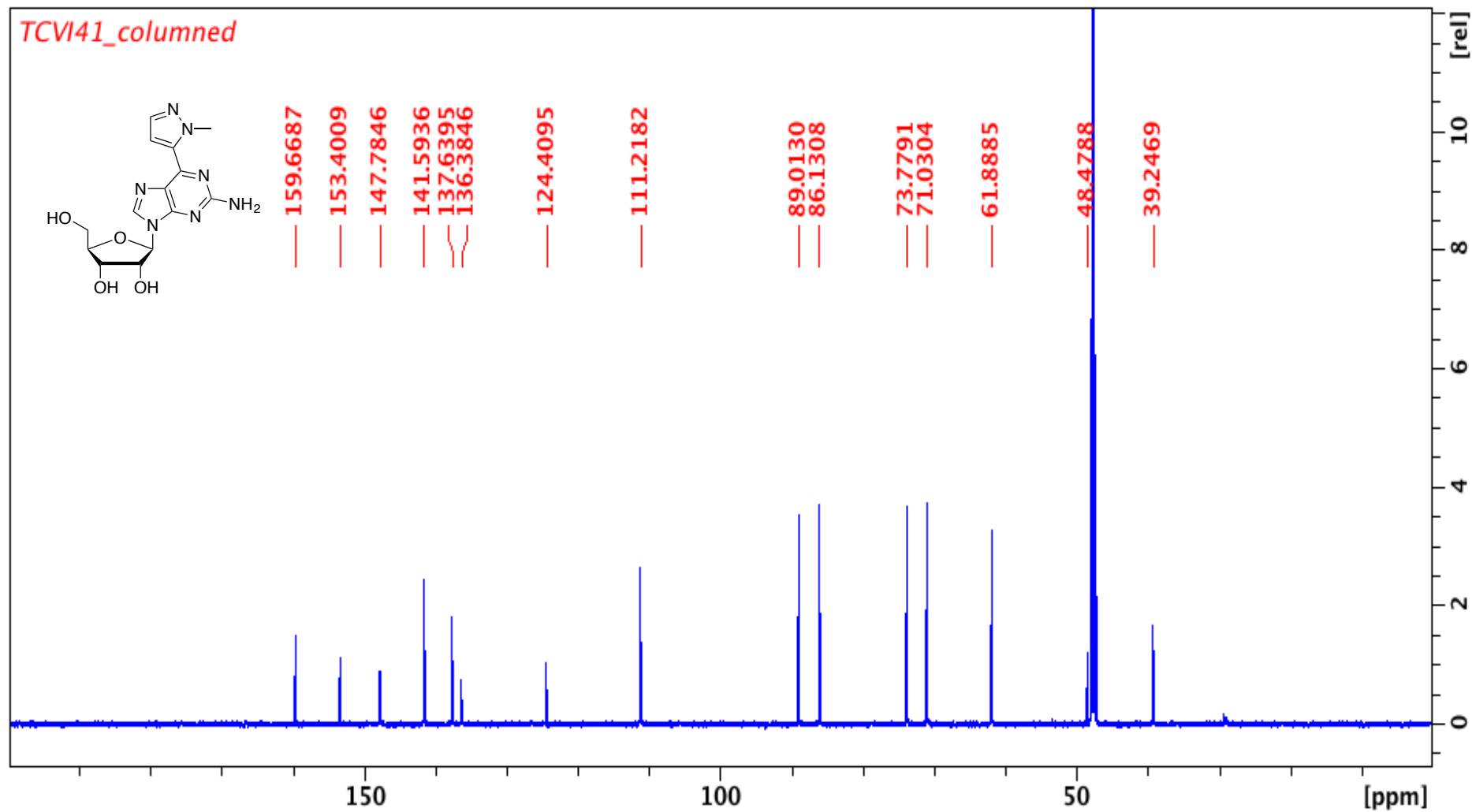
Appendix 68. $^1\text{H-NMR}$ spectrum of 172. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peaks at 1.15 and 1.17 ppm come from a pinacol related impurity.



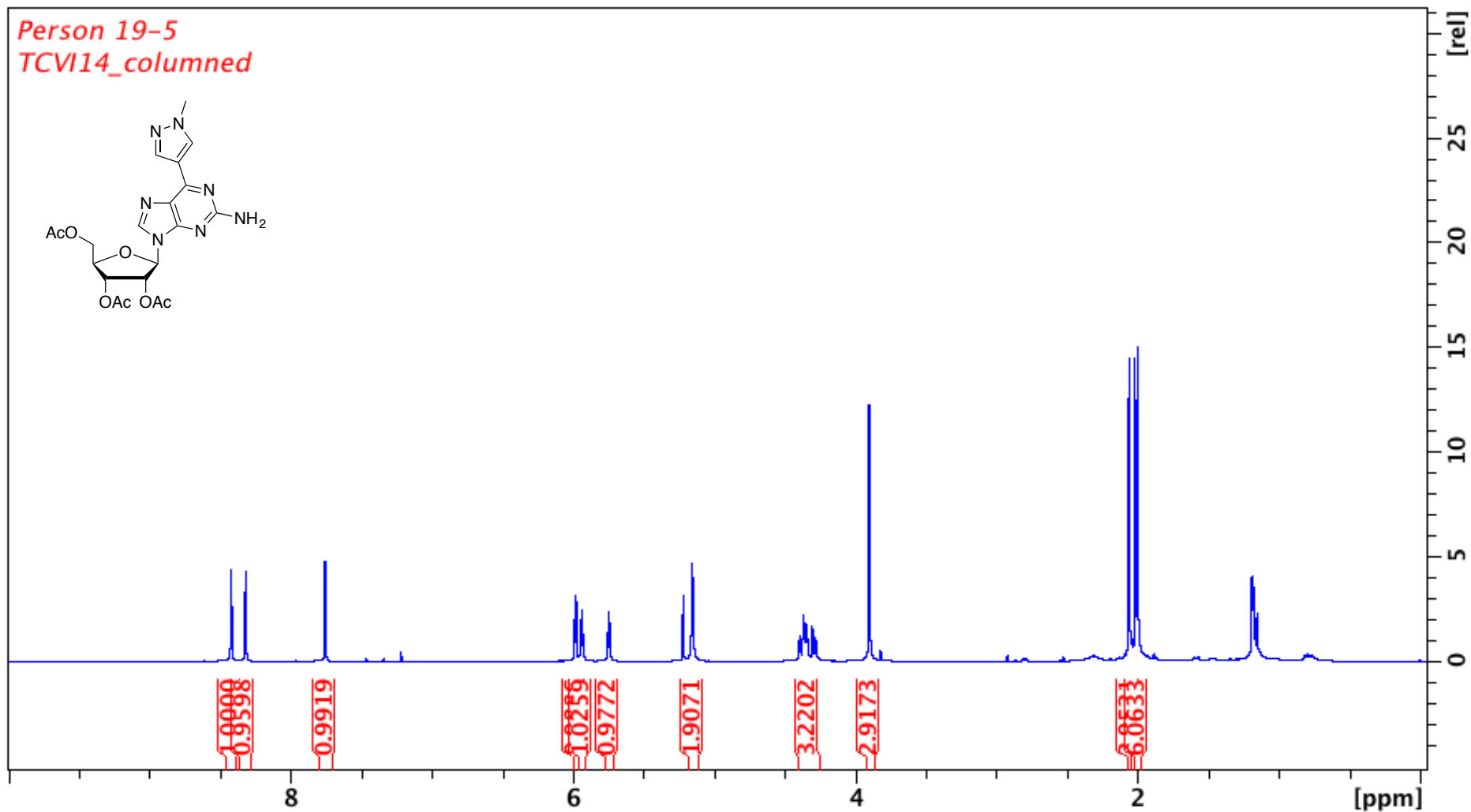
Appendix 69. ¹³C-NMR spectrum of 172. 100 MHz, CDCl₃. Peak at 78 ppm comes from CHCl₃.



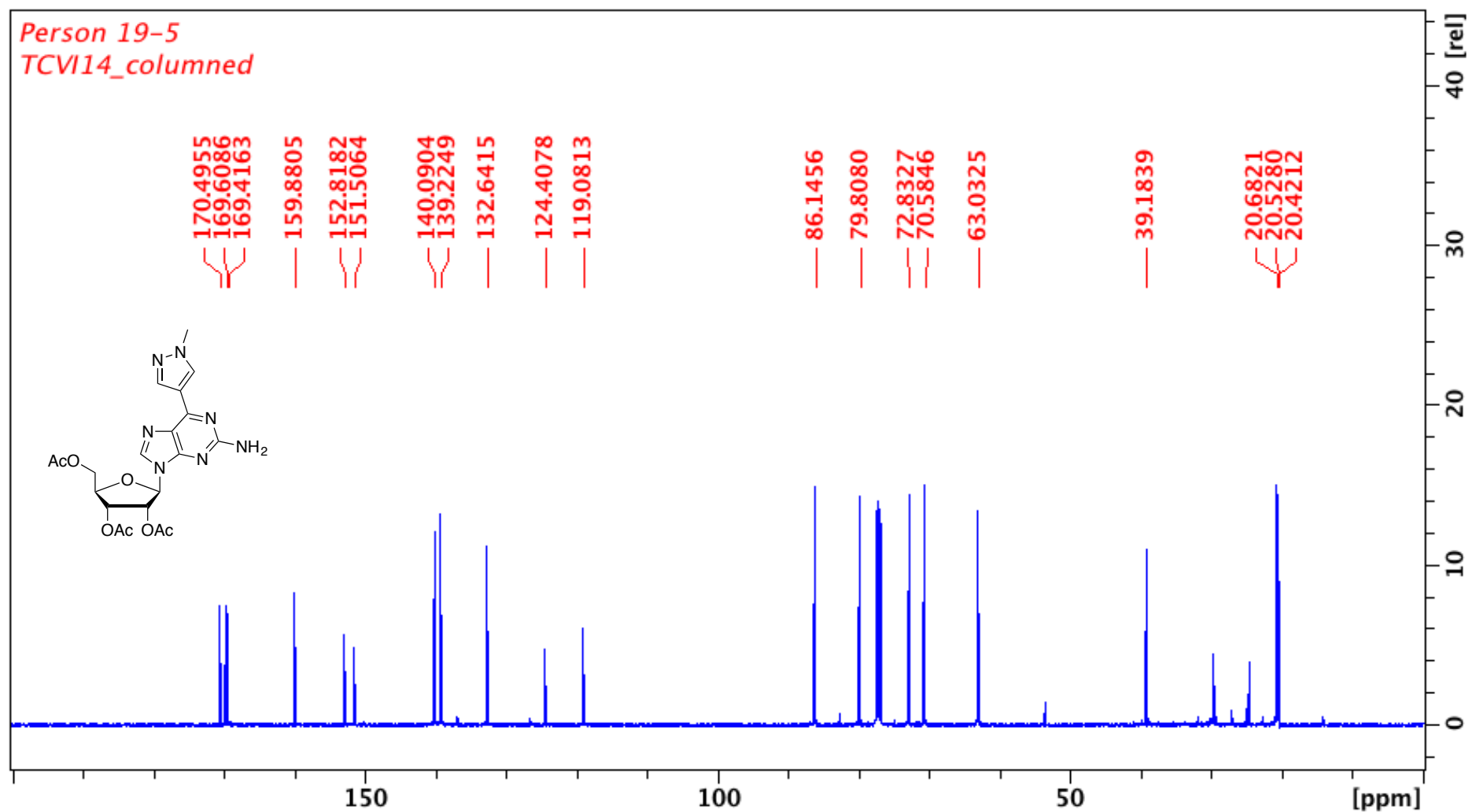
Appendix 70. ^1H -NMR spectrum of **169**. 500 MHz, MeOD. Peaks at 3.3 ppm come from MeOH. Peak at 4.8 ppm comes from water.



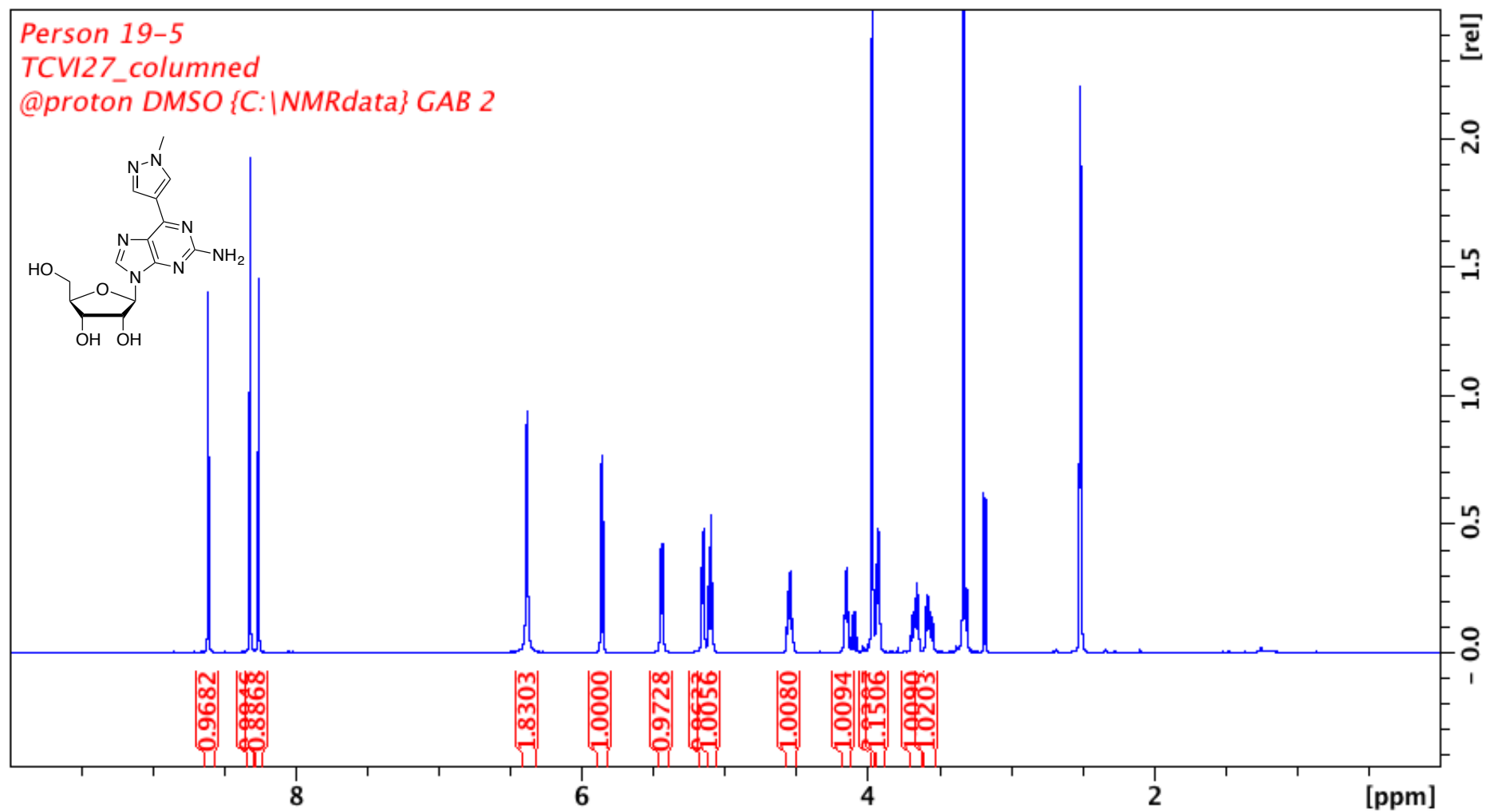
Appendix 71. ^{13}C -NMR spectrum of **169**. 150 MHz, MeOD. Peak at 48 ppm comes from MeOH.



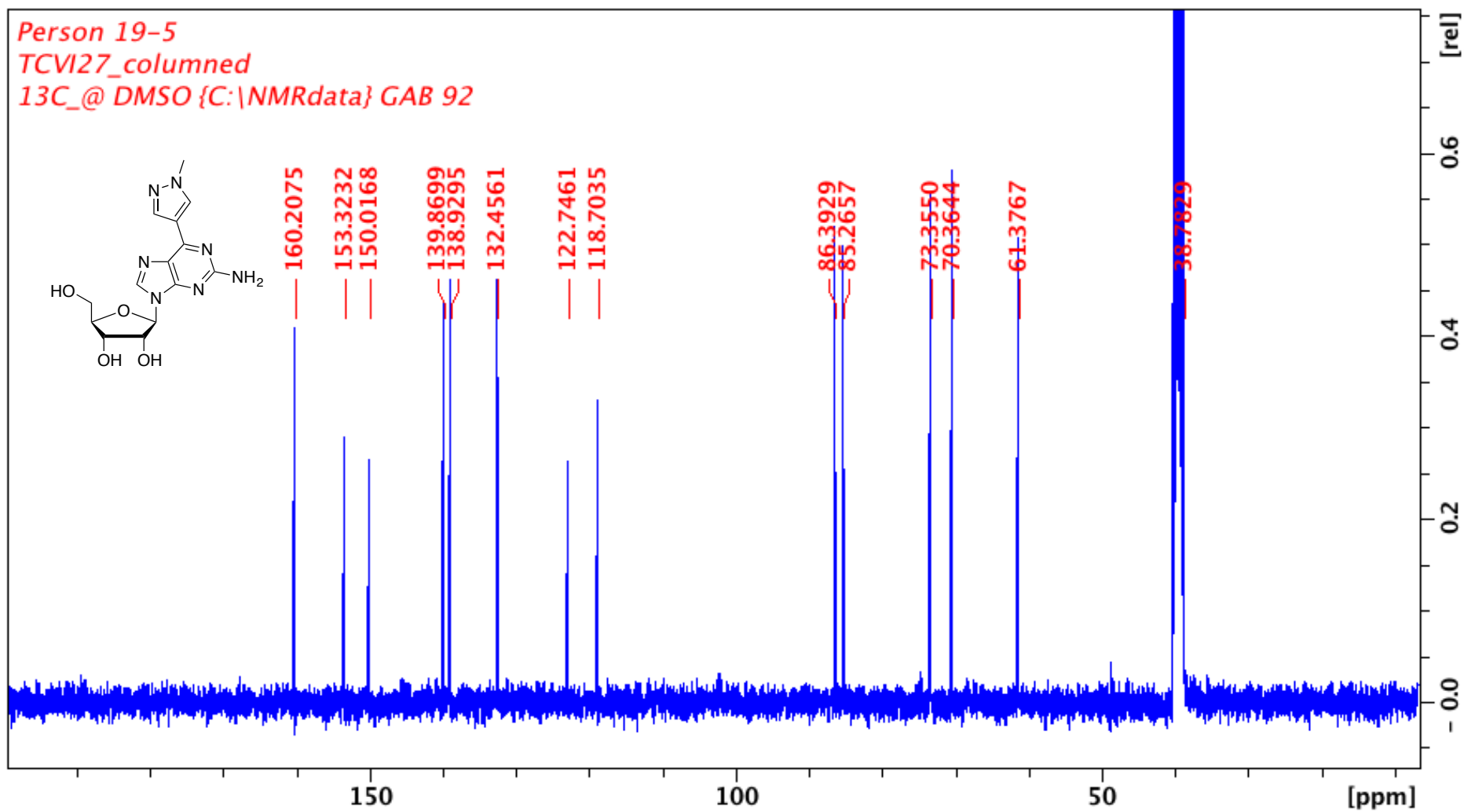
Appendix 72. ¹H-NMR spectrum of 177. 500 MHz, CDCl₃. Peak at 7.26 ppm comes from CHCl₃. Peaks at 1.15 and 1.17 ppm come from a pinacol related impurity.



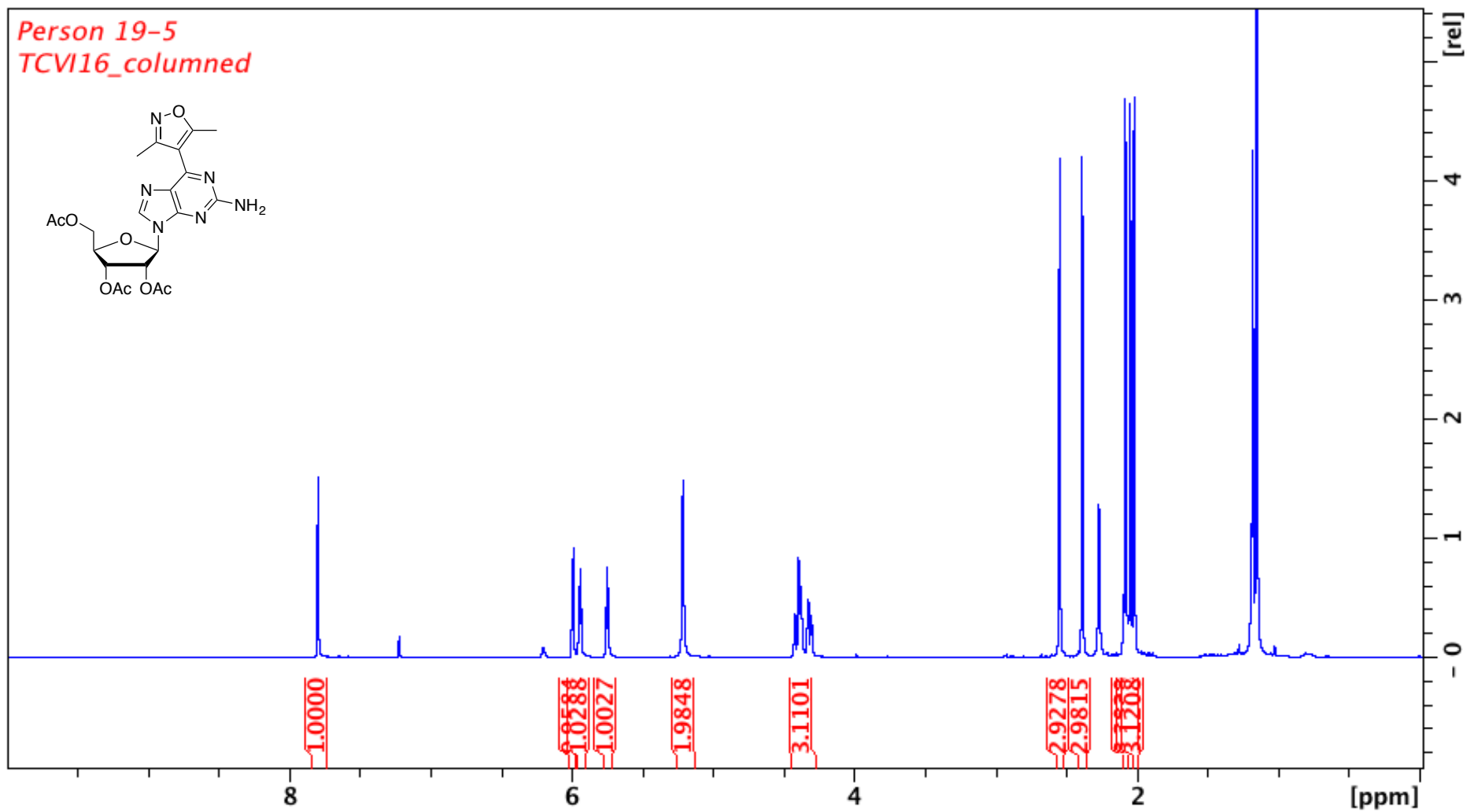
Appendix 73. ^{13}C -NMR spectrum of 177. 125 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 53 ppm comes from DCM. Peaks at 25 and 29 ppm come from a pinacol related impurity.



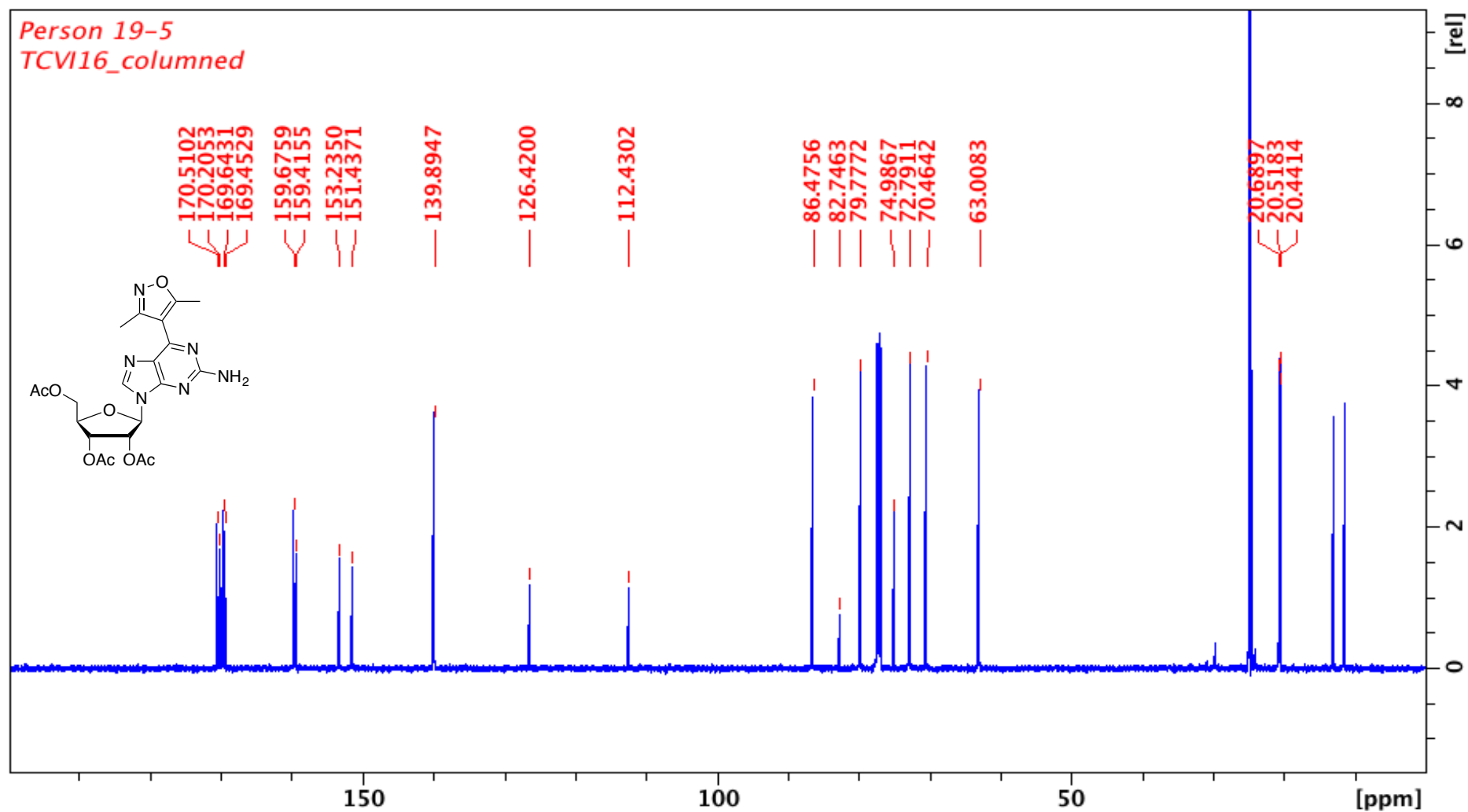
Appendix 74. $^1\text{H-NMR}$ spectrum of **170**. 500 MHz, DMSO- d_6 . Peak at 2.5 ppm comes from DMSO. Peak at 3.1 ppm comes from water. Peak at 3.1 and 4.1 ppm comes from MeOH



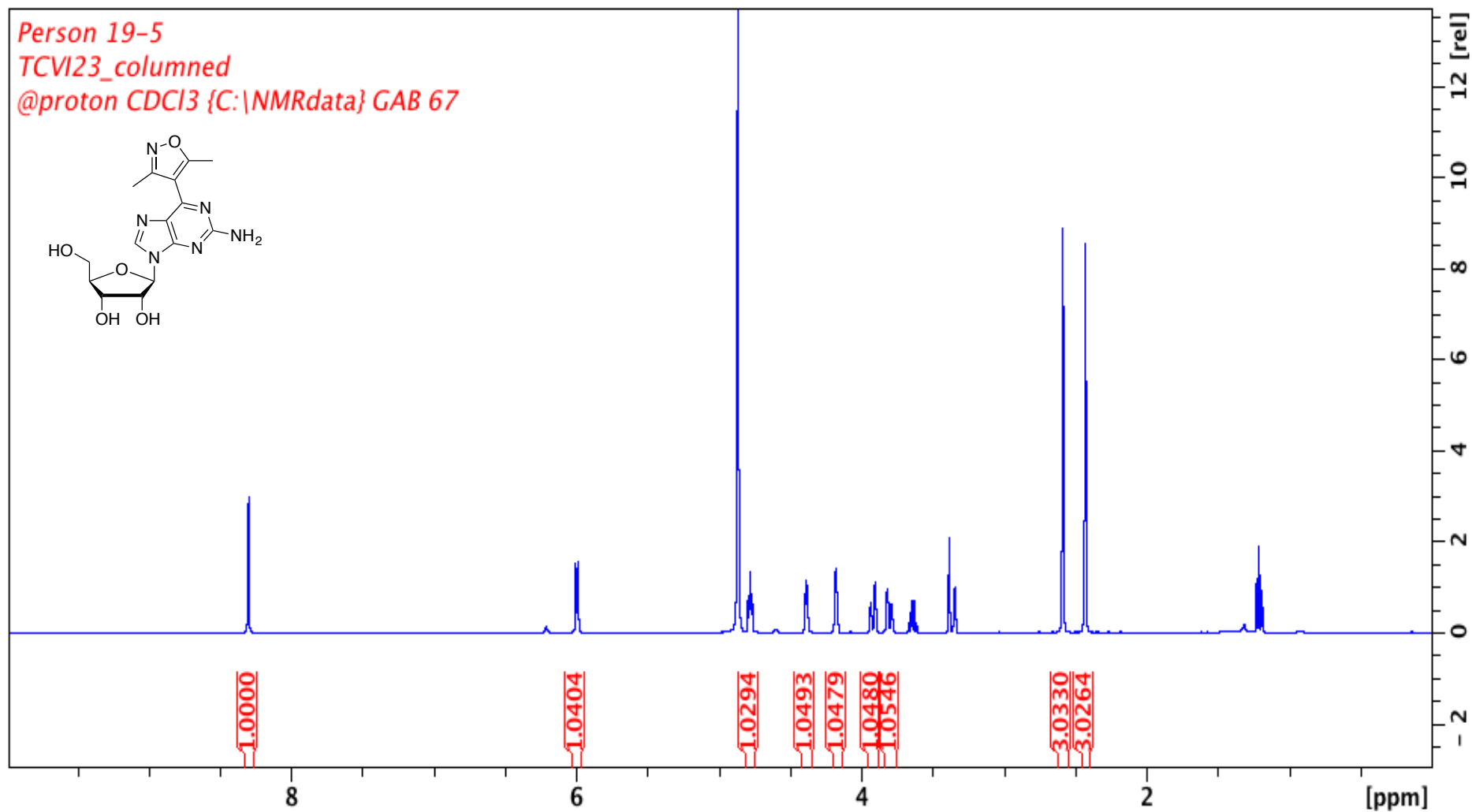
Appendix 75. ^{13}C -NMR spectrum of **170**. 100 MHz, DMSO- d_6 . Peak at 38 ppm comes from DMSO.



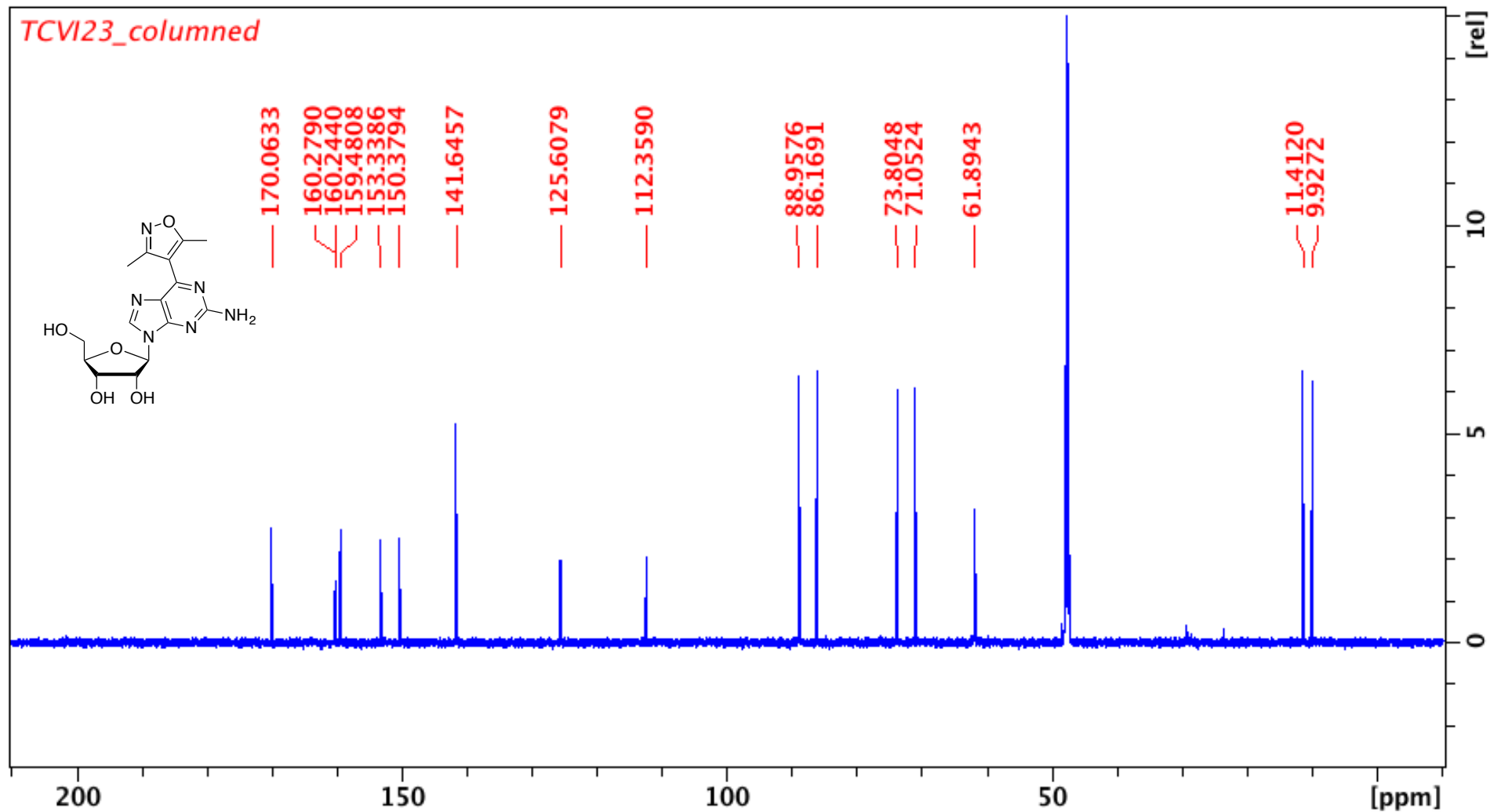
Appendix 76. ^1H -NMR spectrum of **178**. 500 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peak at 2.1 ppm comes from acetone. Peaks at 1.15 and 1.17 ppm come from a pinacol related impurity.



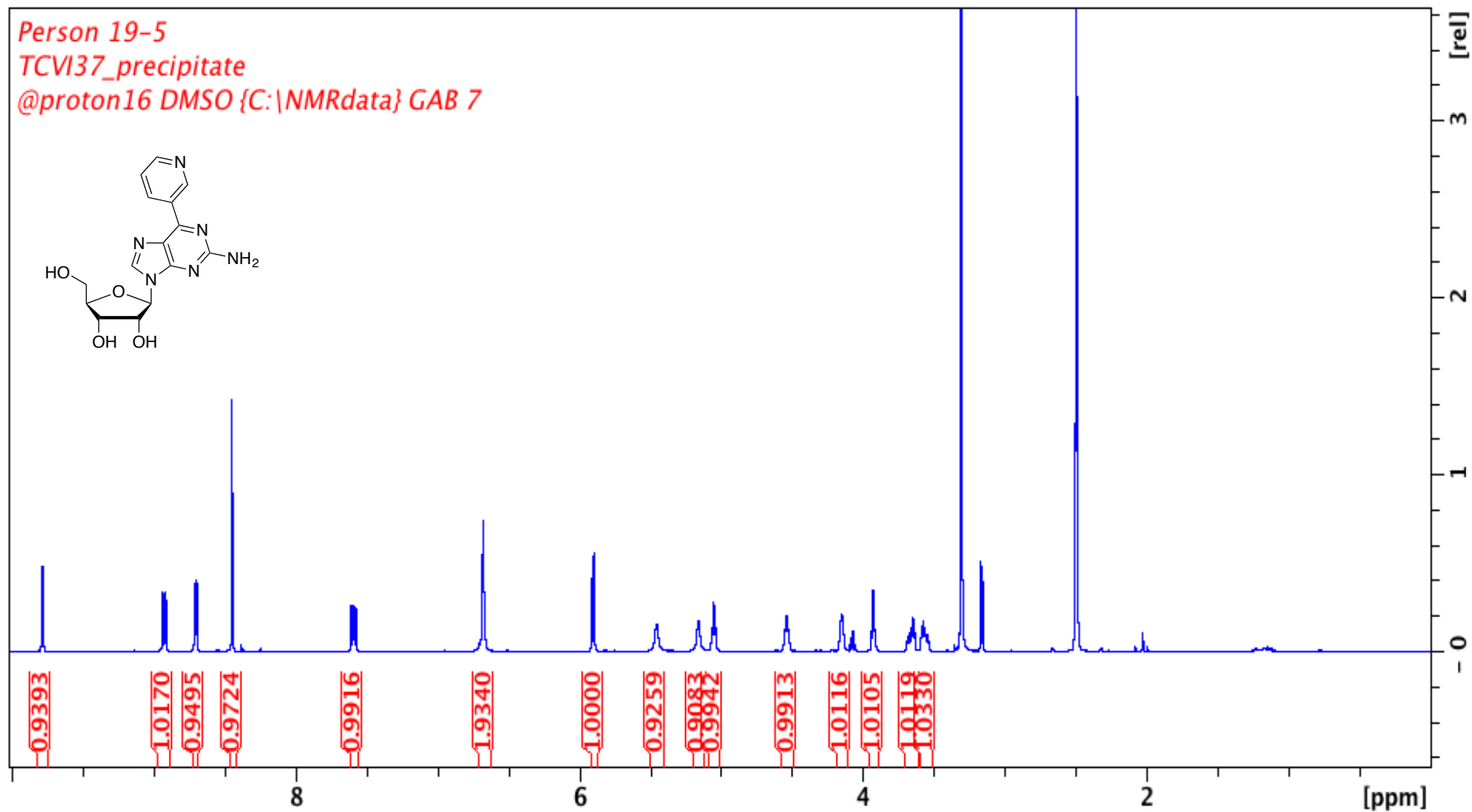
Appendix 77. ^{13}C -NMR spectrum of 178. 125 MHz, CDCl_3 . Peaks at 80 ppm comes from CHCl_3 . Peaks at 12 and 13 ppm come from a pinacol related impurity.



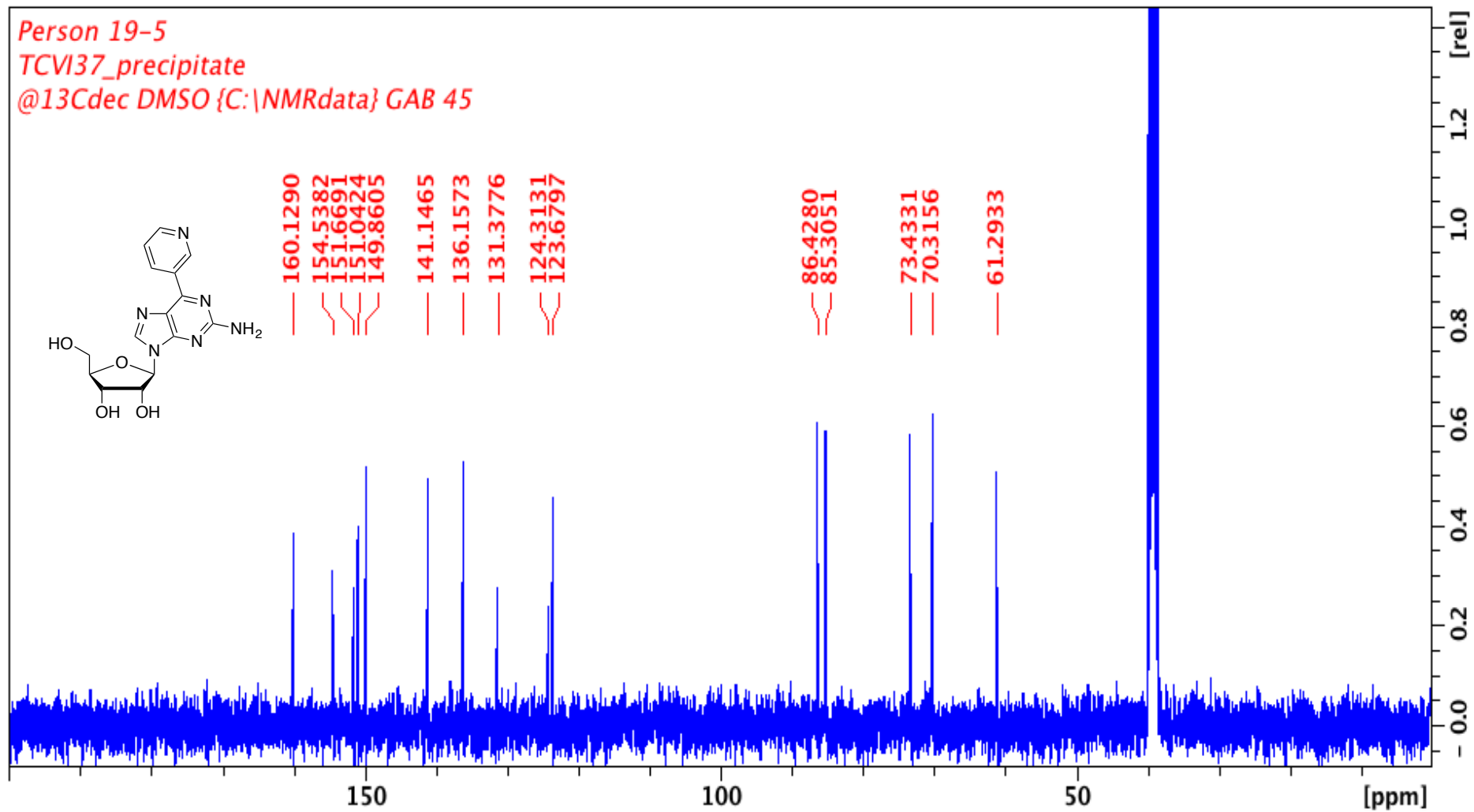
Appendix 78. $^1\text{H-NMR}$ spectrum of **183**. 500 MHz, MeOD. Peak at 4.8 ppm comes from water. Peak at 3.3 ppm come from MeOH. Peaks at 1.2 and 3.6 ppm come from ethanol used as stabilizer in CHCl_3 .



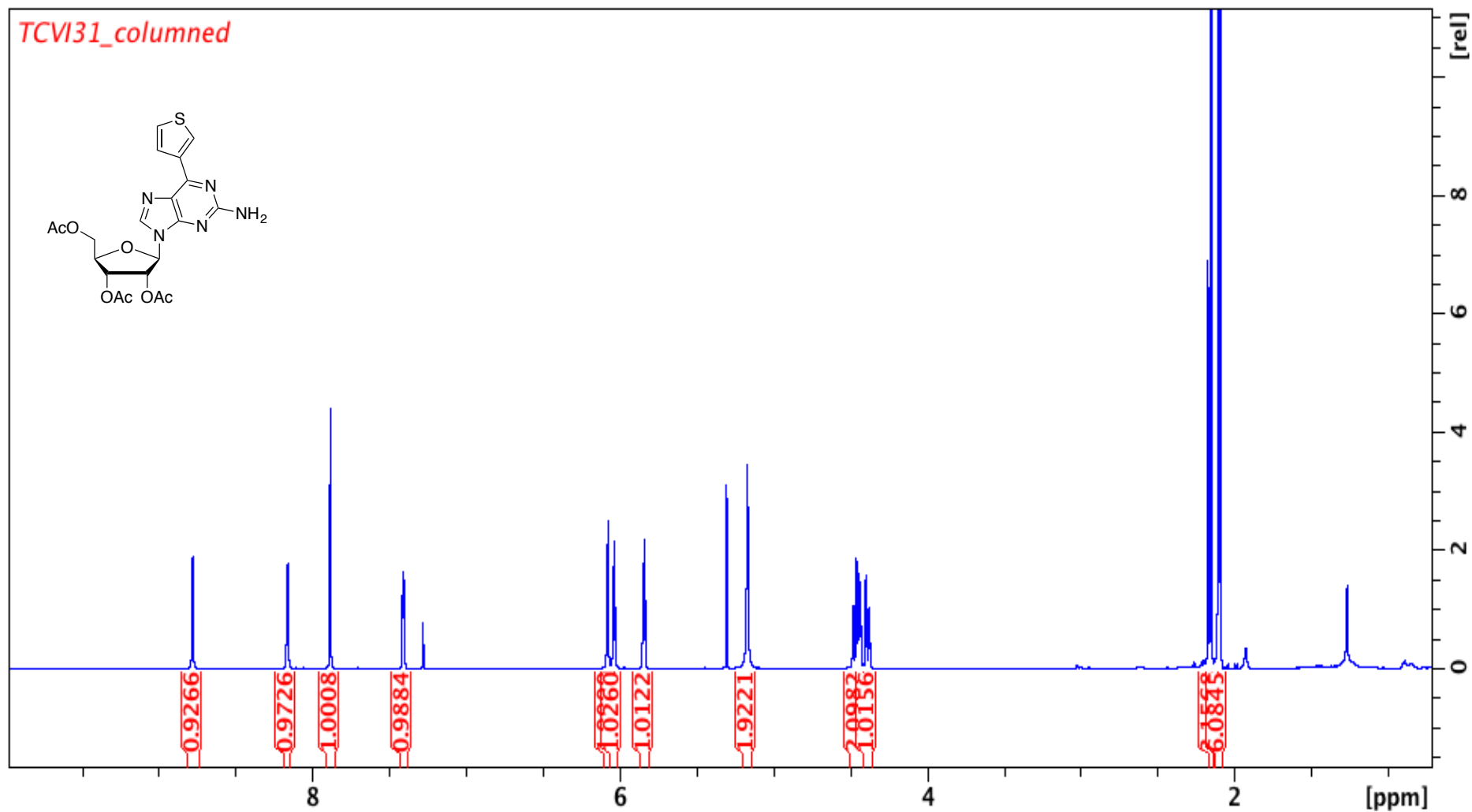
Appendix 79. ^{13}C -NMR spectrum of 183. 150 MHz, MeOD. Peak at 49 ppm come from MeOH.



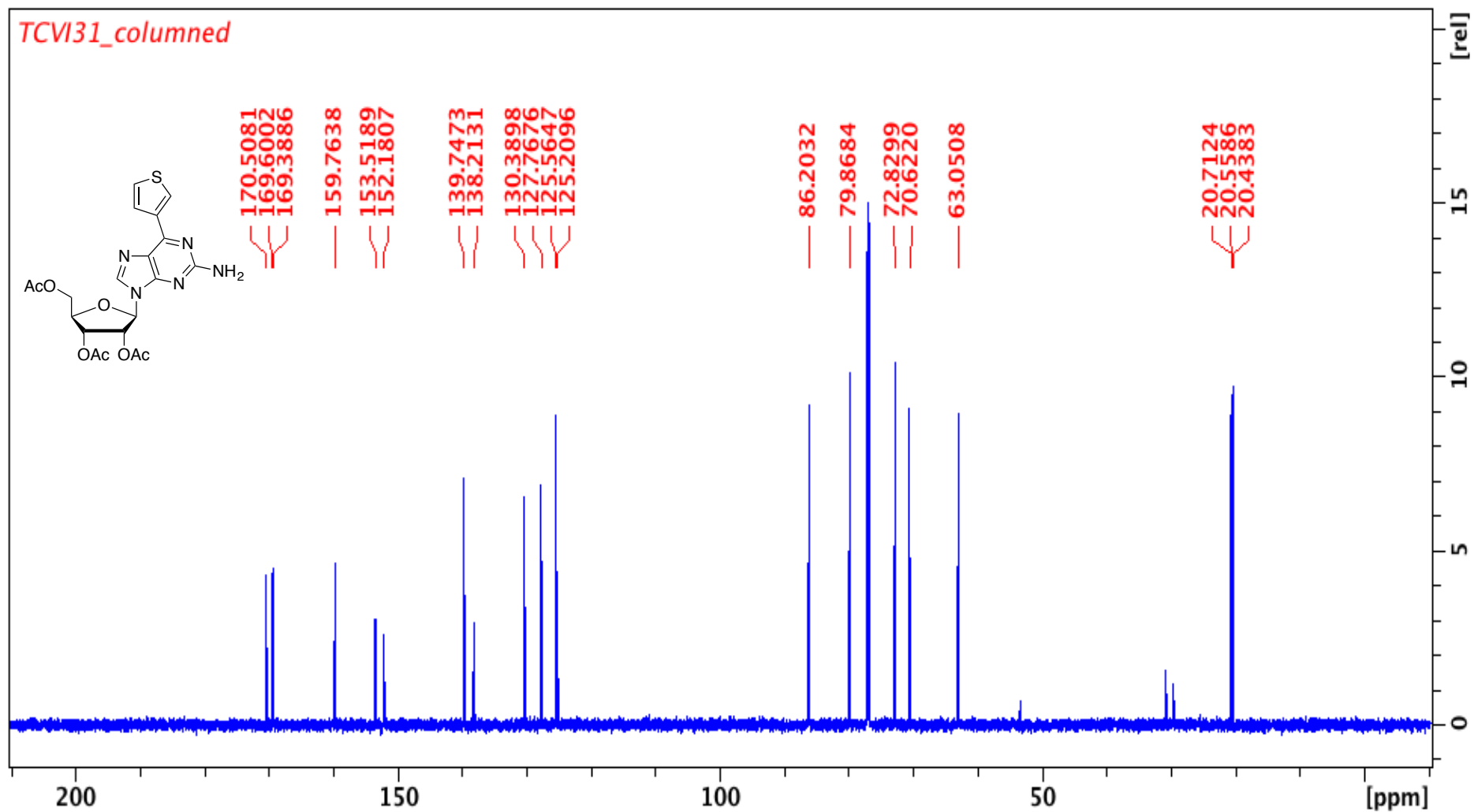
Appendix 80. $^1\text{H-NMR}$ spectrum of **184**. 400 MHz, DMSO- d_6 . Peak at 3.3 ppm comes from water. Peak at 2.5 comes from DMSO. Peaks at 4.0 and 3.1 ppm come from ethanol.



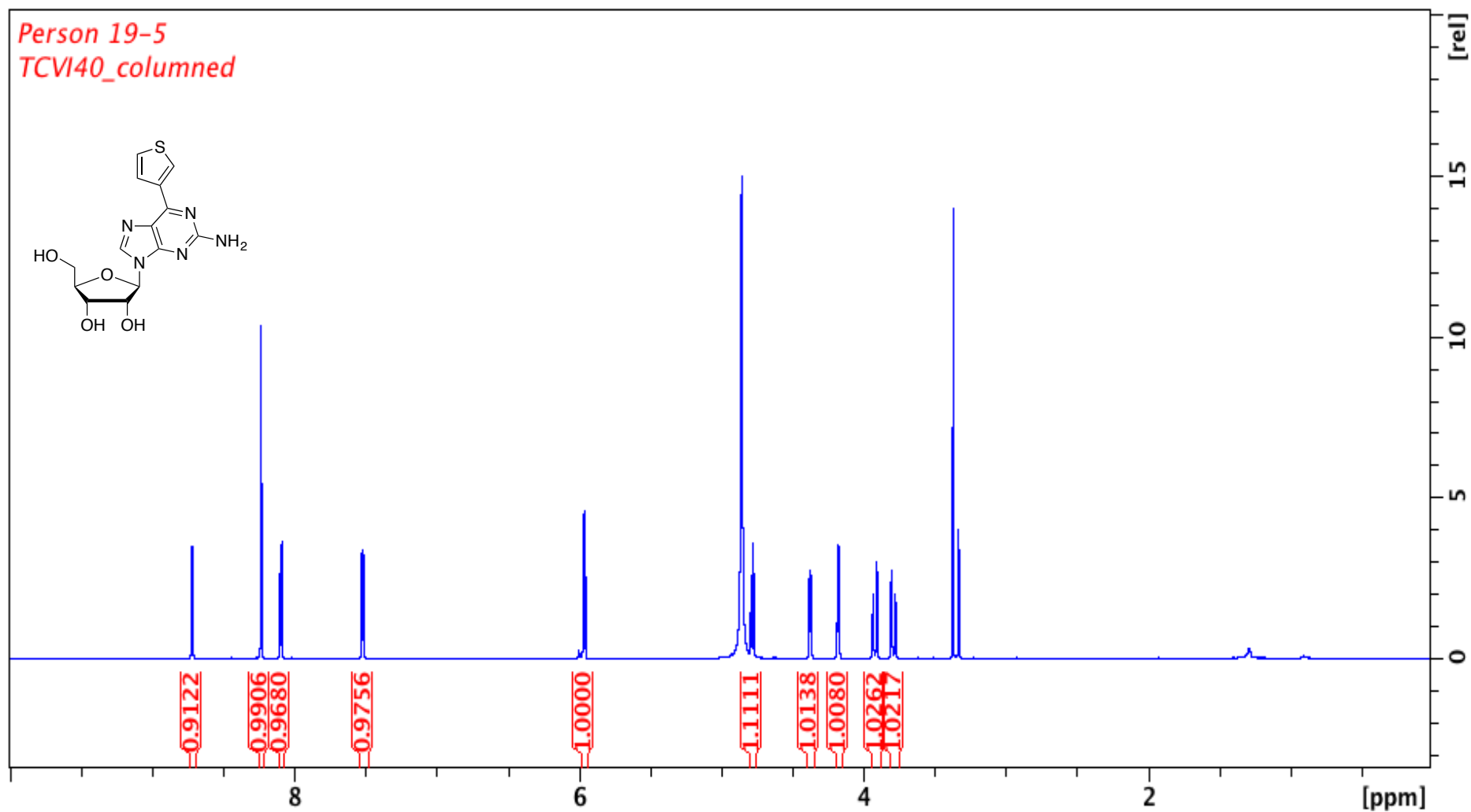
Appendix 81. ^{13}C -NMR spectrum of **184**. 100 MHz, DMSO- d_6 . Peak at 38 ppm comes from DMSO.



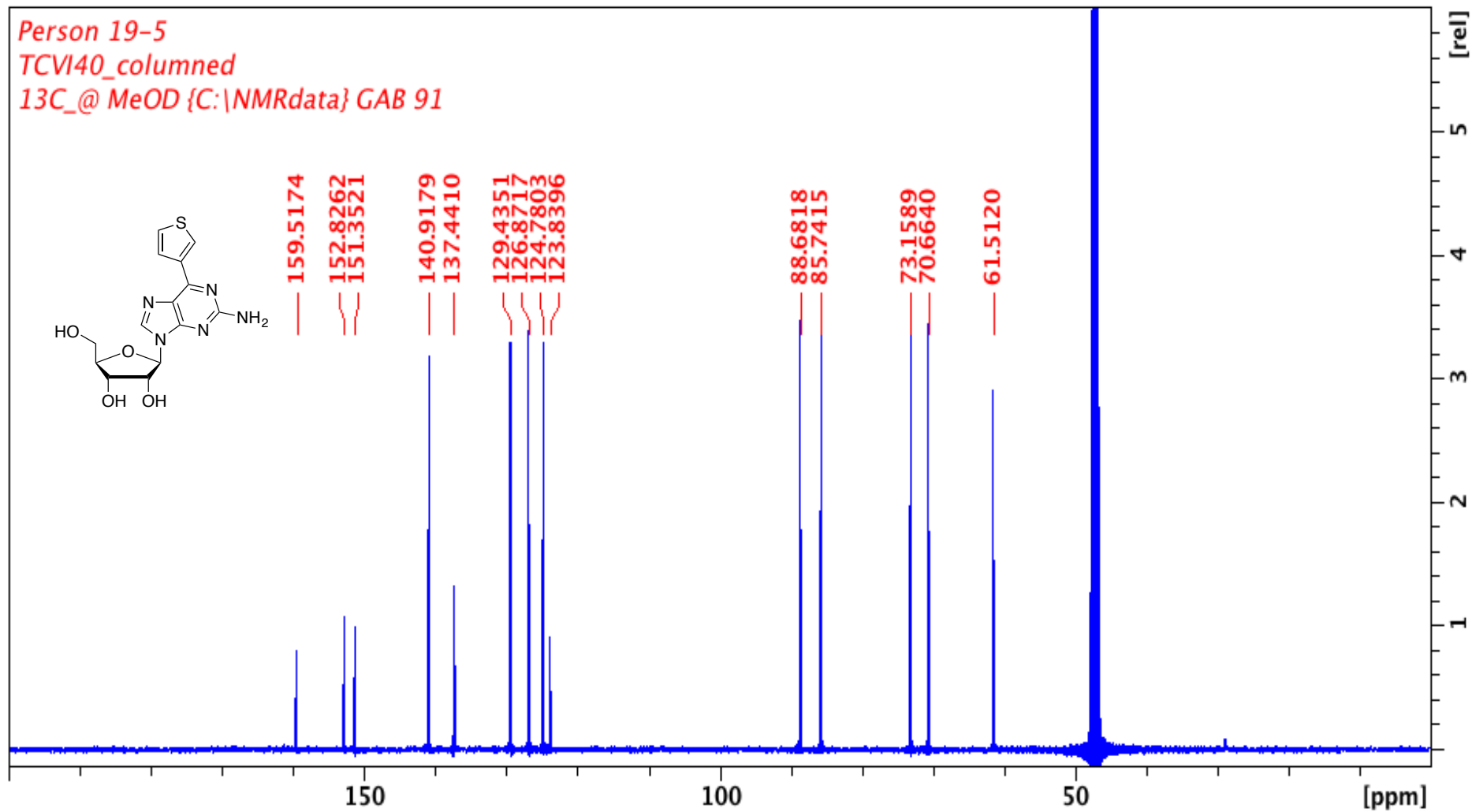
Appendix 82. $^1\text{H-NMR}$ spectrum of **180**. 600 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peak at 5.25 ppm comes from DCM. Peak at 2.1 comes from acetone. Peak at 1.4 ppm comes from grease.



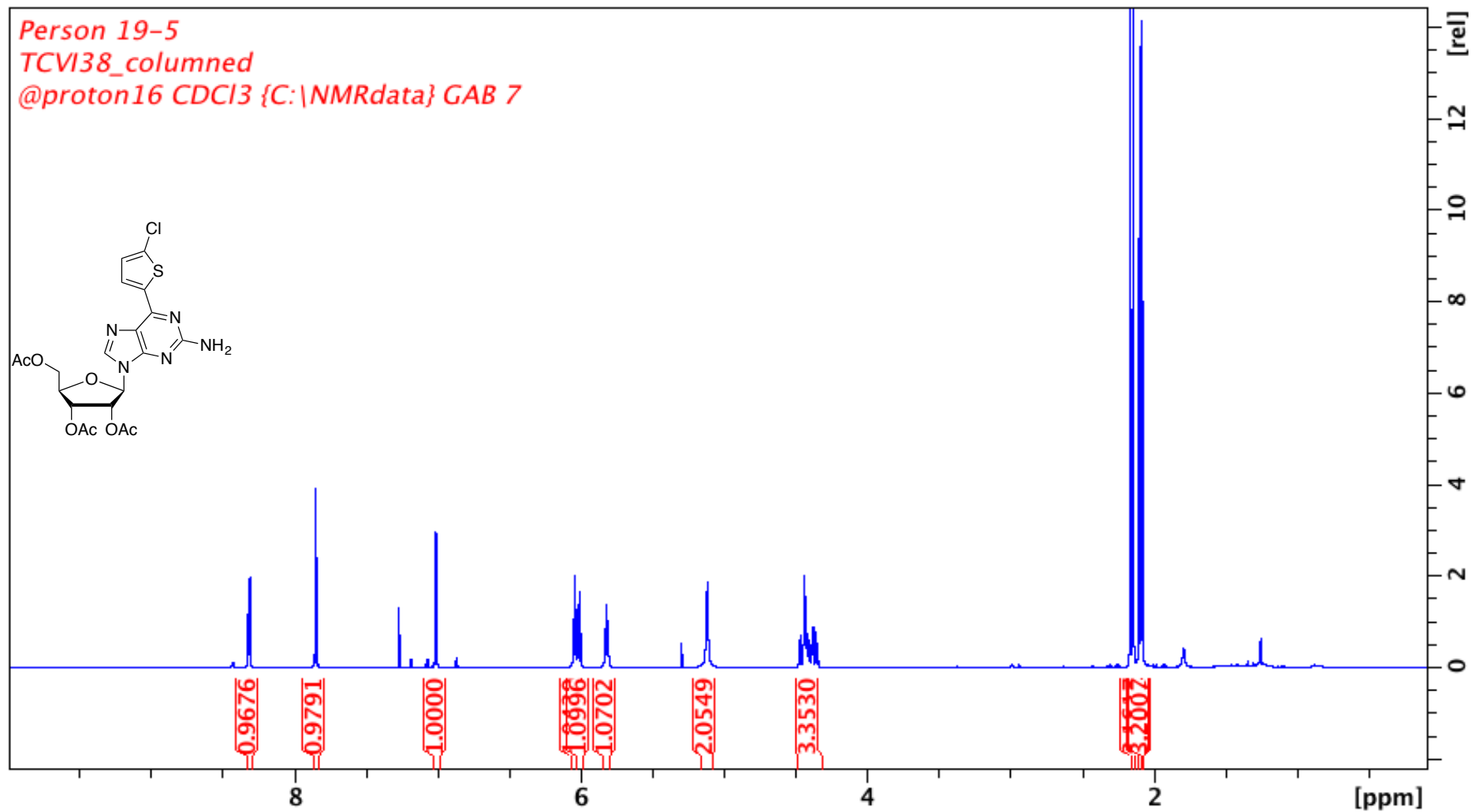
Appendix 83. ^{13}C -NMR spectrum of **180**. 150 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 53 ppm comes from DCM. Peak at 31 comes from acetone. Peak at 30 ppm comes from grease.



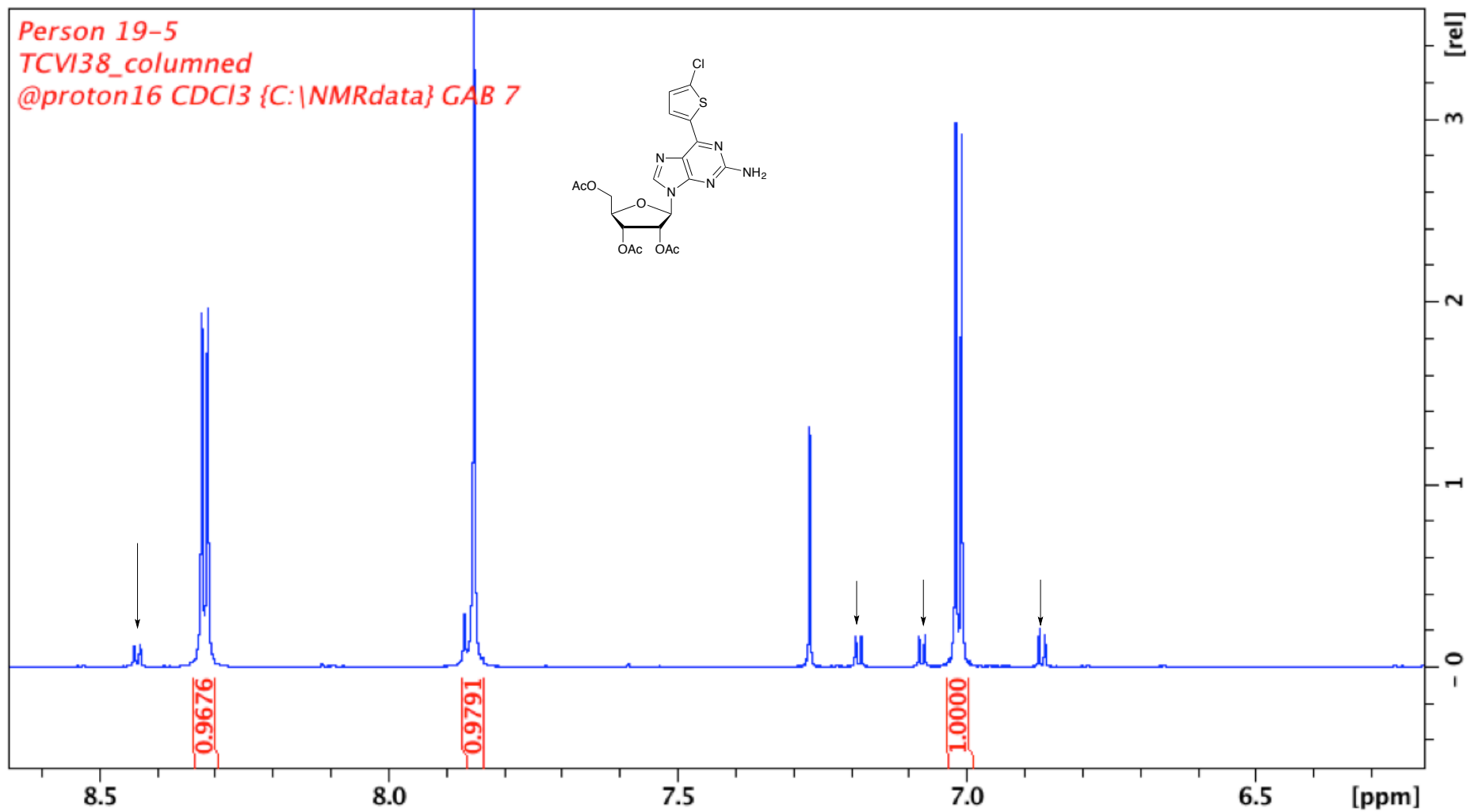
Appendix 84. $^1\text{H-NMR}$ spectrum of **185**. 500 MHz, MeOD. Peak at 3.8 ppm comes from water. Peaks at 3.3 ppm come from MeOH.



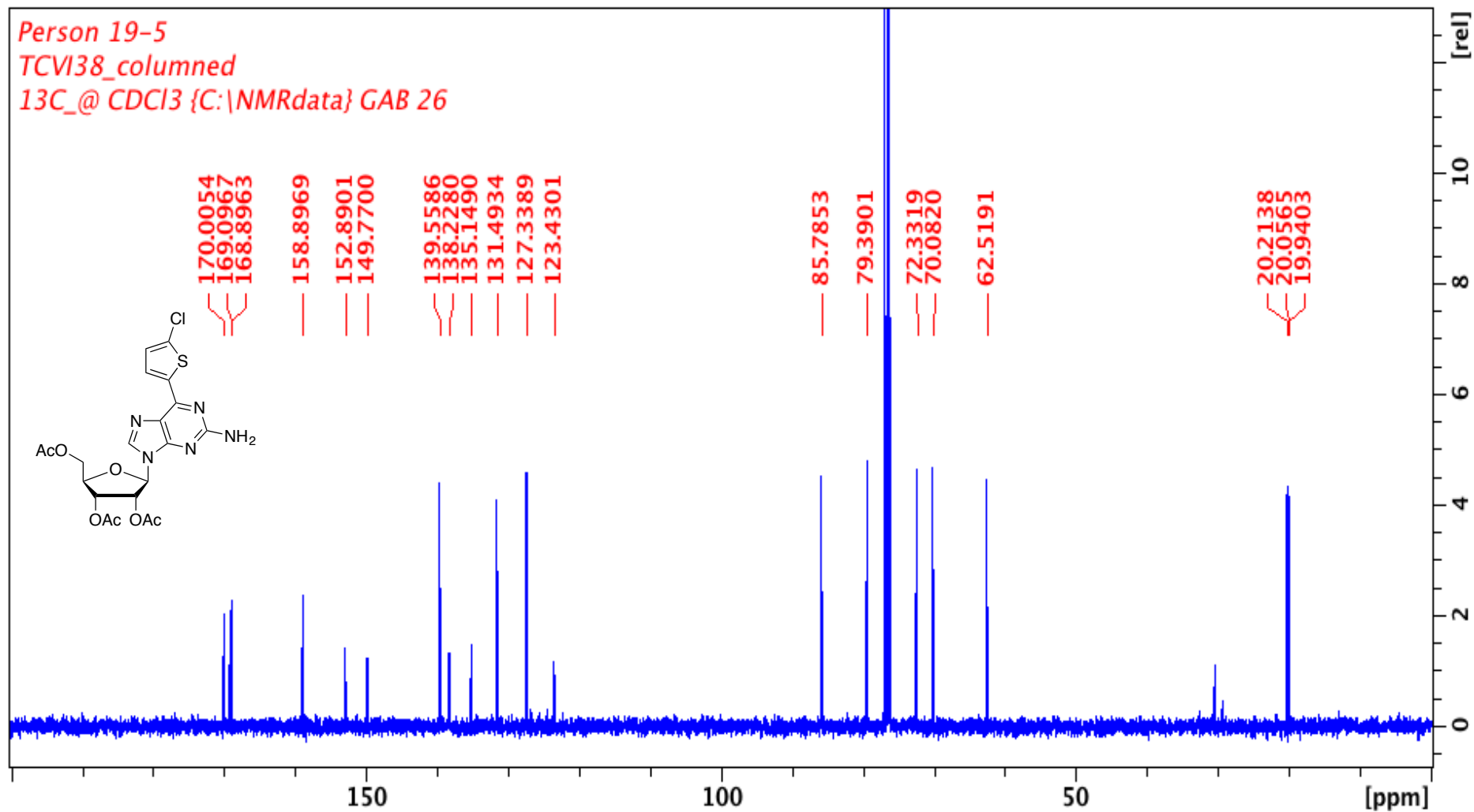
Appendix 85. ^{13}C -NMR spectrum of 185. 100 MHz, MeOD. Peak at 49 ppm comes from MeOH.



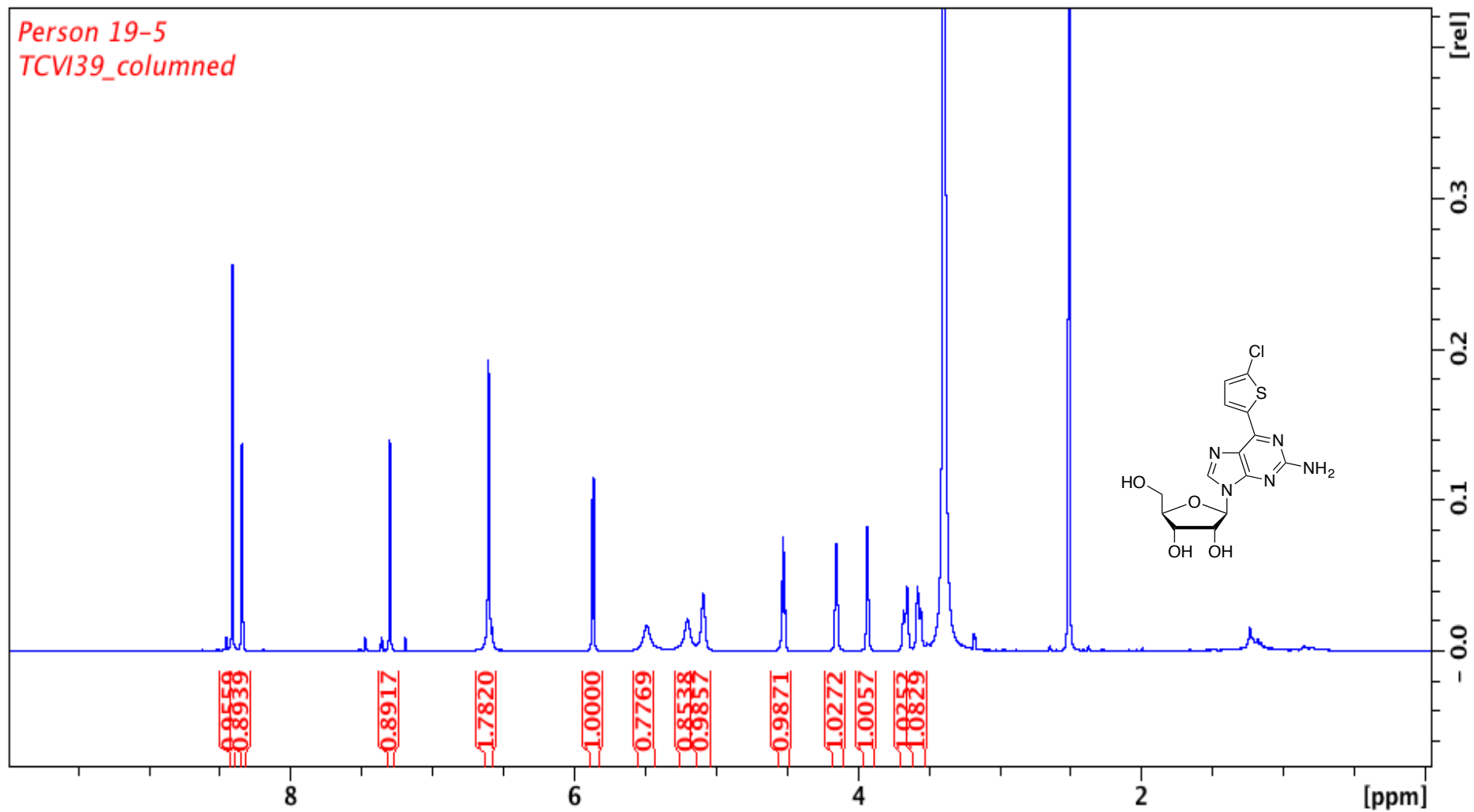
Appendix 86. $^1\text{H-NMR}$ spectrum of **181**. 400 MHz, CDCl_3 . Peak at 7.26 ppm comes from CHCl_3 . Peak at 5.25 ppm comes from DCM.



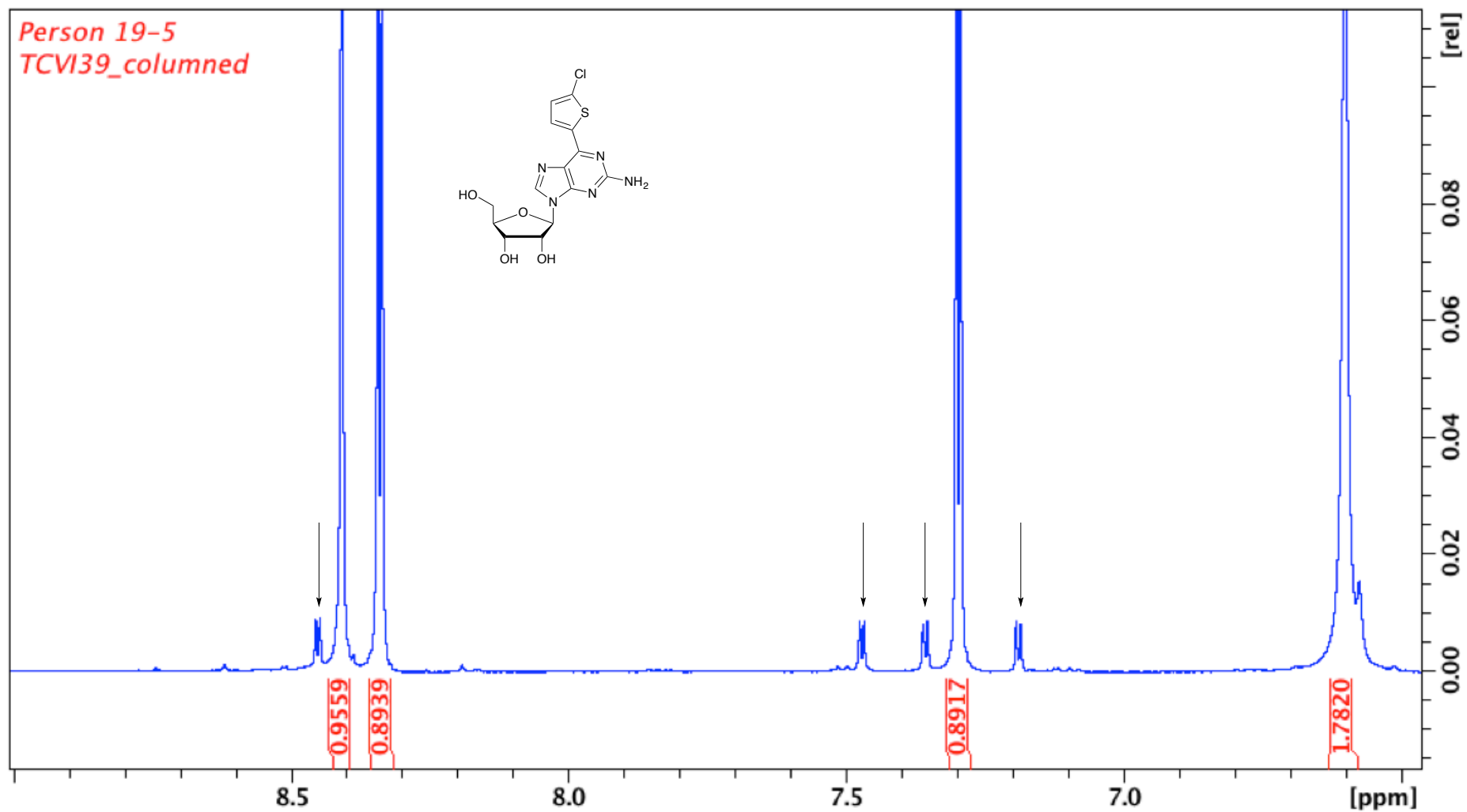
Appendix 87. Expanded section of Appendix 81 showing the peaks corresponding to bis-thiophene **182**. Peak at 7.26 comes from CHCl₃.



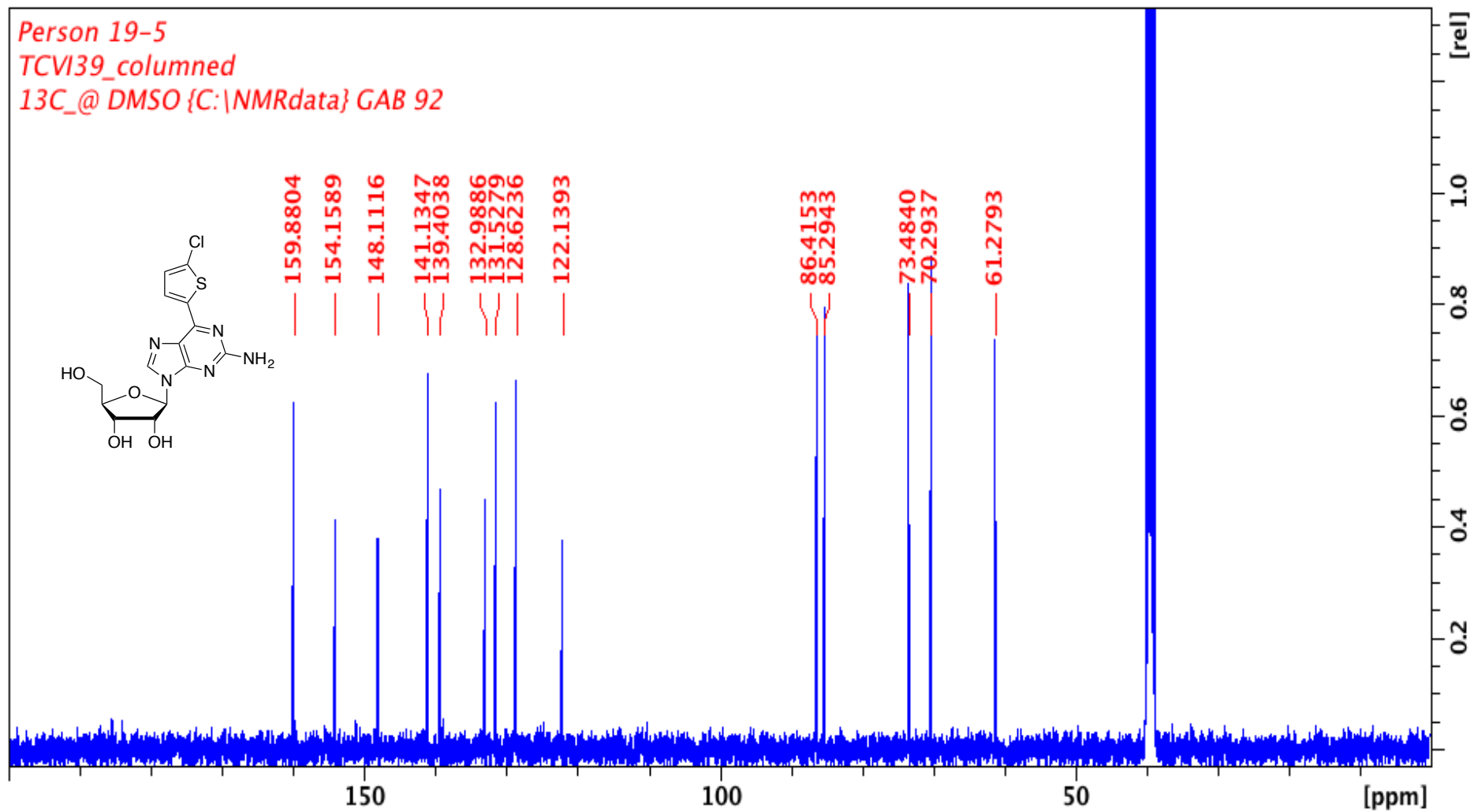
Appendix 88. ^{13}C -NMR spectrum of 185. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 31 ppm comes from acetone.



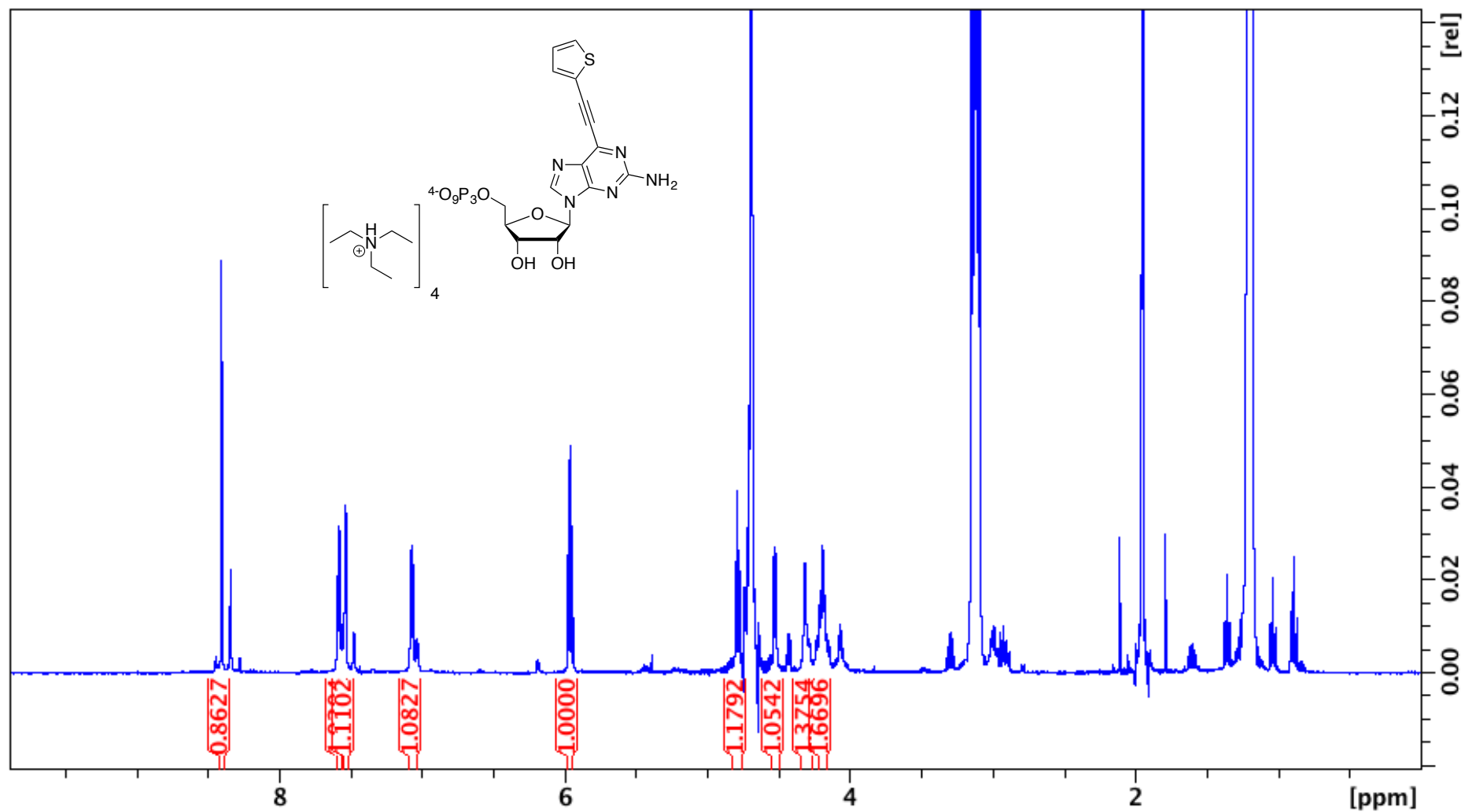
Appendix 89. $^1\text{H-NMR}$ spectrum of 185. 500 MHz, DMSO- d_6 . Peak at 3.4 ppm comes from water. Peak at 2.5 ppm come from DMSO.



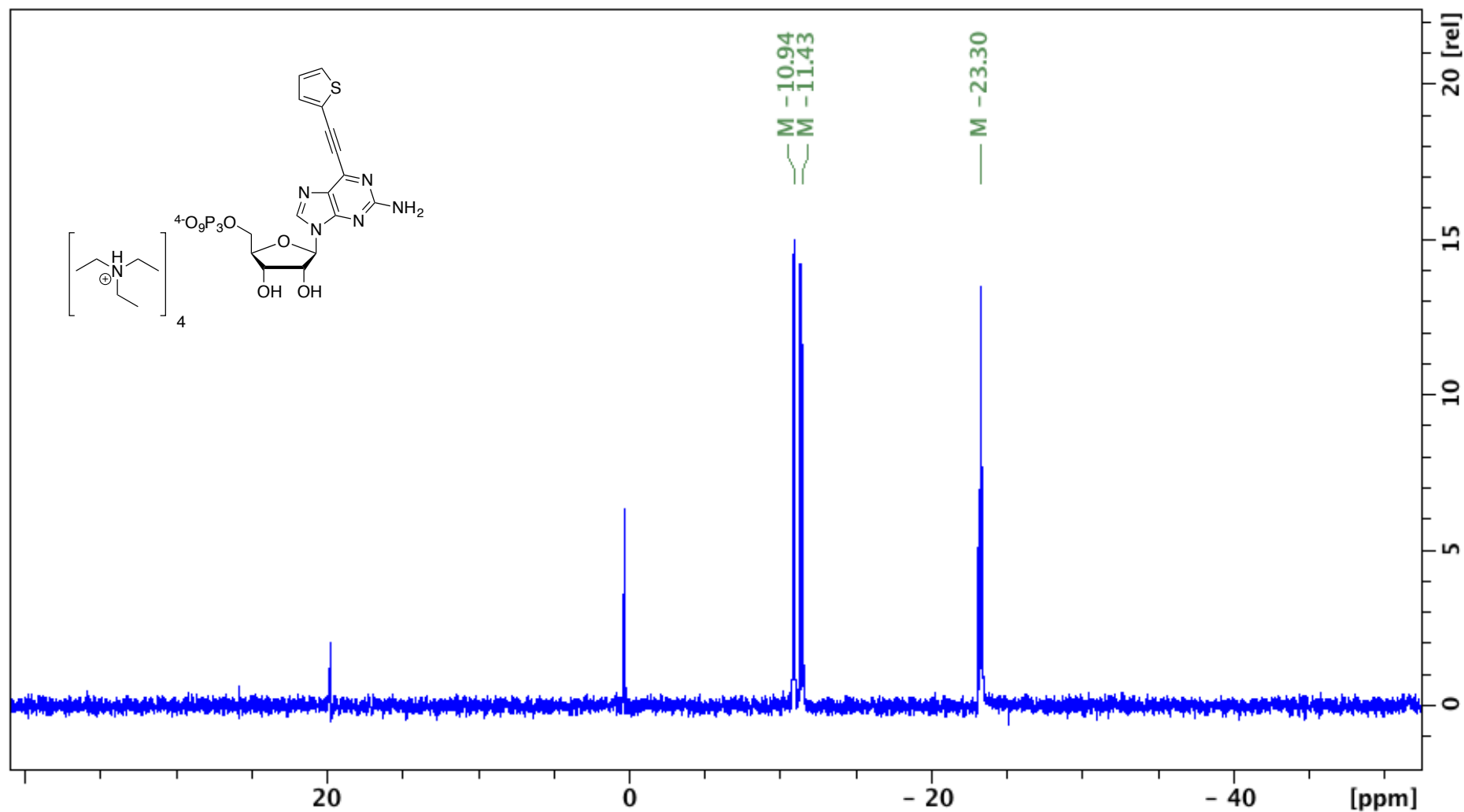
Appendix 90. Expanded section of Appendix 84 showing the peaks corresponding to bis-thiophene **182**. Peak at 7.26 comes from CHCl_3



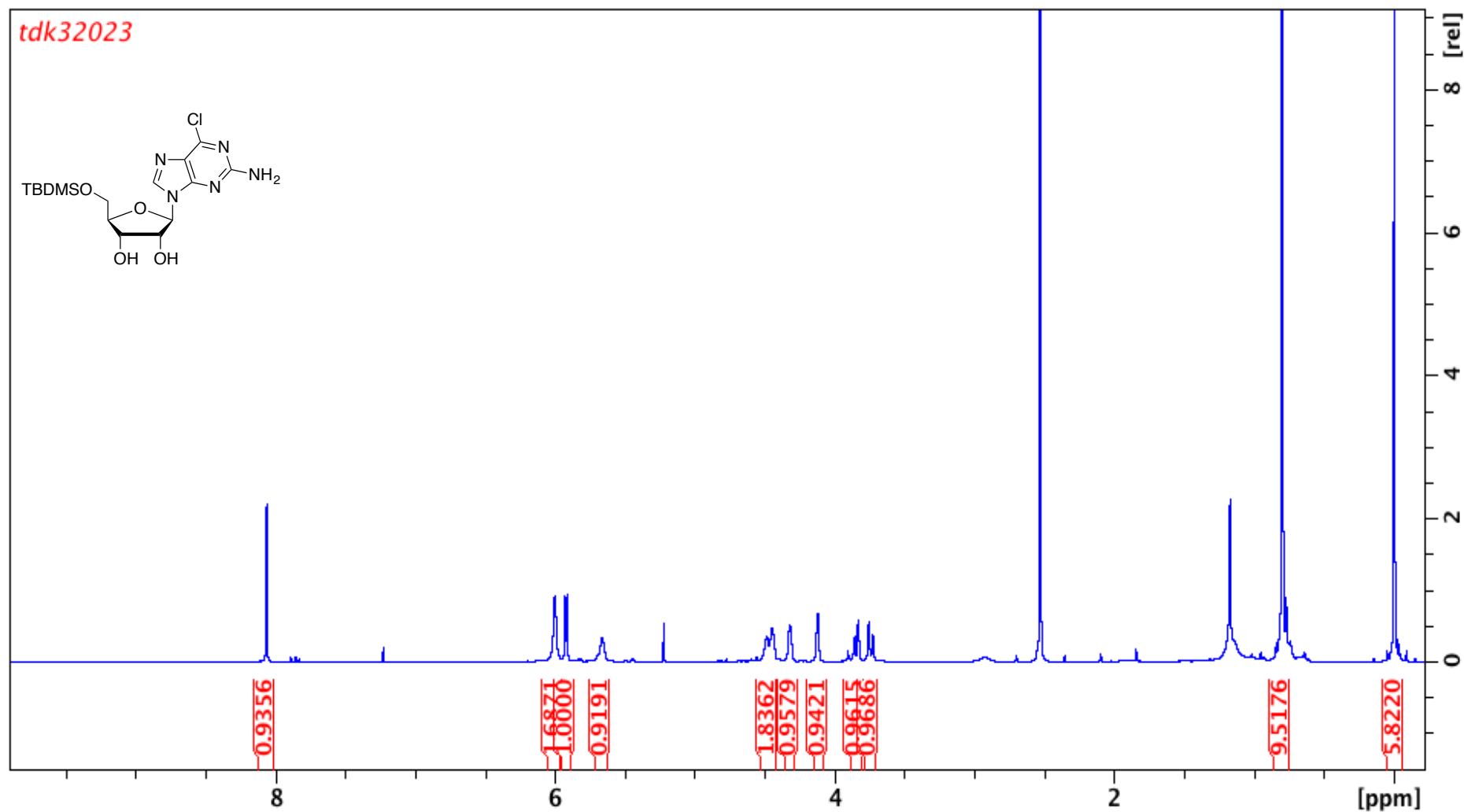
Appendix 91. ^{13}C -NMR spectrum of 185. 100 MHz, DMSO- d_6 . Peak at 38 ppm comes from DMSO.



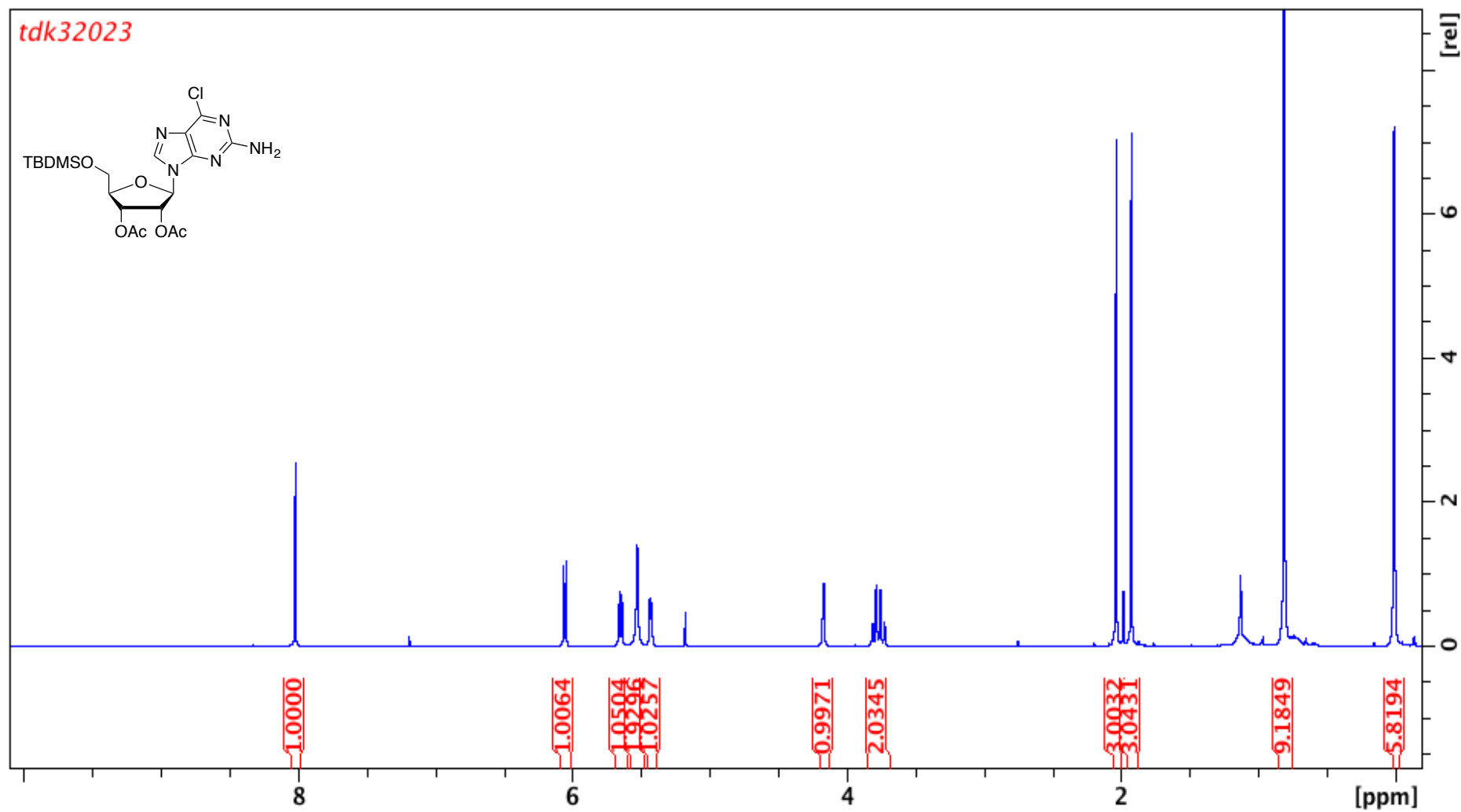
Appendix 92. $^1\text{H-NMR}$ spectrum of **200**. 400 MHz, D_2O . Peak at 4.7 ppm comes from water. Peaks at 2.0, 1.2 and 3.10 ppm come from triethylammonium acetate.



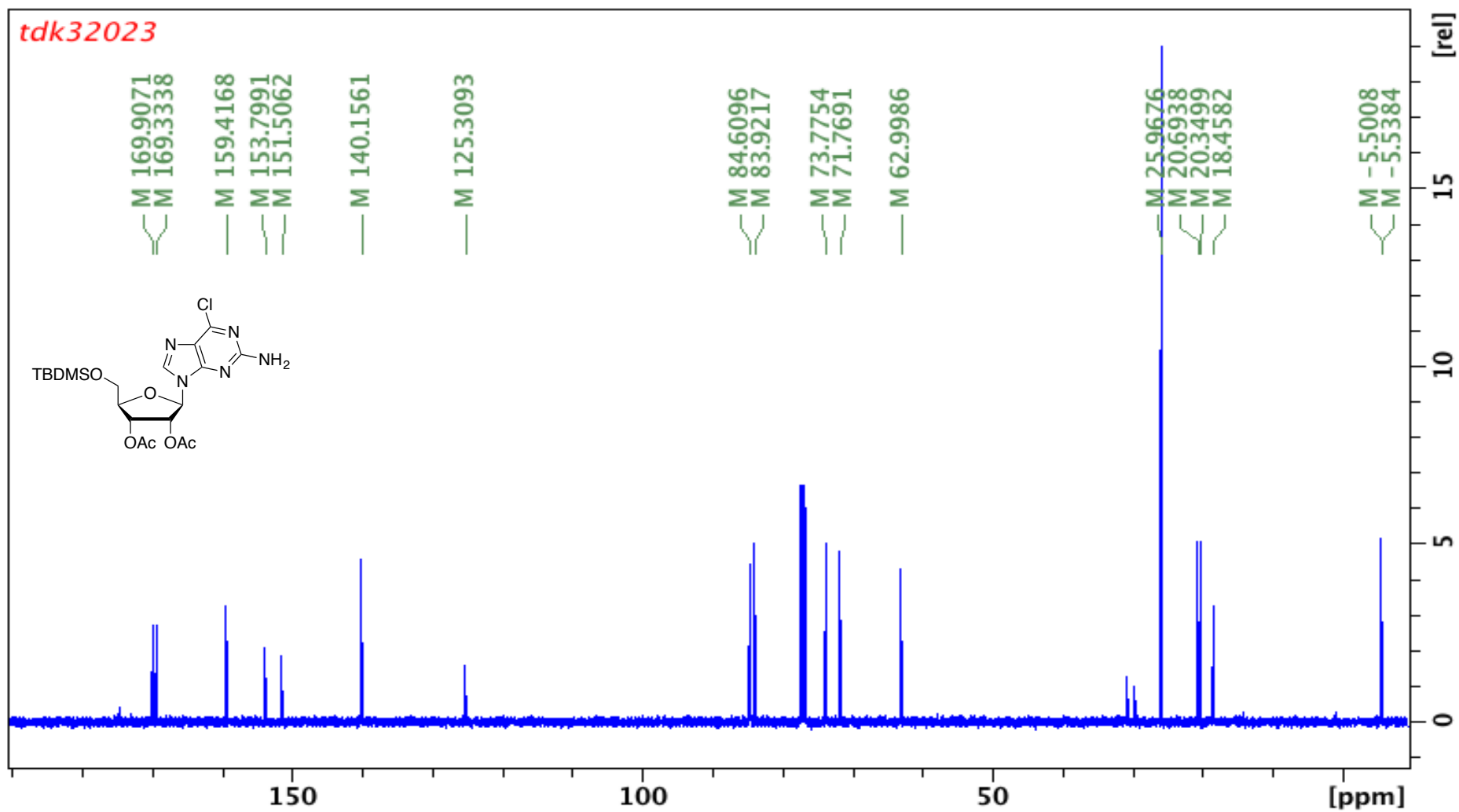
Appendix 93. ^{31}P -NMR spectrum of **200**. 161 MHz, D_2O . Peak at 6.0 ppm comes from unknown pyrophosphoate related impurity. Peak at 0.3 ppm comes from the corresponding monophosphate. Peak at 19.8 ppm comes from cyclic trimetaphosphate.



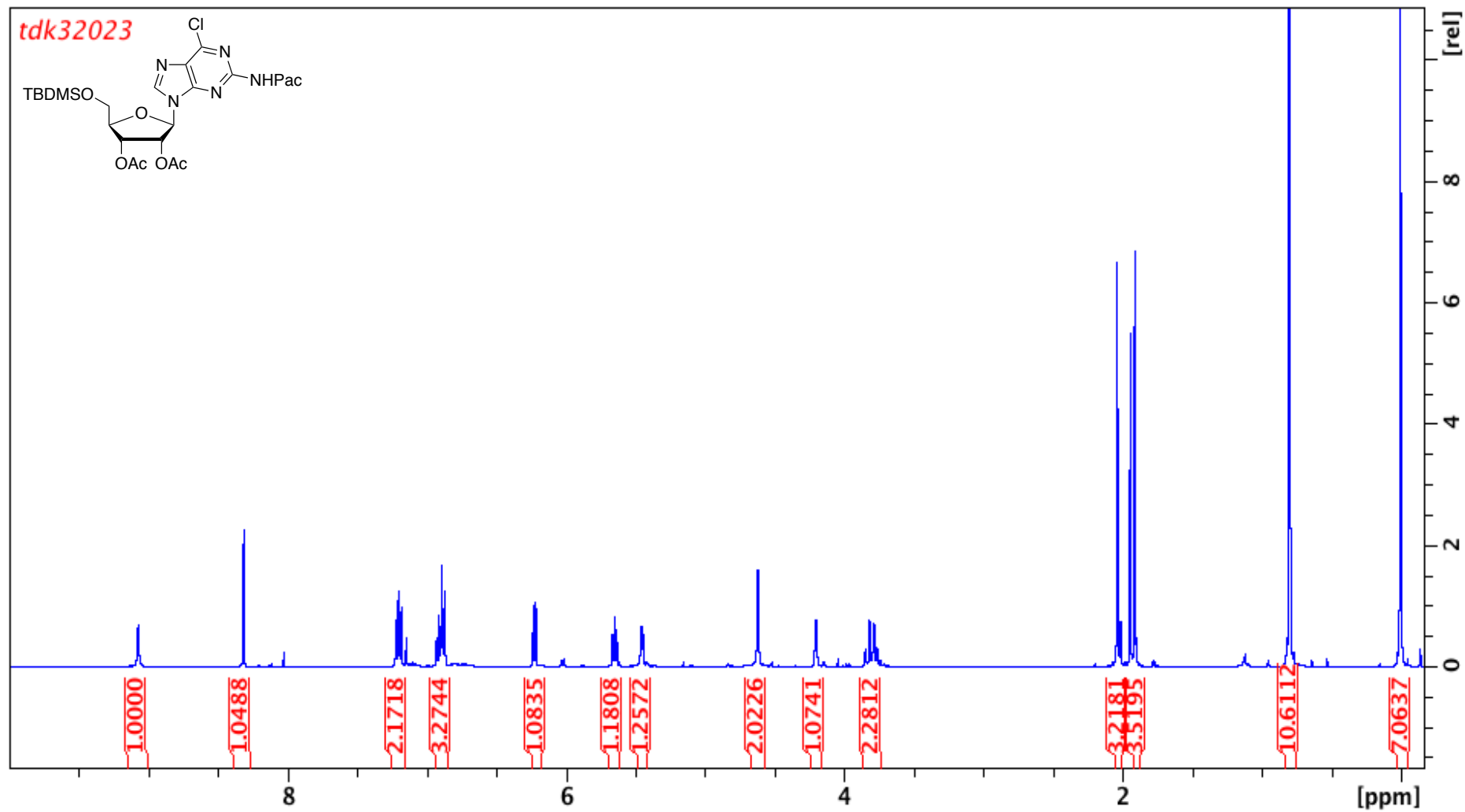
Appendix 94. ¹H-NMR spectrum of **217**. 400 MHz, CDCl₃. Peak at 2.5 ppm comes from DMSO. Peak at 1.2 ppm comes from grease. Peak at 5.3 ppm comes DCM.



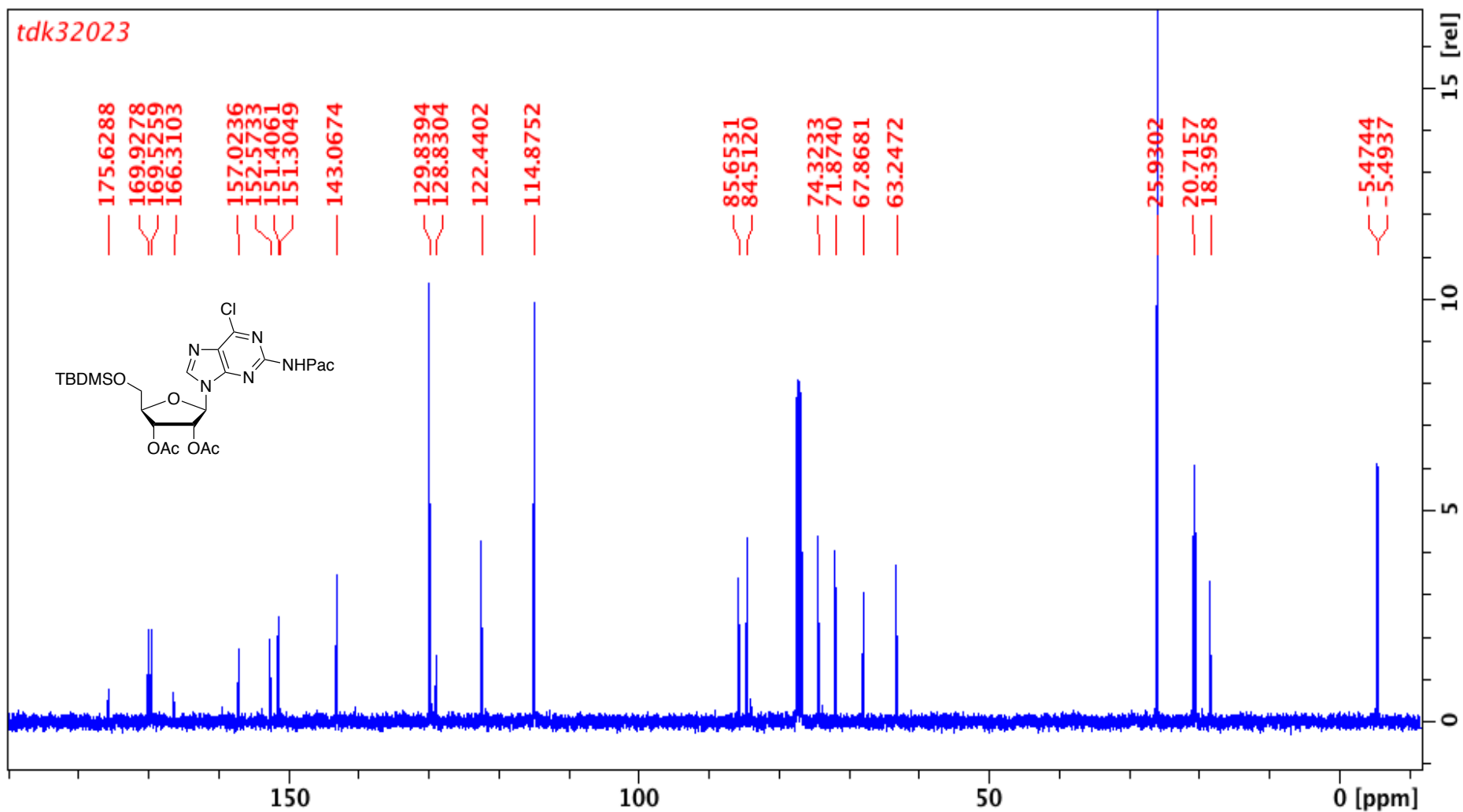
Appendix 96. $^1\text{H-NMR}$ spectrum of **218**. 400 MHz, CDCl_3 . Peaks at 1.2 ppm comes from grease. Peak at 2.0 ppm comes from acetone.



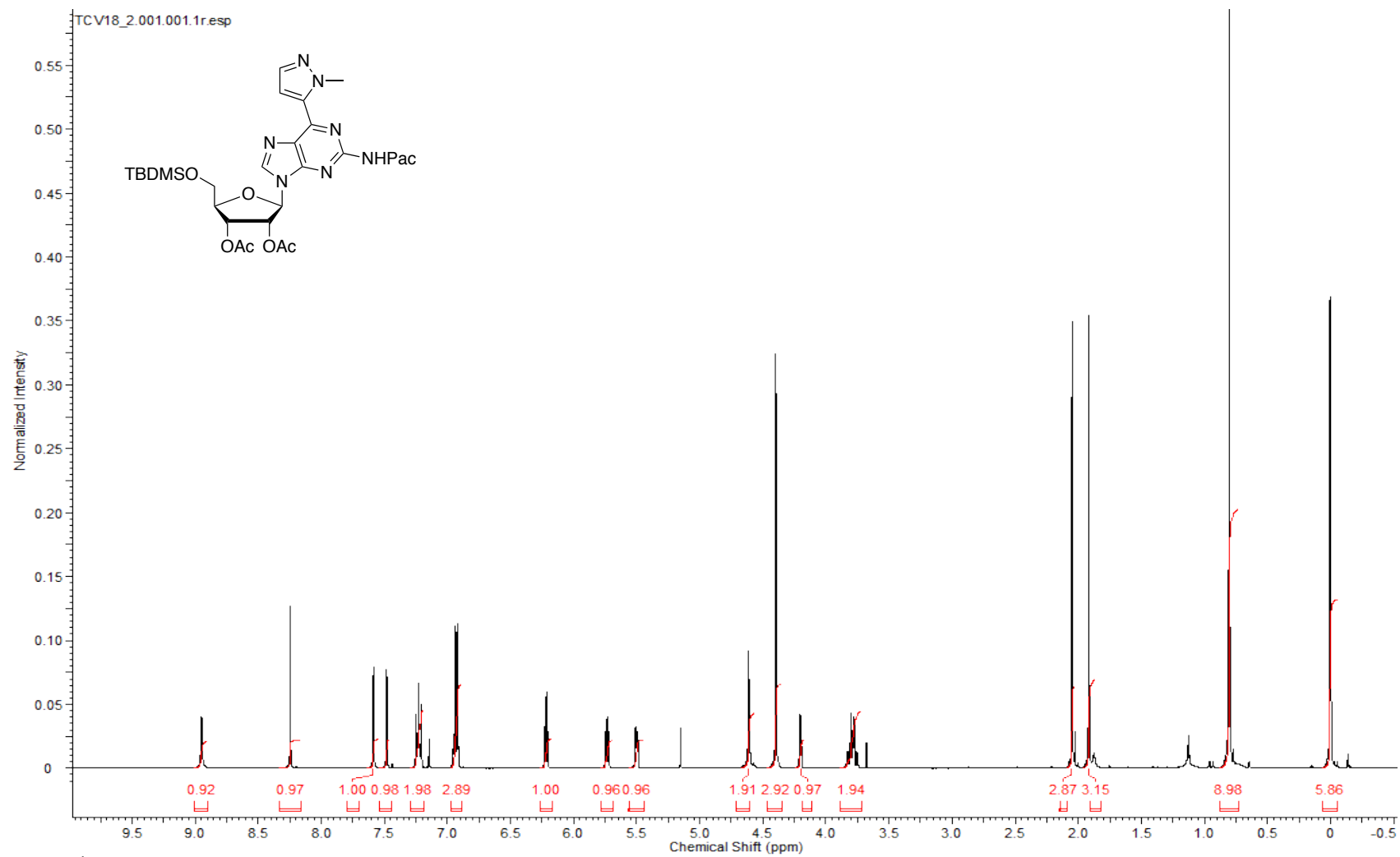
Appendix 97. ^{13}C -NMR spectrum of **218**. 100 MHz, CDCl_3 . Peaks at 29 and 175 ppm come from acetone.



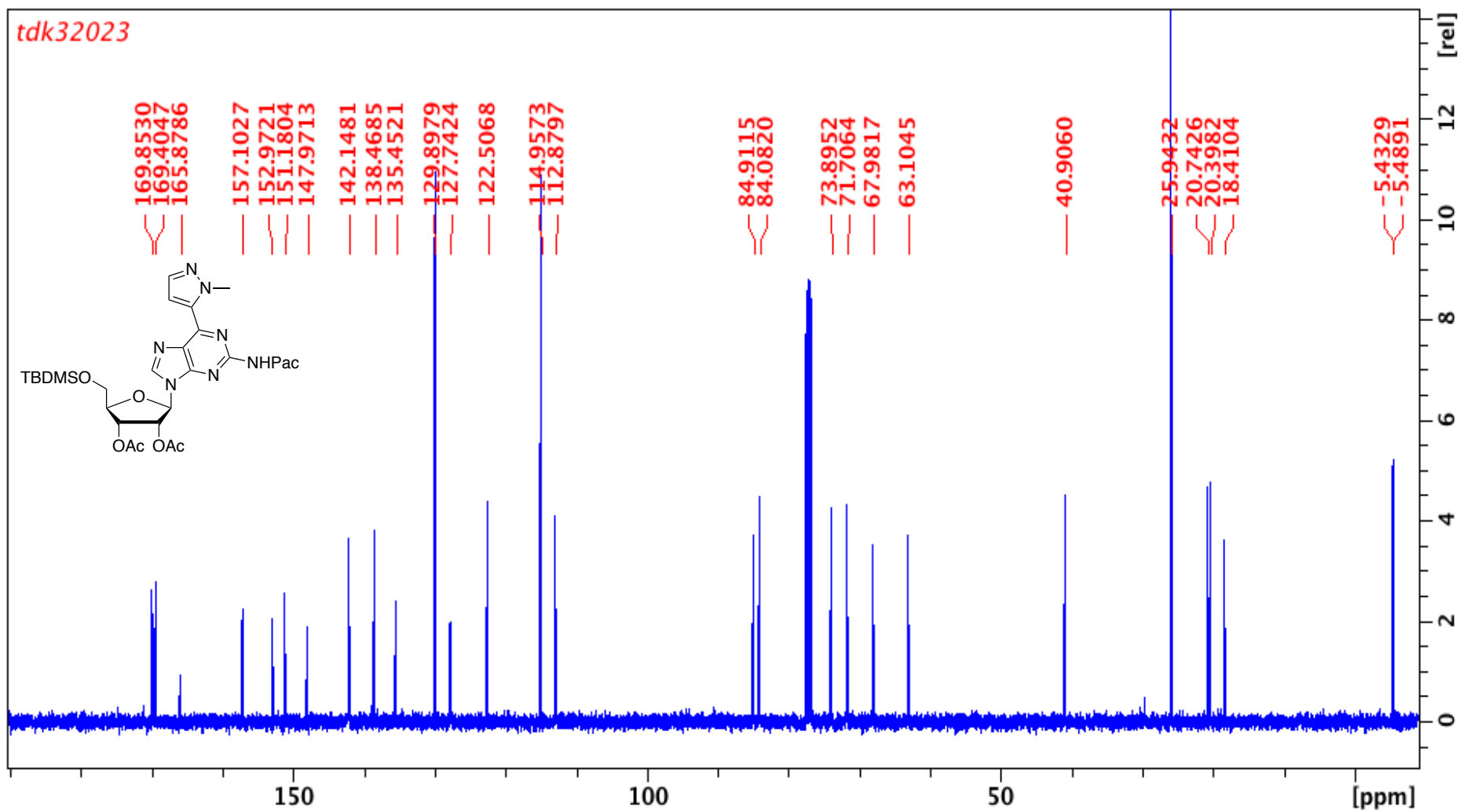
Appendix 98. $^1\text{H-NMR}$ spectrum of **219**. 400 MHz, CDCl_3 . Peak at 7.26 comes from CHCl_3 . Peak at 2.0 ppm comes from acetone.



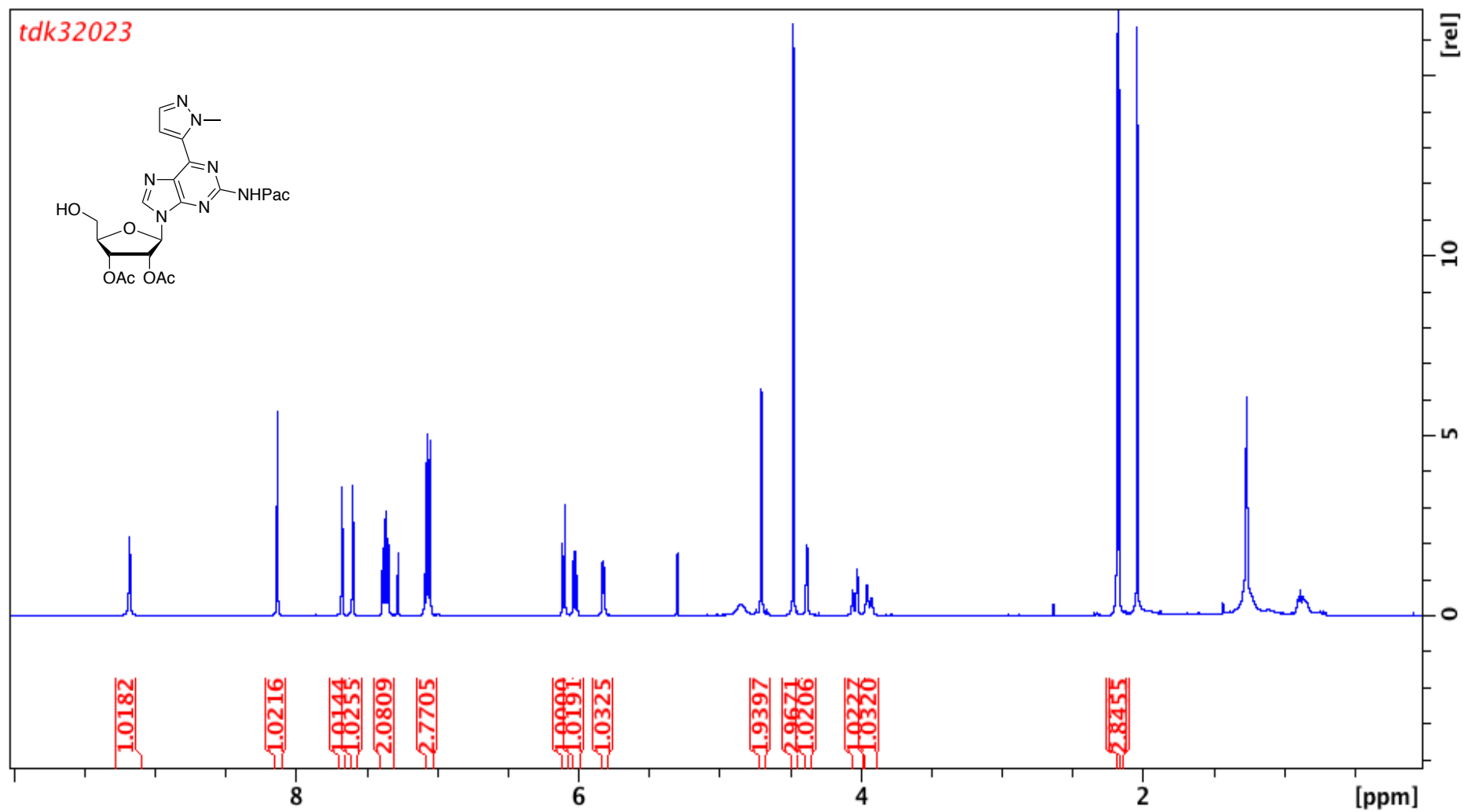
Appendix 99. ^{13}C -NMR spectrum of **219**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .



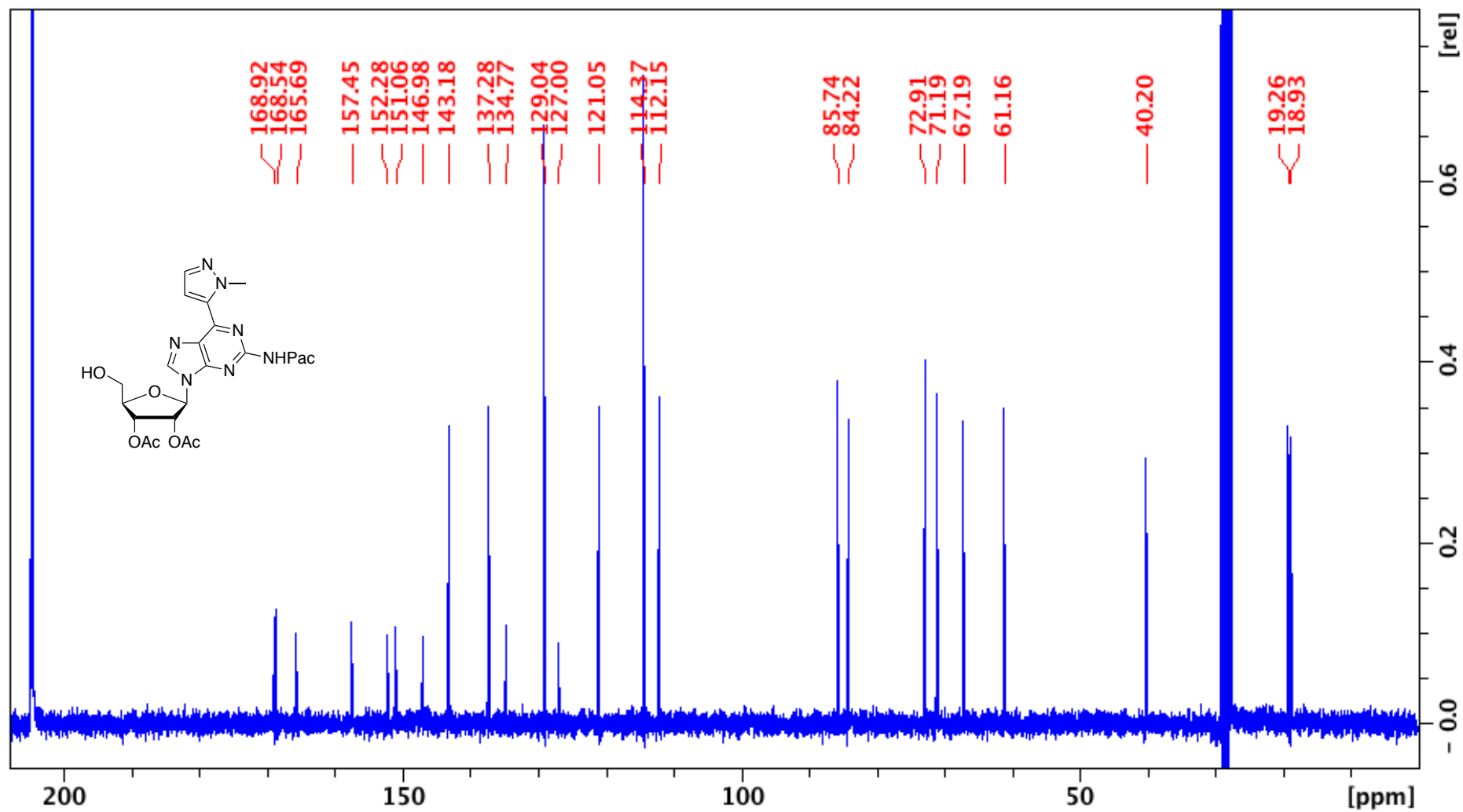
Appendix 100. $^1\text{H-NMR}$ spectrum of **221**. 400 MHz, CDCl_3 . Peak at 7.2 ppm comes from CHCl_3 . Peak at 5.3 ppm comes from DCM.



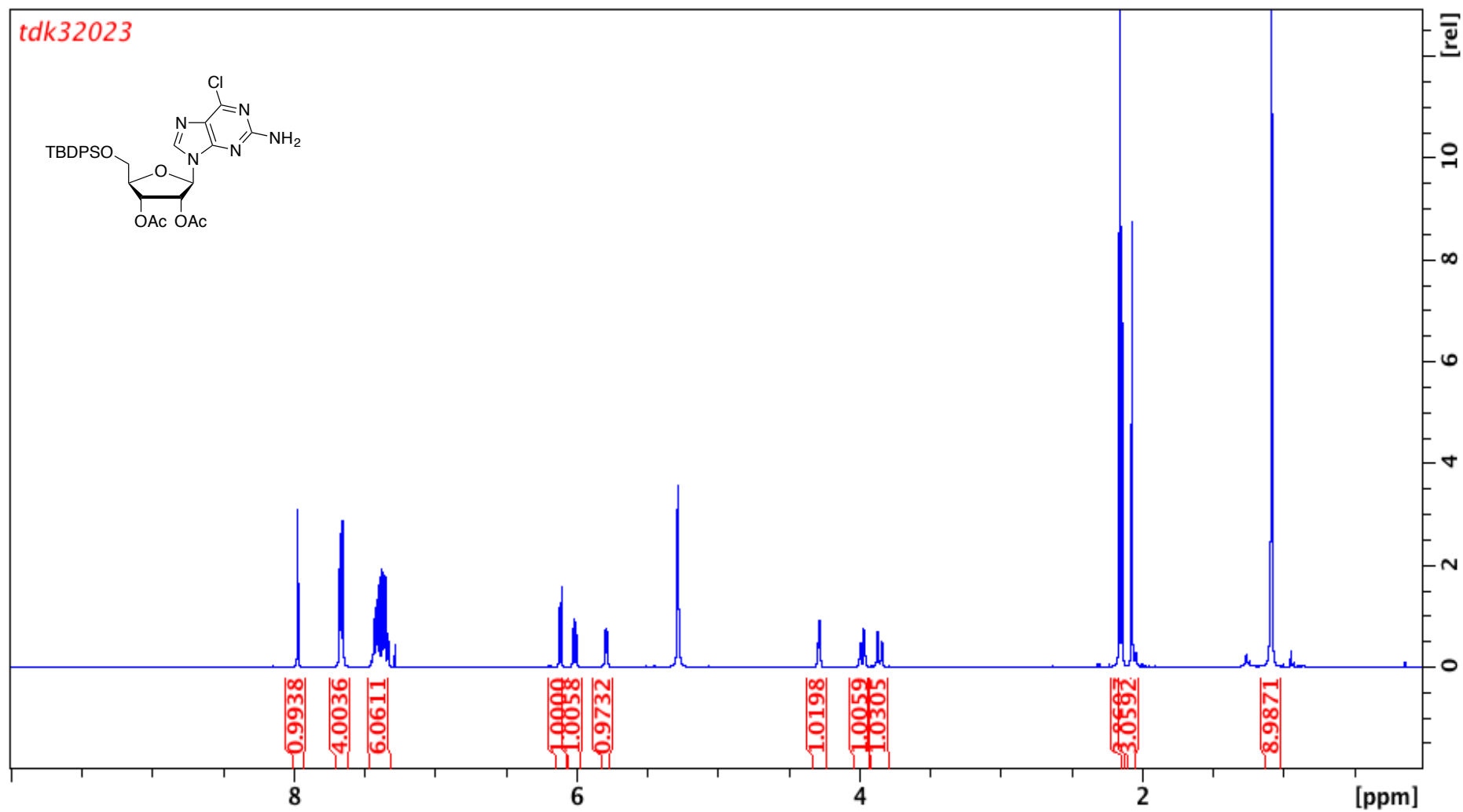
Appendix 101. ^{13}C -NMR spectrum of **221**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .



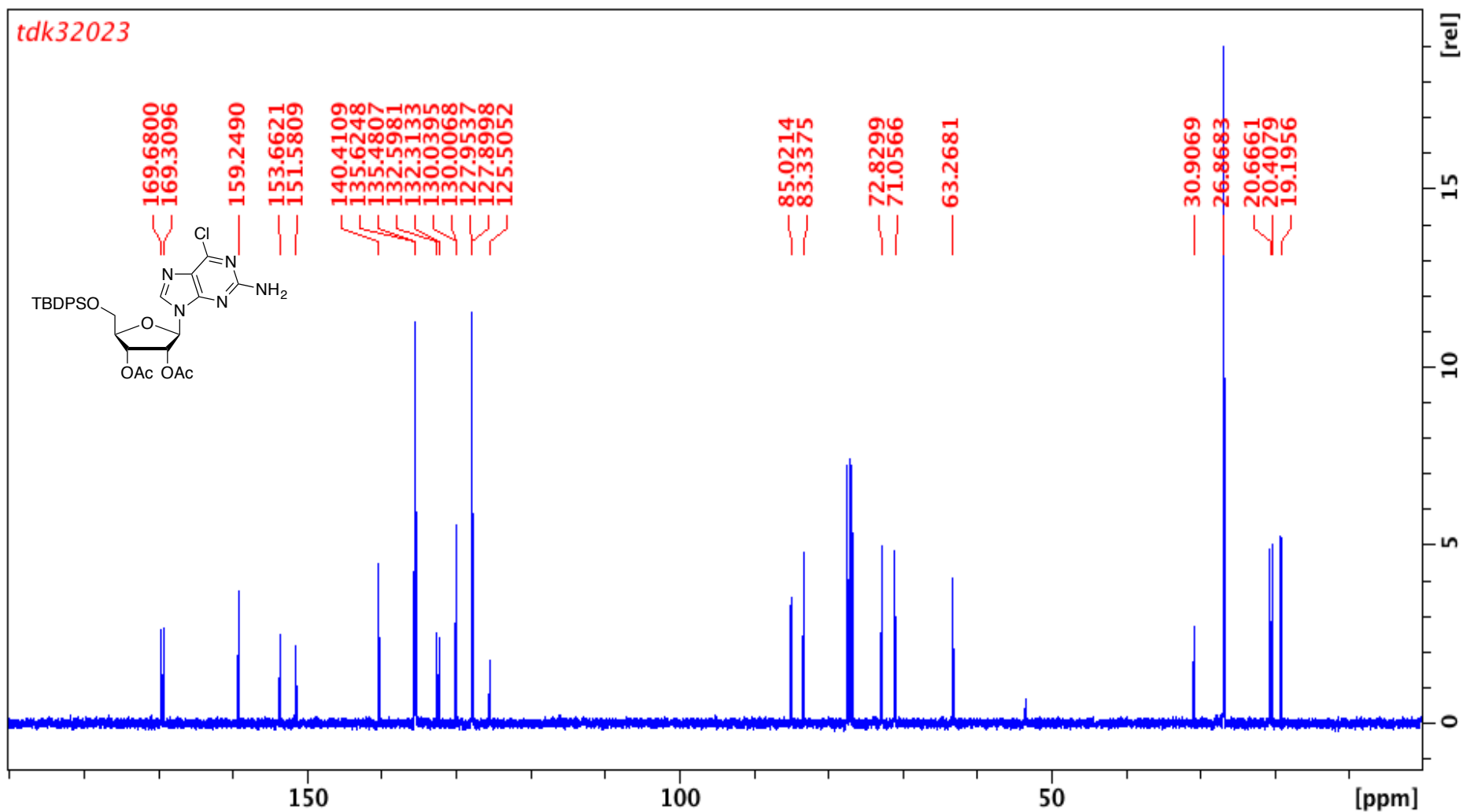
Appendix 102. $^1\text{H-NMR}$ spectrum of **222**. 400 MHz, CDCl_3 . Peak at 7.2 ppm comes from CHCl_3 . Peak at 1.2 ppm comes from grease.



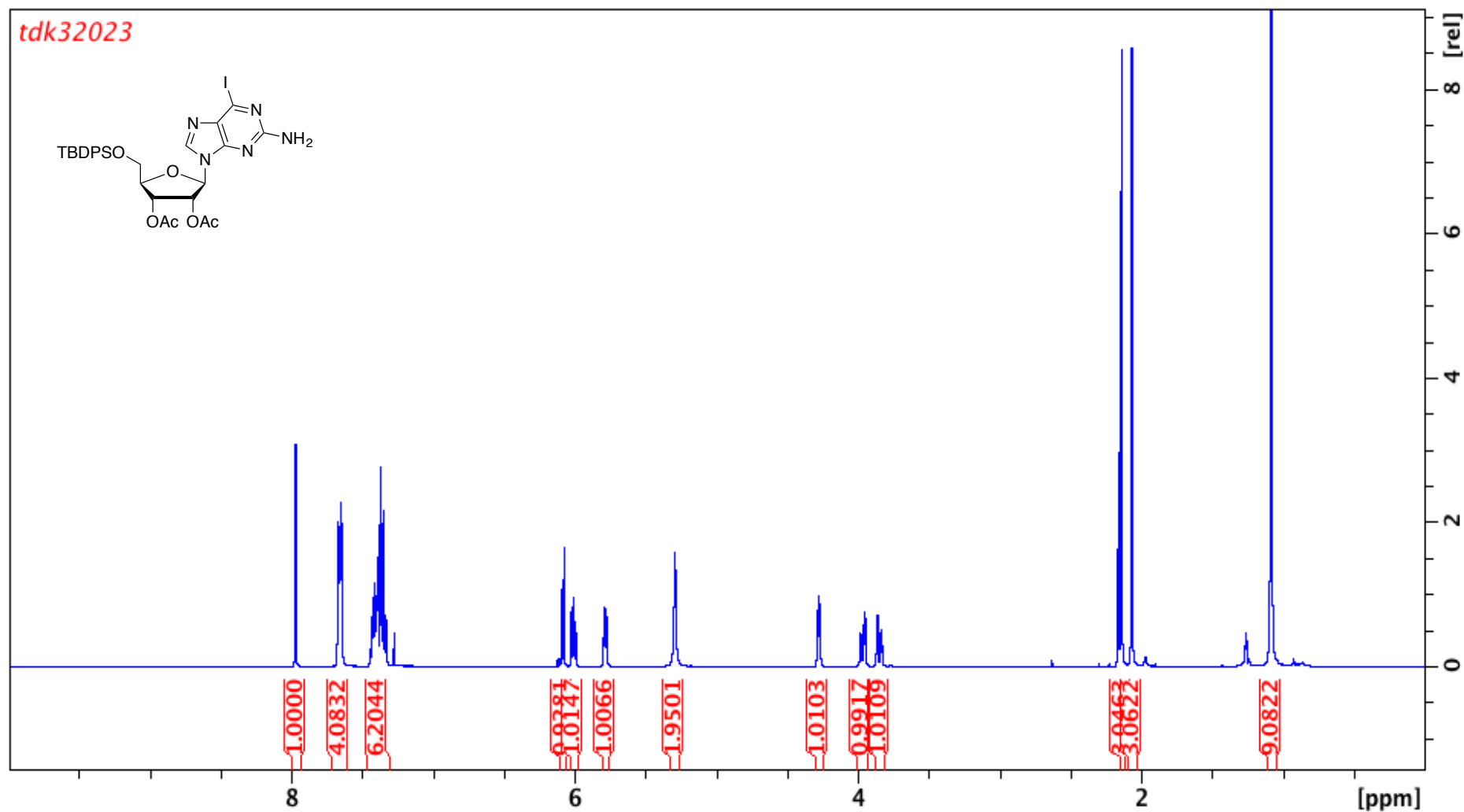
Appendix 103. ^{13}C -NMR spectrum of 222. 100 MHz, acetone- d_6 . Peaks at 28 and 205 ppm come from acetone.



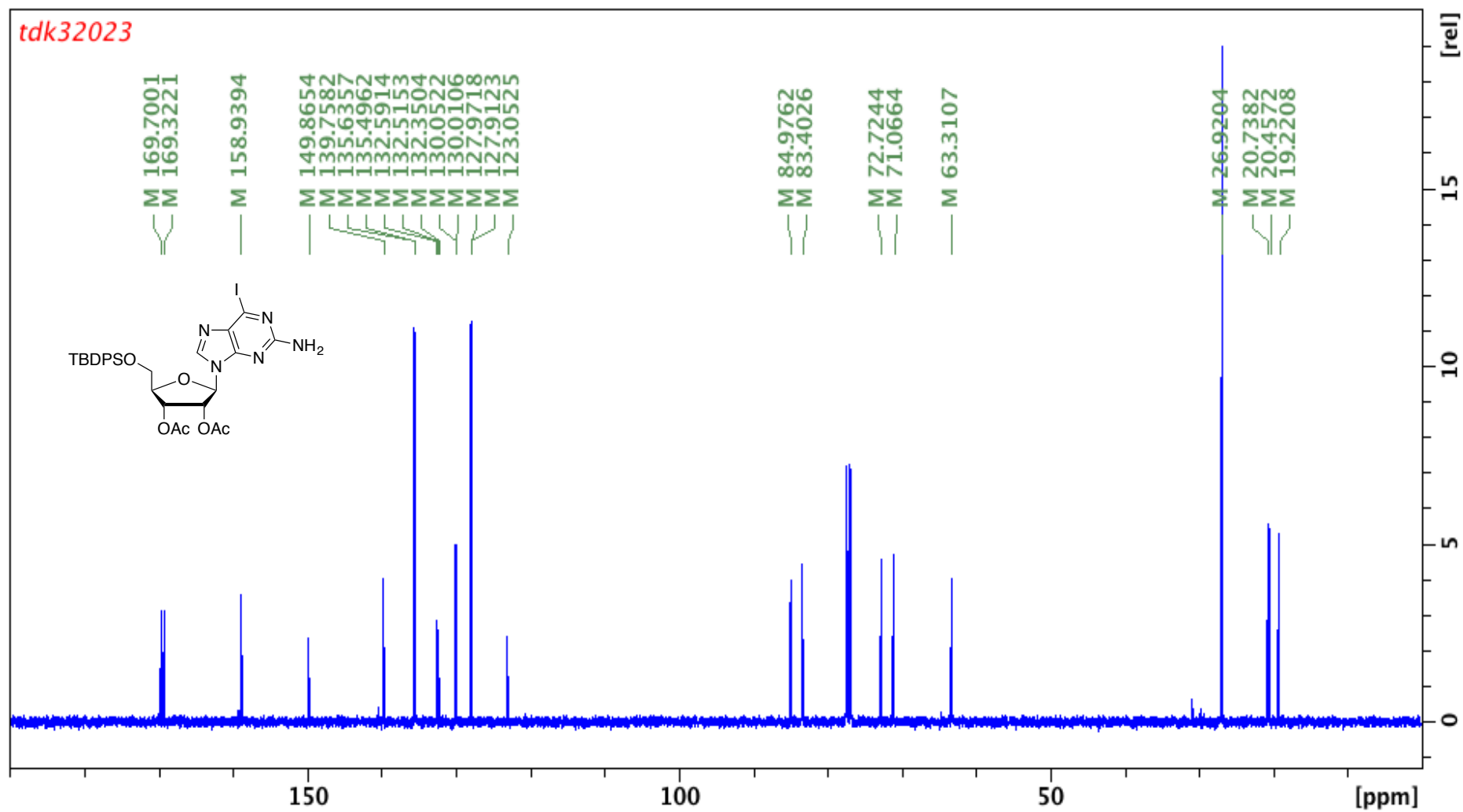
Appendix 106. $^1\text{H-NMR}$ spectrum of **224**. 400 MHz, CDCl_3 . Peak at 7.2 comes from CHCl_3 . Peak at 5.3 comes from DCM.



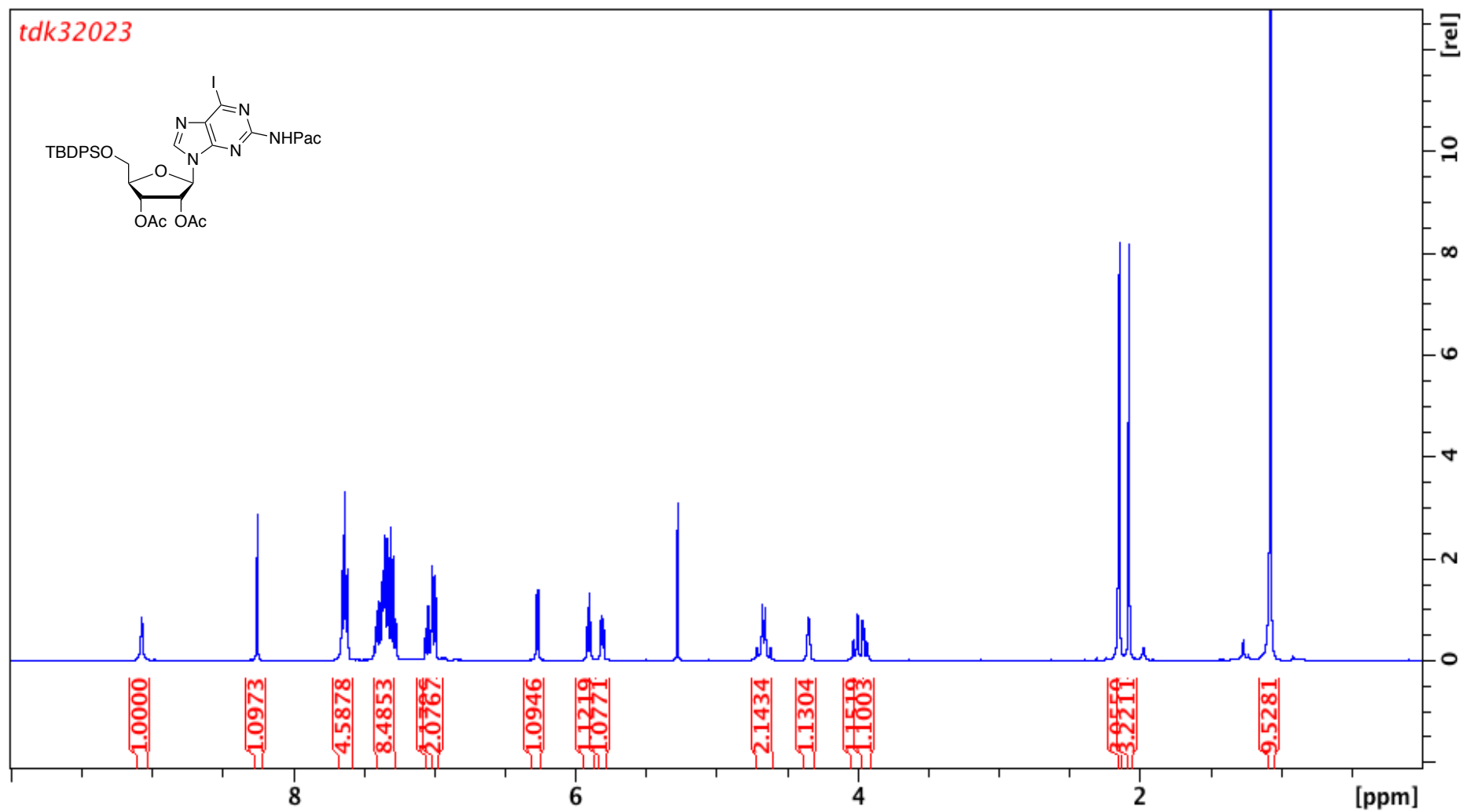
Appendix 107. ^{13}C -NMR spectrum of **224**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 53 ppm comes from DCM.



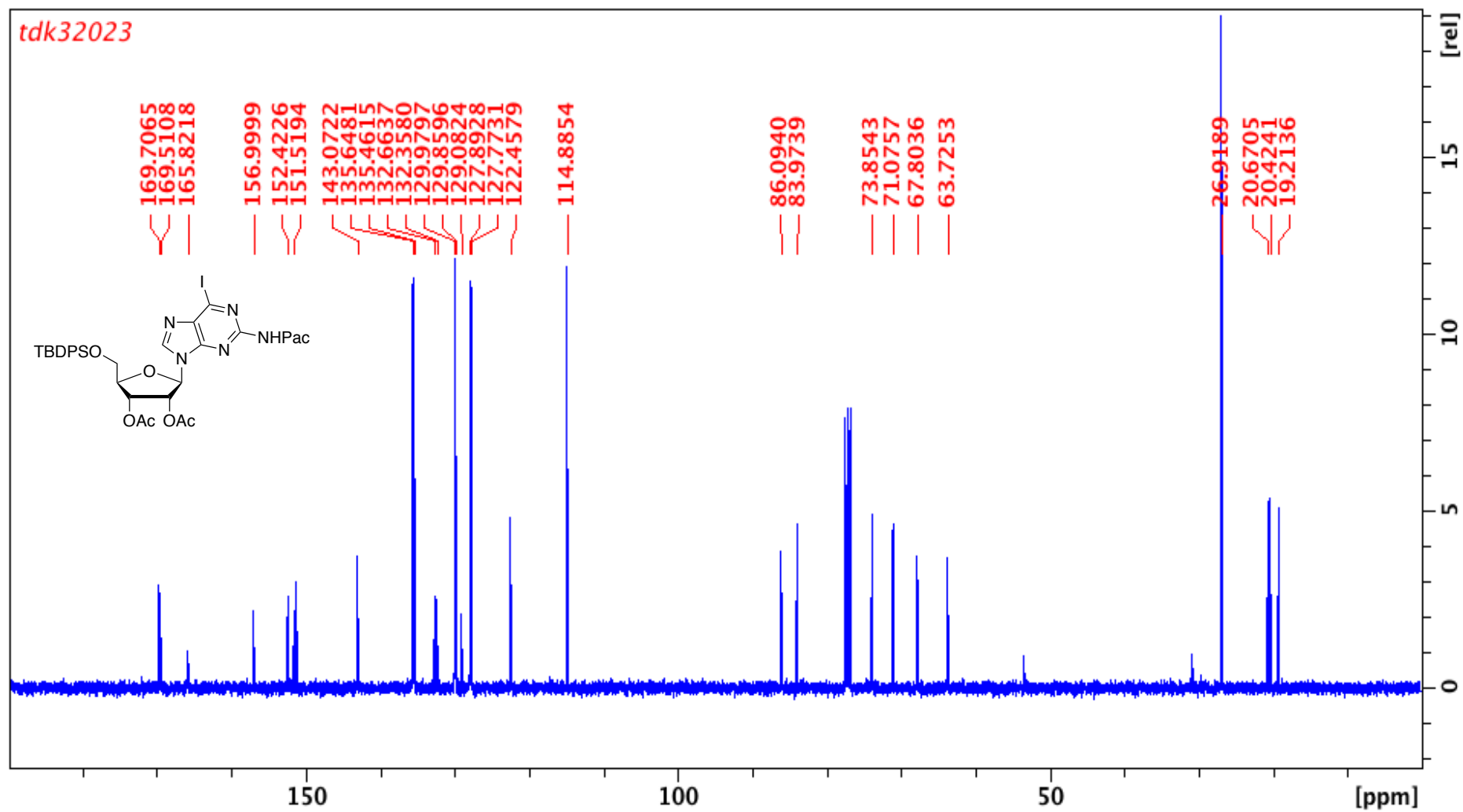
Appendix 108. $^1\text{H-NMR}$ spectrum of **225**. 400 MHz, CDCl_3 . Peak at 7.2 ppm comes from CHCl_3 .



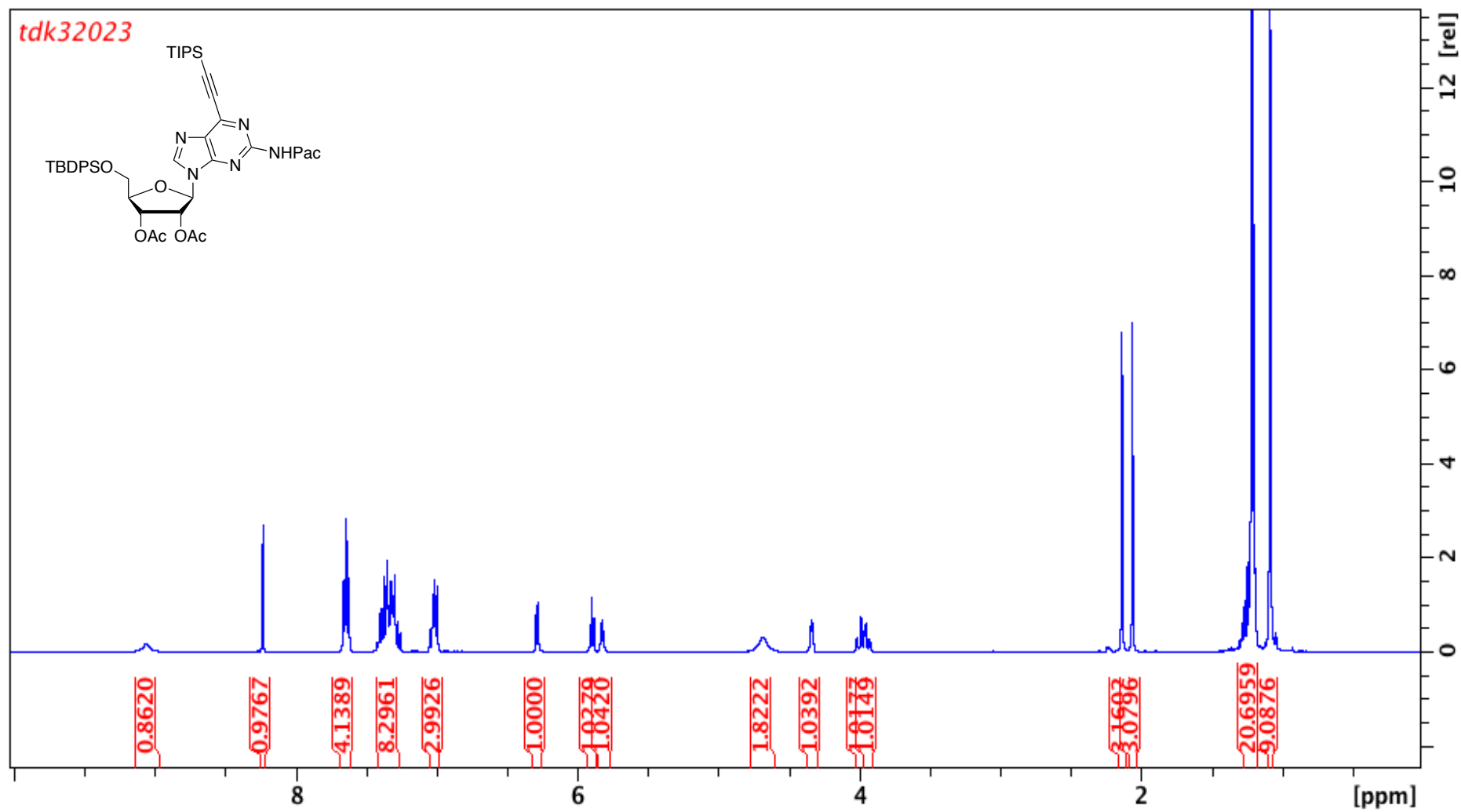
Appendix 109. ^{13}C -NMR spectrum of **225**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .

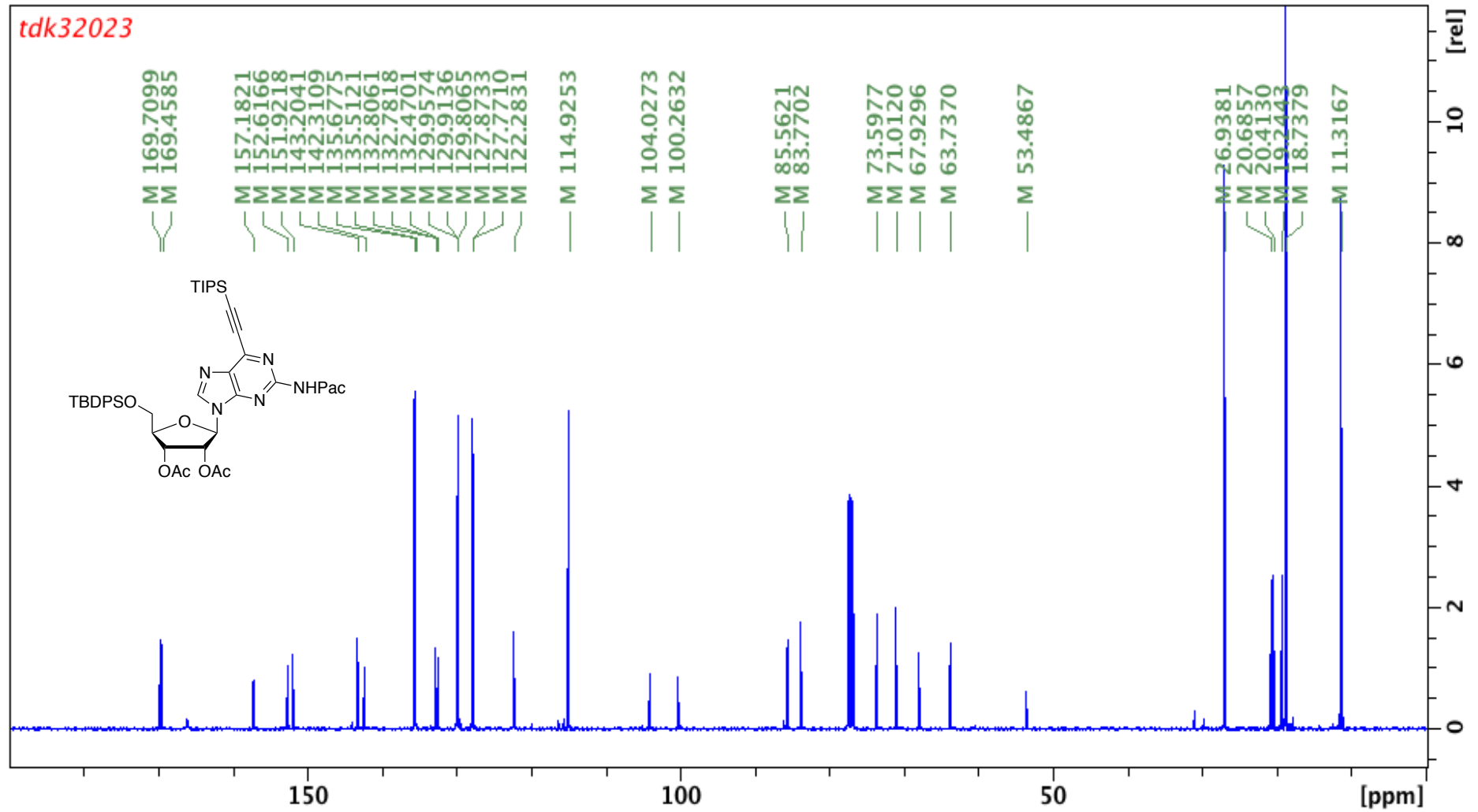


Appendix 110. $^1\text{H-NMR}$ spectrum of 226. 400 MHz, CDCl_3 . Peak at 5.3 ppm comes from DCM.

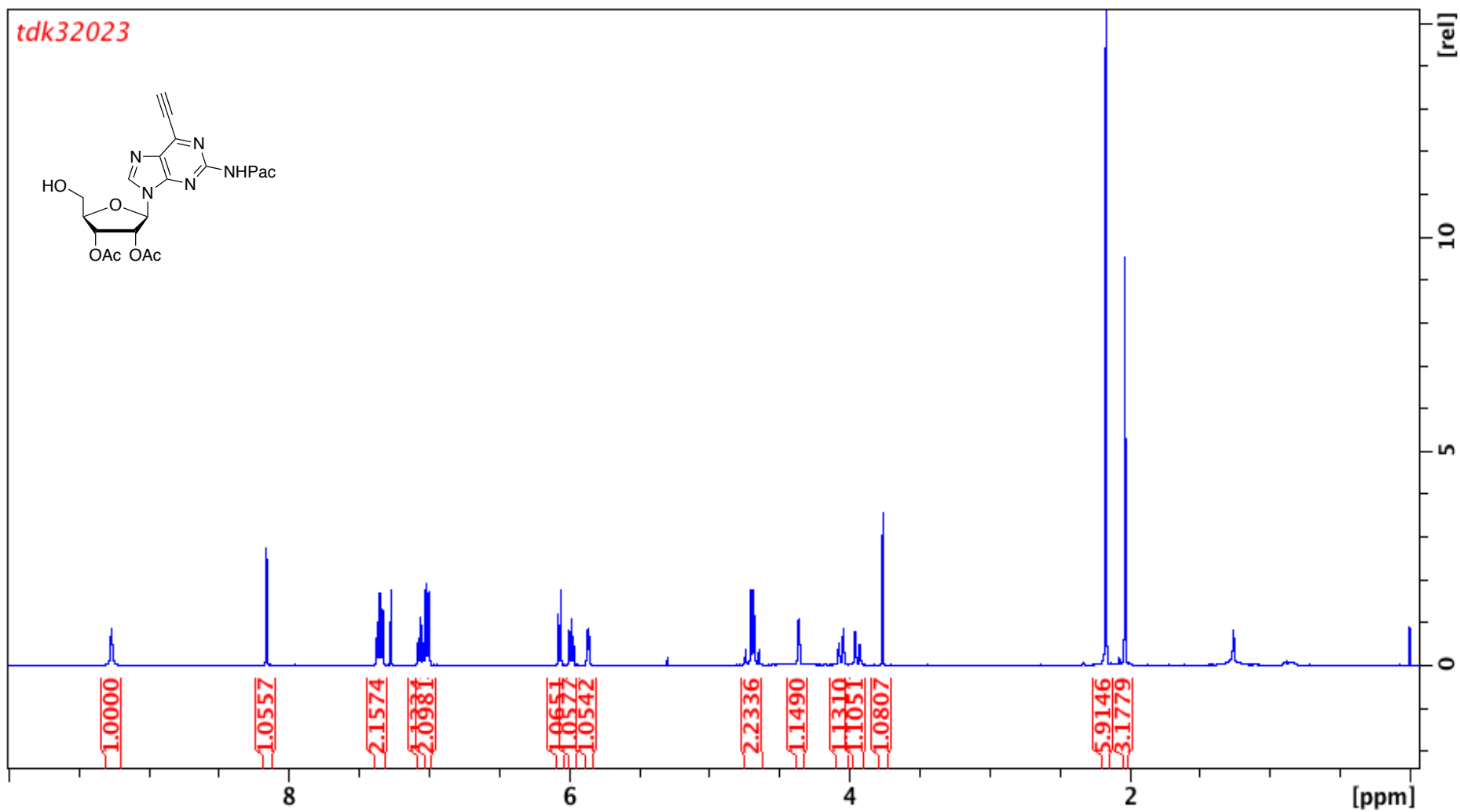


Appendix 111. ^{13}C -NMR spectrum of **226**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 53 ppm comes from DCM.

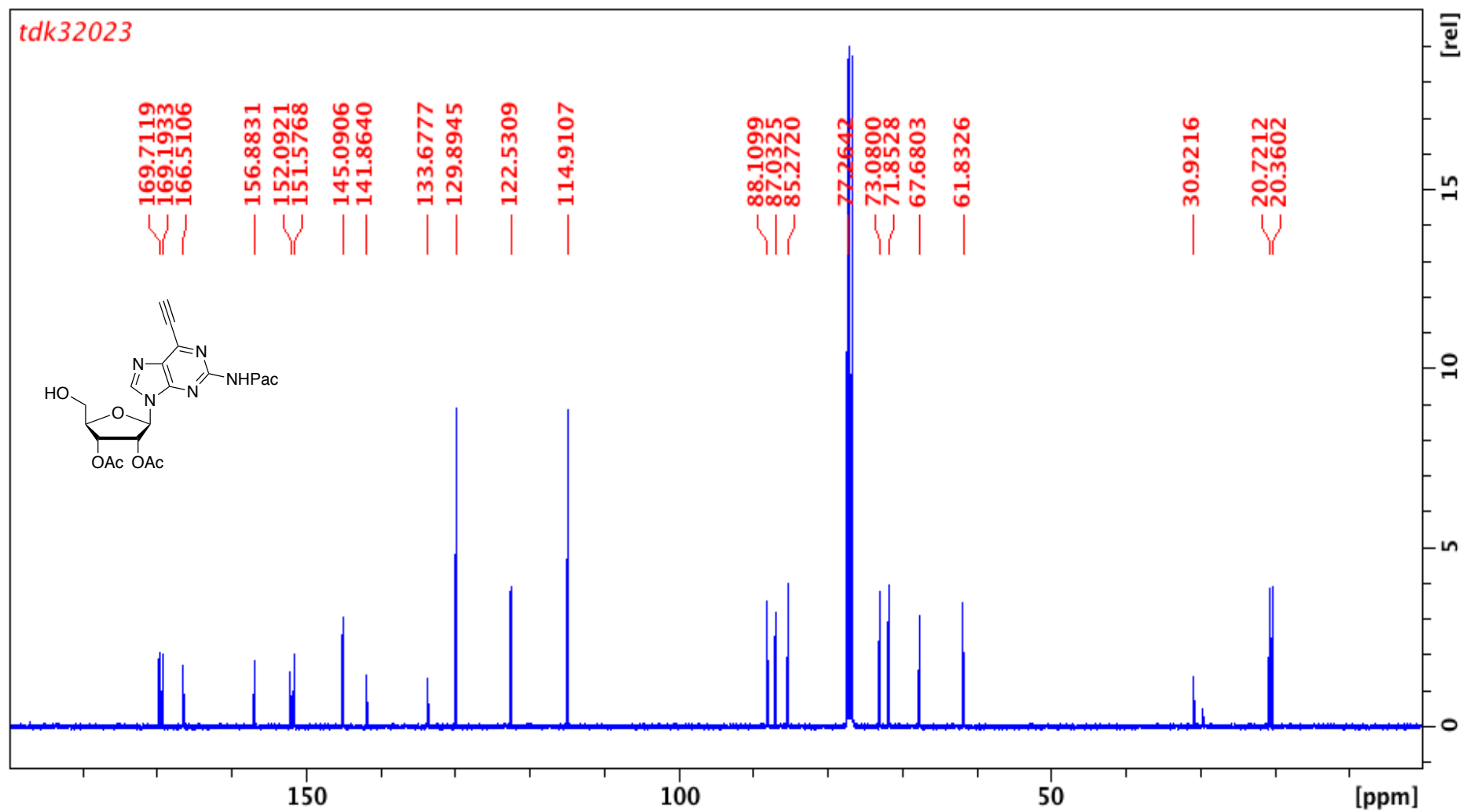
Appendix 112. ¹H-NMR spectrum of **227**. 400 MHz, CDCl₃.



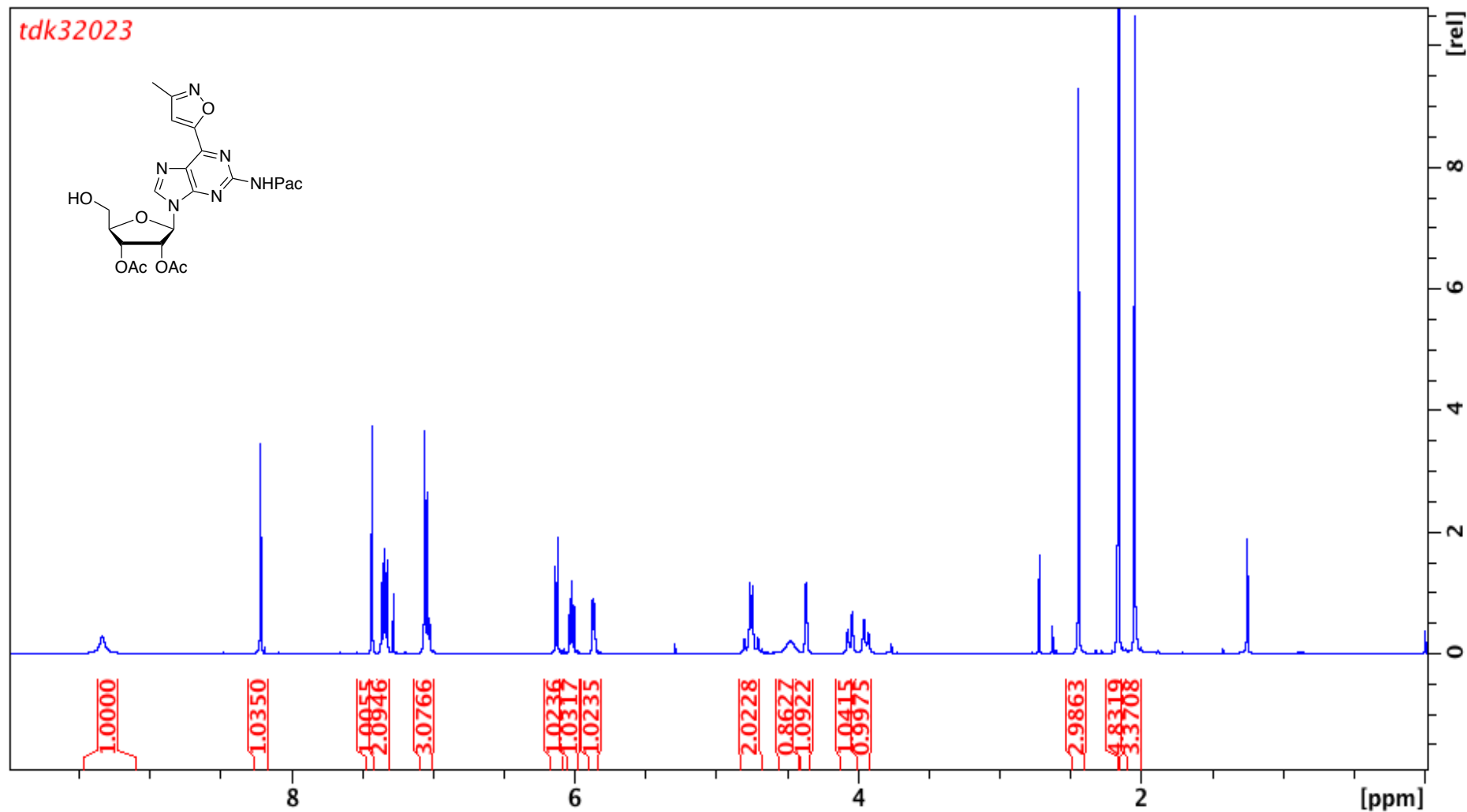
Appendix 113. ^{13}C -NMR spectrum of **227**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .



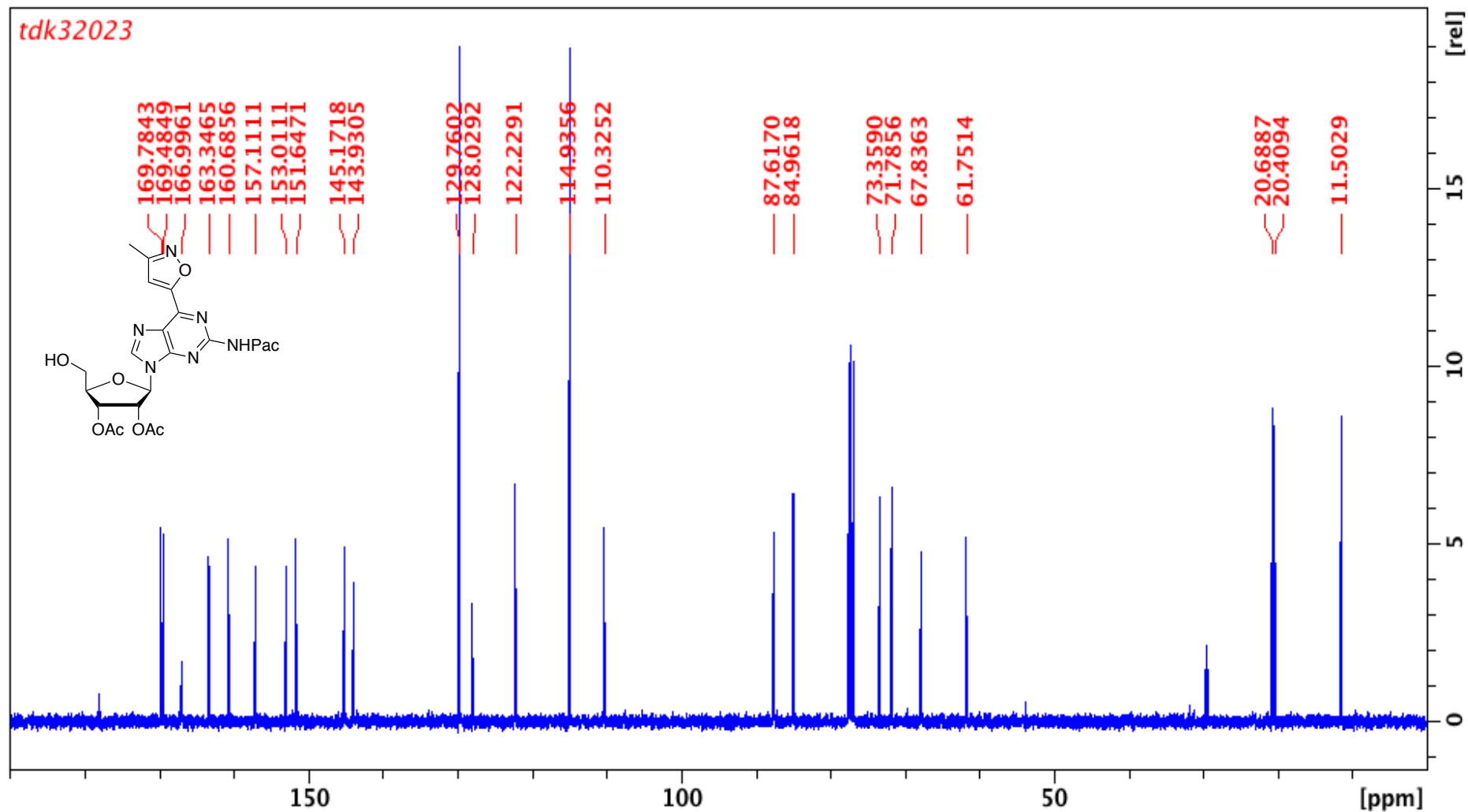
Appendix 114. $^1\text{H-NMR}$ spectrum of **210**. 400 MHz, CDCl_3 . Peak at 7.2 ppm comes from CHCl_3 .



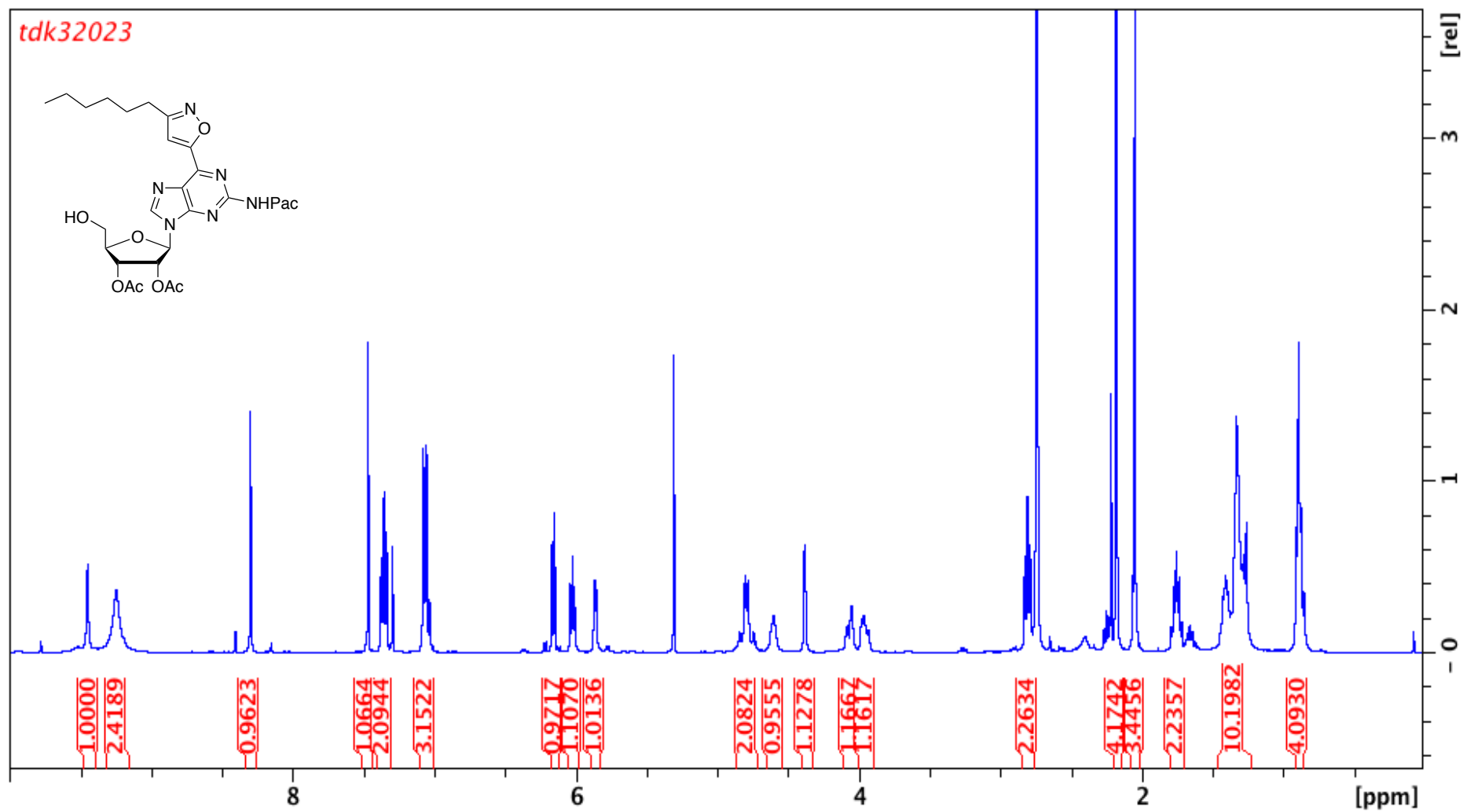
Appendix 115. ^{13}C -NMR spectrum of **210**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 .



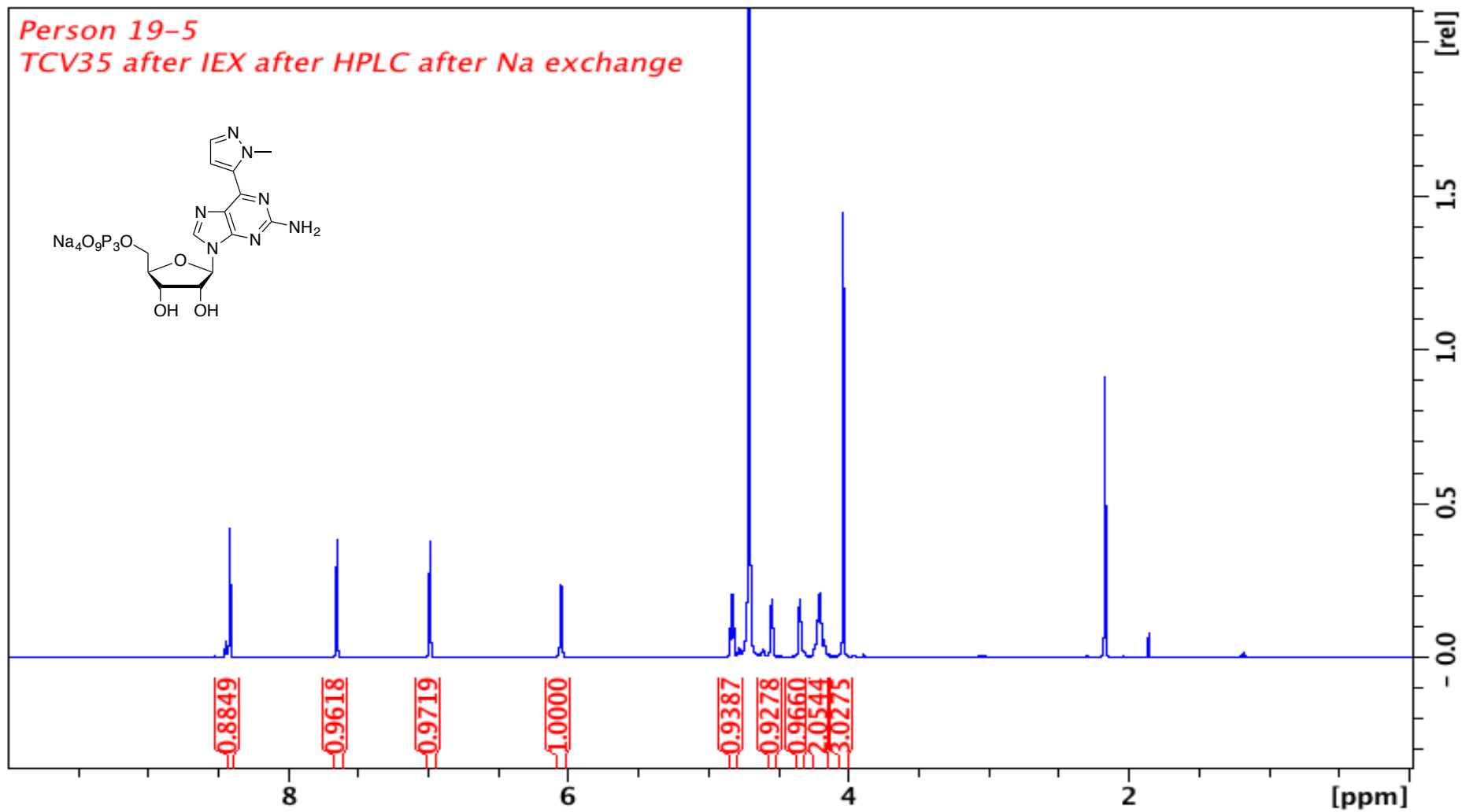
Appendix 116. $^1\text{H-NMR}$ spectrum of **228**. 400 MHz, CDCl_3 . Peak at 7.2 ppm comes from CHCl_3 . Peak at 2.7 ppm comes from succinimide. Peak at 1.26 ppm comes from grease.



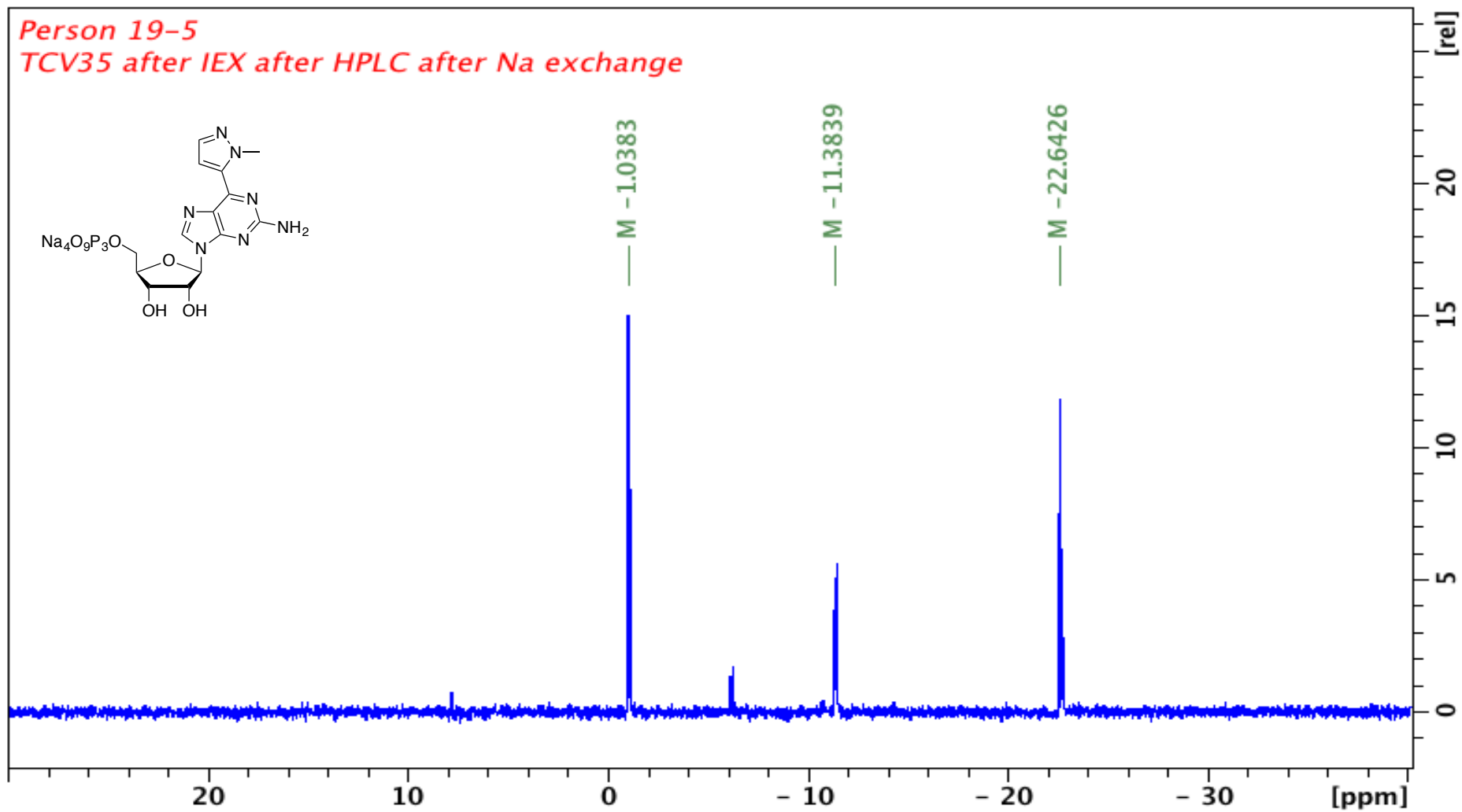
Appendix 117. ^{13}C -NMR spectrum of **228**. 100 MHz, CDCl_3 . Peak at 78 ppm comes from CHCl_3 . Peak at 30 and 177 ppm comes from succinimide.



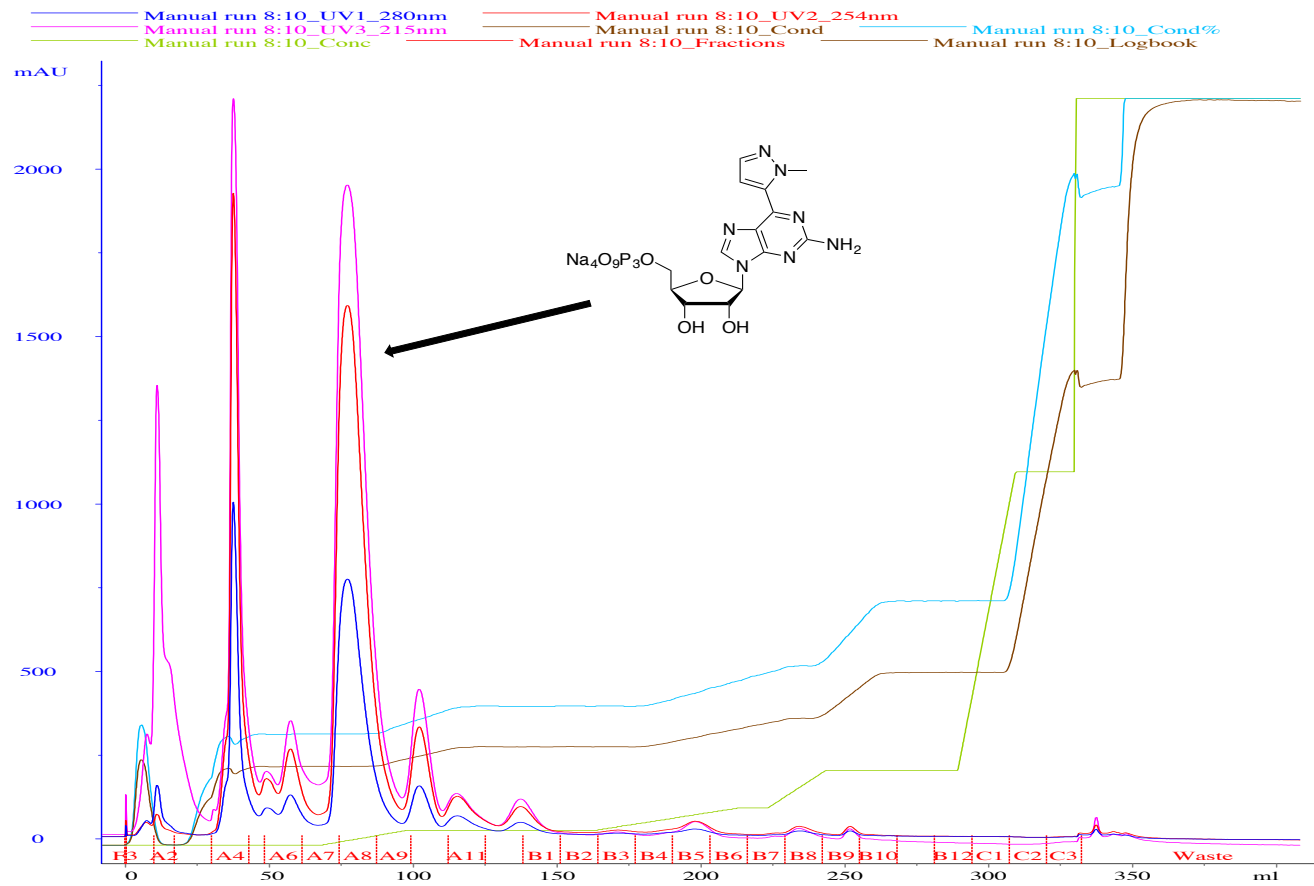
Appendix 118. $^1\text{H-NMR}$ spectrum of **229**. 400 MHz, CDCl_3 . Peak at 2.7 comes from succinimide. Peak at 5.3 comes from DCM.



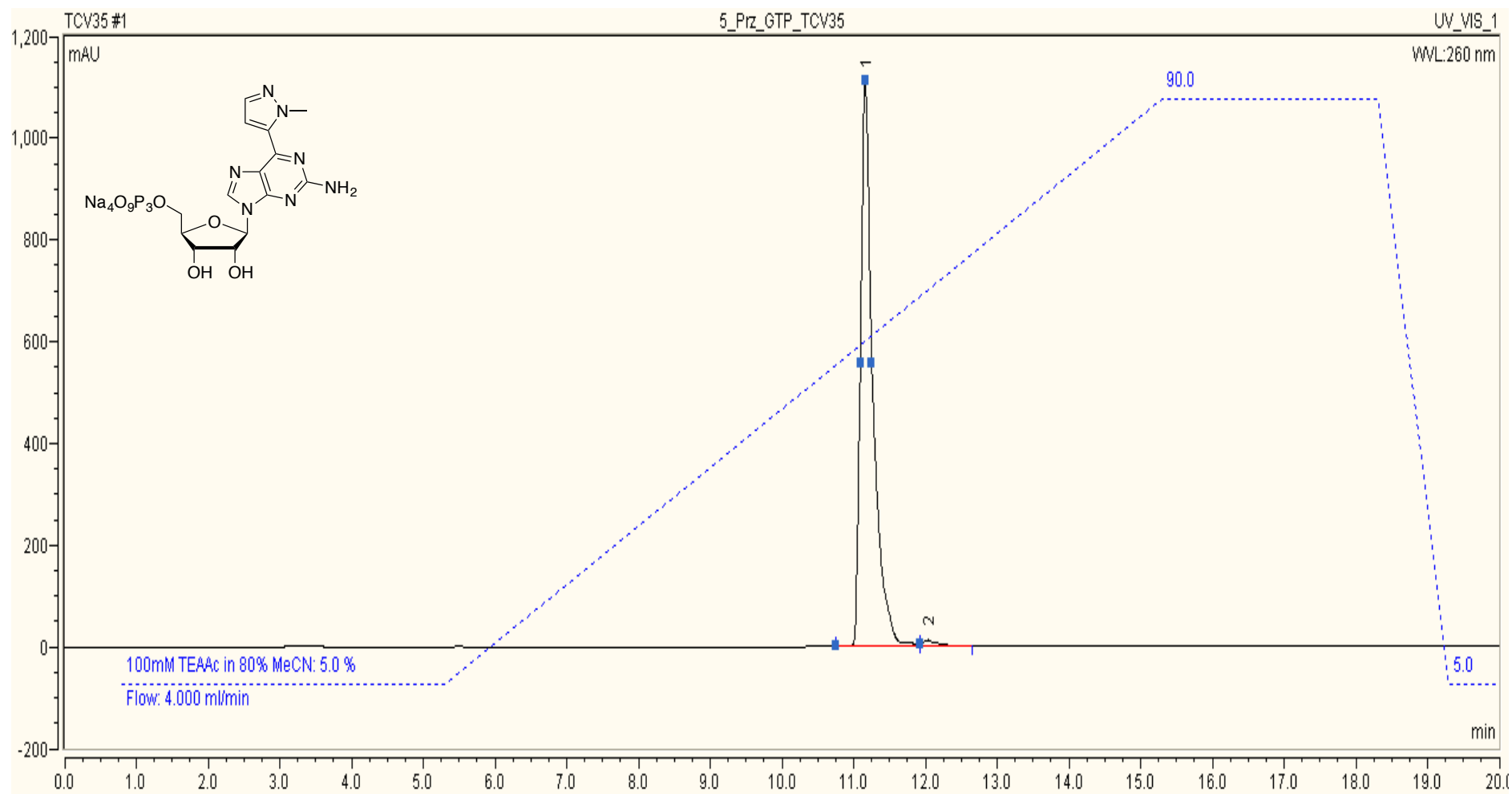
Appendix 119. ¹H-NMR spectrum of **201**. 500 MHz, D₂O. Peak at 4.79 ppm comes from water. Peak at 2.2 ppm comes from acetone.



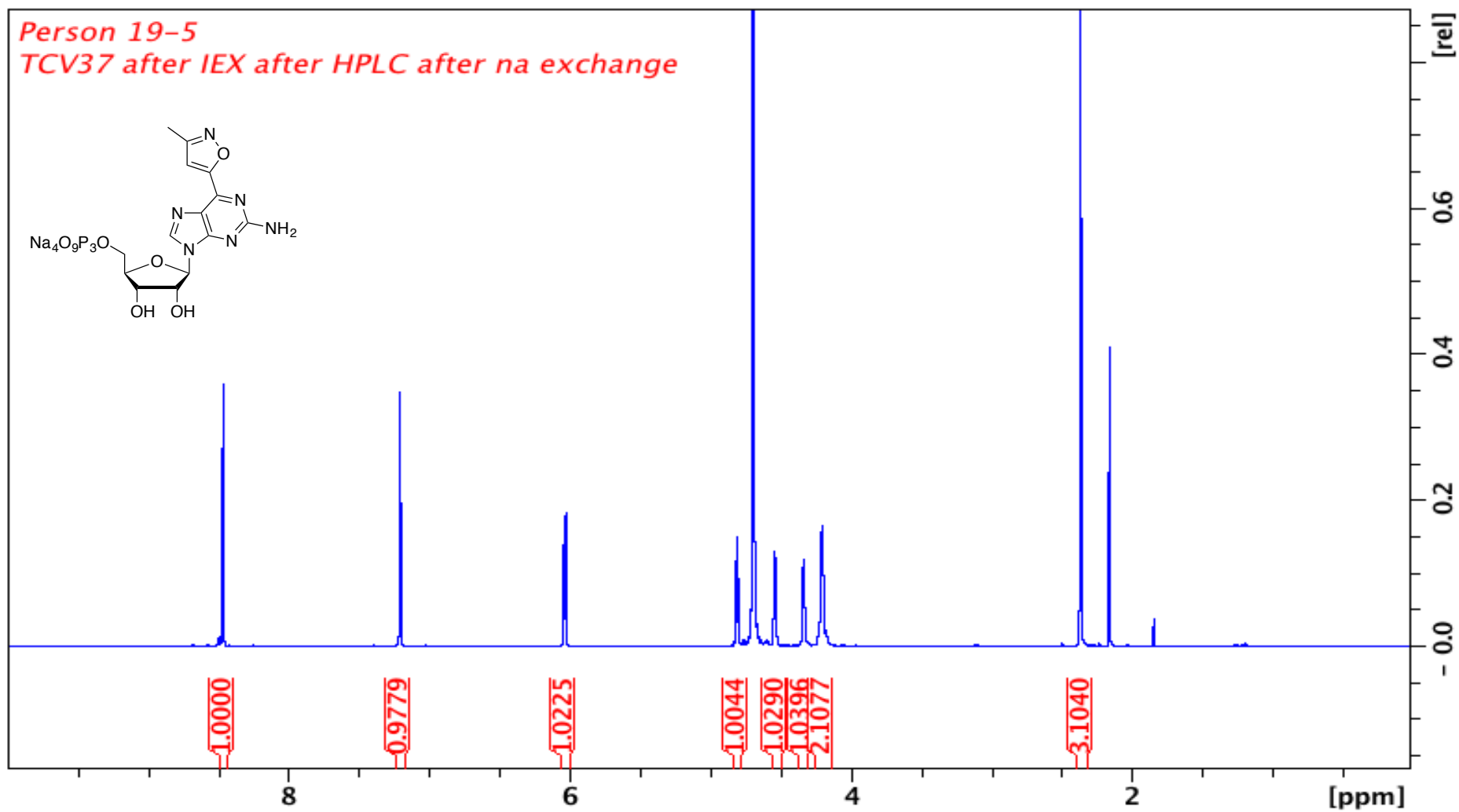
Appendix 120. ^{31}P -NMR spectrum of **201**. 202 MHz, D_2O . Peak at 6.0 ppm comes from unknown pyrophosphate related impurity.



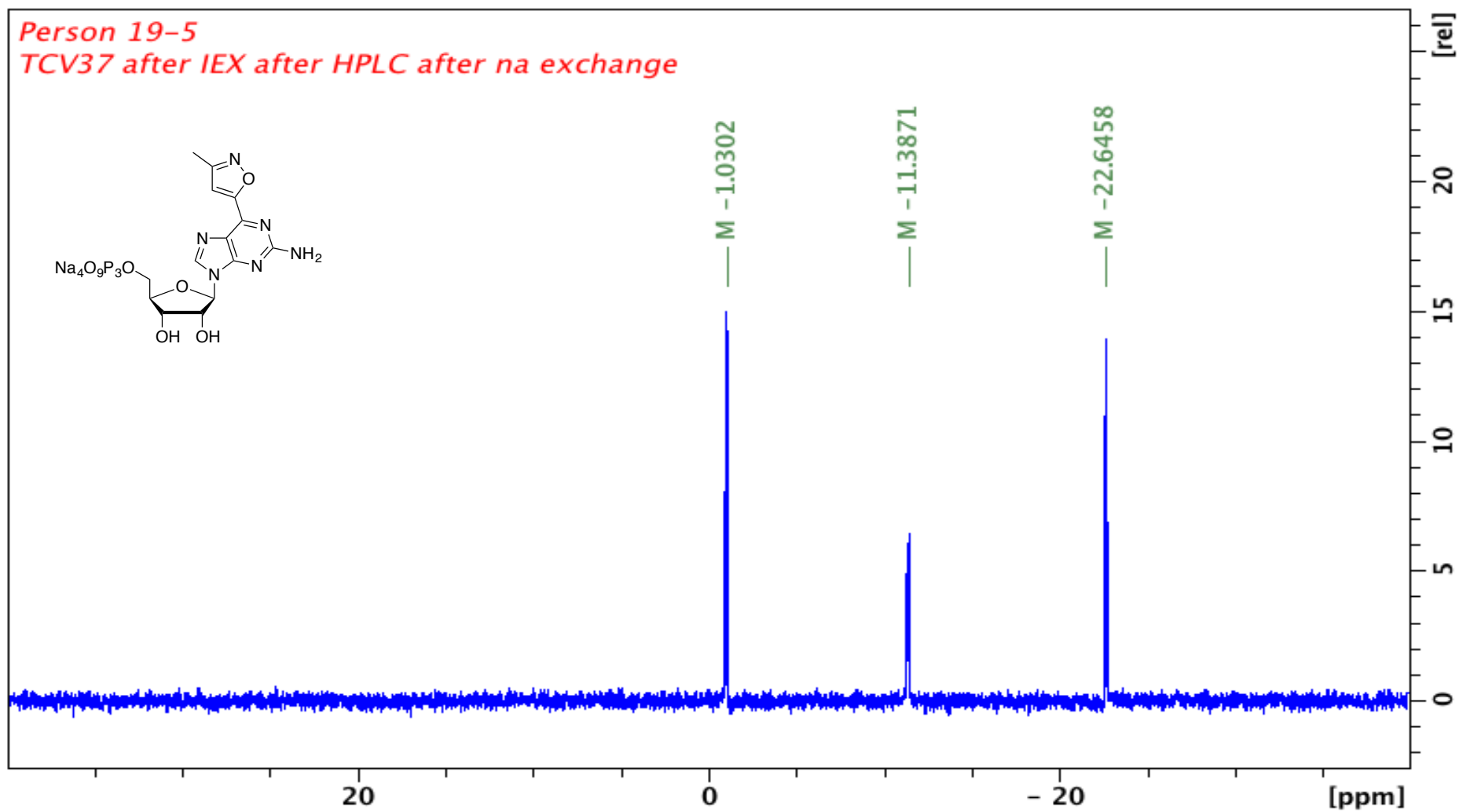
Appendix 121. Ion-Exchange chromatograph of triphosphate 201.



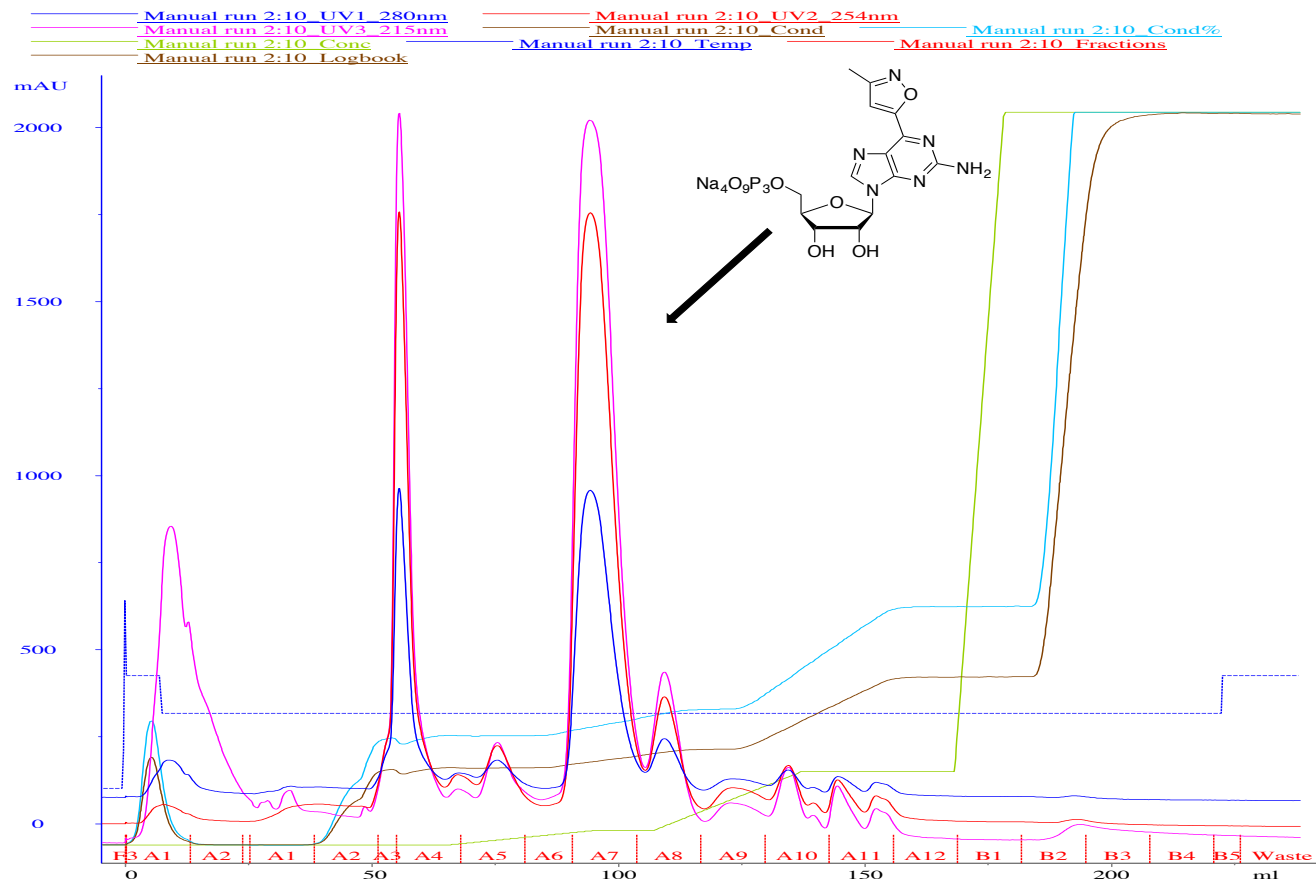
Appendix 122. Semi-preparative chromatograph of triphosphate **201**.



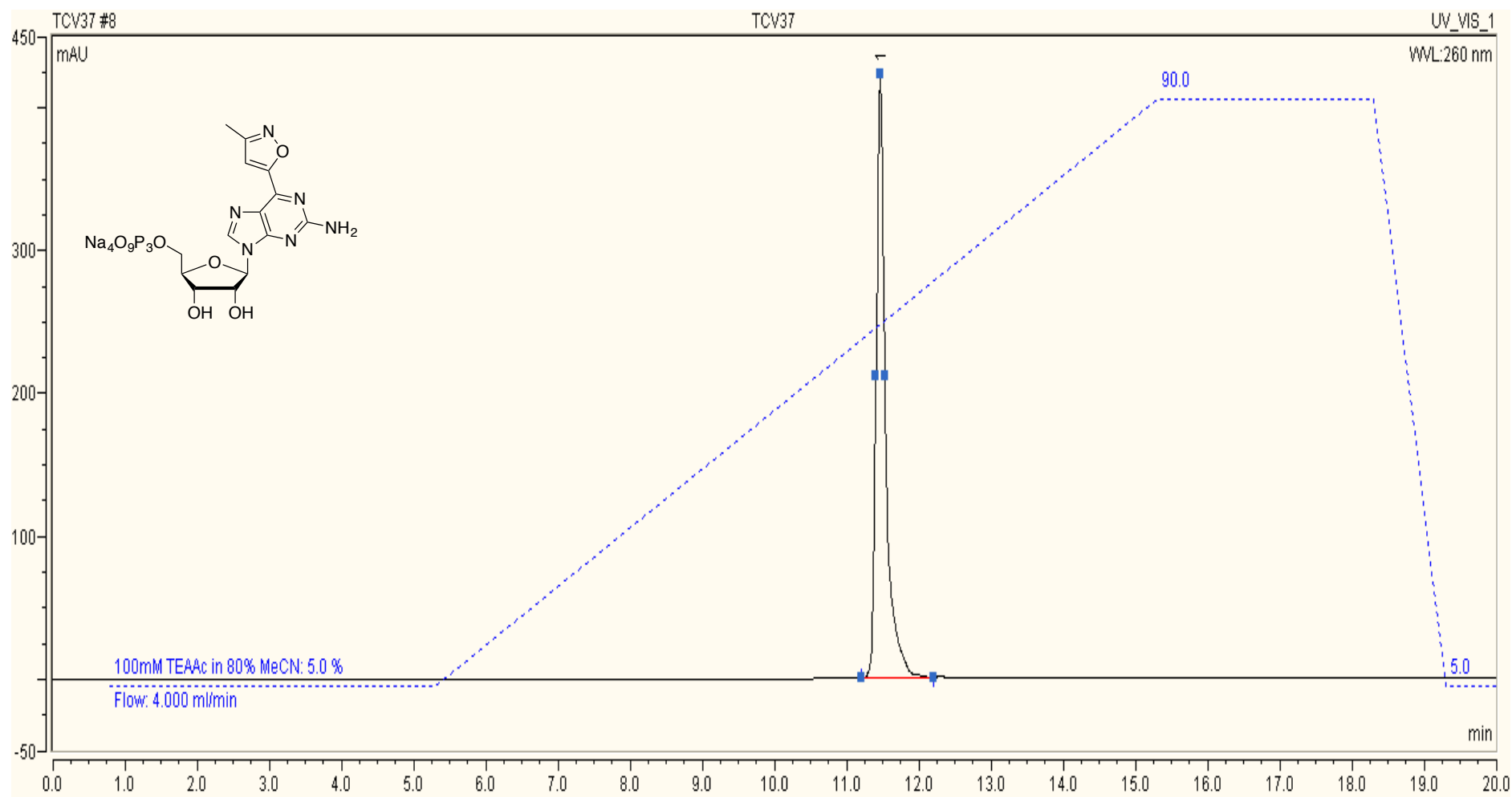
Appendix 123. $^1\text{H-NMR}$ spectrum of **203**. 500 MHz, D_2O . Peak at 4.79 ppm comes from water. Peak at 2.2 ppm comes from acetone.



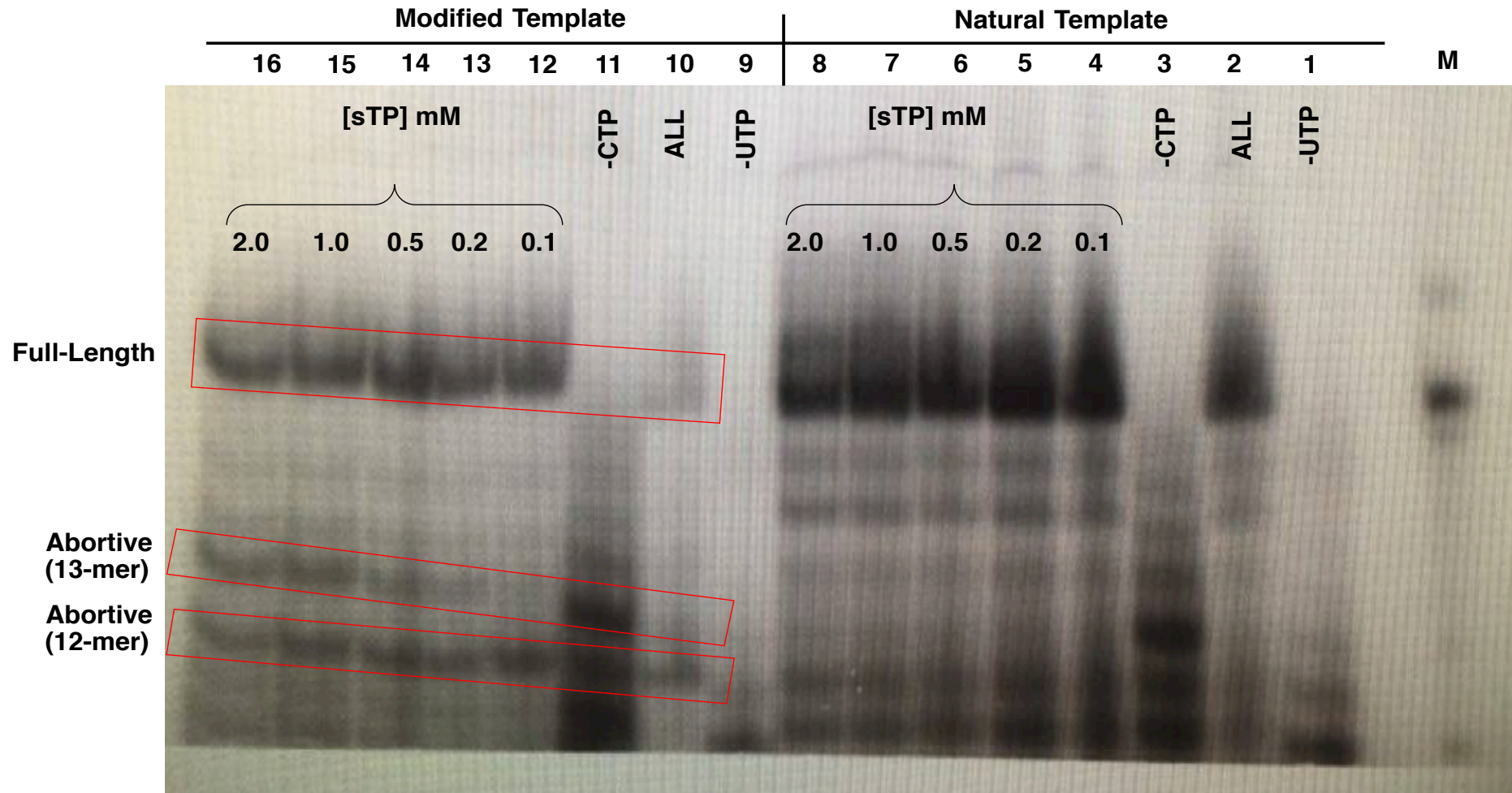
Appendix 124. ^{31}P -NMR spectrum of **203**. 202 MHz, D_2O .



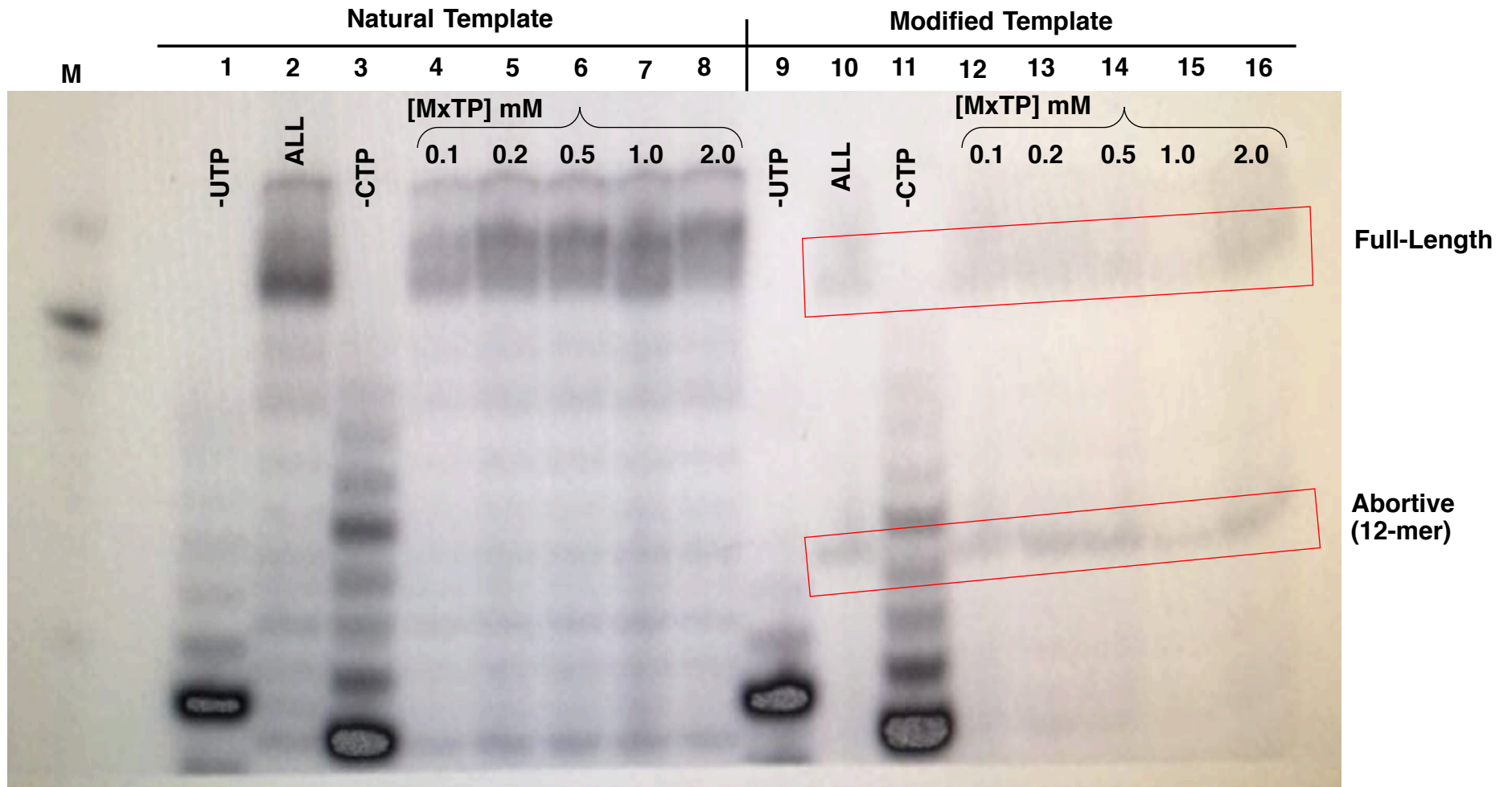
Appendix 125. Ion-Exchange chromatograph of triphosphate 203.



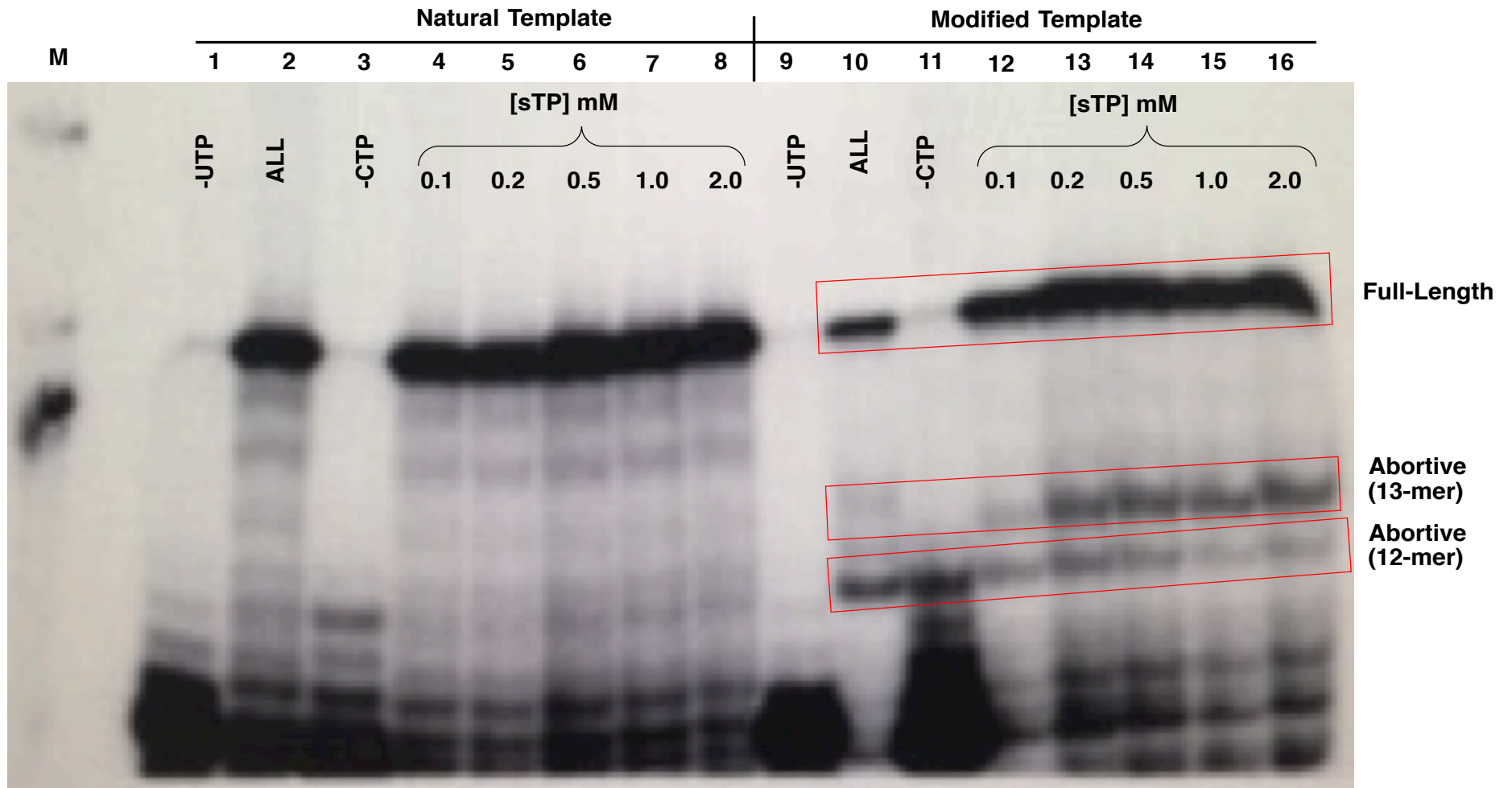
Appendix 126. Semi-preparative chromatograph of triphosphate **203**.



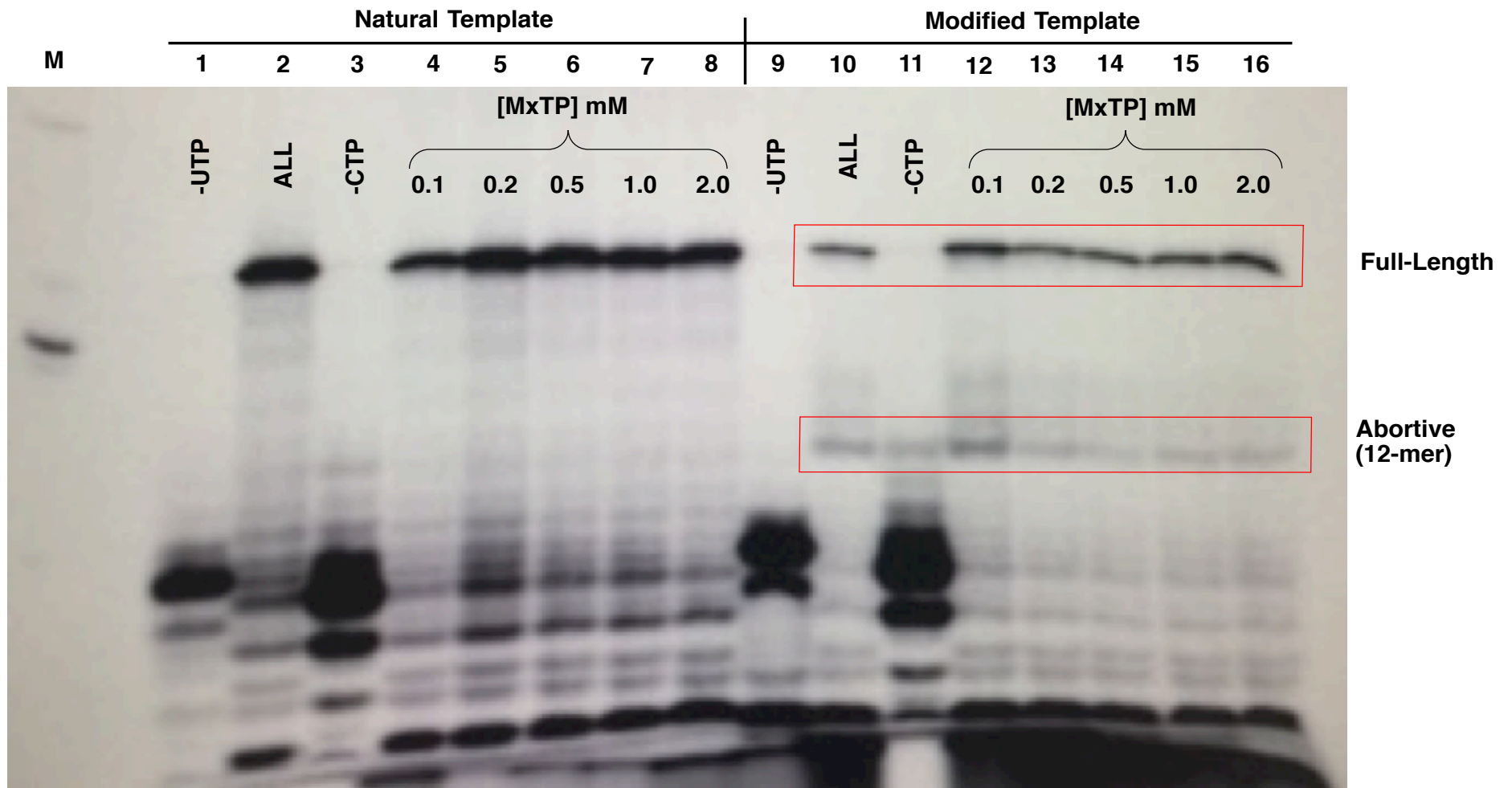
Appendix 127. PAGE image of reactions run with sTP, 0.12 mM NTPs. ALL = All natural NTPs. M = Marker.



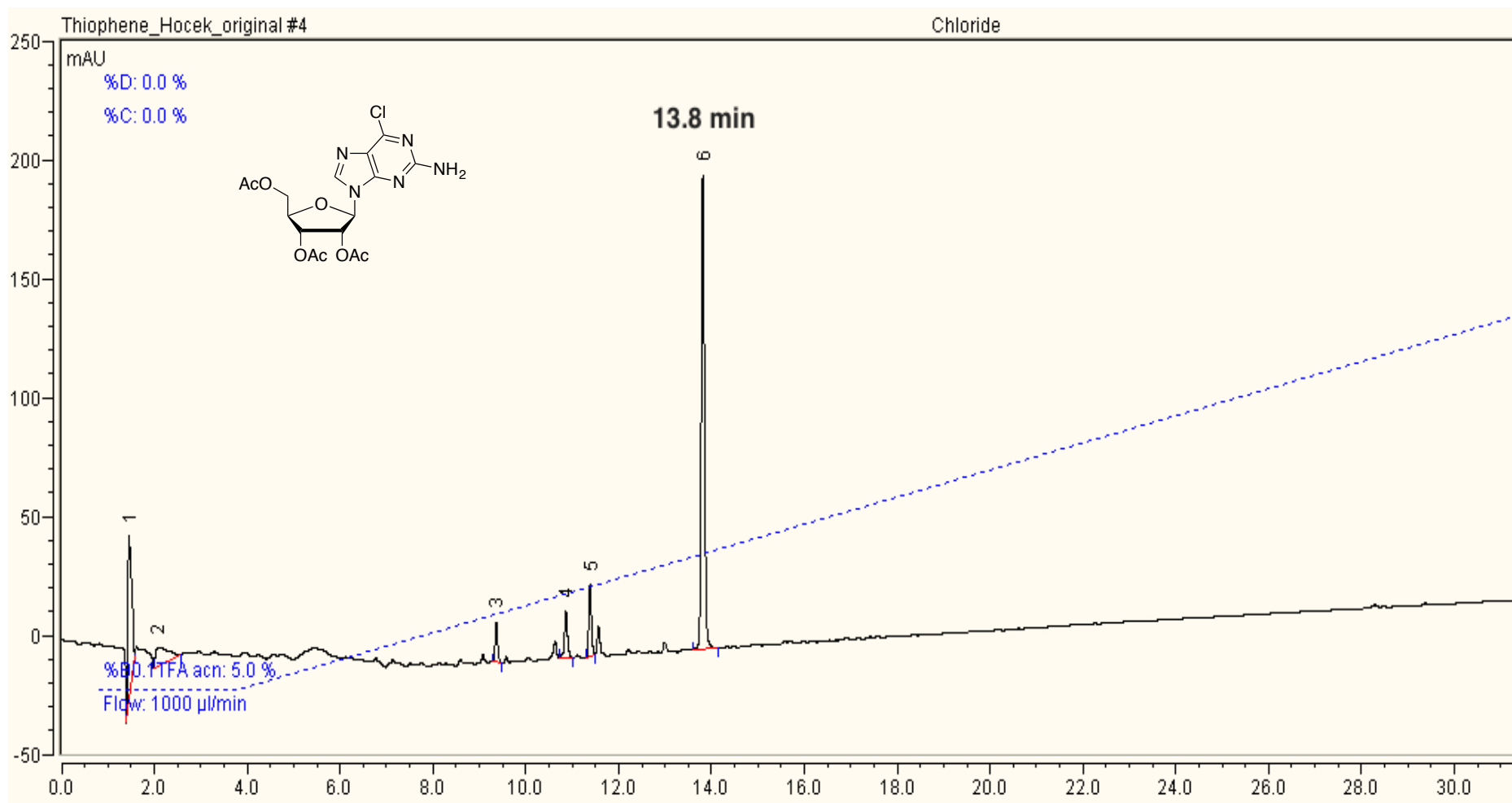
Appendix 128. PAGE image of reactions run with **MxTP**, 0.12 mM NTPs. ALL = All natural NTPs. M = Marker.



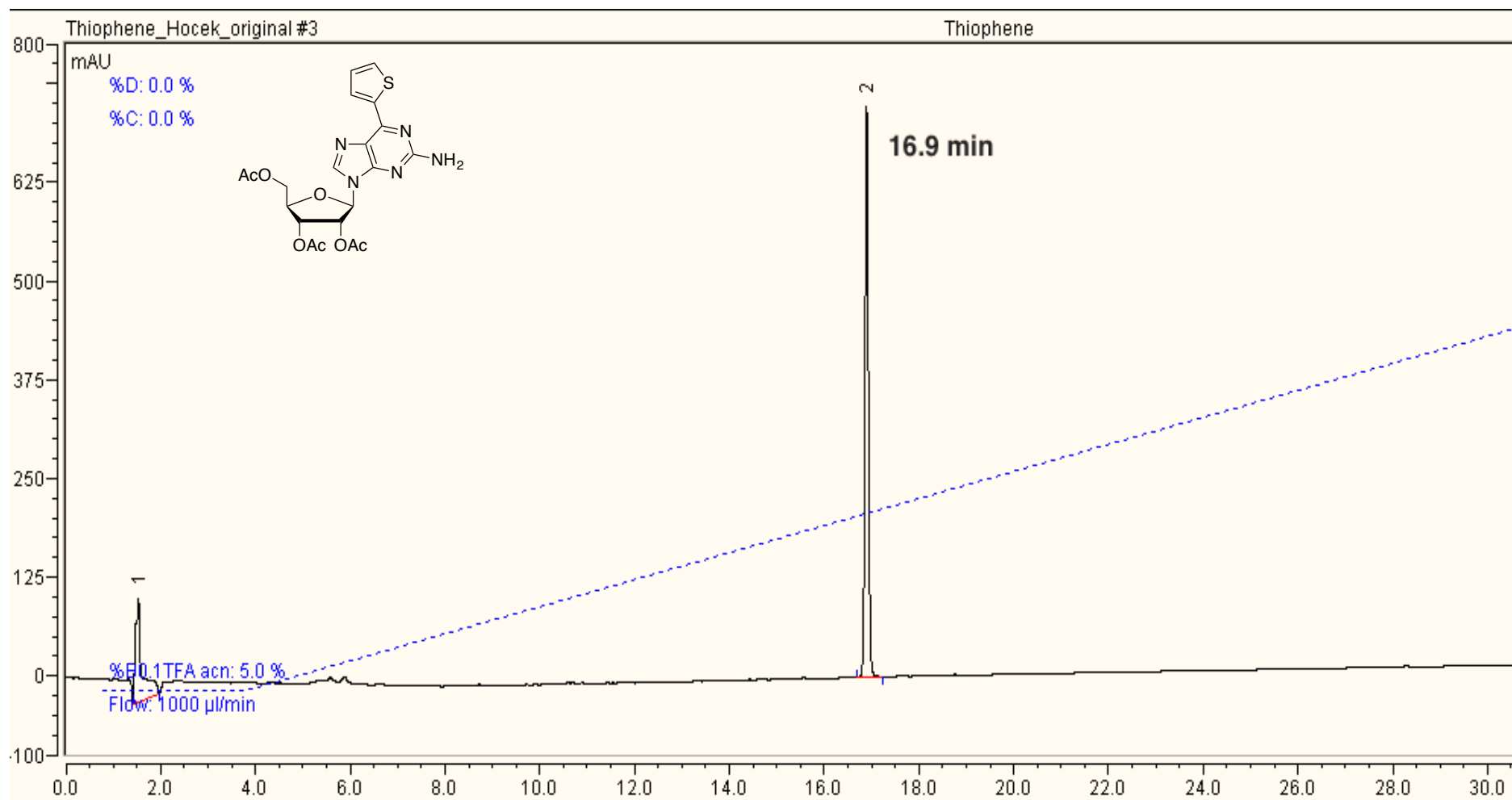
Appendix 129. PAGE image of reactions run with sTP, 0.5 mM MnCl₂, 0.12 mM NTPs. ALL = All natural NTPs. M = Marker.



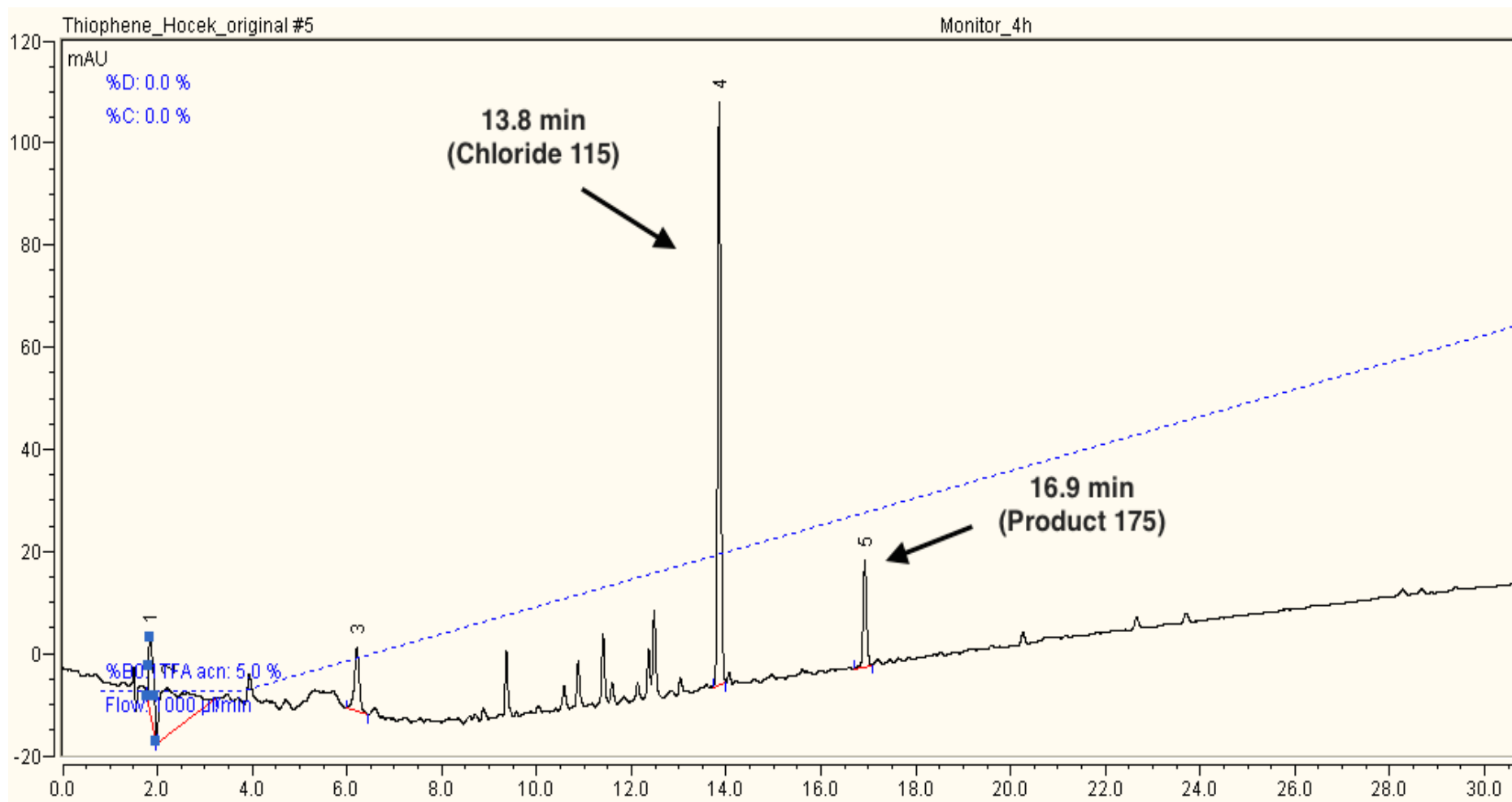
Appendix 130. PAGE image of reactions run with **MxTP**, 0.5 mM $MnCl_2$, 0.12 mM NTPs. ALL = All natural NTPs. M = Marker.



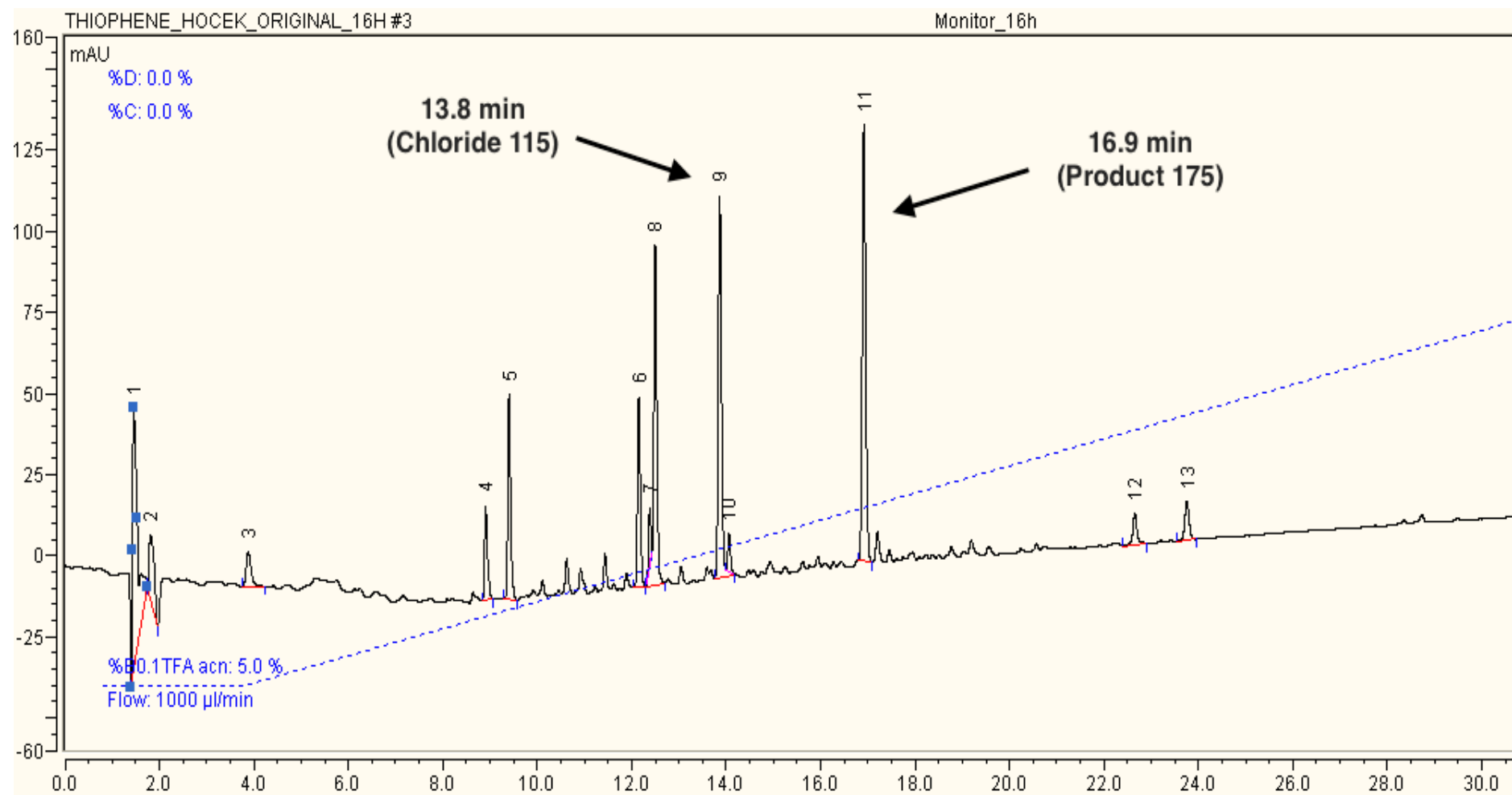
Appendix 131. Analytical HPLC chromatograph of chloride **115**.



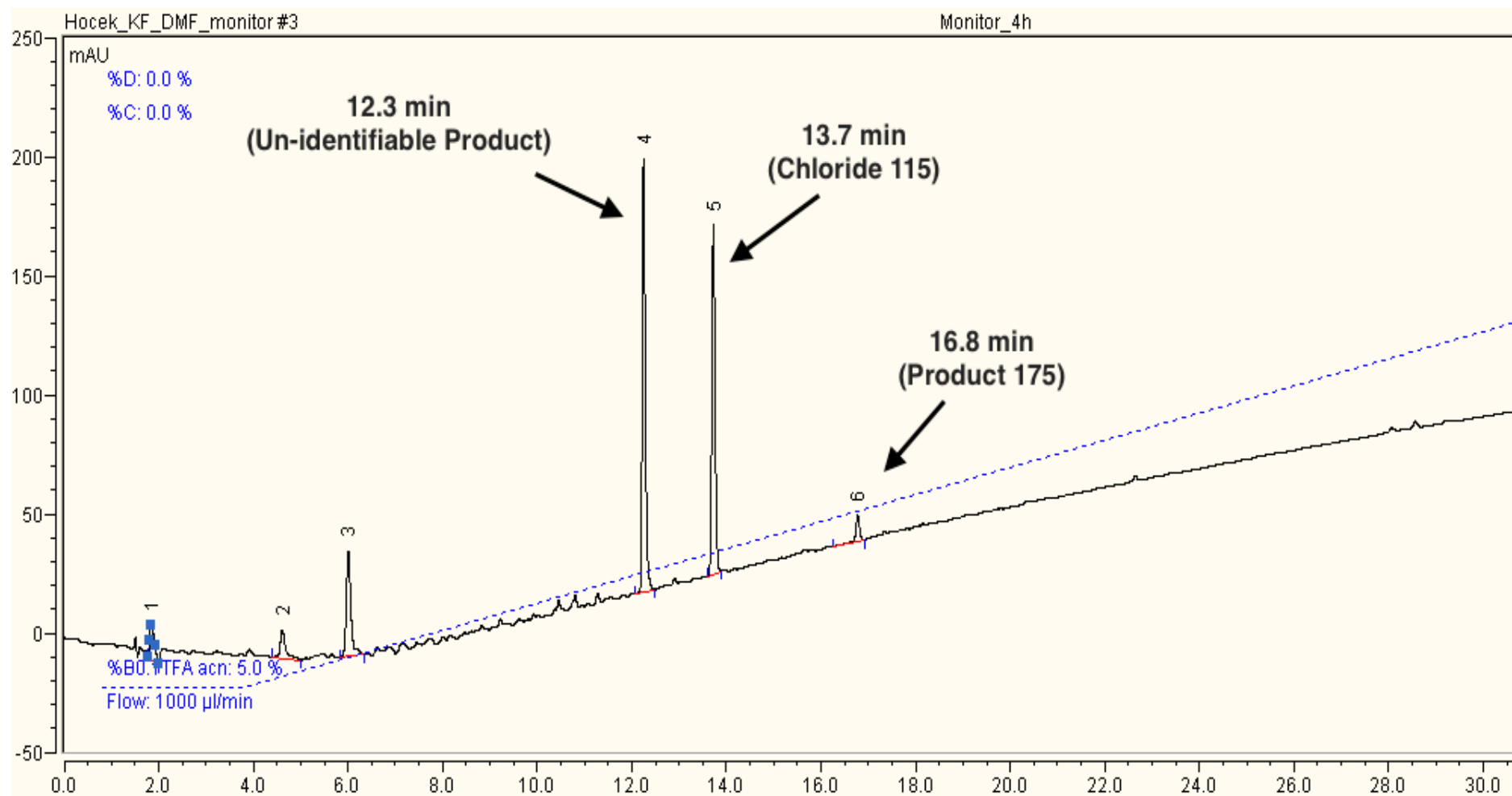
Appendix 132. Analytical HPLC chromatograph of thiophene 175.



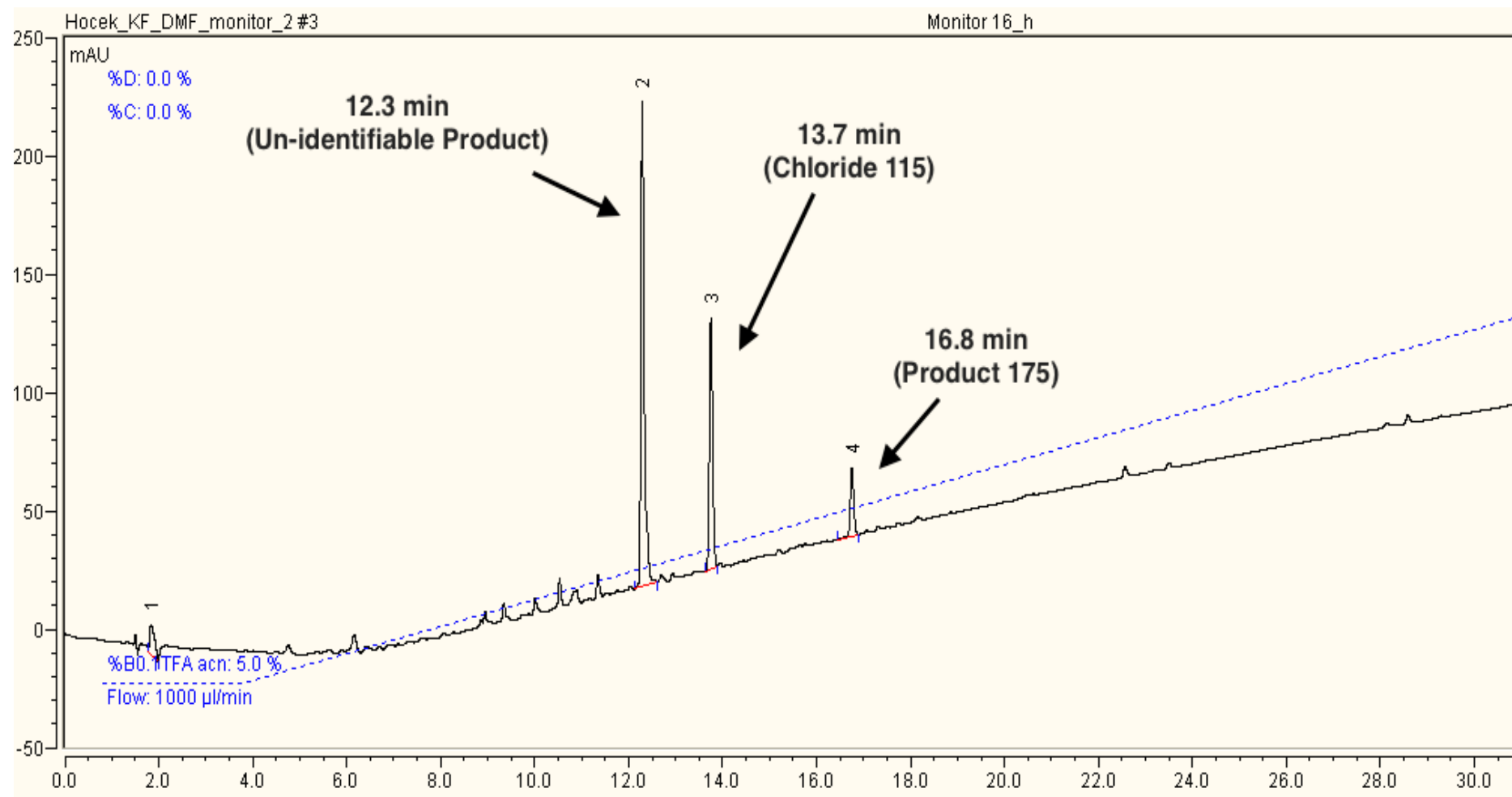
Appendix 133. Analytical HPLC chromatograph of the Suzuki reaction between **115** and **175**. Conditions are described in Table 3-2, Entry 1. Reaction time = 4 h.



Appendix 134. Analytical HPLC chromatograph of the Suzuki reaction between **115** and **175**. Conditions are described in Table 3-2, Entry 1. Reaction time = 16 h.



Appendix 135. Analytical HPLC chromatograph of the Suzuki reaction between **115** and **175**. Conditions are described in Table 3-2, Entry 2. Reaction time = 4 h.



Appendix 136. Analytical HPLC chromatograph of the Suzuki reaction between **115** and **175**. Conditions are described in Table 3-2, Entry 2. Reaction time = 16 h.